

Teradata Vantage™ - Advanced SQL Engine International Character Set Support

Release 17.10

July 2021

Copyright and Trademarks

Copyright © 2000 - 2021 by Teradata. All Rights Reserved.

All copyrights and trademarks used in Teradata documentation are the property of their respective owners. For more information, see [Trademark Information](#).

Product Safety

Safety type	Description
	Indicates a situation which, if not avoided, could result in damage to property, such as to equipment or data, but not related to personal injury.
	Indicates a hazardous situation which, if not avoided, could result in minor or moderate personal injury.
	Indicates a hazardous situation which, if not avoided, could result in death or serious personal injury.

Third-Party Materials

Non-Teradata (i.e., third-party) sites, documents or communications ("Third-party Materials") may be accessed or accessible (e.g., linked or posted) in or in connection with a Teradata site, document or communication. Such Third-party Materials are provided for your convenience only and do not imply any endorsement of any third party by Teradata or any endorsement of Teradata by such third party. Teradata is not responsible for the accuracy of any content contained within such Third-party Materials, which are provided on an "AS IS" basis by Teradata. Such third party is solely and directly responsible for its sites, documents and communications and any harm they may cause you or others.

Warranty Disclaimer

Except as may be provided in a separate written agreement with Teradata or required by applicable law, the information available from the Teradata Documentation website or contained in Teradata information products is provided on an "as-is" basis, without warranty of any kind, either express or implied, including the implied warranties of merchantability, fitness for a particular purpose, or noninfringement.

The information available from the Teradata Documentation website or contained in Teradata information products may contain references or cross-references to features, functions, products, or services that are not announced or available in your country. Such references do not imply that Teradata Corporation intends to announce such features, functions, products, or services in your country. Please consult your local Teradata Corporation representative for those features, functions, products, or services available in your country.

The information available from the Teradata Documentation website or contained in Teradata information products may be changed or updated by Teradata at any time without notice. Teradata may also make changes in the products or services described in this information at any time without notice.

Machine-Assisted Translation

Certain materials on this website have been translated using machine-assisted translation software/tools. Machine-assisted translations of any materials into languages other than English are intended solely as a convenience to the non-English-reading users and are not legally binding. Anybody relying on such information does so at his or her own risk. No automated translation is perfect nor is it intended to replace human translators. Teradata does not make any promises, assurances, or guarantees as to the accuracy of the machine-assisted translations provided. Teradata accepts no responsibility and shall not be liable for any damage or issues that may result from using such translations. Users are reminded to use the English contents.

Feedback

To maintain the quality of our products and services, e-mail your comments on the accuracy, clarity, organization, and value of this document to: docs@teradata.com.

Any comments or materials (collectively referred to as "Feedback") sent to Teradata Corporation will be deemed nonconfidential. Without any payment or other obligation of any kind and without any restriction of any kind, Teradata and its affiliates are hereby free to (1) reproduce, distribute, provide access to, publish, transmit, publicly display, publicly perform, and create derivative works of, the Feedback, (2) use any ideas, concepts, know-how, and techniques contained in such Feedback for any purpose whatsoever, including developing, manufacturing, and marketing products and services incorporating the Feedback, and (3) authorize others to do any or all of the above.

Contents

Chapter 1: Introduction to International Character Set Support	7
Overview	7
Chapter 2: Managing International Language Support	8
Overview	8
Implementation Overview	8
About Character Sets	9
System Default Character Support	9
About Object Names	12
Chapter 3: Client Character Set Options	17
Overview	17
About Client Character Set Translation Codes	17
Teradata Support for Common Client Character Sets	18
ASCII Client Character Set Support	21
EBCDIC Client Character Set Support	22
UTF8 Client Character Set Support	22
UTF16 Client Character Set Support	24
Japanese Client Character Set Support	24
Japanese Encoding Schemes	25
Teradata Supported Japanese Character Sets	26
IBM-Compatible Japanese Character Sets	26
KANJI EBCDIC5026_0I	30
KANJI EBCDIC5035_0I	31
KATAKANA EBCDIC	31
UNIX Compatible Japanese Character Set (KANJI EUC_0U)	32
Windows-Compatible Japanese Character Sets	34
KANJI SJIS_0S	34
KANJI 932_1S0	35
Chinese Character Sets	37
SCHEBCDIC935_2IJ	38
TCHEBCDIC937_3IB	39
SCHGB2312_1T0	40
TCHBIG5_1R0	40
SCHINESE936_6R0	41
TCHINESE950_8R0	42
Korean Character Sets	43
HANGUL EBCDIC933_1II	44
HANGUL KSC5601_2R4	45

HANGUL949_7R0	46
Single-Byte International Character Sets	46
Client Character Information in System Tables	47
Accessing Client Character Set Information	48
Chapter 4: Using Standard Teradata Client Character Sets	49
Overview	49
Implementation Process	49
About the DBC.CharTranslationsV View	49
Enabling a Client Character Set	51
About the Default Client Character Set	51
System Determination of Client Character Set	52
Assigning the Default Client Character Set During Client Setup	52
Assigning the Default Client Character Set by HostId	53
Assigning the Default Client Character Set for a Session	56
Chapter 5: Server Character Sets	57
About Character Data Storage	57
Working with Server Character Set Defaults	57
LATIN Server Character Set	58
UNICODE Server Character Set	60
KANJI1 Character Set	63
KANJIISJIS Character Set	66
GRAPHIC Character Set	67
Chapter 6: Site-Defined Client Character Sets	69
Overview	69
About Custom Single-Byte Client Character Sets	69
Implementing Custom Character Sets	70
Creating a Custom Client Character Set	70
Enabling a Custom Client Character Set	72
Chapter 7: Extended Site-Defined Client Character Sets	79
Overview	79
About Extended Character Sets	79
Implementing Extended Character Sets	80
Naming Character Sets	81
Mapping File for a Single-Byte Character Set	84
Mapping File for a Multibyte Character Set	85
Example Mapping File	88
Chapter 8: Collation Sequences	105
Overview	105
About Collation Sequences	105
ASCII and EBCDIC Collations	107

CHARSET_COLL Collation	108
JIS_COLL Collation	109
MULTINATIONAL Collation	110
Changing the Standard Multinational Default Collation	113
Defining Your Own Collation Sequence	117
MULTINATIONAL Collation for Extended Site-Defined Character Sets	122
Chapter 9: Character Conversion	124
Overview	124
Character Conversions	124
Hard-Coded and Table Conversions	124
Application Model With Single Form-of-Use	126
Working with Export Widths	128
Exception Handling	135
Illegal Form-of-Use	136
E2I Conversion Exceptions	137
I2E Conversion Exceptions	137
Error Characters	138
Truncation Exceptions	141
Truncation Exceptions in ANSI Mode	142
Truncation Exceptions in Teradata Mode	143
Truncation Examples	143
Unwanted Export Truncation	144
Chapter 10: Unicode Pass Through	147
Unicode Pass Through Overview	147
Enabling and Disabling Unicode Pass Through	148
Unicode Pass Through Functionality	149
Importing Unicode Characters in a Pass Through Session	152
Exporting From a UNICODE Server Character Set	152
Ill-formed Code Unit Sequences	152
Usage Notes	153
UPT Restrictions	156
Appendix A: Character Shorthand Notation	158
Appendix B: Japanese Encodings and Mapping Standards	160
Appendix C: ASCII Mappings	173
Appendix D: EBCDIC Mappings	192
Appendix E: Mappings for String Functions	211
Appendix F: KanjiShiftJIS Mappings	347

Appendix G: Character Set Files 370

Appendix H: Additional Information 374

Introduction to International Character Set Support

Overview

Teradata Vantage™ is our flagship analytic platform offering, which evolved from our industry-leading Teradata® Database. Until references in content are updated to reflect this change, the term Teradata Database is synonymous with Teradata Vantage.

Advanced SQL Engine is a core capability of Teradata Vantage, based on our best-in-class Teradata Database. Advanced SQL refers to the ability to run advanced analytic functions beyond that of standard SQL.

This document explains how to use international character sets with the Teradata Database and includes information on:

- Client and server character sets
- Site-defined and extended site-defined client character sets
- Collation sequences
- Teradata Database character mapping

Managing International Language Support

Overview

Teradata Database supports the use of a variety of international character sets, which allows:

- Users to load and retrieve data in their local character set.
- Data to be exchanged through the database between different client character formats.

You must perform various setup tasks to enable and use character set options.

Implementation Overview

Tasks 1 through 3 are required on all systems. Other tasks are optional, based on system needs.

1. Make sure you understand how the Teradata Database system handles character data. See [About Character Sets](#) and [About Object Names](#).
2. Determine whether you can use system character set defaults or if you need to configure character set options. See [System Default Character Support](#).
3. Review the object naming topic to understand the rules that determine which characters are valid in object names for your system. See [About Object Names](#).
4. Select and enable the client character sets that you need:
 - For information about the client character sets for which standard Teradata translation codes are available, see [Client Character Set Options](#)
 - For information about enabling standard Teradata client character sets, see [Using Standard Teradata Client Character Sets](#)
 - For information about database (server) character set defaults and options, including setup and usage instructions, see [Server Character Sets](#)
5. If none of the Teradata-standard client character sets meets site needs, you can define a custom (site-defined) client character set:
 - To create a simple variation of a Teradata standard client character set, see [Site-Defined Client Character Sets](#)
 - To create an extended client character set with custom character mapping, see [Extended Site-Defined Client Character Sets](#)
6. For information on default collation sequences and how to define and implement custom collation sequences, see [Collation Sequences](#)
7. For detailed information on character conversion between client character sets and the server character set, including the option to define custom export widths for use during conversion to the client character set, see [Character Conversion](#)

About Character Sets

There are two categories of character sets in a Teradata Database environment.

Category	Description
Client (session) character set	<ul style="list-style-type: none"> Used by a client to log on and make requests to the database. Used by the database for data returned to the client.
Server or internal (database) character set	Used by the database to store user data.

System Processing of Character Data

The system processes the character data in client requests to the database as follows:

1. A user submits a request in the character set specified as the client default.
2. The database converts the client characters it receives to the default server character set (external-to-internal, or E2I conversion) for processing and data storage.
3. The database converts data from the internal database character set back to the client character set (internal-to-external, or I2E conversion) when returning answers sets to the requesting client.

For information on character conversion options, see [Character Conversion](#).

System Default Character Support

Certain character set options are enabled by default during initial system setup.

Default Client Character Set Support

Teradata clients must use a character set for which translation codes are enabled on the Teradata Database system. The translation codes convert client character data to the applicable default server character set.

Upon installation of Teradata client software, translation codes for the following client character sets are automatically and permanently enabled:

- ASCII. See [ASCII Client Character Set Support](#).
- EBCDIC. See [EBCDIC Client Character Set Support](#).
- UTF8. See [UTF8 Client Character Set Support](#).
- UTF16. See [UTF16 Client Character Set Support](#).

Note:

Depending on the client character set you choose, some configuration may be required. See [Using Standard Teradata Client Character Sets](#)

Default Language Support Mode

Each Teradata Database system is configured for a language support mode as part of sysinit:

- Standard (For details, see [Standard Language Support Mode](#).)
- Japanese (For details, see [Japanese Language Support Mode](#).)

The language support mode determines the default server character set, which in turn determines which character set is used to store data.

IF you enable this language support mode ...	Teradata Database stores object names using ...	and the user default server character set is...
Standard	UNICODE	LATIN
Japanese	UNICODE	UNICODE

The default server character set also determines which client character sets are supported. [Client Character Set Compatibility with Server Character Sets](#).

You can override the global default server character set associated with the language mode by specifying a default server character set for specific users or table columns.

For details about server character set options, see [Server Character Sets](#).

Enabling the Language Support Mode

During the sysinit process, at the following prompt appears:

```
Enable Japanese language support?
```

To optimize the database for Japanese language support, answer YES. Otherwise, answer NO.

If you answer NO, the Teradata Database uses standard language support.

Determining the Current Language Support Mode

To determine the current language support mode on your system, enter this command:

```
SELECT * FROM dbc.dbcinfoV WHERE infokey = 'LANGUAGE SUPPORT MODE';
```

For Japanese systems, the system returns:

```
*** Query completed. One row found. One column returned.
*** Total elapsed time was 1 second.
InfoData
-----
Japanese
```

For Non-Japanese systems, the system returns:

```
*** Query completed. One row found. One column returned.
*** Total elapsed time was 1 second.
InfoData
-----
Standard
```

Standard Language Support Mode

If you enable standard language support, then by default Teradata Database stores:

- Object names in the Data Dictionary using the UNICODE server character set
- User data in Teradata Database tables using the LATIN server character set.

Note:

You can override the global data storage default by specifying a different default character set for individual users and table columns. See [Specifying the Default Server Character Set](#).

For the standard language support mode, the default server character set for user DBC is also LATIN.

About Standard Language LATIN Character Coding

Standard language support provides Teradata Database internal coding for the entire set of printable characters from the ISO 8859-1 (Latin1) and ISO 8859-15 (Latin9) standard, including diacritical marks such as ä, ñ, Ÿ, Œ, and œ, though the Z with caron in Latin9 is not supported. ASCII control characters are also supported for the standard language set.

For a definition of the Teradata LATIN character set used to represent ASCII and EBCDIC characters, see [LATIN Server Character Set](#).

Compatible Languages

The LATIN server character set used in standard language support mode is sufficient to support the following languages.

International Languages That are Compatible with Standard Language Support			
Albanian	English	Germanic	Portuguese
Basque	Estonian	Greenlandic	Rhaeto-Romantic
Breton	Faroese	Icelandic	Romance
Catalonian	Finnish	Irish Gaelic (new orthography)	Samoan
Celtic	French	Italian	Scottish Gaelic
Cornish	Frisian	Latin	Spanish

International Languages That are Compatible with Standard Language Support

Danish	Galician	Luxemburgish	Swahili
Dutch	German	Norwegian	Swedish

Note:

For support of languages not shown in the preceding table, the server character set should be set either:

- System-wide, to [Japanese Language Support Mode](#) (for Japanese clients only)
- To UNICODE for the default character set for users or table columns that require the special language support. See [Specifying the Default Server Character Set](#).

Japanese Language Support Mode

If you enable Japanese language support, by default Teradata Database stores:

- Object names in the Data Dictionary using the UNICODE server character set
- User data in Teradata Database tables using the UNICODE server character set.

In Japanese Language Mode, the default server character set for user DBC is UNICODE.

Note:

The default server character set from user DBC (UNICODE) is inherited when the Teradata SQL DATABASE statement is used to change the default database.

About Using UNICODE to Store User Data

Unicode is a 16-bit encoding of virtually all characters in all current international languages. User data stored as Unicode can be easily shared among clients that use diverse character sets.

For more information about Teradata support for UNICODE, see [UNICODE Server Character Set](#).

About Object Names

Object names are always stored in the data dictionary in UNICODE.

Specification of object names is subject to these character usage rules:

- The rules are based on [Object Naming Controls](#), which define allowable characters.
- You can specify an object name as a regular identifier, that is, without quotation mark delimiters, if the name uses only characters in the current character repertoire.
- You can specify an object name as a delimited identifier, using characters outside the current character repertoire, by enclosing the name in quotations marks.
- Certain characters are explicitly disallowed from appearing in object names.

Related Information

For information on...	See...
Detailed considerations for object naming, including strategies for using the DBS Control fields that determine object naming rules	<i>Teradata Vantage™ - SQL Fundamentals</i> , B035-1141.
Setting DBS Control object naming controls.	<i>Teradata Vantage™ - Database Utilities</i> , B035-1102.

Object Naming Controls

The system determines the characters available for use in object names based on the rules determined by the settings of DBS Control object naming control fields.

Object naming controls determine several naming parameters including available character repertoire and name length.

Optional name validation rules can impose additional character restrictions.

Note:

Changing object naming control settings only affects objects that are created after the setting change.

Rules for Object Naming

The following summarizes object naming rules:

Parameter	Description
Object name length	A maximum of 128 characters when expressed in UNICODE normalization form D. The DBS Control NameValidationRule field can be set to apply additional character restrictions.
Characters allowed in object names not enclosed in quotation marks	An object name not enclosed in quotation marks must be composed of an identifier-start character followed by a sequence of identifier-start or identifier extend characters, up to the maximum object name length limit.

Parameter	Description
	<p>Note: Characters in object names not enclosed in quotation marks must also be in the session character set. Identifier start characters must be contained in the session character set and belong to one of the following Unicode General Category classes:</p> <ul style="list-style-type: none"> • Upper-case letters [Lu] • Lower-case letters [Ll] • Title-case letters [Lt] • Modifier letters [Lm] • Other letters [Lo] • Letter numbers [NI] <p>...Or be one of the following characters:</p> <ul style="list-style-type: none"> • NUMBER SIGN (U+0023) • DOLLAR SIGN (U+0024) • LOW LINE (U+005F) • INVERTED EXCLAMATION MARK (U+001A) • OVERLINE (U+203E) • EURO SIGN (U+20AC) • KATAKANA-HIRAGANA VOICED SOUND MARK (U+309B) • KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK (U+309C) • FULLWIDTH NUMBER SIGN (U+FF03) • FULLWIDTH DOLLAR SIGN (U+FF04) • FULLWIDTH LOW LINE (U+FF3F) <p>Identifier-extender characters must be in the session character set and belong to one of the following Unicode General Category classes:</p> <ul style="list-style-type: none"> • Non-spacing marks [Mn] • Spacing combining marks [Mc] • Decimal numbers [Nd] • Connector punctuations [Pc] • Formatting codes [Cf] <p>Note: The MIDDLE DOT character is also a valid identifier-extender character.</p>
Characters allowed only in object names that are enclosed in quotation marks	<p>A quoted string is required for object names that:</p> <ul style="list-style-type: none"> • Have an identifier-extender character as the first character. • Include the white space character, SPACE (U+0020) • Are Teradata keywords <p>In addition, object names that contain any character from the following classes must be enclosed in quotation marks, unless the character explicitly appears in the list of allowed characters:</p> <ul style="list-style-type: none"> • Other, Control [Cc] • Other, Not Assigned [Cn] <p>Note: No characters in this category appear in UNICODE character repertoire.</p>

Parameter	Description
	<ul style="list-style-type: none"> • Other, Private Use [Co] • Other, Surrogate [Cs] • Letter, Cased [LC] • Mark, Enclosing [Me] • Number, Other [No] • Punctuation, Dash [Pd] • Punctuation, Close [Pe] • Punctuation, Final quote [Pf] (may behave like Ps or Pe depending on usage) • Punctuation, Initial quote [Pi] (may behave like Ps or Pe depending on usage) • Punctuation, Other [Po] • Punctuation, Open [Ps] • Symbol, Currency [Sc] • Symbol, Modifier [Sk] • Symbol, Math [Sm] • Symbol, Other [So] • Separator, Line [Zl] • Separator, Paragraph [Zp] • Separator, Space [Zs] <p>Note: When used to enclose an object name, the beginning and ending quotation marks must be represented as a sequence of two QUOTATION MARK characters (U+0022). Each set of two quotation marks is counted as one character when calculating the name size limit.</p>
Disallowed characters	<p>The following characters cannot appear in an object name:</p> <ul style="list-style-type: none"> • NULL (U+0000) • SUBSTITUTE character (U+001A) • REPLACEMENT CHARACTER (U+FFFD) • Compatibility ideographs (U+FA6C, U+FA6F, U+FAD0, FAD1, FAD5, FAD6, and FAD7) <p>Note: The setting of the NameValidationRule field may define additional character restrictions. For more information, see <i>Teradata Vantage™ - Database Utilities</i>, B035-1102.</p>
Other considerations	<p>These additional restrictions apply:</p> <ul style="list-style-type: none"> • An object name consisting entirely of white spaces is not allowed. • A trailing white space is not considered part of an object name • You can use the NameValidationRule field to restrict object name allowable characters to a subset of those normally allowed with object naming. For more information, see <i>Teradata Vantage™ - Database Utilities</i>, B035-1102.

UNICODE Characters Allowed in Object Names

All object names are stored in the data dictionary in UNICODE. Teradata provides a file, UOBJNEXT.txt, which lists the UNICODE characters valid for object names.

You can download the file here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

Use of these UNICODE characters in object names is subject to limitations imposed by:

- The session (client) character set. See [Client Character Set Options](#).
- Object naming controls. See *Teradata Vantage™ - SQL Fundamentals*, B035-1141.

Display of Incompatible Object Names and Literals

Object names and literals returned by HELP and SHOW statements, and by requests that specify the EXPLAIN modifier, may not be sharable among clients with differing client (session) character sets.

- For SQL SHOW and EXPLAIN, any literals and object names that include characters not in the repertoire of the session character set, or that are otherwise unprintable, are shown as UNICODE delimited literals or identifiers. Specific incompatible characters are escaped to hexadecimal.
- For SQL HELP, the Column SQL Name and Title fields identify names/strings containing characters that are not translatable into the client character set, and show replacement characters for individual untranslatable characters.

For details, see *Teradata Vantage™ - SQL Data Definition Language Syntax and Examples*, B035-1144.

Related Information

For information on...	See...
Using the SHOW, HELP, and EXPLAIN commands	<i>Teradata Vantage™ - SQL Data Definition Language Syntax and Examples</i> , B035-1144.
UNICODE delimited literals and identifiers	<i>Teradata Vantage™ - SQL Fundamentals</i> , B035-1141.

Client Character Set Options

Overview

The following describes the standard Teradata-supplied character translation codes that support the use of the corresponding client character sets. Teradata does not supply any character sets for installation on a client.

Note:

Some client character sets are not supported with some Teradata client applications and drivers. For information on supported client character sets, see the *user guide* for the client application or driver you want to use.

About Client Character Set Translation Codes

When Teradata Database processes a client request, it translates client character data to the default server character set. Then, when returning an answer set, the database translates character data from the server character set back to the client character set.

Teradata Database provides the necessary translation codes for processing data from several common client character sets. Each translation code maps the characters in a client character set to the equivalent characters in a compatible Teradata Database server character set.

Translation codes also define how to convert characters to uppercase, for example when the Teradata Database stores characters in a column that is defined as `UPPERCASE`.

Four hexadecimal translation codes are required for each combination of a client character set and compatible server character set:

- External (Client) to Internal (Teradata Database)
- External-to-internal uppercase
- Internal (Teradata Database) to External (Client)
- Internal-to-external uppercase

Note:

A client character set is usable only if it is specified as the client default, and if the corresponding character translation code is installed and enabled on the Teradata Database system. Only 16 character sets can be active at a time.

Related Information

For information on...	See...
Teradata support for commonly used client character sets	<ul style="list-style-type: none"> • Permanently Enabled Client Character Set Support • Support of Other Client Character Sets
Enabling use of a client character set	Enabling a Client Character Set.
Setting the default client character set	System Determination of Client Character Set.
Creating a custom (site-defined) client character set	<ul style="list-style-type: none"> • Site-Defined Client Character Sets • Extended Site-Defined Client Character Sets

Teradata Support for Common Client Character Sets

Permanently Enabled Client Character Set Support

Teradata translation codes for the following client character sets are permanently enabled:

- [ASCII Client Character Set Support.](#)
- [EBCDIC Client Character Set Support.](#)
- [UTF8 Client Character Set Support.](#)
- [UTF16 Client Character Set Support.](#)

To use a client character set supported by permanently enabled Teradata translation codes, specify it as the client default. See [System Determination of Client Character Set.](#)

Support of Other Client Character Sets

Teradata also provides translation codes for support of these additional client character sets. However, you must enable each translation code you want to use before setting it as the client default. See [Enabling a Client Character Set.](#)

Type	Name
Japanese Client Character Set Support	<ul style="list-style-type: none"> • KATAKANAEBCDIC • KANJIEBCDIC5035_01 • KANJISJIS_0S • KANJIEBCDIC5026_01 • KANJIEUC_0U • KANJI932_1S0
Chinese Character Sets	<ul style="list-style-type: none"> • SCHEBCDIC935_2IJ • TCHEBCDIC937_3IB

Type	Name
	<ul style="list-style-type: none"> • SCHGB2312_1T0 • TCHBIG5_1R0 • SCHINESE936_6R0 • TCHINESE950_8R0
Korean Character Sets	<ul style="list-style-type: none"> • HANGULEBCDIC933_1II • HANGULKSC5601_2R4 • HANGUL949_7R0
Single-Byte International Character Sets	<ul style="list-style-type: none"> • THAI874_4A0 (Thai) • LATIN1250_1A0 Includes: Czech, Croatian, Albanian, Hungarian, Polish, Romanian, and Serbian • CYRILLIC1251_2A0 (Cyrillic and Russian) • HEBREW1255_5A0 • ARABIC1256_6A0 • LATIN1254_7A0 (Turkish) • LATIN1258_8A0 (Vietnamese) • LATIN1252_3A0 • LATIN1252_0A

Client Character Sets Listed in DBC.CharTranslationV View

The DBC.CharTranslationV view lists the client character sets for which Teradata supplies translation codes, and shows the code sequences the system uses to translate character data for each set from client form to server form and from server form back to client form.

The character set list in DBC.CharTranslationV does not include ASCII, EBCDIC, UTF8 and UTF16, for which the Teradata translation codes are permanently enabled. For an example, see [About the DBC.CharTranslationsV View](#).

Note:

You can create custom (site-defined) character sets, but you must add the names of any custom character set to DBC.CharTranslationV before you can enable and use it. See [Site-Defined Client Character Sets](#)

Client Character Set Compatibility with Server Character Sets

To be usable, a client character set must be compatible with the applicable default server character set.

The default server character set is determined in the following hierarchical order. If a default is not defined, the system defers to the next lower item on the list:

1. Table column default character set
2. User default character set
3. System (global) default character set.

Note:

The current system default is defined in the DBSControl DefaultCharacterSet field.

For more information about defining the default server character set, see [Specifying the Default Server Character Set](#).

The following table lists client-server character set compatibilities:

If the default server character set is...	The client character set can be...
LATIN	<ul style="list-style-type: none"> • ASCII • EBCDIC • EBCDIC037_OE • EBCDIC277_OE • EBCDIC273_OE • LATIN1_OA • LATIN9_OA • LATIN1252_OA • A simple site-defined ASCII-based character set <p>Note: Do not use client character sets other than those listed to insert data into a LATIN column, even if the character set appears to be represented in LATIN (for example, Turkish). The system may incorrectly translate such characters when converting them to UNICODE.</p>
UNICODE	Any valid single- or multibyte client character set.
KANJISJIS	KANJISJIS supports mixed single and multibyte characters and its form-of-use is identical with that of its client character set analog
GRAPHIC	Multibyte characters from the defined Graphic repertoire of the following character sets: <ul style="list-style-type: none"> • KanjiEBCDIC • KanjiSJIS_OS • KanjiEUC_OU

Client Character Set Compatibility by Client Type

The following table provides general guidelines for choosing client character sets, based on how each set functions on certain client types.

IF the client type is ...	THEN the client character sets that may work best are ...	
Mainframe	<ul style="list-style-type: none"> • EBCDIC • EBCDIC037_0E • KANJIEBCDIC5026_0I • KANJIEBCDIC5035_0I 	<ul style="list-style-type: none"> • KATAKANAEBCDIC • SCHEBCDIC935_2IJ • TCHEBCDIC937_3IB • HANGULEBCDIC933_1II
Network connected, running the UNIX® operating system	<ul style="list-style-type: none"> • ASCII • KANJIEUC_0U • LATIN1_0A • LATIN9_0A 	<ul style="list-style-type: none"> • UTF8 • SCHGB2312_1T0 • TCHBIG5_1R0 • HANGULKSC5601_2R4
Network connected, running the Windows operating system	<ul style="list-style-type: none"> • ASCII • KANJISJIS_0S • LATIN1252_0A • UTF8 • UTF16 • SCHGB2312_1T0 • TCHBIG5_1R0 • HANGULKSC5601_2R4 • SCHINESE936_6R0 • HANGUL949_7R0 	<ul style="list-style-type: none"> • TCHINESE950_8R0 • KANJI932_1S0 • THAI874_4A0 • LATIN1250_1A0 • CYRILLIC1251_2A0 • LATIN1252_3A0 • HEBREW1255_5A0 • ARABIC1256_6A0 • LATIN1254_7A0 • LATIN1258_8A0

ASCII Client Character Set Support

ASCII is a character set standard in which each byte represents one character.

The Teradata translation codes for the ASCII client character set are permanently enabled in Teradata Database.

Various industrial, national, and international organizations have developed custom character sets that are extensions of the ASCII standard. If you use accented and non-Latin characters, do not use the ASCII client character set. Instead, enable use of one of the other supported ASCII-compatible client character sets or a custom (site-defined) character set that exactly matches your character set needs.

Teradata provides following popular extended ASCII-based client character sets, which you must enable before using. Also see [Enabling a Client Character Set](#).

ASCII Client Character Set Name	Description
Latin1252_3A0	Windows Latin extension to ASCII, officially known as code page 1252.

ASCII Client Character Set Name	Description
	Note: Latin1252_3A0 is 100% compatible with Windows Code Page 1252.
Latin1252_0A	Windows Latin extension to ASCII, officially known as code page 1252. Note: Latin1252_0A is not 100% compatible with Windows Code Page 1252. Some characters in the range of 0x80-8E and 0x90-9E in Latin1252_0A are not supported by Teradata Latin and require use of UNICODE.
Latin1_0A	International standard ISO 8859-1 (Latin 1)
Latin9_0A	International standard ISO 8859-15 (Latin 9)

Note:

If you use or access any characters outside the 7-bit ASCII range, that is, characters with values from 0x80 to 0xff, use a language-specific session character set, for example, a Chinese or Korean character set, a Unicode character set (UTF8 or UTF16), or Latin1252_3A0 for all Western European Windows users.

EBCDIC Client Character Set Support

In the EBCDIC client character set, each byte represents one character. EBCDIC is available for mainframe clients.

The Teradata EBCDIC client character set is suitable for applications that use simple Roman characters, however, accented characters and certain special symbols require an exact match between application and client character set. In such cases it is normally best to use a custom (user-defined) character set. See [Site-Defined Client Character Sets](#)

The translation codes for the EBCDIC character set are permanently enabled for use in Teradata Database.

Teradata also provides support for several commonly used EBCDIC character set variations, which you must enable before using. See [Enabling a Client Character Set](#).

This EBCDIC client character set...	Most commonly used in...
EBCDIC037_0E	United States and Canada
EBCDIC273_0E	Austria and Germany
EBCDIC277_0E	Denmark and Norway

UTF8 Client Character Set Support

The Teradata Universal Coded Character Set Transformation Format (UTF8) client character set supports UTF8, a standard way of encoding Unicode character data that is optimized for backward compatibility with ASCII. This character set is usable for all languages. In Teradata UTF8, a character can consist of one to three bytes.

The UTF8 client character set is permanently enabled for use in Teradata Database.

IF a byte in a UTF8 string is ...	THEN it ...
less than 0x80	represents the same character defined by standard ASCII.
greater than or equal to 0x80	is part of a multibyte sequence and is not a standard ASCII character.

UTF8 Multibyte Sequences

To determine the length of a byte sequence in UTF8, examine the first byte.

IF the...	THEN the sequence is...
high order bit is zero	one byte long. This leaves seven bits to encode information. If a character has a Unicode value that can be represented in seven bits, it is represented as a byte containing the Unicode value. For example, Unicode value 0x0041 is transformed to UTF8 byte 0x41.
three high order bits are 110	a two-byte sequence. The second byte has the two high order bits set to 10. There are five free bits in the first byte and six free bits in the second byte. This allows eleven bits to represent a numeric value. If a character has a Unicode value that can be represented in eleven bits, and cannot be represented by a shorter UTF8 sequence, then it is represented as two bytes, where the free bits contain the Unicode value. For example, Unicode value 0x03F1 is transformed to UTF8 byte sequence 0xCF 0xB1.
four high order bits are 1110	a three byte sequence. The second and third bytes have the two high order bits set to 10. There are four free bits in the first byte and six free bits in each of the second and third bytes. This allows sixteen bits to represent a numeric value. If a character has a Unicode value that can be represented in sixteen bits, and can not be represented by a shorter UTF8 sequence, it is represented as three bytes, where the free bits contain the Unicode value (for example, Unicode value 0x3000 is transformed to UTF8 byte sequence 0xE3 0x80 0x80).

Untranslatable Characters

The KANJI1 server character set is ambiguous with regards to multibyte characters and some single-byte characters, and cannot be reliably translated to UTF8.

For best results, use the UNICODE server character set with the UTF8 session character set, unless the KANJI1 server character set is required.

UTF16 Client Character Set Support

The UTF16 client character set is permanently enabled for use in Teradata Database.

The Teradata UTF16 client character set supports UTF16, one of the standard ways of encoding Unicode character data, based on the Unicode® standard. Each UTF16 character requires exactly 16 bits.

Untranslatable Characters

The KANJI1 server character set is ambiguous with regards to multibyte characters and some single-byte characters, and cannot be reliably translated to UTF16.

For best results, use the UNICODE server character set with the UTF16 session character set, unless the KANJI1 server character set is required.

Endianness

The endianness of UTF16 is the same as endianness of numbers for the session.

Endianness is the byte ordering convention of data that is represented with multiple bytes. The ordering method is either big endian or little endian. For example, the big endian method indicates the number 256 as the sequence 0x01 0x00. The little endian method indicates the number 256 as 0x00 0x01.

Error Messages

The maximum length of an SQL error message is 255 bytes. Because UTF16 has two bytes per character, the maximum length of an SQL error message is 127 characters.

For example, message number 5468 in UTF8 appears like this in BTEQ:

```
BTEQ -- Enter your DBC/SQL request or BTEQ command:
DELETE USER tn2;
*** Failure 5468 Cannot delete database because either this database has a join
or hash index in it or one of its tables has a join or hash index defined on it.
```

In UTF16, the message text is truncated:

```
BTEQ -- Enter your DBC/SQL request or BTEQ command:
DELETE USER tn2;
*** Failure 5468 Cannot delete database because either this database has a join
or hash index in it or one of its tables has a join or hash index.
```


Japanese Client Character Set Support

Teradata Database allows you to submit requests using Japanese character data containing Kanji, Katakana, Hiragana, Romaji, and various other characters.

Characters	Description
Kanji	An ideographic script.
Katakana	A syllabic script used for foreign words.
Hiragana	A syllabic script that can be mapped one-for-one with Katakana. Hiragana is used for Japanese articles and words that do not have Kanji forms.
Romaji	The Japanese name for the standard Roman (Latin) alphabet.

Basic Requirements for Use of Japanese Character Sets

All Japanese client character set must be manually enabled for use in Teradata Database, as shown in [Using Standard Teradata Client Character Sets](#).

Hankaku and Zenkaku

Japanese characters tend to be drawn to fill a square, and text is often written vertically. Standard fixed-width Western writing tends to require two characters to fill a square, and the style is called Hankaku, or half-square, as opposed to the standard Zenkaku, or full-square, mode of writing.

SQL Keywords

SQL keywords, such as SELECT, cannot be translated into Japanese and be expected to act as keywords. Only Hankaku Romaji characters may be used to type a keyword, not Zenkaku Romaji characters.

Japanese Encoding Schemes

Because thousands of characters are required to write Japanese, it is not possible to represent all characters as a single-byte. For this reason, Japanese character sets use either:

- A multibyte mapping standard.
- The combination of a multibyte standard to handle most of the enormous number of required characters, and a single-byte standard to efficiently code a smaller number of frequently used characters.

Mapping Standards

There are several mapping standards used in the character sets supported under the Teradata Database Japanese character support.

Standard	Description
JIS X 0201	Similar to the ISO 8859 family of protocols with the exception that there are some changes in the ASCII region. The area from 0xA1-0xDF is used mainly for Hankaku Katakana.
JIS X 0208	A double-byte standard that includes the more common Kanji characters along with many uncommon ones. It also includes Hiragana, Katakana and Zenkaku Romaji characters, as well as Greek, Cyrillic, and various other characters.
JIS X 0212	A double-byte standard that was designed to include many of the rarer Kanji characters.
IBM Code Page 300	A double-byte standard similar in content to JIS X 0208, but designed for an EBCDIC platform.
IBM-provided single-byte standards for Japanese	Based on EBCDIC, but include Hankaku Katakana characters. These mapping standards are described in more detail in the descriptions of individual supported character sets.
UTF-8	A version of Unicode optimized for backward compatibility with ASCII. In Teradata UTF8, a character can consist of from one to three bytes.

For more information on Japanese encodings and mapping standards, see [Japanese Encodings and Mapping Standards](#)

Teradata Supported Japanese Character Sets

Teradata provides the translation codes to support several common Japanese character sets:

- Three [IBM-Compatible Japanese Character Sets](#) (KanjiEBCDIC)
- A [UNIX Compatible Japanese Character Set \(KANJI EUC_0U\)](#), which emulates the standard Extended UNIX Code (EUC) style of mixed single- and multibyte character data.
- Several [Windows-Compatible Japanese Character Sets](#), which emulate the Shift-JIS style of mixed single- and multibyte character data

IBM-Compatible Japanese Character Sets

The term KanjiEBCDIC refers collectively to these character sets:

- [KANJI EBCDIC5026_0I](#)
- [KANJI EBCDIC5035_0I](#)
- [KATAKANA EBCDIC](#)

These character sets emulate the IBM style, where the multibyte character sequence in a character string is bracketed by Shift-Out and Shift-In characters.

IBM Mainframe Character Sets

Character Set Name	CCSID (1)	GCSGID (2)	CPGID (3)	Description
KANJIEBDIC5026_OI	05026	01172	00290	Single-byte
		00370	00300	Multibyte; 1880 user-defined characters
KANJIEBDIC5035_OI	05035	01172	01027	Single-byte
		00370	00300	Multibyte; 1880 user-defined characters
KATAKANAEBDIC	13218	00332	00290	Single-byte; no 'a'-'z'
		00370	00300	Multibyte; 1880 user-defined characters

(1) Coded Character Set ID.

(2) Character Set ID. Describes the permitted characters.

(3) Code Page. Describes the encoding of the permitted characters.

Special Characters in KanjiEBDIC to KANJI1 Translations

Hexadecimal Value	Character	Description
0x10	¢	Cent sign
0x11	£	Pound Sterling currency symbol
0x12	¬	IBM logical NOT sign
0x13	\	Backslash

Because other translations map the values to control characters, they cannot be shared with other client character sets.

Extended Characters in KanjiEBDIC to UNICODE Translations

The table represents the Unicode hexadecimal for the five characters which were treated specially in Kanji1. The five characters already exist in Unicode and no custom mapping is needed.

Hexadecimal Value	Character	Description
0x00A2	¢	Cent sign
0x00A3	£	Pound Sterling currency symbol
0x00AC	¬	IBM logical NOT sign

Hexadecimal Value	Character	Description
0x005C	\	Backslash
0x007E	~	Tilde

Object Names

Object names on systems enabled with Japanese language support can contain single-byte Latin and Katakana characters from the JIS X 0201 standard, and double-byte characters from the JIS X 0208 standard.

For information on the rules and restrictions for naming Teradata Database objects, see the topics beginning with [About Object Names](#).

For additional rules and restrictions that apply to naming Teradata Database objects, see *Teradata Vantage™ - SQL Fundamentals*, B035-1141.

The following table shows where to find information on characters allowed in object names.

IF the client character set is ...	THEN object names can contain single-byte Latin and Katakana characters shown in ...	AND double-byte characters shown in ...
KATAKANAEBCDIC	EBCDIC KATAKANA Character Set .	the following shaded areas in EBCDIC Kanji Character Set : <ul style="list-style-type: none"> • IBM Extended non-Kanji • IBM Basic Kanji • IBM Extended Kanji
KANJIEBCDIC5026_0I	EBCDIC Japanese Extended Katakana Character Set .	
KANJIEBCDIC5035_0I	EBCDIC Japanese Extended English Character Set .	

Double-byte characters in the User and Marker areas in [EBCDIC Kanji Character Set](#) are not valid for object names. Additionally, double-byte characters in the following table are not in JIS X 0208 and are not valid for object names under KANJIEBCDIC5026_0I, KANJIEBCDIC5035_0I, and KATAKANAEBCDIC.

First Byte	Second Byte						
0x40	40						
0x41	41 - FE						
0x42	41 - 6C	6E - 7A	7C - 80	AA - C0	EA - EF	FA - FE	
0x43	41 - 45	4A - 4F	5B - 7A	7C - 7F	A1	B0 - B9	DC - DD
	E0 - F2						

First Byte		Second Byte												
0x44	42 - 45		4A - 4E		50	5A - 7F		DC - DD		E0 - F3				
0x51	BF	DA	F7											
0x52	4B	7B	94	9E	EB									
0x53	86	94	97	C7										
0x56	44	5D	62	65 - 66		6A	6E	70 - 71		74	7B	86	8A	91
	94	98	A2 - A4		A8	B2	B5	C9	CE	DC	E8	EF	F7	
0x57	4F	50	56	58	61	63	6B	7C	80	87	8B	8E	A9	AE
	BD	C0	DE											
0x58	60	63	6C	70	73	77	82	8A	8C	A1	AD	B0	B1 - B2	
	BE	C0	EC	FC	FE									
0x59	42	44	51	67 - 68		70	78	7F	80	84	87	93	95	97
	C7	CB	D8	E8	F1									
0x5A	44	51	57	5F	65	67 - 68		76	79	7E	94	B5	DA	FB
0x5B	50	53 - 54		7E	97 - 99		9D	A0 - A1		A4	A7 - A8		AB	F9
	AE - AF		B4	B6	B8	BB	C7 - C9		D8	E3	EC	F0	F3	
0x5C	5F	66	6E	78 - 79		90	9A - 9B		A6	A9	AF	B9	BC	EA
	F6	F9	FD - FE											
0x5D	4E	58	60	62	6B	72	74	7A	7F	80	8A	8E	93	95
	A6	C1	CA	D5	D7 - D8		E3	EA	ED	F5 - F6		FB - FD		
0x5E	49	4B	5A	73	75	86	8A	94	99	9E - 9F		A1 - A2		
	A7 - A8		AA - AB		AD - AE		BC - BD		C0	D1	E7			
0x5F	69	70 - 73		7A	8E	AA	AE	B2 - B3		B6	D1 - D2		D5	D9
	DC - DD		DF											
0x60	4F	58 - 5A		82	AD	C4	C6	D4	DA	E1	FD			
0x61	57	6E	77	F8										
0x62	48	5B	67	74	7E	8A	9E	A1	A8	B9	BB	CE	E2	
0x63	5E	90	C5	CB	D4	DF	E1	ED - EF						
0x64	45	50	7C - 7D		85	8D - 8E		C9	F1					
0x65	49	53	64	66 - 67		6B	83	85 - 86		8A	8C - 8D		8F	90

First Byte			Second Byte											
	93	95	9B	9E - 9F		A1	A3	A5 - A6		A8	AB - AD		B4 - B9	
	BC - BD		C1	C3	C5 - C6		C8 - C9		CC	CE	D0	D5	D8	DD
	E4	E8	F0	F4	F6									
0x66	4A	74	78 - 79		7F	97	99	9E	A2 - A5		CB - CC		DD	DF
	E5 - E6		F9											
0x67	59	6B - 6C		8F	91	93	9B	9F	B1	E2	E4	F4	FE	
0x68	56													

KANJIEBCDIC5026_0I

This character set is based on the IBM Japanese Extended Katakana character set.

Under this character set, character data types support mixed single- and multibyte characters.

Valid Characters in KANJIEBCDIC5026_0I

This character set includes:

- Single-byte characters from code page 290.
- Multibyte characters from code page 300.

For a list of valid and not valid characters in object names, see [Object Names](#).

Usage Notes

KANJIEBCDIC5026_0I supports the same characters as KANJIEBCDIC5035_0I, differing only in the encoding of lowercase Latin letters and Hankaku Katakana characters.

Lowercase characters in KANJIEBCDIC5026_0I occur in different positions than in EBCDIC. This means that lowercase characters are misinterpreted on input and garbled on output if the incorrect client character is chosen.

Related Information

For more information on ...	See ...
the IBM Japanese Extended Katakana character set on which KANJIEBCDIC5026_0I is based	EBCDIC Japanese Extended Katakana Character Set .
the character set on which double-byte characters in KANJIEBCDIC5026_0I are based	EBCDIC Kanji Character Set .

KANJIEBCDIC5035_0I

This character set is based on the IBM Japanese Extended English character set.

Under this character set, character data types support mixed single- and multibyte characters.

Valid Characters in KANJIEBCDIC5035_0I

This character set includes:

- Single-byte characters from code page 1027.
- Multibyte characters from code page 300.

For a list of valid and not valid characters in object names, see [Object Names](#).

Usage Notes

KANJIEBCDIC5035_0I supports the same characters as KANJIEBCDIC5026_0I, differing only in the encoding of lowercase Latin letters and Hankaku Katakana characters.

Lowercase KANJIEBCDIC5035_0I characters match standard EBCDIC.

Related Information

For more information on ...	See ...
the IBM Japanese Extended English character set on which KANJIEBCDIC5035_0I is based	EBCDIC Japanese Extended English Character Set .
the character set on which double-byte characters in KANJIEBCDIC5035_0I are based	EBCDIC Kanji Character Set .

KATAKANAEBCDIC

This character set is based on Japanese Katakana EBCDIC.

Under this character set, character data types support mixed single- and multibyte characters.

Valid Characters in KATAKANAEBCDIC

This character set includes:

- Single-byte characters from code page 290, except for characters a–z.
- Multibyte characters from code page 300.

For a list of valid and not valid characters in object names, see [Object Names](#).

Usage Notes

KATAKANAEBCDIC supports the same characters as KanjiEBCDIC5026_0I, except that it lacks lowercase single-byte Latin letters.

The KATAKANAEBCDIC client character set maps Hankaku lowercase characters to Hankaku uppercase characters on output from the Teradata Database.

Related Information

For more information on ...	See ...
the Japanese Katakana EBCDIC character set on which KATAKANAEBCDIC is based	EBCDIC KATAKANA Character Set.
the character set on which double-byte characters in KATAKANAEBCDIC are based	EBCDIC Kanji Character Set.

UNIX Compatible Japanese Character Set (KANJI EUC_0U)

KanjiEUC refers to the character set KANJI EUC_0U, which is compatible with the UNIX operating system.

KanjiEUC emulates the standard Extended UNIX Code style of mixed single- and multibyte character data, where the most significant bit of each byte classifies the byte as a single-byte character or part of a multibyte character.

Valid Characters in KanjiEUC

The KanjiEUC character set includes all characters in the JIS X 0201, JIS X 0208, and JIS 0212 standards, plus extensions.

The valid ranges for JIS X 0201 characters in KanjiEUC dictionary data include all of U.S. ASCII and the portion of JIS X 0201 for which the second byte ranges from A1 through DF. See the rows for code set 0 (cs0) and code set 2 (cs2) in the succeeding table.

External Encodings

KanjiEUC uses the four external code sets defined in the following table.

Code Set	Description	Note
0	Single-byte character data Not permitted for the GRAPHIC server character set	JIS X 0201 For the detailed encoding, see Shift-JIS Encoding: Detailed View.
1	Two-byte character data	JIS X 0208

Code Set	Description	Note
2	Two-byte multibyte character with first byte ss ₂ =0x8E Not permitted for the GRAPHIC server character set	JIS X 0201 Hankaku Katakana
3	Three-byte multibyte character with first byte ss ₃ =0x8F	JIS X 0212

Object Names

Object names on systems enabled with Japanese language support can contain single-byte Latin and Katakana characters from the JIS X 0201 standard, and double-byte characters from the JIS X 0208 standard.

The valid ranges for JIS X 0201 characters in object names under the KanjiEUC client character set appear in rows cs0 and cs2 in [KanjiEUC Code Set Localization](#). The set does not permit Katakana symbols 0x8EA1—0x8EA5 nor Unicode symbols other than \$, #, and _.

The valid ranges for JIS X 0208 characters in object names under the KanjiEUC client character set appear in row cs1 in [KanjiEUC Code Set Localization](#). Characters in the reserved regions of the standard are not allowed.

Characters from JIS X 0212 (row cs3) are not valid in object names. Additionally, some characters that are valid in JIS X 0208 do not map to the KanjiEBCDIC encoding and are not valid in KanjiEUC object names. The following table provides a complete list of multibyte character codes that are not valid for object names under the KANJIEUC_0U character set.

First Byte	Second Byte	Third Byte
0xA1	0xA1 - 0xAA	0xAD - 0xB1
	0xB3 - 0xBB	0xBD - 0xEF
	0xF1 - 0xF3	0xF5 - 0xFE
0xA2	0xA1 - 0xFE	
0xA6 - 0xA8	0xA1 - 0xFE	
0xF4	0xA5 - 0xA6	
0x8E	0xA1 - 0xA5	
0x8F	0xA1 - 0xFE	0xA1 - 0xFE

For information on the rules and restrictions for naming Teradata Database objects, see the topics beginning with [About Object Names](#).

Also see *Teradata Vantage™ - SQL Fundamentals*, B035-1141, which covers topics such as:

- Translation conventions for storing object names in the data dictionary

- Rules for object name comparison

Related Information

For more information on ...	See ...
the JIS X 0201 standard	JIS X 0201.
the JIS X 0208 standard	JIS X 0208.
the standard Extended UNIX Code (EUC) style of mixed single- and multibyte character data	Extended UNIX Code (EUC).

Windows-Compatible Japanese Character Sets

The following Windows compatible Japanese character sets are provided for Windows PC platforms:

- [KANJIISJIS_0S](#)
- [KANJI932_1S0](#)

KANJISJIS_0S

KANJISJIS_0S is provided for PC platforms running Windows, and emulates the Shift-JIS style of mixed single- and multibyte character data, where the range of the first byte in a character determines if the character is represented as one byte or two bytes.

Valid Characters in KANJISJIS_0S

The KANJISJIS_0S character set includes all the characters in the JIS X 0201 and JIS X 0208 standards, plus extensions, including single-byte Hankaku Katakana characters and multibyte Kanji characters.

KANJISJIS_0S is not 100% compatible with the latest Windows code page 932 as the mapping to Unicode is different for several Japanese characters including the Yen symbol. Therefore, the best practice is to use KANJI932_1S0 and UTF8/UTF16 session character sets with the UNICODE server character set.

Object Names

Object names on systems enabled with Japanese language support can contain single-byte Latin and Katakana characters from the JIS X 0201 standard, and double-byte characters from the JIS X 0208 standard.

The valid ranges for JIS X 0201 and JIS X 0208 characters in KANJISJIS_0S object names appear in [Shift-JIS Encoding](#). Characters in the reserved regions of the JIS X 0208 standard are not allowed.

Gaiji characters with first byte values 0xF0-0xF9 and IBM characters with first byte values 0xFA-0xFC are not allowed in object names. Additionally, some characters that are valid in JIS X 0208 do not map to the KanjiEBCDIC encoding and are not valid in KANJISJIS_0S object names.

The following table provides a complete list of multibyte character codes that are not valid for object names under the KANJISJIS_OS character set.

First Byte	Second Byte	
0x81	0x40 - 0x49	0x4C - 0x50
	0x52 - 0x5A	0x5C - 0x8F
	0x91 - 0x93	0x95 - 0xFC
0x83	0x9F - 0xFC	
0x84	0x40 - 0xFC	
0xEA	0xA3 - 0xA4	
0xF0 - 0xFC	0x40 - 0x7E	0x80 - 0xFC

For information on the rules and restrictions for naming Teradata Database objects, see the topics beginning with [About Object Names](#).

Also see *Teradata Vantage™ - SQL Fundamentals*, B035-1141, which covers topics such as:

- Translation conventions for storing object names in the data dictionary
- Rules for object name comparison

Related Information

For more information on ...	See ...
the JIS X 0201 standard	JIS X 0201 .
the JIS X 0208 standard	JIS X 0208 .
the Shift-JIS encoding	Shift-JIS (DOS Kanji) Encoding .

KANJI932_1S0

KANJI932_1S0 character set is based on Windows Code Page 932. It emulates the Shift-JIS style of mixed single- and multibyte character data, where the range of the first byte in a character determines if the character is represented as one byte or two bytes.

Valid Characters in KANJI932_1S0

The KANJI932_1S0 character set includes all the characters in the JIS X 0201 and JIS X 0208 standards, plus extensions. Supported characters include single-byte Hankaku Katakana characters and multibyte Kanji characters.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
K1S0SUCD.txt	KANJI932_1S0 Single Byte to Unicode	Maps KANJI932 to Unicode.
K1S0MUCD.txt	KANJI932_1S0 Multibyte to Unicode	Maps the multibyte character portion of KANJI932 to Unicode.

Object Names

Object names on systems enabled with Japanese language support can contain single-byte Latin and Katakana characters from the JIS X 0201 standard, and double-byte characters from the JIS X 0208 standard.

The valid ranges for JIS X 0201 and JIS X 0208 characters in KANJI932_1S0 object names appear in [Shift-JIS Encoding](#). Characters in the reserved regions of the JIS X 0208 standard are not allowed.

Gaiji characters with first byte values 0xF0-0xF9 and IBM characters with first byte values 0xFA-0xFC are not allowed in object names. Additionally, some characters that are valid in JIS X 0208 do not map to the KanjiEBCDIC encoding and are not valid in KANJI932_1S0 object names. The following table provides a complete list of multibyte character codes that are not valid for object names under the KANJI932_1S0 character set.

First Byte	Second Byte	
0x81	0x40 - 0x49	0x4C - 0x50
	0x52 - 0x5A	0x5C - 0x8F
	0x91 - 0x93	0x95 - 0xFC
0x83	0x9F - 0xFC	
0x84	0x40 - 0xFC	
0xEA	0xA3 - 0xA4	
0xF0 - 0xFC	0x40 - 0x7E	0x80 - 0xFC

For details on the rules and restrictions for naming Teradata Database objects, see *Teradata Vantage™ - SQL Fundamentals*, B035-1141.

Related Information

For more information on ...	See ...
the JIS X 0201 standard	JIS X 0201.
the JIS X 0208 standard	JIS X 0208.
the Shift-JIS encoding	Shift-JIS (DOS Kanji) Encoding.

Chinese Character Sets

Teradata supplies several multibyte character sets to support Simplified Chinese and Traditional Chinese on mainframe and network-attached clients.

Multibyte Character Set Encoding Forms

Each character set uses a specific encoding form to distinguish single-byte characters from multibyte characters.

Character Set	Description	Encoding Form
SCHEBCDIC935_2IJ	Simplified Chinese (IBM CCSID 935) for mainframe clients.	EBCDIC Shift-Out/Shift-In. Shift-out character 0x0E and shift-in character 0x0F bracket each string of double-byte characters.
TCHEBCDIC937_3IB	Traditional Chinese (IBM CCSID 937) for mainframe clients.	
SCHGB2312_1T0	Simplified Chinese (mixed GB2312) for network-attached clients.	Extended UNIX Code (EUC) composed of two code sets: cs0 for single-byte characters and cs1 for double-byte characters.
TCHBIG5_1R0	Traditional Chinese (Big5) for network-attached clients.	Value of first byte in sequence distinguishes single-byte characters from double-byte characters.
SCHINESE936_6R0	Simplified Chinese (mixed GB2312) for network-attached clients.	Value of first byte in sequence distinguishes single-byte characters from double-byte characters.
TCHINESE950_8R0	Traditional Chinese (Big5) for network-attached clients.	Value of first byte in sequence distinguishes single-byte characters from double-byte characters.

Object Names

To determine whether a Chinese character is valid in an object name:

1. Find the text file that maps the character set to UNICODE.

You can find the file here:

- a. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
- b. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

2. In the text file, find the Unicode character to which the client character in question maps.

3. Find the file that identifies valid Unicode characters, UOBJNEXT.txt.

You can download the file here:

- a. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
- b. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

4. If the Unicode character appears in the file that is applicable to your system, you can use the client character that maps to it in an object name.

Internal Storage Limitations

Character data entered using Chinese client character sets should be stored in columns defined as Unicode. The UNICODE server character set requires two bytes of storage per character so that a CHAR(5) CHARACTER SET UNICODE field occupies 10 bytes of storage.

Given the 64000 byte limit on column size, a column cannot exceed 32000 characters. Furthermore, the combination of character data and other data types cannot exceed the 64000 byte limit on row size.

SCHEBCDIC935_2IJ

The SCHEBCDIC935_2IJ character set is based on IBM Coded Character Set ID (CCSID) 935, and provides support for simplified Chinese on mainframe clients. Under this character set, character data types support mixed single- and multibyte characters.

Character Set Name	Description
SCHEBCDIC935_2IJ	IBM Shift-Out/Shift-In (SO/SI) protocol. Shift-out (0x0E) and shift-in (0x0F) characters distinguish single-byte characters from double-byte characters. A shift-out, shift-in pair must bracket each string of double-byte characters.

Valid Characters in SCHEBCDIC935_2IJ

The SCHEBCDIC935_2IJ character set includes definitions for the single-byte characters in IBM Code Page (CPGID) 836, an EBCDIC-based code page similar to IBM Code Page (CPGID) 037.

The value of each byte of a double-byte character is in the range 0x41 through 0xFD, except for the IDEOGRAPHIC SPACE character, which is represented by 0x4040.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
C2IMUNCD.txt	SCHEBCDIC935_2IJ Multibyte to Unicode	Maps the multibyte character portion of SCHEBCDIC935 to Unicode.
C2ISUNCD.txt	SCHEBCDIC935_2IJ Single Byte to Unicode	Maps the single-byte character portion of SCHEBCDIC935 to Unicode.

TCHEBCDIC937_3IB

The TCHEBCDIC937_3IB character set is based on IBM Coded Character Set ID (CCSID) 937, and provides support for traditional Chinese on mainframe clients. Under this character set, character data types support mixed single- and multibyte characters.

Character Set Name	Description
TCHEBCDIC937_3IB	IBM Shift-Out/Shift-In (SO/SI) protocol. Shift-out (0x0E) and shift-in (0x0F) characters distinguish single-byte characters from double-byte characters. A shift-out, shift-in pair must bracket each string of double-byte characters.

Valid Characters in TCHEBCDIC937_3IB

The TCHEBCDIC937_3IB character set includes definitions for the single-byte characters in IBM Code Page (CPGID) 037, an EBCDIC code page.

The value of each byte of a double-byte character is in the range 0x41 through 0xFD, except for the IDEOGRAPHIC SPACE character, which is represented by 0x4040.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
C3IMUNCD.txt	TCHEBCDIC937_3IB Multibyte to Unicode	Maps the multibyte character portion of TCHEBCDIC937 to Unicode.
C3ISUNCD.txt	TCHEBCDIC937_3IB Single Byte to Unicode	Maps the single-byte character portion of TCHEBCDIC937 to Unicode.

SCHGB2312_1T0

The SCHGB2312_1T0 character set provides support for simplified Chinese (mixed ASCII/GB 2312-1980) on network-attached clients. Under this character set, character data types support mixed single- and multibyte characters.

SCHGB2312_1T0 is an Extended UNIX Code (EUC) style, composed of two code sets: cs0 and cs1.

Code Set	Representation	Description
cs0	0xxxxxxx	One byte per character, assigned in accordance with 7-bit ASCII standard. The high-order bit must be zero
cs1	1xxxxxxx 1xxxxxxx	Two bytes per character. All bytes have high-order bit set to one

Valid Characters in SCHGB2312_1T0

The SCHGB2312_1T0 character set includes definitions for the single-byte characters corresponding to seven-bit ASCII.

The value of each byte of a double-byte character is in the range 0xA1 through 0xFE.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
C1T0UNCD.txt	SCHGB2312_1T0 Code Set 0 to Unicode	Maps SCHGB2312 Code Set 0 to corresponding Latin letters of Unicode.
C1T1UNCD.txt	SCHGB2312_1T0 Code Set 1 to Unicode	Maps SCHGB2312 Code Set 1 to Unicode.

TCHBIG5_1R0

The TCHBIG5_1R0 character set provides support for traditional Chinese (Big5) on network-attached clients.

For TCHBIG5_1R0, a CHARACTER data type supports mixed single- and multibyte characters. The value of the first byte in the sequence distinguishes single-byte characters from double-byte characters.

IF the value of the first byte is ...	THEN the length of the character is ...
less than 0x81	one byte
greater than or equal to 0x81	two bytes

Valid Characters in TCHBIG5_1R0

The TCHBIG5_1R0 character set includes definitions for the single-byte characters corresponding to seven-bit ASCII.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
C1RMUNCD.txt	TCHBIG5_1R0 Multibyte to Unicode	Maps the multibyte character portion of TCHBIG5 to Unicode.
C1RSUNCD.txt	TCHBIG5_1R0 Single Byte to Unicode	Maps the single-byte character portion of TCHBIG5 to Unicode.

SCHINESE936_6R0

The SCHINESE936_6R0 character set provides support for simplified Chinese (Windows Code Page 936) on network-attached clients.

Under this character set, character data types support mixed single- and multibyte characters. The value of the first byte in a sequence distinguishes single-byte characters from double-byte characters.

IF the value of the first byte is ...	THEN the length of the character is ...
less than 0x81	one byte
greater than or equal to 0x81	two bytes

Valid Characters in SCHINESE936_6R0

The SCHINESE936_6R0 character set includes definitions for the single-byte characters corresponding to seven-bit ASCII.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
S6R0MUCD.txt	SCHINESE936_6R0 Multibyte to Unicode	Maps the multibyte character portion of SCHINESE936 to Unicode.
S6R0SUCD.txt	SCHINESE936_6R0 Single Byte to Unicode	Maps the single-byte character portion of SCHINESE936 to Unicode.

TCHINESE950_8R0

The TCHINESE950_8R0 character set provides support for traditional Chinese (Windows Code Page 950) on network-attached clients.

Under this character set, a CHARACTER data type supports mixed single- and multibyte characters. The high order bit of the first byte in a sequence distinguishes single-byte characters from double-byte characters.

IF the value of the first byte is ...	THEN the length of the character is ...
less than 0x81	one byte
greater than or equal to 0x81	two bytes

Valid Characters in TCHINESE950_8R0

The TCHINESE950_8R0 character set includes definitions for the single-byte characters corresponding to seven-bit ASCII.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
T8R0MUCD.txt	TCHINESE950_8R0 Multibyte to Unicode	Maps the multibyte character portion of TCHINESE950 to Unicode.
T8R0SUCD.txt	TCHINESE950_8R0 Single Byte to Unicode	Maps the single-byte character portion of TCHINESE950 to Unicode.

Korean Character Sets

Teradata supplies three multibyte character sets to support Korean on mainframe and network-attached clients.

Multibyte Character Set Encoding Forms

Each character set uses a specific encoding form to distinguish single-byte characters from multibyte characters.

Character Set	Description	Encoding Form
HANGULEBCDIC933_1II	Korean (IBM CCSID 933) for mainframe clients.	EBCDIC Shift-Out/Shift-In. Shift-out character 0x0E and shift-in character 0x0F bracket each string of double-byte characters.
HANGULKSC5601_2R4	Korean (mixed KSC5601) for network-attached clients.	Value of first byte in sequence distinguishes single-byte characters from double-byte characters.
HANGUL949_7R0	Korean (Windows Code Page 949) for network-attached clients.	Value of first byte in sequence distinguishes single-byte characters from double-byte characters.

Object Names

To determine whether a Korean character is valid in a Teradata Database object name:

1. Find the text file that maps the character set to UNICODE.

You can find the file here:

- a. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
 - b. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.
2. In the text file, find the Unicode character to which the client character in question maps.
 3. Find the file that identifies valid Unicode characters, UOBJNEXT.txt.

You can download the file here:

- a. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
 - b. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.
4. If the Unicode character appears in the file that is applicable to your system, you can use the client character that maps to it in an object name.

Note:

Korean Object Names containing multibyte Hangul characters are currently not supported in Teradata Database under any client character set.

Internal Storage Limitations

Character data entered using Korean client character sets should be stored in columns defined as Unicode. The UNICODE server character set requires two bytes of storage per character so that a CHAR(5) CHARACTER SET UNICODE field occupies 10 bytes of storage.

Given the 64000 byte limit on column size, a column cannot exceed 32000 characters. Furthermore, the combination of character data and other data types cannot exceed the 64000 byte limit on row size.

HANGULEBCDIC933_1II

The HANGULEBCDIC933_1II character set is based on IBM Coded Character Set ID (CCSID) 933, and provides support for Korean on mainframe clients.

Under this character set, character data types support mixed single-byte/multibyte characters.

Character Set Name	Description
HANGULEBCDIC933_1II	<p>IBM Shift-Out/Shift-In (SO/SI) protocol.</p> <p>Shift-out and shift-in characters distinguish single-byte characters from double-byte characters. The shift-out character is a byte with a value of 0x0E and the shift-in character is a byte with a value of 0x0F.</p> <p>A shift-out, shift-in pair must bracket each string of double-byte characters.</p>

Valid Characters in HANGULEBCDIC933_1II

The HANGULEBCDIC933_1II character set includes definitions for the single-byte characters in IBM Code Page (CPGID) 833, an EBCDIC code page similar to IBM Code Page (CPGID) 037.

The value of each byte of a double-byte character is in the range 0x41 through 0xFD, except for the IDEOGRAPHIC SPACE character, which is represented by 0x40 0x40.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
H1IMUNCD.txt	HANGULEBCDIC933_1II Multibyte to Unicode	Maps the multibyte character portion of HANGULEBCDIC933 to Unicode.
H1ISUNCD.txt	HANGULEBCDIC933_1II Single Byte to Unicode	Maps the single-byte character portion of HANGULEBCDIC933 to Unicode.

HANGULKSC5601_2R4

The HANGULKSC5601_2R4 character set provides support for Korean (mixed KSC5601) on network-attached clients.

Under this character set, character data types support mixed single-byte characters/multibyte characters. The high order bit of the first byte in a sequence distinguishes single-byte characters from double-byte characters.

IF the value of the first byte is ...	THEN the length of the character is ...
less than 0x81	one byte
greater than or equal to 0x81	two bytes

Valid Characters in HANGULKSC5601_2R4

The HANGULKSC5601_2R4 character set includes definitions for the single-byte characters corresponding to seven bit ASCII minus REVERSE SOLIDUS.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
H2RMUNCD.txt	HANGULKSC5601_2R4 Multibyte to Unicode	Maps the multibyte character portion of HANGULKSC5601 to Unicode.
H2RSUNCD.txt	HANGULKSC5601_2R4 Single Byte to Unicode	Maps the single-byte character portion of HANGULKSC5601 to Unicode.

HANGUL949_7R0

The HANGUL949_7R0 character set provides support for Korean (Windows Code Page 949) on network-attached clients.

Under this character set, character data types support mixed single-byte characters/multibyte characters. The value of the first byte in a sequence distinguishes single-byte characters from double-byte characters.

IF the value of the first byte is ...	THEN the length of the character is ...
less than 0x81	one byte
greater than or equal to 0x81	two bytes

Valid Characters in HANGUL949_7R0

The HANGUL949_7R0 character set includes definitions for the single-byte characters corresponding to seven bit ASCII minus REVERSE SOLIDUS.

The character mapping tables are described in the following text files.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

File Name (on CD)	Title (on the Web)	Description
H7R0MUCD.txt	HANGUL949_7R0 Multibyte to Unicode	Maps the multibyte character portion of HANGUL949 to Unicode.
H7R0SUCD.txt	HANGUL949_7R0 Single Byte to Unicode	Maps the single-byte character portion of HANGUL949 to Unicode.

Single-Byte International Character Sets

Teradata supplies translation codes for the following single-byte international client character sets. Teradata also provides files showing the valid single-byte characters for each set, and the conversion from the character set to UNICODE.

You can download the files here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

Character Set Name	Description	File Name
Arabic1256_6A0	Windows Code Page 1256 (Arabic)	A6A0SUCD.txt
Cyrillic1251_2A0	Windows Code Page 1251 (Cyrillic and Russian)	C2A0SUCD.txt
Hebrew1255_5A0	Windows Code Page 1255 (Hebrew)	H5A0SUCD.txt
Latin1250_1A0	Windows Code Page 1250 (Czech, Croatian, Albanian, Hungarian, Polish, Romanian, and Serbian)	L1A0SUCD.txt
Latin1254_7A0	Windows Code Page 1254 (Turkish)	L3A0SUCD.txt
Latin1258_8A0	Windows Code Page 1258 (Vietnamese)	L7A0SUCD.txt
Thai874_4A0	Windows Code Page 874 (Thai)	L8A0SUCD.txt

Object Names

To determine characters that are valid in object names:

1. Find the text file that maps the client character set to UNICODE.

You can find the file here:

- a. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
- b. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

2. In the text file, find the Unicode character to which the client character in question maps.

3. Find the file, UOBNEXT.txt, that lists the Unicode characters valid in object names.

You can download the file here:

- a. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
- b. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

4. If the Unicode character appears in the file that is applicable to your system, you can use the client character that maps to it in an object name.

Client Character Information in System Tables

The following user views provide access to the Teradata Database system tables containing client character set information.

View Name	Contents of the Underlying System Table
DBC.CharTranslationsV	International client character set definitions, including the character set names and translation codes that the Teradata Database uses to translate character data between client and server forms.
DBC.CharSetsV	Names of character sets that are flagged as the character sets to activate during a restart of the Teradata Database. These character sets are currently active and available to specify at the session level only if a restart of the Teradata Database was performed after the flags were set.
DBC.HostsInfoV	The default client character set name for each host, if any.

You may need to access these tables to:

- Activate a client character set to make it available at the session level.
- Define your own site-defined or extended site-defined client character sets.
- Change the default client character set for a host.

For details on activating client character sets and changing the default client character set for a host, see [Using Standard Teradata Client Character Sets](#).

Accessing Client Character Set Information

IF you want to...	THEN...
display the client character set currently in effect for your session	use the HELP SESSION statement.
review the contents of the system tables associated with language support	query the DBC.HostsInfoV, DBC.CharSetsV, or DBC.CharTranslationsV views.

Usage Notes

The presence of a character set in DBC.CharSetsV or DBC.CharTranslationsV does not mean it is active. The InstallFlag column in DBC.CharSetsV or DBC.CharTranslationsV is a good indication, but can be misleading if the table has been changed without performing a restart of the Teradata Database.

ASCII, EBCDIC, UTF16, and UTF8 character sets, which are predefined and permanently loaded into the system, are always available. The character sets are not listed in DBC.CharTranslationsV.

Using Standard Teradata Client Character Sets

Overview

The following describes how to use the standard Teradata-supplied client character set translation functions described in [Client Character Set Options](#)

You can also create and enable custom (site-defined) character sets. See [Site-Defined Client Character Sets](#) and [Extended Site-Defined Client Character Sets](#)

Implementation Process

The requirements for using client character sets vary by character set.

- To use one of the four permanently enabled client character sets (ASCII, EBCDIC, UTF8, and UTF16), enable the character set as a client default. See [System Determination of Client Character Set](#).
- To use a Teradata standard client character set that is not permanently enabled:
 1. Enable the character set in the DBC.CharTranslationsV view. See [Enabling a Client Character Set](#).
 2. Enable the character set as a client default. See [System Determination of Client Character Set](#).

About the DBC.CharTranslationsV View

To be available for use, a client character set must be listed in the DBC.CharTranslationsV view and the Install Flag must be set to Y.

Note:

The ASCII, EBCDIC, UTF8 and UTF16 client character sets are permanently enabled and do not appear in DBC.CharTranslationsV.

For more information on DBC.CharTranslationsV, see *Teradata Vantage™ - Data Dictionary*, B035-1092.

Example DBC.CharTranslationsV

The following example shows the DBC.CharTranslationsV view in the default state, that is, with none of the standard Teradata client character set translation codes enabled.

```
SELECT * FROM DBC.CharTranslationsV;
*** Query completed. 29 rows found. 7 columns returned.
*** Time was 3.60 seconds.
  CharSetName      CharSetId  InstallFlag E2I
  -----
  -----
```

KATAKANAEBDIC	111	N	000102039409967F..
SCHINESE936_6R0	115	N	0001020304050607..
HANGULKSC5601_2R4	120	N	0001020304050607..
EBCDIC273_0E	97	N	000102039C09867F..
EBCDIC037_0E	96	N	000102039C09867F..
KANJIEBDIC5035_0I	113	N	000102039C09867F..
LATIN1_0A	126	N	0001020304050607..
TCHBCDIC937_3IB	110	N	0001020304050607..
KANJISJIS_0S	119	N	0001020304050607..
LATIN1252_3A0	123	N	0001020304050607..
KANJIEBDIC5026_0I	112	N	000102039C09867F..
HEBREW1255_5A0	103	N	0001020304050607..
ARABIC1256_6A0	104	N	0001020304050607..
KANJIEUC_0U	118	N	0001020304050607..
LATIN9_0A	125	N	0001020304050607..
CYRILLIC1251_2A0	102	N	0001020304050607..
THAI874_4A0	100	N	0001020304050607..
LATIN1250_1A0	101	N	0001020304050607..
HANGUL949_7R0	116	N	0001020304050607..
SCHGB2312_1T0	121	N	0001020304050607..
LATIN1254_7A0	105	N	0001020304050607..
SCHEBCDIC935_2IJ	109	N	0001020304050607..
TCHINESE950_8R0	117	N	0001020304050607..
HANGULEBCDIC933_1II	108	N	0001020304050607..
EBCDIC277_0E	98	N	000102039C09867F..
LATIN1258_8A0	106	N	0001020304050607..
LATIN1252_0A	124	N	0001020304050607..
TCHBIG5_1R0	122	N	0001020304050607..
KANJI932_1S0	114	N	0001020304050607..

Explanation of Example

Field	Attributes	Description
CharSetName (TranslateName)	VARCHAR(128) NOT NULL	A name that uniquely identifies a character set. Character set names should start with an unaccented uppercase Roman letter and consist of only unaccented uppercase Roman letters, digits, underscore, dollar sign, or number sign.
CharSetID (TranslateID)	BYTEINT NOT NULL	A number that uniquely identifies the character set for which these translation codes are being inserted. Each character set ID must be in the range 65–126.
InstallFlag (Install)	CHAR(1) NOT NULL	A value of either Y or N, where:

Field	Attributes	Description
		<ul style="list-style-type: none"> Y= Enables the translation codes for the CharSetName at the next restart of the Teradata Database. N=The translation codes for the CharSetName are disabled. <p>A maximum of 16 character sets can be enabled at one time.</p>
E2I (TranslateIn)	BYTE(256) NOT NULL	The external (host)-to-internal (Teradata Database) codes for each character in the set.
E2IUp (TranslateInUp)	BYTE(256) NOT NULL	The external-to-internal uppercase codes for each character in the set.
I2E (TranslateOut)	BYTE(256) NOT NULL	The internal (Teradata Database)-to-external (client) codes for each character in the set.
I2EUp (TranslateOutUp)	BYTE(256) NOT NULL	The internal-to-external uppercase codes for each character in the set.

Enabling a Client Character Set

1. Set the InstallFlag column to Y for all character sets you want to activate. The limit is 16 active character sets.

```
UPDATE DBC.CharTranslationsV SET InstallFlag='Y' WHERE CharSetId
= charactersetId;
```

or

```
UPDATE DBC.CharTranslationsV SET InstallFlag='Y' WHERE CharSetName
= 'charactersetName';
```

2. Perform a full restart of Teradata Database, using the tpareset utility.

Note:

If more than 16 rows are flagged as Y at tpareset, the system sorts CharSetName values in ascending ASCII sequence and enables the rows in alphabetical order until the 16 set limit is reached.

For usage information, see tpareset in *Teradata Vantage™ - Database Utilities*, B035-1102.

About the Default Client Character Set

Specifying the default client character set (session character set) identifies to Teradata Database which translation codes to use when converting the client request into the server character set for processing, and when returning an answer set to client.

System Determination of Client Character Set

Connection Type	How the system determines the Session Character Set
CLLv2	<p>You can specify the client character set in several ways. The system determines the client character set for a session according to the following order of precedence. Each higher numbered item, if it exists, supersedes the previous item.</p> <p>For Mainframe connections (TDP) the following determines the default character set:</p> <ol style="list-style-type: none"> 1. The default character set is defined by the Teradata Database for the Logical Host. 2. The default may be overridden for a particular use of CLLv2 by specifying it in the CLLv2 HSHSPB file provided for that purpose. 3. Applications may override these defaults by explicitly specifying the character set for a particular use. <p>For Gateway connections the following determines the default character set:</p> <ol style="list-style-type: none"> 1. If no default client character set is specified, the system defaults to ASCII for gateway connections. 2. You can specify the default character set for client host groups by HostID in the DBC. HostsInfoV view. See Assigning the Default Client Character Set by HostId. 3. Users can specify the default character set on the client computer when setting up application connections to the database. See Assigning the Default Client Character Set During Client Setup. 4. You can specify the character set for a session using application settings, which override any defaults set by other methods. See Assigning the Default Client Character Set for a Session.
Teradata JDBC Driver	<p>The Teradata JDBC Driver determines the session character set solely by the CHARSET connection parameter. If the CHARSET connection parameter is omitted, the Teradata JDBC Driver uses the ASCII session character set. See Assigning the Default Client Character Set for a Session.</p>
ODBC Driver for Teradata	<p>An application or a user can specify the session character set option in the connection string for the connection. See Assigning the Default Client Character Set During Client Setup.</p>
.NET Data Provider for Teradata	<p>An application or a user can specify the session character set option in the connection string for the connection. The .NET Data Provider defaults the session character set to ASCII when an application or user does not specify a session character set connection string option. See Assigning the Default Client Character Set During Client Setup.</p>

Assigning the Default Client Character Set During Client Setup

You can specify the default character set when setting up client connections to the database.

Connection Type	Method of Specifying Default
CLIV2 Connections	On network-attached clients: Specify the character set in the clipsb.bat, using the character set ID shown in the DBC.CharTranslationsV view. For more information, see <i>Teradata® Call-Level Interface Version 2 Reference for Workstation-Attached Systems</i> , B035-2418.
	On mainframe clients: Specify the character set in the System Parameter Block (HSHSPB) on the client (parameters IBCCSN, IBOSCS and IBCCSC). For more information, see <i>Teradata® Call-Level Interface Version 2 Reference for Mainframe-Attached Systems</i> , B035-2417.
ODBC Driver Connections	Use the Session Character Set parameter of the Teradata ODBC driver setup dialog box to set the default client character set. For more information, see <i>ODBC Driver for Teradata® User Guide</i> .
JDBC Driver Connections	Specify the default character set in the CHARSET connection parameter for the Teradata JDBC driver. For more information, see <i>Teradata JDBC Driver Reference</i> , available at https://teradata-docs.s3.amazonaws.com/doc/connectivity/jdbc/reference/current/frameset.html .
.NET Data Provider for Teradata	Use the Session Character Set connection-string attribute to set the client character set. See <i>.NET Data Provider for Teradata Developer's Guide</i> , available at https://teradata-docs.s3.amazonaws.com/doc/connectivity/tdnetdp/16.20/help/webframe.html .

Assigning the Default Client Character Set by HostId

If clients connect to the Teradata Database system using multiple host connections, you can assign a default character set for each HostID, by adding entries to the DBC.HostsInfoV view.

About the DBC.HostsInfoV View

The DBC.HostsInfoV displays default character set assignments by HostID. See [Assigning Default Client Character Sets in DBC.HostsInfoV](#).

Example DBC.HostsInfoV

```
SELECT * FROM DBC.HostsInfoV;
*** Query completed. Nine rows found. 3 columns returned.
```

```
*** Time was 2.14 seconds.
```

```
LogicalHostId HostName DefaultCharSet
```

```
-----
      136 VM      SWEDISH_EBCDIC
      137 VM      German_EBCDIC
      138 VM      Spanish_EBCDIC
      139 VM      French_EBCDIC
      140 VM      Italian_EBCDIC
      141 VM      Norwegian_EBCDIC
      142 VM      EBCDIC037_0E
    1076 PC_LAN    LATIN1_0A
```

Explanation of Example DBC.HostsInfoV

Field	Data Type	Comment
LogicalHostId	SMALLINT	Identifies the client connection (mainframe or LAN) for which a default character set is being defined. This value must be the same as the identifier established for this connection by the hardware configuration data. For details on how to determine the value, see Finding the LogicalHostId .
HostName	VARCHAR(128)	An arbitrary, user-assigned name for this host, for example, Mainframe1, TokyoLAN, or RetailWebsite.
DefaultCharSet	VARCHAR(128)	The name of the session character set to be used by default when passing data to and from the clients in the host group. The value in this column may be: <ul style="list-style-type: none"> A permanently enabled client character set <ul style="list-style-type: none"> EDCDIC ASCII UTF8 UTF16 A client character set name from the CharSetName column in DBC.CharTranslationsV for which the Install Flag is set to Y.

Finding the LogicalHostId

To assign a default character set for a host group, you must know the LogicalHostID for the group. The LogicalHostId is a 16-bit value equal to the host number + the client CPU architecture number.

You can find:

- All the LogicalHostIds for a database system using the Configuration utility LIST HOST command. The output is similar to:

```
HOSTs  ARRAY
```

```
-----
```

HostNo	LogicalHostID	Type	Connected PE Range
-----	-----	----	-----
52	1076	COP	16379-16383
821	821	IBM	16377
829	829	IBM	16378

For more information, see *Teradata Vantage™ - Database Utilities*, B035-1102.

- The LogicalHostId through which an individual client logs on to the database by querying the DBC.SessionInfoV view, for example:

```
SELECT hostno, logicalhostid, username FROM DBC.SESSIONINFOV;
```

The output is similar to:

```
HostNo LogicalHostId UserName
-----
1      2049          MRN01END
```

Assigning Default Client Character Sets in DBC.HostsInfoV

After determining the LogicalHostId of the host for which you want to specify a default character set, you can make the assignments in the DBC.HostsInfoV view.

- Use an INSERT statement to add a row in the DBC.HostsInfoV view for each host, for example:

```
INS INTO DBC.HostsInfoV (1076, 'PC_LAN', 'LATIN1_0A');
```

where:

Syntax Element	Description.
1076	LogicalHostId
'PC_LAN'	Host name
'LATIN1_0A'	Default character set

- Perform a restart of the Teradata Database, using the tpareset utility, to activate the defaults.

IF the restart...	THEN...
successfully activates the default character sets	the character set name and identifier for each client system are sent to the appropriate TDP during startup.
encounters errors with a default character set	the system defaults to this host character set: <ul style="list-style-type: none"> EBCDIC for mainframe clients ASCII for all other clients

3. Go to `/var/log/messages` to check for errors associated with loading default character sets into `DBC.HostsInfoV`.

When the system attempts to validate the rows in `DBC.HostsInfoV`, if a character set in the view is not enabled, the system logs an error in the `/var/adm/messages` log, for example:

```
3133: Default Character Set is invalid.
On Fri Apr 18 10:51:05 2003 on NODE 001-01, VPROC 16383, partition 13,
task disabrt
Default Character Set name: 'Portuguese_EBCDIC'
```

4. If the log contains any errors, correct the problems and re-run steps 2 and 3.

Assigning the Default Client Character Set for a Session

You can control the session parameter that defines the character set to be used for external representation from within the client application submitting the request. Specifying an alternate character set overrides the default client character set for a session.

For example, in BTEQ you can use the `SESSION CHARSET` command to set the character set for the current session:

```
.SET SESSION CHARSET 'ascii'
```

Make sure the character set you use (for example, 'ascii') exactly matches the name of a valid, enabled client character set.

For further information on how to use a session character set, see the users guide for the particular application.

Server Character Sets

About Character Data Storage

The Teradata Database language support mode defines the global character set default, which determines how object names and user data are stored in the database.

IF you enable this language support mode ...	Teradata Database stores object names using ...	and the user default server character set is...
Standard	UNICODE	LATIN
Japanese	UNICODE	UNICODE

You can optionally specify the default character set for individual users and table columns, which overrides the global default character set for storing user data.

See [Specifying the Default Server Character Set](#).

Working with Server Character Set Defaults

In addition to the global default server character set indicated by the language mode (Standard or Japanese), you can specify the default character set for specific table columns and users using SQL DDL statements.

Default Server Character Set Options

You can specify one of the following character sets as the default. Make sure that the specified default is compatible with the client character sets with which it interacts.

- LATIN
- UNICODE
- KANJISJIS

Note:

You cannot specify GRAPHIC or KANJI1 as the default server character set.

Specifying the Default Server Character Set

You can specify the default server character set:

- For the data stored in a specific table column, using the CHARACTER SET option in a CREATE TABLE or ALTER TABLE statement.
- For a user, using the DEFAULT CHARACTER SET options in a CREATE USER or MODIFY USER statement.

The default character set for data storage is determined in the following order. If a default is not defined, the system defers to the next lower item on the list:

1. Table column default
2. User default
3. System default (as defined by the language mode)

LATIN Server Character Set

The Teradata LATIN server character set is an extension of the R2 international character set. This extension is designed to provide support for the ISO 8859-1 (Latin1) and ISO 8859-15 (Latin9) repertoires.

This internal Teradata Database form-of-use is also the client form-of-use when the client character set is ASCII (although seven-bit ASCII is not affected).

Client Character Sets Supported with LATIN

Clients accessing the LATIN server character set can use any of the following single-byte client character sets:

- ASCII
- LATIN1_0A
- LATIN9_0A
- LATIN1252_0A
- EBCDIC
- EBCDIC037_0E
- EBCDIC277_0E
- EBCDIC273_0E
- Simple custom (site-defined) client character sets based ASCII or EBCDIC

For details on LATIN-compatible client character sets, see [Client Character Set Options](#).

Differences Between ASCII and Teradata LATIN

When used in the Teradata Database, ASCII and Teradata LATIN are identical on all code points except the 80-FF range, where Teradata LATIN defines additional West European letters.

Code points outside the seven-bit ASCII range result in data that may not behave as intended.

Detecting ASCII and Teradata LATIN Differences

You can detect code points that conflict with ASCII semantics either before or after migrating to Teradata LATIN. Corrective action can then be taken under either environment.

```
SELECT cchar
FROM tab1
WHERE
INDEX(cchar, '80'XC) <> 0
OR INDEX(cchar, '81'XC) <> 0
OR INDEX(cchar, '82'XC) <> 0
.
/* all codes 83-FE */
.
OR INDEX(cchar, 'FF'XC) <> 0;
```

Using LATIN Characters Above ASCII X'7F'

If you intend to use characters with code points above ASCII X'7F', take the following steps:

1. Install, as appropriate, LATIN1_0A, LATIN9_0A, EBCDIC037_0E, or a site-defined character set that matches the character set used on your client.
2. Set the client character set to LATIN1_0A, LATIN9_0A, EBCDIC037_0E, or the name of the site-defined character set that you installed in Step 1 in order to use the encodings.

The preceding applies if non-ASCII characters are desired.

ISO 8859-1 (Latin1) Support

Teradata Database Japanese language facilities support ISO 8859-1 (Latin1) with the restriction that the code points 80-85 are not allowed.

The characters at 80-85 are in the C1 control region and are not considered to be essential.

ISO 8859-15 (Latin9) Support

Teradata Database Japanese language facilities support ISO 8859-15 (Latin9) with the restriction that the code points 80-85 and the LETTER Z WITH CARON are not allowed.

The characters at 80-85 are in the C1 control region and are not considered to be essential. The character LETTER Z WITH CARON is part of the support for Finnish, but is not considered part of the Finnish alphabet and is rarely used even in Finland.

Supported Characters

For supported LATIN characters, see LATIN Server Character Set text file, latin_server.txt.

You can find the file here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

Related Information

For additional information on the LATIN server character set, see *Teradata Vantage™ - Data Types and Literals*, B035-1143.

UNICODE Server Character Set

The Unicode standard is a 16-bit encoding of virtually all characters in all current world languages.

Sharing Data Among Heterogeneous Clients

UNICODE is a canonical character set. Data stored as UNICODE can be shared among heterogeneous clients.

Supported Client Character Sets

UNICODE supports all client character sets.

Definition

Teradata Database supports the BMP characters from Unicode® 6.0, referred to as UNICODE. For a list of the supported characters, see the UNICODE Server Character Set text file, UNCDUNCD.txt.

You can find the file here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.

Note:

Teradata allows the storage and searching of all other Unicode characters. See [Unicode Pass Through](#).

Areas of Unicode

It is often useful to divide the UNICODE character set into eight areas.

Area	Description
General scripts	Latin, Greek, Cyrillic, Hebrew, Arabic, Indic, and other characters.
Symbols	Arrows, mathematical symbols, and punctuation.
CJK Phonetics and Symbols	Hiragana, Katakana, and Bopomofo. CJK stands for Chinese, Japanese, and Korean.
CJK Ideographs	Chinese, Japanese, and Korean ideographs.
Hangul Syllables	Complete set of modern Hangul.
Surrogates	Code points designed to extend the range of Unicode within the ISO 10646 encoding scheme.
Private Use Area	The Private Use Area contains characters for sharing site-defined characters from the KanjiEBCDIC, KanjiEUC, and KanjiShift-JIS client character sets.
Compatibility Zone	The Compatibility Zone contains halfwidth and fullwidth variants of characters defined by Japanese standards and, among others, includes Hankaku (halfwidth) Katakana and fullwidth ASCII characters.

2nd Byte of UNICODE	
00	ASCII (0000-007F) Latin (0080-00FF)
	General Scripts (0000-19FF)
20	Symbols (2000-2BFF)
2C	Glagolitic, Coptic, Georgian, Tifinagh, Ethiopic (2C00-2DDF)
2E	Symbols and Chinese Radicals (2E00-33FF)
34	CJK Ideographic Extension A (3400-4DFF)
4E	CJK Ideographs (4E00-9FFF)
A0	Yi Syllables (A000-A4C6)
A9	Undefined Area
AC	Hangul Syllables (AC00-D7A3)
D8	Surrogates (D800-DFFF)
E0	Private Use (E000-F8FF)
F9	Compatibility and Specials (F900-FFFF)
FF	Fullwidth ASCII Halfwidth Katakana

The first 1880 characters of the Private Use Area are used for sharing site-defined characters from KanjiEBCDIC, KanjiEUC, and KanjiShift-JIS client character sets.

The following table defines the Teradata Database use of the Private Use Area.

Name	Unicode Range	KanjiShift-JIS Range	KanjiEUC Range	KanjiEBCDIC Range	Comments
Gaiji-1	U+ E000 to U+ E3AB	0xF040 to 0xF4FC (1st 940 Gaiji)	0xF5A1 to 0xFEFE (Rows 85-94 JIS X 0208)	0x6941 to 0x6DF4 (1st 940 Gaiji)	Shared by all. (940 characters)
Gaiji-2	U+ E3AC to U+ E757	0xF540 to 0xF9FC (2nd 940 Gaiji)	0x8FF5A1 to 0x8FFFEFE (Rows 85-94 JIS X 0212)	0x6DF5 to 0x72EA (2nd 940 Gaiji)	Shared by all. (940 characters)

Name	Unicode Range	KanjiShift-JIS Range	KanjiEUC Range	KanjiEBCDIC Range	Comments
Graphic Error Character	U+ F8FF	Not applicable	Not applicable	Not applicable	Associated with the VARGRAPHIC function

Related Information

For additional characteristics of the UNICODE server character set, see *Teradata Vantage™ - Data Types and Literals*, B035-1143.

KANJI1 Character Set

The KANJI1 server character set is designed for Japanese applications that must remain compatible with Teradata Kanji data from Teradata Database releases prior to V2R3.0.

The semantics and limitations of KANJI1 are identical to those of the CHARACTER server character set in releases prior to V2R3.0.

NOTICE

In accordance with Teradata internationalization plans, KANJI1 support is deprecated and is to be discontinued in the near future. KANJI1 is not allowed as a default character set; the system changes the KANJI1 default character set to the UNICODE character set. Creation of new KANJI1 objects is highly restricted. Although many KANJI1 queries and applications may continue to operate, sites using KANJI1 should convert to another character set as soon as possible.

See [KANJI1 Restrictions](#).

Mapping Standard

The JIS X 0201 mapping standard is the basis for the KANJI1 server character set.

Sharing Data Among Heterogeneous Clients

Because it stores client character data in the form-of-use in which it receives it, KANJI1 is non-canonical and cannot be shared among heterogeneous clients.

KANJI1 Restrictions

NOTICE

In accordance with Teradata internationalization plans, KANJI1 support is deprecated and is to be discontinued in the near future. KANJI1 is not allowed as a default character set; the system changes the KANJI1 default character set to the UNICODE character set. Creation of new KANJI1 objects is highly restricted. Although many KANJI1 queries and applications may continue to operate, sites using KANJI1 should convert to another character set as soon as possible.

Note:

Upon upgrading to Teradata Database 14.0 or greater, the system automatically replaces DEFAULT CHARACTER SET KANJI1 with DEFAULT CHARACTER SET UNICODE in existing user definitions.

As part of the plans for discontinuing Kanji1 support, the creation of new Kanji1 objects is highly restricted. Inclusion of the phrase CHARACTER SET KANJI1 in the following statements returns a syntax error:

- CREATE USER/MODIFY USER
- CREATE TABLE/ALTER TABLE
- CREATE FUNCTION/REPLACE FUNCTION
- CREATE TYPE/ALTER TYPE
- CREATE PROCEDURE/REPLACE PROCEDURE
- CREATE MACRO/REPLACE MACRO
- CREATE VIEW/REPLACE VIEW
- CAST function

Plan to use the TRANSLATE function to convert existing Kanji1 data to Unicode or another supported server character set. For details, see “TRANSLATE” in *Teradata Vantage™ - SQL Functions, Expressions, and Predicates*, B035-1145.

Supported Client Character Sets

The KANJI1 server character set is designed to support Japanese characters only when using specifically designed client character sets.

KANJI1 supports single-byte characters from the following client character sets:

- EBCDIC
- ASCII
- UTF8

Only single-byte characters that are coded the same in the LATIN and JIS X 0201 standard can be used.

KANJI1 supports mixed single- and multibyte characters from the following client character sets:

- KanjiEBCDIC
 - Single-byte data uses JIS X 0201
 - Double-byte data uses SO/SI
- KanjiSJIS_0S
 - JIS X 0201
 - JIS X 0208
- Kanji1932_1S0
 - JIS X 0201
 - JIS X 0208
- KanjiEUC_0U
 - JIS X 0201
 - Bytes 00-7F as defined
 - Byte 8E mapped to 80
 - JIS X 0208 (converted to KanjiSJIS)
 - JIS X 0212
 - Byte 8F mapped to FF

Only a limited set of characters stored as KANJI1 can be retrieved correctly by a client character set other than the one that entered the data. This limited set includes most 7-bit ASCII characters:

- The letters A-Z and a-z
- The digits 0-9
- Various punctuation, symbols, and control characters

In general, it does not include:

- Japanese characters
- The backslash (\)
- Yen sign
- Tilde
- Overline

Attempts to retrieve Japanese or other non-sharable characters from a KANJI1 field may result in error messages for the following character sets:

- UTF8
- UTF16
- Most site-defined client character sets using map files

For other character sets, attempts to retrieve Japanese or other non-sharable characters from a KANJI1 field may result in improperly translated data.

KANJI1 Character Expectations

KANJI1 character data is usually a mixture of single- and multibyte characters. Therefore, even when a session uses a non-Japanese client character set, such as standard EBCDIC or ASCII, certain character configurations are interpreted either as starting a multibyte character string or as control characters.

The following interpretations apply to name and data characters for all client character sets when the server character set is KANJI1.

Characters with a client encoding of ...	Are ...
0x0E 0x0F	interpreted as the Shift-Out or Shift-In character respectively, which delimit the start or end of a multibyte character string.
0x80 0xFF	translated internally into characters that are reserved for the ss ₂ and ss ₃ escape characters of KanjiEUC data.

Related Information

For details about the KANJI1 server character set, see *Teradata Vantage™ - Data Types and Literals*, B035-1143.

KANJISJIS Character Set

The KANJISJIS server character set is designed for Japanese applications that rely on the semantics of KanjiSJIS for characteristics such as string length and physical space allocation.

Sharing Data Among Heterogeneous Clients

KANJISJIS is a canonical character set. Data stored as KANJISJIS can be shared among heterogeneous clients.

Supported Client Character Sets

KANJISJIS supports mixed single and multibyte characters and its form-of-use is identical with that of its client character set analog.

Limitations

Storage space for this server character set is allocated on a byte basis rather than a character basis. This means that updates to a column defined as KANJISJIS can truncate because of this physical space.

SQL string manipulations of KANJISJIS data operate on a logical character basis and are performed by means of temporary conversion to Unicode. For this reason, KANJISJIS strings are limited to 32000 characters.

Related Information

For details about the KANJISJIS server character set, see *Teradata Vantage™ - Data Types and Literals*, B035-1143.

GRAPHIC Character Set

The GRAPHIC server character set is designed for IBM DB2 compatibility.

Sharing Data Among Heterogeneous Clients

GRAPHIC is a canonical character set. Data stored as GRAPHIC can be shared among heterogeneous clients.

Supported Client Character Sets

GRAPHIC supports multibyte characters from the defined Graphic repertoire of the following character sets:

- KanjiEBCDIC
- KanjiSJIS_OS
- KanjiEUC_0U

Characters stored as type GRAPHIC cannot be any of the following:

- ASCII
- Halfwidth Japanese
- Various special characters

GRAPHIC data is encoded as Unicode. Because of this, GRAPHIC data from Teradata Database releases prior to V2R3.0 must be migrated to the V2R3.0 server character set.

Unsupported Client Character Sets

GRAPHIC is designed to support client character sets where excluding non-double-byte characters is desired. This is not true of UTF8, Chinese, or Korean.

Moreover, the GRAPHIC pad character is three bytes in UTF8 so that standard padding techniques cannot be used.

For these reasons, the GRAPHIC server character set is not compatible with UTF8, Chinese, or Korean sessions.

Restriction on Specification of GRAPHIC as the Default Character Set

You can specify GRAPHIC as the character set for a table column, but not as the default character set for a user.

Related Information

For details about the GRAPHIC server character set, see *Teradata Vantage™ - Data Types and Literals*, B035-1143.

Site-Defined Client Character Sets

Overview

If an external client character set has characters that do not appear in the corresponding Teradata-supplied client character set translation codes, you can create your own site-defined client character set.

The following sections describes how to create and implement custom single-byte client character sets, based on the ASCII and EBCDIC character sets.

For information on how to create and implement character sets that client character set that define non-western European characters and require custom character mapping, see [Extended Site-Defined Client Character Sets](#).

About Custom Single-Byte Client Character Sets

If the character set you use on a client system is represented by a single-byte character set, and the Teradata-supplied ASCII-compatible or EBCDIC-compatible character sets do not provide all the characters you need, you can create a custom (site-defined) client character set.

A custom client character set that you create using the procedures in the following sections must be able to translate to the Teradata LATIN server character set.

Example Character Sets

Teradata ships the following single-byte character sets that you can use as examples when creating a custom single-byte client character set:

- EBCDIC037_0E
- EBCDIC273_0E
- EBCDIC277_0E
- LATIN1_0A
- LATIN9_0A
- LATIN1252_0A

Custom Japanese Character Sets

If the single-byte component of the Teradata-supplied Japanese client character sets do not provide all the characters you need, you can create your own site-defined client character set. The single-byte component of the client character set you create must be able to translate to the JIS X 0201 standard.

To create a client character set that defines single- and multibyte Japanese characters and translates in the Teradata Database to Unicode, you must create an extended site-defined character set. For details, see [Extended Site-Defined Client Character Sets](#).

Implementing Custom Character Sets

1. Create the character set by defining:
 - The external-to-internal and internal-to-external hexadecimal codes for both lowercase and uppercase translation for each character in the set.
 - A unique character set name.
 - A unique identifier (use a value in the range 65 to 126).
 See [Creating a Custom Client Character Set](#).
2. Load the character set information into the DBC.CharTranslationsV. See [Loading a Custom Client Character Set](#).
3. Enable the character set. See [Activating a Client Character Set](#).

Note:

A maximum of 16 character sets can be active at one time.

4. Optionally assign the character set as the default. See the section beginning with [System Determination of Client Character Set](#).

Creating a Custom Client Character Set

To create a custom character set, do the following:

1. Observe applicable restrictions. See [Restrictions on Redefining Characters](#).
2. Define the character set translation code. See [Defining the Custom Character Translation Codes](#).
3. Name the character set. See [Naming a Custom Character Set](#).

Restrictions on Redefining Characters

Site-defined character set restrictions for networked and mainframe clients are described in the following sections.

Note:

Network clients use ASCII-compatible character sets and mainframe clients are often EBCDIC-based.

ASCII

Site-defined character sets based on the ASCII character set must preserve the definition of code points 0-127. This means that you may only redefine those characters whose numeric value is 128 or greater.

EBCDIC

When you create a custom character set based on the EBCDIC character set, you must retain a certain subset of the EBCDIC characters. The required characters are identified as shaded code points in the table.

The shaded code points must remain as defined, and the site-defined translation tables must map these code points into the correct internal Teradata Database representations.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0					S P	&	-									0
1							/		a	j			A	J		1
2									b	k	s		B	K	S	2
3									c	l	t		C	L	T	3
4									d	m	u		D	M	U	4
5									e	n	v		E	N	V	5
6									f	o	w		F	O	W	6
7									g	p	x		G	P	X	7
8									h	q	y		H	Q	Y	8
9									i	r	z		I	R	Z	9
A								:								
B					.	\$,	#								
C					<	*	%	@								
D					()	-	'								
E					+	;	>	=								
F							¬ ^a	?	"							

a. The logical not sign ¬ may also be shown as a caret (^).

Defining the Custom Character Translation Codes

1. Copy the existing translation codes for the Teradata character set that is closest to the one you want to define
2. Make changes.
3. Save.

Collation

When you create a custom character translation code/character set, the Teradata Database automatically generates a collation matching the order of the site-defined character set. To use the collation when you use the site-defined character set, choose the CHARSET_COLL collation.

Naming a Custom Character Set

Each custom character set must have a unique name (CharSetName), 1-30 characters in length, in the form:

```
CHARACTERSETNAME_0X
```

where:

CHARACTERSETNAME identifies the character set.

- The name must begin with an unaccented, uppercase Roman letter.
- The rest of the name can use only unaccented uppercase Roman letters, digits, the underscore, the dollar sign, or the number sign.

_0x is the suffix (required), composed of a _ (LOWLINE) character, followed by a numeric character and then an alpha character.

- The numeric character must be zero. Other values are reserved for future use.
- The alpha character indicates the type of encoding used in the character set, as shown in the following table:

Code	Description
a	Single-byte character ASCII.
e	Single-byte character EBCDIC.
i	IBM (SO/SI style) mixed character single- and multibyte characters.
u	EUC mixed character single-byte characters/multibyte characters.
s	Shift-JIS mixed character single-byte characters/multibyte characters and graphic multibyte characters.
b-d, f-h, j-r, t, v-z	Single- and multibyte extended site-defined character sets.

For example, the suffix _0i in the name KANJIEBCDIC5026_0i implies IBM mixed single- and multibyte characters.

Common character set names, such as ASCII, EBCDIC and UTF8 cannot be used for site-defined character sets. It also recognizes the special character set name KATAKANAEDBCDIC that is used to specify characteristics similar to the suffix "_0i". Other site-defined character set names are given characteristics as if the name ends in either "_0a" or "_0e".

Referencing a Client Character Set Name

The suffix is part of the client character set name. The complete name must be included in any references to the character set, for example, in a BTEQ SET SESSION CHARSET command.

Enabling a Custom Client Character Set

You perform the following setup tasks, using the DBC.CharTranslationsV view, to enable a client character set:

1. [Loading a Custom Client Character Set](#).
2. [Activating a Client Character Set](#)

About the DBC.CharTranslationsV View

To be available for use, a client character set must be listed in the DBC.CharTranslationsV view.

All Teradata-supplied client character sets are listed by default in the DBC.CharTranslationsV, except the four defaults, ASCII, EBCDIC, UTF8 and UTF16..

Example DBC.CharTranslationsV Example DBC.CharTranslationsV

Note:

The example that follows shows only the first 4 of 8 columns. The last 4 columns are translation codes similar to the E2I column. For details, see the explanation table that follows the example.

```
SELECT * FROM DBC.CharTranslationsV;
*** Query completed. 29 rows found. 7 columns returned.
*** Time was 3.60 seconds.
```

CharSetName	CharSetId	InstallFlag	E2I
KATAKANAEBCDIC	111	N	000102039409967F..
SCHINESE936_6R0	115	N	0001020304050607..
HANGULKSC5601_2R4	120	N	0001020304050607..
EBCDIC273_0E	97	N	000102039C09867F..
EBCDIC037_0E	96	N	000102039C09867F..
KANJIEBCDIC5035_0I	113	N	000102039C09867F..
LATIN1_0A	126	N	0001020304050607..
TCHEBCDIC937_3IB	110	N	0001020304050607..
KANJISJIS_0S	119	N	0001020304050607..
LATIN1252_3A0	123	N	0001020304050607..
KANJIEBCDIC5026_0I	112	N	000102039C09867F..
HEBREW1255_5A0	103	N	0001020304050607..
ARABIC1256_6A0	104	N	0001020304050607..
KANJIEUC_0U	118	N	0001020304050607..
LATIN9_0A	125	N	0001020304050607..
CYRILLIC1251_2A0	102	N	0001020304050607..
THAI874_4A0	100	N	0001020304050607..

LATIN1250_1A0	101	N	0001020304050607..
HANGUL949_7R0	116	N	0001020304050607..
SCHGB2312_1T0	121	N	0001020304050607..
LATIN1254_7A0	105	N	0001020304050607..
SCHEBCDIC935_2IJ	109	N	0001020304050607..
TCHINESE950_8R0	117	N	0001020304050607..
HANGULEBCDIC933_1II	108	N	0001020304050607..
EBCDIC277_0E	98	N	000102039C09867F..
LATIN1258_8A0	106	N	0001020304050607..
LATIN1252_0A	124	N	0001020304050607..
TCHBIG5_1R0	122	N	0001020304050607..
KANJI932_1S0	114	N	0001020304050607..

Explanation of Example

Field	Attributes	Description
CharSetName (TranslateName)	VARCHAR(128) NOT NULL	A name that uniquely identifies a character set. Japanese character set names should start with an unaccented uppercase Roman letter and consist of only unaccented uppercase Roman letters, digits, underscore, dollar sign, or number sign.
CharSetID (TranslateID)	BYTEINT NOT NULL	A number that uniquely identifies the character set for which these translation codes are being inserted. Each character set ID must be in the range 65–126.
InstallFlag (Install)	CHAR(1) NOT NULL	A value of either Y or N, where: <ul style="list-style-type: none"> Y= Enables the translation codes for the CharSetName at the next restart of the Teradata Database. N=The translation codes for the CharSetName are disabled. A maximum of 16 character sets can be enabled at one time. If more than 16 rows are flagged as Y at reset time, the system sorts CharSetName values in ascending ASCII sequence and enables the rows in alphabetical order until the 16 set limit is reached.
E2I (TranslateIn)	BYTE(256) NOT NULL	The external (host)-to-internal (Teradata Database) codes for each character in the set.
E2IUp (TranslateInUp)	BYTE(256) NOT NULL	The external-to-internal uppercase codes for each character in the set.
I2E (TranslateOut)	BYTE(256) NOT NULL	The internal (Teradata Database)-to-external (client) codes for each character in the set.
I2EUp (TranslateOutUp)	BYTE(256) NOT NULL	The internal-to-external uppercase codes for each character in the set.

For more information on DBC.CharTranslationsV, see *Teradata Vantage™ - Data Dictionary*, B035-1092.

Loading a Custom Client Character Set

You must load a character set before you can activate it for use.

1. Log on to the Teradata Database as a user with INSERT privileges on database DBC.
2. Use the INSERT statement to load each needed character set into the DBC.CharTranslationsV view, using a BTEQ batch file or other method.

Ensure that the client configuration matches the translation entry and table just created.

Note:

There is no limit on the number of character sets you can load onto your system, but no more than 16 character sets can be active at a time.

Example Load Script

This example shows a BTEQ batch job used to insert a row defining the translation codes for a Finnish/Swedish character set. The external hexadecimal codes are EBCDIC and agree with Character Sequence Number 24 for an IBM 3270 Control Unit displaying to an IBM 3270 Information Display System.

```

/*****
/* This is the Finnish/Swedish Character Set Codes for IBM 3270 */
/* information display system. The IBM manuals title "IBM 3270 */
/* Information Display System. 3270 Control Unit" and "IBM 3270 */
/* Information Display System. Character Set Reference". */
/* Document numbers GA23-0065-6 and GA27-2837-9 correspondingly. */
/* The sequence number for this character set is 24 - see 2-55 */
/* of GA23-0065-6 and 10-17 and 10-45 of GA27-2837-9. Here is */
/* the map between DBC and IBM character codes: */
/*
/*      Character Name          DBC   IBM      */
/*                               HEX   HEX      */
/*
/*      Lowercase a with umlaut (diaresis)...E4    C0    */
/*      Uppercase A with umlaut (diaresis)...C4    7B    */
/*      Lowercase a with ring.....E5            D0    */
/*      Uppercase A with ring.....C5            5B    */
/*      Lowercase e with acute accent.....E9     79    */
/*      Uppercase E with acute accent.....C9     E0    */
/*      Lowercase o with umlaut (diaresis)...F6    05    */
/*      Uppercase O with umlaut (diaresis)...D6    7C    */
/*      Lowercase u with umlaut (diaresis)...FC    A1    */
*****/

```

```

INSERT INTO DBC.CharTranslationsV
( CharSetName
, CharSetId
, InstallFlag
, E2I
, E2IUp
, I2E
, I2EUp )
VALUES
( 'SWEDISH_EBCDIC'
, 71
, 'Y'
/* TranslateIn.  External -> Internal      */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F          */
, '00010203A8F6A97FD1D2D30B0C0D0E0F'xb/* 0 */
|| '10111213D4D508401819D7D81C1D1E1F'xb/* 1 */
|| '7DE6D9E7E80A171B60F0F1F2F3050607'xb/* 2 */
|| 'F4F51609F75CF804F9E2E37B1415E01A'xb/* 3 */
|| '208182838485868788898B2E3C282B8D'xb/* 4 */
|| '2691929394959697989921C52A293B5E'xb/* 5 */
|| '2D2F9B9DA2A3A4A5A6A77C2C255F3E3F'xb/* 6 */
|| 'C1C2C32324C6C7C87EE93AC4D6273D22'xb/* 7 */
|| '806162636465666768698AC08CAD8E8F'xb/* 8 */
|| '906A6B6C6D6E6F7071729AD09CBD9E9F'xb/* 9 */
|| 'A0FC737475767778797AAAABAC5BAEAF'xb/* A */
|| 'B0B1B2B3B4B5B6B7B8B9BABBBBC5DBEBF'xb/* B */
|| 'E4414243444546474849CACBCCCDCECF'xb/* C */
|| 'E54A4B4C4D4E4F505152DADBDCDDDEDF'xb/* D */
|| 'C9E1535455565758595AEAEBECEDEEEF'xb/* E */
|| '30313233343536373839FAFBA1FDFEFF'xb/* F */
/* TranslateInUp.  External -> Internal      */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F          */
, '00010203A8D6A97FD1D2D30B0C0D0E0F'xb/* 0 */
|| '10111213D4D508401819D7D81C1D1E1F'xb/* 1 */
|| '7DE6D9E7E80A171B60F0F1F2F3050607'xb/* 2 */
|| 'F4F51609F75CF804F9E2E37B1415E01A'xb/* 3 */
|| '208182838485868788898B2E3C282B8D'xb/* 4 */
|| '2691929394959697989921C52A293B5E'xb/* 5 */
|| '2D2F9B9DA2A3A4A5A6A77C2C255F3E3F'xb/* 6 */
|| 'C1C2C32324C6C7C87EC93AC4D6273D22'xb/* 7 */
|| '804142434445464748498AC08CAD8E8F'xb/* 8 */
|| '904A4B4C4D4E4F5051529AD09CBD9E9F'xb/* 9 */
|| 'A0FC535455565758595AAAABAC5BAEAF'xb/* A */
|| 'B0B1B2B3B4B5B6B7B8B9BABBBBC5DBEBF'xb/* B */

```

```

||'C4414243444546474849CACBCCCDCECF'xb/* C */
||'C54A4B4C4D4E4F505152DADBDCDDDEDF'xb/* D */
||'C9E1535455565758595AEAEBECEDEEEF'xb/* E */
||'30313233343536373839FAFBA1FDFEFF'xb/* F */
/* TranslateOut. Internal -> External */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
, '00010203372D2E2F1633250B0C0D0E0F'xb/* 0 */
||'101112133C3D322618193F271C1D1E1F'xb/* 1 */
||'405A7F73746C507D4D5D5C4E6B604B61'xb/* 2 */
||'F0F1F2F3F4F5F6F7F8F97A5E4C7E6E6F'xb/* 3 */
||'17C1C2C3C4C5C6C7C8C9D1D2D3D4D5D6'xb/* 4 */
||'D7D8D9E2E3E4E5E6E7E8E9AD35BD5F6D'xb/* 5 */
||'28818283848586878889919293949596'xb/* 6 */
||'979899A2A3A4A5A6A7A8A93B6A207807'xb/* 7 */
||'804142434445464748498A4A8C4F8E8F'xb/* 8 */
||'905152535455565758599A629C639E9F'xb/* 9 */
||'A0FC6465666768690406AAABAC8DAEAF'xb/* A */
||'B0B1B2B3B4B5B6B7B8B9BABBBBC9DBEBF'xb/* B */
||'8B7071727B5B757677E0CACBCCCDCECF'xb/* C */
||'9B08090A14157C1A1B22DADBDCDDDEDF'xb/* D */
||'3EE1393AC0D021232479EAEBECEDEEEF'xb/* E */
||'292A2B2C303105343638FAFBA1FDFEFF'xb/* F */
/* TranslateOutUp. Internal -> External */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
, '00010203372D2E2F1633250B0C0D0E0F'xb/* 0 */
||'101112133C3D322618193F271C1D1E1F'xb/* 1 */
||'405A7F73746C507D4D5D5C4E6B604B61'xb/* 2 */
||'F0F1F2F3F4F5F6F7F8F97A5E4C7E6E6F'xb/* 3 */
||'17C1C2C3C4C5C6C7C8C9D1D2D3D4D5D6'xb/* 4 */
||'D7D8D9E2E3E4E5E6E7E8E9AD35BD5F6D'xb/* 5 */
||'28C1C2C3C4C5C6C7C8C9D1D2D3D4D5D6'xb/* 6 */
||'D7D8D9E2E3E4E5E6E7E8E93B6A207807'xb/* 7 */
||'804142434445464748498A4A8C4F8E8F'xb/* 8 */
||'905152535455565758599A629C639E9F'xb/* 9 */
||'A0FC6465666768690406AAABAC8DAEAF'xb/* A */
||'B0B1B2B3B4B5B6B7B8B9BABBBBC9DBEBF'xb/* B */
||'8B7071727B5B757677E0CACBCCCDCECF'xb/* C */
||'9B08090A14157C1A1B22DADBDCDDDEDF'xb/* D */
||'3EE1393A7B5B212324E0EAEBECEDEEEF'xb/* E */
||'292A2B2C30317C343638FAFBA1FDFEFF'xb/* F */
);

```

Activating a Client Character Set

After loading the needed character sets into the DBC.CharTranslationsV view, you must activate the character set.

1. Set the InstallFlag column to Y for the character set you want to activate. The limit is 16 active character sets.

```
UPDATE DBC.CharTranslationsV SET InstallFlag='Y' WHERE CharSetId  
= custom_characterset_id;
```

or

```
UPDATE DBC.CharTranslationsV SET InstallFlag='Y' WHERE CharSetName  
= 'custom_characterset_name';
```

2. Perform a full restart of Teradata Database, using the tpareset utility. For usage information, see tpareset in *Teradata Vantage™ - Database Utilities*, B035-1102.

Extended Site-Defined Client Character Sets

Overview

The following sections describe how to create and implement extended site-defined client character sets.

About Extended Character Sets

An extended site-defined character set defines the mapping of hexadecimal values to characters for single- and multibyte components of a character set. If you use a non-Western European language such as Russian, Arabic, or Urdu, you can define and install your own single-byte character set. If you use a non-Western European language such as Japanese, Korean, or Chinese, and the Teradata-supplied character sets are not entirely sufficient for your site, you can define and install your own multibyte character set.

Extended site-defined character sets can support, with certain constraints, any subset of the Unicode repertoire.

A user who is sufficiently privileged can define the relevant client character set, mapping bytes from the client to their corresponding Unicode values.

Internal Storage Limitations

Character data entered using extended site-defined client character sets should be stored in columns defined as Unicode. The UNICODE server character set requires two bytes of storage per character so that a CHAR(5) CHARACTER SET UNICODE field occupies 10 bytes of storage.

Given the 64000 byte limit on column size, a column cannot exceed 32000 characters. Furthermore, the combination of character data and other data types cannot exceed the 64000 byte limit on row size.

Collations

A user who is sufficiently privileged can also define an appropriate collation.

If a custom collation is required, and CHARSET_COLL collation does not produce the desired result, then you can modify the MULTINATIONAL collation. For information, see [MULTINATIONAL Collation for Extended Site-Defined Character Sets](#).

Privileges Required to Define Character Sets and Collations

Only sufficiently privileged users can define extended site-defined character sets and collations. A sufficiently privileged user is one who can:

- Edit and place files in the appropriate directories on every node in the Teradata Database.

- Modify records in DBC.CharTranslationsV and DBC.CollationsV.
- Restart the Teradata Database.

Examples to Use in Constructing Extended Site-Defined Character Sets

Teradata ships the following characters sets that you can use as examples of extended site-defined character sets:

- SCHEBCDIC935_2IJ
- TCHEBCDIC937_3IB
- HANGULEBCDIC933_1II
- SCHGB2312_1T0
- TCHBIG5_1R0
- HANGULKSC5601_2R4
- KANJI932_1S0
- LATIN1252_3A0
- LATIN1250_1A0
- LATIN1254_7A0
- LATIN1258_8A0
- SCHINESE936_6R0
- HANGUL949_7R0
- HEBREW1255_5A0
- ARABIC1256_6A0
- CYRILLIC1251_2A0
- THAI1874_4A0
- TCHINESE950_8R0

Implementing Extended Character Sets

To implement an extended site-defined client character set, you must create an entry in DBC.CharTranslationsV and a corresponding mapping file.

DBC.CharTranslationsV Entry

To define an extended site-defined client character set, you must create an entry in DBC.CharTranslationsV.

Follow the rules and procedures in [Implementing Custom Character Sets](#), with the following exceptions:

- The name of the character set follows the conventions described in [Naming Character Sets](#).
- Rather than mapping the client character set to the LATIN or KANJI1 internal form, the mapping is defined in terms of an internal transition form.

- For all extended site defined client character sets, the E2IUp and I2EUp fields in DBC.CharTranslationsV are ignored; the uppercase properties of the character set are derived automatically.

Mapping File

You must also create a corresponding file on every node in one of the following directories in Teradata Database:

- TPA etc

You can identify the location of the directory by using `pdepath -e` on the command line.

- TPA cfg

You can identify the location of the directory by using `pdepath -c` on the command line.

If the mapping file is stored in the TPA etc directory, during version change, any changes to this mapping file are lost. To avoid losing the changes, the best practice is to store the mapping and collation files in the TPA cfg directory, which is not updated during the installation of a new version of Teradata.

A map file definition in the cfg overrides the definition in etc.

The file defines the mapping between an internal transition form and Unicode for the character set. A single definition file may be used to support multiple client character sets. For example, if there is a need to handle both extended EBCDIC and extended ASCII character sets for support of the language, a single file may be used to define the internal transitional form. Two entries might be placed in the DBC.CharTranslationsV view that define the conversion between the client character set and the internal transitional form.

Naming Character Sets

Choosing a Mnemonic for an Extended Single-Byte Character Set

Choose a letter as a mnemonic for the group of character sets being defined. Use the mnemonic as the last letter in the suffix of the character set name and in the name of the mapping file. If you choose 'x', for example, the suffix of the character set name is `_0x`, and the name of the mapping file is `map_0x`. The system uses the letter as a link between the name of the client character set, as specified in DBC.CharTranslationsV, and the name of the mapping file.

For example, Cyrillic character sets can use the mnemonic 'c'. If `CYRILLIC_0C` is the name of a client character set, the system uses the internal transitional form specified in `map_0c`.

If you omit the mnemonic in the name of the character set, the system does not look for a mapping file, and assumes the character set is based on standard Latin characters. A client character set named `CYRILLIC` does not produce Cyrillic characters, although it may appear to be working correctly due to the display characteristics of the client.

Note:

The characters 'a', 'e', 'i', 'r', 's', 't', and 'u' are reserved and cannot be used for extended single-byte character set mapping file names.

Choosing a Name for a Chinese Multibyte Character Set

If the Teradata-defined Chinese character sets described in [Chinese Character Sets](#) are not appropriate for your site, you can define your own character sets using the following names, IDs, and encodings.

Character Set Name	ID	Description
SDSCHEBCDIC935_6IJ	75	Simplified Chinese for mainframe clients. The encoding form is EBCDIC Shift-Out/Shift-In, where the shift-out character 0x0E and shift-in character 0x0F bracket zero or more double-byte characters.
SDTCHEBCDIC937_7IB	76	Traditional Chinese for mainframe clients. The encoding form is EBCDIC Shift-Out/Shift-In, where the shift-out character 0x0E and shift-in character 0x0F bracket zero or more double-byte characters.
SDSCHGB2312_2T0	94	Simplified Chinese for network-attached clients. The encoding form is Extended UNIX Code (EUC), composed of two code sets: cs0 for single-byte characters and cs1 for double-byte characters.
SDTCHBIG5_3R0	95	Traditional Chinese for network-attached clients. The value of the first byte in a sequence distinguishes single-byte characters from double-byte characters.

The system uses the two characters following the underscore (_) in the character set name as a link to the mapping file you create in the TPA etc or TPA cfg directory. The name of the mapping file must start with "map_" and end with the first two characters following the underscore in the character set name.

For example, if you define a character set for SDTCHBIG5_3R0, you must create a mapping file named map_3R that provides the translation tables between the transitional forms and Unicode.

Choosing a Name for a Korean Multibyte Character Set

If the Teradata-defined Korean character sets described in [Korean Character Sets](#) are not appropriate for your site, you can define your own character sets using the following names, IDs, and encodings.

Character Set Name	ID	Description
SDHANGULEBCDIC933_5II	74	Korean for mainframe clients. The encoding form is EBCDIC Shift-Out/Shift-In, where the shift-out character 0x0E and shift-in character 0x0F bracket zero or more double-byte characters.

Character Set Name	ID	Description
SDHANGULKSC5601_4R4	93	Korean for network-attached clients. The value of the first byte in a sequence distinguishes single-byte characters from double-byte characters.

The system uses the two characters following the underscore (`_`) in the character set name as a link to the mapping file you create in the TPA etc or TPA cfg directory. The name of the mapping file must start with “map_” and end with the first two characters following the underscore in the character set name.

For example, if you define a character set for SDHANGULKSC5601_4R4, you must create a mapping file named map_4R that provides the translation tables between the transitional forms and Unicode.

Choosing a Name for a Japanese Multibyte Character Set

If the Teradata-defined Japanese character sets described in [Japanese Client Character Set Support](#) are not appropriate for your site, you can define your own character sets using the following names, IDs, and encodings.

Character Set Name	ID	Description
SDKATAKANAEBCDIC_4IF	77	Japanese Katakana EBCDIC for mainframe clients. The encoding form is EBCDIC Shift-Out/Shift-In, where the shift-out character 0x0E and shift-in character 0x0F bracket zero or more double-byte characters.
SDKANJIEBCDIC5026_4IG	78	IBM Japanese Extended Katakana character set for mainframe clients. The high order bit of the first byte in a sequence distinguishes single-byte characters from double-byte characters.
SDKANJIEBCDIC5035_4IH	79	IBM Japanese Extended English character set for mainframe clients. The high order bit of the first byte in a sequence distinguishes single-byte characters from double-byte characters.
SDKANJIEUC_1U3	91	Japanese character set for network-attached clients that is compatible with the UNIX operating system. The encoding form is Extended UNIX Code (EUC), composed of four code sets: cs0 for one-byte characters, cs1 and cs2 for two-byte characters, and cs3 for three-byte characters.
SDKANJISJIS_1S3	92	Windows-compatible Japanese character set for network-attached clients. The first byte in a sequence distinguishes single-byte characters from double-byte characters.

The system uses the two characters following the underscore (`_`) in the character set name as a link to the mapping file you create in the TPA etc or TPA cfg directory. The name of the mapping file must start with “map_” and end with the first two characters following the underscore in the character set name.

For example, if you define a character set for SDKANJISJIS_1S3, you must create a mapping file named map_1S that provides the translation tables between the transitional forms and Unicode.

Mapping File for a Single-Byte Character Set

The mapping file for a new single-byte client character set, which the user creates, should contain up to four translation tables, analogous to the translation tables in DBC.CharTranslationV.

- The first table defines the translation from the single-byte transitional form to the double-byte Unicode form.
- The second table defines the translation from the single-byte transitional form to the corresponding uppercase double-byte Unicode form.

Only characters that are mapped differently from the first table should be included.

- The third table is optional, and allows for the user to create a many-to-one mapping from Unicode to the external character set if desired.

If the third table is not present, then an inverse mapping is computed from the translation defined in the first table.

- The fourth table defines the translation from double-byte Unicode form to the corresponding upper case single-byte internal transitional form. Only characters that are mapped differently from the optional third table (or the inverse mapping computed from the first table if the third table is not present) should be included.

STATEMACHINE Statement

The mapping file for a single-byte client character set can optionally contain the statement:

```
#STATEMACHINE SBC
```

Note:

If a mapping file does not specify the STATEMACHINE statement, then the default is STATEMACHINE SBC.

The STATEMACHINE statement describes the encoding form of the character set. STATEMACHINE SBC means that the character set is a single-byte character set.

Format of the Mapping File

The format of the mapping file is multiple lines, with each line terminated by a linefeed character. This may be problematic for editors that expect carriage-return or carriage-return followed by linefeed to terminate a line.

Note:

Linefeed termination is the UNIX convention. Carriage-return linefeed is the Windows convention.

Each translation table starts with a #BEGINMAP line and ends with an #ENDMAP line. Other than those two commands, the # indicates the start of a comment that continues to the end of line. Blank lines are ignored.

The #BEGINMAP and #ENDMAP statements also include the name of the map being defined.

For an example, see [Example Mapping File](#).

Mapping File for a Multibyte Character Set

The mapping file for a multibyte client character set contains single-byte and multibyte translation tables that define the mapping between an internal transition form and Unicode for the character set.

The single-byte translation tables are analogous to the E2I and I2E fields in the translation tables in DBC.CharTranslationV, which define the conversion between the client character set and the internal transition form.

STATEMACHINE Statement

The mapping file for a multibyte client character set must contain the following statement:

```
#STATEMACHINE  smachine
```

where the client character set name determines the value of *smachine*.

IF the character set name is...	THEN the value of <i>smachine</i> is...	AND the mapping file provides translation tables for a character set that uses this encoding form...
<ul style="list-style-type: none"> SDSCHEBCDIC935_6IJ SDTCHEBCDIC937_7IB SDKATAKANAEBBCDIC_4IF SDKANJIEBCDIC5026_4IG SDKANJIEBCDIC5035_4IH SDHANGULEBCDIC933_5II 	SOSI0E0F	EBCDIC Shift-Out/Shift-In. Shift-out character 0x0E and shift-in character 0x0F bracket zero or more double-byte characters.
<ul style="list-style-type: none"> SDTCHBIG5_3R0 SDHANGULKSC5601_4R4 	S81	The value of first byte in the sequence, which distinguishes single-byte characters from double-byte characters. If the value of the first byte is: <ul style="list-style-type: none"> less than 0x81, the length of the character is one byte equal to or greater than 0x81, the length of the character is 2 bytes

IF the character set name is...	THEN the value of smachine is...	AND the mapping file provides translation tables for a character set that uses this encoding form...
		Note: S80 is no longer supported. S80 behaves like S81. You should change your map files to reflect this change (distributed map files are changed by Teradata).
SDSCHGB2312_2T0	EUC1211	Extended UNIX Code (EUC), composed of two code sets: cs0 for single-byte characters and cs1 for double-byte characters
SDKANJISJIS_1S3	S80A1E0	The value of first byte in the sequence, which distinguishes single-byte characters from double-byte characters. If the value of the first byte is: <ul style="list-style-type: none"> less than 0x81, the length of the character is one byte equal to or greater than 0x81, the length of the character is 2 bytes greater than or equal to 0xA1 and less than 0xE0, the length of the character is one byte greater than or equal to 0xE0, the length of the character is 2 bytes
SDKANJIEUC_1U3	EUC1223	The encoding form is Extended UNIX Code (EUC), composed of four code sets: cs0 for one-byte characters, cs1 and cs2 for two-byte characters, and cs3 for three-byte characters.

Translation Tables

Translation tables map characters between an extended site-defined client character set and Unicode.

A mapping file must minimally provide translation tables that map characters from the client character set to Unicode.

A mapping file may optionally provide translation tables that map characters from Unicode to the client character set. If the optional translation tables are not defined, the system derives them by inverting the corresponding mandatory tables.

If one of the following conditions exists, however, the optional translation tables become mandatory:

- More than one character from the client character set maps to a single Unicode character.
- More than one Unicode character maps to a single character of the client character set.

BEGINMAP Statement

Each translation table starts with the following statement:

```
#BEGINMAP  table_name
```

and ends with the following statement:

```
#ENDMAP  table_name
```

where:

The value of *table_name* is determined by the client character set name.

IF the character set name is ...	THEN the mapping file must define these tables ...	AND optionally define these tables ...
SDHANGULEBCDIC933_5II	<ul style="list-style-type: none"> 5I_SBC_2_UNICODE 5I_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_5I_SBC UNICODE_2_5I_MBC
SDHANGULKSC5601_4R4	<ul style="list-style-type: none"> 4R_SBC_2_UNICODE 4R_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_4R_SBC UNICODE_2_4R_MBC
SDKANJIEBCDIC5026_4IG	<ul style="list-style-type: none"> 4I_SBC_2_UNICODE 4I_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_4I_SBC UNICODE_2_4I_MBC
SDKANJIEBCDIC5035_4IH	<ul style="list-style-type: none"> 4I_SBC_2_UNICODE 4I_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_4I_SBC UNICODE_2_4I_MBC
SDKANJIEUC_1U3	<ul style="list-style-type: none"> 1U_CS0_2_UNICODE 1U_CS1_2_UNICODE 1U_CS2_2_UNICODE 1U_CS3_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_1U_CS0 UNICODE_2_1U_CS1 UNICODE_2_1U_CS2 UNICODE_2_1U_CS3
SDKANJISJIS_1S3	<ul style="list-style-type: none"> 1S_SBC_2_UNICODE 1S_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_1S_SBC UNICODE_2_1S_MBC
SDKATAKANAEBDIC_4IF	<ul style="list-style-type: none"> 4I_SBC_2_UNICODE 4I_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_4I_SBC UNICODE_2_4I_MBC
SDSCHEBCDIC935_6IJ	<ul style="list-style-type: none"> 6I_SBC_2_UNICODE 6I_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_6I_SBC UNICODE_2_6I_MBC
SDSCHGB2312_2T0	<ul style="list-style-type: none"> 2T_CS0_2_UNICODE 2T_CS1_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_2T_CS0 UNICODE_2_2T_CS1
SDTCHBIG5_3R0	<ul style="list-style-type: none"> 3R_SBC_2_UNICODE 3R_MBC_2_UNICODE 	<ul style="list-style-type: none"> UNICODE_2_3R_SBC UNICODE_2_3R_MBC

IF the character set name is ...	THEN the mapping file must define these tables ...	AND optionally define these tables ...
SDTCHEBCDIC937_7IB	<ul style="list-style-type: none"> • 7I_SBC_2_UNICODE • 7I_MBC_2_UNICODE 	<ul style="list-style-type: none"> • UNICODE_2_7I_SBC • UNICODE_2_7I_MBC

Example

A mapping file named map_4R has the following statement:

```
#STATEMACHINE S81
```

and defines the following translation tables:

```
4R_SBC_2_UNICODE
4R_MBC_2_UNICODE
```

and optionally defines the following translation tables:

```
UNICODE_2_4R_SBC
UNICODE_2_4R_MBC
```

Format of the Mapping File

The format of the mapping file is multiple lines, with each line terminated by a linefeed character. This may be problematic for editors that expect carriage-return or carriage-return followed by linefeed to terminate a line.

Note:

Linefeed termination is the UNIX convention. Carriage-return linefeed is the Windows convention.

Use the # to start a comment that continues to the end of a line. Blank lines are ignored.

Example Mapping File

Here is an example map_0r file for Arabic/Urdu support. Note that the name of the map makes use of the mnemonic letter used in the file name, so the example map names would be modified depending on the name of the mapping file.

Between the #BEGINMAP and #ENDMAP lines are lines defining the translation.

The first hexadecimal number indicates the code being mapped from, and the second value indicates the code point being mapped to. The internal transitional form requires two hexadecimal digits, and the Unicode

form requires four hexadecimal digits. Internal transitional form must always map the bytes 0xXY as 0x00XY for values from 0x00 through 0x7F, except for the error character 0x1A that maps to Unicode 0xFFFFD.

Example File

```
#####
# Name: map_0r
#
# Description: Define the mapping arrays used by isfinitbcbx:
#             0R_2_UNICODE
#             0R_2_UNICODE_UC
#             UNICODE_2_0R_UC
#
# History: 1999Aug23 Create an Arabic mapping based upon code # page 1256
#####
#BEGINMAP 0R_2_UNICODE
#
#   Name: CP1256(Arabic)to the corresponding letters of Unicode.
#   Unicode version: 2.1
#   Format: Three tab-separated columns
#       Column #1 is the CP1256 input set (in hex as 0xXXXX)
#       Column #2 is the Unicode output set (in hex as 0xXXXX)
#       Column#3:the Unicode(ISO 10646)name(follows a comment sign)
#       Each name for column #1 & column #2, if difference
#       between them
#
#   History: 14Aug1999Created for V2R4
#
0x00      0x0000# <control>
0x01      0x0001# <control>
0x02      0x0002# <control>
0x03      0x0003# <control>
0x04      0x0004# <control>
0x05      0x0005# <control>
0x06      0x0006# <control>
0x07      0x0007# <control>
0x08      0x0008# <control>
0x09      0x0009# <control>
0x0A      0x000A# <control>
0x0B      0x000B# <control>
0x0C      0x000C# <control>
0x0D      0x000D# <control>
0x0E      0x000E# <control>
```

0x0F	0x000F# <control>
0x10	0x0010# <control>
0x11	0x0011# <control>
0x12	0x0012# <control>
0x13	0x0013# <control>
0x14	0x0014# <control>
0x15	0x0015# <control>
0x16	0x0016# <control>
0x17	0x0017# <control>
0x18	0x0018# <control>
0x19	0x0019# <control>
0x1A	0xFFFFD# <control> & REPLACEMENT CHARACTER
0x1B	0x001B# <control>
0x1C	0x001C# <control>
0x1D	0x001D# <control>
0x1E	0x001E# <control>
0x1F	0x001F# <control>
0x20	0x0020# SPACE
0x21	0x0021# EXCLAMATION MARK
0x22	0x0022# QUOTATION MARK
0x23	0x0023# NUMBER SIGN
0x24	0x0024# DOLLAR SIGN
0x25	0x0025# PERCENT SIGN
0x26	0x0026# AMPERSAND
0x27	0x0027# APOSTROPHE
0x28	0x0028# LEFT PARENTHESIS
0x29	0x0029# RIGHT PARENTHESIS
0x2A	0x002A# ASTERISK
0x2B	0x002B# PLUS SIGN
0x2C	0x002C# COMMA
0x2D	0x002D# HYPHEN-MINUS
0x2E	0x002E# FULL STOP
0x2F	0x002F# SOLIDUS
0x30	0x0030# DIGIT ZERO
0x31	0x0031# DIGIT ONE
0x32	0x0032# DIGIT TWO
0x33	0x0033# DIGIT THREE
0x34	0x0034# DIGIT FOUR
0x35	0x0035# DIGIT FIVE
0x36	0x0036# DIGIT SIX
0x37	0x0037# DIGIT SEVEN
0x38	0x0038# DIGIT EIGHT
0x39	0x0039# DIGIT NINE
0x3A	0x003A# COLON

0x3B	0x003B# SEMICOLON
0x3C	0x003C# LESS-THAN SIGN
0x3D	0x003D# EQUALS SIGN
0x3E	0x003E# GREATER-THAN SIGN
0x3F	0x003F# QUESTION MARK
0x40	0x0040# COMMERCIAL AT
0x41	0x0041# LATIN CAPITAL LETTER A
0x42	0x0042# LATIN CAPITAL LETTER B
0x43	0x0043# LATIN CAPITAL LETTER C
0x44	0x0044# LATIN CAPITAL LETTER D
0x45	0x0045# LATIN CAPITAL LETTER E
0x46	0x0046# LATIN CAPITAL LETTER F
0x47	0x0047# LATIN CAPITAL LETTER G
0x48	0x0048# LATIN CAPITAL LETTER H
0x49	0x0049# LATIN CAPITAL LETTER I
0x4A	0x004A# LATIN CAPITAL LETTER J
0x4B	0x004B# LATIN CAPITAL LETTER K
0x4C	0x004C# LATIN CAPITAL LETTER L
0x4D	0x004D# LATIN CAPITAL LETTER M
0x4E	0x004E# LATIN CAPITAL LETTER N
0x4F	0x004F# LATIN CAPITAL LETTER O
0x50	0x0050# LATIN CAPITAL LETTER P
0x51	0x0051# LATIN CAPITAL LETTER Q
0x52	0x0052# LATIN CAPITAL LETTER R
0x53	0x0053# LATIN CAPITAL LETTER S
0x54	0x0054# LATIN CAPITAL LETTER T
0x55	0x0055# LATIN CAPITAL LETTER U
0x56	0x0056# LATIN CAPITAL LETTER V
0x57	0x0057# LATIN CAPITAL LETTER W
0x58	0x0058# LATIN CAPITAL LETTER X
0x59	0x0059# LATIN CAPITAL LETTER Y
0x5A	0x005A# LATIN CAPITAL LETTER Z
0x5B	0x005B# LEFT SQUARE BRACKET
0x5C	0x005C# REVERSE SOLIDUS
0x5D	0x005D# RIGHT SQUARE BRACKET
0x5E	0x005E# CIRCUMFLEX ACCENT
0x5F	0x005F# LOW LINE
0x60	0x0060# GRAVE ACCENT
0x61	0x0061# LATIN SMALL LETTER A
0x62	0x0062# LATIN SMALL LETTER B
0x63	0x0063# LATIN SMALL LETTER C
0x64	0x0064# LATIN SMALL LETTER D
0x65	0x0065# LATIN SMALL LETTER E
0x66	0x0066# LATIN SMALL LETTER F

0x67	0x0067#	LATIN SMALL LETTER G
0x68	0x0068#	LATIN SMALL LETTER H
0x69	0x0069#	LATIN SMALL LETTER I
0x6A	0x006A#	LATIN SMALL LETTER J
0x6B	0x006B#	LATIN SMALL LETTER K
0x6C	0x006C#	LATIN SMALL LETTER L
0x6D	0x006D#	LATIN SMALL LETTER M
0x6E	0x006E#	LATIN SMALL LETTER N
0x6F	0x006F#	LATIN SMALL LETTER O
0x70	0x0070#	LATIN SMALL LETTER P
0x71	0x0071#	LATIN SMALL LETTER Q
0x72	0x0072#	LATIN SMALL LETTER R
0x73	0x0073#	LATIN SMALL LETTER S
0x74	0x0074#	LATIN SMALL LETTER T
0x75	0x0075#	LATIN SMALL LETTER U
0x76	0x0076#	LATIN SMALL LETTER V
0x77	0x0077#	LATIN SMALL LETTER W
0x78	0x0078#	LATIN SMALL LETTER X
0x79	0x0079#	LATIN SMALL LETTER Y
0x7A	0x007A#	LATIN SMALL LETTER Z
0x7B	0x007B#	LEFT CURLY BRACKET
0x7C	0x007C#	VERTICAL LINE
0x7D	0x007D#	RIGHT CURLY BRACKET
0x7E	0x007E#	TILDE
0x7F	0x007F#	<control>
0x80	0x20AC#	EURO SIGN
0x81	0x067E#	ARABIC LETTER PEH
0x82	0x201A#	SINGLE LOW-9 QUOTATION MARK
0x83	0x0192#	LATIN SMALL LETTER F WITH HOOK
0x84	0x201E#	DOUBLE LOW-9 QUOTATION MARK
0x85	0x2026#	HORIZONTAL ELLIPSIS
0x86	0x2020#	DAGGER
0x87	0x2021#	DOUBLE DAGGER
0x88	0x02C6#	MODIFIER LETTER CIRCUMFLEX ACCENT
0x89	0x2030#	PER MILLE SIGN
0x8A	0x0679#	ARABIC LETTER TTEH
0x8B	0x2039#	SINGLE LEFT-POINTING ANGLE QUOTATION MARK
0x8C	0x0152#	LATIN CAPITAL LIGATURE OE
0x8D	0x0686#	ARABIC LETTER TCHEH
0x8E	0x0698#	ARABIC LETTER JEH
0x8F	0x0688#	ARABIC LETTER DDAL
0x90	0x06AF#	ARABIC LETTER GAF
0x91	0x2018#	LEFT SINGLE QUOTATION MARK
0x92	0x2019#	RIGHT SINGLE QUOTATION MARK

0x93	0x201C#	LEFT DOUBLE QUOTATION MARK
0x94	0x201D#	RIGHT DOUBLE QUOTATION MARK
0x95	0x2022#	BULLET
0x96	0x2013#	EN DASH
0x97	0x2014#	EM DASH
0x98	0x06A9#	ARABIC LETTER KEHEH
0x99	0x2122#	TRADE MARK SIGN
0x9A	0x0691#	ARABIC LETTER RREH
0x9B	0x203A#	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
0x9C	0x0153#	LATIN SMALL LIGATURE OE
0x9D	0x200C#	ZERO WIDTH NON-JOINER
0x9E	0x200D#	ZERO WIDTH JOINER
0x9F	0x06BA#	ARABIC LETTER NOON GHUNNA
0xA0	0x00A0#	NO-BREAK SPACE
0xA1	0x060C#	ARABIC COMMA
0xA2	0x00A2#	CENT SIGN
0xA3	0x00A3#	POUND SIGN
0xA4	0x00A4#	CURRENCY SIGN
0xA5	0x00A5#	YEN SIGN
0xA6	0x00A6#	BROKEN BAR
0xA7	0x00A7#	SECTION SIGN
0xA8	0x00A8#	DIAERESIS
0xA9	0x00A9#	COPYRIGHT SIGN
0xAA	0x06BE#	ARABIC LETTER HEH DOACHASHMEE
0xAB	0x00AB#	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
0xAC	0x00AC#	NOT SIGN
0xAD	0x00AD#	SOFT HYPHEN
0xAE	0x00AE#	REGISTERED SIGN
0xAF	0x00AF#	MACRON
0xB0	0x00B0#	DEGREE SIGN
0xB1	0x00B1#	PLUS-MINUS SIGN
0xB2	0x00B2#	SUPERSCRIP TWO
0xB3	0x00B3#	SUPERSCRIP THREE
0xB4	0x00B4#	ACUTE ACCENT
0xB5	0x00B5#	MICRO SIGN
0xB6	0x00B6#	PILCROW SIGN
0xB7	0x00B7#	MIDDLE DOT
0xB8	0x00B8#	CEDILLA
0xB9	0x00B9#	SUPERSCRIP ONE
0xBA	0x061B#	ARABIC SEMICOLON
0xBB	0x00BB#	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
0xBC	0x00BC#	VULGAR FRACTION ONE QUARTER
0xBD	0x00BD#	VULGAR FRACTION ONE HALF
0xBE	0x00BE#	VULGAR FRACTION THREE QUARTERS

0xBF	0x061F#	ARABIC QUESTION MARK
0xC0	0x06C1#	ARABIC LETTER HEH GOAL
0xC1	0x0621#	ARABIC LETTER HAMZA
0xC2	0x0622#	ARABIC LETTER ALEF WITH MADDA ABOVE
0xC3	0x0623#	ARABIC LETTER ALEF WITH HAMZA ABOVE
0xC4	0x0624#	ARABIC LETTER WAW WITH HAMZA ABOVE
0xC5	0x0625#	ARABIC LETTER ALEF WITH HAMZA BELOW
0xC6	0x0626#	ARABIC LETTER YEH WITH HAMZA ABOVE
0xC7	0x0627#	ARABIC LETTER ALEF
0xC8	0x0628#	ARABIC LETTER BEH
0xC9	0x0629#	ARABIC LETTER TEH MARBUTA
0xCA	0x062A#	ARABIC LETTER TEH
0xCB	0x062B#	ARABIC LETTER THEH
0xCC	0x062C#	ARABIC LETTER JEEM
0xCD	0x062D#	ARABIC LETTER HAH
0xCE	0x062E#	ARABIC LETTER KHAH
0xCF	0x062F#	ARABIC LETTER DAL
0xD0	0x0630#	ARABIC LETTER THAL
0xD1	0x0631#	ARABIC LETTER REH
0xD2	0x0632#	ARABIC LETTER ZAIN
0xD3	0x0633#	ARABIC LETTER SEEN
0xD4	0x0634#	ARABIC LETTER SHEEN
0xD5	0x0635#	ARABIC LETTER SAD
0xD6	0x0636#	ARABIC LETTER DAD
0xD7	0x00D7#	MULTIPLICATION SIGN
0xD8	0x0637#	ARABIC LETTER TAH
0xD9	0x0638#	ARABIC LETTER ZAH
0xDA	0x0639#	ARABIC LETTER AIN
0xDB	0x063A#	ARABIC LETTER GHAIN
0xDC	0x0640#	ARABIC TATWEEL
0xDD	0x0641#	ARABIC LETTER FEH
0xDE	0x0642#	ARABIC LETTER QAF
0xDF	0x0643#	ARABIC LETTER KAF
0xE0	0x00E0#	LATIN SMALL LETTER A WITH GRAVE
0xE1	0x0644#	ARABIC LETTER LAM
0xE2	0x00E2#	LATIN SMALL LETTER A WITH CIRCUMFLEX
0xE3	0x0645#	ARABIC LETTER MEEM
0xE4	0x0646#	ARABIC LETTER NOON
0xE5	0x0647#	ARABIC LETTER HEH
0xE6	0x0648#	ARABIC LETTER WAW
0xE7	0x00E7#	LATIN SMALL LETTER C WITH CEDILLA
0xE8	0x00E8#	LATIN SMALL LETTER E WITH GRAVE
0xE9	0x00E9#	LATIN SMALL LETTER E WITH ACUTE
0xEA	0x00EA#	LATIN SMALL LETTER E WITH CIRCUMFLEX

```

0xEB      0x00EB# LATIN SMALL LETTER E WITH DIAERESIS
0xEC      0x0649# ARABIC LETTER ALEF MAKSURA
0xED      0x064A# ARABIC LETTER YE
0xEE      0x00EE# LATIN SMALL LETTER I WITH CIRCUMFLEX
0xEF      0x00EF# LATIN SMALL LETTER I WITH DIAERESIS
0xF0      0x064B# ARABIC FATHATAN
0xF1      0x064C# ARABIC DAMMATAN
0xF2      0x064D# ARABIC KASRATAN
0xF3      0x064E# ARABIC FATHA
0xF4      0x00F4# LATIN SMALL LETTER O WITH CIRCUMFLEX
0xF5      0x064F# ARABIC DAMMA
0xF6      0x0650# ARABIC KASRA
0xF7      0x00F7# DIVISION SIGN
0xF8      0x0651# ARABIC SHADDA
0xF9      0x00F9# LATIN SMALL LETTER U WITH GRAVE
0xFA      0x0652# ARABIC SUKUN
0xFB      0x00FB# LATIN SMALL LETTER U WITH CIRCUMFLEX
0xFC      0x00FC# LATIN SMALL LETTER U WITH DIAERESIS
0xFD      0x200E# LEFT-TO-RIGHT MARK
0xFE      0x200F# RIGHT-TO-LEFT MARK
0xFF      0x06D2# ARABIC LETTER YEH BARREE
#ENDMAP0R_2_UNICODE
#BEGINMAP0R_2_UNICODE_UC
#
#      Name:The lower case letters of CP1256 (Arabic) to the corresponding
#              upper case letters of Unicode.
#      Unicode version:  2.1
#      Format:Three tab-separated columns
#          Column #1 is the CP1256 input set (in hex as 0xFFFF)
#          Column #2 is the Unicode output set (in hex as 0xFFFF)
#          Column #3 is the Unicode(ISO 10646) name(follows comment sign)
#          Each name for column #1 & column #2, if difference
#          between them
#
#      History: 14Aug1999Created for V2R4
#
0x61      0x0041# LATIN SMALL LETTER A & LATIN CAPITAL LETTER A
0x62      0x0042# LATIN SMALL LETTER B & LATIN CAPITAL LETTER B
0x63      0x0043# LATIN SMALL LETTER C & LATIN CAPITAL LETTER C
0x64      0x0044# LATIN SMALL LETTER D & LATIN CAPITAL LETTER D
0x65      0x0045# LATIN SMALL LETTER E & LATIN CAPITAL LETTER E
0x66      0x0046# LATIN SMALL LETTER F & LATIN CAPITAL LETTER F
0x67      0x0047# LATIN SMALL LETTER G & LATIN CAPITAL LETTER G
0x68      0x0048# LATIN SMALL LETTER H & LATIN CAPITAL LETTER H

```

0x69	0x0049#	LATIN SMALL LETTER I & LATIN CAPITAL LETTER I
0x6A	0x004A#	LATIN SMALL LETTER J & LATIN CAPITAL LETTER J
0x6B	0x004B#	LATIN SMALL LETTER K & LATIN CAPITAL LETTER K
0x6C	0x004C#	LATIN SMALL LETTER L & LATIN CAPITAL LETTER L
0x6D	0x004D#	LATIN SMALL LETTER M & LATIN CAPITAL LETTER M
0x6E	0x004E#	LATIN SMALL LETTER N & LATIN CAPITAL LETTER N
0x6F	0x004F#	LATIN SMALL LETTER O & LATIN CAPITAL LETTER O
0x70	0x0050#	LATIN SMALL LETTER P & LATIN CAPITAL LETTER P
0x71	0x0051#	LATIN SMALL LETTER Q & LATIN CAPITAL LETTER Q
0x72	0x0052#	LATIN SMALL LETTER R & LATIN CAPITAL LETTER R
0x73	0x0053#	LATIN SMALL LETTER S & LATIN CAPITAL LETTER S
0x74	0x0054#	LATIN SMALL LETTER T & LATIN CAPITAL LETTER T
0x75	0x0055#	LATIN SMALL LETTER U & LATIN CAPITAL LETTER U
0x76	0x0056#	LATIN SMALL LETTER V & LATIN CAPITAL LETTER V
0x77	0x0057#	LATIN SMALL LETTER W & LATIN CAPITAL LETTER W
0x78	0x0058#	LATIN SMALL LETTER X & LATIN CAPITAL LETTER X
0x79	0x0059#	LATIN SMALL LETTER Y & LATIN CAPITAL LETTER Y
0x7A	0x005A#	LATIN SMALL LETTER Z & LATIN CAPITAL LETTER Z
0x83	0x0191#	LATIN SMALL LETTER F WITH HOOK & LATIN CAPITAL LETTER F WITH HOOK
0x9C	0x0152#	LATIN SMALL LIGATURE OE & LATIN CAPITAL LIGATURE OE
0xE0	0x00C0#	LATIN SMALL LETTER A WITH GRAVE & LATIN CAPITAL LETTER A WITH GRAVE
0xE2	0x00C2#	LATIN SMALL LETTER A WITH CIRCUMFLEX & LATIN CAPITAL LETTER A WITH CIRCUMFLEX
0xE7	0x00C7#	LATIN SMALL LETTER C WITH CEDILLA & LATIN CAPITAL LETTER C WITH CEDILLA
0xE8	0x00C8#	LATIN SMALL LETTER E WITH GRAVE & LATIN CAPITAL LETTER E WITH GRAVE
0xE9	0x00C9#	LATIN SMALL LETTER E WITH ACUTE & LATIN CAPITAL LETTER E WITH ACUTE
0xEA	0x00CA#	LATIN SMALL LETTER E WITH CIRCUMFLEX & LATIN CAPITAL LETTER E WITH CIRCUMFLEX
0xEB	0x00CB#	LATIN SMALL LETTER E WITH DIAERESIS & LATIN CAPITAL LETTER E WITH DIAERESIS
0xEE	0x00CE#	LATIN SMALL LETTER I WITH CIRCUMFLEX & LATIN CAPITAL LETTER I WITH CIRCUMFLEX
0xEF	0x00CF#	LATIN SMALL LETTER I WITH DIAERESIS & LATIN CAPITAL LETTER I WITH DIAERESIS
0xF4	0x00D4#	LATIN SMALL LETTER O WITH CIRCUMFLEX & LATIN CAPITAL LETTER O WITH CIRCUMFLEX
0xF9	0x00D9#	LATIN SMALL LETTER U WITH GRAVE & LATIN CAPITAL LETTER U WITH GRAVE
0xFB	0x00DB#	LATIN SMALL LETTER U WITH CIRCUMFLEX & LATIN CAPITAL LETTER U WITH CIRCUMFLEX


```

0xFC      0x00DC# LATIN SMALL LETTER U WITH DIAERESIS & LATIN CAPITAL
#LETTER U WITH DIAERESIS
#ENDMAP      0R_2_UNICODE_UC
#BEGINMAP    UNICODE_2_0R
#
#      Name: Unicode to the corresponding letters of CP1256 (Arabic).
#      Unicode version: 2.1
#      Format:Three tab-separated columns
#      Column #1 is the Unicode input set (in hex as 0xFFFF)
#      Column #2 is the CP1256 output set (in hex as 0xFFFF)
#      Column #3 is the Unicode (ISO 10646) name (follows a comment
#      sign)
#      Each name for column #1 & column #2, if difference
#      between them
#
#      History: 14Aug1999Created for V2R4
#
0x0000      0x00# <control>
0x0001      0x01# <control>
0x0002      0x02# <control>
0x0003      0x03# <control>
0x0004      0x04# <control>
0x0005      0x05# <control>
0x0006      0x06# <control>
0x0007      0x07# <control>
0x0008      0x08# <control>
0x0009      0x09# <control>
0x000A      0x0A# <control>
0x000B      0x0B# <control>
0x000C      0x0C# <control>
0x000D      0x0D# <control>
0x000E      0x0E# <control>
0x000F      0x0F# <control>
0x0010      0x10# <control>
0x0011      0x11# <control>
0x0012      0x12# <control>
0x0013      0x13# <control>
0x0014      0x14# <control>
0x0015      0x15# <control>
0x0016      0x16# <control>
0x0017      0x17# <control>
0x0018      0x18# <control>
0x0019      0x19# <control>
0x001B      0x1B# <control>

```

0x001C	0x1C# <control>
0x001D	0x1D# <control>
0x001E	0x1E# <control>
0x001F	0x1F# <control>
0x0020	0x20# SPACE
0x0021	0x21# EXCLAMATION MARK
0x0022	0x22# QUOTATION MARK
0x0023	0x23# NUMBER SIGN
0x0024	0x24# DOLLAR SIGN
0x0025	0x25# PERCENT SIGN
0x0026	0x26# AMPERSAND
0x0027	0x27# APOSTROPHE
0x0028	0x28# LEFT PARENTHESIS
0x0029	0x29# RIGHT PARENTHESIS
0x002A	0x2A# ASTERISK
0x002B	0x2B# PLUS SIGN
0x002C	0x2C# COMMA
0x002D	0x2D# HYPHEN-MINUS
0x002E	0x2E# FULL STOP
0x002F	0x2F# SOLIDUS
0x0030	0x30# DIGIT ZERO
0x0031	0x31# DIGIT ONE
0x0032	0x32# DIGIT TWO
0x0033	0x33# DIGIT THREE
0x0034	0x34# DIGIT FOUR
0x0035	0x35# DIGIT FIVE
0x0036	0x36# DIGIT SIX
0x0037	0x37# DIGIT SEVEN
0x0038	0x38# DIGIT EIGHT
0x0039	0x39# DIGIT NINE
0x003A	0x3A# COLON
0x003B	0x3B# SEMICOLON
0x003C	0x3C# LESS-THAN SIGN
0x003D	0x3D# EQUALS SIGN
0x003E	0x3E# GREATER-THAN SIGN
0x003F	0x3F# QUESTION MARK
0x0040	0x40# COMMERCIAL AT
0x0041	0x41# LATIN CAPITAL LETTER A
0x0042	0x42# LATIN CAPITAL LETTER B
0x0043	0x43# LATIN CAPITAL LETTER C
0x0044	0x44# LATIN CAPITAL LETTER D
0x0045	0x45# LATIN CAPITAL LETTER E
0x0046	0x46# LATIN CAPITAL LETTER F
0x0047	0x47# LATIN CAPITAL LETTER G

0x0048	0x48#	LATIN CAPITAL LETTER H
0x0049	0x49#	LATIN CAPITAL LETTER I
0x004A	0x4A#	LATIN CAPITAL LETTER J
0x004B	0x4B#	LATIN CAPITAL LETTER K
0x004C	0x4C#	LATIN CAPITAL LETTER L
0x004D	0x4D#	LATIN CAPITAL LETTER M
0x004E	0x4E#	LATIN CAPITAL LETTER N
0x004F	0x4F#	LATIN CAPITAL LETTER O
0x0050	0x50#	LATIN CAPITAL LETTER P
0x0051	0x51#	LATIN CAPITAL LETTER Q
0x0052	0x52#	LATIN CAPITAL LETTER R
0x0053	0x53#	LATIN CAPITAL LETTER S
0x0054	0x54##	LATIN CAPITAL LETTER T
0x0055	0x55#	LATIN CAPITAL LETTER U
0x0056	0x56#	LATIN CAPITAL LETTER V
0x0057	0x57#	LATIN CAPITAL LETTER W
0x0058	0x58#	LATIN CAPITAL LETTER X
0x0059	0x59#	LATIN CAPITAL LETTER Y
0x005A	0x5A#	LATIN CAPITAL LETTER Z
0x005B	0x5B#	LEFT SQUARE BRACKET
0x005C	0x5C#	REVERSE SOLIDUS
0x005D	0x5D#	RIGHT SQUARE BRACKET
0x005E	0x5E#	CIRCUMFLEX ACCENT
0x005F	0x5F#	LOW LINE
0x0060	0x60#	GRAVE ACCENT
0x0061	0x61#	LATIN SMALL LETTER A
0x0062	0x62#	LATIN SMALL LETTER B
0x0063	0x63#	LATIN SMALL LETTER C
0x0064	0x64#	LATIN SMALL LETTER D
0x0065	0x65#	LATIN SMALL LETTER E
0x0066	0x66#	LATIN SMALL LETTER F
0x0067	0x67#	LATIN SMALL LETTER G
0x0068	0x68#	LATIN SMALL LETTER H
0x0069	0x69#	LATIN SMALL LETTER I
0x006A	0x6A#	LATIN SMALL LETTER J
0x006B	0x6B#	LATIN SMALL LETTER K
0x006C	0x6C#	LATIN SMALL LETTER L
0x006D	0x6D#	LATIN SMALL LETTER M
0x006E	0x6E#	LATIN SMALL LETTER N
0x006F	0x6F#	LATIN SMALL LETTER O
0x0070	0x70#	LATIN SMALL LETTER P
0x0071	0x71#	LATIN SMALL LETTER Q
0x0072	0x72#	LATIN SMALL LETTER R
0x0073	0x73#	LATIN SMALL LETTER S

0x0074	0x74#	LATIN SMALL LETTER T
0x0075	0x75#	LATIN SMALL LETTER U
0x0076	0x76#	LATIN SMALL LETTER V
0x0077	0x77#	LATIN SMALL LETTER W
0x0078	0x78#	LATIN SMALL LETTER X
0x0079	0x79#	LATIN SMALL LETTER Y
0x007A	0x7A#	LATIN SMALL LETTER Z
0x007B	0x7B#	LEFT CURLY BRACKET
0x007C	0x7C#	VERTICAL LINE
0x007D	0x7D#	RIGHT CURLY BRACKET
0x007E	0x7E#	TILDE
0x007F	0x7F#	<control>
0x00A0	0xA0#	NO-BREAK SPACE
0x00A2	0xA2#	CENT SIGN
0x00A3	0xA3#	POUND SIGN
0x00A4	0xA4#	CURRENCY SIGN
0x00A5	0xA5#	YEN SIGN
0x00A6	0xA6#	BROKEN BAR
0x00A7	0xA7#	SECTION SIGN
0x00A8	0xA8#	DIAERESIS
0x00A9	0xA9#	COPYRIGHT SIGN
0x00AB	0xAB#	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
0x00AC	0xAC#	NOT SIGN
0x00AD	0xAD#	SOFT HYPHEN
0x00AE	0xAE#	REGISTERED SIGN
0x00AF	0xAF#	MACRON
0x00B0	0xB0#	DEGREE SIGN
0x00B1	0xB1#	PLUS-MINUS SIGN
0x00B2	0xB2#	SUPERSCRIFT TWO
0x00B3	0xB3#	SUPERSCRIFT THREE
0x00B4	0xB4#	ACUTE ACCENT
0x00B5	0xB5#	MICRO SIGN
0x00B6	0xB6#	PILCROW SIGN
0x00B7	0xB7#	MIDDLE DOT
0x00B8	0xB8#	CEDILLA
0x00B9	0xB9#	SUPERSCRIFT ONE
0x00BB	0xBB#	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
0x00BC	0xBC#	VULGAR FRACTION ONE QUARTER
0x00BD	0xBD#	VULGAR FRACTION ONE HALF
0x00BE	0xBE#	VULGAR FRACTION THREE QUARTERS
0x00C0	0xE0#	LATIN CAPITAL LETTER A WITH GRAVE & LATIN SMALL #LETTER A WITH GRAVE
0x00C2	0xE2#	LATIN CAPITAL LETTER A WITH CIRCUMFLEX & LATIN #SMALL LETTER A WITH CIRCUMFLEX

```

0x00C7    0xE7# LATIN CAPITAL LETTER C WITH CEDILLA & LATIN SMALL
#LETTER C WITH CEDILLA
0x00C8    0xE8# LATIN CAPITAL LETTER E WITH GRAVE & LATIN SMALL
#LETTER E WITH GRAVE
0x00C9    0xE9# LATIN CAPITAL LETTER E WITH ACUTE & LATIN SMALL
#LETTER E WITH ACUTE
0x00CA    0xEA# LATIN CAPITAL LETTER E WITH CIRCUMFLEX & LATIN
#SMALL LETTER E WITH CIRCUMFLEX
0x00CB    0xEB# LATIN CAPITAL LETTER E WITH DIAERESIS & LATIN SMALL
#LETTER E WITH DIAERESIS
0x00CE    0xEE# LATIN CAPITAL LETTER I WITH CIRCUMFLEX & LATIN
#SMALL LETTER I WITH CIRCUMFLEX
0x00CF    0xEF# LATIN CAPITAL LETTER I WITH DIAERESIS & LATIN SMALL
#LETTER I WITH DIAERESIS
0x00D4    0xF4# LATIN CAPITAL LETTER O WITH CIRCUMFLEX & LATIN
#SMALL LETTER O WITH CIRCUMFLEX
0x00D7    0xD7# MULTIPLICATION SIGN
0x00D9    0xF9# LATIN CAPITAL LETTER U WITH GRAVE & LATIN SMALL
#LETTER U WITH GRAVE
0x00DB    0xFB# LATIN CAPITAL LETTER U WITH CIRCUMFLEX & LATIN
#SMALL LETTER U WITH CIRCUMFLEX
0x00DC    0xFC# LATIN CAPITAL LETTER U WITH DIAERESIS & LATIN SMALL
#LETTER U WITH DIAERESIS
0x00E0    0xE0# LATIN SMALL LETTER A WITH GRAVE
0x00E2    0xE2# LATIN SMALL LETTER A WITH CIRCUMFLEX
0x00E7    0xE7# LATIN SMALL LETTER C WITH CEDILLA
0x00E8    0xE8# LATIN SMALL LETTER E WITH GRAVE
0x00E9    0xE9# LATIN SMALL LETTER E WITH ACUTE
0x00EA    0xEA# LATIN SMALL LETTER E WITH CIRCUMFLEX
0x00EB    0xEB# LATIN SMALL LETTER E WITH DIAERESIS
0x00EE    0xEE# LATIN SMALL LETTER I WITH CIRCUMFLEX
0x00EF    0xEF# LATIN SMALL LETTER I WITH DIAERESIS
0x00F4    0xF4# LATIN SMALL LETTER O WITH CIRCUMFLEX
0x00F7    0xF7# DIVISION SIGN
0x00F9    0xF9# LATIN SMALL LETTER U WITH GRAVE
0x00FB    0xFB# LATIN SMALL LETTER U WITH CIRCUMFLEX
0x00FC    0xFC# LATIN SMALL LETTER U WITH DIAERESIS
0x0152    0x8C# LATIN CAPITAL LIGATURE OE
0x0153    0x9C# LATIN SMALL LIGATURE OE
0x0191    0x83# LATIN CAPITAL LETTER F WITH HOOK & LATIN SMALL
#LETTER F WITH HOOK
0x0192    0x83# LATIN SMALL LETTER F WITH HOOK
0x02C6    0x88# MODIFIER LETTER CIRCUMFLEX ACCENT
0x060C    0xA1# ARABIC COMMA

```

0x061B	0xBA#	ARABIC SEMICOLON
0x061F	0xBF#	ARABIC QUESTION MARK
0x0621	0xC1#	ARABIC LETTER HAMZA
0x0622	0xC2#	ARABIC LETTER ALEF WITH MADDA ABOVE
0x0623	0xC3#	ARABIC LETTER ALEF WITH HAMZA ABOVE
0x0624	0xC4#	ARABIC LETTER WAW WITH HAMZA ABOVE
0x0625	0xC5#	ARABIC LETTER ALEF WITH HAMZA BELOW
0x0626	0xC6#	ARABIC LETTER YEH WITH HAMZA ABOVE
0x0627	0xC7#	ARABIC LETTER ALEF
0x0628	0xC8#	ARABIC LETTER BEH
0x0629	0xC9#	ARABIC LETTER TEH MARBUTA
0x062A	0xCA#	ARABIC LETTER TEH
0x062B	0xCB#	ARABIC LETTER THEH
0x062C	0xCC#	ARABIC LETTER JEEM
0x062D	0xCD#	ARABIC LETTER HAH
0x062E	0xCE#	ARABIC LETTER KHAH
0x062F	0xCF#	ARABIC LETTER DAL
0x0630	0xD0#	ARABIC LETTER THAL
0x0631	0xD1#	ARABIC LETTER REH
0x0632	0xD2#	ARABIC LETTER ZAIN
0x0633	0xD3#	ARABIC LETTER SEEN
0x0634	0xD4#	ARABIC LETTER SHEEN
0x0635	0xD5#	ARABIC LETTER SAD
0x0636	0xD6#	ARABIC LETTER DAD
0x0637	0xD8#	ARABIC LETTER TAH
0x0638	0xD9#	ARABIC LETTER ZAH
0x0639	0xDA#	ARABIC LETTER AIN
0x063A	0xDB#	ARABIC LETTER GHAIN
0x0640	0xDC#	ARABIC TATWEEL
0x0641	0xDD#	ARABIC LETTER FEH
0x0642	0xDE#	ARABIC LETTER QAF
0x0643	0xDF#	ARABIC LETTER KAF
0x0644	0xE1#	ARABIC LETTER LAM
0x0645	0xE3#	ARABIC LETTER MEEM
0x0646	0xE4#	ARABIC LETTER NOON
0x0647	0xE5#	ARABIC LETTER HEH
0x0648	0xE6#	ARABIC LETTER WAW
0x0649	0xEC#	ARABIC LETTER ALEF MAKSURA
0x064A	0xED#	ARABIC LETTER YEH
0x064B	0xF0#	ARABIC FATHATAN
0x064C	0xF1#	ARABIC DAMMATAN
0x064D	0xF2#	ARABIC KASRATAN
0x064E	0xF3#	ARABIC FATHA
0x064F	0xF5#	ARABIC DAMMA

```

0x0650    0xF6# ARABIC KASRA
0x0651    0xF8# ARABIC SHADDA
0x0652    0xFA# ARABIC SUKUN
0x0679    0x8A# ARABIC LETTER TTEH
0x067E    0x81# ARABIC LETTER PEH
0x0686    0x8D# ARABIC LETTER TCHEH
0x0688    0x8F# ARABIC LETTER DDAL
0x0691    0x9A# ARABIC LETTER RREH
0x0698    0x8E# ARABIC LETTER JEH
0x06A9    0x98# ARABIC LETTER KEHEH
0x06AF    0x90# ARABIC LETTER GAF
0x06BA    0x9F# ARABIC LETTER NOON GHUNNA
0x06BE    0xAA# ARABIC LETTER HEH DOACHASHMEE
0x06C1    0xC0# ARABIC LETTER HEH GOAL
0x06D2    0xFF# ARABIC LETTER YEH BARREE
0x200C    0x9D# ZERO WIDTH NON-JOINER
0x200D    0x9E# ZERO WIDTH JOINER
0x200E    0xFD# LEFT-TO-RIGHT MARK
0x200F    0xFE# RIGHT-TO-LEFT MARK
0x2013    0x96# EN DASH
0x2014    0x97# EM DASH
0x2018    0x91# LEFT SINGLE QUOTATION MARK
0x2019    0x92# RIGHT SINGLE QUOTATION MARK
0x201A    0x82# SINGLE LOW-9 QUOTATION MARK
0x201C    0x93# LEFT DOUBLE QUOTATION MARK
0x201D    0x94# RIGHT DOUBLE QUOTATION MARK
0x201E    0x84# DOUBLE LOW-9 QUOTATION MARK
0x2020    0x86# DAGGER
0x2021    0x87# DOUBLE DAGGER
0x2022    0x95# BULLET
0x2026    0x85# HORIZONTAL ELLIPSIS
0x2030    0x89# PER MILLE SIGN
0x2039    0x8B# SINGLE LEFT-POINTING ANGLE QUOTATION MARK
0x203A    0x9B# SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
0x20AC    0x80# EURO SIGN
0x2122    0x99# TRADE MARK SIGN
#ENDMAP      UNICODE_2_0R
#BEGINMAP    UNICODE_2_0R_UC
#
#   Name:The lower case CP1256 letters of Unicode to the
#   corresponding upper case letters of CP1256.
#   Unicode version:  2.1
#   Format:Three tab-separated columns
#   Column #1 is the Unicode input set (in hex as 0xFFFF)

```

```

#      Column #2 is the CP1256 output set (in hex as 0xFFFF)
#      Column #3 is the Unicode(ISO 10646) name(follows comment sign)
#      Each name for column #1 & column #2, if difference
#      between them
#
#      History: 14Aug1999Created for V2R4
#
0x0061    0x41# LATIN SMALL LETTER A & LATIN CAPITAL LETTER A
0x0062    0x42# LATIN SMALL LETTER B & LATIN CAPITAL LETTER B
0x0063    0x43# LATIN SMALL LETTER C & LATIN CAPITAL LETTER C
0x0064    0x44# LATIN SMALL LETTER D & LATIN CAPITAL LETTER D
0x0065    0x45# LATIN SMALL LETTER E & LATIN CAPITAL LETTER E
0x0066    0x46# LATIN SMALL LETTER F & LATIN CAPITAL LETTER F
0x0067    0x47# LATIN SMALL LETTER G & LATIN CAPITAL LETTER G
0x0068    0x48# LATIN SMALL LETTER H & LATIN CAPITAL LETTER H
0x0069    0x49# LATIN SMALL LETTER I & LATIN CAPITAL LETTER I
0x006A    0x4A# LATIN SMALL LETTER J & LATIN CAPITAL LETTER J
0x006B    0x4B# LATIN SMALL LETTER K & LATIN CAPITAL LETTER K
0x006C    0x4C# LATIN SMALL LETTER L & LATIN CAPITAL LETTER L
0x006D    0x4D# LATIN SMALL LETTER M & LATIN CAPITAL LETTER M
0x006E    0x4E# LATIN SMALL LETTER N & LATIN CAPITAL LETTER N
0x006F    0x4F# LATIN SMALL LETTER O & LATIN CAPITAL LETTER O
0x0070    0x50# LATIN SMALL LETTER P & LATIN CAPITAL LETTER P
0x0071    0x51# LATIN SMALL LETTER Q & LATIN CAPITAL LETTER Q
0x0072    0x52# LATIN SMALL LETTER R & LATIN CAPITAL LETTER R
0x0073    0x53# LATIN SMALL LETTER S & LATIN CAPITAL LETTER S
0x0074    0x54# LATIN SMALL LETTER T & LATIN CAPITAL LETTER T
0x0075    0x55# LATIN SMALL LETTER U & LATIN CAPITAL LETTER U
0x0076    0x56# LATIN SMALL LETTER V & LATIN CAPITAL LETTER V
0x0077    0x57# LATIN SMALL LETTER W & LATIN CAPITAL LETTER W
0x0078    0x58# LATIN SMALL LETTER X & LATIN CAPITAL LETTER X
0x0079    0x59# LATIN SMALL LETTER Y & LATIN CAPITAL LETTER Y
0x007A    0x5A# LATIN SMALL LETTER Z & LATIN CAPITAL LETTER Z
0x0152    0x9C# LATIN CAPITAL LIGATURE OE & LATIN SMALL LIGATURE OE
#ENDMAP UNICODE_2_0R_UC

```


Collation Sequences

Overview

The following sections describe the default collation sequences supplied by Teradata and how to define and install site-defined collation sequences.

About Collation Sequences

Collations control character ordering and comparison operations during Teradata Database sessions.

Collations are designed as single level or two level. A two-level collation orders character strings according to a two-level comparison.

Characters are first partitioned into equivalence classes that have the same collating value. The relative ordering of classes and characters within a class is significant.

Comparisons obey the following rules:

- All characters in a class have the same collation value.
- A character from class i is less than any character from class $i+1$.

The process is as follows:

1. Convert characters in the strings to be compared into equivalence classes.
2. Compare the strings.

IF the strings are...	THEN processing ...
not equal	stops.
equal	continues.

3. Order characters within each class using criteria defined for the collation sequence, and compare.

The MULTINATIONAL Norwegian standard collation sequence is an example of a two-level collation.

Standard Teradata Database Collation Sequences

The Teradata Database offers five standard collation sequences in which data can be defined as CASESPECIFIC or NOT CASESPECIFIC. This affects how the five collation sequences collate and compare data.

The five collations, determined either by default or explicit use of the SET SESSION COLLATION statement, are:

- ASCII

- EBCDIC
- CHARSET_COLL
- JIS_COLL
- MULTINATIONAL

CASESPECIFIC or NOT CASESPECIFIC can be chosen at table definition time, or specified as part of the SQL statement.

The default collation sequence is based upon the client type:

- EBCDIC for mainframe clients
- ASCII for all other clients

Collation Sequence Ordering

Collation sequence ordering is as follows:

- ASCII collation orders the data essentially as would converting the data to Teradata extended ASCII (the LATIN server character set), and then using a binary ordering of the resulting byte string.
- EBCDIC collation orders the data essentially as would converting the data to Teradata extended EBCDIC, and then using a binary ordering of the resulting bytes.
- CHARSET_COLL collation orders the data essentially as would converting the string to the current client character set, and then using a binary ordering of the resulting byte string.
- JIS_Coll collation collates in the order of the Japanese Industrial Standards.
- MULTINATIONAL collation provides more culturally aware ordering of data.

The predefined multinational collation options are:

- Teradata Standard Multinational (the initial default)
- Swedish
- Norwegian
- Katakana_Standard
- Kanji5026_Standard
- Kanji5035_Standard

Further multinational collation options can be loaded using scripts. The database administrator can alter MULTINATIONAL collation.

When collation is set to MULTINATIONAL, the default sequence currently installed is used. This can either be one of the predefined sequences, supplied with the Teradata Database, or a sequence you have defined and installed.

You can execute predefined macros to change the default to Swedish, Norwegian, or the appropriate Japanese standard collation. You can also define and install your own collation, as explained in [Defining Your Own Collation Sequence](#).

If all the items being compared or collated are determined to be NOT CASESPECIFIC, the collation works as if all characters that have an uppercase counterpart were converted to uppercase before being processed through ASCII, EBCDIC, CHARSET_COLL or JIS_COLL collation.

Setting and Changing Collations

Collation can be set or changed several different ways:

- At the user-definition level with the CREATE USER or MODIFY USER statements.
- At the session level with the SQL SET SESSION COLLATION statement.

You can use predefined macros to change the collation default to Swedish, Norwegian, or the appropriate Japanese standard collation.

Note:

Katakana_Standard, Kanji5026_Standard, and Kanji5035_Standard are designed for the KANJI1 server character set and should not be used with other server character sets. Similarly, the other predefined collation options should not be used for KANJI1 data.

You can also define and install your own collation sequences, as explained in [Defining Your Own Collation Sequence](#).

HELP SESSION Statement

Use the HELP SESSION statement to display the collation currently in effect for your session.

Further Collation Sequences and Character Mappings

Selected collation and character mapping tables are described in text files that are available for download. See [Character Set Files](#).

ASCII and EBCDIC Collations

The ASCII collation sequence and the EBCDIC collation sequence are always available.

ASCII Collation

ASCII collation orders the data essentially as would converting the data to Teradata extended ASCII (the LATIN server character set), and then using a binary ordering of the resulting byte string.

EBCDIC Collation

EBCDIC collation orders the data essentially as would converting the data to Teradata extended EBCDIC, and then using a binary ordering of the resulting bytes.

Default Collation

If the COLLATION parameter is not defined, the default is HOST (the collation that is compatible with the logon client).

HOST collation gives you ASCII collation on network-attached clients and EBCDIC collation on mainframe clients.

Related Information

For details on Teradata Database collating conventions, see:

- “Comparison Operators” and “Comparison of KANJI1 Characters” in *Teradata Vantage™ - SQL Functions, Expressions, and Predicates*, B035-1145
- “ORDER BY Clause” in *Teradata Vantage™ - SQL Data Manipulation Language*, B035-1146

CHARSET_COLL Collation

The CHARSET_COLL collation produces a binary ordering based upon the current client character set. The NOT CASESPECIFIC version is designed to produce the results if the strings were converted to uppercase and then sorted in binary order on the client.

CHARSET_COLL gives you a collation that matches the client character set. For example, if the client character set is KANJIIEBCDIC5035_OI, then the collation matches KANJIIEBCDIC5035_OI order (rather than EBCDIC order).

Comparison Rules

Strings are compared character-by-character.

The comparison rules for CHARSET_COLL are:

- If one string is shorter, it is padded with the pad character for the character set.
- If the comparison is not case specific, lowercase characters are mapped to their uppercase counterparts.
- If the strings are now identical, the equality relation holds. Otherwise, the first pair of characters that are not equal determine the collating sequence.
- If both characters are in the repertoire of the current client character set, then the binary ordering of the two characters in the client form-of-use becomes the ordering of the two strings.
- If one of the characters is not within the repertoire of the current client character set, then the error character is used as the collation point for that character.
- If both characters being compared are outside the repertoire of the current client character set, then the binary ordering of the characters (case blind or case specific, as appropriate) in the Unicode form-of-use becomes the ordering of the two strings.
- Kanji data

- `CHARSET_COLL` is of limited use with the KANJI1 server character set.
KANJI1 character data can contain mixed single-byte/multibyte characters. Single-byte characters are translated into the Teradata Database form-of-use and multibyte characters are not translated.
- Single-byte characters are collated based on the current character set and multibyte characters based on their internal value.
- For KanjiEBCDIC and KanjiShift-JIS client character sets, the collation is like a binary sort on the client.
- For a KanjiEUC client character set, the collation is like Kanji Phase I ASCII collation.
The distinction between this and a binary sort on the client is that the JIS X 0208 characters collate before, rather than after, the JIS X 0212 characters.

Making `CHARSET_COLL` the Default Collation Sequence

You can specify `CHARSET_COLL` as the default user collation with the `CREATE USER` or `MODIFY USER` statements.

You can also use the `SQL SET SESSION COLLATION CHARSET_COLL` statement to override any user defaults.

Related Information

For information on...	See...
Teradata Database collating conventions	“Comparison Operators” and “Comparison of KANJI1 Characters” in <i>Teradata Vantage™ - SQL Functions, Expressions, and Predicates</i> , B035-1145.
	“ORDER BY Clause” in <i>Teradata Vantage™ - SQL Data Manipulation Language</i> , B035-1146.

JIS_COLL Collation

`JIS_COLL` collation provides a character set independent ordering based on the Japanese Industrial Standards.

JIS_COLL Collation Sequence

`JIS_COLL` collation orders data as follows:

1. Characters and symbols from the JIS X 0201 standard (in JIS X 0201 order).
2. Ideographs, characters and symbols from the JIS X 0208 standard (in JIS X 0208 order).
3. Ideographs, characters and symbols from the JIS X 0212 standard (in JIS X 0212 order).
4. IBM Kanji ideographs not present in JIS X 0201, JIS X 0208, and JIS X 0212 (in KanjiEBCDIC order).

5. Site-defined ideographs (in Unicode order).
6. The remaining characters in Unicode (in Unicode order).

Collation of the KANJISJIS, GRAPHIC, and LATIN character sets is as if the data were first converted to Unicode and then the appropriate JIS_COLL ordering were applied.

Making JIS_COLL the Default Collation Sequence

You can specify JIS_COLL as the default user collation with the CREATE or MODIFY USER statement, or use SET SESSION COLLATION JIS_COLL to override any user defaults.

Related Information

For information on...	See...
Teradata Database collating conventions	“Comparison Operators” and “Comparison of KANJI1 Characters” in <i>Teradata Vantage™ - SQL Functions, Expressions, and Predicates</i> , B035-1145.
	“ORDER BY Clause” in <i>Teradata Vantage™ - SQL Data Manipulation Language</i> , B035-1146.

MULTINATIONAL Collation

Available Collation Sequences

The collation sequence for MULTINATIONAL collation can be any one of the following:

- Teradata Standard Multinational (a two-level comparison based on the Unicode collation standard).

This is the initial default sequence for any user or session with [SESSION] COLLATION set to MULTINATIONAL.

It can be altered using the macro CollAddStandard.

- Norwegian Standard (based on the Norwegian collation sequence).
- Swedish Standard (based on the Swedish collation sequence).
- Any one of three predefined Japanese collations:
 - Katakana_Standard
 - Kanji5026_Standard
 - Kanji5035_Standard
- A site-defined collation sequence.

Multilevel Collation

MULTINATIONAL collation is two level except when used with KANJI1 data, where the collation is single level. Teradata, Norwegian, and Swedish standard collations are designed as two level, while all Japanese collations are single level.

Use collations designed as single level on KANJI1 columns only. Do not use collations designed as two level on KANJI1 columns.

MULTINATIONAL Collation Rules

The comparison of two character strings under MULTINATIONAL collation obeys the following rules:

- All characters in an equivalent class have the same collation value. A character from class i is less than any character from class $i+1$.
- If two strings are equal, based upon the class, additional processing takes place to order the characters within a class.

Example: Teradata Standard Multinational Collation Equivalence Class

As an example using Teradata Standard Multinational collation, consider the following characters: b, B, c, C, ç, Ç, d and D.

Teradata Standard Multinational collation considers these characters ...	To be in the following equivalence class ...
b B	B
c C ç Ç	C
d D	D

Within the equivalence class of ...	The case-specific ordering is ...
B	1. b 2. B
C	1. c 2. C 3. ç

Within the equivalence class of ...	The case-specific ordering is ...
	4. Ç
D	1. d 2. D

Equivalence Classes and Ordering Within Class

The following table shows the internal values of the characters b, B, c, C, ç, Ç, d and D, and their corresponding values in CollEqvClass, CollOrderCS, and CollOrderUC.

In this table, the values assigned to the equivalence class of b, c, and d are 0x62, 0x63, and 0x64, respectively. This assures that b sorts before c, and c before d. The case-specific (CollOrderCS) value for c is 0x00, and the case-specific value for C is 0x01. This assures that, using the second-level comparison, c sorts before C.

For details on ordering characters within classes as performed by the Teradata Standard Multinational collation, see “ORDER BY Clause” in “The SELECT Statement” in *Teradata Vantage™ - SQL Data Manipulation Language*, B035-1146.

Character	Hexadecimal Value			
	Teradata Internal Value	CollEqv Class	CollOrder Casespecific	CollOrder Uppercase
b	0x62	0x62	0x00	0x01
B	0x42	0x62	0x01	0x01
c	0x63	0x63	0x00	0x01
C	0x43	0x63	0x01	0x01
ç	0xE7	0x63	0x02	0x01
Ç	0xC7	0x63	0x03	0x01
d	0x64	0x64	0x00	0x01
D	0x44	0x64	0x01	0x01

Single-Level Collations: KANJI1

For Japanese language support, MULTINATIONAL collation is single level.

Making MULTINATIONAL the Default Collation Sequence

You can specify MULTINATIONAL as the default user collation with the CREATE and MODIFY USER statements.

You can also use the SQL SET SESSION COLLATION MULTINATIONAL statement to override any user defaults.

Related Information

For information on...	See...
Teradata Database collating conventions	“Comparison Operators” and “Comparison of KANJI1 Characters” in <i>Teradata Vantage™ - SQL Functions, Expressions, and Predicates</i> , B035-1145.
	“ORDER BY Clause” in <i>Teradata Vantage™ - SQL Data Manipulation Language</i> , B035-1146.

Changing the Standard Multinational Default Collation

After running the DIP utility, a privileged user can execute the CollInstallMulti macro to change the default sequence for multinational collation. The EXECUTE privilege on this macro is initially assigned to system users SYSADMIN and DBC.

After the change is installed, the new ordering is enforced for the session of any user whose COLLATION attribute is MULTINATIONAL, or who submits the following SQL statement:

```
SET SESSION COLLATION MULTINATIONAL;
```

Tools for Collation Control

During a multinational session, the Teradata Database checks a system table for a row containing the value MULTINATIONAL in the CollName column. The hexadecimal codes in that row determine the sort sequence for the session.

The Teradata Database provides a set of objects in database DBC with which you can manipulate, build, and maintain codes for the various collation options for Multinational users. These objects include:

- A system table that you can access by means of the DBC.CollationsV view in which each row contains the definition of one collation sequence.

Note that only one definition can be in effect at any given time. The default collation is based on the codes in the row where:

```
CollName=MULTINATIONAL
CollInstall=Y
```

- A system view named DBC.CollationsV.
Insert site-defined collation codes through this view to populate a row in the underlying system table. The view is created with the clause GRANT SELECT TO PUBLIC to make it accessible to all Teradata Database users.
- A system macro named DBC.CollInstallMulti.
This macro is invoked by a privileged user to replace the current MULTINATIONAL collation with Teradata Database or site-defined codes that have been loaded into a row of the system table.
A system restart is required before the new collation becomes the current MULTINATIONAL collation.
- A system macro named DBC.CollAddStandard.
This macro is run automatically when the DIP utility is executed during the Teradata Database installation procedure.

Procedure

To change the Teradata standard MULTINATIONAL collation to the standard Swedish or Norwegian collation, perform the following procedure:

1. Log on to a BTEQ session as user SYSADMIN or user DBC.
2. Execute the DBC.CollInstallMulti macro, specifying the desired standard as the CollationName parameter for the macro.

For example, any of the following:

```
EXECUTE DBC.CollInstallMulti ('NORWEGIAN_STANDARD');
EXECUTE DBC.CollInstallMulti ('SWEDISH_STANDARD');
EXECUTE DBC.CollInstallMulti ('KANJI5026_STANDARD');
EXECUTE DBC.CollInstallMulti ('KANJI5035_STANDARD');
EXECUTE DBC.CollInstallMulti ('KATAKANA_STANDARD');
```

3. On the Command Line of the Database Window of the system console, enter the restart tpa command to restart the Teradata Database.

During restart processing, the Teradata Database displays a message indicating whether the specified collation was installed.

Uninstalling a Nonstandard Collation

To uninstall a nonstandard collation and return to the Teradata standard MULTINATIONAL collation, perform the following procedure:

1. Log on to a BTEQ session as system user SYSADMIN or DBC.

2. Do one of the following:
 - Delete the row from DBC.CollationsV where CollName='MULTINATIONAL'
 - Change to N (no) the value of the CollInstall column in the row where CollName='MULTINATIONAL'
3. To reload the Teradata Standard Multinational collation, enter the following command to execute the DBC.CollAddStandard macro:

```
EXECUTE DBC.CollAddStandard;
```

4. Restart the Teradata Database.

You can also execute the DBC.CollAddStandard macro to reload the Teradata Standard Multinational collation if the contents of the system table become corrupted. Remember to restart the Teradata Database to register the codes.

Example of Installing a Japanese Character Standard Collation

The following script is an example of uninstalling and installing a standard Japanese character collation:

```
DELETE FROM DBC.CollationsV
  WHERE CollName = 'KANJI5026_STANDARD';
INSERT INTO DBC.CollationsV
  ( CollName
  , CollInstall
  , CollEqvClass
  , CollOrderCS
  , CollOrderUC )
VALUES
  ( 'KANJI5026_STANDARD'
  , 'N'
  , /* CollEqvClass. Equivalence class. */
  , /* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
  , '00000000000000000000000000000000'xb /* 00__ */
  , ||'00000000000000000000000000000000'xb /* 01__ */
  , ||'00000000000000000000000000000000'xb /* 02__ */
  , ||'00000000000000000000000000000000'xb /* 03__ */
  , ||'00000000000000000000000000000000'xb /* 04__ */
  , ||'00000000000000000000000000000000'xb /* 05__ */
  , ||'00000000000000000000000000000000'xb /* 06__ */
  , ||'00000000000000000000000000000000'xb /* 07__ */
  , ||'00000000000000000000000000000000'xb /* 08__ */
  , ||'00000000000000000000000000000000'xb /* 09__ */
  , ||'00000000000000000000000000000000'xb /* 0a__ */
  , ||'00000000000000000000000000000000'xb /* 0b__ */
```

```

|'|'00000000000000000000000000000000'xb /* 0c__ */
|'|'00000000000000000000000000000000'xb /* 0d__ */
|'|'00000000000000000000000000000000'xb /* 0e__ */
|'|'00000000000000000000000000000000'xb /* 0f__ */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
/* CollOrderCS. Case-specific ordering */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
, '00010203372d2e2f1605250b0c0d0e0f'xb /* 00__ */
|'|'b14a5fb2a03d322618193f271c1d1e1f'xb /* 01__ */
|'|'405a7f7be06c507d4d5d5c4e6b604b61'xb /* 02__ */
|'|'f0f1f2f3f4f5f6f7f8f97a5e4c7e6e6f'xb /* 03__ */
|'|'7cc1c2c3c4c5c6c7c8c9d1d2d3d4d5d6'xb /* 04__ */
|'|'d7d8d9e2e3e4e5e6e7e8e9705b80b06d'xb /* 05__ */
|'|'79626364656667686971727374757677'xb /* 06__ */
|'|'788b9babb3b4b5b6b7b8b9c04fd0a107'xb /* 07__ */
|'|'102122232415061728292a2b2c090a1b'xb /* 08__ */
|'|'30311a333435360838393a3b04143eff'xb /* 09__ */
|'|'11414243444546474849515253545556'xb /* 0a__ */
|'|'588182838485868788898a8c8d8e8f90'xb /* 0b__ */
|'|'9192939495969798999a9d9e9fa2a3a4'xb /* 0c__ */
|'|'a5a6a7a8a9aaacadaeafbabbcbdbebf'xb /* 0d__ */
|'|'2057596a9ccacbccdcecfdadbdcdde'xb /* 0e__ */
|'|'dfe1eaebeceeeffafbfcd3c1213fe'xb /* 0f__ */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
/* CollOrderUC. uppercase ordering */
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
, '00010203372d2e2f1605250b0c0d0e0f'xb /* 00__ */
|'|'b14a5fb2a03d322618193f271c1d1e1f'xb /* 01__ */
|'|'405a7f7be06c507d4d5d5c4e6b604b61'xb /* 02__ */
|'|'f0f1f2f3f4f5f6f7f8f97a5e4c7e6e6f'xb /* 03__ */
|'|'7cc1c2c3c4c5c6c7c8c9d1d2d3d4d5d6'xb /* 04__ */
|'|'d7d8d9e2e3e4e5e6e7e8e9705b80b06d'xb /* 05__ */
|'|'79c1c2c3c4c5c6c7c8c9d1d2d3d4d5d6'xb /* 06__ */
|'|'d7d8d9e2e3e4e5e6e7e8e9c04fd0a107'xb /* 07__ */
|'|'102122232415061728292a2b2c090a1b'xb /* 08__ */
|'|'30311a333435360838393a3b04143eff'xb /* 09__ */
|'|'11414243444546474849515253545556'xb /* 0a__ */
|'|'588182838485868788898a8c8d8e8f90'xb /* 0b__ */
|'|'9192939495969798999a9d9e9fa2a3a4'xb /* 0c__ */
|'|'a5a6a7a8a9aaacadaeafbabbcbdbebf'xb /* 0d__ */
|'|'2057596a9ccacbccdcecfdadbdcdde'xb /* 0e__ */
|'|'dfe1eaebeceeeffafbfcd3c1213fe'xb /* 0f__ */

```

```
/* 0 1 2 3 4 5 6 7 8 9 A B C D E F */
);
```

Defining Your Own Collation Sequence

To create your own collation sequence, define for each character of the alphabet the hexadecimal representation for each of the following:

- The equivalence class
- The casespecific value
- The uppercase value

It is useful to have the hexadecimal representation for the Teradata Database internal value of the character so that the preceding items can be placed in the appropriate position within the defining byte string, as shown in [Example of Installing a Japanese Character Standard Collation](#).

Each group of values is inserted into a column in one row of DBC.CollationsV.

Site-defined collations are valid only for single-byte characters.

Customizing the Defined Collations

A defined collation is active for all server character sets, but they are not applied equally.

You can change the MULTINATIONAL collation for all 256 characters in LATIN, and this definition holds true for those characters whether they occur in GRAPHIC, KANJISJIS, or UNICODE. Note that KANJISJIS and UNICODE have many more characters than LATIN and that the “extra” characters continue to follow the default multinational collation rules.

The collation is two level, which, in the following case specific example, allows you to specify a collation that always produces the following ordering:

1. aa
2. Aa
3. ab
4. Ab

A single-level, case blind collation might give that answer sometimes, but could theoretically also produce several other orderings, such as the following, on consecutive queries:

1. Aa
2. aa
3. ab
4. Ab

KANJI1 Exception

The KANJI1 server character type is different because its collation is single level, based on the CollOrderCS and CollOrderUC fields. In other words, KANJI1 uses the second-level collation values as its only level. Because of this, it is not possible to define a collation that works well with both KANJI1 and other server character sets.

Note also that for KANJI1, only single-byte characters are affected by the MULTINATIONAL collation definition. Multibyte characters continue to collate as always by means of binary comparison of the internal representation of those characters. This should be very similar to how those characters would collate in the client character set they were entered from, with the exception that EUC cs3 characters on the client would sort before, rather than after, cs1 characters.

Definition of the CollInstallMulti Macro

The CollInstallMulti macro takes the codes you load into the DBC.CollationsV view and inserts them into the underlying system table as the new default collation sequence.

The :CollationName parameter you submit to the macro must match the CollName value you specified for DBC.CollationsV.

During this process, CollInstallMulti performs the following actions.

1. Verifies that the value of :CollationName is not MULTINATIONAL.
2. Verifies that :CollationName is found in the CollName column of DBC.CollationTbl.
3. Deletes the row where CollName is MULTINATIONAL.
4. Inserts (through DBC.CollationsV) a row with the following values:
 - The string MULTINATIONAL in CollName
 - The letter Y in CollInstall
 - The codes for CollEqvClass, CollOrderCS, and CollOrderUC, which are selected from the row you loaded (where CollName should be equal to :CollationName).

The SQL code for CollInstallMulti is as follows:

```
REPLACE MACRO CollInstallMulti (CollationName CHAR(30)) AS
( ABORT 'Argument cannot be MULTINATIONAL collation'
WHERE :CollationName = 'MULTINATIONAL';
ABORT 'Specified collation has not been defined'
WHERE :CollationName NOT IN
( SELECT CollName
FROM DBC.CollationsV );
DELETE FROM DBC.CollationsV
WHERE CollName = 'MULTINATIONAL';
INSERT INTO DBC.CollationsV
SELECT 'MULTINATIONAL', 'Y', CollEqvClass,
```

```
CollOrderCS, CollOrderUC
FROM DBC.CollationsV
WHERE CollName = :CollationName; );
```

DBC.CollationsV View Columns

The name, attributes, and use of each column in DBC.CollationsV are defined as follows.

Column Name	Data Type	Description
CollName	VARCHAR(128) NOT NULL	Enter any legal name except MULTINATIONAL. When executed, the CollInstallMulti macro accepts any legal name other than MULTINATIONAL.
CollInstall	CHAR(1) NOT NULL	Enter the letter N. When executed, CollInstallMulti inserts a Y indicating that this row is to be installed.
CollEqvClass	BYTE (256) BETWEEN 0x00 AND 0xFF NOT NULL	Specify, for each character, the value of its equivalent class. For standard Japanese collations, each value in this field must be 0x00.
CollOrderCS	BYTE (256) BETWEEN 0x00 AND 0xFF NOT NULL	Specify, for each character, the relative order of the character within its class for case-specific comparison.
CollOrderUC	BYTE (256) BETWEEN 0x00 AND 0xFF NOT NULL	Specify, for each character, the relative order of the character within its class for uppercase comparison.

For more information about DBC.CollationsV, see *Teradata Vantage™ - Data Dictionary*, B035-1092.

Inserting Collation Codes

Use the DBC.CollationsV view to populate the system table with your collation codes.

To populate the table through the view, perform the following procedure:

1. Create a BTEQ batch file containing a CollName, a CollInstall value, and the hexadecimal definition of every character in the new collation sequence. Be sure that entries exist for every column of the DBC.CollationsV view.

The value of CollName cannot be MULTINATIONAL. Also, the first SQL statement in the job file should be a DELETE statement to remove any existing row that might have the same CollName. For details, see [Example of Inserting and Activating a Site-Defined Collation Sequence](#).

2. Log on to the Teradata Database as user SYSADMIN or as user DBC.
3. Use the following statement to verify the contents of database DBC:

```
HELP DATABASE DBC;
```

The following objects must appear in the HELP listing for DBC to implement a site-defined collation sequence:

- A table named CollationTbl
- A view named CollationsV
- The following macros:
 - CollAddStandard
 - CollInstallMulti

If these objects are not in the DBC database, use the DIP utility to create them.

4. Run the BTEQ batch job you created in step 1.

The row you insert into the DBC.CollationsV view automatically populates the DBC.CollationTbl.

Activating Collation Codes

To activate the collation codes you inserted using the DBC.CollationsV view, perform the following procedure:

1. Execute the CollInstallMulti macro, as follows:

```
EXECUTE CollInstallMulti ('collation_name');
```

where *collation_name* is the CollName you specified in your BTEQ job.

2. Restart the Teradata Database (enter the restart tpa command from the Database Window of the system console) to install the codes as the new default collation sequence.

From now on, the collation sequence you just installed is the default when the COLLATION option is set to MULTINATIONAL.

Example of Inserting and Activating a Site-Defined Collation Sequence

The following example shows a BTEQ batch job that inserts into the DBC.CollationsV view a 256-character row containing the hexadecimal codes for a site-defined collation sequence, and then executes the CollInstallMulti macro to activate those codes as the new default.

You must restart the Teradata Database to complete the installation.

```
/* First delete from DBC.CollationsV any      */
/* existing definitions and then insert      */
/* the new definitions                        */
DELETE FROM DBC.CollationsV
  WHERE CollName = 'NON_STANDARD';
INSERT INTO DBC.CollationsV
( CollName
```



```

, CollInstall
, CollEqvClass
, CollOrderCS
, CollOrderUC )
VALUES
( 'NON_STANDARD'
, 'N'
/* CollEqvClass. Equivalence class.          */
/*  0 1 2 3 4 5 6 7 8 9 A B C D E F          */
, '000102030405060708090A0B0C0D0E0F'xb /* 0 */
|| '101112131415161718191A1B1C1D1E1F'xb /* 1 */
|| '202122232425262728292A2B2C2D2E2F'xb /* 2 */
|| '303132333435363738393A3B3C3D3E3F'xb /* 3 */
|| '406162636465666768696A6B6C6D6E6F'xb /* 4 */
|| '707172737475767778797A7B7C7D7E7F'xb /* 5 */
|| 'B56162636465666768696A6B6C6D6E6F'xb /* 6 */
|| '707172737475767778797A7B7C7D7E7F'xb /* 7 */
|| 'EAEBCEDBBBCBD80BFC0C1C2C3C4C5C6'xb /* 8 */
|| 'C7C8C9CACBCCDCDEEEFF0CFD0D1D2D3'xb /* 9 */
|| 'F1D4D5D6F2D7F3D8D9DADBDCF4F5F6F7'xb /* A */
|| 'DDDEDFE0F8E1E2E3F9E4E5E6E7E8FAE9'xb /* B */
|| '6161616161AEAB63656565656565656565'xb /* C */
|| 'FC6E6F6F6F6F6F6FAC7575757579FB73'xb /* D */
|| '6161616161AEAB63656565656565656565'xb /* E */
|| 'FD6E6F6F6F6F6F6FAC75757579FEFF'xb /* F */
/* CollOrderCS. Case-specific ordering      */
/*  0 1 2 3 4 5 6 7 8 9 A B C D E F          */
, '00000000000000000000000000000000'xb /* 0 */
|| '00000000000000000000000000000000'xb /* 1 */
|| '00000000000000000000000000000000'xb /* 2 */
|| '00000000000000000000000000000000'xb /* 3 */
|| '00010101010101010101010101010101'xb /* 4 */
|| '0101010101010101010101010000000000'xb /* 5 */
|| '00000000000000000000000000000000'xb /* 6 */
|| '00000000000000000000000000000000'xb /* 7 */
|| '00000000000000000000000000000000'xb /* 8 */
|| '00000000000000000000000000000000'xb /* 9 */
|| '00000000000000000000000000000000'xb /* A */
|| '00000000000000000000000000000000'xb /* B */
|| '030507090B0101030305070903050709'xb /* C */
|| '0003030507090B0D0103050709030002'xb /* D */
|| '020406080A0000020204060802040608'xb /* E */
|| '0002020406080A0C0002040608020000'xb /* F */
/* CollOrderUC. uppercase ordering          */

```

```

/*  0 1 2 3 4 5 6 7 8 9 A B C D E F          */
, '00000000000000000000000000000000'xb /* 0 */
|| '00000000000000000000000000000000'xb /* 1 */
|| '00000000000000000000000000000000'xb /* 2 */
|| '00000000000000000000000000000000'xb /* 3 */
|| '00010101010101010101010101010101'xb /* 4 */
|| '01010101010101010101010100000000'xb /* 5 */
|| '00010101010101010101010101010101'xb /* 6 */
|| '01010101010101010101010100000000'xb /* 7 */
|| '00000000000000000000000000000000'xb /* 8 */
|| '00000000000000000000000000000000'xb /* 9 */
|| '00000000000000000000000000000000'xb /* A */
|| '00000000000000000000000000000000'xb /* B */
|| '030507090B0101030305070903050709'xb /* C */
|| '0003030507090B0D0503050709030001'xb /* D */
|| '030507090B0101030305070903050709'xb /* E */
|| '0003030507090B0D0503050709030001'xb /* F */);
/* Now install this collation as the collation default */
EXECUTE CollInstallMulti ('NON_STANDARD');

```

Changing a Site-Defined Default

To change the default from your defined collation back to either Swedish, Norwegian, or one of the Japanese collations, execute the CollInstallMulti macro as explained in the previous passages.

To return to the Teradata Standard Multinational collation, execute the CollAddStandard macro as explained in previous passages.

Executing the CollAddStandard macro deletes your defined codes before reloading the codes for the Teradata Standard Multinational collation.

MULTINATIONAL Collation for Extended Site-Defined Character Sets

If the CHARSET_COLL collation and the supplied default MULTINATIONAL collation are not sufficient to support the desired language, the MULTINATIONAL collation can be modified to provide support for an extended site-defined character set.

Because modifying MULTINATIONAL collation is a complex task, avoid it if at all possible.

Modifying MULTINATIONAL Collation

Modify the following files in the TPA etc or TPA cfg directory:

- latin1multinationalcb.z

- latin1multinationalcs.z
- unicodemultinationalcb.z
- unicodemultinationalcs.z

Back up these files so that you can restore them, if required.

The latin1 files determine the Standard MULTINATIONAL default collation for LATIN and KANJI1 data, and the unicode files determine the MULTINATIONAL collation for Unicode data.

Note:

A site-defined MULTINATIONAL collation overrides the definition in the preceding files. For more information, see [Changing the Standard Multinational Default Collation](#) and [Defining Your Own Collation Sequence](#).

The files with suffix cb.z handle case blind (NOT CASESPECIFIC) collation. The files with suffix cs.z handle case specific (CASESPECIFIC) comparison.

Generate the cs.z files first, and then create the cb.z files from copies of the cs.z files. Next, modify the cb.z file so that one of the characters in a case pair matches the weight of the other character in the same case pair. If the files are not properly synchronized, then unpredictable results may ensue.

The format for each file is one character weighting definition per line. The first item is the character to be weighted (in a special hexadecimal Unicode-based format), followed by the primary and secondary weights (also expressed in a special hexadecimal Unicode-based format), which are separated by the semicolon character.

When used as a weight, the line of the file in which the Unicode character first occurs determines the relative weight, earlier characters indicate earlier weights. Forward references are not allowed.

As Unicode is a very large set, it may be best to attempt modifying existing files rather than starting from scratch. Since the startup routine that reads these files is very format sensitive, formatting should be followed precisely.

A % indicates the start of a comment that continues to the end of line. Blank lines and comment only lines are illegal.

Although part of the system, you can also find the collation files (multnatl.txt and blinddef.txt) here:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
 2. In the left pane, select **Attachments** to download the *International_Character_Set_Text_Files* zip file.
-

Note:

Do not install site-defined MULTINATIONAL collation if you take the file modification approach described here. That way MULTINATIONAL collation is exactly as defined by modified Standard Multinational Default collation files.

After you set up the files and DBC.CollationsV, you must perform a tpareset for the MULTINATIONAL collation changes to take effect.

Character Conversion

Overview

The following sections describe how the server:

- Converts inbound client character set data into server character set data
- Converts outbound server character set data into client character set data
- Determines export widths
- Handles conversion errors

Character Conversions

External Single-Byte Character Conversion

Conversion of client single-byte characters is handled as follows:

- Mapping is performed using ASCII, EBCDIC or active client character sets (as of the last restart of Teradata Database) in the DBC.CharTranslationsV view.
- The E2I conversion between client single-byte characters and Unicode is done as if the characters were first converted to LATIN strings and then from LATIN to Unicode.
- The E2I conversion between client single-byte characters and KANJISJIS is done as if the characters were first converted to LATIN strings and then the `TRANSLATE(string USING Latin_TO_KanjiSJIS)` function were applied.

External Mixed Single- and Multibyte Character Conversion

Conversion of a mixed single- and multibyte client character set is handled as follows:

- The conversion of mixed single-byte/multibyte character sets is performed using hard-coded tables, or algorithms, for the 3 supported character sets (KanjiEBCDIC, KanjiEUC, and KanjiShift-JIS).
- The DBC.CharTranslationsV view is only used for conversion of the single-byte characters in mixed single-byte/multibyte character sets. The table is not used for KANJISJIS.

Hard-Coded and Table Conversions

The following tables show when DBC.Translation, hard-coded algorithms, or other tables are used to perform conversions.

Single-Byte Client Character Set

IF the server character set is ...	THEN the DBC.Translation table defines ...
LATIN	a translation between the client character set and LATIN. All characters are accounted for using the table for conversion.
UNICODE	a translation between the client character set and UNICODE. All characters are accounted for using the table for conversion.
KANJISJIS	the translation for single-byte characters (those characters below 0x7F, that is, the single-byte characters of KANJISJIS). Multibyte character translation is hard-coded.
GRAPHIC	no translation. Conversion between single-byte character sets and GRAPHIC generates an error.
KANJI1	a translation between single-byte characters (those characters below 0x7F, that is, the single-byte characters of KANJISJIS) and Teradata extended JIS X 0201. Multibyte characters are not converted.

KanjiEUC

IF the server character set is ...	THEN the DBC.Translation table defines the translation ...
LATIN	but only for the Teradata extended JIS X 0201 characters of LATIN (those characters below 0x7F). Multibyte character translation is hard coded.
UNICODE	but only for the Teradata extended JIS X 0201 characters of UNICODE (single-byte characters below 0x7F). Multibyte character translation is hard coded.
KANJISJIS	but only for the Teradata extended JIS X 0201 part of KANJISJIS (single-byte characters below 0x7F). Multibyte character translation is hard-coded.
GRAPHIC	but only for the Teradata extended JIS X 0201 characters of GRAPHIC. Multibyte character translation is hard coded.
KANJI1	between single-byte characters (those characters below 0x7F, that is, the single-byte characters of KANJISJIS) and Teradata extended JIS X 0201. Multibyte characters are not converted.

KanjiShift-JIS

Translations between the KanjiShift-JIS client character set and server character sets are hard-coded.

KanjiEBCDIC

IF the server character set is ...	THEN the DBC.Translation table defines the translation ...
LATIN	but only for the Teradata extended JIS X 0201 characters of LATIN (those characters below 0x7F). Multibyte character translation is hard coded.
UNICODE	but only for the Teradata extended JIS X 0201 characters of UNICODE (single-byte characters below 0x7F). Multibyte character translation is hard coded.
KANJISJIS	but only for the Teradata extended JIS X 0201 part of KANJISJIS (single-byte characters below 0x7F). Multibyte character translation is hard-coded.
GRAPHIC	but only for the Teradata extended JIS X 0201 characters of Graphic. Multibyte character translation is hard coded.
KANJI1	between single-byte characters (those characters below 0x7F, that is, the single-byte characters of KANJISJIS) and Teradata extended JIS X 0201. Multibyte characters are not converted.

Application Model With Single Form-of-Use

Whenever a client application talks to the Teradata Database, it indicates its character set (form-of-use for character data) and the Teradata Database returns all character types to the application in that form.

Note that in application programs (within ANSI C programs, for example), Kanji strings can be represented as character strings with mixed single-byte character/multibyte character form-of-use or as strings with fixed width form-of-use.

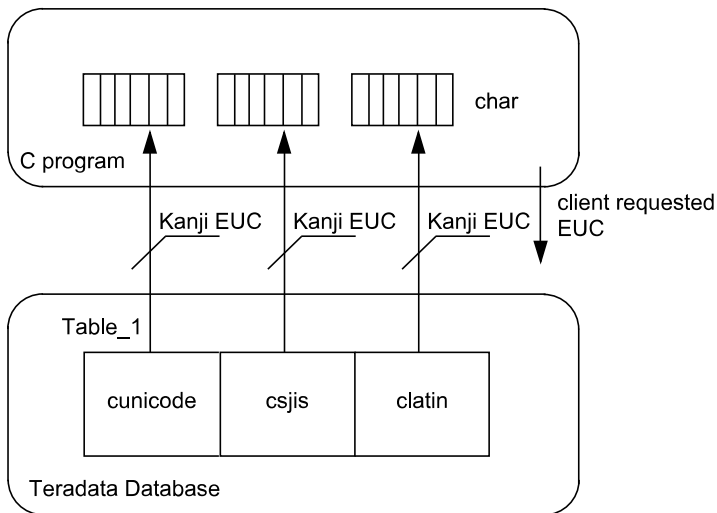
Example: KanjiEUC Form-of-Use

Suppose that table_1 has three columns named cunicode, csjis, and clatin with the respective corresponding repertoires UNICODE, KANJISJIS, and LATIN. Suppose further that the current client character set is KanjiEUC.

If you submit the following statement, then the data from all columns will arrive into the application in the KanjiEUC form-of-use.

```
SELECT cunicode, csjis, clatin FROM table_1;
```

Because of this, the application should use variables declared as mixed single-byte character/multibyte character strings in order to store the data from the Teradata Database.

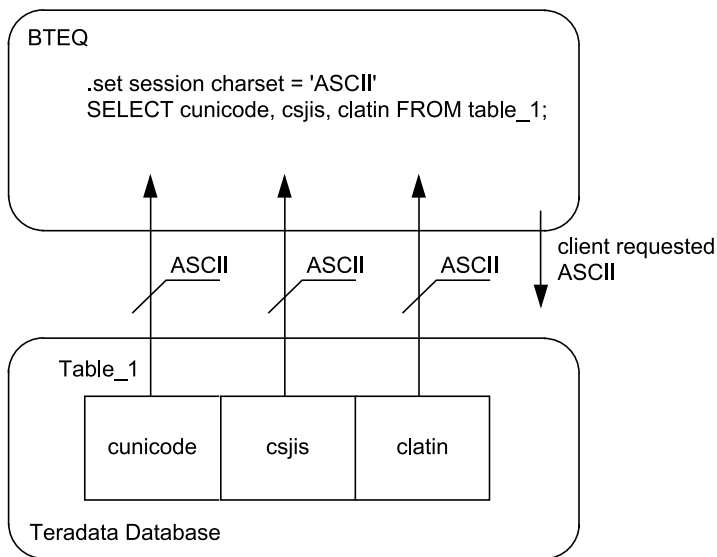


Example: Data Returned in ASCII

Consider a BTEQ session with the ASCII client character set.

The data from the cunicode, csjis, and clatin columns are returned to the BTEQ session in ASCII, as illustrated by the following figure.

Note that your application should declare its variables as mixed single-byte strings to store the data from the Teradata Database.



Because the ASCII repertoire is much smaller than the repertoires of cunicode or csjis columns, some characters might be converted to error characters on input, some might be converted to something else on export, or the transaction might abort.

Working with Export Widths

About Export Widths

The number of characters or bytes the Teradata Database exports can exceed or be less than the number of characters or bytes the client application expects to receive unless the export width is set properly.

The export width defines how the system reserves space for each character field in result set rows returned to Teradata clients. The global export width is set by default, but you can change the setting. In addition, users with the required privileges can override the global setting with a session-level export width setting.

Note:

64000 bytes is the maximum possible for an exported character string.

Setting the Global Export Width

You can use the DBS Control utility to set field 23, the Export Width Table ID field, to define the default export width definition the system uses when returning a result set to a client.

Export Width Table ID	DBSControl Setting	Description
Expected Defaults (default setting)	0	Provides reasonable default widths for the character data type and client form of use.

Export Width Table ID	DBSControl Setting	Description
Compatibility Default	1	Enables Unicode data to be processed by applications that written to process Latin or KANJI1 data.
Maximum Default	2	Provides maximum default width of the character data type and client form of use.

Also see the next topic, [About Export Width Rules](#), and “Export Width Table ID” in *Teradata Vantage™ - Database Utilities*, B035-1102.

About Export Width Rules

The character conversion parameters for all export width definitions, including any custom definitions you create, are stored as a rule set in the DBC.ExportWidth table, for example:

```
SELECT * FROM DBC.ExportWidth;
```

ExportDefinitionName	ExportWidthRuleSet
-----	-----
EXPECTED	1112211111222232222211121111112222322222
MAXIMUM	1322323221332232322321123111122322323222
COMPATIBILITY	1112211111111231111111211111112222322222
CustomDefinition1	4112211111222232222211121111112222322224

Note:

Also see [Creating a Custom Export Width Definition](#).

Interpreting the Export Width Rule Set for EXPECTED

To determine the export width for a character set combination:

1. Identify the ExportWidthRuleSet values that correspond to the server character set; for example, for the DBS Control export definition setting EXPECTED.

Server Character Set	Applicable Export Width Rule Set
LATIN	First set of 10 digits (1112211111)
UNICODE	Second set of 10 digits (2222322222)
KANJISJIS	Third set of 10 digits (1112111111)
GRAPHIC	Fourth set of 10 digits (2222322222)

Note:

The ExportWidthRuleSet for a server character set is always in the same position for any ExportDefinitionName. For example, LATIN is always the first set of 10 digits.

- Within each set of 10 digits for a server character set, you can find the export width conversion multiplier for a session character set.

The position in each group of 10 digits...	Is the Conversion Multiplier for...
First	<ul style="list-style-type: none"> Any session character set that ends in the string '_0I' The session character set 'KATAKANAEBDIC'
Second	Any session character set that ends in the string '_0U'
Third	Any session character set that ends in the string '_0S'
Fourth	The session character set 'UTF16'
Fifth	The session character set 'UTF8'
Sixth	Any site-defined session character set with STATEMACHINE EUC1211
Seventh	Any site-defined session character set with STATEMACHINE EUC1223
Eighth	Any site-defined session character set with STATEMACHINE S80
Ninth	Any site-defined session character set with STATEMACHINE S80A1E0
Tenth	Any site-defined session character set with STATEMACHINE SOSI0E0F

For example:

- When exporting from the server character set LATIN (first set of 10 digits, 1112211111), for the session character set UTF8 (the fifth position in the set), the export width is 2.
- When exporting from the server character set UNICODE (second set of 10 digits, 2222322222) for the session character set UTF8 (the fifth position in the set), the export width 3.
- When exporting from the server character set KANJISJIS (third set of 10 digits, 1112111111), for the session character set UTF16 (the fourth position in the set), the export width is 2.

Export Width Space Considerations

The export width value is a multiplier, so that if data in the server character set is n bytes, and the export width is 2, the same data in the result set occupies $2n$ bytes.

Viewing the Current Export Widths for a Session

You can use `HELP SESSION` to view the current export widths for the session. Based on the rule set shown in [About Export Width Rules](#), and an EXPECTED user default export width, the request returns output similar to the following:

```

HELP SESSION ;
User Name SPEC2
Account Name DBC
Logon Date 02/08/11
Logon Time 10:58:56
Current DataBase DBC
Collation UTF8
Character Set UTF8
Current DateForm IntegerDate
Session Time Zone -08:00
Default Character Type LATIN
Export Latin 2
Export Unicode 3
Export Unicode Adjust 0
Export KanjiSJIS 1
Export Graphic 3
Export Width Rules 2313
...

```

where:

Value	Description
Export Latin 2	Reports the export width value for each server character data type for the session character set (UTF8). Note: A 0 value means that character conversion between server and client character sets is not possible, for example, from Graphic (double-byte) to ASCII (single-byte).
Export Unicode 3	
Export Unicode Adjust 0	
Export KanjiSJIS 1	
Export Graphic 3	
Export Width Rules 2313	The export width values for converting the session character set to Latin, Unicode, KanjiSJIS, and Graphic.

Note:

No export width is reported for the Kanji1 character data type because the export width is always 1.

Creating a Custom Export Width Definition

You can create a custom export width definition by executing the ReplaceExportDefinition stored procedure, which is included with an installation of or upgrade to the current Teradata Database release.

Note:

The user executing the stored procedure must have EXECUTE privileges on the stored procedure or on the SYSLIB database.

1. Supply values for the required stored procedure arguments. For example:

```
Ex. call syslib.ReplaceExportDefinition('export_definition_name',
'export_width_rule_set'xb,msg);
```

where:

Stored Procedure Argument	Description
<i>export_definition_name</i>	A unique name for the custom export definition that conforms to the requirements for Teradata Database object names.
<i>export_width_rule_set</i>	A byte string of 40 hexadecimal digits; 10 digits for each of the 4 server character sets. For the meaning of each position n the string, see About Export Width Rules . Each position can have a value of 1, 2, 3, or 4, except the fourth position in each 10 character set (UTF16), which must have a value of either 2 or 4.
xb	The standard literal suffix. <ul style="list-style-type: none"> • x = hexadecimal • b = bytes
msg	The output value of the stored procedure.

2. Execute the stored procedure.

The system returns an error message for any arguments that are not valid. If all arguments are valid, the system writes the new rule set to the DBC.ExportWidth table.

Error Messages for CreateExportDefinition

The following standard error messages are generated by the ReplaceExportDefinition stored procedure when it detects syntax that is not valid.

Export Definition Name Not Valid

Example input:

```
Ex. call syslib.ReplaceExportDefinition('EXPECTED',
'4112211111222232222211121111112222322224'xb,results);
```

Error message:

Cannot create or update the reserve ExportDefinitionName (EXPECTED/COMPATIBILITY/MAXIMUM)

Export Width Range Not Valid

Example input:

```
Ex. call syslib.ReplaceExportDefinition('aaa',
'5672211111222232222211121111112222322224'xb,results);
```

Error message:

Valid range of values for the ExportWidthRuleSet are 1,2,3,4'

Export Width Not Valid for a Location

Example input:

```
Ex. call syslib.CreateExportDefinition('testing',
'4113211111222232222211121111112222322224'xb,results);
```

Error message:

At position fourth of ExportWidthRuleSet, valid value is 2 or 4)

Modifying a Custom Export Width Definition

You can modify an export definition with the ReplaceExportDefinition stored procedure, using the same arguments shown in [Creating a Custom Export Width Definition](#).

1. Specify the name of an existing definition that you want to modify.

Note:

You can modify only custom definitions that you create, and cannot modify the global default definitions EXPECTED, COMPATIBILITY and MAXIMUM.

2. Specify a new export width rule set. All 40 digits are required, even if some of them have not changed.
3. Execute the stored procedure.

The new definition overwrites the old definition in the DBC.ExportWidth table.

The new export width definition takes effect at the next transaction for each affected user. In process transactions continue with the existing export width definition.

For information about the DBC.ExportWidth table, see *Teradata Vantage™ - Data Dictionary*, B035-1092.

Deleting a Custom Export Width Definition

You can delete an existing custom export width definition using the `RemoveExportDefinition` stored procedure, for example:

```
Ex. call syslib.RemoveExportDefinition('exportname',msg);
```

The system deletes the specified definition from the `DBC.ExportWidth` table. In process transactions using the removed definition continue, but it is unavailable to new transactions.

Note:

The user executing the stored procedure must have `EXECUTE` privileges on the stored procedure or on the `SYSLIB` database.

Setting the Export Width for a User

You can define the source of export widths for a user with the `CREATE USER` or `MODIFY USER` statement, for example:

```
CREATE USER username
...
EXPORTWIDTH = 'export_width_definition_name';
```

where `export_width_definition_name` can be:

- One of the standard export width tables used for `DBSControl` settings:
 - `EXPECTED`
 - `COMPATIBILITY`
 - `MAXIMUM`
- `DEFAULT` (defers to the current `DBSControl` global setting)
- A custom `ExportDefinitionName` name. See [Creating a Custom Export Width Definition](#).

See “`CREATE USER`” in *Teradata Vantage™ - SQL Data Definition Language Syntax and Examples*, B035-1144.

System Handling of the Export Width for a Session

When determining the export width for a session:

- If the user initiating the session has an assignment in the `EXPORTWIDTH` option for the user definition, the system uses the assigned export width definition.

- If the user initiating the session does not have an assignment in the EXPORTWIDTH option in the user definition, the system uses the global default export width definition specified in the DBSControl Export Width Table ID field.

Note:

The session-level table ID is automatically saved in *DBC.SessionTbl* and is restored after a restart.

Exception Handling

Types of Exceptions

Conversion of character data type between a client form-of-use and the Teradata Database form-of-use may encounter exception situations.

There are three possible exceptions.

Type	Description	See
Illegal client data form-of-use	An illegal form-of-use occurs when the client data violates the structure of the client form-of-use, for example, an unmatched SI character in KanjiEBCDIC.	Illegal Form-of-Use
Conversion exception	A conversion exception occurs when there is no character in the target repertoire that corresponds to a character from the source string.	<ul style="list-style-type: none"> • E2I Conversion Exceptions • I2E Conversion Exceptions
Truncation exception	A truncation exception occurs when the source cannot fit into the target field without truncation of a non-pad character.	Truncation Exceptions

For information on how exceptions are handled in sessions where Unicode Pass Through is enabled, see [Unicode Pass Through](#).

Exception Precedence

There is no precedence of one exception over the other. When the illegal form-of-use, truncation-exception, and conversion-exception occur simultaneously, whichever exception is recognized first is the one reported.

Errors and Transaction Modes

For an SQL session operating in Teradata mode, an error aborts the current transaction.

An ANSI mode session follows the ANSI transaction semantics.

Errors and Load Utilities

In the case of the load utilities (MultiLoad and FastLoad), the offending record is placed into the error tables used by that utility.

The offending characters, rather than any error characters, are stored in the error tables. The Teradata Database records the field that caused the problem, and the load utility is allowed to proceed.

Illegal Form-of-Use

See [Ill-formed Code Unit Sequences](#) for information about how ill-formed code unit sequences are handled in sessions with Unicode Pass Through (UPT) enabled.

In sessions where UPT is not enabled, the illegal form-of-use exception always generates an illegal form-of-use error. External single-byte character sets do not have an illegal form-of-use.

Illegal Form-of-Use Examples

Consider the `s1`string for the examples under this topic, where:

`s1=92 年 03 月 10 日`

If the string was somewhat modified as in the following examples, an illegal form-of-use error is returned because it is malformed.

Example: KanjiEBCDIC

The first SI is missing, as the following table illustrates.

F9	F2	0E	45	60	F0	F3	0E
9	2	<	年		0	3	<

Example: KanjiEUC

The second byte of 日 is missing.

39	32	C7	AF	30	33	B7	EE	31	30	C6
9	2	年		0	3	月		1	0	not valid

Example: KanjiSJIS

The second byte of 年 is missing.

39	32	94	30	33
9	2	z	0	3

Example: UTF8

The second and third byte of 年 are missing.

00	30	00	32	E5
9	2			not valid

E2I Conversion Exceptions

For information about External-to-Internal (E2I) conversions and how exceptions are handled in sessions where Unicode Pass Through is enabled, see [Importing Unicode Characters in a Pass Through Session](#).

E2I conversion generates an exception when a character in the client repertoire cannot be converted to the server repertoire.

For example, an attempt to insert a non-Latin character into a character column where the server character set is LATIN generates a conversion exception.

E2I conversion exceptions generate an SQL error.

Conversion to KANJI1

Conversion from non-Japanese multibyte client character sets to KANJI1 generates an exception when a character is outside the 7-bit ASCII range of KANJI1.

Conversion of multibyte characters from the Japanese client character sets to KANJI1 is more forgiving than the conversion to UNICODE. This is because earlier Teradata Database Japanese language releases only range-checked multibyte characters during E2I conversion, allowing undefined code points from the client character sets.

Conversion to GRAPHIC

The E2I conversion from single-byte characters in any client character set to GRAPHIC always generates an SQL error. This error may be masked when constants are used because constants are processed internally as UNICODE.

For example, it is possible to insert GRAPHIC data (via '..' XC) from an ASCII character set into a GRAPHIC column; however, selecting that GRAPHIC column from the same ASCII character set returns an error.

I2E Conversion Exceptions

For information about Internal-to-External (I2E) conversions and how exceptions are handled in sessions where Unicode Pass Through is enabled, see [Exporting From a UNICODE Server Character Set](#).

I2E conversion generates an exception when a character in the server repertoire cannot be converted to the client repertoire.

For example, an attempt to convert a multibyte KANJISJIS character to an ASCII character generates a conversion exception.

For most I2E character conversion, exceptions are ignored and the offending character is converted to the predefined error character for the target client character set.

For information on error characters, see [Error Characters](#).

Conversion from GRAPHIC

The I2E conversion from GRAPHIC to single-byte characters in any client character set always generates an SQL error.

To get the hexadecimal representation for GRAPHIC characters, you can use CHAR2HEXINT. For more information on CHAR2HEXINT, see *Teradata Vantage™ - SQL Functions, Expressions, and Predicates*, B035-1145.

Conversion from KANJI1

Conversion from KANJI1 to non-Japanese multibyte client character sets generates an exception when a character is outside the 7-bit ASCII range of KANJI1.

Error Characters

Each client character set and each server character set includes an error character.

For information about how error characters are used and handled in sessions with Unicode Pass Through enabled, see [Unicode Pass Through](#).

Client Error Characters

The following table lists the error characters associated with each client character set.

Client Character Set	Error Character
<ul style="list-style-type: none"> • ASCII character set • KanjiSJIS_0S character set • KanjiEUC_0U character set 	0x1A
<ul style="list-style-type: none"> • EBCDIC character set • KanjiEBCDIC character sets 	0x3F

Client Character Set	Error Character
<ul style="list-style-type: none"> • KanjiSJIS_0S character set, GRAPHIC data (Returned only when exporting in Record or Indicator modes.) • KanjiEUC_0U character set, GRAPHIC data (Record and Field modes only.) • KanjiEBCDIC character sets, GRAPHIC data (Record mode only.) 	0xFFFD
<ul style="list-style-type: none"> • Site-defined single-byte client character sets • TCHBIG5_1R0 • SCHGB2312_1T0 • HANGULKSC5601_2R4 • SDSCHGB2312_2T0 • SDTCHBIG5_3R0 • SDHANGULKSC5601_4R4 • KANJI932_1S0 • SCHINESE936_6R0 • TCHINESE950_8R0 • HANGUL949_7R0 • THAI874_4A0 • LATIN1250_1A0 • CYRILLIC1251_2A0 • HEBREW1255_5A0 • ARABIC1256_6A0 • LATIN1254_7A0 • LATIN1258_8A0 • LATIN1252_3A0 	DBC.Translation. TranslateOut of 0x1A
<ul style="list-style-type: none"> • SCHEBCDIC935_2IJ • TCHEBCDIC937_3IB • HANGULEBCDIC933_1II • SDSCHEBCDIC935_6IJ • SDTCHEBCDIC937_7IB • SDHANGULEBCDIC933_5II 	DBC.Translation. TranslateOut of 0x3F

Server Error Characters

The following table lists the error characters for the server character sets.

Form-of-Use	Error Character
LATIN server character set	0x1A
KANJI1 server character set	0x1A
KANJISJIS server character set	0x1A
UNICODE server character set	U+ FFFD

Form-of-Use	Error Character
GRAPHIC server character set	U+ FFFD
GRAPHIC server character set, VARGRAPHIC function	U+ F8FF

Note:

0x1A in LATIN/KANJI1/KANJISJIS and U+FFFD in UNICODE/GRAPHIC are used internally by Teradata as the error character; therefore, they are unusable as user data. The user cannot store or retrieve these values through Teradata.

Site-Defined Character Sets

For other site-defined character sets, the DBC.Translation table defines the translation of single-byte character sets and the single-byte character portion of Kanji character sets. The error character for these character sets is derived from how they translate to the LATIN and KANJI1 error character 0x1A.

This translation is defined in the column TranslateOut of the DBC.Translation table.

Finding an Unknown Error Character

You can determine this character using the following SELECT statement.

```
SELECT SUBSTR(TRANSLATEOUT,27,1)
FROM DBC.TRANSLATION
WHERE TRANSLATENAME = character_set;
```

Conversion of Error Characters

Internal conversion between the server character sets preserves the error character.

For example, when translating from LATIN to UNICODE to KANJISJIS, 0x1A translates to U+FFFD which translates to 0x1A.

For the KANJI1 character set, conversion of multibyte characters never causes an error.

Example: Character Cannot Be Represented in a Character Set

Suppose that a column declared as CHAR(2) CHARACTER SET UNICODE stores CAPITAL LETTER A WITH MACRON (Ä).

The Teradata Database representation is as follows.

01	00	00	20
A			

Selection from a KanjiEUC session successfully returns Ā encoded in code set 3 (cs3) of EUC. The KanjiEUC representation is as follows.

8F	AA	A7	20
A			

Assume that an ASCII session does a SELECT on the column containing the Ā. The export width for the session is set to the expected defaults and the conversion options are set to ignore conversion.

Because the character cannot be represented in ASCII, it is converted to the ASCII error character 0x1A. The representation of the results in ASCII is as follows.

1A	20
not valid	

Selection from KanjiSJIS returns an error character.

1A	20	20	20
not valid			

Selection from KanjiEBCDIC also returns an error character.

3F	40	40	40	40	40
not valid					

Truncation Exceptions

Truncation Exceptions and Fixed Length Columns

Fixed length columns are awkward to use in Kanji environments. The reason is that ANSI SQL treats fixed length strings in terms of characters, while most applications treat fixed length strings in terms of physical bytes. The difference cannot be reconciled.

Teradata SQL addresses this problem by treating LATIN, UNICODE and GRAPHIC character data type columns according to ANSI, but also providing the export width rules to export them as fixed byte length fields. For more information about export width, see [Working with Export Widths](#).

Truncation Exceptions When Converting Fixed Length Columns

Sharing the fixed length KANJI SJIS server character set fields involves a different problem. The applications need to place enough padding within the field so that when it is converted to a different

form-of-use, say KANJISJIS to KanjiEUC character set, the meaningful characters in the string are not truncated.

Truncation due to conversion and length change does not occur for KANJI1 server character set fields.

Because of the length change problem, you should use varying character fields when different representations can vary in length.

Truncation of a Pass Through Character (PTC)

In sessions where Unicode Pass Through is enabled, truncation of a PTC occurs when the destination buffer does not have 4 bytes to contain the entire surrogate pair. If only 2 bytes remain in a fixed CHAR string, then the string is padded with a SPACE (U+0020) in Teradata mode. In ANSI mode, the truncation mode error is returned.

Other Truncation Exceptions

Truncation exceptions can occur for character data types during the following operations:

- During external-to-internal conversion and assignment of USING data into a column with the SQL INSERT statement or using load utilities.
- Explicit conversion using the SQL CAST or CHARACTER clauses.
- On internal-to-external conversion of data exported to the client.

Truncation and Transaction Mode

The behavior of the truncation depends upon whether the conversion occurs while the Teradata Database is operating in Teradata mode or ANSI mode. For more information, see [Truncation Exceptions in ANSI Mode](#) and [Truncation Exceptions in Teradata Mode](#).

In this mode ...	Truncation of this server character set is as follows ...
ANSI	non-pad character data causes an error.
Teradata	character data is truncated without reporting an error.

Truncation Exceptions in ANSI Mode

In ANSI mode, truncation of non-pad characters assigned to a column returns an error. The pad character is determined by the target column type.

- If during external-to-internal or explicit conversion the data to be placed in a fixed character column expands so that the column length is exceeded with non-pad characters, the Teradata Database returns an error.
- If the fixed character column length is exceeded with pad characters, the Teradata Database truncates the input at the length of the column.

- If data to be placed in a varying column expands so that the maximum column length is exceeded with non-pad characters, the Teradata Database returns an error.
- If the varying character column maximum length is exceeded with pad characters, the Teradata Database truncates excess trailing pad characters.

Internal-to-external conversion is not affected by the mode of the Teradata Database. ANSI mode works the same as Teradata mode in this respect.

Truncation Exceptions in Teradata Mode

In Teradata mode, truncation can occur without error in many operations, for example:

- If, during external-to-internal or explicit conversions, the data to be placed in a fixed character column expands so that the column length is exceeded, Teradata Database truncates on the logical character boundary, and if necessary, pads to the full length of the column. Padding is done with the pad character of the target column type and occurs only on fix columns.
- If data to be placed in a varying column expands so that the maximum column length is exceeded, Teradata Database truncates the input on the logical character boundary so that the maximum length of the column is not exceeded.
- During internal-to-external conversion, a truncation exception can occur due to the limited output length imposed by the defined export width. Those exceptions are ignored as they were in previous releases.
- Suppose the length designated by the export width is defined to be *l*. If during internal-to-external conversion data expands so that *l* is exceeded, then the Teradata Database truncates on the logical character boundary and for fixed characters, pads to length *l*. Truncation on the logical boundary is with respect to the client form-of-use. Padding is done with the pad character of the client form-of-use.

For an illustration of truncation with padding, see [Example: KanjiEUC](#). For other truncation examples, see [Example: KanjiEBCDIC](#) and [Example: KanjiSJIS](#).

Truncation Examples

Scenario 1

Consider a half-square Katakana string s_4 which means 'text': テキスト (pronunciation te-ki-su-to). Suppose that a KanjiSJIS session stores s_4 in a column csjis declared as CHAR(4) CHARACTER SET UNICODE.

Example: Unicode

The Teradata Database representation of the s_4 string in UNICODE.

FF	83	FF	77	FF	7D	FF	84
テ		キ		ス		ト	

Example: KanjiEBCDIC

Because Katakana characters are single-byte character in KanjiEBCDIC, a selection from a KanjiEBCDIC session is complete.

94	87	8E	95
テ	キ	ス	ト

Scenario 2

Consider string s_5 containing a half-square Katakana and a Kanji character ア 山 (the half-square Katakana 'a' followed by the Kanji 'yama' for mountain).

Suppose that a KanjiSJIS session stores s_5 in a column csjis declared as CHAR(3) CHARACTER SET KANJISJIS.

Example: KanjiSJIS

The Teradata Database representation of the s_5 string in KANJISJIS is:

B1	8E	52
ア	山	

Example: KanjiEUC

If column csjis is exported from a KanjiEUC session with the export width rule as n, the following data is exported (observe that ア expanded to two bytes and truncation on logical character boundary occurred for 山 and padding took place).

8E	B1	20
ア		

Unwanted Export Truncation

While the defaults are set to avoid this in most cases, it is possible for the EXPORT facility to produce unwanted data truncation.

Under the expected-defaults, this can only occur in the following circumstances:

- During a KanjiEUC session where cs3 characters exist in a string without a corresponding number of cs0 characters.

- During a KanjiEBCDIC session where a string exists with highly interspersed single-byte and multibyte characters.

The examples that follow illustrate how unwanted truncation can occur during the export process.

Note that you can avoid these types of truncation by using the CAST function. For information, see *Teradata Vantage™ - Data Types and Literals*, B035-1143.

KanjiEUC Example

For example, consider the following SELECT statement from a KanjiEUC session.

```
SELECT 'Âgé';
```

The string Âgé is translated internally to UCS2.

00	C2	00	67	00	E9
Â		g		é	

Because this is three characters, the expected-defaults exports at most six bytes, but the representation of Âgé in KanjiEUC requires seven bytes.

8F	AA	A4	67	8F	AB	B1
Â			g	é		

Therefore, the output must be truncated to six bytes.

Truncation occurs at a character boundary, so the output is truncated to Âg, which requires four bytes.

If this were a fixed width field, these four bytes would be padded with SPACE characters to produce six bytes. Because the field is not fixed width, only four bytes are output (excluding the length information which is always included in variable width data).

8F	AA	A4	67
Â			g

KanjiEBCDIC Example

For KanjiEBCDIC, the problem occurs when a single-byte character is placed between two sequences of multibyte characters.

For example, consider the following SELECT statement.

```
SELECT '平成 1 年';
```

平成 1 年, where the 1 is a single-byte character, is translated internally to UCS2.

5E	73	62	10	00	31	5E	74
平		成		1		年	

Because this is four characters, the expected-defaults exports at most ten bytes, but the representation of 平成 1 年 in KanjiEBCDIC requires eleven bytes.

0E	45	8D	45	BA	0F	31	0E	45	60	0F
<	平		成		>	1	<	年		>

Therefore the output must be truncated to ten bytes.

Truncation occurs at a character boundary, so the output is truncated to 平成 1, which requires seven bytes.

If this were a fixed width field, these seven bytes would be padded with SPACE characters to produce ten bytes. Because the field is not fixed width, only seven bytes are output (excluding the length information which is always included in variable width data).

0E	45	8D	45	BA	0F	31
<	平		成		>	1

Unicode Pass Through

Unicode Pass Through Overview

Unicode Pass Through (UPT) is a Unicode error handling feature. The feature gives users the ability to allow Pass Through Characters (PTCs) to be imported into and exported from Teradata. This increases the repertoire of characters available to Unicode users and provides the ability to store, retrieve, and analyze Emoji and other ideographs.

PTCs consist of the following:

- Characters not supported by Teradata:
 - Characters in BMP (Basic Multilingual Plane) from Unicode versions 6.1.0 to 9.0.0
 - Characters in SMP (Supplementary Planes 1 to 16) from all Unicode versions

Note:

These are the current unsupported characters in Teradata. This may change in the future if Teradata adds support for additional characters.

- Unassigned characters
- Private use characters

Unassigned characters are not assigned in the Unicode Standard and not supported by any implementation. Private use characters will never be assigned in the Unicode Standard and the usage is determined by a private agreement among cooperating implementations.

The following are not PTCs:

- 6.0 BMP characters supported by Teradata
- The set of noncharacter code points

Noncharacters are code points that are permanently reserved for internal use.

Note:

You should not use UPT if you rely on Teradata to screen out Teradata unsupported characters or the REPLACEMENT CHARACTER (U+FFFD).

Examples of PTCs and Noncharacters

Unicode Code Point	Plane	UTF-16 in hex	Unicode Name, <noncharacter>, or <unassigned>	Notes
U+20BD	0	20BD	RUBLE SIGN	Basic Multilingual Plane (BMP)

Unicode Code Point	Plane	UTF-16 in hex	Unicode Name, <noncharacter>, or <unassigned>	Notes
U+1F600	1	D83D DE00	GRINNING FACE	Supplementary Multilingual Plane (SMP)
U+20000	2	D840 DC00	CJK Unified Ideographs Extension B	Supplementary Ideographic Plane (SIP)
U+35555	3	D895 DD55	<unassigned>	Tertiary Ideographic Plane (TIP). No assigned characters.
U+46666	4	D8D9 DE66	<unassigned>	No characters defined on plane 4
U+5FFFE	5	D93F DFFE	<noncharacter>	No characters defined on plane 5
U+6FFFF	6	D97F DFFF	<noncharacter>	No characters defined on plane 6
U+E0103	14	DB40 DD03	VARIATION SELECTOR-20	Supplementary S
U+FFFF0	15	DBBF DFF0	<unassigned>	Supplementary Private Use Area-A
U+10FFFD	16	DBFF DFFD	<unassigned>	Supplementary Private Use Area-B

Enabling and Disabling Unicode Pass Through

Unicode Pass Through is enabled or disabled at the session level and is valid only for Unicode sessions.

For sessions using the UTF8 or UTF16 client character sets, you can use the `SET SESSION CHARACTER SET UNICODE PASS THROUGH [ON|OFF]` statement to enable or disable UPT.

For example, `SET SESSION CHARACTER SET UNICODE PASS THROUGH ON` establishes a Pass Through Session (PTS).

The `SET SESSION` statement can also be added to the startup string as follows:

```
CREATE USER Unicode_upt_user AS PERM = 1e6,
PASSWORD = emoji,
STARTUP = 'SET SESSION CHARACTER SET UNICODE PASS THROUGH ON;';
```

By default, Unicode Pass Through is not enabled.

Example: Enabling UPT Using BTEQ

You can enable UPT using the `SET SESSION SQL` statement.

```
.SESSIONS 5
.LOGON mysystem/uptuser,emoji
```

```
.REPEAT 5
SET SESSION CHARACTER SET UNICODE PASS THROUGH ON;
```

Example: Enabling UPT in Teradata Parallel Transporter (TPT)

Use the following TPT syntax to enable UPT:

```
VARCHAR UnicodePassThrough = 'ON',
```

This syntax only applies to the following TPT operators:

- Load operator
- Update operator
- Export operator
- Stream operator
- DDL operator
- SQL Inserter operator
- SQL Selector operator

Note that the default value is OFF.

Example: Enabling UPT Using JDBC

You can enable UPT using a user startup string:

```
create user uptuser as password = emoji, startup = 'set session character set
unicode pass through on';
```

```
Connection con =
DriverManager.getConnection("jdbc:teradata://dbs/RUNSTARTUP=ON",
"uptuser", "emoji");
```

You can enable UPT using the SET SESSION SQL statement:

```
stmt.executeUpdate("set session character set unicode pass through on");
```

Unicode Pass Through Functionality

UPT provides the following functionality in Pass Through sessions:

- Proper storage of all assigned and unassigned Unicode characters from U+0000 to U+10FFFD.
- Ability to store PTCs in Unicode columns as data and retrieve it as entered. For example:

```
INSERT INTO currency (U&'#20BD' UESCAPE '#');
```

- Ability to use PTCs in queries. For example:

```
SELECT feed FROM twitter WHERE feed LIKE U&'%#+01F602%' UESCAPE '#';
```

- Noncharacters and ill-formed code unit sequences are changed to the REPLACEMENT CHARACTER (U+FFFD) which can pass through and be stored in Teradata.

With UPT enabled, users can expect translation errors to be reduced or eliminated.

Example: Using Pass Through Character Literals

This example shows how you can use pass through character (PTC) literals in a session with Unicode Pass Through enabled. The PTC literals are emoji characters, and the session character set is UTF16.

To run these examples, replace the emoji graphics with the corresponding Unicode code points (such as U+1F601) inserted using your client application.

```

/*****
/* Enable Unicode Pass Through                                     */
*****/

```

```
SET SESSION CHARACTER SET UNICODE PASS THROUGH ON;
```

```

/*****
/* CREATE TABLE                                                 */
*****/

```

```

CREATE TABLE upt_tb10
(
  col1    INTEGER,
  col2    VARCHAR(6) CHARACTER SET UNICODE
) UNIQUE PRIMARY INDEX (col1);

```

```

/*****
/* Insert Pass Through Characters into the table                 */
*****/

```

```
/* U+1F601 = GRINNING FACE WITH SMILING EYES */
```

```
INSERT INTO upt_tb10 VALUES (1, '😄');
```

```
/* U+1F602 = FACE WITH TEARS OF JOY */
```

```
INSERT INTO upt_tb10 VALUES (2, '😂');
```

```
/* U+1F603 = SMILING FACE WITH OPEN MOUTH */
```

```
INSERT INTO upt_tb10 VALUES (3, '😊');
```

```
/* U+1F604 = SMILING FACE WITH OPEN MOUTH AND SMILING EYES */
```

```

INSERT INTO upt_tbl0 VALUES (4, ' 😊 ');
/* U+1F605 = SMILING FACE WITH OPEN MOUTH AND COLD SWEAT */
INSERT INTO upt_tbl0 VALUES (5, ' 😊 ');

```

```

/*****
/* DISPLAY ROWS                                     */
*****/

```

```
SELECT ' 😊 ';
```

Result:

```
, 😊 ,
```

```
-----
```

```
😊
```

```
SELECT * FROM upt_tbl0 WHERE col2 = ' 😊 ';
```

Result:

```
col1 col2
```

```
-----
```

```
2      😊
```

```
SELECT * FROM upt_tbl0 ORDER BY col2;
```

Result:

```
col1 col2
```

```
-----
```

```
1      😊
```

```
2      😊
```

```
3      😊
```

```
4      😊
```

```
5      😊
```

Importing Unicode Characters in a Pass Through Session

In a Pass Through session, Unicode characters (including PTCs) are imported to a UNICODE server character set as follows:

- Teradata supported characters in 6.0 BMP are processed as 2-byte UTF16.
- Teradata unsupported characters in 6.1 to 9.0 BMP and all SMP are converted to and processed as 2 or 4-byte UTF16.
- Unassigned and Private Use characters in the BMP and SMP are converted to and processed as 2 or 4-byte UTF16.
- Noncharacters are changed to the REPLACEMENT CHARACTER (U+FFFD).
- Ill-formed code unit sequences are changed to the REPLACEMENT CHARACTER (U+FFFD).
- CLOBs are supported.

In sessions where UPT is not enabled, importing a PTC results in an External-to-Internal (E2I) conversion exception and the query fails.

Note:

UPT only applies to Unicode data and sessions, meaning UTF8 and UTF16 sessions and the UNICODE server character set. For example, importing the 0x1A replacement character from an ASCII character set to a Unicode column will still be rejected even in a session enabled with the UPT feature.

Users should also be aware that once PTCs are imported into the database, they can percolate through the system so there is a possibility that sessions with UPT disabled may still encounter PTCs. If you want to remove PTCs from your system, you must delete or replace the PTCs.

Exporting From a UNICODE Server Character Set

In a Pass Through Session, PTCs exported to a UTF8 or UTF16 client character set are converted and exported without error.

In a session in which Unicode Pass Through is not enabled, PTCs to be exported to a UTF8 or UTF16 client character set are converted to the UTF8 or UTF16 client error character and exported without error.

In all sessions, Pass Through or not, PTCs that are exported to client character sets that are not UTF8 or UTF16 are converted to the client error character and exported without error.

Ill-formed Code Unit Sequences

Ill-formed code unit sequence errors occur when the byte values in a character string do not follow the encoding rules. These exceptions are detected while importing characters from a UTF8 or UTF16 session to a Unicode character string, or exporting from a Unicode character string to a UTF16 or UTF8 session.

UTF8 and Unicode

In a session where UPT is disabled, an error is returned when an ill-formed sequence error is encountered and processing stops.

In a Pass Through session, processing continues without error and a UNICODE REPLACEMENT CHARACTER (U+FFFD) is stored in the destination string for each ill-formed byte in the source string sequence.

The following table shows how UTF8 source strings are converted to UTF16 destination strings with replacement characters. Note that the replacement stops once a well-formed byte sequence is encountered.

Ill-formed UTF8 Source String	UTF16 Destination String With Replacement Characters
80	FFFD
C261	FFFD0061
E18065	FFFDFFFD0065
F1808062	FFFDFFFDFFFD0062
63F48164	0063FFFDFFFD0064
C3679068A0BF69	FFFD0067FFFD0068FFFDFFFD0069

Ill-formed sequence exceptions during export to UTF8 are rare and can only occur if the source Unicode string is not well-formed.

UTF16 and Unicode

In a session where UPT is disabled, an error is returned when an ill-formed sequence error is encountered and processing stops.

In a Pass Through session, the following occurs for ill-formed sequence errors:

- In cases where a High Surrogate is not followed by a Low Surrogate or a Low Surrogate is not preceded by a High Surrogate, processing continues without error and a UNICODE REPLACEMENT CHARACTER (U+FFFD) is stored in the destination string for each inappropriate surrogate. For example, the ill-formed UTF16 source string of 'D800D800' is converted to the UTF16 destination string with replacement characters of 'FFDFFFD'.
- If the length of the source UTF16 character string is odd on import, processing continues without error and a trailing UNICODE REPLACEMENT CHARACTER (U+FFFD) is stored in the destination string for the partial source byte. For example, the ill-formed UTF16 source string of 'D800' is converted to the UTF16 destination string with replacement characters of 'FFFD'.

Usage Notes

- You may have existing processes in place to cleanse Unicode data before loading it into Teradata Database, such as access modules and UDFs. To take full advantage of Unicode Pass Through, you must change or eliminate prior methods of cleansing Unicode data before loading. See the Teradata Parallel Transporter (TPT) documentation for details on configuring how TPT uses access modules. For details about configuring access modules to allow pass through characters, see *Teradata® Tools and Utilities Access Module Reference*, B035-2425, in particular the sections regarding automatic character conversions.
- PTCs have default character properties; therefore, case sensitivity, object name support, collation, and inequality comparisons are not supported.
- Unicode delimited character literals support PTCs.

For example: `SELECT feed FROM twitter WHERE feed LIKE U&'%#+01F602%' UESCAPE '#';`

- PTCs in SMP require two 16-bit UTF-16 code units, or 4 bytes. Therefore, VARCHAR(2) or CHAR(2) are the minimum sizes required to store a PTC.
- CHARACTER(4) is the minimum size required in order to import a PTC using a USING clause.
- PTCs have no impact on compression. The TransUnicodeToUTF8 compression function will compress surrogate pairs to a single 4 byte UTF8 encoded value. TransUTF8ToUnicode will decompress the value to surrogate pairs.
- The SDF language used by the tdlocaledef utility to define the system level localization will only support PTCs in the BMP.

Truncation

In sessions where Unicode Pass Through is enabled, truncation of a PTC occurs when the destination buffer does not have 4 bytes to contain the entire surrogate pair. If only 2 bytes remain in a fixed CHAR string, then the string is padded with a SPACE (U+0020) in Teradata mode. In ANSI mode, the truncation mode error is returned.

Format Strings

The length of an 'X' format string is in UTF-16 code units. So FORMAT 'X(2)' is the minimum size format string required in order to display a PTC.

Export Width

The representation of supplementary code points is four bytes long in the UTF8 and UTF16 character sets. This might suggest that the export width for the session must be four bytes per character to correctly process these characters. However, the representation internally is UTF-16 as two code units. Export width will see this as two 'characters', and so unless a site-defined export width is specified, export width will reserve four bytes for two characters exported to UTF16 and six bytes when exported to UTF8. Therefore there is no export width issue for supplementary code points unless a site-tailored export width is employed. For sessions with tailored export width that might cause truncation of strings with supplementary code

points, truncation was already possible unless the repertoire is known. If the repertoire of characters is being managed, pass through sessions are not an issue.

Collation

For Unicode data, PTCs sort at the position of the error character but in Unicode code point order. Characters off the BMP, such as those represented as surrogate pairs, would collate based on their UTF-16 code point order. This is true even for `CHARSET_COLL` collation for UTF8 sessions.

Unicode Pass Through does not provide full support for PTCs. With respect to collation, pass through surrogate pairs sort as if they are composed of two characters. In the current system, for all ASCII, EBCDIC, JIS_COLL, `CHARSET_COLL`, and `MULTINATIONAL` collations, code points in the range 0xD800 to 0xDFFF sort before U+FFFF. For example, under ASCII case-specific collation, a pass through surrogate pair 0xD800 0xDC00 is treated as two Unicode UTF-16 characters, U+D800 and U+DC00 and will be sorted before U+FFFF.

Comparisons

- Equality and the LIKE operators work correctly with PTCs.
- Inequality predicates may not work correctly with PTCs.

String Functions

PTCs are not affected by most string functions, but the following are exceptions:

- Indexing and counting functions operate on UTF-16 code units.
- Substring can split surrogate pairs resulting in U+FFFD.
- `UNICODE_TO_UNICODE_NF(K)C` and `UNICODE_TO_UNICODE_NF(K)D` TRANSLATE functions support the normalization of PTCs producing a normalized result string.
- The `TRANSLATE_CHK` function used with Unicode to Unicode translations will not stop processing when a PTC is encountered and will only stop when a U+FFFD is encountered.
- CASE translations may not work correctly.

Stored Procedures

In a session where UPT is disabled, you cannot create a stored procedure that references a PTC inside the stored procedure body. Such a procedure can only be created within a Pass Through session.

When calling a stored procedure that is defined with a PTC parameter or references a PTC in the stored procedure body, the following applies:

- In a session where UPT is disabled, you cannot call a stored procedure that is defined with a PTC parameter.
- In a Pass Through session, you can successfully call a stored procedure that is defined with a PTC parameter and pass a PTC literal to it as an argument.
- In all sessions (UPT enabled or disabled), you can call a stored procedure that references a PTC in the body of the the stored procedure.

The following table summarizes the usage for stored procedures with PTCs.

Stored Procedure Creation or Execution	Current Session Type	Expected Result
Creating a stored procedure that references a PTC in the procedure body	Pass Through session	Successful
	UPT is disabled	Error
Calling a stored procedure that is defined with a PTC parameter	Pass Through session	Successful
	UPT is disabled	Error
Calling a stored procedure that references a PTC in the procedure body	Pass Through session	Successful
	UPT is disabled	Successful
Calling a stored procedure that is defined with a PTC parameter and also references a PTC in the procedure body	Pass Through session	Successful
	UPT is disabled	Error

UPT Restrictions

Although UPT allows the storage and retrieval of PTCs, full support is not available for these characters in Teradata. For example, the following are not supported for PTCs:

- Collation
- Case sensitivity
 - Uppercase function, uppercase storage, and lowercase function may produce incorrect results.
- Object name support
 - PTCs cannot be used as part of an object name.
- Inequality comparisons may produce incorrect results.
- PTCs should not be stored in a column that is part of an index.

Note:

Because casing, collation, and object support are not included with UPT, there are some scenarios where the behavior will differ from what is expected with fully supported characters. For example, the following may produce unexpected results:

- Value ordered NUSIs or join index with ORDER BY clauses
 - RANGE PPI definitions
 - Case blind equality comparison and uniqueness constraints
 - Upper and lower case translation
 - Interaction with PIVOT (PTCs cannot be column names)
-

UPT does not resolve these conversion errors:

- NonUnicode client to Unicode (on error character)
- Extended site-defined client character set (for example, ARABIC1256_6A0) to the Latin server character set
- Unicode to Latin
- Latin to Unicode (error on 1A)
- TRANSLATE without WITH ERROR with nonUnicode
- KanjiSJIS to Unicode or Latin (error on 1A)

The following additional restrictions and limitations apply:

- UPT does not provide a per column granularity.
- Standalone FastExport, FastLoad, MultiLoad and T pump do not support PTCs. You can use Teradata Parallel Transporter to load and unload data containing PTCs.

Character Shorthand Notation

Overview

This document uses the Unicode naming convention for characters. For example, the lowercase character 'a' is more formally specified as either LATIN SMALL LETTER A or U+0041. The U+xxxx notation refers to a particular code point in the Unicode standard, where xxxx stands for the hexadecimal representation of the 16-bit value defined in the standard.

In parts of the document it is convenient to use a symbol to represent a special character or a particular class of characters. This is particularly true in discussion of the following Japanese character encodings.

- KanjiEBCDIC
- KanjiEUC
- KanjiShift-JIS

Character Symbols

The symbols, along with character sets with which they are used, are defined in the following table.

Symbol	Encoding	Meaning
a-z A-Z 0-9	Any	Any single byte Latin letter or digit.
<u>a-z</u> <u>A-Z</u> <u>0-9</u>	Any	Any fullwidth Latin letter or digit.
<	KanjiEBCDIC	Shift Out [SO] (0x0E). Indicates transition from single to multibyte character in KanjiEBCDIC.
>	KanjiEBCDIC	Shift In [SI] (0x0F). Indicates transition from multibyte to single byte KanjiEBCDIC.
T	Any	Any multibyte character. The encoding depends on the current character set. For KanjiEUC, code set 3 characters are always preceded by "ss 3".
!	Any	Any single byte Hankaku Katakana character. In KanjiEUC, it must be preceded by "ss 2", forming an individual multibyte character.
<u>Δ</u>	Any	Represents the graphic pad character.

Symbol	Encoding	Meaning
Δ	Any	Represents a single or multibyte pad character, depending on context.
ss ₂	KanjiEUC	Represents the EUC code set 2 introducer (0x8E).
ss ₃	KanjiEUC	Represents the EUC code set 3 introducer (0x8F).

For example, string “TEST”, where each letter is intended to be a fullwidth character, is written as **TEST**. Occasionally, when encoding is important, hexadecimal representation is used.

For example, the following mixed single byte/multibyte character data in KanjiEBCDIC character set

```
LMN<TEST>QRS
```

is represented as:

```
D3 D4 D5 0E 42E3 42C5 42E2 42E3 0F D8 D9 E2
```

Pad Characters

The following table lists the pad characters for the various server character sets.

Server Character Set	Pad Character Name	Pad Character Value
LATIN	SPACE	0x20
UNICODE	SPACE	U+0020
GRAPHIC	IDEOGRAPHIC SPACE	U+3000
KANJISJIS	ASCII SPACE	0x20
KANJI1	ASCII SPACE	0x20

Japanese Encodings and Mapping Standards

Overview

The text, tables, and figures in the following sections describe the internal and external Japanese character set ranges supported by the Teradata Database, with any combination of IBM, Microsoft Windows, and UNIX OS client systems.

JIS X 0201

The Teradata Database uses JIS X 0201, an 8-bit Japanese standard that defines encoding for 8-bit Latin and Katakana characters.

Related Teradata Database Client Character Sets

JIS X 0201 is used by the following Teradata Database client character sets:

- KANJIEUC_0U, as the encoding for EUC code set 0 and code set 2
- KANJISJIS_0S as Shift-JIS single-byte characters

Related Teradata Database Server Character Set

JIS X 0201 is the basis for the KANJI1 server character set.

JIS X 0201 Standard

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	`	p				ー	タ	ミ		
1			!	1	A	Q	a	q			。	ア	チ	ム		
2			"	2	B	R	b	r			「	イ	ツ	メ		
3			#	3	C	S	c	s			」	ウ	テ	モ		
4			\$	4	D	T	d	t			、	エ	ト	ヤ		
5			%	5	E	U	e	u			・	オ	ナ	ユ		
6			&	6	F	V	f	v			ヲ	カ	ニ	ヨ		
7			'	7	G	W	g	w			ヲ	キ	ヌ	ラ		
8			(8	H	X	h	x			イ	ク	ネ	リ		
9)	9	I	Y	i	y			ウ	ケ	ノ	ル		
A			*	:	J	Z	j	z			エ	コ	ハ	レ		
B			+	;	K	[k	{			オ	サ	ヒ	ロ		
C			,	<	L	¥	l				ヤ	シ	フ	ワ		
D			-	=	M]	m	}			ユ	ス	ヘ	ン		
E			.	>	N	^	n	~			ヨ	セ	ホ	ッ		
F			/	?	O	_	o				ツ	ソ	マ	°		

Related Information

For more information on ...	See ...
how the KANJIEUC_0U client character set uses the JIS X 0201 mapping standard	UNIX Compatible Japanese Character Set (KANJI EUC_0U).
how the KANJISJIS_0S client character set uses the JIS X 0201 mapping standard	Windows-Compatible Japanese Character Sets.
the KANJI1 server character set	KANJI1 Character Set.

JIS X 0208

JIS X 0208, the Japanese Industrial Standard Code for Japanese Graphic Set for Information Exchange, is the Japanese multibyte Kanji standard.

JIS X 0208 has three components.

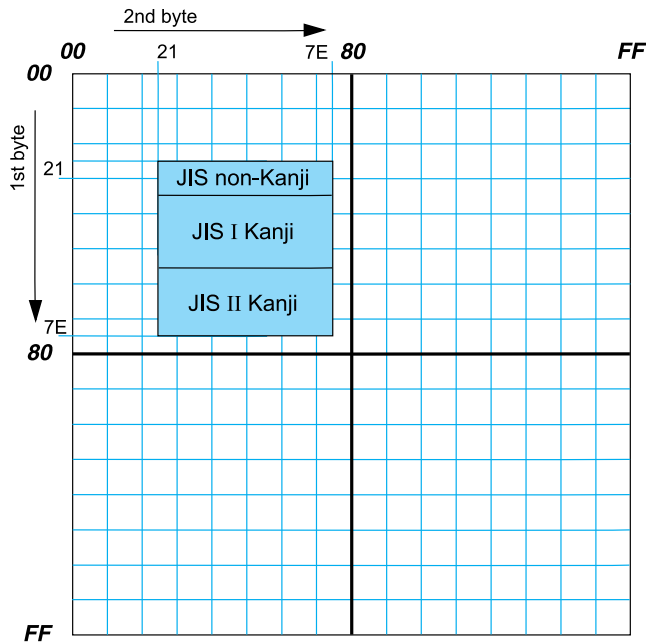
Component	Contents
JIS non-Kanji	Multibyte versions of Latin letters, Katakana, Hiragana, Cyrillic letters, Greek letters, and other special non-Kanji characters, like punctuation characters, and so on
JIS I Kanji	Commonly used Kanji characters
JIS II Kanji	Less commonly used Kanji characters

Related Teradata Database Client Character Sets

The following Teradata Database client character sets use the JIS X 0208 mapping standard:

- KANJIEUC_0U, as the encoding for EUC code set 1
- KANJISJIS_0S

The JIS X 0208 Standard



Related Information

For more information on ...	See ...
how the KANJIEUC_0U client character set uses the JIS X 0208 mapping standard	UNIX Compatible Japanese Character Set (KANJIEUC_0U).
how the KANJISJIS_0S client character set uses the JIS X 0208 mapping standard	Windows-Compatible Japanese Character Sets.

IBM Mainframe Character Sets for Japan

The topics that follow provide tables illustrating the following character sets:

- EBCDIC Katakana
- EBCDIC Japanese Extended English

- EBCDIC Japanese Extended Katakana
- EBCDIC Kanji

EBCDIC KATAKANA Character Set

Related Teradata Database Client Character Set

The Teradata Database KATAKANAEBCDIC client character set is based on the EBCDIC Katakana character set.

EBCDIC KATAKANA Character Set

In the EBCDIC KATAKANA character set that follows, non-blank spaces represent single-byte Latin and Katakana characters that can appear in object names under the KATAKANAEBCDIC client character set.

The blank spaces represent single-byte characters that should be used with caution as they are either undefined or control characters.

The 0x0E control character indicates the transition from single-byte to multibyte characters and the 0x0F control character indicates the transition from multibyte characters to single-byte characters.

	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0					SP	&	-		ソ						\$	0
-1					。	エ	/		ア	タ	-		A	J		1
-2					「	オ			イ	チ	ヘ		B	K	S	2
-3					」	ヤ			ウ	ツ	ホ		C	L	T	3
-4					、	ユ			エ	テ	マ		D	M	U	4
-5					・	ヨ			オ	ト	ミ		E	N	V	5
-6					ヲ	ヅ			カ	ナ	ム		F	O	W	6
-7					ヲ				キ	ニ	メ		G	P	X	7
-8					イ	ー			ク	ヌ	モ		H	Q	Y	8
-9					ウ				ケ	ネ	ヤ		I	R	Z	9
-A					£	!		:	コ	ノ	ユ	レ				
-B					・	¥	,	#				ロ				
-C					<	*	%	@	サ		ヨ	ワ				
-D					()	-	'	シ	ハ	ラ	ン				
-E					+	;	>	=	ス	ヒ	リ	・				
-F						〒	?	"	セ	フ	ル	。				

Related Information

For details on the related KATAKANAEBCDIC client character set, see [KATAKANAEBCDIC](#).

EBCDIC Japanese Extended Katakana Character Set

Related Teradata Database Client Character Set

The KANJIEBCDIC5026_01 client character set is based on the EBCDIC Japanese Extended Katakana character set.

EBCDIC Japanese Extended Katakana Character Set

In the Japanese Extended Katakana character set (IBM single-byte character CCSID=290) that follows, the non-blank spaces represent single-byte Latin and Katakana characters that can appear in object names under the KANJIEBCDIC5026_01 client character set.

The blank spaces represent single-byte characters that should be used with caution as they are either undefined or control characters.

The 0x0E control character indicates the transition from single-byte to multibyte characters and the 0x0F control character indicates the transition from multibyte characters to single-byte characters.

	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0					SP	&	—	[]	ソ	~	^	{	}	\$	0
-1					o	エ	/	i	ア	タ	—	¢	A	J		1
-2					「	オ	a	j	イ	チ	へ	\	B	K	S	2
-3					」	ヤ	b	k	ウ	ツ	ホ	t	C	L	T	3
-4					、	ユ	c	l	エ	テ	マ	u	D	M	U	4
-5					・	ヨ	d	m	オ	ト	ミ	v	E	N	V	5
-6					ヲ	ツ	e	n	カ	ナ	ム	w	F	O	W	6
-7					ア		f	o	キ	ニ	メ	x	G	P	X	7
-8					イ	ー	g	p	ク	ヌ	モ	y	H	Q	Y	8
-9					ウ		h	ゝ	ケ	ネ	ヤ	z	I	R	Z	9
-A					£	!		:	コ	ノ	ユ	レ				
-B					・	¥	,	#	q	r	s	ロ				
-C					<	*	%	@	サ		ヨ	ワ				
-D					()	_	'	シ	ハ	ラ	ン				
-E					+	;	>	=	ス	ヒ	リ	ゝ				
-F						ゝ	?	"	セ	フ	ル	。				

Related Information

For details on the related KANJIEBCDIC5026_01 client character set, see [KANJIEBCDIC5026_01](#).

EBCDIC Japanese Extended English Character Set

Related Teradata Database Client Character Set

The KANJIEBCDIC5035_01 client character set is based on the EBCDIC Japanese Extended English character set.

EBCDIC Japanese Extended English Character Set

In the Japanese Extended English character set (IBM single-byte character CCSID=1027) that follows, non-blank spaces represent single-byte Latin and Katakana characters that can appear in object names under the KANJI EBCDIC5035_01 client character set.

The blank spaces represent single-byte characters that should be used with caution as they are either undefined or control characters.

The 0x0E control character indicates the transition from single-byte to multibyte characters and the 0x0F control character indicates the transition from multibyte characters to single-byte characters.

	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
-0					SP	&	ー	コ			ー	^	{	}	\	0
-1						ウ	/	サ	a	j	~	£	A	J		1
-2					o	エ	イ	シ	b	k	s	¥	B	K	S	2
-3					「	オ	ウ	ス	c	l	t	ヤ	C	L	T	3
-4					」	ヤ	エ	セ	d	m	u	ユ	D	M	U	4
-5					、	ユ	オ	ソ	e	n	v	ヨ	E	N	V	5
-6					・	ヨ	カ	タ	f	o	w	ラ	F	O	W	6
-7					ヲ	ツ	キ	チ	g	p	x	リ	G	P	X	7
-8					ア	一	ク	ツ	h	q	y	ル	H	Q	Y	8
-9					イ	ア	ケ	、	i	r	z	レ	I	R	Z	9
-A					¢	!		:	テ	ノ	マ	ロ				
-B					・	\$,	#	ト	ハ	ミ	ワ				
-C					<	*	%	@	ナ	ヒ	ム	ン				
-D					()	_	'	ニ	フ	[]				
-E					+	;	>	=	ヌ	ヘ	メ	、				
-F						¬	?	"	ネ	ホ	モ	。				

Related Information

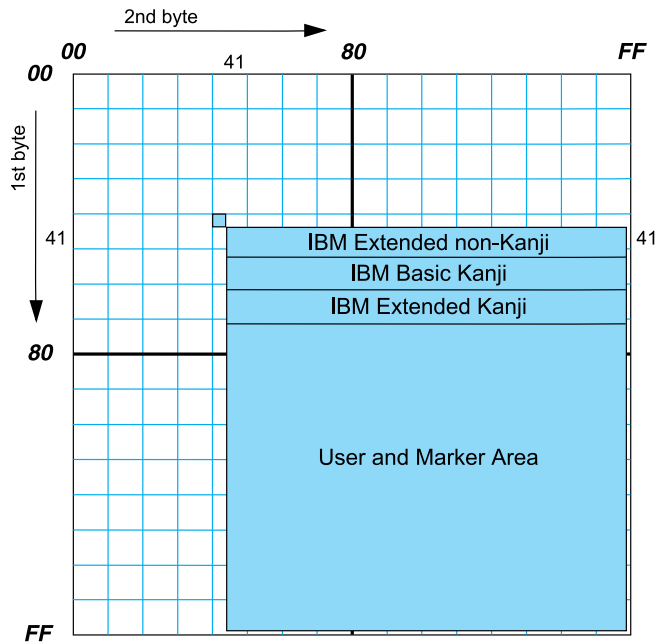
For details on the related KANJI EBCDIC5035_01 client character set, see [KANJI EBCDIC5035_01](#).

EBCDIC Kanji Character Set

Related Teradata Database Client Character Sets

The EBCDIC Kanji character set is the basis for double-byte characters in the KATAKANA EBCDIC, KANJI EBCDIC5026_01, and KANJI EBCDIC5035_01 client character sets, and any other character set with a name ending in _01.

EBCDIC Kanji Character Distribution



The area in the shaded portion is distributed as follows.

1st Byte	2nd Byte	Description
40	40	Kanji blank
41	41-FE	IBM Extended non-Kanji set
42		Alphanumeric
43		Basic non-Kanji set Zen-Katakana
44		Basic non-Kanji set Hiragana
45-55		IBM Basic Kanji set (arranged based on frequency of use)
56-68		IBM Extended Kanji set (arranged according to Kanji ideographical order)
69-7F	41-FE	User Area (private use area stops at 72EA)
80-FE		Marker reserved area

With the exception of those characters identified in [Object Names](#), double-byte characters in the following shaded portions of the preceding diagram are valid for object names under the KanjiEBCDIC client character sets:

- IBM Extended non-Kanji
- IBM Basic Kanji
- IBM Extended Kanji

Double-byte characters in the User and Marker area are not valid for object names.

Selected Characters for EBCDIC Kanji

The following table identifies selected characters in EBCDIC Kanji.

Double-Byte Space	Double-Byte Underscore	Double-Byte Percent
0x4040	0x426D	0x426C

The following control characters are the bytes that introduce or terminate interpretation of bytes as EBCDIC Kanji.

Shift Out (SO)	Shift In (SI)
0x0E	0x0F

Related Information

For more information on the related ...	See ...
KATAKANAEBCDIC client character set	KATAKANAEBCDIC.
KANJIEBCDIC5026_0I client character set	KANJIEBCDIC5026_0I.
KANJIEBCDIC5035_0I client character set	KANJIEBCDIC5035_0I.

Extended UNIX Code (EUC)

For UNIX client systems, the Teradata Database supports the Extended UNIX Code (EUC).

Related Teradata Database Client Character Set

The KANJIEUC_0U client character set is based on EUC.

Definitions

EUC is composed of one primary and three supplementary codesets.

The primary codeset, codeset 0, is used for ASCII characters.

The three supplementary code sets, code sets 1, 2, and 3, can be assigned to different character sets by the user.

There is a system default assignment for these codesets.

The primary code set is defined to be a single-byte with the most significant (high-order) bit set to 0. The supplementary codesets can be multiple bytes, and the most significant bit of each is set to 1.

Code sets 2 and 3 have a preceding single-shift character, known as ss₂ and ss₃, respectively, where ss₂ is 0x8E and ss₃ is 0x8F. Differentiation between codesets is as follows.

IF the most significant bit is this value ...	THEN ...
0	the code set is one-byte ASCII.
1	the byte is checked for ss ₂ or ss ₃ to determine the code set. The length in bytes of characters from that code set is retrieved from an ANSI localization table governing character classification, and that number of bytes is read in.

KanjiEUC Code Set Localization

The following table shows the KanjiEUC Code Set Localization.

Code Set	EUC Representation (In Bits)	Japanese Language Environment Implementation
cs0	0xxxxxxx	JIS X 0201
cs1	1xxxxxxx 1xxxxxxx	JIS X 0208 (Kanji Characters). The first 1xxxxxxx must not be ss ₂ or ss ₃ . The valid range of the first byte is A1-FE and the valid range of the second byte is A1-FE. Those ranges are implied by the JIS X 0208 standard.
cs2	ss ₂ 1xxxxxxx	JIS X 0201 (half-size Katakana). The valid range of the second byte is A1-DF.
cs3	ss ₃ 1xxxxxxx 1xxxxxxx	JIS X 0212. The valid range of the first byte is A1-FE and the valid range of the second byte is A1-FE. These ranges are implied by the JIS X 0212 standard.

Selected Characters for KanjiEUC

The following table identifies selected characters in the KanjiEUC character set.

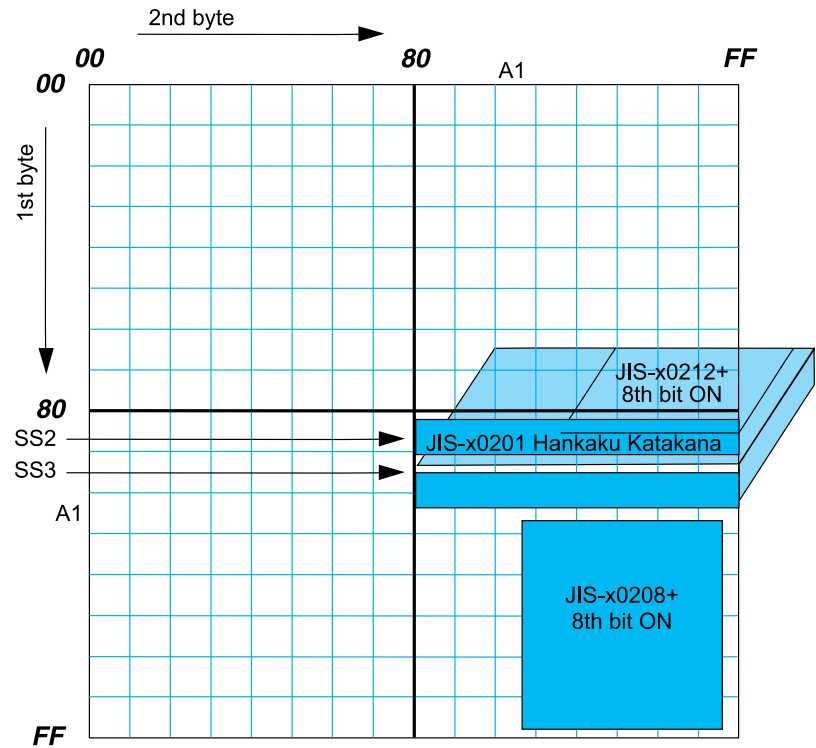
Double-Byte Space	Double-Byte Underscore	Double-Byte Percent
0xA1A1	0xA1B2	0xA1F3

The following table identifies the EUC code set 2 introducer (ss₂) and code set 3 introducer (ss₃).

SS 2	SS 3
0x8E	0x8F

KanjiEUC Encoding for Kanji Alphabets

The following graphic illustrates the KanjiEUC encoding for Kanji alphabets.



Related Information

For more information on ...	See ...
the JIS X 0201 standard	JIS X 0201 .
the JIS X 0208 standard	JIS X 0208 .
the KANJIEUC_OU client character set that is based on EUC	UNIX Compatible Japanese Character Set (KANJIEUC_OU) .

Shift-JIS (DOS Kanji) Encoding

For Windows client systems, the Teradata Database supports Shift-JIS encoding.

Related Teradata Database Client Character Set

The KANJISJIS_0S client character set emulates the Shift-JIS style of mixed single- and multibyte character data, where the range of the first byte in a character determines if the character is represented as one byte or two bytes.

Notes

DOS/V is an implementation of a Japanese character set that uses the undefined columns of JIS X 0201; those bytes are the first bytes for 2-byte Kanji characters. This encoding is referred to as the Shift-JIS encoding.

The following tables show the Shift-JIS encoding according to character values and selected Shift-JIS characters. The figures in [Shift-JIS Encoding for Kanji](#) and [Shift-JIS Encoding: Detailed View](#) illustrate the encoding ranges.

Shift-JIS Encoding

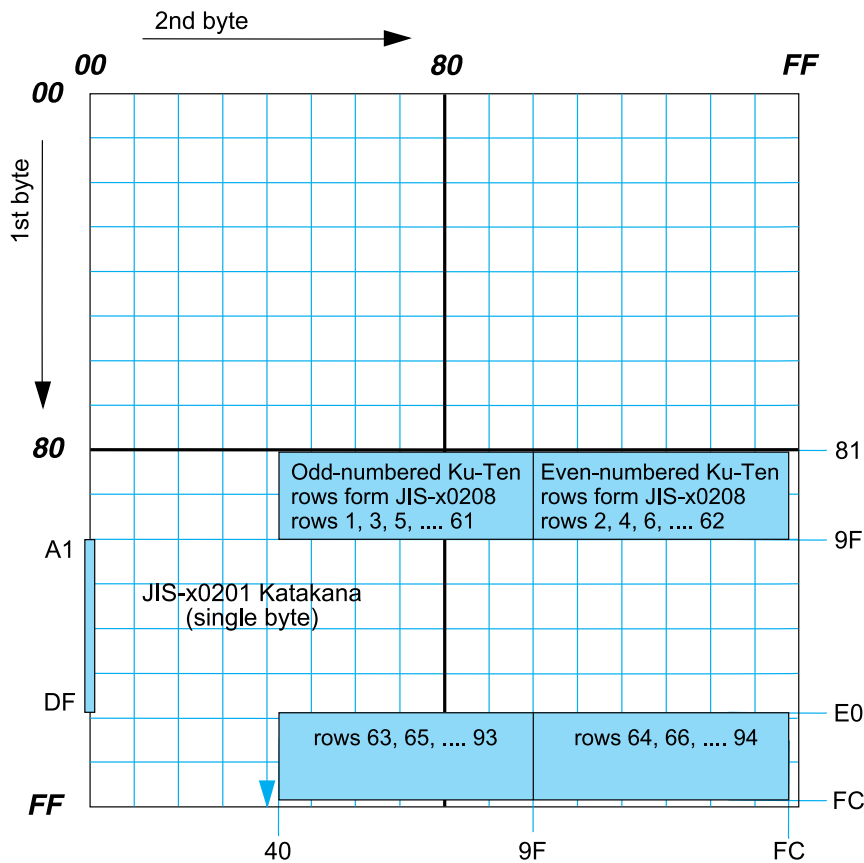
Hex Representation of Shift-JIS	Shift-JIS Implementation
0x00-0x7E, 0xA1-0xDF	JIS X 0201
0x81-0x9F, 0xE0-0xFC	First byte of double-byte representation. Its mapping is as follows: <ol style="list-style-type: none"> 1. 0x81-0x9F--Contains rows 1 to 62 from JIS X 0208. 2. 0xE0-0xEF--Contains rows 63 to 94 from JIS X 0208. 3. 0xF0-0xF9--Contains 1,880 Gaiji characters. 4. 0xFA-0xFC--Contains IBM-defined characters.
0x40-0x7E, 0x80-0xFC	Second byte of double-byte representation.

Selected Characters for Shift-JIS Kanji

Double-byte Space	Double-byte Underscore	Double-byte Percent
0x8140	0x8151	0x8193

Shift-JIS Encoding for Kanji

Note that although this graphic shows data at second byte 0x7F, there is none, as documented in the Shift-JIS encoding table (see [Shift-JIS Encoding](#)).

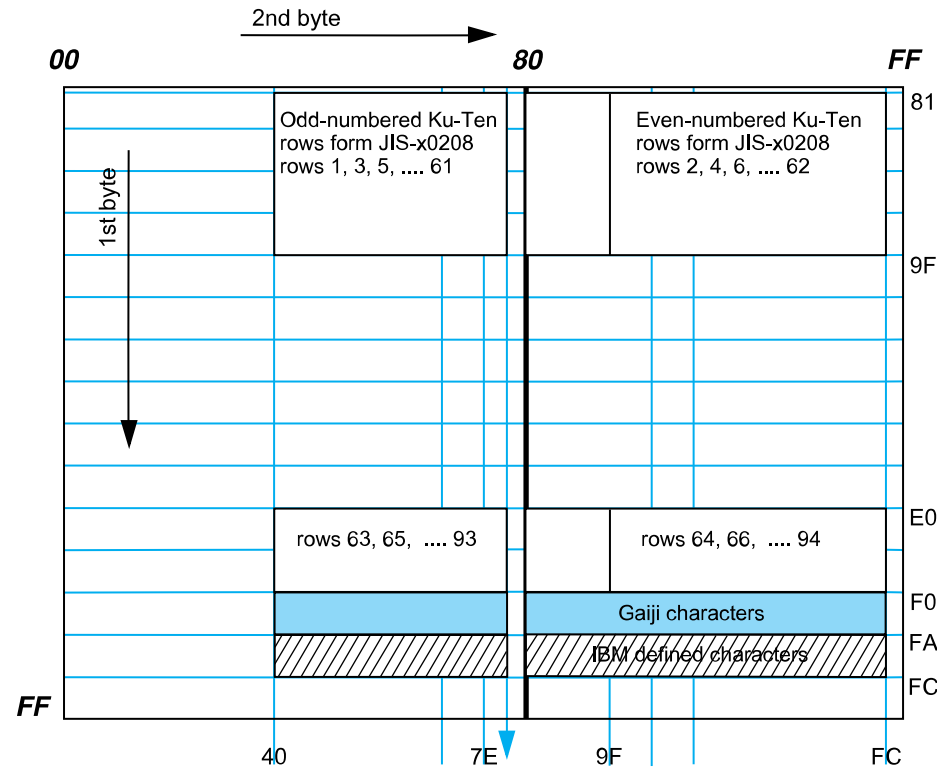


The next figure ([Shift-JIS Encoding: Detailed View](#)) shows a more detailed view of the Shift-JIS encoding. The shaded regions show the JIS X 0208 area, the Gaiji area, and the IBM-defined character area. Observe that only the range 0x8000-0xFFFF is shown.

This figure illustrates the following:

- Shift-JIS range 0x8140-0x9FFC contains rows 1 to 62 from JIS X 0208.
- Shift-JIS range 0xE040-0xEFFC contains rows 63 to 94 from JIS X 0208.
- Shift-JIS range 0xF040-0xF9FC contains 1,880 Gaiji characters. This range is equivalent to EUC ss3 range 0xA1A1-0xB4FE.
- Shift-JIS range 0xFA40-0xFCFC contains IBM-defined characters. IBM characters exist in IBM EBCDIC Kanji, but not in JIS X 0208.

Shift-JIS Encoding: Detailed View



Related Information

For more information on ...	See ...
the JIS X 0201 standard	JIS X 0201 .
the JIS X 0208 standard	JIS X 0208 .
the KANJISJIS_0S client character set that emulates Shift-JIS	Windows-Compatible Japanese Character Sets .

ASCII Mappings

ASCII Mappings

The ASCII mappings are as follows:

- [Client ASCII to Server LATIN](#)
- [Client ASCII to Server UNICODE](#)

For more ASCII mappings, such as ASCII to KANJI1, see the mapascii file in the etc directory of the Teradata software distribution.

Client ASCII to Server LATIN

The following table shows how Teradata Database maps characters from the ASCII client character set to characters in the LATIN server character set. The translation works in both directions, from ASCII to LATIN and from LATIN to ASCII.

ASCII	LATIN	Unicode Character Name
0x00	0x00	<control> NULL
0x01	0x01	<control> START OF HEADING
0x02	0x02	<control> START OF TEXT
0x03	0x03	<control> END OF TEXT
0x04	0x04	<control> END OF TRANSMISSION
0x05	0x05	<control> ENQUIRY
0x06	0x06	<control> ACKNOWLEDGE
0x07	0x07	<control> BELL
0x08	0x08	<control> BACKSPACE
0x09	0x09	<control> CHARACTER TABULATION (horizontal tabulation)
0x0A	0x0A	<control> LINE FEED
0x0B	0x0B	<control> LINE TABULATION (vertical tabulation)
0x0C	0x0C	<control> FORM FEED
0x0D	0x0D	<control> CARRIAGE RETURN
0x0E	0x0E	<control> SHIFT OUT

ASCII	LATIN	Unicode Character Name
0x0F	0x0F	<control> SHIFT IN
0x10	0x10	<control> DATA LINK ESCAPE
0x11	0x11	<control> DEVICE CONTROL ONE
0x12	0x12	<control> DEVICE CONTROL TWO
0x13	0x13	<control> DEVICE CONTROL THREE
0x14	0x14	<control> DEVICE CONTROL FOUR
0x15	0x15	<control> NEGATIVE ACKNOWLEDGE
0x16	0x16	<control> SYNCHRONOUS IDLE
0x17	0x17	<control> END OF TRANSMISSION BLOCK
0x18	0x18	<control> CANCEL
0x19	0x19	<control> END OF MEDIUM
0x1A	0x1A	<control> SUBSTITUTE
0x1B	0x1B	<control> ESCAPE
0x1C	0x1C	<control> INFORMATION SEPARATOR FOUR (file separator)
0x1D	0x1D	<control> INFORMATION SEPARATOR THREE (group separator)
0x1E	0x1E	<control> INFORMATION SEPARATOR TWO (record separator)
0x1F	0x1F	<control> INFORMATION SEPARATOR ONE (unit separator)
0x20	0x20	SPACE
0x21	0x21	EXCLAMATION MARK
0x22	0x22	QUOTATION MARK
0x23	0x23	NUMBER SIGN
0x24	0x24	DOLLAR SIGN
0x25	0x25	PERCENT SIGN
0x26	0x26	AMPERSAND
0x27	0x27	APOSTROPHE
0x28	0x28	LEFT PARENTHESIS
0x29	0x29	RIGHT PARENTHESIS
0x2A	0x2A	ASTERISK
0x2B	0x2B	PLUS SIGN

ASCII	LATIN	Unicode Character Name
0x2C	0x2C	COMMA
0x2D	0x2D	HYPHEN-MINUS
0x2E	0x2E	FULL STOP
0x2F	0x2F	SOLIDUS
0x30	0x30	DIGIT ZERO
0x31	0x31	DIGIT ONE
0x32	0x32	DIGIT TWO
0x33	0x33	DIGIT THREE
0x34	0x34	DIGIT FOUR
0x35	0x35	DIGIT FIVE
0x36	0x36	DIGIT SIX
0x37	0x37	DIGIT SEVEN
0x38	0x38	DIGIT EIGHT
0x39	0x39	DIGIT NINE
0x3A	0x3A	COLON
0x3B	0x3B	SEMICOLON
0x3C	0x3C	LESS-THAN SIGN
0x3D	0x3D	EQUALS SIGN
0x3E	0x3E	GREATER-THAN SIGN
0x3F	0x3F	QUESTION MARK
0x40	0x40	COMMERCIAL AT
0x41	0x41	LATIN CAPITAL LETTER A
0x42	0x42	LATIN CAPITAL LETTER B
0x43	0x43	LATIN CAPITAL LETTER C
0x44	0x44	LATIN CAPITAL LETTER D
0x45	0x45	LATIN CAPITAL LETTER E
0x46	0x46	LATIN CAPITAL LETTER F
0x47	0x47	LATIN CAPITAL LETTER G
0x48	0x48	LATIN CAPITAL LETTER H

ASCII	LATIN	Unicode Character Name
0x49	0x49	LATIN CAPITAL LETTER I
0x4A	0x4A	LATIN CAPITAL LETTER J
0x4B	0x4B	LATIN CAPITAL LETTER K
0x4C	0x4C	LATIN CAPITAL LETTER L
0x4D	0x4D	LATIN CAPITAL LETTER M
0x4E	0x4E	LATIN CAPITAL LETTER N
0x4F	0x4F	LATIN CAPITAL LETTER O
0x50	0x50	LATIN CAPITAL LETTER P
0x51	0x51	LATIN CAPITAL LETTER Q
0x52	0x52	LATIN CAPITAL LETTER R
0x53	0x53	LATIN CAPITAL LETTER S
0x54	0x54	LATIN CAPITAL LETTER T
0x55	0x55	LATIN CAPITAL LETTER U
0x56	0x56	LATIN CAPITAL LETTER V
0x57	0x57	LATIN CAPITAL LETTER W
0x58	0x58	LATIN CAPITAL LETTER X
0x59	0x59	LATIN CAPITAL LETTER Y
0x5A	0x5A	LATIN CAPITAL LETTER Z
0x5B	0x5B	LEFT SQUARE BRACKET
0x5C	0x5C	REVERSE SOLIDUS
0x5D	0x5D	RIGHT SQUARE BRACKET
0x5E	0x5E	CIRCUMFLEX ACCENT
0x5F	0x5F	LOW LINE
0x60	0x60	GRAVE ACCENT
0x61	0x61	LATIN SMALL LETTER A
0x62	0x62	LATIN SMALL LETTER B
0x63	0x63	LATIN SMALL LETTER C
0x64	0x64	LATIN SMALL LETTER D
0x65	0x65	LATIN SMALL LETTER E

ASCII	LATIN	Unicode Character Name
0x66	0x66	LATIN SMALL LETTER F
0x67	0x67	LATIN SMALL LETTER G
0x68	0x68	LATIN SMALL LETTER H
0x69	0x69	LATIN SMALL LETTER I
0x6A	0x6A	LATIN SMALL LETTER J
0x6B	0x6B	LATIN SMALL LETTER K
0x6C	0x6C	LATIN SMALL LETTER L
0x6D	0x6D	LATIN SMALL LETTER M
0x6E	0x6E	LATIN SMALL LETTER N
0x6F	0x6F	LATIN SMALL LETTER O
0x70	0x70	LATIN SMALL LETTER P
0x71	0x71	LATIN SMALL LETTER Q
0x72	0x72	LATIN SMALL LETTER R
0x73	0x73	LATIN SMALL LETTER S
0x74	0x74	LATIN SMALL LETTER T
0x75	0x75	LATIN SMALL LETTER U
0x76	0x76	LATIN SMALL LETTER V
0x77	0x77	LATIN SMALL LETTER W
0x78	0x78	LATIN SMALL LETTER X
0x79	0x79	LATIN SMALL LETTER Y
0x7A	0x7A	LATIN SMALL LETTER Z
0x7B	0x7B	LEFT CURLY BRACKET
0x7C	0x7C	VERTICAL LINE
0x7D	0x7D	RIGHT CURLY BRACKET
0x7E	0x7E	TILDE
0x7F	0x7F	<control> DELETE
0x80	0x80	NO-BREAK SPACE
0x81	0x81	DIAERESIS
0x82	0x82	ACUTE ACCENT

ASCII	LATIN	Unicode Character Name
0x83	0x83	CEDILLA
0x84	0x84	MULTIPLICATION SIGN
0x85	0x85	DIVISION SIGN
0x86	0x86	<control> START OF SELECTED AREA
0x87	0x87	<control> END OF SELECTED AREA
0x88	0x88	<control> CHARACTER TABULATION SET
0x89	0x89	<control> CHARACTER TABULATION WITH JUSTIFICATION
0x8A	0x8A	<control> LINE TABULATION SET
0x8B	0x8B	<control> PARTIAL LINE FORWARD
0x8C	0x8C	<control> PARTIAL LINE BACKWARD
0x8D	0x8D	<control> REVERSE LINE FEED
0x8E	0x8E	<control> SINGLE SHIFT TWO
0x8F	0x8F	<control> SINGLE SHIFT THREE
0x90	0x90	<control> DEVICE CONTROL STRING
0x91	0x91	<control> PRIVATE USE ONE
0x92	0x92	<control> PRIVATE USE TWO
0x93	0x93	<control> SET TRANSMIT STATE
0x94	0x94	<control> CANCEL CHARACTER
0x95	0x95	<control> MESSAGE WAITING
0x96	0x96	<control> START OF GUARDED AREA
0x97	0x97	<control> END OF GUARDED AREA
0x98	0x98	<control> START OF STRING
0x99	0x99	<control>
0x9A	0x9A	<control> SINGLE CHARACTER INTRODUCER
0x9B	0x9B	<control> CONTROL SEQUENCE INTRODUCER
0x9C	0x9C	<control> STRING TERMINATOR
0x9D	0x9D	<control> OPERATING SYSTEM COMMAND
0x9E	0x9E	<control> PRIVACY MESSAGE
0x9F	0x9F	<control> APPLICATION PROGRAM COMMAND

ASCII	LATIN	Unicode Character Name
0xA0	0xA0	LATIN SMALL LETTER S WITH CARON
0xA1	0xA1	INVERTED EXCLAMATION MARK
0xA2	0xA2	CENT SIGN
0xA3	0xA3	POUND SIGN
0xA4	0xA4	LATIN SMALL LETTER Y WITH ACUTE
0xA5	0xA5	YEN SIGN
0xA6	0xA6	BROKEN BAR
0xA7	0xA7	SECTION SIGN
0xA8	0xA8	CURRENCY SIGN
0xA9	0xA9	COPYRIGHT SIGN
0xAA	0xAA	FEMININE ORDINAL INDICATOR
0xAB	0xAB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
0xAC	0xAC	NOT SIGN
0xAD	0xAD	SOFT HYPHEN
0xAE	0xAE	REGISTERED SIGN
0xAF	0xAF	MACRON
0xB0	0xB0	DEGREE SIGN
0xB1	0xB1	PLUS-MINUS SIGN
0xB2	0xB2	SUPERSCRIFT TWO
0xB3	0xB3	SUPERSCRIFT THREE
0xB4	0xB4	LATIN CAPITAL LETTER Y WITH ACUTE
0xB5	0xB5	MICRO SIGN
0xB6	0xB6	PILCROW SIGN
0xB7	0xB7	MIDDLE DOT
0xB8	0xB8	LATIN CAPITAL LETTER S WITH CARON
0xB9	0xB9	SUPERSCRIFT ONE
0xBA	0xBA	MASCULINE ORDINAL INDICATOR
0xBB	0xBB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
0xBC	0xBC	VULGAR FRACTION ONE QUARTER

ASCII	LATIN	Unicode Character Name
0xBD	0xBD	VULGAR FRACTION ONE HALF
0xBE	0xBE	VULGAR FRACTION THREE QUARTERS
0xBF	0xBF	INVERTED QUESTION MARK
0xC0	0xC0	LATIN CAPITAL LETTER A WITH GRAVE
0xC1	0xC1	LATIN CAPITAL LETTER A WITH ACUTE
0xC2	0xC2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
0xC3	0xC3	LATIN CAPITAL LETTER A WITH TILDE
0xC4	0xC4	LATIN CAPITAL LETTER A WITH DIAERESIS
0xC5	0xC5	LATIN CAPITAL LETTER A WITH RING ABOVE
0xC6	0xC6	LATIN CAPITAL LETTER AE
0xC7	0xC7	LATIN CAPITAL LETTER C WITH CEDILLA
0xC8	0xC8	LATIN CAPITAL LETTER E WITH GRAVE
0xC9	0xC9	LATIN CAPITAL LETTER E WITH ACUTE
0xCA	0xCA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
0xCB	0xCB	LATIN CAPITAL LETTER E WITH DIAERESIS
0xCC	0xCC	LATIN CAPITAL LETTER I WITH GRAVE
0xCD	0xCD	LATIN CAPITAL LETTER I WITH ACUTE
0xCE	0xCE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
0xCF	0xCF	LATIN CAPITAL LETTER I WITH DIAERESIS
0xD0	0xD0	LATIN SMALL LETTER ETH
0xD1	0xD1	LATIN CAPITAL LETTER N WITH TILDE
0xD2	0xD2	LATIN CAPITAL LETTER O WITH GRAVE
0xD3	0xD3	LATIN CAPITAL LETTER O WITH ACUTE
0xD4	0xD4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
0xD5	0xD5	LATIN CAPITAL LETTER O WITH TILDE
0xD6	0xD6	LATIN CAPITAL LETTER O WITH DIAERESIS
0xD7	0xD7	LATIN CAPITAL LIGATURE OE
0xD8	0xD8	LATIN CAPITAL LETTER O WITH STROKE
0xD9	0xD9	LATIN CAPITAL LETTER U WITH GRAVE

ASCII	LATIN	Unicode Character Name
0xDA	0xDA	LATIN CAPITAL LETTER U WITH ACUTE
0xDB	0xDB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
0xDC	0xDC	LATIN CAPITAL LETTER U WITH DIAERESIS
0xDD	0xDD	LATIN CAPITAL LETTER Y WITH DIAERESIS
0xDE	0xDE	LATIN SMALL LETTER THORN
0xDF	0xDF	LATIN SMALL LETTER SHARP S
0xE0	0xE0	LATIN SMALL LETTER A WITH GRAVE
0xE1	0xE1	LATIN SMALL LETTER A WITH ACUTE
0xE2	0xE2	LATIN SMALL LETTER A WITH CIRCUMFLEX
0xE3	0xE3	LATIN SMALL LETTER A WITH TILDE
0xE4	0xE4	LATIN SMALL LETTER A WITH DIAERESIS
0xE5	0xE5	LATIN SMALL LETTER A WITH RING ABOVE
0xE6	0xE6	LATIN SMALL LETTER AE
0xE7	0xE7	LATIN SMALL LETTER C WITH CEDILLA
0xE8	0xE8	LATIN SMALL LETTER E WITH GRAVE
0xE9	0xE9	LATIN SMALL LETTER E WITH ACUTE
0xEA	0xEA	LATIN SMALL LETTER E WITH CIRCUMFLEX
0xEB	0xEB	LATIN SMALL LETTER E WITH DIAERESIS
0xEC	0xEC	LATIN SMALL LETTER I WITH GRAVE
0xED	0xED	LATIN SMALL LETTER I WITH ACUTE
0xEE	0xEE	LATIN SMALL LETTER I WITH CIRCUMFLEX
0xEF	0xEF	LATIN SMALL LETTER I WITH DIAERESIS
0xF0	0xF0	LATIN CAPITAL LETTER ETH
0xF1	0xF1	LATIN SMALL LETTER N WITH TILDE
0xF2	0xF2	LATIN SMALL LETTER O WITH GRAVE
0xF3	0xF3	LATIN SMALL LETTER O WITH ACUTE
0xF4	0xF4	LATIN SMALL LETTER O WITH CIRCUMFLEX
0xF5	0xF5	LATIN SMALL LETTER O WITH TILDE
0xF6	0xF6	LATIN SMALL LETTER O WITH DIAERESIS

ASCII	LATIN	Unicode Character Name
0xF7	0xF7	LATIN SMALL LIGATURE OE
0xF8	0xF8	LATIN SMALL LETTER O WITH STROKE
0xF9	0xF9	LATIN SMALL LETTER U WITH GRAVE
0xFA	0xFA	LATIN SMALL LETTER U WITH ACUTE
0xFB	0xFB	LATIN SMALL LETTER U WITH CIRCUMFLEX
0xFC	0xFC	LATIN SMALL LETTER U WITH DIAERESIS
0xFD	0xFD	LATIN SMALL LETTER Y WITH DIAERESIS
0xFE	0xFE	LATIN CAPITAL LETTER THORN
0xFF	0xFF	EURO SIGN

Client ASCII to Server UNICODE

The following table shows how Teradata Database maps Latin characters from the ASCII client character set to the UNICODE server character set. The translation works in both directions, from ASCII to UNICODE and from UNICODE to ASCII.

ASCII	UNICODE Latin	Unicode Character Name
0x00	U+0000	<control> NULL
0x01	U+0001	<control> START OF HEADING
0x02	U+0002	<control> START OF TEXT
0x03	U+0003	<control> END OF TEXT
0x04	U+0004	<control> END OF TRANSMISSION
0x05	U+0005	<control> ENQUIRY
0x06	U+0006	<control> ACKNOWLEDGE
0x07	U+0007	<control> BELL
0x08	U+0008	<control> BACKSPACE
0x09	U+0009	<control> CHARACTER TABULATION (horizontal tabulation)
0x0A	U+000A	<control> LINE FEED
0x0B	U+000B	<control> LINE TABULATION (vertical tabulation)
0x0C	U+000C	<control> FORM FEED
0x0D	U+000D	<control> CARRIAGE RETURN

ASCII	UNICODE Latin	Unicode Character Name
0x0E	U+000E	<control> SHIFT OUT
0x0F	U+000F	<control> SHIFT IN
0x10	U+0010	<control> DATA LINK ESCAPE
0x11	U+0011	<control> DEVICE CONTROL ONE
0x12	U+0012	<control> DEVICE CONTROL TWO
0x13	U+0013	<control> DEVICE CONTROL THREE
0x14	U+0014	<control> DEVICE CONTROL FOUR
0x15	U+0015	<control> NEGATIVE ACKNOWLEDGE
0x16	U+0016	<control> SYNCHRONOUS IDLE
0x17	U+0017	<control> END OF TRANSMISSION BLOCK
0x18	U+0018	<control> CANCEL
0x19	U+0019	<control> END OF MEDIUM
0x1A	U+001A	<control> SUBSTITUTE
0x1B	U+001B	<control> ESCAPE
0x1C	U+001C	<control> INFORMATION SEPARATOR FOUR (file separator)
0x1D	U+001D	<control> INFORMATION SEPARATOR THREE (group separator)
0x1E	U+001E	<control> INFORMATION SEPARATOR TWO (record separator)
0x1F	U+001F	<control> INFORMATION SEPARATOR ONE (unit separator)
0x20	U+0020	SPACE
0x21	U+0021	EXCLAMATION MARK
0x22	U+0022	QUOTATION MARK
0x23	U+0023	NUMBER SIGN
0x24	U+0024	DOLLAR SIGN
0x25	U+0025	PERCENT SIGN
0x26	U+0026	AMPERSAND
0x27	U+0027	APOSTROPHE
0x28	U+0028	LEFT PARENTHESIS
0x29	U+0029	RIGHT PARENTHESIS
0x2A	U+002A	ASTERISK

ASCII	UNICODE Latin	Unicode Character Name
0x2B	U+002B	PLUS SIGN
0x2C	U+002C	COMMA
0x2D	U+002D	HYPHEN-MINUS
0x2E	U+002E	FULL STOP
0x2F	U+002F	SOLIDUS
0x30	U+0030	DIGIT ZERO
0x31	U+0031	DIGIT ONE
0x32	U+0032	DIGIT TWO
0x33	U+0033	DIGIT THREE
0x34	U+0034	DIGIT FOUR
0x35	U+0035	DIGIT FIVE
0x36	U+0036	DIGIT SIX
0x37	U+0037	DIGIT SEVEN
0x38	U+0038	DIGIT EIGHT
0x39	U+0039	DIGIT NINE
0x3A	U+003A	COLON
0x3B	U+003B	SEMICOLON
0x3C	U+003C	LESS-THAN SIGN
0x3D	U+003D	EQUALS SIGN
0x3E	U+003E	GREATER-THAN SIGN
0x3F	U+003F	QUESTION MARK
0x40	U+0040	COMMERCIAL AT
0x41	U+0041	LATIN CAPITAL LETTER A
0x42	U+0042	LATIN CAPITAL LETTER B
0x43	U+0043	LATIN CAPITAL LETTER C
0x44	U+0044	LATIN CAPITAL LETTER D
0x45	U+0045	LATIN CAPITAL LETTER E
0x46	U+0046	LATIN CAPITAL LETTER F
0x47	U+0047	LATIN CAPITAL LETTER G

ASCII	UNICODE Latin	Unicode Character Name
0x48	U+0048	LATIN CAPITAL LETTER H
0x49	U+0049	LATIN CAPITAL LETTER I
0x4A	U+004A	LATIN CAPITAL LETTER J
0x4B	U+004B	LATIN CAPITAL LETTER K
0x4C	U+004C	LATIN CAPITAL LETTER L
0x4D	U+004D	LATIN CAPITAL LETTER M
0x4E	U+004E	LATIN CAPITAL LETTER N
0x4F	U+004F	LATIN CAPITAL LETTER O
0x50	U+0050	LATIN CAPITAL LETTER P
0x51	U+0051	LATIN CAPITAL LETTER Q
0x52	U+0052	LATIN CAPITAL LETTER R
0x53	U+0053	LATIN CAPITAL LETTER S
0x54	U+0054	LATIN CAPITAL LETTER T
0x55	U+0055	LATIN CAPITAL LETTER U
0x56	U+0056	LATIN CAPITAL LETTER V
0x57	U+0057	LATIN CAPITAL LETTER W
0x58	U+0058	LATIN CAPITAL LETTER X
0x59	U+0059	LATIN CAPITAL LETTER Y
0x5A	U+005A	LATIN CAPITAL LETTER Z
0x5B	U+005B	LEFT SQUARE BRACKET
0x5C	U+005C	REVERSE SOLIDUS
0x5D	U+005D	RIGHT SQUARE BRACKET
0x5E	U+005E	CIRCUMFLEX ACCENT
0x5F	U+005F	LOW LINE
0x60	U+0060	GRAVE ACCENT
0x61	U+0061	LATIN SMALL LETTER A
0x62	U+0062	LATIN SMALL LETTER B
0x63	U+0063	LATIN SMALL LETTER C
0x64	U+0064	LATIN SMALL LETTER D

ASCII	UNICODE Latin	Unicode Character Name
0x65	U+0065	LATIN SMALL LETTER E
0x66	U+0066	LATIN SMALL LETTER F
0x67	U+0067	LATIN SMALL LETTER G
0x68	U+0068	LATIN SMALL LETTER H
0x69	U+0069	LATIN SMALL LETTER I
0x6A	U+006A	LATIN SMALL LETTER J
0x6B	U+006B	LATIN SMALL LETTER K
0x6C	U+006C	LATIN SMALL LETTER L
0x6D	U+006D	LATIN SMALL LETTER M
0x6E	U+006E	LATIN SMALL LETTER N
0x6F	U+006F	LATIN SMALL LETTER O
0x70	U+0070	LATIN SMALL LETTER P
0x71	U+0071	LATIN SMALL LETTER Q
0x72	U+0072	LATIN SMALL LETTER R
0x73	U+0073	LATIN SMALL LETTER S
0x74	U+0074	LATIN SMALL LETTER T
0x75	U+0075	LATIN SMALL LETTER U
0x76	U+0076	LATIN SMALL LETTER V
0x77	U+0077	LATIN SMALL LETTER W
0x78	U+0078	LATIN SMALL LETTER X
0x79	U+0079	LATIN SMALL LETTER Y
0x7A	U+007A	LATIN SMALL LETTER Z
0x7B	U+007B	LEFT CURLY BRACKET
0x7C	U+007C	VERTICAL LINE
0x7D	U+007D	RIGHT CURLY BRACKET
0x7E	U+007E	TILDE
0x7F	U+007F	<control> DELETE
0x80	U+00A0	NO-BREAK SPACE
0x81	U+00A8	DIAERESIS

ASCII	UNICODE Latin	Unicode Character Name
0x82	U+00B4	ACUTE ACCENT
0x83	U+00B8	CEDILLA
0x84	U+00D7	MULTIPLICATION SIGN
0x85	U+00F7	DIVISION SIGN
0x86	U+0086	<control> START OF SELECTED AREA
0x87	U+0087	<control> END OF SELECTED AREA
0x88	U+0088	<control> CHARACTER TABULATION SET
0x89	U+0089	<control> CHARACTER TABULATION WITH JUSTIFICATION
0x8A	U+008A	<control> LINE TABULATION SET
0x8B	U+008B	<control> PARTIAL LINE FORWARD
0x8C	U+008C	<control> PARTIAL LINE BACKWARD
0x8D	U+008D	<control> REVERSE LINE FEED
0x8E	U+008E	<control> SINGLE SHIFT TWO
0x8F	U+008F	<control> SINGLE SHIFT THREE
0x90	U+0090	<control> DEVICE CONTROL STRING
0x91	U+0091	<control> PRIVATE USE ONE
0x92	U+0092	<control> PRIVATE USE TWO
0x93	U+0093	<control> SET TRANSMIT STATE
0x94	U+0094	<control> CANCEL CHARACTER
0x95	U+0095	<control> MESSAGE WAITING
0x96	U+0096	<control> START OF GUARDED AREA
0x97	U+0097	<control> END OF GUARDED AREA
0x98	U+0098	<control> START OF STRING
0x99	U+0099	<control>
0x9A	U+009A	<control> SINGLE CHARACTER INTRODUCER
0x9B	U+009B	<control> CONTROL SEQUENCE INTRODUCER
0x9C	U+009C	<control> STRING TERMINATOR
0x9D	U+009D	<control> OPERATING SYSTEM COMMAND
0x9E	U+009E	<control> PRIVACY MESSAGE

ASCII	UNICODE Latin	Unicode Character Name
0x9F	U+009F	<control> APPLICATION PROGRAM COMMAND
0xA0	U+0161	LATIN SMALL LETTER S WITH CARON
0xA1	U+00A1	INVERTED EXCLAMATION MARK
0xA2	U+00A2	CENT SIGN
0xA3	U+00A3	POUND SIGN
0xA4	U+00FD	LATIN SMALL LETTER Y WITH ACUTE
0xA5	U+00A5	YEN SIGN
0xA6	U+00A6	BROKEN BAR
0xA7	U+00A7	SECTION SIGN
0xA8	U+00A4	CURRENCY SIGN
0xA9	U+00A9	COPYRIGHT SIGN
0xAA	U+00AA	FEMININE ORDINAL INDICATOR
0xAB	U+00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
0xAC	U+00AC	NOT SIGN
0xAD	U+00AD	SOFT HYPHEN
0xAE	U+00AE	REGISTERED SIGN
0xAF	U+00AF	MACRON
0xB0	U+00B0	DEGREE SIGN
0xB1	U+00B1	PLUS-MINUS SIGN
0xB2	U+00B2	SUPERSCRPT TWO
0xB3	U+00B3	SUPERSCRPT THREE
0xB4	U+00DD	LATIN CAPITAL LETTER Y WITH ACUTE
0xB5	U+00B5	MICRO SIGN
0xB6	U+00B6	PILCROW SIGN
0xB7	U+00B7	MIDDLE DOT
0xB8	U+0160	LATIN CAPITAL LETTER S WITH CARON
0xB9	U+00B9	SUPERSCRPT ONE
0xBA	U+00BA	MASCULINE ORDINAL INDICATOR
0xBB	U+00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK

ASCII	UNICODE Latin	Unicode Character Name
0xBC	U+00BC	VULGAR FRACTION ONE QUARTER
0xBD	U+00BD	VULGAR FRACTION ONE HALF
0xBE	U+00BE	VULGAR FRACTION THREE QUARTERS
0xBF	U+00BF	INVERTED QUESTION MARK
0xC0	U+00C0	LATIN CAPITAL LETTER A WITH GRAVE
0xC1	U+00C1	LATIN CAPITAL LETTER A WITH ACUTE
0xC2	U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
0xC3	U+00C3	LATIN CAPITAL LETTER A WITH TILDE
0xC4	U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS
0xC5	U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE
0xC6	U+00C6	LATIN CAPITAL LETTER AE
0xC7	U+00C7	LATIN CAPITAL LETTER C WITH CEDILLA
0xC8	U+00C8	LATIN CAPITAL LETTER E WITH GRAVE
0xC9	U+00C9	LATIN CAPITAL LETTER E WITH ACUTE
0xCA	U+00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
0xCB	U+00CB	LATIN CAPITAL LETTER E WITH DIAERESIS
0xCC	U+00CC	LATIN CAPITAL LETTER I WITH GRAVE
0xCD	U+00CD	LATIN CAPITAL LETTER I WITH ACUTE
0xCE	U+00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
0xCF	U+00CF	LATIN CAPITAL LETTER I WITH DIAERESIS
0xD0	U+00F0	LATIN SMALL LETTER ETH
0xD1	U+00D1	LATIN CAPITAL LETTER N WITH TILDE
0xD2	U+00D2	LATIN CAPITAL LETTER O WITH GRAVE
0xD3	U+00D3	LATIN CAPITAL LETTER O WITH ACUTE
0xD4	U+00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
0xD5	U+00D5	LATIN CAPITAL LETTER O WITH TILDE
0xD6	U+00D6	LATIN CAPITAL LETTER O WITH DIAERESIS
0xD7	U+0152	LATIN CAPITAL LIGATURE OE
0xD8	U+00D8	LATIN CAPITAL LETTER O WITH STROKE

ASCII	UNICODE Latin	Unicode Character Name
0xD9	U+00D9	LATIN CAPITAL LETTER U WITH GRAVE
0xDA	U+00DA	LATIN CAPITAL LETTER U WITH ACUTE
0xDB	U+00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
0xDC	U+00DC	LATIN CAPITAL LETTER U WITH DIAERESIS
0xDD	U+0178	LATIN CAPITAL LETTER Y WITH DIAERESIS
0xDE	U+00FE	LATIN SMALL LETTER THORN
0xDF	U+00DF	LATIN SMALL LETTER SHARP S
0xE0	U+00E0	LATIN SMALL LETTER A WITH GRAVE
0xE1	U+00E1	LATIN SMALL LETTER A WITH ACUTE
0xE2	U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX
0xE3	U+00E3	LATIN SMALL LETTER A WITH TILDE
0xE4	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
0xE5	U+00E5	LATIN SMALL LETTER A WITH RING ABOVE
0xE6	U+00E6	LATIN SMALL LETTER AE
0xE7	U+00E7	LATIN SMALL LETTER C WITH CEDILLA
0xE8	U+00E8	LATIN SMALL LETTER E WITH GRAVE
0xE9	U+00E9	LATIN SMALL LETTER E WITH ACUTE
0xEA	U+00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX
0xEB	U+00EB	LATIN SMALL LETTER E WITH DIAERESIS
0xEC	U+00EC	LATIN SMALL LETTER I WITH GRAVE
0xED	U+00ED	LATIN SMALL LETTER I WITH ACUTE
0xEE	U+00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX
0xEF	U+00EF	LATIN SMALL LETTER I WITH DIAERESIS
0xF0	U+00D0	LATIN CAPITAL LETTER ETH
0xF1	U+00F1	LATIN SMALL LETTER N WITH TILDE
0xF2	U+00F2	LATIN SMALL LETTER O WITH GRAVE
0xF3	U+00F3	LATIN SMALL LETTER O WITH ACUTE
0xF4	U+00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX
0xF5	U+00F5	LATIN SMALL LETTER O WITH TILDE

ASCII	UNICODE Latin	Unicode Character Name
0xF6	U+00F6	LATIN SMALL LETTER O WITH DIAERESIS
0xF7	U+0153	LATIN SMALL LIGATURE OE
0xF8	U+00F8	LATIN SMALL LETTER O WITH STROKE
0xF9	U+00F9	LATIN SMALL LETTER U WITH GRAVE
0xFA	U+00FA	LATIN SMALL LETTER U WITH ACUTE
0xFB	U+00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX
0xFC	U+00FC	LATIN SMALL LETTER U WITH DIAERESIS
0xFD	U+00FF	LATIN SMALL LETTER Y WITH DIAERESIS
0xFE	U+00DE	LATIN CAPITAL LETTER THORN
0xFF	U+20AC	EURO SIGN

EBCDIC Mappings

EBCDIC Mappings

The EBCDIC mappings are as follows:

- [Client EBCDIC to Server LATIN](#)
- [Client EBCDIC to Server UNICODE](#)

For more EBCDIC mappings, such as EBCDIC to KANJI1, see the mapebcdic file in the etc directory of the Teradata software distribution.

Client EBCDIC to Server LATIN

The following table shows how characters from the EBCDIC client character set map to characters in the LATIN server character set. The translation works for both directions, from EBCDIC to LATIN and from LATIN to EBCDIC.

EBCDIC	LATIN	Unicode Character Name
0x00	0x00	<control> NULL
0x01	0x01	<control> START OF HEADING
0x02	0x02	<control> START OF TEXT
0x03	0x03	<control> END OF TEXT
0x04	0xA8	CURRENCY SIGN
0x05	0x09	<control> CHARACTER TABULATION (horizontal tabulation)
0x06	0xA9	COPYRIGHT SIGN
0x07	0x7F	<control> DELETE
0x08	0xD1	LATIN CAPITAL LETTER N WITH TILDE
0x09	0xD2	LATIN CAPITAL LETTER O WITH GRAVE
0x0A	0xD3	LATIN CAPITAL LETTER O WITH ACUTE
0x0B	0x0B	<control> LINE TABULATION (vertical tabulation)
0x0C	0x0C	<control> FORM FEED
0x0D	0x0D	<control> CARRIAGE RETURN
0x0E	0x0E	<control> SHIFT OUT

EBCDIC	LATIN	Unicode Character Name
0x0F	0x0F	<control> SHIFT IN
0x10	0x10	<control> DATA LINK ESCAPE
0x11	0x11	<control> DEVICE CONTROL ONE
0x12	0x12	<control> DEVICE CONTROL TWO
0x13	0x13	<control> DEVICE CONTROL THREE
0x14	0xD4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
0x15	0xD5	LATIN CAPITAL LETTER O WITH TILDE
0x16	0x08	<control> BACKSPACE
0x17	0xD6	LATIN CAPITAL LETTER O WITH DIAERESIS
0x18	0x18	<control> CANCEL
0x19	0x19	<control> END OF MEDIUM
0x1A	0xD7	LATIN CAPITAL LIGATURE OE
0x1B	0xD8	LATIN CAPITAL LETTER O WITH STROKE
0x1C	0x1C	<control> INFORMATION SEPARATOR FOUR (file separator)
0x1D	0x1D	<control> INFORMATION SEPARATOR THREE (group separator)
0x1E	0x1E	<control> INFORMATION SEPARATOR TWO (record separator)
0x1F	0x1F	<control> INFORMATION SEPARATOR ONE (unit separator)
0x20	0xE5	LATIN SMALL LETTER A WITH RING ABOVE
0x21	0xE6	LATIN SMALL LETTER AE
0x22	0xD9	LATIN CAPITAL LETTER U WITH GRAVE
0x23	0xE7	LATIN SMALL LETTER C WITH CEDILLA
0x24	0xE8	LATIN SMALL LETTER E WITH GRAVE
0x25	0x0A	<control> LINE FEED
0x26	0x17	<control> END OF TRANSMISSION BLOCK
0x27	0x1B	<control> ESCAPE
0x28	0xE9	LATIN SMALL LETTER E WITH ACUTE
0x29	0xF0	LATIN CAPITAL LETTER ETH
0x2A	0xF1	LATIN SMALL LETTER N WITH TILDE
0x2B	0xF2	LATIN SMALL LETTER O WITH GRAVE

EBCDIC	LATIN	Unicode Character Name
0x2C	0xF3	LATIN SMALL LETTER O WITH ACUTE
0x2D	0x05	<control> ENQUIRY
0x2E	0x06	<control> ACKNOWLEDGE
0x2F	0x07	<control> BELL
0x30	0xF4	LATIN SMALL LETTER O WITH CIRCUMFLEX
0x31	0xF5	LATIN SMALL LETTER O WITH TILDE
0x32	0x16	<control> SYNCHRONOUS IDLE
0x33	0xF6	LATIN SMALL LETTER O WITH DIAERESIS
0x34	0xF7	LATIN SMALL LIGATURE OE
0x35	0xC9	LATIN CAPITAL LETTER E WITH ACUTE
0x36	0xF8	LATIN SMALL LETTER O WITH STROKE
0x37	0x04	<control> END OF TRANSMISSION
0x38	0xF9	LATIN SMALL LETTER U WITH GRAVE
0x39	0xE2	LATIN SMALL LETTER A WITH CIRCUMFLEX
0x3A	0xE3	LATIN SMALL LETTER A WITH TILDE
0x3B	0xE4	LATIN SMALL LETTER A WITH DIAERESIS
0x3C	0x14	<control> DEVICE CONTROL FOUR
0x3D	0x15	<control> NEGATIVE ACKNOWLEDGE
0x3E	0xE0	LATIN SMALL LETTER A WITH GRAVE
0x3F	0x1A	<control> SUBSTITUTE
0x40	0x20	SPACE
0x41	0x81	DIAERESIS
0x42	0x82	ACUTE ACCENT
0x43	0x83	CEDILLA
0x44	0x84	MULTIPLICATION SIGN
0x45	0x85	DIVISION SIGN
0x46	0x86	<control> START OF SELECTED AREA
0x47	0x87	<control> END OF SELECTED AREA
0x48	0x88	<control> CHARACTER TABULATION SET

EBCDIC	LATIN	Unicode Character Name
0x49	0x89	<control> CHARACTER TABULATION WITH JUSTIFICATION
0x4A	0x8B	<control> PARTIAL LINE FORWARD
0x4B	0x2E	FULL STOP
0x4C	0x3C	LESS-THAN SIGN
0x4D	0x28	LEFT PARENTHESIS
0x4E	0x2B	PLUS SIGN
0x4F	0x8D	<control> REVERSE LINE FEED
0x50	0x26	AMPERSAND
0x51	0x91	<control> PRIVATE USE ONE
0x52	0x92	<control> PRIVATE USE TWO
0x53	0x93	<control> SET TRANSMIT STATE
0x54	0x94	<control> CANCEL CHARACTER
0x55	0x95	<control> MESSAGE WAITING
0x56	0x96	<control> START OF GUARDED AREA
0x57	0x97	<control> END OF GUARDED AREA
0x58	0x98	<control> START OF STRING
0x59	0x99	<control>
0x5A	0x21	EXCLAMATION MARK
0x5B	0x24	DOLLAR SIGN
0x5C	0x2A	ASTERISK
0x5D	0x29	RIGHT PARENTHESIS
0x5E	0x3B	SEMICOLON
0x5F	0x5E	CIRCUMFLEX ACCENT
0x60	0x2D	HYPHEN-MINUS
0x61	0x2F	SOLIDUS
0x62	0x9B	<control> CONTROL SEQUENCE INTRODUCER
0x63	0x9D	<control> OPERATING SYSTEM COMMAND
0x64	0xA2	CENT SIGN
0x65	0xA3	POUND SIGN

EBCDIC	LATIN	Unicode Character Name
0x66	0xA4	LATIN SMALL LETTER Y WITH ACUTE
0x67	0xA5	YEN SIGN
0x68	0xA6	BROKEN BAR
0x69	0xA7	SECTION SIGN
0x6A	0x7C	VERTICAL LINE
0x6B	0x2C	COMMA
0x6C	0x25	PERCENT SIGN
0x6D	0x5F	LOW LINE
0x6E	0x3E	GREATER-THAN SIGN
0x6F	0x3F	QUESTION MARK
0x70	0xC1	LATIN CAPITAL LETTER A WITH ACUTE
0x71	0xC2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
0x72	0xC3	LATIN CAPITAL LETTER A WITH TILDE
0x73	0xC4	LATIN CAPITAL LETTER A WITH DIAERESIS
0x74	0xC5	LATIN CAPITAL LETTER A WITH RING ABOVE
0x75	0xC6	LATIN CAPITAL LETTER AE
0x76	0xC7	LATIN CAPITAL LETTER C WITH CEDILLA
0x77	0xC8	LATIN CAPITAL LETTER E WITH GRAVE
0x78	0xA1	INVERTED EXCLAMATION MARK
0x79	0x60	GRAVE ACCENT
0x7A	0x3A	COLON
0x7B	0x23	NUMBER SIGN
0x7C	0x40	COMMERCIAL AT
0x7D	0x27	APOSTROPHE
0x7E	0x3D	EQUALS SIGN
0x7F	0x22	QUOTATION MARK
0x80	0x80	NO-BREAK SPACE
0x81	0x61	LATIN SMALL LETTER A
0x82	0x62	LATIN SMALL LETTER B

EBCDIC	LATIN	Unicode Character Name
0x83	0x63	LATIN SMALL LETTER C
0x84	0x64	LATIN SMALL LETTER D
0x85	0x65	LATIN SMALL LETTER E
0x86	0x66	LATIN SMALL LETTER F
0x87	0x67	LATIN SMALL LETTER G
0x88	0x68	LATIN SMALL LETTER H
0x89	0x69	LATIN SMALL LETTER I
0x8A	0x8A	<control> LINE TABULATION SET
0x8B	0xC0	LATIN CAPITAL LETTER A WITH GRAVE
0x8C	0x8C	<control> PARTIAL LINE BACKWARD
0x8D	0xAD	SOFT HYPHEN
0x8E	0x8E	<control> SINGLE SHIFT TWO
0x8F	0x8F	<control> SINGLE SHIFT THREE
0x90	0x90	<control> DEVICE CONTROL STRING
0x91	0x6A	LATIN SMALL LETTER J
0x92	0x6B	LATIN SMALL LETTER K
0x93	0x6C	LATIN SMALL LETTER L
0x94	0x6D	LATIN SMALL LETTER M
0x95	0x6E	LATIN SMALL LETTER N
0x96	0x6F	LATIN SMALL LETTER O
0x97	0x70	LATIN SMALL LETTER P
0x98	0x71	LATIN SMALL LETTER Q
0x99	0x72	LATIN SMALL LETTER R
0x9A	0x9A	<control> SINGLE CHARACTER INTRODUCER
0x9B	0xD0	LATIN SMALL LETTER ETH
0x9C	0x9C	<control> STRING TERMINATOR
0x9D	0xBD	VULGAR FRACTION ONE HALF
0x9E	0x9E	<control> PRIVACY MESSAGE
0x9F	0x9F	<control> APPLICATION PROGRAM COMMAND

EBCDIC	LATIN	Unicode Character Name
0xA0	0xA0	LATIN SMALL LETTER S WITH CARON
0xA1	0x7E	TILDE
0xA2	0x73	LATIN SMALL LETTER S
0xA3	0x74	LATIN SMALL LETTER T
0xA4	0x75	LATIN SMALL LETTER U
0xA5	0x76	LATIN SMALL LETTER V
0xA6	0x77	LATIN SMALL LETTER W
0xA7	0x78	LATIN SMALL LETTER X
0xA8	0x79	LATIN SMALL LETTER Y
0xA9	0x7A	LATIN SMALL LETTER Z
0xAA	0xAA	FEMININE ORDINAL INDICATOR
0xAB	0xAB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
0xAC	0xAC	NOT SIGN
0xAD	0x5B	LEFT SQUARE BRACKET
0xAE	0xAE	REGISTERED SIGN
0xAF	0xAF	MACRON
0xB0	0xB0	DEGREE SIGN
0xB1	0xB1	PLUS-MINUS SIGN
0xB2	0xB2	SUPERSCRPT TWO
0xB3	0xB3	SUPERSCRPT THREE
0xB4	0xB4	LATIN CAPITAL LETTER Y WITH ACUTE
0xB5	0xB5	MICRO SIGN
0xB6	0xB6	PILCROW SIGN
0xB7	0xB7	MIDDLE DOT
0xB8	0xB8	LATIN CAPITAL LETTER S WITH CARON
0xB9	0xB9	SUPERSCRPT ONE
0xBA	0xBA	MASCULINE ORDINAL INDICATOR
0xBB	0xBB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
0xBC	0xBC	VULGAR FRACTION ONE QUARTER

EBCDIC	LATIN	Unicode Character Name
0xBD	0x5D	RIGHT SQUARE BRACKET
0xBE	0xBE	VULGAR FRACTION THREE QUARTERS
0xBF	0xBF	INVERTED QUESTION MARK
0xC0	0x7B	LEFT CURLY BRACKET
0xC1	0x41	LATIN CAPITAL LETTER A
0xC2	0x42	LATIN CAPITAL LETTER B
0xC3	0x43	LATIN CAPITAL LETTER C
0xC4	0x44	LATIN CAPITAL LETTER D
0xC5	0x45	LATIN CAPITAL LETTER E
0xC6	0x46	LATIN CAPITAL LETTER F
0xC7	0x47	LATIN CAPITAL LETTER G
0xC8	0x48	LATIN CAPITAL LETTER H
0xC9	0x49	LATIN CAPITAL LETTER I
0xCA	0xCA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
0xCB	0xCB	LATIN CAPITAL LETTER E WITH DIAERESIS
0xCC	0xCC	LATIN CAPITAL LETTER I WITH GRAVE
0xCD	0xCD	LATIN CAPITAL LETTER I WITH ACUTE
0xCE	0xCE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
0xCF	0xCF	LATIN CAPITAL LETTER I WITH DIAERESIS
0xD0	0x7D	RIGHT CURLY BRACKET
0xD1	0x4A	LATIN CAPITAL LETTER J
0xD2	0x4B	LATIN CAPITAL LETTER K
0xD3	0x4C	LATIN CAPITAL LETTER L
0xD4	0x4D	LATIN CAPITAL LETTER M
0xD5	0x4E	LATIN CAPITAL LETTER N
0xD6	0x4F	LATIN CAPITAL LETTER O
0xD7	0x50	LATIN CAPITAL LETTER P
0xD8	0x51	LATIN CAPITAL LETTER Q
0xD9	0x52	LATIN CAPITAL LETTER R

EBCDIC	LATIN	Unicode Character Name
0xDA	0xDA	LATIN CAPITAL LETTER U WITH ACUTE
0xDB	0xDB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
0xDC	0xDC	LATIN CAPITAL LETTER U WITH DIAERESIS
0xDD	0xDD	LATIN CAPITAL LETTER Y WITH DIAERESIS
0xDE	0xDE	LATIN SMALL LETTER THORN
0xDF	0xDF	LATIN SMALL LETTER SHARP S
0xE0	0x5C	REVERSE SOLIDUS
0xE1	0xE1	LATIN SMALL LETTER A WITH ACUTE
0xE2	0x53	LATIN CAPITAL LETTER S
0xE3	0x54	LATIN CAPITAL LETTER T
0xE4	0x55	LATIN CAPITAL LETTER U
0xE5	0x56	LATIN CAPITAL LETTER V
0xE6	0x57	LATIN CAPITAL LETTER W
0xE7	0x58	LATIN CAPITAL LETTER X
0xE8	0x59	LATIN CAPITAL LETTER Y
0xE9	0x5A	LATIN CAPITAL LETTER Z
0xEA	0xEA	LATIN SMALL LETTER E WITH CIRCUMFLEX
0xEB	0xEB	LATIN SMALL LETTER E WITH DIAERESIS
0xEC	0xEC	LATIN SMALL LETTER I WITH GRAVE
0xED	0xED	LATIN SMALL LETTER I WITH ACUTE
0xEE	0xEE	LATIN SMALL LETTER I WITH CIRCUMFLEX
0xEF	0xEF	LATIN SMALL LETTER I WITH DIAERESIS
0xF0	0x30	DIGIT ZERO
0xF1	0x31	DIGIT ONE
0xF2	0x32	DIGIT TWO
0xF3	0x33	DIGIT THREE
0xF4	0x34	DIGIT FOUR
0xF5	0x35	DIGIT FIVE
0xF6	0x36	DIGIT SIX

EBCDIC	LATIN	Unicode Character Name
0xF7	0x37	DIGIT SEVEN
0xF8	0x38	DIGIT EIGHT
0xF9	0x39	DIGIT NINE
0xFA	0xFA	LATIN SMALL LETTER U WITH ACUTE
0xFB	0xFB	LATIN SMALL LETTER U WITH CIRCUMFLEX
0xFC	0xFC	LATIN SMALL LETTER U WITH DIAERESIS
0xFD	0xFD	LATIN SMALL LETTER Y WITH DIAERESIS
0xFE	0xFE	LATIN CAPITAL LETTER THORN
0xFF	0xFF	EURO SIGN

Client EBCDIC to Server UNICODE

The following table shows how single-byte Latin characters from the EBCDIC client character set map to characters in the UNICODE server character set. The translation works both ways, from EBCDIC to UNICODE and from UNICODE to EBCDIC.

EBCDIC	UNICODE	Unicode Character Name
0x00	U+0000	<control> NULL
0x01	U+0001	<control> START OF HEADING
0x02	U+0002	<control> START OF TEXT
0x03	U+0003	<control> END OF TEXT
0x04	U+00A4	CURRENCY SIGN
0x05	U+0009	<control> CHARACTER TABULATION (horizontal tabulation)
0x06	U+00A9	COPYRIGHT SIGN
0x07	U+007F	<control> DELETE
0x08	U+00D1	LATIN CAPITAL LETTER N WITH TILDE
0x09	U+00D2	LATIN CAPITAL LETTER O WITH GRAVE
0x0A	U+00D3	LATIN CAPITAL LETTER O WITH ACUTE
0x0B	U+000B	<control> LINE TABULATION (vertical tabulation)
0x0C	U+000C	<control> FORM FEED
0x0D	U+000D	<control> CARRIAGE RETURN

EBCDIC	UNICODE	Unicode Character Name
0x0E	U+000E	<control> SHIFT OUT
0x0F	U+000F	<control> SHIFT IN
0x10	U+0010	<control> DATA LINK ESCAPE
0x11	U+0011	<control> DEVICE CONTROL ONE
0x12	U+0012	<control> DEVICE CONTROL TWO
0x13	U+0013	<control> DEVICE CONTROL THREE
0x14	U+00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
0x15	U+00D5	LATIN CAPITAL LETTER O WITH TILDE
0x16	U+0008	<control> BACKSPACE
0x17	U+00D6	LATIN CAPITAL LETTER O WITH DIAERESIS
0x18	U+0018	<control> CANCEL
0x19	U+0019	<control> END OF MEDIUM
0x1A	U+0152	LATIN CAPITAL LIGATURE OE
0x1B	U+00D8	LATIN CAPITAL LETTER O WITH STROKE
0x1C	U+001C	<control> INFORMATION SEPARATOR FOUR (file separator)
0x1D	U+001D	<control> INFORMATION SEPARATOR THREE (group separator)
0x1E	U+001E	<control> INFORMATION SEPARATOR TWO (record separator)
0x1F	U+001F	<control> INFORMATION SEPARATOR ONE (unit separator)
0x20	U+00E5	LATIN SMALL LETTER A WITH RING ABOVE
0x21	U+00E6	LATIN SMALL LETTER AE
0x22	U+00D9	LATIN CAPITAL LETTER U WITH GRAVE
0x23	U+00E7	LATIN SMALL LETTER C WITH CEDILLA
0x24	U+00E8	LATIN SMALL LETTER E WITH GRAVE
0x25	U+000A	<control> LINE FEED
0x26	U+0017	<control> END OF TRANSMISSION BLOCK
0x27	U+001B	<control> ESCAPE
0x28	U+00E9	LATIN SMALL LETTER E WITH ACUTE
0x29	U+00D0	LATIN CAPITAL LETTER ETH
0x2A	U+00F1	LATIN SMALL LETTER N WITH TILDE

EBCDIC	UNICODE	Unicode Character Name
0x2B	U+00F2	LATIN SMALL LETTER O WITH GRAVE
0x2C	U+00F3	LATIN SMALL LETTER O WITH ACUTE
0x2D	U+0005	<control> ENQUIRY
0x2E	U+0006	<control> ACKNOWLEDGE
0x2F	U+0007	<control> BELL
0x30	U+00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX
0x31	U+00F5	LATIN SMALL LETTER O WITH TILDE
0x32	U+0016	<control> SYNCHRONOUS IDLE
0x33	U+00F6	LATIN SMALL LETTER O WITH DIAERESIS
0x34	U+0153	LATIN SMALL LIGATURE OE
0x35	U+00C9	LATIN CAPITAL LETTER E WITH ACUTE
0x36	U+00F8	LATIN SMALL LETTER O WITH STROKE
0x37	U+0004	<control> END OF TRANSMISSION
0x38	U+00F9	LATIN SMALL LETTER U WITH GRAVE
0x39	U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX
0x3A	U+00E3	LATIN SMALL LETTER A WITH TILDE
0x3B	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
0x3C	U+0014	<control> DEVICE CONTROL FOUR
0x3D	U+0015	<control> NEGATIVE ACKNOWLEDGE
0x3E	U+00E0	LATIN SMALL LETTER A WITH GRAVE
0x3F	U+001A	<control> SUBSTITUTE
0x40	U+0020	SPACE
0x41	U+00A8	DIAERESIS
0x42	U+00B4	ACUTE ACCENT
0x43	U+00B8	CEDILLA
0x44	U+00D7	MULTIPLICATION SIGN
0x45	U+00F7	DIVISION SIGN
0x46	U+0086	<control> START OF SELECTED AREA
0x47	U+0087	<control> END OF SELECTED AREA

EBCDIC	UNICODE	Unicode Character Name
0x48	U+0088	<control> CHARACTER TABULATION SET
0x49	U+0089	<control> CHARACTER TABULATION WITH JUSTIFICATION
0x4A	U+008B	<control> PARTIAL LINE FORWARD
0x4B	U+002E	FULL STOP
0x4C	U+003C	LESS-THAN SIGN
0x4D	U+0028	LEFT PARENTHESIS
0x4E	U+002B	PLUS SIGN
0x4F	U+008D	<control> REVERSE LINE FEED
0x50	U+0026	AMPERSAND
0x51	U+0091	<control> PRIVATE USE ONE
0x52	U+0092	<control> PRIVATE USE TWO
0x53	U+0093	<control> SET TRANSMIT STATE
0x54	U+0094	<control> CANCEL CHARACTER
0x55	U+0095	<control> MESSAGE WAITING
0x56	U+0096	<control> START OF GUARDED AREA
0x57	U+0097	<control> END OF GUARDED AREA
0x58	U+0098	<control> START OF STRING
0x59	U+0099	<control>
0x5A	U+0021	EXCLAMATION MARK
0x5B	U+0024	DOLLAR SIGN
0x5C	U+002A	ASTERISK
0x5D	U+0029	RIGHT PARENTHESIS
0x5E	U+003B	SEMICOLON
0x5F	U+005E	CIRCUMFLEX ACCENT
0x60	U+002D	HYPHEN-MINUS
0x61	U+002F	SOLIDUS
0x62	U+009B	<control> CONTROL SEQUENCE INTRODUCER
0x63	U+009D	<control> OPERATING SYSTEM COMMAND
0x64	U+00A2	CENT SIGN

EBCDIC	UNICODE	Unicode Character Name
0x65	U+00A3	POUND SIGN
0x66	U+00FD	LATIN SMALL LETTER Y WITH ACUTE
0x67	U+00A5	YEN SIGN
0x68	U+00A6	BROKEN BAR
0x69	U+00A7	SECTION SIGN
0x6A	U+007C	VERTICAL LINE
0x6B	U+002C	COMMA
0x6C	U+0025	PERCENT SIGN
0x6D	U+005F	LOW LINE
0x6E	U+003E	GREATER-THAN SIGN
0x6F	U+003F	QUESTION MARK
0x70	U+00C1	LATIN CAPITAL LETTER A WITH ACUTE
0x71	U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
0x72	U+00C3	LATIN CAPITAL LETTER A WITH TILDE
0x73	U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS
0x74	U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE
0x75	U+00C6	LATIN CAPITAL LETTER AE
0x76	U+00C7	LATIN CAPITAL LETTER C WITH CEDILLA
0x77	U+00C8	LATIN CAPITAL LETTER E WITH GRAVE
0x78	U+00A1	INVERTED EXCLAMATION MARK
0x79	U+0060	GRAVE ACCENT
0x7A	U+003A	COLON
0x7B	U+0023	NUMBER SIGN
0x7C	U+0040	COMMERCIAL AT
0x7D	U+0027	APOSTROPHE
0x7E	U+003D	EQUALS SIGN
0x7F	U+0022	QUOTATION MARK
0x80	U+00A0	NO-BREAK SPACE
0x81	U+0061	LATIN SMALL LETTER A

EBCDIC	UNICODE	Unicode Character Name
0x82	U+0062	LATIN SMALL LETTER B
0x83	U+0063	LATIN SMALL LETTER C
0x84	U+0064	LATIN SMALL LETTER D
0x85	U+0065	LATIN SMALL LETTER E
0x86	U+0066	LATIN SMALL LETTER F
0x87	U+0067	LATIN SMALL LETTER G
0x88	U+0068	LATIN SMALL LETTER H
0x89	U+0069	LATIN SMALL LETTER I
0x8A	U+008A	<control> LINE TABULATION SET
0x8B	U+00C0	LATIN CAPITAL LETTER A WITH GRAVE
0x8C	U+008C	<control> PARTIAL LINE BACKWARD
0x8D	U+00AD	SOFT HYPHEN
0x8E	U+008E	<control> SINGLE SHIFT TWO
0x8F	U+008F	<control> SINGLE SHIFT THREE
0x90	U+0090	<control> DEVICE CONTROL STRING
0x91	U+006A	LATIN SMALL LETTER J
0x92	U+006B	LATIN SMALL LETTER K
0x93	U+006C	LATIN SMALL LETTER L
0x94	U+006D	LATIN SMALL LETTER M
0x95	U+006E	LATIN SMALL LETTER N
0x96	U+006F	LATIN SMALL LETTER O
0x97	U+0070	LATIN SMALL LETTER P
0x98	U+0071	LATIN SMALL LETTER Q
0x99	U+0072	LATIN SMALL LETTER R
0x9A	U+009A	<control> SINGLE CHARACTER INTRODUCER
0x9B	U+00F0	LATIN SMALL LETTER ETH
0x9C	U+009C	<control> STRING TERMINATOR
0x9D	U+00BD	VULGAR FRACTION ONE HALF
0x9E	U+009E	<control> PRIVACY MESSAGE

EBCDIC	UNICODE	Unicode Character Name
0x9F	U+009F	<control> APPLICATION PROGRAM COMMAND
0xA0	U+0161	LATIN SMALL LETTER S WITH CARON
0xA1	U+007E	TILDE
0xA2	U+0073	LATIN SMALL LETTER S
0xA3	U+0074	LATIN SMALL LETTER T
0xA4	U+0075	LATIN SMALL LETTER U
0xA5	U+0076	LATIN SMALL LETTER V
0xA6	U+0077	LATIN SMALL LETTER W
0xA7	U+0078	LATIN SMALL LETTER X
0xA8	U+0079	LATIN SMALL LETTER Y
0xA9	U+007A	LATIN SMALL LETTER Z
0xAA	U+00AA	FEMININE ORDINAL INDICATOR
0xAB	U+00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
0xAC	U+00AC	NOT SIGN
0xAD	U+005B	LEFT SQUARE BRACKET
0xAE	U+00AE	REGISTERED SIGN
0xAF	U+00AF	MACRON
0xB0	U+00B0	DEGREE SIGN
0xB1	U+00B1	PLUS-MINUS SIGN
0xB2	U+00B2	SUPERSCRPT TWO
0xB3	U+00B3	SUPERSCRPT THREE
0xB4	U+00DD	LATIN CAPITAL LETTER Y WITH ACUTE
0xB5	U+00B5	MICRO SIGN
0xB6	U+00B6	PILCROW SIGN
0xB7	U+00B7	MIDDLE DOT
0xB8	U+0160	LATIN CAPITAL LETTER S WITH CARON
0xB9	U+00B9	SUPERSCRPT ONE
0xBA	U+00BA	MASCULINE ORDINAL INDICATOR
0xBB	U+00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK

EBCDIC	UNICODE	Unicode Character Name
0xBC	U+00BC	VULGAR FRACTION ONE QUARTER
0xBD	U+005D	RIGHT SQUARE BRACKET
0xBE	U+00BE	VULGAR FRACTION THREE QUARTERS
0xBF	U+00BF	INVERTED QUESTION MARK
0xC0	U+007B	LEFT CURLY BRACKET
0xC1	U+0041	LATIN CAPITAL LETTER A
0xC2	U+0042	LATIN CAPITAL LETTER B
0xC3	U+0043	LATIN CAPITAL LETTER C
0xC4	U+0044	LATIN CAPITAL LETTER D
0xC5	U+0045	LATIN CAPITAL LETTER E
0xC6	U+0046	LATIN CAPITAL LETTER F
0xC7	U+0047	LATIN CAPITAL LETTER G
0xC8	U+0048	LATIN CAPITAL LETTER H
0xC9	U+0049	LATIN CAPITAL LETTER I
0xCA	U+00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
0xCB	U+00CB	LATIN CAPITAL LETTER E WITH DIAERESIS
0xCC	U+00CC	LATIN CAPITAL LETTER I WITH GRAVE
0xCD	U+00CD	LATIN CAPITAL LETTER I WITH ACUTE
0xCE	U+00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
0xCF	U+00CF	LATIN CAPITAL LETTER I WITH DIAERESIS
0xD0	U+007D	RIGHT CURLY BRACKET
0xD1	U+004A	LATIN CAPITAL LETTER J
0xD2	U+004B	LATIN CAPITAL LETTER K
0xD3	U+004C	LATIN CAPITAL LETTER L
0xD4	U+004D	LATIN CAPITAL LETTER M
0xD5	U+004E	LATIN CAPITAL LETTER N
0xD6	U+004F	LATIN CAPITAL LETTER O
0xD7	U+0050	LATIN CAPITAL LETTER P
0xD8	U+0051	LATIN CAPITAL LETTER Q

EBCDIC	UNICODE	Unicode Character Name
0xD9	U+0052	LATIN CAPITAL LETTER R
0xDA	U+00DA	LATIN CAPITAL LETTER U WITH ACUTE
0xDB	U+00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
0xDC	U+00DC	LATIN CAPITAL LETTER U WITH DIAERESIS
0xDD	U+0178	LATIN CAPITAL LETTER Y WITH DIAERESIS
0xDE	U+00FE	LATIN SMALL LETTER THORN
0xDF	U+00DF	LATIN SMALL LETTER SHARP S
0xE0	U+005C	REVERSE SOLIDUS
0xE1	U+00E1	LATIN SMALL LETTER A WITH ACUTE
0xE2	U+0053	LATIN CAPITAL LETTER S
0xE3	U+0054	LATIN CAPITAL LETTER T
0xE4	U+0055	LATIN CAPITAL LETTER U
0xE5	U+0056	LATIN CAPITAL LETTER V
0xE6	U+0057	LATIN CAPITAL LETTER W
0xE7	U+0058	LATIN CAPITAL LETTER X
0xE8	U+0059	LATIN CAPITAL LETTER Y
0xE9	U+005A	LATIN CAPITAL LETTER Z
0xEA	U+00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX
0xEB	U+00EB	LATIN SMALL LETTER E WITH DIAERESIS
0xEC	U+00EC	LATIN SMALL LETTER I WITH GRAVE
0xED	U+00ED	LATIN SMALL LETTER I WITH ACUTE
0xEE	U+00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX
0xEF	U+00EF	LATIN SMALL LETTER I WITH DIAERESIS
0xF0	U+0030	DIGIT ZERO
0xF1	U+0031	DIGIT ONE
0xF2	U+0032	DIGIT TWO
0xF3	U+0033	DIGIT THREE
0xF4	U+0034	DIGIT FOUR
0xF5	U+0035	DIGIT FIVE

EBCDIC	UNICODE	Unicode Character Name
0xF6	U+0036	DIGIT SIX
0xF7	U+0037	DIGIT SEVEN
0xF8	U+0038	DIGIT EIGHT
0xF9	U+0039	DIGIT NINE
0xFA	U+00FA	LATIN SMALL LETTER U WITH ACUTE
0xFB	U+00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX
0xFC	U+00FC	LATIN SMALL LETTER U WITH DIAERESIS
0xFD	U+00FF	LATIN SMALL LETTER Y WITH DIAERESIS
0xFE	U+00DE	LATIN CAPITAL LETTER THORN
0xFF	U+20AC	EURO SIGN

Mappings for String Functions

Overview

The following provides internal mappings that Teradata Database uses for the TRANSLATE, LOWER, and UPPER functions.

The following mappings apply to the TRANSLATE function:

- [Pad Character Translation](#)
- [Space Folding](#)
- [Non-GRAPHIC Characters](#)
- [UNICODE Halfwidth to Fullwidth](#)
- [UNICODE Fullwidth to Halfwidth](#)
- [UNICODE to LATIN](#)

For the UNICODE to KANJISJIS mapping used by the TRANSLATE function, see [UNICODE to KanjiShiftJIS Single-Byte](#).

The following mapping applies to the LOWER function:

- [LATIN Uppercase to Lowercase](#)
- [UNICODE Uppercase to Lowercase](#)

The following mappings apply to the UPPER function:

- [Multibyte KANJISJIS Lowercase to Uppercase](#)
- [Single-Byte KANJISJIS Lowercase to Uppercase](#)
- [UNICODE Lowercase to Uppercase](#)

For details on the TRANSLATE, LOWER, and UPPER functions, see *Teradata Vantage™ - SQL Functions, Expressions, and Predicates*, B035-1145.

Pad Character Translation

The following table shows how the IDEOGRAPHIC SPACE character in the GRAPHIC server character set maps to the SPACE character in the UNICODE server character set. This mapping applies to character strings that you translate from GRAPHIC to UNICODE using the TRANSLATE string function and specifying GRAPHIC_TO_UNICODE_PadSpace.

GRAPHIC	Unicode Character Name	UNICODE	Unicode Character Name
0x3000	IDEOGRAPHIC SPACE	U+0020	SPACE

The following table shows how the SPACE character in the UNICODE server character set maps to the IDEOGRAPHIC SPACE character in the GRAPHIC server character set. This mapping applies to character strings that you translate from UNICODE to GRAPHIC using the TRANSLATE string function and specifying UNICODE_TO_GRAPHIC_PadGraphic.

UNICODE	Unicode Character Name	GRAPHIC	Unicode Character Name
U+0020	SPACE	0x3000	IDEOGRAPHIC SPACE

Space Folding

The following table shows which characters map to the SPACE character in the UNICODE server character when you translate UNICODE character strings using the TRANSLATE string function and specify UNICODE_TO_UNICODE_FoldSpace.

UNICODE Input	Unicode Character Name	UNICODE Fold Output	Unicode Character Name
U+00A0	NO-BREAK SPACE	U+0020	SPACE
U+0x00A0	OGHAM SPACE MARK	U+0x0020	SPACE
U+0x180E	MONGOLIAN VOWEL SEPARATOR	U+0x0020	SPACE
U+2000	EN QUAD	U+0020	SPACE
U+2001	EM QUAD	U+0020	SPACE
U+2002	EN SPACE	U+0020	SPACE
U+2003	EM SPACE	U+0020	SPACE
U+2004	THREE-PER-EM SPACE	U+0020	SPACE
U+2005	FOUR-PER-EM SPACE	U+0020	SPACE
U+2006	SIX-PER-EM SPACE	U+0020	SPACE
U+2007	FIGURE SPACE	U+0020	SPACE
U+2008	PUNCTUATION SPACE	U+0020	SPACE
U+2009	THIN SPACE	U+0020	SPACE
U+200A	HAIR SPACE	U+0020	SPACE
U+202F	NARROW NO-BREAK SPACE	U+0020	SPACE
U+205F	MEDIUM MATHEMATICAL SPACE	U+0020	SPACE
U+3000	IDEOGRAPHIC SPACE	U+0020	SPACE

Non-GRAPHIC Characters

The following table identifies non-GRAPHIC characters in the UNICODE server character set.

Teradata Database substitutes the GRAPHIC REPLACEMENT CHARACTER (0xFFFD) for non-GRAPHIC characters when you use the TRANSLATE function and specify the following:

- WITH ERROR option
- UNICODE_TO_GRAPHIC

Code	Unicode Character Name	Code	Unicode Character Name
U+0000	<control> NULL	U+0092	<control> PRIVATE USE TWO
U+0001	<control> START OF HEADING	U+0093	<control> SET TRANSMIT RATE
U+0002	<control> START OF TEXT	U+0094	<control> CANCEL CHARACTER
U+0003	<control> END OF TEXT	U+0095	<control> MESSAGE WAITING
U+0004	<control> END OF TRANSMISSION	U+0096	<control> START OF GUARDED AREA
U+0005	<control> ENQUIRY	U+0097	<control> END OF GUARDED AREA
U+0006	<control> ACKNOWLEDGE	U+0098	<control> START OF STRING
U+0007	<control> BELL	U+0099	<control>
U+0008	<control> BACKSPACE	U+009A	<control> SINGLE CHARACTER INTRODUCER
U+0009	<control> CHARACTER TABULATION (horizontal tabulation)	U+009B	<control> CONTROL SEQUENCE INTRODUCER
U+000A	<control> LINE FEED	U+009C	<control> STRING TERMINATOR
U+000B	<control> LINE TABULATION (vertical tabulation)	U+009D	<control> OPERATING SYSTEM COMMAND
U+000C	<control> FORM FEED	U+009E	<control> PRIVACY MESSAGE
U+000D	<control> CARRIAGE RETURN	U+009F	<control> APPLICATION PROGRAM COMMAND
U+000E	<control> SHIFT OUT	U+00A2	CENT SIGN
U+000F	<control> SHIFT IN	U+00A3	POUND SIGN
U+0010	<control> DATA LINK ESCAPE	U+00A5	YEN SIGN
U+0011	<control> DEVICE CONTROL ONE	U+00A6	BROKEN BAR
U+0012	<control> DEVICE CONTROL TWO	U+00AC	NOT SIGN
U+0013	<control> DEVICE CONTROL THREE	U+00AF	MACRON
U+0014	<control> DEVICE CONTROL FOUR	U+203E	OVERLINE

Code	Unicode Character Name	Code	Unicode Character Name
U+0015	<control> NEGATIVE ACKNOWLEDGE	U+20A9	WON SIGN
U+0016	<control> SYNCHRONOUS IDLE	0x2985	LEFT WHITE PARENTHESIS
U+0017	<control> END OF TRANSMISSION BLOCK	0x2986	RIGHT WHITE PARENTHESIS
U+0018	<control> CANCEL	U+FF61	HALFWIDTH IDEOGRAPHIC FULL STOP
U+0019	<control> END OF MEDIUM	U+FF62	HALFWIDTH LEFT CORNER BRACKET
U+001A	<control> SUBSTITUTE	U+FF63	HALFWIDTH RIGHT CORNER BRACKET
U+001B	<control> ESCAPE	U+FF64	HALFWIDTH IDEOGRAPHIC COMMA
U+001C	<control> INFORMATION SEPARATOR FOUR (file separator)	U+FF65	HALFWIDTH KATAKANA MIDDLE DOT
U+001D	<control> INFORMATION SEPARATOR THREE (group separator)	U+FF66	HALFWIDTH KATAKANA LETTER WO
U+001E	<control> INFORMATION SEPARATOR TWO (record separator)	U+FF67	HALFWIDTH KATAKANA LETTER SMALL A
U+001F	<control> INFORMATION SEPARATOR ONE (unit separator)	U+FF68	HALFWIDTH KATAKANA LETTER SMALL I
U+0020	SPACE	U+FF69	HALFWIDTH KATAKANA LETTER SMALL U
U+0021	EXCLAMATION MARK	U+FF6A	HALFWIDTH KATAKANA LETTER SMALL E
U+0022	QUOTATION MARK	U+FF6B	HALFWIDTH KATAKANA LETTER SMALL O
U+0023	NUMBER SIGN	U+FF6C	HALFWIDTH KATAKANA LETTER SMALL YA
U+0024	DOLLAR SIGN	U+FF6D	HALFWIDTH KATAKANA LETTER SMALL YU
U+0025	PERCENT SIGN	U+FF6E	HALFWIDTH KATAKANA LETTER SMALL YO
U+0026	AMPERSAND	U+FF6F	HALFWIDTH KATAKANA LETTER SMALL TU
U+0027	APOSTROPHE	U+FF70	HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK
U+0028	LEFT PARENTHESIS	U+FF71	HALFWIDTH KATAKANA LETTER A

Code	Unicode Character Name	Code	Unicode Character Name
U+0029	RIGHT PARENTHESIS	U+FF72	HALFWIDTH KATAKANA LETTER I
U+002A	ASTERISK	U+FF73	HALFWIDTH KATAKANA LETTER U
U+002B	PLUS SIGN	U+FF74	HALFWIDTH KATAKANA LETTER E
U+002C	COMMA	U+FF75	HALFWIDTH KATAKANA LETTER O
U+002D	HYPHEN-MINUS	U+FF76	HALFWIDTH KATAKANA LETTER KA
U+002E	FULL STOP	U+FF77	HALFWIDTH KATAKANA LETTER KI
U+002F	SOLIDUS	U+FF78	HALFWIDTH KATAKANA LETTER KU
U+0030	DIGIT ZERO	U+FF79	HALFWIDTH KATAKANA LETTER KE
U+0031	DIGIT ONE	U+FF7A	HALFWIDTH KATAKANA LETTER KO
U+0032	DIGIT TWO	U+FF7B	HALFWIDTH KATAKANA LETTER SA
U+0033	DIGIT THREE	U+FF7C	HALFWIDTH KATAKANA LETTER SI
U+0034	DIGIT FOUR	U+FF7D	HALFWIDTH KATAKANA LETTER SU
U+0035	DIGIT FIVE	U+FF7E	HALFWIDTH KATAKANA LETTER SE
U+0036	DIGIT SIX	U+FF7F	HALFWIDTH KATAKANA LETTER SO
U+0037	DIGIT SEVEN	U+FF80	HALFWIDTH KATAKANA LETTER TA
U+0038	DIGIT EIGHT	U+FF81	HALFWIDTH KATAKANA LETTER TI
U+0039	DIGIT NINE	U+FF82	HALFWIDTH KATAKANA LETTER TU
U+003A	COLON	U+FF83	HALFWIDTH KATAKANA LETTER TE
U+003B	SEMICOLON	U+FF84	HALFWIDTH KATAKANA LETTER TO
U+003C	LESS-THAN SIGN	U+FF85	HALFWIDTH KATAKANA LETTER NA
U+003D	EQUALS SIGN	U+FF86	HALFWIDTH KATAKANA LETTER NI
U+003E	GREATER-THAN SIGN	U+FF87	HALFWIDTH KATAKANA LETTER NU
U+003F	QUESTION MARK	U+FF88	HALFWIDTH KATAKANA LETTER NE
U+0040	COMMERCIAL AT	U+FF89	HALFWIDTH KATAKANA LETTER NO
U+0041	LATIN CAPITAL LETTER A	U+FF8A	HALFWIDTH KATAKANA LETTER HA
U+0042	LATIN CAPITAL LETTER B	U+FF8B	HALFWIDTH KATAKANA LETTER HI
U+0043	LATIN CAPITAL LETTER C	U+FF8C	HALFWIDTH KATAKANA LETTER HU
U+0044	LATIN CAPITAL LETTER D	U+FF8D	HALFWIDTH KATAKANA LETTER HE
U+0045	LATIN CAPITAL LETTER E	U+FF8E	HALFWIDTH KATAKANA LETTER HO

Code	Unicode Character Name	Code	Unicode Character Name
U+0046	LATIN CAPITAL LETTER F	U+FF8F	HALFWIDTH KATAKANA LETTER MA
U+0047	LATIN CAPITAL LETTER G	U+FF90	HALFWIDTH KATAKANA LETTER MI
U+0048	LATIN CAPITAL LETTER H	U+FF91	HALFWIDTH KATAKANA LETTER MU
U+0049	LATIN CAPITAL LETTER I	U+FF92	HALFWIDTH KATAKANA LETTER ME
U+004A	LATIN CAPITAL LETTER J	U+FF93	HALFWIDTH KATAKANA LETTER MO
U+004B	LATIN CAPITAL LETTER K	U+FF94	HALFWIDTH KATAKANA LETTER YA
U+004C	LATIN CAPITAL LETTER L	U+FF95	HALFWIDTH KATAKANA LETTER YU
U+004D	LATIN CAPITAL LETTER M	U+FF96	HALFWIDTH KATAKANA LETTER YO
U+004E	LATIN CAPITAL LETTER N	U+FF97	HALFWIDTH KATAKANA LETTER RA
U+004F	LATIN CAPITAL LETTER O	U+FF98	HALFWIDTH KATAKANA LETTER RI
U+0050	LATIN CAPITAL LETTER P	U+FF99	HALFWIDTH KATAKANA LETTER RU
U+0051	LATIN CAPITAL LETTER Q	U+FF9A	HALFWIDTH KATAKANA LETTER RE
U+0052	LATIN CAPITAL LETTER R	U+FF9B	HALFWIDTH KATAKANA LETTER RO
U+0053	LATIN CAPITAL LETTER S	U+FF9C	HALFWIDTH KATAKANA LETTER WA
U+0054	LATIN CAPITAL LETTER T	U+FF9D	HALFWIDTH KATAKANA LETTER N
U+0055	LATIN CAPITAL LETTER U	U+FF9E	HALFWIDTH KATAKANA VOICED SOUND MARK
U+0056	LATIN CAPITAL LETTER V	U+FF9F	HALFWIDTH KATAKANA SEMI-VOICED SOUND MARK
U+0057	LATIN CAPITAL LETTER W	U+FFA0	HALFWIDTH HANGUL FILLER
U+0058	LATIN CAPITAL LETTER X	U+FFA1	HALFWIDTH HANGUL LETTER KIYEOK
U+0059	LATIN CAPITAL LETTER Y	U+FFA2	HALFWIDTH HANGUL LETTER SSANGKIYEOK
U+005A	LATIN CAPITAL LETTER Z	U+FFA3	HALFWIDTH HANGUL LETTER KIYEOK-SIOS
U+005B	LEFT SQUARE BRACKET	U+FFA4	HALFWIDTH HANGUL LETTER NIEUN
U+005C	REVERSE SOLIDUS	U+FFA5	HALFWIDTH HANGUL LETTER NIEUN- CIEUC
U+005D	RIGHT SQUARE BRACKET	U+FFA6	HALFWIDTH HANGUL LETTER NIEUN- HIEUH
U+005E	CIRCUMFLEX ACCENT	U+FFA7	HALFWIDTH HANGUL LETTER TIKEUT

Code	Unicode Character Name	Code	Unicode Character Name
U+005F	LOW LINE	U+FFA8	HALFWIDTH HANGUL LETTER SSANGTIKEUT
U+0060	GRAVE ACCENT	U+FFA9	HALFWIDTH HANGUL LETTER RIEUL
U+0061	LATIN SMALL LETTER A	U+FFAA	HALFWIDTH HANGUL LETTER RIEUL-KIYEOK
U+0062	LATIN SMALL LETTER B	U+FFAB	HALFWIDTH HANGUL LETTER RIEUL-MIEUM
U+0063	LATIN SMALL LETTER C	U+FFAC	HALFWIDTH HANGUL LETTER RIEUL-PIEUP
U+0064	LATIN SMALL LETTER D	U+FFAD	HALFWIDTH HANGUL LETTER RIEUL-SIOS
U+0065	LATIN SMALL LETTER E	U+FFAE	HALFWIDTH HANGUL LETTER RIEUL-THIEUTH
U+0066	LATIN SMALL LETTER F	U+FFAF	HALFWIDTH HANGUL LETTER RIEUL-PHIEUPH
U+0067	LATIN SMALL LETTER G	U+FFB0	HALFWIDTH HANGUL LETTER RIEUL-HIEUH
U+0068	LATIN SMALL LETTER H	U+FFB1	HALFWIDTH HANGUL LETTER MIEUM
U+0069	LATIN SMALL LETTER I	U+FFB2	HALFWIDTH HANGUL LETTER PIEUP
U+006A	LATIN SMALL LETTER J	U+FFB3	HALFWIDTH HANGUL LETTER SSANGPIEUP
U+006B	LATIN SMALL LETTER K	U+FFB4	HALFWIDTH HANGUL LETTER PIEUP-SIOS
U+006C	LATIN SMALL LETTER L	U+FFB5	HALFWIDTH HANGUL LETTER SIOS
U+006D	LATIN SMALL LETTER M	U+FFB6	HALFWIDTH HANGUL LETTER SSANGSIOS
U+006E	LATIN SMALL LETTER N	U+FFB7	HALFWIDTH HANGUL LETTER IEUNG
U+006F	LATIN SMALL LETTER O	U+FFB8	HALFWIDTH HANGUL LETTER CIEUC
U+0070	LATIN SMALL LETTER P	U+FFB9	HALFWIDTH HANGUL LETTER SSANGCIEUC
U+0071	LATIN SMALL LETTER Q	U+FFBA	HALFWIDTH HANGUL LETTER CHIEUCH
U+0072	LATIN SMALL LETTER R	U+FFBB	HALFWIDTH HANGUL LETTER KHIEUKH
U+0073	LATIN SMALL LETTER S	U+FFBC	HALFWIDTH HANGUL LETTER THIEUTH

Code	Unicode Character Name	Code	Unicode Character Name
U+0074	LATIN SMALL LETTER T	U+FFBD	HALFWIDTH HANGUL LETTER PHIEUPH
U+0075	LATIN SMALL LETTER U	U+FFBE	HALFWIDTH HANGUL LETTER HIEUH
U+0076	LATIN SMALL LETTER V	U+FFC2	HALFWIDTH HANGUL LETTER A
U+0077	LATIN SMALL LETTER W	U+FFC3	HALFWIDTH HANGUL LETTER AE
U+0078	LATIN SMALL LETTER X	U+FFC4	HALFWIDTH HANGUL LETTER YA
U+0079	LATIN SMALL LETTER Y	U+FFC5	HALFWIDTH HANGUL LETTER YAE
U+007A	LATIN SMALL LETTER Z	U+FFC6	HALFWIDTH HANGUL LETTER EO
U+007B	LEFT CURLY BRACKET	U+FFC7	HALFWIDTH HANGUL LETTER E
U+007C	VERTICAL LINE	U+FFCA	HALFWIDTH HANGUL LETTER YEO
U+007D	RIGHT CURLY BRACKET	U+FFCB	HALFWIDTH HANGUL LETTER YE
U+007E	TILDE	U+FFCC	HALFWIDTH HANGUL LETTER O
U+007F	<control> DELETE	U+FFCD	HALFWIDTH HANGUL LETTER WA
U+0080	<control>	U+FFCE	HALFWIDTH HANGUL LETTER WAE
U+0081	<control>	U+FFCF	HALFWIDTH HANGUL LETTER OE
U+0082	<control> BREAK PERMITTED HERE	U+FFD2	HALFWIDTH HANGUL LETTER YO
U+0083	<control> NO BREAK HERE	U+FFD3	HALFWIDTH HANGUL LETTER U
U+0084	<control>	U+FFD4	HALFWIDTH HANGUL LETTER WEO
U+0085	<control> NEXT LINE (NEL)	U+FFD5	HALFWIDTH HANGUL LETTER WE
U+0086	<control> START OF SELECTED AREA	U+FFD6	HALFWIDTH HANGUL LETTER WI
U+0087	<control> END OF SELECTED AREA	U+FFD7	HALFWIDTH HANGUL LETTER YU
U+0088	<control> CHARACTER TABULATION SET	U+FFDA	HALFWIDTH HANGUL LETTER EU
U+0089	<control> CHARACTER TABULATION WITH JUSTIFICATION	U+FFDB	HALFWIDTH HANGUL LETTER YI
U+008A	<control> LINE TABULATION SET	U+FFDC	HALFWIDTH HANGUL LETTER I
U+008B	<control> PARTIAL LINE FORWARD	U+FFE8	HALFWIDTH FORMS LIGHT VERTICAL
U+008C	<control> PARTIAL LINE BACKWARD	U+FFE9	HALFWIDTH LEFTWARDS ARROW
U+008D	<control> REVERSE LINE FEED	U+FFEA	HALFWIDTH UPWARDS ARROW
U+008E	<control> SINGLE SHIFT TWO	U+FFEB	HALFWIDTH RIGHTWARDS ARROW

Code	Unicode Character Name	Code	Unicode Character Name
U+008F	<control> SINGLE SHIFT THREE	U+FFEC	HALFWIDTH DOWNWARDS ARROW
U+0090	<control> DEVICE CONTROL STRING	U+FFED	HALFWIDTH BLACK SQUARE
U+0091	<control> PRIVATE USE ONE	U+FFEE	HALFWIDTH WHITE CIRCLE

UNICODE Halfwidth to Fullwidth

UNICODE_TO_UNICODE_Fullwidth maps the halfwidth characters of the UNICODE server character set to their fullwidth equivalents.

Teradata Database uses this mapping when you translate UNICODE character strings using the TRANSLATE string function and specify UNICODE_TO_UNICODE_Fullwidth.

The UNICODE_TO_UNICODE_Fullwidth mapping table is described in a text file that is available for download. See [Character Set Files](#).

File Name	Title	Description
UNCDH2F.txt	UNICODE to UNICODE_Fullwidth	Maps the halfwidth characters of Unicode to the fullwidth characters of Unicode. Characters that are marked as HALFWIDTH are replaced with their unmarked form. Characters that are unmarked, but have a FULLWIDTH form, are mapped to that FULLWIDTH form. All other characters are unaffected.

UNICODE Fullwidth to Halfwidth

UNICODE_TO_UNICODE_Halfwidth maps the fullwidth characters of the UNICODE server character set to their halfwidth equivalents.

Teradata Database uses this mapping when you translate UNICODE character strings using the TRANSLATE string function and specify UNICODE_TO_UNICODE_Halfwidth.

The UNICODE_TO_UNICODE_Halfwidth mapping table is described in a text file that is available for download. See [Character Set Files](#).

File Name	Title	Description
UNCDF2H.txt	UNICODE to UNICODE_Halfwidth	Maps the fullwidth characters of Unicode to the halfwidth characters of Unicode. Characters that are marked as FULLWIDTH are replaced with their unmarked form. Characters that are unmarked, but have a HALFWIDTH form, are mapped to that HALFWIDTH form. All other characters are unaffected.

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2190	LEFTWARDS ARROW	U+FFE9	HALFWIDTH LEFTWARDS ARROW
U+2191	UPWARDS ARROW	U+FFEA	HALFWIDTH UPWARDS ARROW
U+2192	RIGHTWARDS ARROW	U+FFEB	HALFWIDTH RIGHTWARDS ARROW
U+2193	DOWNWARDS ARROW	U+FFEC	HALFWIDTH DOWNWARDS ARROW
U+2502	FORMS LIGHT VERTICAL	U+FFE8	HALFWIDTH FORMS LIGHT VERTICAL
U+25A0	BLACK SQUARE	U+FFED	HALFWIDTH BLACK SQUARE
U+25CB	WHITE CIRCLE	U+FFEE	HALFWIDTH WHITE CIRCLE
U+3000	IDEOGRAPHIC SPACE	U+0020	SPACE
U+3001	IDEOGRAPHIC COMMA	U+FF64	HALFWIDTH IDEOGRAPHIC COMMA
U+3002	IDEOGRAPHIC FULL STOP	U+FF61	HALFWIDTH IDEOGRAPHIC FULL STOP
U+300C	LEFT CORNER BRACKET	U+FF62	HALFWIDTH LEFT CORNER BRACKET
U+300D	RIGHT CORNER BRACKET	U+FF63	HALFWIDTH RIGHT CORNER BRACKET
U+3099	COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK	U+FF9E	HALFWIDTH KATAKANA VOICED SOUND MARK
U+309a	COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK	U+FF9F	HALFWIDTH KATAKANA SEMI- VOICED SOUND MARK
U+30A1	KATAKANA LETTER SMALL A	U+FF67	HALFWIDTH KATAKANA LETTER SMALL A
U+30A2	KATAKANA LETTER A	U+FF71	HALFWIDTH KATAKANA LETTER A
U+30A3	KATAKANA LETTER SMALL I	U+FF68	HALFWIDTH KATAKANA LETTER SMALL I
U+30A4	KATAKANA LETTER I	U+FF72	HALFWIDTH KATAKANA LETTER I
U+30A5	KATAKANA LETTER SMALL U	U+FF69	HALFWIDTH KATAKANA LETTER SMALL U
U+30A6	KATAKANA LETTER U	U+FF73	HALFWIDTH KATAKANA LETTER U
U+30A7	KATAKANA LETTER SMALL E	U+FF6A	HALFWIDTH KATAKANA LETTER SMALL E
U+30A8	KATAKANA LETTER E	U+FF74	HALFWIDTH KATAKANA LETTER E

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+30A9	KATAKANA LETTER SMALL O	U+FF6B	HALFWIDTH KATAKANA LETTER SMALL O
U+30AA	KATAKANA LETTER O	U+FF75	HALFWIDTH KATAKANA LETTER O
U+30AB	KATAKANA LETTER KA	U+FF76	HALFWIDTH KATAKANA LETTER KA
U+30AD	KATAKANA LETTER KI	U+FF77	HALFWIDTH KATAKANA LETTER KI
U+30AF	KATAKANA LETTER KU	U+FF78	HALFWIDTH KATAKANA LETTER KU
U+30B1	KATAKANA LETTER KE	U+FF79	HALFWIDTH KATAKANA LETTER KE
U+30B3	KATAKANA LETTER KO	U+FF7A	HALFWIDTH KATAKANA LETTER KO
U+30B5	KATAKANA LETTER SA	U+FF7B	HALFWIDTH KATAKANA LETTER SA
U+30B7	KATAKANA LETTER SI	U+FF7C	HALFWIDTH KATAKANA LETTER SI
U+30B9	KATAKANA LETTER SU	U+FF7D	HALFWIDTH KATAKANA LETTER SU
U+30BB	KATAKANA LETTER SE	U+FF7E	HALFWIDTH KATAKANA LETTER SE
U+30BD	KATAKANA LETTER SO	U+FF7F	HALFWIDTH KATAKANA LETTER SO
U+30BF	KATAKANA LETTER TA	U+FF80	HALFWIDTH KATAKANA LETTER TA
U+30C1	KATAKANA LETTER TI	U+FF81	HALFWIDTH KATAKANA LETTER TI
U+30C3	KATAKANA LETTER SMALL TU	U+FF6F	HALFWIDTH KATAKANA LETTER SMALL TU
U+30C4	KATAKANA LETTER TU	U+FF82	HALFWIDTH KATAKANA LETTER TU
U+30C6	KATAKANA LETTER TE	U+FF83	HALFWIDTH KATAKANA LETTER TE
U+30C8	KATAKANA LETTER TO	U+FF84	HALFWIDTH KATAKANA LETTER TO
U+30CA	KATAKANA LETTER NA	U+FF85	HALFWIDTH KATAKANA LETTER NA
U+30CB	KATAKANA LETTER NI	U+FF86	HALFWIDTH KATAKANA LETTER NI
U+30CC	KATAKANA LETTER NU	U+FF87	HALFWIDTH KATAKANA LETTER NU
U+30CD	KATAKANA LETTER NE	U+FF88	HALFWIDTH KATAKANA LETTER NE
U+30CE	KATAKANA LETTER NO	U+FF89	HALFWIDTH KATAKANA LETTER NO
U+30CF	KATAKANA LETTER HA	U+FF8A	HALFWIDTH KATAKANA LETTER HA
U+30D2	KATAKANA LETTER HI	U+FF8B	HALFWIDTH KATAKANA LETTER HI
U+30D5	KATAKANA LETTER HU	U+FF8C	HALFWIDTH KATAKANA LETTER HU
U+30D8	KATAKANA LETTER HE	U+FF8D	HALFWIDTH KATAKANA LETTER HE

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+30DB	KATAKANA LETTER HO	U+FF8E	HALFWIDTH KATAKANA LETTER HO
U+30DE	KATAKANA LETTER MA	U+FF8F	HALFWIDTH KATAKANA LETTER MA
U+30DF	KATAKANA LETTER MI	U+FF90	HALFWIDTH KATAKANA LETTER MI
U+30E0	KATAKANA LETTER MU	U+FF91	HALFWIDTH KATAKANA LETTER MU
U+30E1	KATAKANA LETTER ME	U+FF92	HALFWIDTH KATAKANA LETTER ME
U+30E2	KATAKANA LETTER MO	U+FF93	HALFWIDTH KATAKANA LETTER MO
U+30E3	KATAKANA LETTER SMALL YA	U+FF6C	HALFWIDTH KATAKANA LETTER SMALL YA
U+30E4	KATAKANA LETTER YA	U+FF94	HALFWIDTH KATAKANA LETTER YA
U+30E5	KATAKANA LETTER SMALL YU	U+FF6D	HALFWIDTH KATAKANA LETTER SMALL YU
U+30E6	KATAKANA LETTER YU	U+FF95	HALFWIDTH KATAKANA LETTER YU
U+30E7	KATAKANA LETTER SMALL YO	U+FF6E	HALFWIDTH KATAKANA LETTER SMALL YO
U+30E8	KATAKANA LETTER YO	U+FF96	HALFWIDTH KATAKANA LETTER YO
U+30E9	KATAKANA LETTER RA	U+FF97	HALFWIDTH KATAKANA LETTER RA
U+30EA	KATAKANA LETTER RI	U+FF98	HALFWIDTH KATAKANA LETTER RI
U+30EB	KATAKANA LETTER RU	U+FF99	HALFWIDTH KATAKANA LETTER RU
U+30EC	KATAKANA LETTER RE	U+FF9A	HALFWIDTH KATAKANA LETTER RE
U+30ED	KATAKANA LETTER RO	U+FF9B	HALFWIDTH KATAKANA LETTER RO
U+30EF	KATAKANA LETTER WA	U+FF9C	HALFWIDTH KATAKANA LETTER WA
U+30F2	KATAKANA LETTER WO	U+FF66	HALFWIDTH KATAKANA LETTER WO
U+30F3	KATAKANA LETTER N	U+FF9D	HALFWIDTH KATAKANA LETTER N
U+30FB	KATAKANA MIDDLE DOT	U+FF65	HALFWIDTH KATAKANA MIDDLE DOT
U+30FC	KATAKANA-HIRAGANA PROLONGED SOUND MARK	U+FF70	HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK
U+3131	HANGUL LETTER KIYEOK	U+FFA1	HALFWIDTH HANGUL LETTER KIYEOK
U+3132	HANGUL LETTER SSANGKIYEOK	U+FFA2	HALFWIDTH HANGUL LETTER SSANGKIYEOK

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+3133	HANGUL LETTER KIYEOK-SIOS	U+FFA3	HALFWIDTH HANGUL LETTER KIYEOK-SIOS
U+3134	HANGUL LETTER NIEUN	U+FFA4	HALFWIDTH HANGUL LETTER NIEUN
U+3135	HANGUL LETTER NIEUN-CIEUC	U+FFA5	HALFWIDTH HANGUL LETTER NIEUN-CIEUC
U+3136	HANGUL LETTER NIEUN-HIEUH	U+FFA6	HALFWIDTH HANGUL LETTER NIEUN-HIEUH
U+3137	HANGUL LETTER TKEUT	U+FFA7	HALFWIDTH HANGUL LETTER TKEUT
U+3138	HANGUL LETTER SSANGTKEUT	U+FFA8	HALFWIDTH HANGUL LETTER SSANGTKEUT
U+3139	HANGUL LETTER RIEUL	U+FFA9	HALFWIDTH HANGUL LETTER RIEUL
U+313A	HANGUL LETTER RIEUL-KIYEOK	U+FFAA	HALFWIDTH HANGUL LETTER RIEUL-KIYEOK
U+313B	HANGUL LETTER RIEUL-MIEUM	U+FFAB	HALFWIDTH HANGUL LETTER RIEUL-MIEUM
U+313C	HANGUL LETTER RIEUL-PIEUP	U+FFAC	HALFWIDTH HANGUL LETTER RIEUL-PIEUP
U+313D	HANGUL LETTER RIEUL-SIOS	U+FFAD	HALFWIDTH HANGUL LETTER RIEUL-SIOS
U+313E	HANGUL LETTER RIEUL-THIEUTH	U+FFAE	HALFWIDTH HANGUL LETTER RIEUL-THIEUTH
U+313F	HANGUL LETTER RIEUL-PHIEUPH	U+FFAF	HALFWIDTH HANGUL LETTER RIEUL-PHIEUPH
U+3140	HANGUL LETTER RIEUL-HIEUH	U+FFB0	HALFWIDTH HANGUL LETTER RIEUL-HIEUH
U+3141	HANGUL LETTER MIEUM	U+FFB1	HALFWIDTH HANGUL LETTER MIEUM
U+3142	HANGUL LETTER PIEUP	U+FFB2	HALFWIDTH HANGUL LETTER PIEUP
U+3143	HANGUL LETTER SSANGPIEUP	U+FFB3	HALFWIDTH HANGUL LETTER SSANGPIEUP
U+3144	HANGUL LETTER PIEUP-SIOS	U+FFB4	HALFWIDTH HANGUL LETTER PIEUP-SIOS
U+3145	HANGUL LETTER SIOS	U+FFB5	HALFWIDTH HANGUL LETTER SIOS

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+3146	HANGUL LETTER SSANGSIOS	U+FFB6	HALFWIDTH HANGUL LETTER SSANGSIOS
U+3147	HANGUL LETTER IEUNG	U+FFB7	HALFWIDTH HANGUL LETTER IEUNG
U+3148	HANGUL LETTER CIEUC	U+FFB8	HALFWIDTH HANGUL LETTER CIEUC
U+3149	HANGUL LETTER SSANGCIEUC	U+FFB9	HALFWIDTH HANGUL LETTER SSANGCIEUC
U+314A	HANGUL LETTER CHIEUCH	U+FFBA	HALFWIDTH HANGUL LETTER CHIEUCH
U+314B	HANGUL LETTER KHIEUKH	U+FFBB	HALFWIDTH HANGUL LETTER KHIEUKH
U+314C	HANGUL LETTER THIEUTH	U+FFBC	HALFWIDTH HANGUL LETTER THIEUTH
U+314D	HANGUL LETTER PHIEUPH	U+FFBD	HALFWIDTH HANGUL LETTER PHIEUPH
U+314E	HANGUL LETTER HIEUH	U+FFBE	HALFWIDTH HANGUL LETTER HIEUH
U+314F	HANGUL LETTER A	U+FFC2	HALFWIDTH HANGUL LETTER A
U+3150	HANGUL LETTER AE	U+FFC3	HALFWIDTH HANGUL LETTER AE
U+3151	HANGUL LETTER YA	U+FFC4	HALFWIDTH HANGUL LETTER YA
U+3152	HANGUL LETTER YAE	U+FFC5	HALFWIDTH HANGUL LETTER YAE
U+3153	HANGUL LETTER EO	U+FFC6	HALFWIDTH HANGUL LETTER EO
U+3154	HANGUL LETTER E	U+FFC7	HALFWIDTH HANGUL LETTER E
U+3155	HANGUL LETTER YEO	U+FFCA	HALFWIDTH HANGUL LETTER YEO
U+3156	HANGUL LETTER YE	U+FFCB	HALFWIDTH HANGUL LETTER YE
U+3157	HANGUL LETTER O	U+FFCC	HALFWIDTH HANGUL LETTER O
U+3158	HANGUL LETTER WA	U+FFCD	HALFWIDTH HANGUL LETTER WA
U+3159	HANGUL LETTER WAE	U+FFCE	HALFWIDTH HANGUL LETTER WAE
U+315A	HANGUL LETTER OE	U+FFCF	HALFWIDTH HANGUL LETTER OE
U+315B	HANGUL LETTER YO	U+FFD2	HALFWIDTH HANGUL LETTER YO
U+315C	HANGUL LETTER U	U+FFD3	HALFWIDTH HANGUL LETTER U

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+315D	HANGUL LETTER WEO	U+FFD4	HALFWIDTH HANGUL LETTER WEO
U+315E	HANGUL LETTER WE	U+FFD5	HALFWIDTH HANGUL LETTER WE
U+315F	HANGUL LETTER WI	U+FFD6	HALFWIDTH HANGUL LETTER WI
U+3160	HANGUL LETTER YU	U+FFD7	HALFWIDTH HANGUL LETTER YU
U+3161	HANGUL LETTER EU	U+FFDA	HALFWIDTH HANGUL LETTER EU
U+3162	HANGUL LETTER YI	U+FFDB	HALFWIDTH HANGUL LETTER YI
U+3163	HANGUL LETTER I	U+FFDC	HALFWIDTH HANGUL LETTER I
U+3164	HANGUL FILLER	U+FFA0	HALFWIDTH HANGUL FILLER
U+FF01	FULLWIDTH EXCLAMATION MARK	U+0021	EXCLAMATION MARK
U+FF02	FULLWIDTH QUOTATION MARK	U+0022	QUOTATION MARK
U+FF03	FULLWIDTH NUMBER SIGN	U+0023	NUMBER SIGN
U+FF04	FULLWIDTH DOLLAR SIGN	U+0024	DOLLAR SIGN
U+FF05	FULLWIDTH PERCENT SIGN	U+0025	PERCENT SIGN
U+FF06	FULLWIDTH AMPERSAND	U+0026	AMPERSAND
U+FF07	FULLWIDTH APOSTROPHE	U+0027	APOSTROPHE
U+FF08	FULLWIDTH LEFT PARENTHESIS	U+0028	LEFT PARENTHESIS
U+FF09	FULLWIDTH RIGHT PARENTHESIS	U+0029	RIGHT PARENTHESIS
U+FF0A	FULLWIDTH ASTERISK	U+002A	ASTERISK
U+FF0B	FULLWIDTH PLUS SIGN	U+002B	PLUS SIGN
U+FF0C	FULLWIDTH COMMA	U+002C	COMMA
U+FF0D	FULLWIDTH HYPHEN-MINUS	U+002D	HYPHEN-MINUS
U+FF0E	FULLWIDTH FULL STOP	U+002E	FULL STOP
U+FF0F	FULLWIDTH SOLIDUS	U+002F	SOLIDUS
U+FF10	FULLWIDTH DIGIT ZERO	U+0030	DIGIT ZERO
U+FF11	FULLWIDTH DIGIT ONE	U+0031	DIGIT ONE
U+FF12	FULLWIDTH DIGIT TWO	U+0032	DIGIT TWO
U+FF13	FULLWIDTH DIGIT THREE	U+0033	DIGIT THREE
U+FF14	FULLWIDTH DIGIT FOUR	U+0034	DIGIT FOUR

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF15	FULLWIDTH DIGIT FIVE	U+0035	DIGIT FIVE
U+FF16	FULLWIDTH DIGIT SIX	U+0036	DIGIT SIX
U+FF17	FULLWIDTH DIGIT SEVEN	U+0037	DIGIT SEVEN
U+FF18	FULLWIDTH DIGIT EIGHT	U+0038	DIGIT EIGHT
U+FF19	FULLWIDTH DIGIT NINE	U+0039	DIGIT NINE
U+FF1A	FULLWIDTH COLON	U+003A	COLON
U+FF1B	FULLWIDTH SEMICOLON	U+003B	SEMICOLON
U+FF1C	FULLWIDTH LESS-THAN SIGN	U+003C	LESS-THAN SIGN
U+FF1D	FULLWIDTH EQUALS SIGN	U+003D	EQUALS SIGN
U+FF1E	FULLWIDTH GREATER-THAN SIGN	U+003E	GREATER-THAN SIGN
U+FF1F	FULLWIDTH QUESTION MARK	U+003F	QUESTION MARK
U+FF20	FULLWIDTH COMMERCIAL AT	U+0040	COMMERCIAL AT
U+FF21	FULLWIDTH LATIN CAPITAL LETTER A	U+0041	LATIN CAPITAL LETTER A
U+FF22	FULLWIDTH LATIN CAPITAL LETTER B	U+0042	LATIN CAPITAL LETTER B
U+FF23	FULLWIDTH LATIN CAPITAL LETTER C	U+0043	LATIN CAPITAL LETTER C
U+FF24	FULLWIDTH LATIN CAPITAL LETTER D	U+0044	LATIN CAPITAL LETTER D
U+FF25	FULLWIDTH LATIN CAPITAL LETTER E	U+0045	LATIN CAPITAL LETTER E
U+FF26	FULLWIDTH LATIN CAPITAL LETTER F	U+0046	LATIN CAPITAL LETTER F
U+FF27	FULLWIDTH LATIN CAPITAL LETTER G	U+0047	LATIN CAPITAL LETTER G
U+FF28	FULLWIDTH LATIN CAPITAL LETTER H	U+0048	LATIN CAPITAL LETTER H
U+FF29	FULLWIDTH LATIN CAPITAL LETTER I	U+0049	LATIN CAPITAL LETTER I
U+FF2A	FULLWIDTH LATIN CAPITAL LETTER J	U+004A	LATIN CAPITAL LETTER J
U+FF2B	FULLWIDTH LATIN CAPITAL LETTER K	U+004B	LATIN CAPITAL LETTER K

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF2C	FULLWIDTH LATIN CAPITAL LETTER L	U+004C	LATIN CAPITAL LETTER L
U+FF2D	FULLWIDTH LATIN CAPITAL LETTER M	U+004D	LATIN CAPITAL LETTER M
U+FF2E	FULLWIDTH LATIN CAPITAL LETTER N	U+004E	LATIN CAPITAL LETTER N
U+FF2F	FULLWIDTH LATIN CAPITAL LETTER O	U+004F	LATIN CAPITAL LETTER O
U+FF30	FULLWIDTH LATIN CAPITAL LETTER P	U+0050	LATIN CAPITAL LETTER P
U+FF31	FULLWIDTH LATIN CAPITAL LETTER Q	U+0051	LATIN CAPITAL LETTER Q
U+FF32	FULLWIDTH LATIN CAPITAL LETTER R	U+0052	LATIN CAPITAL LETTER R
U+FF33	FULLWIDTH LATIN CAPITAL LETTER S	U+0053	LATIN CAPITAL LETTER S
U+FF34	FULLWIDTH LATIN CAPITAL LETTER T	U+0054	LATIN CAPITAL LETTER T
U+FF35	FULLWIDTH LATIN CAPITAL LETTER U	U+0055	LATIN CAPITAL LETTER U
U+FF36	FULLWIDTH LATIN CAPITAL LETTER V	U+0056	LATIN CAPITAL LETTER V
U+FF37	FULLWIDTH LATIN CAPITAL LETTER W	U+0057	LATIN CAPITAL LETTER W
U+FF38	FULLWIDTH LATIN CAPITAL LETTER X	U+0058	LATIN CAPITAL LETTER X
U+FF39	FULLWIDTH LATIN CAPITAL LETTER Y	U+0059	LATIN CAPITAL LETTER Y
U+FF3A	FULLWIDTH LATIN CAPITAL LETTER Z	U+005A	LATIN CAPITAL LETTER Z
U+FF3B	FULLWIDTH LEFT SQUARE BRACKET	U+005B	LEFT SQUARE BRACKET
U+FF3C	FULLWIDTH REVERSE SOLIDUS	U+005C	REVERSE SOLIDUS
U+FF3D	FULLWIDTH RIGHT SQUARE BRACKET	U+005D	RIGHT SQUARE BRACKET
U+FF3E	FULLWIDTH CIRCUMFLEX ACCENT	U+005E	CIRCUMFLEX ACCENT

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF3F	FULLWIDTH LOW LINE	U+005F	LOW LINE
U+FF40	FULLWIDTH GRAVE ACCENT	U+0060	GRAVE ACCENT
U+FF41	FULLWIDTH LATIN SMALL LETTER A	U+0061	LATIN SMALL LETTER A
U+FF42	FULLWIDTH LATIN SMALL LETTER B	U+0062	LATIN SMALL LETTER B
U+FF43	FULLWIDTH LATIN SMALL LETTER C	U+0063	LATIN SMALL LETTER C
U+FF44	FULLWIDTH LATIN SMALL LETTER D	U+0064	LATIN SMALL LETTER D
U+FF45	FULLWIDTH LATIN SMALL LETTER E	U+0065	LATIN SMALL LETTER E
U+FF46	FULLWIDTH LATIN SMALL LETTER F	U+0066	LATIN SMALL LETTER F
U+FF47	FULLWIDTH LATIN SMALL LETTER G	U+0067	LATIN SMALL LETTER G
U+FF48	FULLWIDTH LATIN SMALL LETTER H	U+0068	LATIN SMALL LETTER H
U+FF49	FULLWIDTH LATIN SMALL LETTER I	U+0069	LATIN SMALL LETTER I
U+FF4A	FULLWIDTH LATIN SMALL LETTER J	U+006A	LATIN SMALL LETTER J
U+FF4B	FULLWIDTH LATIN SMALL LETTER K	U+006B	LATIN SMALL LETTER K
U+FF4C	FULLWIDTH LATIN SMALL LETTER L	U+006C	LATIN SMALL LETTER L
U+FF4D	FULLWIDTH LATIN SMALL LETTER M	U+006D	LATIN SMALL LETTER M
U+FF4E	FULLWIDTH LATIN SMALL LETTER N	U+006E	LATIN SMALL LETTER N
U+FF4F	FULLWIDTH LATIN SMALL LETTER O	U+006F	LATIN SMALL LETTER O
U+FF50	FULLWIDTH LATIN SMALL LETTER P	U+0070	LATIN SMALL LETTER P
U+FF51	FULLWIDTH LATIN SMALL LETTER Q	U+0071	LATIN SMALL LETTER Q
U+FF52	FULLWIDTH LATIN SMALL LETTER R	U+0072	LATIN SMALL LETTER R
U+FF53	FULLWIDTH LATIN SMALL LETTER S	U+0073	LATIN SMALL LETTER S
U+FF54	FULLWIDTH LATIN SMALL LETTER T	U+0074	LATIN SMALL LETTER T
U+FF55	FULLWIDTH LATIN SMALL LETTER U	U+0075	LATIN SMALL LETTER U
U+FF56	FULLWIDTH LATIN SMALL LETTER V	U+0076	LATIN SMALL LETTER V
U+FF57	FULLWIDTH LATIN SMALL LETTER W	U+0077	LATIN SMALL LETTER W
U+FF58	FULLWIDTH LATIN SMALL LETTER X	U+0078	LATIN SMALL LETTER X
U+FF59	FULLWIDTH LATIN SMALL LETTER Y	U+0079	LATIN SMALL LETTER Y
U+FF5A	FULLWIDTH LATIN SMALL LETTER Z	U+007A	LATIN SMALL LETTER Z

UNICODE Fullwidth Input		UNICODE Halfwidth Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF5B	FULLWIDTH LEFT CURLY BRACKET	U+007B	LEFT CURLY BRACKET
U+FF5C	FULLWIDTH VERTICAL LINE	U+007C	VERTICAL LINE
U+FF5D	FULLWIDTH RIGHT CURLY BRACKET	U+007D	RIGHT CURLY BRACKET
U+FF5E	FULLWIDTH TILDE	U+007E	TILDE
U+FF5F	FULL WIDTH LEFT WHITE PARENTHESIS	U+2985	LEFT WHITE PARENTHESIS
U+FF60	FULL WIDTH RIGHT WHITE PARENTHESIS	U+2986	RIGHT WHITE PARENTHESIS
U+FFE0	FULLWIDTH CENT SIGN	U+00A2	CENT SIGN
U+FFE1	FULLWIDTH POUND SIGN	U+00A3	POUND SIGN
U+FFE2	FULLWIDTH NOT SIGN	U+00AC	NOT SIGN
U+FFE3	FULLWIDTH MACRON	U+00AF	MACRON
U+FFE4	FULLWIDTH BROKEN BAR	U+00A6	BROKEN BAR
U+FFE5	FULLWIDTH YEN SIGN	U+00A5	YEN SIGN
U+FFE6	FULLWIDTH WON SIGN	U+20A9	WON SIGN

UNICODE to LATIN

This section maps UNICODE characters to their LATIN equivalents. Teradata Database uses this mapping when you translate a UNICODE string using the TRANSLATE function and specify UNICODE_TO_LATIN.

UNICODE	LATIN	Unicode Character Name
U+0000	0x00	<control> NULL
U+0001	0x01	<control> START OF HEADING
U+0002	0x02	<control> START OF TEXT
U+0003	0x03	<control> END OF TEXT
U+0004	0x04	<control> END OF TRANSMISSION
U+0005	0x05	<control> ENQUIRY
U+0006	0x06	<control> ACKNOWLEDGE
U+0007	0x07	<control> BELL

UNICODE	LATIN	Unicode Character Name
U+0008	0x08	<control> BACKSPACE
U+0009	0x09	<control> CHARACTER TABULATION (horizontal tabulation)
U+000A	0x0A	<control> LINE FEED
U+000B	0x0B	<control> LINE TABULATION (vertical tabulation)
U+000C	0x0C	<control> FORM FEED
U+000D	0x0D	<control> CARRIAGE RETURN
U+000E	0x0E	<control> SHIFT OUT
U+000F	0x0F	<control> SHIFT IN
U+0010	0x10	<control> DATA LINK ESCAPE
U+0011	0x11	<control> DEVICE CONTROL ONE
U+0012	0x12	<control> DEVICE CONTROL TWO
U+0013	0x13	<control> DEVICE CONTROL THREE
U+0014	0x14	<control> DEVICE CONTROL FOUR
U+0015	0x15	<control> NEGATIVE ACKNOWLEDGE
U+0016	0x16	<control> SYNCHRONOUS IDLE
U+0017	0x17	<control> END OF TRANSMISSION BLOCK
U+0018	0x18	<control> CANCEL
U+0019	0x19	<control> END OF MEDIUM
U+001A	0x1A	<control> SUBSTITUTE
U+001B	0x1B	<control> ESCAPE
U+001C	0x1C	<control> INFORMATION SEPARATOR FOUR (file separator)
U+001D	0x1D	<control> INFORMATION SEPARATOR THREE (group separator)
U+001E	0x1E	<control> INFORMATION SEPARATOR TWO (record separator)
U+001F	0x1F	<control> INFORMATION SEPARATOR ONE (unit separator)
U+0020	0x20	SPACE
U+0021	0x21	EXCLAMATION MARK
U+0022	0x22	QUOTATION MARK
U+0023	0x23	NUMBER SIGN
U+0024	0x24	DOLLAR SIGN

UNICODE	LATIN	Unicode Character Name
U+0025	0x25	PERCENT SIGN
U+0026	0x26	AMPERSAND
U+0027	0x27	APOSTROPHE
U+0028	0x28	LEFT PARENTHESIS
U+0029	0x29	RIGHT PARENTHESIS
U+002A	0x2A	ASTERISK
U+002B	0x2B	PLUS SIGN
U+002C	0x2C	COMMA
U+002D	0x2D	HYPHEN-MINUS
U+002E	0x2E	FULL STOP
U+002F	0x2F	SOLIDUS
U+0030	0x30	DIGIT ZERO
U+0031	0x31	DIGIT ONE
U+0032	0x32	DIGIT TWO
U+0033	0x33	DIGIT THREE
U+0034	0x34	DIGIT FOUR
U+0035	0x35	DIGIT FIVE
U+0036	0x36	DIGIT SIX
U+0037	0x37	DIGIT SEVEN
U+0038	0x38	DIGIT EIGHT
U+0039	0x39	DIGIT NINE
U+003A	0x3A	COLON
U+003B	0x3B	SEMICOLON
U+003C	0x3C	LESS-THAN SIGN
U+003D	0x3D	EQUALS SIGN
U+003E	0x3E	GREATER-THAN SIGN
U+003F	0x3F	QUESTION MARK
U+0040	0x40	COMMERCIAL AT
U+0041	0x41	LATIN CAPITAL LETTER A

UNICODE	LATIN	Unicode Character Name
U+0042	0x42	LATIN CAPITAL LETTER B
U+0043	0x43	LATIN CAPITAL LETTER C
U+0044	0x44	LATIN CAPITAL LETTER D
U+0045	0x45	LATIN CAPITAL LETTER E
U+0046	0x46	LATIN CAPITAL LETTER F
U+0047	0x47	LATIN CAPITAL LETTER G
U+0048	0x48	LATIN CAPITAL LETTER H
U+0049	0x49	LATIN CAPITAL LETTER I
U+004A	0x4A	LATIN CAPITAL LETTER J
U+004B	0x4B	LATIN CAPITAL LETTER K
U+004C	0x4C	LATIN CAPITAL LETTER L
U+004D	0x4D	LATIN CAPITAL LETTER M
U+004E	0x4E	LATIN CAPITAL LETTER N
U+004F	0x4F	LATIN CAPITAL LETTER O
U+0050	0x50	LATIN CAPITAL LETTER P
U+0051	0x51	LATIN CAPITAL LETTER Q
U+0052	0x52	LATIN CAPITAL LETTER R
U+0053	0x53	LATIN CAPITAL LETTER S
U+0054	0x54	LATIN CAPITAL LETTER T
U+0055	0x55	LATIN CAPITAL LETTER U
U+0056	0x56	LATIN CAPITAL LETTER V
U+0057	0x57	LATIN CAPITAL LETTER W
U+0058	0x58	LATIN CAPITAL LETTER X
U+0059	0x59	LATIN CAPITAL LETTER Y
U+005A	0x5A	LATIN CAPITAL LETTER Z
U+005B	0x5B	LEFT SQUARE BRACKET
U+005C	0x5C	REVERSE SOLIDUS
U+005D	0x5D	RIGHT SQUARE BRACKET
U+005E	0x5E	CIRCUMFLEX ACCENT

UNICODE	LATIN	Unicode Character Name
U+005F	0x5F	LOW LINE
U+0060	0x60	GRAVE ACCENT
U+0061	0x61	LATIN SMALL LETTER A
U+0062	0x62	LATIN SMALL LETTER B
U+0063	0x63	LATIN SMALL LETTER C
U+0064	0x64	LATIN SMALL LETTER D
U+0065	0x65	LATIN SMALL LETTER E
U+0066	0x66	LATIN SMALL LETTER F
U+0067	0x67	LATIN SMALL LETTER G
U+0068	0x68	LATIN SMALL LETTER H
U+0069	0x69	LATIN SMALL LETTER I
U+006A	0x6A	LATIN SMALL LETTER J
U+006B	0x6B	LATIN SMALL LETTER K
U+006C	0x6C	LATIN SMALL LETTER L
U+006D	0x6D	LATIN SMALL LETTER M
U+006E	0x6E	LATIN SMALL LETTER N
U+006F	0x6F	LATIN SMALL LETTER O
U+0070	0x70	LATIN SMALL LETTER P
U+0071	0x71	LATIN SMALL LETTER Q
U+0072	0x72	LATIN SMALL LETTER R
U+0073	0x73	LATIN SMALL LETTER S
U+0074	0x74	LATIN SMALL LETTER T
U+0075	0x75	LATIN SMALL LETTER U
U+0076	0x76	LATIN SMALL LETTER V
U+0077	0x77	LATIN SMALL LETTER W
U+0078	0x78	LATIN SMALL LETTER X
U+0079	0x79	LATIN SMALL LETTER Y
U+007A	0x7A	LATIN SMALL LETTER Z
U+007B	0x7B	LEFT CURLY BRACKET

UNICODE	LATIN	Unicode Character Name
U+007C	0x7C	VERTICAL LINE
U+007D	0x7D	RIGHT CURLY BRACKET
U+007E	0x7E	TILDE
U+007F	0x7F	<control> DELETE
U+0086	0x86	<control> START OF SELECTED AREA
U+0087	0x87	<control> END OF SELECTED AREA
U+0088	0x88	<control> CHARACTER TABULATION SET
U+0089	0x89	<control> CHARACTER TABULATION WITH JUSTIFICATION
U+008A	0x8A	<control> LINE TABULATION SET
U+008B	0x8B	<control> PARTIAL LINE FORWARD
U+008C	0x8C	<control> PARTIAL LINE BACKWARD
U+008D	0x8D	<control> REVERSE LINE FEED
U+008E	0x8E	<control> SINGLE SHIFT TWO
U+008F	0x8F	<control> SINGLE SHIFT THREE
U+0090	0x90	<control> DEVICE CONTROL STRING
U+0091	0x91	<control> PRIVATE USE ONE
U+0092	0x92	<control> PRIVATE USE TWO
U+0093	0x93	<control> SET TRANSMIT RATE
U+0094	0x94	<control> CANCEL CHARACTER
U+0095	0x95	<control> MESSAGE WAITING
U+0096	0x96	<control> START OF GUARDED AREA
U+0097	0x97	<control> END OF GUARDED AREA
U+0098	0x98	<control> START OF STRING
U+0099	0x99	<control>
U+009A	0x9A	<control> SINGLE CHARACTER INTRODUCER
U+009B	0x9B	<control> CONTROL SEQUENCE INTRODUCER
U+009C	0x9C	<control> STRING TERMINATOR
U+009D	0x9D	<control> OPERATING SYSTEM COMMAND
U+009E	0x9E	<control> PRIVACY MESSAGE

UNICODE	LATIN	Unicode Character Name
U+009F	0x9F	<control> APPLICATION PROGRAM COMMAND
U+00A0	0x80	NO-BREAK SPACE
U+00A1	0xA1	INVERTED EXCLAMATION MARK
U+00A2	0xA2	CENT SIGN
U+00A3	0xA3	POUND SIGN
U+00A4	0xA4	CURRENCY SIGN
U+00A5	0xA5	YEN SIGN
U+00A6	0xA6	BROKEN BAR
U+00A7	0xA7	SECTION SIGN
U+00A8	0x81	DIAERESIS
U+00A9	0xA9	COPYRIGHT SIGN
U+00AA	0xAA	FEMININE ORDINAL INDICATOR
U+00AB	0xAB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
U+00AC	0xAC	NOT SIGN
U+00AD	0xAD	SOFT HYPHEN
U+00AE	0xAE	REGISTERED SIGN
U+00AF	0xAF	MACRON
U+00B0	0xB0	DEGREE SIGN
U+00B1	0xB1	PLUS-MINUS SIGN
U+00B2	0xB2	SUPERSCRIP TWO
U+00B3	0xB3	SUPERSCRIP THREE
U+00B4	0x82	ACUTE ACCENT
U+00B5	0xB5	MICRO SIGN
U+00B6	0xB6	PILCROW SIGN
U+00B7	0xB7	MIDDLE DOT
U+00B8	0x83	CEDILLA
U+00B9	0xB9	SUPERSCRIP ONE
U+00BA	0xBA	MASCULINE ORDINAL INDICATOR
U+00BB	0xBB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK

UNICODE	LATIN	Unicode Character Name
U+00BC	0xBC	VULGAR FRACTION ONE QUARTER
U+00BD	0xBD	VULGAR FRACTION ONE HALF
U+00BE	0xBE	VULGAR FRACTION THREE QUARTERS
U+00BF	0xBF	INVERTED QUESTION MARK
U+00C0	0xC0	LATIN CAPITAL LETTER A WITH GRAVE
U+00C1	0xC1	LATIN CAPITAL LETTER A WITH ACUTE
U+00C2	0xC2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
U+00C3	0xC3	LATIN CAPITAL LETTER A WITH TILDE
U+00C4	0xC4	LATIN CAPITAL LETTER A WITH DIAERESIS
U+00C5	0xC5	LATIN CAPITAL LETTER A WITH RING ABOVE
U+00C6	0xC6	LATIN CAPITAL LETTER AE
U+00C7	0xC7	LATIN CAPITAL LETTER C WITH CEDILLA
U+00C8	0xC8	LATIN CAPITAL LETTER E WITH GRAVE
U+00C9	0xC9	LATIN CAPITAL LETTER E WITH ACUTE
U+00CA	0xCA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
U+00CB	0xCB	LATIN CAPITAL LETTER E WITH DIAERESIS
U+00CC	0xCC	LATIN CAPITAL LETTER I WITH GRAVE
U+00CD	0xCD	LATIN CAPITAL LETTER I WITH ACUTE
U+00CE	0xCE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
U+00CF	0xCF	LATIN CAPITAL LETTER I WITH DIAERESIS
U+00D0	0xF0	LATIN CAPITAL LETTER ETH
U+00D1	0xD1	LATIN CAPITAL LETTER N WITH TILDE
U+00D2	0xD2	LATIN CAPITAL LETTER O WITH GRAVE
U+00D3	0xD3	LATIN CAPITAL LETTER O WITH ACUTE
U+00D4	0xD4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
U+00D5	0xD5	LATIN CAPITAL LETTER O WITH TILDE
U+00D6	0xD6	LATIN CAPITAL LETTER O WITH DIAERESIS
U+00D7	0x84	MULTIPLICATION SIGN
U+00D8	0xD8	LATIN CAPITAL LETTER O WITH STROKE

UNICODE	LATIN	Unicode Character Name
U+00D9	0xD9	LATIN CAPITAL LETTER U WITH GRAVE
U+00DA	0xDA	LATIN CAPITAL LETTER U WITH ACUTE
U+00DB	0xDB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
U+00DC	0xDC	LATIN CAPITAL LETTER U WITH DIAERESIS
U+00DD	0xB4	LATIN CAPITAL LETTER Y WITH ACUTE
U+00DE	0xFE	LATIN CAPITAL LETTER THORN
U+00DF	0xDF	LATIN SMALL LETTER SHARP S
U+00E0	0xE0	LATIN SMALL LETTER A WITH GRAVE
U+00E1	0xE1	LATIN SMALL LETTER A WITH ACUTE
U+00E2	0xE2	LATIN SMALL LETTER A WITH CIRCUMFLEX
U+00E3	0xE3	LATIN SMALL LETTER A WITH TILDE
U+00E4	0xE4	LATIN SMALL LETTER A WITH DIAERESIS
U+00E5	0xE5	LATIN SMALL LETTER A WITH RING ABOVE
U+00E6	0xE6	LATIN SMALL LETTER AE
U+00E7	0xE7	LATIN SMALL LETTER C WITH CEDILLA
U+00E8	0xE8	LATIN SMALL LETTER E WITH GRAVE
U+00E9	0xE9	LATIN SMALL LETTER E WITH ACUTE
U+00EA	0xEA	LATIN SMALL LETTER E WITH CIRCUMFLEX
U+00EB	0xEB	LATIN SMALL LETTER E WITH DIAERESIS
U+00EC	0xEC	LATIN SMALL LETTER I WITH GRAVE
U+00ED	0xED	LATIN SMALL LETTER I WITH ACUTE
U+00EE	0xEE	LATIN SMALL LETTER I WITH CIRCUMFLEX
U+00EF	0xEF	LATIN SMALL LETTER I WITH DIAERESIS
U+00F0	0xD0	LATIN SMALL LETTER ETH
U+00F1	0xF1	LATIN SMALL LETTER N WITH TILDE
U+00F2	0xF2	LATIN SMALL LETTER O WITH GRAVE
U+00F3	0xF3	LATIN SMALL LETTER O WITH ACUTE
U+00F4	0xF4	LATIN SMALL LETTER O WITH CIRCUMFLEX
U+00F5	0xF5	LATIN SMALL LETTER O WITH TILDE

UNICODE	LATIN	Unicode Character Name
U+00F6	0xF6	LATIN SMALL LETTER O WITH DIAERESIS
U+00F7	0x85	DIVISION SIGN
U+00F8	0xF8	LATIN SMALL LETTER O WITH STROKE
U+00F9	0xF9	LATIN SMALL LETTER U WITH GRAVE
U+00FA	0xFA	LATIN SMALL LETTER U WITH ACUTE
U+00FB	0xFB	LATIN SMALL LETTER U WITH CIRCUMFLEX
U+00FC	0xFC	LATIN SMALL LETTER U WITH DIAERESIS
U+00FD	0xA4	LATIN SMALL LETTER Y WITH ACUTE
U+00FE	0xDE	LATIN SMALL LETTER THORN
U+00FF	0xFD	LATIN SMALL LETTER Y WITH DIAERESIS
U+0152	0xD7	LATIN CAPITAL LIGATURE OE
U+0153	0xF7	LATIN SMALL LIGATURE OE
U+0160	0xB8	LATIN CAPITAL LETTER S WITH CARON
U+0161	0xA0	LATIN SMALL LETTER S WITH CARON
U+0178	0xDD	LATIN CAPITAL LETTER Y WITH DIAERESIS
U+039C	0xB5	GREEK CAPITAL LETTER MU and MICRO SIGN
U+20AC	0xFF	EURO SIGN
U+FFFD	0x1A	REPLACEMENT CHARACTER & <control>

LATIN Uppercase to Lowercase

This section maps uppercase LATIN characters to the lowercase LATIN equivalents. Teradata Database uses this mapping when you use the LOWER function on LATIN character strings.

LATIN Uppercase Input		LATIN Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x41	LATIN CAPITAL LETTER A	0x6	LATIN SMALL LETTER A
0x42	LATIN CAPITAL LETTER B	0x6	LATIN SMALL LETTER B
0x43	LATIN CAPITAL LETTER C	0x6	LATIN SMALL LETTER C
0x44	LATIN CAPITAL LETTER D	0x6	LATIN SMALL LETTER D
0x45	LATIN CAPITAL LETTER E	0x65	LATIN SMALL LETTER E

LATIN Uppercase Input		LATIN Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x46	LATIN CAPITAL LETTER F	0x66	LATIN SMALL LETTER F
0x47	LATIN CAPITAL LETTER G	0x67	LATIN SMALL LETTER G
0x48	LATIN CAPITAL LETTER H	0x68	LATIN SMALL LETTER H
0x49	LATIN CAPITAL LETTER I	0x69	LATIN SMALL LETTER I
0x4A	LATIN CAPITAL LETTER J	0x6A	LATIN SMALL LETTER J
0x4B	LATIN CAPITAL LETTER K	0x6B	LATIN SMALL LETTER K
0x4C	LATIN CAPITAL LETTER L	0x6C	LATIN SMALL LETTER L
0x4D	LATIN CAPITAL LETTER M	0x6D	LATIN SMALL LETTER M
0x4E	LATIN CAPITAL LETTER N	0x6E	LATIN SMALL LETTER N
0x4F	LATIN CAPITAL LETTER O	0x6F	LATIN SMALL LETTER O
0x50	LATIN CAPITAL LETTER P	0x70	LATIN SMALL LETTER P
0x51	LATIN CAPITAL LETTER Q	0x71	LATIN SMALL LETTER Q
0x52	LATIN CAPITAL LETTER R	0x72	LATIN SMALL LETTER R
0x53	LATIN CAPITAL LETTER S	0x73	LATIN SMALL LETTER S
0x54	LATIN CAPITAL LETTER T	0x74	LATIN SMALL LETTER T
0x55	LATIN CAPITAL LETTER U	0x75	LATIN SMALL LETTER U
0x56	LATIN CAPITAL LETTER V	0x76	LATIN SMALL LETTER V
0x57	LATIN CAPITAL LETTER W	0x77	LATIN SMALL LETTER W
0x58	LATIN CAPITAL LETTER X	0x78	LATIN SMALL LETTER X
0x59	LATIN CAPITAL LETTER Y	0x79	LATIN SMALL LETTER Y
0x5A	LATIN CAPITAL LETTER Z	0x7A	LATIN SMALL LETTER Z
0xB8	LATIN CAPITAL LETTER S WITH CARON	0xA0	LATIN SMALL LETTER S WITH CARON
0xB4	LATIN CAPITAL LETTER Y WITH ACUTE	0xA4	LATIN SMALL LETTER Y WITH ACUTE
0xC0	LATIN CAPITAL LETTER A WITH GRAVE	0xE0	LATIN SMALL LETTER A WITH GRAVE
0xC1	LATIN CAPITAL LETTER A WITH ACUTE	0xE1	LATIN SMALL LETTER A WITH ACUTE
0xC2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX	0xE2	LATIN SMALL LETTER A WITH CIRCUMFLEX
0xC3	LATIN CAPITAL LETTER A WITH TILDE	0xE3	LATIN SMALL LETTER A WITH TILDE

LATIN Uppercase Input		LATIN Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0xC4	LATIN CAPITAL LETTER A WITH DIAERESIS	0xE4	LATIN SMALL LETTER A WITH DIAERESIS
0xC5	LATIN CAPITAL LETTER A WITH RING ABOVE	0xE5	LATIN SMALL LETTER A WITH RING ABOVE
0xC6	LATIN CAPITAL LETTER AE	0xE6	LATIN SMALL LETTER AE
0xC7	LATIN CAPITAL LETTER C WITH CEDILLA	0xE7	LATIN SMALL LETTER C WITH CEDILLA
0xC8	LATIN CAPITAL LETTER E WITH GRAVE	0xE8	LATIN SMALL LETTER E WITH GRAVE
0xC9	LATIN CAPITAL LETTER E WITH ACUTE	0xE9	LATIN SMALL LETTER E WITH ACUTE
0xCA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX	0xEA	LATIN SMALL LETTER E WITH CIRCUMFLEX
0xCB	LATIN CAPITAL LETTER E WITH DIAERESIS	0xEB	LATIN SMALL LETTER E WITH DIAERESIS
0xCC	LATIN CAPITAL LETTER I WITH GRAVE	0xEC	LATIN SMALL LETTER I WITH GRAVE
0xCD	LATIN CAPITAL LETTER I WITH ACUTE	0xED	LATIN SMALL LETTER I WITH ACUTE
0xCE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX	0xEE	LATIN SMALL LETTER I WITH CIRCUMFLEX
0xCF	LATIN CAPITAL LETTER I WITH DIAERESIS	0xEF	LATIN SMALL LETTER I WITH DIAERESIS
0xD1	LATIN CAPITAL LETTER N WITH TILDE	0xF1	LATIN SMALL LETTER N WITH TILDE
0xD2	LATIN CAPITAL LETTER O WITH GRAVE	0xF2	LATIN SMALL LETTER O WITH GRAVE
0xD3	LATIN CAPITAL LETTER O WITH ACUTE	0xF3	LATIN SMALL LETTER O WITH ACUTE
0xD4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX	0xF4	LATIN SMALL LETTER O WITH CIRCUMFLEX
0xD5	LATIN CAPITAL LETTER O WITH TILDE	0xF5	LATIN SMALL LETTER O WITH TILDE
0xD6	LATIN CAPITAL LETTER O WITH DIAERESIS	0xF6	LATIN SMALL LETTER O WITH DIAERESIS
0xD7	LATIN CAPITAL LIGATURE OE	0xF7	LATIN SMALL LIGATURE OE
0xD8	LATIN CAPITAL LETTER O WITH STROKE	0xF8	LATIN SMALL LETTER O WITH STROKE
0xD9	LATIN CAPITAL LETTER U WITH GRAVE	0xF9	LATIN SMALL LETTER U WITH GRAVE
0xDA	LATIN CAPITAL LETTER U WITH ACUTE	0xFA	LATIN SMALL LETTER U WITH ACUTE

LATIN Uppercase Input		LATIN Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0xDB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX	0xFB	LATIN SMALL LETTER U WITH CIRCUMFLEX
0xDC	LATIN CAPITAL LETTER U WITH DIAERESIS	0xFC	LATIN SMALL LETTER U WITH DIAERESIS
0xDD	LATIN CAPITAL LETTER Y WITH DIAERESIS	0xFD	LATIN SMALL LETTER Y WITH DIAERESIS
0xF0	LATIN CAPITAL LETTER ETH	0xD0	LATIN SMALL LETTER ETH
0xFE	LATIN CAPITAL LETTER THORN	0xDE	LATIN SMALL LETTER THORN

Multibyte KANJISJIS Lowercase to Uppercase

This section maps the lowercase multibyte character portion of KanjiSJIS to its uppercase multibyte character equivalent. Teradata Database uses this mapping when you use the UPPER function on character strings that are defined with the KANJISJIS server character set.

KanjiSJIS Lowercase Input		KanjiSJIS Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x8281	FULLWIDTH LATIN SMALL LETTER A	0x8260	FULLWIDTH LATIN CAPITAL LETTER A
0x8282	FULLWIDTH LATIN SMALL LETTER B	0x8261	FULLWIDTH LATIN CAPITAL LETTER B
0x8283	FULLWIDTH LATIN SMALL LETTER C	0x8262	FULLWIDTH LATIN CAPITAL LETTER C
0x8284	FULLWIDTH LATIN SMALL LETTER D	0x8263	FULLWIDTH LATIN CAPITAL LETTER D
0x8285	FULLWIDTH LATIN SMALL LETTER E	0x8264	FULLWIDTH LATIN CAPITAL LETTER E
0x8286	FULLWIDTH LATIN SMALL LETTER F	0x8265	FULLWIDTH LATIN CAPITAL LETTER F
0x8287	FULLWIDTH LATIN SMALL LETTER G	0x8266	FULLWIDTH LATIN CAPITAL LETTER G
0x8288	FULLWIDTH LATIN SMALL LETTER H	0x8267	FULLWIDTH LATIN CAPITAL LETTER H
0x8289	FULLWIDTH LATIN SMALL LETTER I	0x8268	FULLWIDTH LATIN CAPITAL LETTER I
0x828A	FULLWIDTH LATIN SMALL LETTER J	0x8269	FULLWIDTH LATIN CAPITAL LETTER J
0x828B	FULLWIDTH LATIN SMALL LETTER K	0x826A	FULLWIDTH LATIN CAPITAL LETTER K
0x828C	FULLWIDTH LATIN SMALL LETTER L	0x826B	FULLWIDTH LATIN CAPITAL LETTER L

KanjiSJIS Lowercase Input		KanjiSJIS Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x828D	FULLWIDTH LATIN SMALL LETTER M	0x826C	FULLWIDTH LATIN CAPITAL LETTER M
0x828E	FULLWIDTH LATIN SMALL LETTER N	0x826D	FULLWIDTH LATIN CAPITAL LETTER N
0x828F	FULLWIDTH LATIN SMALL LETTER O	0x826E	FULLWIDTH LATIN CAPITAL LETTER O
0x8290	FULLWIDTH LATIN SMALL LETTER P	0x826F	FULLWIDTH LATIN CAPITAL LETTER P
0x8291	FULLWIDTH LATIN SMALL LETTER Q	0x8270	FULLWIDTH LATIN CAPITAL LETTER Q
0x8292	FULLWIDTH LATIN SMALL LETTER R	0x8271	FULLWIDTH LATIN CAPITAL LETTER R
0x8293	FULLWIDTH LATIN SMALL LETTER S	0x8272	FULLWIDTH LATIN CAPITAL LETTER S
0x8294	FULLWIDTH LATIN SMALL LETTER T	0x8273	FULLWIDTH LATIN CAPITAL LETTER T
0x8295	FULLWIDTH LATIN SMALL LETTER U	0x8274	FULLWIDTH LATIN CAPITAL LETTER U
0x8296	FULLWIDTH LATIN SMALL LETTER V	0x8275	FULLWIDTH LATIN CAPITAL LETTER V
0x8297	FULLWIDTH LATIN SMALL LETTER W	0x8276	FULLWIDTH LATIN CAPITAL LETTER W
0x8298	FULLWIDTH LATIN SMALL LETTER X	0x8277	FULLWIDTH LATIN CAPITAL LETTER X
0x8299	FULLWIDTH LATIN SMALL LETTER Y	0x8278	FULLWIDTH LATIN CAPITAL LETTER Y
0x829A	FULLWIDTH LATIN SMALL LETTER Z	0x8279	FULLWIDTH LATIN CAPITAL LETTER Z
0x83BF	GREEK SMALL LETTER ALPHA	0x839F	GREEK CAPITAL LETTER ALPHA
0x83C0	GREEK SMALL LETTER BETA	0x83A0	GREEK CAPITAL LETTER BETA
0x83C1	GREEK SMALL LETTER GAMMA	0x83A1	GREEK CAPITAL LETTER GAMMA
0x83C2	GREEK SMALL LETTER DELTA	0x83A2	GREEK CAPITAL LETTER DELTA
0x83C3	GREEK SMALL LETTER EPSILON	0x83A3	GREEK CAPITAL LETTER EPSILON
0x83C4	GREEK SMALL LETTER ZETA	0x83A4	GREEK CAPITAL LETTER ZETA
0x83C5	GREEK SMALL LETTER ETA	0x83A5	GREEK CAPITAL LETTER ETA
0x83C6	GREEK SMALL LETTER THETA	0x83A6	GREEK CAPITAL LETTER THETA
0x83C7	GREEK SMALL LETTER IOTA	0x83A7	GREEK CAPITAL LETTER IOTA
0x83C8	GREEK SMALL LETTER KAPPA	0x83A8	GREEK CAPITAL LETTER KAPPA

KanjiSJIS Lowercase Input		KanjiSJIS Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x83C9	GREEK SMALL LETTER LAMDA	0x83A9	GREEK CAPITAL LETTER LAMDA
0x83CA	GREEK SMALL LETTER MU	0x83AA	GREEK CAPITAL LETTER MU
0x83CB	GREEK SMALL LETTER NU	0x83AB	GREEK CAPITAL LETTER NU
0x83CC	GREEK SMALL LETTER XI	0x83AC	GREEK CAPITAL LETTER XI
0x83CD	GREEK SMALL LETTER OMICRON	0x83AD	GREEK CAPITAL LETTER OMICRON
0x83CE	GREEK SMALL LETTER PI	0x83AE	GREEK CAPITAL LETTER PI
0x83CF	GREEK SMALL LETTER RHO	0x83AF	GREEK CAPITAL LETTER RHO
0x83D0	GREEK SMALL LETTER SIGMA	0x83B0	GREEK CAPITAL LETTER SIGMA
0x83D1	GREEK SMALL LETTER TAU	0x83B1	GREEK CAPITAL LETTER TAU
0x83D2	GREEK SMALL LETTER UPSILON	0x83B2	GREEK CAPITAL LETTER UPSILON
0x83D3	GREEK SMALL LETTER PHI	0x83B3	GREEK CAPITAL LETTER PHI
0x83D4	GREEK SMALL LETTER CHI	0x83B4	GREEK CAPITAL LETTER CHI
0x83D5	GREEK SMALL LETTER PSI	0x83B5	GREEK CAPITAL LETTER PSI
0x83D6	GREEK SMALL LETTER OMEGA	0x83B6	GREEK CAPITAL LETTER OMEGA
0x8470	CYRILLIC SMALL LETTER A	0x8440	CYRILLIC CAPITAL LETTER A
0x8471	CYRILLIC SMALL LETTER BE	0x8441	CYRILLIC CAPITAL LETTER BE
0x8472	CYRILLIC SMALL LETTER VE	0x8442	CYRILLIC CAPITAL LETTER VE
0x8473	CYRILLIC SMALL LETTER GHE	0x8443	CYRILLIC CAPITAL LETTER GHE
0x8474	CYRILLIC SMALL LETTER DE	0x8444	CYRILLIC CAPITAL LETTER DE
0x8475	CYRILLIC SMALL LETTER IE	0x8445	CYRILLIC CAPITAL LETTER IE
0x8476	CYRILLIC SMALL LETTER IO	0x8446	CYRILLIC CAPITAL LETTER IO
0x8477	CYRILLIC SMALL LETTER ZHE	0x8447	CYRILLIC CAPITAL LETTER ZHE
0x8478	CYRILLIC SMALL LETTER ZE	0x8448	CYRILLIC CAPITAL LETTER ZE
0x8479	CYRILLIC SMALL LETTER I	0x8449	CYRILLIC CAPITAL LETTER I
0x847A	CYRILLIC SMALL LETTER SHORT I	0x844A	CYRILLIC CAPITAL LETTER SHORT I
0x847B	CYRILLIC SMALL LETTER KA	0x844B	CYRILLIC CAPITAL LETTER KA
0x847C	CYRILLIC SMALL LETTER EL	0x844C	CYRILLIC CAPITAL LETTER EL

KanjiSJIS Lowercase Input		KanjiSJIS Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x847D	CYRILLIC SMALL LETTER EM	0x844D	CYRILLIC CAPITAL LETTER EM
0x847E	CYRILLIC SMALL LETTER EN	0x844E	CYRILLIC CAPITAL LETTER EN
0x8480	CYRILLIC SMALL LETTER O	0x844F	CYRILLIC CAPITAL LETTER O
0x8481	CYRILLIC SMALL LETTER PE	0x8450	CYRILLIC CAPITAL LETTER PE
0x8482	CYRILLIC SMALL LETTER ER	0x8451	CYRILLIC CAPITAL LETTER ER
0x8483	CYRILLIC SMALL LETTER ES	0x8452	CYRILLIC CAPITAL LETTER ES
0x8484	CYRILLIC SMALL LETTER TE	0x8453	CYRILLIC CAPITAL LETTER TE
0x8485	CYRILLIC SMALL LETTER U	0x8454	CYRILLIC CAPITAL LETTER U
0x8486	CYRILLIC SMALL LETTER EF	0x8455	CYRILLIC CAPITAL LETTER EF
0x8487	CYRILLIC SMALL LETTER HA	0x8456	CYRILLIC CAPITAL LETTER HA
0x8488	CYRILLIC SMALL LETTER TSE	0x8457	CYRILLIC CAPITAL LETTER TSE
0x8489	CYRILLIC SMALL LETTER CHE	0x8458	CYRILLIC CAPITAL LETTER CHE
0x848A	CYRILLIC SMALL LETTER SHA	0x8459	CYRILLIC CAPITAL LETTER SHA
0x848B	CYRILLIC SMALL LETTER SHCHA	0x845A	CYRILLIC CAPITAL LETTER SHCHA
0x848C	CYRILLIC SMALL LETTER HARD SIGN	0x845B	CYRILLIC CAPITAL LETTER HARD SIGN
0x848D	CYRILLIC SMALL LETTER YERU	0x845C	CYRILLIC CAPITAL LETTER YERU
0x848E	CYRILLIC SMALL LETTER SOFT SIGN	0x845D	CYRILLIC CAPITAL LETTER SOFT SIGN
0x848F	CYRILLIC SMALL LETTER E	0x845E	CYRILLIC CAPITAL LETTER E
0x8490	CYRILLIC SMALL LETTER YU	0x845F	CYRILLIC CAPITAL LETTER YU
0x8491	CYRILLIC SMALL LETTER YA	0x8460	CYRILLIC CAPITAL LETTER YA

Single-Byte KANJISJIS Lowercase to Uppercase

This section maps the lowercase single-byte portion of the KANJISJIS server character set to its uppercase single-byte equivalent. Teradata Database uses this mapping when you use the UPPER function on character strings defined with the KANJISJIS server character set.

KANJIISJIS Lowercase Input		KANJIISJIS Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
0x61	LATIN SMALL LETTER A	0x41	LATIN CAPITAL LETTER A
0x62	LATIN SMALL LETTER B	0x42	LATIN CAPITAL LETTER B
0x63	LATIN SMALL LETTER C	0x43	LATIN CAPITAL LETTER C
0x64	LATIN SMALL LETTER D	0x44	LATIN CAPITAL LETTER D
0x65	LATIN SMALL LETTER E	0x45	LATIN CAPITAL LETTER E
0x66	LATIN SMALL LETTER F	0x46	LATIN CAPITAL LETTER F
0x67	LATIN SMALL LETTER G	0x47	LATIN CAPITAL LETTER G
0x68	LATIN SMALL LETTER H	0x48	LATIN CAPITAL LETTER H
0x69	LATIN SMALL LETTER I	0x49	LATIN CAPITAL LETTER I
0x6A	LATIN SMALL LETTER J	0x4A	LATIN CAPITAL LETTER J
0x6B	LATIN SMALL LETTER K	0x4B	LATIN CAPITAL LETTER K
0x6C	LATIN SMALL LETTER L	0x4C	LATIN CAPITAL LETTER L
0x6D	LATIN SMALL LETTER M	0x4D	LATIN CAPITAL LETTER M
0x6E	LATIN SMALL LETTER N	0x4E	LATIN CAPITAL LETTER N
0x6F	LATIN SMALL LETTER O	0x4F	LATIN CAPITAL LETTER O
0x70	LATIN SMALL LETTER P	0x50	LATIN CAPITAL LETTER P
0x71	LATIN SMALL LETTER Q	0x51	LATIN CAPITAL LETTER Q
0x72	LATIN SMALL LETTER R	0x52	LATIN CAPITAL LETTER R
0x73	LATIN SMALL LETTER S	0x53	LATIN CAPITAL LETTER S
0x74	LATIN SMALL LETTER T	0x54	LATIN CAPITAL LETTER T
0x75	LATIN SMALL LETTER U	0x55	LATIN CAPITAL LETTER U
0x76	LATIN SMALL LETTER V	0x56	LATIN CAPITAL LETTER V
0x77	LATIN SMALL LETTER W	0x57	LATIN CAPITAL LETTER W
0x78	LATIN SMALL LETTER X	0x58	LATIN CAPITAL LETTER X
0x79	LATIN SMALL LETTER Y	0x59	LATIN CAPITAL LETTER Y
0x7A	LATIN SMALL LETTER Z	0x5A	LATIN CAPITAL LETTER Z

UNICODE Lowercase to Uppercase

The following table shows the mapping between UNICODE lowercase characters and their uppercase equivalents. Teradata Database uses this mapping when a user calls the UPPER function on character strings defined with the UNICODE server character set.

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0061	LATIN SMALL LETTER A	U+0041	LATIN CAPITAL LETTER A
U+0062	LATIN SMALL LETTER B	U+0042	LATIN CAPITAL LETTER B
U+0063	LATIN SMALL LETTER C	U+0043	LATIN CAPITAL LETTER C
U+0064	LATIN SMALL LETTER D	U+0044	LATIN CAPITAL LETTER D
U+0065	LATIN SMALL LETTER E	U+0045	LATIN CAPITAL LETTER E
U+0066	LATIN SMALL LETTER F	U+0046	LATIN CAPITAL LETTER F
U+0067	LATIN SMALL LETTER G	U+0047	LATIN CAPITAL LETTER G
U+0068	LATIN SMALL LETTER H	U+0048	LATIN CAPITAL LETTER H
U+0069	LATIN SMALL LETTER I	U+0049	LATIN CAPITAL LETTER I
U+006A	LATIN SMALL LETTER J	U+004A	LATIN CAPITAL LETTER J
U+006B	LATIN SMALL LETTER K	U+004B	LATIN CAPITAL LETTER K
U+006C	LATIN SMALL LETTER L	U+004C	LATIN CAPITAL LETTER L
U+006D	LATIN SMALL LETTER M	U+004D	LATIN CAPITAL LETTER M
U+006E	LATIN SMALL LETTER N	U+004E	LATIN CAPITAL LETTER N
U+006F	LATIN SMALL LETTER O	U+004F	LATIN CAPITAL LETTER O
U+0070	LATIN SMALL LETTER P	U+0050	LATIN CAPITAL LETTER P
U+0071	LATIN SMALL LETTER Q	U+0051	LATIN CAPITAL LETTER Q
U+0072	LATIN SMALL LETTER R	U+0052	LATIN CAPITAL LETTER R
U+0073	LATIN SMALL LETTER S	U+0053	LATIN CAPITAL LETTER S
U+0074	LATIN SMALL LETTER T	U+0054	LATIN CAPITAL LETTER T
U+0075	LATIN SMALL LETTER U	U+0055	LATIN CAPITAL LETTER U
U+0076	LATIN SMALL LETTER V	U+0056	LATIN CAPITAL LETTER V
U+0077	LATIN SMALL LETTER W	U+0057	LATIN CAPITAL LETTER W
U+0078	LATIN SMALL LETTER X	U+0058	LATIN CAPITAL LETTER X

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0079	LATIN SMALL LETTER Y	U+0059	LATIN CAPITAL LETTER Y
U+007A	LATIN SMALL LETTER Z	U+005A	LATIN CAPITAL LETTER Z
U+00B5	MICRO SIGN	U+039C	GREEK CAPITAL LETTER MU
U+00E0	LATIN SMALL LETTER A WITH GRAVE	U+00C0	LATIN CAPITAL LETTER A WITH GRAVE
U+00E1	LATIN SMALL LETTER A WITH ACUTE	U+00C1	LATIN CAPITAL LETTER A WITH ACUTE
U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX	U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
U+00E3	LATIN SMALL LETTER A WITH TILDE	U+00C3	LATIN CAPITAL LETTER A WITH TILDE
U+00E4	LATIN SMALL LETTER A WITH DIAERESIS	U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS
U+00E5	LATIN SMALL LETTER A WITH RING ABOVE	U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE
U+00E6	LATIN SMALL LETTER AE	U+00C6	LATIN CAPITAL LETTER AE
U+00E7	LATIN SMALL LETTER C WITH CEDILLA	U+00C7	LATIN CAPITAL LETTER C WITH CEDILLA
U+00E8	LATIN SMALL LETTER E WITH GRAVE	U+00C8	LATIN CAPITAL LETTER E WITH GRAVE
U+00E9	LATIN SMALL LETTER E WITH ACUTE	U+00C9	LATIN CAPITAL LETTER E WITH ACUTE
U+00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX	U+00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
U+00EB	LATIN SMALL LETTER E WITH DIAERESIS	U+00CB	LATIN CAPITAL LETTER E WITH DIAERESIS
U+00EC	LATIN SMALL LETTER I WITH GRAVE	U+00CC	LATIN CAPITAL LETTER I WITH GRAVE
U+00ED	LATIN SMALL LETTER I WITH ACUTE	U+00CD	LATIN CAPITAL LETTER I WITH ACUTE
U+00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX	U+00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
U+00EF	LATIN SMALL LETTER I WITH DIAERESIS	U+00CF	LATIN CAPITAL LETTER I WITH DIAERESIS
U+00F0	LATIN SMALL LETTER ETH	U+00D0	LATIN CAPITAL LETTER ETH

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+00F1	LATIN SMALL LETTER N WITH TILDE	U+00D1	LATIN CAPITAL LETTER N WITH TILDE
U+00F2	LATIN SMALL LETTER O WITH GRAVE	U+00D2	LATIN CAPITAL LETTER O WITH GRAVE
U+00F3	LATIN SMALL LETTER O WITH ACUTE	U+00D3	LATIN CAPITAL LETTER O WITH ACUTE
U+00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX	U+00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
U+00F5	LATIN SMALL LETTER O WITH TILDE	U+00D5	LATIN CAPITAL LETTER O WITH TILDE
U+00F6	LATIN SMALL LETTER O WITH DIAERESIS	U+00D6	LATIN CAPITAL LETTER O WITH DIAERESIS
U+00F8	LATIN SMALL LETTER O WITH STROKE	U+00D8	LATIN CAPITAL LETTER O WITH STROKE
U+00F9	LATIN SMALL LETTER U WITH GRAVE	U+00D9	LATIN CAPITAL LETTER U WITH GRAVE
U+00FA	LATIN SMALL LETTER U WITH ACUTE	U+00DA	LATIN CAPITAL LETTER U WITH ACUTE
U+00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX	U+00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
U+00FC	LATIN SMALL LETTER U WITH DIAERESIS	U+00DC	LATIN CAPITAL LETTER U WITH DIAERESIS
U+00FD	LATIN SMALL LETTER Y WITH ACUTE	U+00DD	LATIN CAPITAL LETTER Y WITH ACUTE
U+00FE	LATIN SMALL LETTER THORN	U+00DE	LATIN CAPITAL LETTER THORN
U+00FF	LATIN SMALL LETTER Y WITH DIAERESIS	U+0178	LATIN CAPITAL LETTER Y WITH DIAERESIS
U+0101	LATIN SMALL LETTER A WITH MACRON	U+0100	LATIN CAPITAL LETTER A WITH MACRON
U+0103	LATIN SMALL LETTER A WITH BREVE	U+0102	LATIN CAPITAL LETTER A WITH BREVE
U+0105	LATIN SMALL LETTER A WITH OGONEK	U+0104	LATIN CAPITAL LETTER A WITH OGONEK
U+0107	LATIN SMALL LETTER C WITH ACUTE	U+0106	LATIN CAPITAL LETTER C WITH ACUTE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0109	LATIN SMALL LETTER C WITH CIRCUMFLEX	U+0108	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
U+010B	LATIN SMALL LETTER C WITH DOT ABOVE	U+010A	LATIN CAPITAL LETTER C WITH DOT ABOVE
U+010D	LATIN SMALL LETTER C WITH CARON	U+010C	LATIN CAPITAL LETTER C WITH CARON
U+010F	LATIN SMALL LETTER D WITH CARON	U+010E	LATIN CAPITAL LETTER D WITH CARON
U+0111	LATIN SMALL LETTER D WITH STROKE	U+0110	LATIN CAPITAL LETTER D WITH STROKE
U+0113	LATIN SMALL LETTER E WITH MACRON	U+0112	LATIN CAPITAL LETTER E WITH MACRON
U+0115	LATIN SMALL LETTER E WITH BREVE	U+0114	LATIN CAPITAL LETTER E WITH BREVE
U+0117	LATIN SMALL LETTER E WITH DOT ABOVE	U+0116	LATIN CAPITAL LETTER E WITH DOT ABOVE
U+0119	LATIN SMALL LETTER E WITH OGONEK	U+0118	LATIN CAPITAL LETTER E WITH OGONEK
U+011B	LATIN SMALL LETTER E WITH CARON	U+011A	LATIN CAPITAL LETTER E WITH CARON
U+011D	LATIN SMALL LETTER G WITH CIRCUMFLEX	U+011C	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
U+011F	LATIN SMALL LETTER G WITH BREVE	U+011E	LATIN CAPITAL LETTER G WITH BREVE
U+0121	LATIN SMALL LETTER G WITH DOT ABOVE	U+0120	LATIN CAPITAL LETTER G WITH DOT ABOVE
U+0123	LATIN SMALL LETTER G WITH CEDILLA	U+0122	LATIN CAPITAL LETTER G WITH CEDILLA
U+0125	LATIN SMALL LETTER H WITH CIRCUMFLEX	U+0124	LATIN CAPITAL LETTER H WITH CIRCUMFLEX
U+0127	LATIN SMALL LETTER H WITH STROKE	U+0126	LATIN CAPITAL LETTER H WITH STROKE
U+0129	LATIN SMALL LETTER I WITH TILDE	U+0128	LATIN CAPITAL LETTER I WITH TILDE
U+012B	LATIN SMALL LETTER I WITH MACRON	U+012A	LATIN CAPITAL LETTER I WITH MACRON

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+012D	LATIN SMALL LETTER I WITH BREVE	U+012C	LATIN CAPITAL LETTER I WITH BREVE
U+012F	LATIN SMALL LETTER I WITH OGONEK	U+012E	LATIN CAPITAL LETTER I WITH OGONEK
U+0131	LATIN SMALL LETTER DOTLESS I	U+0049	LATIN CAPITAL LETTER I
U+0133	LATIN SMALL LIGATURE IJ	U+0132	LATIN CAPITAL LIGATURE IJ
U+0135	LATIN SMALL LETTER J WITH CIRCUMFLEX	U+0134	LATIN CAPITAL LETTER J WITH CIRCUMFLEX
U+0137	LATIN SMALL LETTER K WITH CEDILLA	U+0136	LATIN CAPITAL LETTER K WITH CEDILLA
U+013A	LATIN SMALL LETTER L WITH ACUTE	U+0139	LATIN CAPITAL LETTER L WITH ACUTE
U+013C	LATIN SMALL LETTER L WITH CEDILLA	U+013B	LATIN CAPITAL LETTER L WITH CEDILLA
U+013E	LATIN SMALL LETTER L WITH CARON	U+013D	LATIN CAPITAL LETTER L WITH CARON
U+0140	LATIN SMALL LETTER L WITH MIDDLE DOT	U+013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT
U+0142	LATIN SMALL LETTER L WITH STROKE	U+0141	LATIN CAPITAL LETTER L WITH STROKE
U+0144	LATIN SMALL LETTER N WITH ACUTE	U+0143	LATIN CAPITAL LETTER N WITH ACUTE
U+0146	LATIN SMALL LETTER N WITH CEDILLA	U+0145	LATIN CAPITAL LETTER N WITH CEDILLA
U+0148	LATIN SMALL LETTER N WITH CARON	U+0147	LATIN CAPITAL LETTER N WITH CARON
U+014B	LATIN SMALL LETTER ENG	U+014A	LATIN CAPITAL LETTER ENG
U+014D	LATIN SMALL LETTER O WITH MACRON	U+014C	LATIN CAPITAL LETTER O WITH MACRON
U+014F	LATIN SMALL LETTER O WITH BREVE	U+014E	LATIN CAPITAL LETTER O WITH BREVE
U+0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE	U+0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
U+0153	LATIN SMALL LIGATURE OE	U+0152	LATIN CAPITAL LIGATURE OE
U+0155	LATIN SMALL LETTER R WITH ACUTE	U+0154	LATIN CAPITAL LETTER R WITH ACUTE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0157	LATIN SMALL LETTER R WITH CEDILLA	U+0156	LATIN CAPITAL LETTER R WITH CEDILLA
U+0159	LATIN SMALL LETTER R WITH CARON	U+0158	LATIN CAPITAL LETTER R WITH CARON
U+015B	LATIN SMALL LETTER S WITH ACUTE	U+015A	LATIN CAPITAL LETTER S WITH ACUTE
U+015D	LATIN SMALL LETTER S WITH CIRCUMFLEX	U+015C	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
U+015F	LATIN SMALL LETTER S WITH CEDILLA	U+015E	LATIN CAPITAL LETTER S WITH CEDILLA
U+0161	LATIN SMALL LETTER S WITH CARON	U+0160	LATIN CAPITAL LETTER S WITH CARON
U+0163	LATIN SMALL LETTER T WITH CEDILLA	U+0162	LATIN CAPITAL LETTER T WITH CEDILLA
U+0165	LATIN SMALL LETTER T WITH CARON	U+0164	LATIN CAPITAL LETTER T WITH CARON
U+0167	LATIN SMALL LETTER T WITH STROKE	U+0166	LATIN CAPITAL LETTER T WITH STROKE
U+0169	LATIN SMALL LETTER U WITH TILDE	U+0168	LATIN CAPITAL LETTER U WITH TILDE
U+016B	LATIN SMALL LETTER U WITH MACRON	U+016A	LATIN CAPITAL LETTER U WITH MACRON
U+016D	LATIN SMALL LETTER U WITH BREVE	U+016C	LATIN CAPITAL LETTER U WITH BREVE
U+016F	LATIN SMALL LETTER U WITH RING ABOVE	U+016E	LATIN CAPITAL LETTER U WITH RING ABOVE
U+0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE	U+0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
U+0173	LATIN SMALL LETTER U WITH OGONEK	U+0172	LATIN CAPITAL LETTER U WITH OGONEK
U+0175	LATIN SMALL LETTER W WITH CIRCUMFLEX	U+0174	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
U+0177	LATIN SMALL LETTER Y WITH CIRCUMFLEX	U+0176	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
U+017A	LATIN SMALL LETTER Z WITH ACUTE	U+0179	LATIN CAPITAL LETTER Z WITH ACUTE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+017C	LATIN SMALL LETTER Z WITH DOT ABOVE	U+017B	LATIN CAPITAL LETTER Z WITH DOT ABOVE
U+017E	LATIN SMALL LETTER Z WITH CARON	U+017D	LATIN CAPITAL LETTER Z WITH CARON
U+017F	LATIN SMALL LETTER LONG S	U+0053	LATIN CAPITAL LETTER S
U+0180	LATIN SMALL LETTER B WITH STROKE	U+0243	LATIN CAPITAL LETTER B WITH STROKE
U+0183	LATIN SMALL LETTER B WITH TOPBAR	U+0182	LATIN CAPITAL LETTER B WITH TOPBAR
U+0185	LATIN SMALL LETTER TONE SIX	U+0184	LATIN CAPITAL LETTER TONE SIX
U+0188	LATIN SMALL LETTER C WITH HOOK	U+0187	LATIN CAPITAL LETTER C WITH HOOK
U+018C	LATIN SMALL LETTER D WITH TOPBAR	U+018B	LATIN CAPITAL LETTER D WITH TOPBAR
U+0192	LATIN SMALL LETTER F WITH HOOK	U+0191	LATIN CAPITAL LETTER F WITH HOOK
U+0195	LATIN SMALL LETTER HV	U+01F6	LATIN CAPITAL LETTER HWAIR
U+0199	LATIN SMALL LETTER K WITH HOOK	U+0198	LATIN CAPITAL LETTER K WITH HOOK
U+019A	LATIN SMALL LETTER L WITH BAR	U+023D	LATIN CAPITAL LETTER L WITH BAR
U+019E	LATIN SMALL LETTER N WITH LONG RIGHT LEG	U+0220	LATIN CAPITAL LETTER N WITH LONG RIGHT LEG
U+01A1	LATIN SMALL LETTER O WITH HORN	U+01A0	LATIN CAPITAL LETTER O WITH HORN
U+01A3	LATIN SMALL LETTER OI	U+01A2	LATIN CAPITAL LETTER OI
U+01A5	LATIN SMALL LETTER P WITH HOOK	U+01A4	LATIN CAPITAL LETTER P WITH HOOK
U+01A8	LATIN SMALL LETTER TONE TWO	U+01A7	LATIN CAPITAL LETTER TONE TWO
U+01AD	LATIN SMALL LETTER T WITH HOOK	U+01AC	LATIN CAPITAL LETTER T WITH HOOK
U+01B0	LATIN SMALL LETTER U WITH HORN	U+01AF	LATIN CAPITAL LETTER U WITH HORN
U+01B4	LATIN SMALL LETTER Y WITH HOOK	U+01B3	LATIN CAPITAL LETTER Y WITH HOOK

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+01B6	LATIN SMALL LETTER Z WITH STROKE	U+01B5	LATIN CAPITAL LETTER Z WITH STROKE
U+01B9	LATIN SMALL LETTER EZH REVERSED	U+01B8	LATIN CAPITAL LETTER EZH REVERSED
U+01BD	LATIN SMALL LETTER TONE FIVE	U+01BC	LATIN CAPITAL LETTER TONE FIVE
U+01BF	LATIN LETTER WYNN	U+01F7	LATIN CAPITAL LETTER WYNN
U+01C5	LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON	U+01C4	LATIN CAPITAL LETTER DZ WITH CARON
U+01C6	LATIN SMALL LETTER DZ WITH CARON	U+01C4	LATIN CAPITAL LETTER DZ WITH CARON
U+01C8	LATIN CAPITAL LETTER L WITH SMALL LETTER J	U+01C7	LATIN CAPITAL LETTER LJ
U+01C9	LATIN SMALL LETTER LJ	U+01C7	LATIN CAPITAL LETTER LJ
U+01CB	LATIN CAPITAL LETTER N WITH SMALL LETTER J	U+01CA	LATIN CAPITAL LETTER NJ
U+01CC	LATIN SMALL LETTER NJ	U+01CA	LATIN CAPITAL LETTER NJ
U+01CE	LATIN SMALL LETTER A WITH CARON	U+01CD	LATIN CAPITAL LETTER A WITH CARON
U+01D0	LATIN SMALL LETTER I WITH CARON	U+01CF	LATIN CAPITAL LETTER I WITH CARON
U+01D2	LATIN SMALL LETTER O WITH CARON	U+01D1	LATIN CAPITAL LETTER O WITH CARON
U+01D4	LATIN SMALL LETTER U WITH CARON	U+01D3	LATIN CAPITAL LETTER U WITH CARON
U+01D6	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON	U+01D5	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON
U+01D8	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE	U+01D7	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE
U+01DA	LATIN SMALL LETTER U WITH DIAERESIS AND CARON	U+01D9	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON
U+01DC	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE	U+01DB	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE
U+01DD	LATIN SMALL LETTER TURNED E	U+018E	LATIN CAPITAL LETTER REVERSED E
U+01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON	U+01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+01E1	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON	U+01E0	LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON
U+01E3	LATIN SMALL LETTER AE WITH MACRON	U+01E2	LATIN CAPITAL LETTER AE WITH MACRON
U+01E5	LATIN SMALL LETTER G WITH STROKE	U+01E4	LATIN CAPITAL LETTER G WITH STROKE
U+01E7	LATIN SMALL LETTER G WITH CARON	U+01E6	LATIN CAPITAL LETTER G WITH CARON
U+01E9	LATIN SMALL LETTER K WITH CARON	U+01E8	LATIN CAPITAL LETTER K WITH CARON
U+01EB	LATIN SMALL LETTER O WITH OGONEK	U+01EA	LATIN CAPITAL LETTER O WITH OGONEK
U+01ED	LATIN SMALL LETTER O WITH OGONEK AND MACRON	U+01EC	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON
U+01EF	LATIN SMALL LETTER EZH WITH CARON	U+01EE	LATIN CAPITAL LETTER EZH WITH CARON
U+01F2	LATIN CAPITAL LETTER D WITH SMALL LETTER Z	U+01F1	LATIN CAPITAL LETTER DZ
U+01F3	LATIN SMALL LETTER DZ	U+01F1	LATIN CAPITAL LETTER DZ
U+01F5	LATIN SMALL LETTER G WITH ACUTE	U+01F4	LATIN CAPITAL LETTER G WITH ACUTE
U+01F9	LATIN SMALL LETTER N WITH GRAVE	U+01F8	LATIN CAPITAL LETTER N WITH GRAVE
U+01FB	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE	U+01FA	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE
U+01FD	LATIN SMALL LETTER AE WITH ACUTE	U+01FC	LATIN CAPITAL LETTER AE WITH ACUTE
U+01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE	U+01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE
U+0201	LATIN SMALL LETTER A WITH DOUBLE GRAVE	U+0200	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE
U+0203	LATIN SMALL LETTER A WITH INVERTED BREVE	U+0202	LATIN CAPITAL LETTER A WITH INVERTED BREVE
U+0205	LATIN SMALL LETTER E WITH DOUBLE GRAVE	U+0204	LATIN CAPITAL LETTER E WITH DOUBLE GRAVE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0207	LATIN SMALL LETTER E WITH INVERTED BREVE	U+0206	LATIN CAPITAL LETTER E WITH INVERTED BREVE
U+0209	LATIN SMALL LETTER I WITH DOUBLE GRAVE	U+0208	LATIN CAPITAL LETTER I WITH DOUBLE GRAVE
U+020B	LATIN SMALL LETTER I WITH INVERTED BREVE	U+020A	LATIN CAPITAL LETTER I WITH INVERTED BREVE
U+020D	LATIN SMALL LETTER O WITH DOUBLE GRAVE	U+020C	LATIN CAPITAL LETTER O WITH DOUBLE GRAVE
U+020F	LATIN SMALL LETTER O WITH INVERTED BREVE	U+020E	LATIN CAPITAL LETTER O WITH INVERTED BREVE
U+0211	LATIN SMALL LETTER R WITH DOUBLE GRAVE	U+0210	LATIN CAPITAL LETTER R WITH DOUBLE GRAVE
U+0213	LATIN SMALL LETTER R WITH INVERTED BREVE	U+0212	LATIN CAPITAL LETTER R WITH INVERTED BREVE
U+0215	LATIN SMALL LETTER U WITH DOUBLE GRAVE	U+0214	LATIN CAPITAL LETTER U WITH DOUBLE GRAVE
U+0217	LATIN SMALL LETTER U WITH INVERTED BREVE	U+0216	LATIN CAPITAL LETTER U WITH INVERTED BREVE
U+0219	LATIN SMALL LETTER S WITH COMMA BELOW	U+0218	LATIN CAPITAL LETTER S WITH COMMA BELOW
U+021B	LATIN SMALL LETTER T WITH COMMA BELOW	U+021A	LATIN CAPITAL LETTER T WITH COMMA BELOW
U+021D	LATIN SMALL LETTER YOGH	U+021C	LATIN CAPITAL LETTER YOGH
U+021F	LATIN SMALL LETTER H	U+021E	LATIN CAPITAL LETTER H WITH CARON
U+0223	LATIN SMALL LETTER OU	U+0222	LATIN CAPITAL LETTER OU
U+0225	LATIN SMALL LETTER Z WITH HOOK	U+0224	LATIN CAPITAL LETTER Z WITH HOOK
U+0227	LATIN SMALL LETTER A WITH DOT ABOVE	U+0226	LATIN CAPITAL LETTER A WITH DOT ABOVE
U+0229	LATIN SMALL LETTER E WITH CEDILLA	U+0228	LATIN CAPITAL LETTER E WITH CEDILLA
U+022B	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON	U+022A	LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON
U+022D	LATIN SMALL LETTER O WITH TILDE AND MACRON	U+022C	LATIN CAPITAL LETTER O WITH TILDE AND MACRON

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+022F	LATIN SMALL LETTER O WITH DOT ABOVE	U+022E	LATIN CAPITAL LETTER O WITH DOT ABOVE
U+0231	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON	U+0230	LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON
U+0233	LATIN SMALL LETTER Y WITH MACRON	U+0232	LATIN CAPITAL LETTER Y WITH MACRON
U+023C	LATIN SMALL LETTER C WITH STROKE	U+023B	LATIN CAPITAL LETTER C WITH STROKE
U+023F	LATIN SMALL LETTER S WITH SWASH TAIL	U+2C7E	LATIN CAPITAL LETTER S WITH SWASH TAIL
U+0240	LATIN SMALL LETTER Z WITH SWASH TAIL	U+2C7F	LATIN CAPITAL LETTER Z WITH SWASH TAIL
U+0242	LATIN SMALL LETTER GLOTTAL STOP	U+0241	LATIN CAPITAL LETTER GLOTTAL STOP
U+0247	LATIN SMALL LETTER E WITH STROKE	U+0246	LATIN CAPITAL LETTER E WITH STROKE
U+0249	LATIN SMALL LETTER J WITH STROKE	U+0248	LATIN CAPITAL LETTER J WITH STROKE
U+024B	LATIN SMALL LETTER Q WITH HOOK TAIL	U+024A	LATIN CAPITAL LETTER SMALL Q WITH HOOK TAIL
U+024D	LATIN SMALL LETTER R WITH STROKE	U+024C	LATIN CAPITAL LETTER R WITH STROKE
U+024F	LATIN SMALL LETTER Y WITH STROKE	U+024E	LATIN CAPITAL LETTER Y WITH STROKE
U+0250	LATIN SMALL LETTER TURNED A	U+2C6F	LATIN CAPITAL LETTER TURNED A
U+0251	LATIN SMALL LETTER ALPHA	U+C6D	LATIN CAPITAL LETTER ALPHA
U+0252	LATIN SMALL LETTER TURNED ALPHA	U+2C70	LATIN CAPITAL LETTER TURNED ALPHA
U+0253	LATIN SMALL LETTER B WITH HOOK	U+0181	LATIN CAPITAL LETTER B WITH HOOK
U+0254	LATIN SMALL LETTER OPEN O	U+0186	LATIN CAPITAL LETTER OPEN O
U+0256	LATIN SMALL LETTER D WITH TAIL	U+0189	LATIN CAPITAL LETTER AFRICAN D
U+0257	LATIN SMALL LETTER D WITH HOOK	U+018A	LATIN CAPITAL LETTER D WITH HOOK
U+01DD	LATIN SMALL LETTER TURNED E	U+018E	LATIN CAPITAL LETTER REVERSED E

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0259	LATIN SMALL LETTER SCHWA	U+018F	LATIN CAPITAL LETTER SCHWA
U+025B	LATIN SMALL LETTER OPEN E	U+0190	LATIN CAPITAL LETTER OPEN E
U+0260	LATIN SMALL LETTER G WITH HOOK	U+0193	LATIN CAPITAL LETTER G WITH HOOK
U+0263	LATIN SMALL LETTER GAMMA	U+0194	LATIN CAPITAL LETTER GAMMA
U+0265	LATIN SMALL LETTER TURNED H	U+A78D	LATIN CAPITAL LETTER TURNED H
U+0268	LATIN SMALL LETTER I WITH STROKE	U+0197	LATIN CAPITAL LETTER I WITH STROKE
U+0269	LATIN SMALL LETTER IOTA	U+0196	LATIN CAPITAL LETTER IOTA
U+026B	LATIN SMALL LETTER L WITH MIDDLE TILDE	U+2C62	LATIN CAPITAL LETTER L WITH MIDDLE TILDE
U+026F	LATIN SMALL LETTER TURNED M	U+019C	LATIN CAPITAL LETTER TURNED M
U+0271	LATIN SMALL LETTER M WITH HOOK	U+2C6E	LATIN CAPITAL LETTER M WITH HOOK
U+0272	LATIN SMALL LETTER N WITH LEFT HOOK	U+019D	LATIN CAPITAL LETTER N WITH LEFT HOOK
U+0275	LATIN SMALL LETTER BARRED O	U+019F	LATIN CAPITAL LETTER O WITH MIDDLE TILDE
U+027D	LATIN SMALL LETTER R WITH TAIL	U+2C64	LATIN CAPITAL LETTER R WITH TAIL
U+0283	LATIN SMALL LETTER ESH	U+01A9	LATIN CAPITAL LETTER ESH
U+0288	LATIN SMALL LETTER T WITH RETROFLEX HOOK	U+01AE	LATIN CAPITAL LETTER T WITH RETROFLEX HOOK
U+0289	LATIN SMALL LETTER U BAR	U+0244	LATIN CAPITAL LETTER U BAR
U+028A	LATIN SMALL LETTER UPSILON	U+01B1	LATIN CAPITAL LETTER UPSILON
U+028B	LATIN SMALL LETTER V WITH HOOK	U+01B2	LATIN CAPITAL LETTER V WITH HOOK
U+028C	LATIN SMALL LETTER TURNED V	U+0245	LATIN CAPITAL LETTER TURNED V
U+0292	LATIN SMALL LETTER EZH	U+01B7	LATIN CAPITAL LETTER EZH
U+0345	COMBINING GREEK YPOGEGRAMMENI	U+0399	GREEK CAPITAL LETTER IOTA
U+0371	GREEK SMALL LETTER HETA	U+0370	GREEK CAPITAL LETTER HETA
U+0373	GREEK SMALL LETTER ARCHAIC SAMPI	U+0372	GREEK CAPITAL LETTER ARCHAIC SAMPI

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0377	GREEK SMALL LETTER PAMPHYLIAN DIGAMMA	U+0376	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA
U+037B	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL	U+03FD	GREEK CAPITAL REVERSED LUNATE SIGMA SYMBOL
U+037C	GREEK SMALL DOTTED LUNATE SIGMA SYMBOL	U+03FE	GREEK CAPITAL DOTTED LUNATE SIGMA SYMBOL
U+037D	GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL	U+03FF	GREEK CAPITAL REVERSED DOTTED LUNATE SIGMA SYMBOL
U+03AC	GREEK SMALL LETTER ALPHA WITH TONOS	U+0386	GREEK CAPITAL LETTER ALPHA WITH TONOS
U+03AD	GREEK SMALL LETTER EPSILON WITH TONOS	U+0388	GREEK CAPITAL LETTER EPSILON WITH TONOS
U+03AE	GREEK SMALL LETTER ETA WITH TONOS	U+0389	GREEK CAPITAL LETTER ETA WITH TONOS
U+03AF	GREEK SMALL LETTER IOTA WITH TONOS	U+038A	GREEK CAPITAL LETTER IOTA WITH TONOS
U+03B1	GREEK SMALL LETTER ALPHA	U+0391	GREEK CAPITAL LETTER ALPHA
U+03B2	GREEK SMALL LETTER BETA	U+0392	GREEK CAPITAL LETTER BETA
U+03B3	GREEK SMALL LETTER GAMMA	U+0393	GREEK CAPITAL LETTER GAMMA
U+03B4	GREEK SMALL LETTER DELTA	U+0394	GREEK CAPITAL LETTER DELTA
U+03B5	GREEK SMALL LETTER EPSILON	U+0395	GREEK CAPITAL LETTER EPSILON
U+03B6	GREEK SMALL LETTER ZETA	U+0396	GREEK CAPITAL LETTER ZETA
U+03B7	GREEK SMALL LETTER ETA	U+0397	GREEK CAPITAL LETTER ETA
U+03B8	GREEK SMALL LETTER THETA	U+0398	GREEK CAPITAL LETTER THETA
U+03B9	GREEK SMALL LETTER IOTA	U+0399	GREEK CAPITAL LETTER IOTA
U+03BA	GREEK SMALL LETTER KAPPA	U+039A	GREEK CAPITAL LETTER KAPPA
U+03BB	GREEK SMALL LETTER LAMDA	U+039B	GREEK CAPITAL LETTER LAMDA
U+00B5	MICRO SIGN	U+039C	GREEK CAPITAL LETTER MU
U+03BC	GREEK SMALL LETTER MU	U+039C	GREEK CAPITAL LETTER MU
U+03BD	GREEK SMALL LETTER NU	U+039D	GREEK CAPITAL LETTER NU
U+03BE	GREEK SMALL LETTER XI	U+039E	GREEK CAPITAL LETTER XI
U+03BF	GREEK SMALL LETTER OMICRON	U+039F	GREEK CAPITAL LETTER OMICRON

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+03C0	GREEK SMALL LETTER PI	U+03A0	GREEK CAPITAL LETTER PI
U+03C1	GREEK SMALL LETTER RHO	U+03A1	GREEK CAPITAL LETTER RHO
U+03C2	GREEK SMALL LETTER FINAL SIGMA	U+03A3	GREEK CAPITAL LETTER SIGMA
U+03C3	GREEK SMALL LETTER SIGMA	U+03A3	GREEK CAPITAL LETTER SIGMA
U+03C4	GREEK SMALL LETTER TAU	U+03A4	GREEK CAPITAL LETTER TAU
U+03C5	GREEK SMALL LETTER UPSILON	U+03A5	GREEK CAPITAL LETTER UPSILON
U+03C6	GREEK SMALL LETTER PHI	U+03A6	GREEK CAPITAL LETTER PHI
U+03C7	GREEK SMALL LETTER CHI	U+03A7	GREEK CAPITAL LETTER CHI
U+03C8	GREEK SMALL LETTER PSI	U+03A8	GREEK CAPITAL LETTER PSI
U+03C9	GREEK SMALL LETTER OMEGA	U+03A9	GREEK CAPITAL LETTER OMEGA
U+03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA	U+03AA	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
U+03CB	GREEK SMALL LETTER UPSILON WITH DIALYTIKA	U+03AB	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
U+03CC	GREEK SMALL LETTER OMICRON WITH TONOS	U+038C	GREEK CAPITAL LETTER OMICRON WITH TONOS
U+03CD	GREEK SMALL LETTER UPSILON WITH TONOS	U+038E	GREEK CAPITAL LETTER UPSILON WITH TONOS
U+03CE	GREEK SMALL LETTER OMEGA WITH TONOS	U+038F	GREEK CAPITAL LETTER OMEGA WITH TONOS
U+03D0	GREEK BETA SYMBOL	U+0392	GREEK CAPITAL LETTER BETA
U+03D1	GREEK THETA SYMBOL	U+0398	GREEK CAPITAL LETTER THETA
U+03D5	GREEK PHI SYMBOL	U+03A6	GREEK CAPITAL LETTER PHI
U+03D6	GREEK PI SYMBOL	U+03A0	GREEK CAPITAL LETTER PI
U+03D7	GREEK KAI SYMBOL	U+03CF	GREEK CAPITAL KAI SYMBOL
U+03D9	GREEK SMALL LETTER ARCHAIC KOPPA	U+03D8	GREEK LETTER ARCHAIC KOPPA
U+03DB	GREEK SMALL LETTER STIGMA	U+03DA	GREEK LETTER STIGMA
U+03DD	GREEK SMALL LETTER DIGAMMA	U+03DC	GREEK LETTER DIGAMMA
U+03DF	GREEK SMALL LETTER KOPPA	U+03DE	GREEK LETTER KOPPA

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+03E1	GREEK SMALL LETTER SAMPI	U+03E0	GREEK LETTER SAMPI
U+03E3	COPTIC SMALL LETTER SHEI	U+03E2	COPTIC CAPITAL LETTER SHEI
U+03E5	COPTIC SMALL LETTER FEI	U+03E4	COPTIC CAPITAL LETTER FEI
U+03E7	COPTIC SMALL LETTER KHEI	U+03E6	COPTIC CAPITAL LETTER KHEI
U+03E9	COPTIC SMALL LETTER HORI	U+03E8	COPTIC CAPITAL LETTER HORI
U+03EB	COPTIC SMALL LETTER GANGIA	U+03EA	COPTIC CAPITAL LETTER GANGIA
U+03ED	COPTIC SMALL LETTER SHIMA	U+03EC	COPTIC CAPITAL LETTER SHIMA
U+03EF	COPTIC SMALL LETTER DEI	U+03EE	COPTIC CAPITAL LETTER DEI
U+03F0	GREEK KAPPA SYMBOL	U+039A	GREEK CAPITAL LETTER KAPPA
U+03F1	GREEK RHO SYMBOL	U+03A1	GREEK CAPITAL LETTER RHO
U+03F2	GREEK LUNATE SIGMA SYMBOL	U+03F9	GREEK CAPITAL LUNATE SIGMA SYMBOL
U+03F5	GREEK LUNATE EPSILON SYMBOL	U+0395	GREEK CAPITAL LETTER EPSILON
U+03F8	GREEK SMALL LETTER SHO	U+03F7	GREEK CAPITAL LETTER SHO
U+03FB	GREEK SMALL LETTER SAN	U+03FA	GREEK CAPITAL LETTER SAN
U+0430	CYRILLIC SMALL LETTER A	U+0410	CYRILLIC CAPITAL LETTER A
U+0431	CYRILLIC SMALL LETTER BE	U+0411	CYRILLIC CAPITAL LETTER BE
U+0432	CYRILLIC SMALL LETTER VE	U+0412	CYRILLIC CAPITAL LETTER VE
U+0433	CYRILLIC SMALL LETTER GHE	U+0413	CYRILLIC CAPITAL LETTER GHE
U+0434	CYRILLIC SMALL LETTER DE	U+0414	CYRILLIC CAPITAL LETTER DE
U+0435	CYRILLIC SMALL LETTER IE	U+0415	CYRILLIC CAPITAL LETTER IE
U+0436	CYRILLIC SMALL LETTER ZHE	U+0416	CYRILLIC CAPITAL LETTER ZHE
U+0437	CYRILLIC SMALL LETTER ZE	U+0417	CYRILLIC CAPITAL LETTER ZE
U+0438	CYRILLIC SMALL LETTER I	U+0418	CYRILLIC CAPITAL LETTER I
U+0439	CYRILLIC SMALL LETTER SHORT I	U+0419	CYRILLIC CAPITAL LETTER SHORT I
U+043A	CYRILLIC SMALL LETTER KA	U+041A	CYRILLIC CAPITAL LETTER KA
U+043B	CYRILLIC SMALL LETTER EL	U+041B	CYRILLIC CAPITAL LETTER EL
U+043C	CYRILLIC SMALL LETTER EM	U+041C	CYRILLIC CAPITAL LETTER EM

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+043D	CYRILLIC SMALL LETTER EN	U+041D	CYRILLIC CAPITAL LETTER EN
U+043E	CYRILLIC SMALL LETTER O	U+041E	CYRILLIC CAPITAL LETTER O
U+043F	CYRILLIC SMALL LETTER PE	U+041F	CYRILLIC CAPITAL LETTER PE
U+0440	CYRILLIC SMALL LETTER ER	U+0420	CYRILLIC CAPITAL LETTER ER
U+0441	CYRILLIC SMALL LETTER ES	U+0421	CYRILLIC CAPITAL LETTER ES
U+0442	CYRILLIC SMALL LETTER TE	U+0422	CYRILLIC CAPITAL LETTER TE
U+0443	CYRILLIC SMALL LETTER U	U+0423	CYRILLIC CAPITAL LETTER U
U+0444	CYRILLIC SMALL LETTER EF	U+0424	CYRILLIC CAPITAL LETTER EF
U+0445	CYRILLIC SMALL LETTER HA	U+0425	CYRILLIC CAPITAL LETTER HA
U+0446	CYRILLIC SMALL LETTER TSE	U+0426	CYRILLIC CAPITAL LETTER TSE
U+0447	CYRILLIC SMALL LETTER CHE	U+0427	CYRILLIC CAPITAL LETTER CHE
U+0448	CYRILLIC SMALL LETTER SHA	U+0428	CYRILLIC CAPITAL LETTER SHA
U+0449	CYRILLIC SMALL LETTER SHCHA	U+0429	CYRILLIC CAPITAL LETTER SHCHA
U+044A	CYRILLIC SMALL LETTER HARD SIGN	U+042A	CYRILLIC CAPITAL LETTER HARD SIGN
U+044B	CYRILLIC SMALL LETTER YERU	U+042B	CYRILLIC CAPITAL LETTER YERU
U+044C	CYRILLIC SMALL LETTER SOFT SIGN	U+042C	CYRILLIC CAPITAL LETTER SOFT SIGN
U+044D	CYRILLIC SMALL LETTER E	U+042D	CYRILLIC CAPITAL LETTER E
U+044E	CYRILLIC SMALL LETTER YU	U+042E	CYRILLIC CAPITAL LETTER YU
U+044F	CYRILLIC SMALL LETTER YA	U+042F	CYRILLIC CAPITAL LETTER YA
U+0450	CYRILLIC SMALL LETTER IE WITH GRAVE	U+0400	CYRILLIC CAPITAL LETTER IE WITH GRAVE
U+0451	CYRILLIC SMALL LETTER IO	U+0401	CYRILLIC CAPITAL LETTER IO
U+0452	CYRILLIC SMALL LETTER DJE	U+0402	CYRILLIC CAPITAL LETTER DJE
U+0453	CYRILLIC SMALL LETTER GJE	U+0403	CYRILLIC CAPITAL LETTER GJE
U+0454	CYRILLIC SMALL LETTER UKRAINIAN IE	U+0404	CYRILLIC CAPITAL LETTER UKRAINIAN IE
U+0455	CYRILLIC SMALL LETTER DZE	U+0405	CYRILLIC CAPITAL LETTER DZE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0456	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I	U+0406	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
U+0457	CYRILLIC SMALL LETTER YI	U+0407	CYRILLIC CAPITAL LETTER YI
U+0458	CYRILLIC SMALL LETTER JE	U+0408	CYRILLIC CAPITAL LETTER JE
U+0459	CYRILLIC SMALL LETTER LJE	U+0409	CYRILLIC CAPITAL LETTER LJE
U+045A	CYRILLIC SMALL LETTER NJE	U+040A	CYRILLIC CAPITAL LETTER NJE
U+045B	CYRILLIC SMALL LETTER TSHE	U+040B	CYRILLIC CAPITAL LETTER TSHE
U+045C	CYRILLIC SMALL LETTER KJE	U+040C	CYRILLIC CAPITAL LETTER KJE
U+045D	CYRILLIC SMALL LETTER I WITH GRAVE	U+040D	CYRILLIC CAPITAL LETTER I WITH GRAVE
U+045E	CYRILLIC SMALL LETTER SHORT U	U+040E	CYRILLIC CAPITAL LETTER SHORT U
U+045F	CYRILLIC SMALL LETTER DZHE	U+040F	CYRILLIC CAPITAL LETTER DZHE
U+0461	CYRILLIC SMALL LETTER OMEGA	U+0460	CYRILLIC CAPITAL LETTER OMEGA
U+0463	CYRILLIC SMALL LETTER YAT	U+0462	CYRILLIC CAPITAL LETTER YAT
U+0465	CYRILLIC SMALL LETTER IOTIFIED E	U+0464	CYRILLIC CAPITAL LETTER IOTIFIED E
U+0467	CYRILLIC SMALL LETTER LITTLE YUS	U+0466	CYRILLIC CAPITAL LETTER LITTLE YUS
U+0469	CYRILLIC SMALL LETTER IOTIFIED LITTLE YUS	U+0468	CYRILLIC CAPITAL LETTER IOTIFIED LITTLE YUS
U+046B	CYRILLIC SMALL LETTER BIG YUS	U+046A	CYRILLIC CAPITAL LETTER BIG YUS
U+046D	CYRILLIC SMALL LETTER IOTIFIED BIG YUS	U+046C	CYRILLIC CAPITAL LETTER IOTIFIED BIG YUS
U+046F	CYRILLIC SMALL LETTER KSI	U+046E	CYRILLIC CAPITAL LETTER KSI
U+0471	CYRILLIC SMALL LETTER PSI	U+0470	CYRILLIC CAPITAL LETTER PSI
U+0473	CYRILLIC SMALL LETTER FITA	U+0472	CYRILLIC CAPITAL LETTER FITA
U+0475	CYRILLIC SMALL LETTER IZHITSA	U+0474	CYRILLIC CAPITAL LETTER IZHITSA
U+0477	CYRILLIC SMALL LETTER IZHITSA WITH DOUBLE GRAVE ACCENT	U+0476	CYRILLIC CAPITAL LETTER IZHITSA WITH DOUBLE GRAVE ACCENT
U+0479	CYRILLIC SMALL LETTER UK	U+0478	CYRILLIC CAPITAL LETTER UK

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+047B	CYRILLIC SMALL LETTER ROUND OMEGA	U+047A	CYRILLIC CAPITAL LETTER ROUND OMEGA
U+047D	CYRILLIC SMALL LETTER OMEGA WITH TITLO	U+047C	CYRILLIC CAPITAL LETTER OMEGA WITH TITLO
U+047F	CYRILLIC SMALL LETTER OT	U+047E	CYRILLIC CAPITAL LETTER OT
U+0481	CYRILLIC SMALL LETTER KOPPA	U+0480	CYRILLIC CAPITAL LETTER KOPPA
U+048B	CYRILLIC SMALL LETTER SHORT I WITH TAIL	U+048A	CYRILLIC CAPITAL LETTER SHORT I WITH TAIL
U+048D	CYRILLIC SMALL LETTER SEMISOFT SIGN	U+048C	CYRILLIC CAPITAL LETTER SEMISOFT SIGN
U+048F	CYRILLIC SMALL LETTER ER WITH TICK	U+048E	CYRILLIC CAPITAL LETTER ER WITH TICK
U+0491	CYRILLIC SMALL LETTER GHE WITH UPTURN	U+0490	CYRILLIC CAPITAL LETTER GHE WITH UPTURN
U+0493	CYRILLIC SMALL LETTER GHE WITH STROKE	U+0492	CYRILLIC CAPITAL LETTER GHE WITH STROKE
U+0495	CYRILLIC SMALL LETTER GHE WITH MIDDLE HOOK	U+0494	CYRILLIC CAPITAL LETTER GHE WITH MIDDLE HOOK
U+0497	CYRILLIC SMALL LETTER ZHE WITH DESCENDER	U+0496	CYRILLIC CAPITAL LETTER ZHE WITH DESCENDER
U+0499	CYRILLIC SMALL LETTER ZE WITH DESCENDER	U+0498	CYRILLIC CAPITAL LETTER ZE WITH DESCENDER
U+049B	CYRILLIC SMALL LETTER KA WITH DESCENDER	U+049A	CYRILLIC CAPITAL LETTER KA WITH DESCENDER
U+049D	CYRILLIC SMALL LETTER KA WITH VERTICAL STROKE	U+049C	CYRILLIC CAPITAL LETTER KA WITH VERTICAL STROKE
U+049F	CYRILLIC SMALL LETTER KA WITH STROKE	U+049E	CYRILLIC CAPITAL LETTER KA WITH STROKE
U+04A1	CYRILLIC SMALL LETTER BASHKIR KA	U+04A0	CYRILLIC CAPITAL LETTER BASHKIR KA
U+04A3	CYRILLIC SMALL LETTER EN WITH DESCENDER	U+04A2	CYRILLIC CAPITAL LETTER EN WITH DESCENDER
U+04A5	CYRILLIC SMALL LIGATURE EN GHE	U+04A4	CYRILLIC CAPITAL LIGATURE EN GHE
U+04A7	CYRILLIC SMALL LETTER PE WITH MIDDLE HOOK	U+04A6	CYRILLIC CAPITAL LETTER PE WITH MIDDLE HOOK

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+04A9	CYRILLIC SMALL LETTER ABKHASIAN HA	U+04A8	CYRILLIC CAPITAL LETTER ABKHASIAN HA
U+04AB	CYRILLIC SMALL LETTER ES WITH DESCENDER	U+04AA	CYRILLIC CAPITAL LETTER ES WITH DESCENDER
U+04AD	CYRILLIC SMALL LETTER TE WITH DESCENDER	U+04AC	CYRILLIC CAPITAL LETTER TE WITH DESCENDER
U+04AF	CYRILLIC SMALL LETTER STRAIGHT U	U+04AE	CYRILLIC CAPITAL LETTER STRAIGHT U
U+04B1	CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE	U+04B0	CYRILLIC CAPITAL LETTER STRAIGHT U WITH STROKE
U+04B3	CYRILLIC SMALL LETTER HA WITH DESCENDER	U+04B2	CYRILLIC CAPITAL LETTER HA WITH DESCENDER
U+04B5	CYRILLIC SMALL LIGATURE TE TSE	U+04B4	CYRILLIC CAPITAL LIGATURE TE TSE
U+04B7	CYRILLIC SMALL LETTER CHE WITH DESCENDER	U+04B6	CYRILLIC CAPITAL LETTER CHE WITH DESCENDER
U+04B9	CYRILLIC SMALL LETTER CHE WITH VERTICAL STROKE	U+04B8	CYRILLIC CAPITAL LETTER CHE WITH VERTICAL STROKE
U+04BB	CYRILLIC SMALL LETTER SHHA	U+04BA	CYRILLIC CAPITAL LETTER SHHA
U+04BD	CYRILLIC SMALL LETTER ABKHASIAN CHE	U+04BC	CYRILLIC CAPITAL LETTER ABKHASIAN CHE
U+04BF	CYRILLIC SMALL LETTER ABKHASIAN CHE WITH DESCENDER	U+04BE	CYRILLIC CAPITAL LETTER ABKHASIAN CHE WITH DESCENDER
U+04C2	CYRILLIC SMALL LETTER ZHE WITH BREVE	U+04C1	CYRILLIC CAPITAL LETTER ZHE WITH BREVE
U+04C4	CYRILLIC SMALL LETTER KA WITH HOOK	U+04C3	CYRILLIC CAPITAL LETTER KA WITH HOOK
U+04C6	CYRILLIC SMALL LETTER EL WITH TAIL	U+04C5	CYRILLIC CAPITAL LETTER EL WITH TAIL
U+04C8	CYRILLIC SMALL LETTER EN WITH HOOK	U+04C7	CYRILLIC CAPITAL LETTER EN WITH HOOK
U+04CA	CYRILLIC SMALL LETTER EN WITH TAIL	U+04C9	CYRILLIC CAPITAL LETTER EN WITH TAIL
U+04CC	CYRILLIC SMALL LETTER KHAKASSIAN CHE	U+04CB	CYRILLIC CAPITAL LETTER KHAKASSIAN CHE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+04CE	CYRILLIC SMALL LETTER EM WITH TAIL	U+04CD	CYRILLIC CAPITAL LETTER EM WITH TAIL
U+04CF	CYRILLIC SMALL LETTER PALOCHKA	U+04C0	CYRILLIC CAPITAL LETTER PALOCHKA
U+04D1	CYRILLIC SMALL LETTER A WITH BREVE	U+04D0	CYRILLIC CAPITAL LETTER A WITH BREVE
U+04D3	CYRILLIC SMALL LETTER A WITH DIAERESIS	U+04D2	CYRILLIC CAPITAL LETTER A WITH DIAERESIS
U+04D5	CYRILLIC SMALL LIGATURE A IE	U+04D4	CYRILLIC CAPITAL LIGATURE A IE
U+04D7	CYRILLIC SMALL LETTER IE WITH BREVE	U+04D6	CYRILLIC CAPITAL LETTER IE WITH BREVE
U+04D9	CYRILLIC SMALL LETTER SCHWA	U+04D8	CYRILLIC CAPITAL LETTER SCHWA
U+04DB	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS	U+04DA	CYRILLIC CAPITAL LETTER SCHWA WITH DIAERESIS
U+04DD	CYRILLIC SMALL LETTER ZHE WITH DIAERESIS	U+04DC	CYRILLIC CAPITAL LETTER ZHE WITH DIAERESIS
U+04DF	CYRILLIC SMALL LETTER ZE WITH DIAERESIS	U+04DE	CYRILLIC CAPITAL LETTER ZE WITH DIAERESIS
U+04E1	CYRILLIC SMALL LETTER ABKHASIAN DZE	U+04E0	CYRILLIC CAPITAL LETTER ABKHASIAN DZE
U+04E3	CYRILLIC SMALL LETTER I WITH MACRON	U+04E2	CYRILLIC CAPITAL LETTER I WITH MACRON
U+04E5	CYRILLIC SMALL LETTER I WITH DIAERESIS	U+04E4	CYRILLIC CAPITAL LETTER I WITH DIAERESIS
U+04E7	CYRILLIC SMALL LETTER O WITH DIAERESIS	U+04E6	CYRILLIC CAPITAL LETTER O WITH DIAERESIS
U+04E9	CYRILLIC SMALL LETTER BARRED O	U+04E8	CYRILLIC CAPITAL LETTER BARRED O
U+04EB	CYRILLIC SMALL LETTER BARRED O WITH DIAERESIS	U+04EA	CYRILLIC CAPITAL LETTER BARRED O WITH DIAERESIS
U+04ED	CYRILLIC SMALL LETTER E WITH DIAERESIS	U+04EC	CYRILLIC CAPITAL LETTER E WITH DIAERESIS
U+04EF	CYRILLIC SMALL LETTER U WITH MACRON	U+04EE	CYRILLIC CAPITAL LETTER U WITH MACRON
U+04F1	CYRILLIC SMALL LETTER U WITH DIAERESIS	U+04F0	CYRILLIC CAPITAL LETTER U WITH DIAERESIS

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+04F3	CYRILLIC SMALL LETTER U WITH DOUBLE ACUTE	U+04F2	CYRILLIC CAPITAL LETTER U WITH DOUBLE ACUTE
U+04F5	CYRILLIC SMALL LETTER CHE WITH DIAERESIS	U+04F4	CYRILLIC CAPITAL LETTER CHE WITH DIAERESIS
U+04F7	CYRILLIC SMALL LETTER GHE WITH DESCENDER	U+04F6	CYRILLIC CAPITAL LETTER GHE WITH DESCENDER
U+04F9	CYRILLIC SMALL LETTER YERU WITH DIAERESIS	U+04F8	CYRILLIC CAPITAL LETTER YERU WITH DIAERESIS
U+04FB	CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK	U+04FA	CYRILLIC CAPITAL LETTER GHE WITH STROKE AND HOOK
U+04FD	CYRILLIC SMALL LETTER HA WITH HOOK	U+04FC	CYRILLIC CAPITAL LETTER HA WITH HOOK
U+04FF	CYRILLIC SMALL LETTER HA WITH STROKE	U+04FE	CYRILLIC CAPITAL LETTER HA WITH STROKE
U+0501	CYRILLIC SMALL LETTER KOMI DE	U+0500	CYRILLIC CAPITAL LETTER KOMI DE
U+0503	CYRILLIC SMALL LETTER KOMI DJE	U+0502	CYRILLIC CAPITAL LETTER KOMI DJE
U+0505	CYRILLIC SMALL LETTER KOMI ZJE	U+0504	CYRILLIC CAPITAL LETTER KOMI ZJE
U+0507	CYRILLIC SMALL LETTER KOMI DZJE	U+0506	CYRILLIC CAPITAL LETTER KOMI DZJE
U+0509	CYRILLIC SMALL LETTER KOMI LJE	U+0508	CYRILLIC CAPITAL LETTER KOMI LJE
U+050B	CYRILLIC SMALL LETTER KOMI NJE	U+050A	CYRILLIC CAPITAL LETTER KOMI NJE
U+050D	CYRILLIC SMALL LETTER KOMI SJE	U+050C	CYRILLIC CAPITAL LETTER KOMI SJE
U+050F	CYRILLIC SMALL LETTER KOMI TJE	U+050E	CYRILLIC CAPITAL LETTER KOMI THE
U+0511	CYRILLIC SMALL LETTER REVERSED ZE	U+0510	CYRILLIC CAPITAL LETTER REVERSED ZE
U+0513	CYRILLIC SMALL LETTER EL WITH HOOK	U+0512	CYRILLIC CAPITAL LETTER EL WITH HOOK
U+0515	CYRILLIC SMALL LETTER LHA	U+0514	CYRILLIC CAPITAL LETTER LHA
U+0517	CYRILLIC SMALL LETTER RHA	U+0516	CYRILLIC CAPITAL LETTER RHA
U+0519	CYRILLIC SMALL LETTER YAE	U+0518	CYRILLIC CAPITAL LETTER YAE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+051B	CYRILLIC SMALL LETTER QA	U+051A	CYRILLIC CAPITAL LETTER QA
U+051D	CYRILLIC SMALL LETTER WE	U+051C	CYRILLIC CAPITAL LETTER WE
U+051F	CYRILLIC SMALL LETTER ALEUT KA	U+051E	CYRILLIC CAPITAL LETTER ALEUT KA
U+0521	CYRILLIC SMALL LETTER EL WITH MIDDLE HOOK	U+0520	CYRILLIC CAPITAL LETTER EL WITH MIDDLE HOOK
U+0523	CYRILLIC SMALL LETTER EN WITH MIDDLE HOOK	U+0522	CYRILLIC CAPITAL LETTER EN WITH MIDDLE HOOK
U+0525	CYRILLIC SMALL LETTER PE WITH DESCENDER	U+0524	CYRILLIC CAPITAL LETTER PE WITH DESCENDER
U+0527	CYRILLIC SMALL LETTER SHHA WITH DESCENDER	U+0526	CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER
U+0561	ARMENIAN SMALL LETTER AYB	U+0531	ARMENIAN CAPITAL LETTER AYB
U+0562	ARMENIAN SMALL LETTER BEN	U+0532	ARMENIAN CAPITAL LETTER BEN
U+0563	ARMENIAN SMALL LETTER GIM	U+0533	ARMENIAN CAPITAL LETTER GIM
U+0564	ARMENIAN SMALL LETTER DA	U+0534	ARMENIAN CAPITAL LETTER DA
U+0565	ARMENIAN SMALL LETTER ECH	U+0535	ARMENIAN CAPITAL LETTER ECH
U+0566	ARMENIAN SMALL LETTER ZA	U+0536	ARMENIAN CAPITAL LETTER ZA
U+0567	ARMENIAN SMALL LETTER EH	U+0537	ARMENIAN CAPITAL LETTER EH
U+0568	ARMENIAN SMALL LETTER ET	U+0538	ARMENIAN CAPITAL LETTER ET
U+0569	ARMENIAN SMALL LETTER TO	U+0539	ARMENIAN CAPITAL LETTER TO
U+056A	ARMENIAN SMALL LETTER ZHE	U+053A	ARMENIAN CAPITAL LETTER ZHE
U+056B	ARMENIAN SMALL LETTER INI	U+053B	ARMENIAN CAPITAL LETTER INI
U+056C	ARMENIAN SMALL LETTER LIWN	U+053C	ARMENIAN CAPITAL LETTER LIWN
U+056D	ARMENIAN SMALL LETTER XEH	U+053D	ARMENIAN CAPITAL LETTER XEH
U+056E	ARMENIAN SMALL LETTER CA	U+053E	ARMENIAN CAPITAL LETTER CA
U+056F	ARMENIAN SMALL LETTER KEN	U+053F	ARMENIAN CAPITAL LETTER KEN
U+0570	ARMENIAN SMALL LETTER HO	U+0540	ARMENIAN CAPITAL LETTER HO
U+0571	ARMENIAN SMALL LETTER JA	U+0541	ARMENIAN CAPITAL LETTER JA
U+0572	ARMENIAN SMALL LETTER GHAD	U+0542	ARMENIAN CAPITAL LETTER GHAD

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0573	ARMENIAN SMALL LETTER CHEH	U+0543	ARMENIAN CAPITAL LETTER CHEH
U+0574	ARMENIAN SMALL LETTER MEN	U+0544	ARMENIAN CAPITAL LETTER MEN
U+0575	ARMENIAN SMALL LETTER YI	U+0545	ARMENIAN CAPITAL LETTER YI
U+0576	ARMENIAN SMALL LETTER NOW	U+0546	ARMENIAN CAPITAL LETTER NOW
U+0577	ARMENIAN SMALL LETTER SHA	U+0547	ARMENIAN CAPITAL LETTER SHA
U+0578	ARMENIAN SMALL LETTER VO	U+0548	ARMENIAN CAPITAL LETTER VO
U+0579	ARMENIAN SMALL LETTER CHA	U+0549	ARMENIAN CAPITAL LETTER CHA
U+057A	ARMENIAN SMALL LETTER PEH	U+054A	ARMENIAN CAPITAL LETTER PEH
U+057B	ARMENIAN SMALL LETTER JHEH	U+054B	ARMENIAN CAPITAL LETTER JHEH
U+057C	ARMENIAN SMALL LETTER RA	U+054C	ARMENIAN CAPITAL LETTER RA
U+057D	ARMENIAN SMALL LETTER SEH	U+054D	ARMENIAN CAPITAL LETTER SEH
U+057E	ARMENIAN SMALL LETTER VEW	U+054E	ARMENIAN CAPITAL LETTER VEW
U+057F	ARMENIAN SMALL LETTER TIWN	U+054F	ARMENIAN CAPITAL LETTER TIWN
U+0580	ARMENIAN SMALL LETTER REH	U+0550	ARMENIAN CAPITAL LETTER REH
U+0581	ARMENIAN SMALL LETTER CO	U+0551	ARMENIAN CAPITAL LETTER CO
U+0582	ARMENIAN SMALL LETTER YIWN	U+0552	ARMENIAN CAPITAL LETTER YIWN
U+0583	ARMENIAN SMALL LETTER PIWR	U+0553	ARMENIAN CAPITAL LETTER PIWR
U+0584	ARMENIAN SMALL LETTER KEH	U+0554	ARMENIAN CAPITAL LETTER KEH
U+0585	ARMENIAN SMALL LETTER OH	U+0555	ARMENIAN CAPITAL LETTER OH
U+0586	ARMENIAN SMALL LETTER FEH	U+0556	ARMENIAN CAPITAL LETTER FEH
U+1D79	LATIN SMALL LETTER INSULAR G	U+A77D	LATIN CAPITAL LETTER INSULAR G
U+1D7D	LATIN SMALL LETTER P WITH STROKE	U+2C63	LATIN CAPITAL LETTER P WITH STROKE
U+1E01	LATIN SMALL LETTER A WITH RING BELOW	U+1E00	LATIN CAPITAL LETTER A WITH RING BELOW
U+1E03	LATIN SMALL LETTER B WITH DOT ABOVE	U+1E02	LATIN CAPITAL LETTER B WITH DOT ABOVE
U+1E05	LATIN SMALL LETTER B WITH DOT BELOW	U+1E04	LATIN CAPITAL LETTER B WITH DOT BELOW

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E07	LATIN SMALL LETTER B WITH LINE BELOW	U+1E06	LATIN CAPITAL LETTER B WITH LINE BELOW
U+1E09	LATIN SMALL LETTER C WITH CEDILLA AND ACUTE	U+1E08	LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE
U+1E0B	LATIN SMALL LETTER D WITH DOT ABOVE	U+1E0A	LATIN CAPITAL LETTER D WITH DOT ABOVE
U+1E0D	LATIN SMALL LETTER D WITH DOT BELOW	U+1E0C	LATIN CAPITAL LETTER D WITH DOT BELOW
U+1E0F	LATIN SMALL LETTER D WITH LINE BELOW	U+1E0E	LATIN CAPITAL LETTER D WITH LINE BELOW
U+1E11	LATIN SMALL LETTER D WITH CEDILLA	U+1E10	LATIN CAPITAL LETTER D WITH CEDILLA
U+1E13	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW	U+1E12	LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW
U+1E15	LATIN SMALL LETTER E WITH MACRON AND GRAVE	U+1E14	LATIN CAPITAL LETTER E WITH MACRON AND GRAVE
U+1E17	LATIN SMALL LETTER E WITH MACRON AND ACUTE	U+1E16	LATIN CAPITAL LETTER E WITH MACRON AND ACUTE
U+1E19	LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW	U+1E18	LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW
U+1E1B	LATIN SMALL LETTER E WITH TILDE BELOW	U+1E1A	LATIN CAPITAL LETTER E WITH TILDE BELOW
U+1E1D	LATIN SMALL LETTER E WITH CEDILLA AND BREVE	U+1E1C	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE
U+1E1F	LATIN SMALL LETTER F WITH DOT ABOVE	U+1E1E	LATIN CAPITAL LETTER F WITH DOT ABOVE
U+1E21	LATIN SMALL LETTER G WITH MACRON	U+1E20	LATIN CAPITAL LETTER G WITH MACRON
U+1E23	LATIN SMALL LETTER H WITH DOT ABOVE	U+1E22	LATIN CAPITAL LETTER H WITH DOT ABOVE
U+1E25	LATIN SMALL LETTER H WITH DOT BELOW	U+1E24	LATIN CAPITAL LETTER H WITH DOT BELOW
U+1E27	LATIN SMALL LETTER H WITH DIAERESIS	U+1E26	LATIN CAPITAL LETTER H WITH DIAERESIS
U+1E29	LATIN SMALL LETTER H WITH CEDILLA	U+1E28	LATIN CAPITAL LETTER H WITH CEDILLA

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E2B	LATIN SMALL LETTER H WITH BREVE BELOW	U+1E2A	LATIN CAPITAL LETTER H WITH BREVE BELOW
U+1E2D	LATIN SMALL LETTER I WITH TILDE BELOW	U+1E2C	LATIN CAPITAL LETTER I WITH TILDE BELOW
U+1E2F	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE	U+1E2E	LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE
U+1E31	LATIN SMALL LETTER K WITH ACUTE	U+1E30	LATIN CAPITAL LETTER K WITH ACUTE
U+1E33	LATIN SMALL LETTER K WITH DOT BELOW	U+1E32	LATIN CAPITAL LETTER K WITH DOT BELOW
U+1E35	LATIN SMALL LETTER K WITH LINE BELOW	U+1E34	LATIN CAPITAL LETTER K WITH LINE BELOW
U+1E37	LATIN SMALL LETTER L WITH DOT BELOW	U+1E36	LATIN CAPITAL LETTER L WITH DOT BELOW
U+1E39	LATIN SMALL LETTER L WITH DOT BELOW AND MACRON	U+1E38	LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON
U+1E3B	LATIN SMALL LETTER L WITH LINE BELOW	U+1E3A	LATIN CAPITAL LETTER L WITH LINE BELOW
U+1E3D	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW	U+1E3C	LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW
U+1E3F	LATIN SMALL LETTER M WITH ACUTE	U+1E3E	LATIN CAPITAL LETTER M WITH ACUTE
U+1E41	LATIN SMALL LETTER M WITH DOT ABOVE	U+1E40	LATIN CAPITAL LETTER M WITH DOT ABOVE
U+1E43	LATIN SMALL LETTER M WITH DOT BELOW	U+1E42	LATIN CAPITAL LETTER M WITH DOT BELOW
U+1E45	LATIN SMALL LETTER N WITH DOT ABOVE	U+1E44	LATIN CAPITAL LETTER N WITH DOT ABOVE
U+1E47	LATIN SMALL LETTER N WITH DOT BELOW	U+1E46	LATIN CAPITAL LETTER N WITH DOT BELOW
U+1E49	LATIN SMALL LETTER N WITH LINE BELOW	U+1E48	LATIN CAPITAL LETTER N WITH LINE BELOW
U+1E4B	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW	U+1E4A	LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW
U+1E4D	LATIN SMALL LETTER O WITH TILDE AND ACUTE	U+1E4C	LATIN CAPITAL LETTER O WITH TILDE AND ACUTE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E4F	LATIN SMALL LETTER O WITH TILDE AND DIAERESIS	U+1E4E	LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS
U+1E51	LATIN SMALL LETTER O WITH MACRON AND GRAVE	U+1E50	LATIN CAPITAL LETTER O WITH MACRON AND GRAVE
U+1E53	LATIN SMALL LETTER O WITH MACRON AND ACUTE	U+1E52	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE
U+1E55	LATIN SMALL LETTER P WITH ACUTE	U+1E54	LATIN CAPITAL LETTER P WITH ACUTE
U+1E57	LATIN SMALL LETTER P WITH DOT ABOVE	U+1E56	LATIN CAPITAL LETTER P WITH DOT ABOVE
U+1E59	LATIN SMALL LETTER R WITH DOT ABOVE	U+1E58	LATIN CAPITAL LETTER R WITH DOT ABOVE
U+1E5B	LATIN SMALL LETTER R WITH DOT BELOW	U+1E5A	LATIN CAPITAL LETTER R WITH DOT BELOW
U+1E5D	LATIN SMALL LETTER R WITH DOT BELOW AND MACRON	U+1E5C	LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON
U+1E5F	LATIN SMALL LETTER R WITH LINE BELOW	U+1E5E	LATIN CAPITAL LETTER R WITH LINE BELOW
U+1E61	LATIN SMALL LETTER S WITH DOT ABOVE	U+1E60	LATIN CAPITAL LETTER S WITH DOT ABOVE
U+1E63	LATIN SMALL LETTER S WITH DOT BELOW	U+1E62	LATIN CAPITAL LETTER S WITH DOT BELOW
U+1E65	LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE	U+1E64	LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE
U+1E67	LATIN SMALL LETTER S WITH CARON AND DOT ABOVE	U+1E66	LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE
U+1E69	LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE	U+1E68	LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE
U+1E6B	LATIN SMALL LETTER T WITH DOT ABOVE	U+1E6A	LATIN CAPITAL LETTER T WITH DOT ABOVE
U+1E6D	LATIN SMALL LETTER T WITH DOT BELOW	U+1E6C	LATIN CAPITAL LETTER T WITH DOT BELOW
U+1E6F	LATIN SMALL LETTER T WITH LINE BELOW	U+1E6E	LATIN CAPITAL LETTER T WITH LINE BELOW
U+1E71	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW	U+1E70	LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E73	LATIN SMALL LETTER U WITH DIAERESIS BELOW	U+1E72	LATIN CAPITAL LETTER U WITH DIAERESIS BELOW
U+1E75	LATIN SMALL LETTER U WITH TILDE BELOW	U+1E74	LATIN CAPITAL LETTER U WITH TILDE BELOW
U+1E77	LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW	U+1E76	LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW
U+1E79	LATIN SMALL LETTER U WITH TILDE AND ACUTE	U+1E78	LATIN CAPITAL LETTER U WITH TILDE AND ACUTE
U+1E7B	LATIN SMALL LETTER U WITH MACRON AND DIAERESIS	U+1E7A	LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS
U+1E7D	LATIN SMALL LETTER V WITH TILDE	U+1E7C	LATIN CAPITAL LETTER V WITH TILDE
U+1E7F	LATIN SMALL LETTER V WITH DOT BELOW	U+1E7E	LATIN CAPITAL LETTER V WITH DOT BELOW
U+1E81	LATIN SMALL LETTER W WITH GRAVE	U+1E80	LATIN CAPITAL LETTER W WITH GRAVE
U+1E83	LATIN SMALL LETTER W WITH ACUTE	U+1E82	LATIN CAPITAL LETTER W WITH ACUTE
U+1E85	LATIN SMALL LETTER W WITH DIAERESIS	U+1E84	LATIN CAPITAL LETTER W WITH DIAERESIS
U+1E87	LATIN SMALL LETTER W WITH DOT ABOVE	U+1E86	LATIN CAPITAL LETTER W WITH DOT ABOVE
U+1E89	LATIN SMALL LETTER W WITH DOT BELOW	U+1E88	LATIN CAPITAL LETTER W WITH DOT BELOW
U+1E8B	LATIN SMALL LETTER X WITH DOT ABOVE	U+1E8A	LATIN CAPITAL LETTER X WITH DOT ABOVE
U+1E8D	LATIN SMALL LETTER X WITH DIAERESIS	U+1E8C	LATIN CAPITAL LETTER X WITH DIAERESIS
U+1E8F	LATIN SMALL LETTER Y WITH DOT ABOVE	U+1E8E	LATIN CAPITAL LETTER Y WITH DOT ABOVE
U+1E91	LATIN SMALL LETTER Z WITH CIRCUMFLEX	U+1E90	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX
U+1E93	LATIN SMALL LETTER Z WITH DOT BELOW	U+1E92	LATIN CAPITAL LETTER Z WITH DOT BELOW
U+1E95	LATIN SMALL LETTER Z WITH LINE BELOW	U+1E94	LATIN CAPITAL LETTER Z WITH LINE BELOW

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E9B	LATIN SMALL LETTER LONG S WITH DOT ABOVE	U+1E60	LATIN CAPITAL LETTER S WITH DOT ABOVE
U+1EA1	LATIN SMALL LETTER A WITH DOT BELOW	U+1EA0	LATIN CAPITAL LETTER A WITH DOT BELOW
U+1EA3	LATIN SMALL LETTER A WITH HOOK ABOVE	U+1EA2	LATIN CAPITAL LETTER A WITH HOOK ABOVE
U+1EA5	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE	U+1EA4	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE
U+1EA7	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE	U+1EA6	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE
U+1EA9	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	U+1EA8	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
U+1EAB	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE	U+1EAA	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE
U+1EAD	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW	U+1EAC	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW
U+1EAF	LATIN SMALL LETTER A WITH BREVE AND ACUTE	U+1EAE	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE
U+1EB1	LATIN SMALL LETTER A WITH BREVE AND GRAVE	U+1EB0	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE
U+1EB3	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE	U+1EB2	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE
U+1EB5	LATIN SMALL LETTER A WITH BREVE AND TILDE	U+1EB4	LATIN CAPITAL LETTER A WITH BREVE AND TILDE
U+1EB7	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	U+1EB6	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW
U+1EB9	LATIN SMALL LETTER E WITH DOT BELOW	U+1EB8	LATIN CAPITAL LETTER E WITH DOT BELOW
U+1EBB	LATIN SMALL LETTER E WITH HOOK ABOVE	U+1EBA	LATIN CAPITAL LETTER E WITH HOOK ABOVE
U+1EBD	LATIN SMALL LETTER E WITH TILDE	U+1EBC	LATIN CAPITAL LETTER E WITH TILDE
U+1EBF	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE	U+1EBE	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE
U+1EC1	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE	U+1EC0	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1EC3	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	U+1EC2	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
U+1EC5	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE	U+1EC4	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE
U+1EC7	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	U+1EC6	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW
U+1EC9	LATIN SMALL LETTER I WITH HOOK ABOVE	U+1EC8	LATIN CAPITAL LETTER I WITH HOOK ABOVE
U+1ECB	LATIN SMALL LETTER I WITH DOT BELOW	U+1ECA	LATIN CAPITAL LETTER I WITH DOT BELOW
U+1ECD	LATIN SMALL LETTER O WITH DOT BELOW	U+1ECC	LATIN CAPITAL LETTER O WITH DOT BELOW
U+1ECF	LATIN SMALL LETTER O WITH HOOK ABOVE	U+1ECE	LATIN CAPITAL LETTER O WITH HOOK ABOVE
U+1ED1	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE	U+1ED0	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE
U+1ED3	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE	U+1ED2	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE
U+1ED5	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	U+1ED4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
U+1ED7	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE	U+1ED6	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE
U+1ED9	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	U+1ED8	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW
U+1EDB	LATIN SMALL LETTER O WITH HORN AND ACUTE	U+1EDA	LATIN CAPITAL LETTER O WITH HORN AND ACUTE
U+1EDD	LATIN SMALL LETTER O WITH HORN AND GRAVE	U+1EDC	LATIN CAPITAL LETTER O WITH HORN AND GRAVE
U+1EDF	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE	U+1EDE	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE
U+1EE1	LATIN SMALL LETTER O WITH HORN AND TILDE	U+1EE0	LATIN CAPITAL LETTER O WITH HORN AND TILDE
U+1EE3	LATIN SMALL LETTER O WITH HORN AND DOT BELOW	U+1EE2	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW
U+1EE5	LATIN SMALL LETTER U WITH DOT BELOW	U+1EE4	LATIN CAPITAL LETTER U WITH DOT BELOW

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1EE7	LATIN SMALL LETTER U WITH HOOK ABOVE	U+1EE6	LATIN CAPITAL LETTER U WITH HOOK ABOVE
U+1EE9	LATIN SMALL LETTER U WITH HORN AND ACUTE	U+1EE8	LATIN CAPITAL LETTER U WITH HORN AND ACUTE
U+1EEB	LATIN SMALL LETTER U WITH HORN AND GRAVE	U+1EEA	LATIN CAPITAL LETTER U WITH HORN AND GRAVE
U+1EED	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE	U+1EEC	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE
U+1EEF	LATIN SMALL LETTER U WITH HORN AND TILDE	U+1EEE	LATIN CAPITAL LETTER U WITH HORN AND TILDE
U+1EF1	LATIN SMALL LETTER U WITH HORN AND DOT BELOW	U+1EF0	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW
U+1EF3	LATIN SMALL LETTER Y WITH GRAVE	U+1EF2	LATIN CAPITAL LETTER Y WITH GRAVE
U+1EF5	LATIN SMALL LETTER Y WITH DOT BELOW	U+1EF4	LATIN CAPITAL LETTER Y WITH DOT BELOW
U+1EF7	LATIN SMALL LETTER Y WITH HOOK ABOVE	U+1EF6	LATIN CAPITAL LETTER Y WITH HOOK ABOVE
U+1EF9	LATIN SMALL LETTER Y WITH TILDE	U+1EF8	LATIN CAPITAL LETTER Y WITH TILDE
U+1EFB	LATIN SMALL LETTER MIDDLE-WELSH LL	U+EFA	LATIN CAPITAL LETTER MIDDLE-WELSH LL
U+1EFD	LATIN SMALL LETTER MIDDLE-WELSH V	U+1EFC	LATIN CAPITAL LETTER MIDDLE-WELSH V
U+1EFF	LATIN SMALL LETTER Y WITH LOOP	U+1EFE	LATIN CAPITAL LETTER Y WITH LOOP
U+1F00	GREEK SMALL LETTER ALPHA WITH PSILI	U+1F08	GREEK CAPITAL LETTER ALPHA WITH PSILI
U+1F01	GREEK SMALL LETTER ALPHA WITH DASIA	U+1F09	GREEK CAPITAL LETTER ALPHA WITH DASIA
U+1F02	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA	U+1F0A	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA
U+1F03	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA	U+1F0B	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA
U+1F04	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA	U+1F0C	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F05	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA	U+1F0D	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA
U+1F06	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI	U+1F0E	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI
U+1F07	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI	U+1F0F	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI
U+1F10	GREEK SMALL LETTER EPSILON WITH PSILI	U+1F18	GREEK CAPITAL LETTER EPSILON WITH PSILI
U+1F11	GREEK SMALL LETTER EPSILON WITH DASIA	U+1F19	GREEK CAPITAL LETTER EPSILON WITH DASIA
U+1F12	GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA	U+1F1A	GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA
U+1F13	GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA	U+1F1B	GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA
U+1F14	GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA	U+1F1C	GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA
U+1F15	GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA	U+1F1D	GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA
U+1F20	GREEK SMALL LETTER ETA WITH PSILI	U+1F28	GREEK CAPITAL LETTER ETA WITH PSILI
U+1F21	GREEK SMALL LETTER ETA WITH DASIA	U+1F29	GREEK CAPITAL LETTER ETA WITH DASIA
U+1F22	GREEK SMALL LETTER ETA WITH PSILI AND VARIA	U+1F2A	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA
U+1F23	GREEK SMALL LETTER ETA WITH DASIA AND VARIA	U+1F2B	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA
U+1F24	GREEK SMALL LETTER ETA WITH PSILI AND OXIA	U+1F2C	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA
U+1F25	GREEK SMALL LETTER ETA WITH DASIA AND OXIA	U+1F2D	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA
U+1F26	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI	U+1F2E	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI
U+1F27	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI	U+1F2F	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI
U+1F30	GREEK SMALL LETTER IOTA WITH PSILI	U+1F38	GREEK CAPITAL LETTER IOTA WITH PSILI

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F31	GREEK SMALL LETTER IOTA WITH DASIA	U+1F39	GREEK CAPITAL LETTER IOTA WITH DASIA
U+1F32	GREEK SMALL LETTER IOTA WITH PSILI AND VARIA	U+1F3A	GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA
U+1F33	GREEK SMALL LETTER IOTA WITH DASIA AND VARIA	U+1F3B	GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA
U+1F34	GREEK SMALL LETTER IOTA WITH PSILI AND OXIA	U+1F3C	GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA
U+1F35	GREEK SMALL LETTER IOTA WITH DASIA AND OXIA	U+1F3D	GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA
U+1F36	GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI	U+1F3E	GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI
U+1F37	GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI	U+1F3F	GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI
U+1F40	GREEK SMALL LETTER OMICRON WITH PSILI	U+1F48	GREEK CAPITAL LETTER OMICRON WITH PSILI
U+1F41	GREEK SMALL LETTER OMICRON WITH DASIA	U+1F49	GREEK CAPITAL LETTER OMICRON WITH DASIA
U+1F42	GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA	U+1F4A	GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA
U+1F43	GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA	U+1F4B	GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA
U+1F44	GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA	U+1F4C	GREEK CAPITAL LETTER OMICRON WITH PSILI AND OXIA
U+1F45	GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA	U+1F4D	GREEK CAPITAL LETTER OMICRON WITH DASIA AND OXIA
U+1F51	GREEK SMALL LETTER UPSILON WITH DASIA	U+1F59	GREEK CAPITAL LETTER UPSILON WITH DASIA
U+1F53	GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA	U+1F5B	GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA
U+1F55	GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA	U+1F5D	GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA
U+1F57	GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI	U+1F5F	GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI
U+1F60	GREEK SMALL LETTER OMEGA WITH PSILI	U+1F68	GREEK CAPITAL LETTER OMEGA WITH PSILI

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F61	GREEK SMALL LETTER OMEGA WITH DASIA	U+1F69	GREEK CAPITAL LETTER OMEGA WITH DASIA
U+1F62	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA	U+1F6A	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA
U+1F63	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA	U+1F6B	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA
U+1F64	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA	U+1F6C	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA
U+1F65	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA	U+1F6D	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA
U+1F66	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI	U+1F6E	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI
U+1F67	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI	U+1F6F	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI
U+1F70	GREEK SMALL LETTER ALPHA WITH VARIA	U+1FBA	GREEK CAPITAL LETTER ALPHA WITH VARIA
U+1F71	GREEK SMALL LETTER ALPHA WITH OXIA	U+1FBB	GREEK CAPITAL LETTER ALPHA WITH OXIA
U+1F72	GREEK SMALL LETTER EPSILON WITH VARIA	U+1FC8	GREEK CAPITAL LETTER EPSILON WITH VARIA
U+1F73	GREEK SMALL LETTER EPSILON WITH OXIA	U+1FC9	GREEK CAPITAL LETTER EPSILON WITH OXIA
U+1F74	GREEK SMALL LETTER ETA WITH VARIA	U+1FCA	GREEK CAPITAL LETTER ETA WITH VARIA
U+1F75	GREEK SMALL LETTER ETA WITH OXIA	U+1FCB	GREEK CAPITAL LETTER ETA WITH OXIA
U+1F76	GREEK SMALL LETTER IOTA WITH VARIA	U+1FDA	GREEK CAPITAL LETTER IOTA WITH VARIA
U+1F77	GREEK SMALL LETTER IOTA WITH OXIA	U+1FDB	GREEK CAPITAL LETTER IOTA WITH OXIA
U+1F78	GREEK SMALL LETTER OMICRON WITH VARIA	U+1FF8	GREEK CAPITAL LETTER OMICRON WITH VARIA
U+1F79	GREEK SMALL LETTER OMICRON WITH OXIA	U+1FF9	GREEK CAPITAL LETTER OMICRON WITH OXIA
U+1F7A	GREEK SMALL LETTER UPSILON WITH VARIA	U+1FEA	GREEK CAPITAL LETTER UPSILON WITH VARIA

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F7B	GREEK SMALL LETTER UPSILON WITH OXIA	U+1FEB	GREEK CAPITAL LETTER UPSILON WITH OXIA
U+1F7C	GREEK SMALL LETTER OMEGA WITH VARIA	U+1FFA	GREEK CAPITAL LETTER OMEGA WITH VARIA
U+1F7D	GREEK SMALL LETTER OMEGA WITH OXIA	U+1FFB	GREEK CAPITAL LETTER OMEGA WITH OXIA
U+1F80	GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI	U+1F88	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENI
U+1F81	GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI	U+1F89	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI
U+1F82	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI	U+1F8A	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI
U+1F83	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI	U+1F8B	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI
U+1F84	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI	U+1F8C	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI
U+1F85	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI	U+1F8D	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI
U+1F86	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI	U+1F8E	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
U+1F87	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI	U+1F8F	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
U+1F90	GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI	U+1F98	GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI
U+1F91	GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI	U+1F99	GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI
U+1F92	GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI	U+1F9A	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI
U+1F93	GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI	U+1F9B	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F94	GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI	U+1F9C	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI
U+1F95	GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI	U+1F9D	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI
U+1F96	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI	U+1F9E	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
U+1F97	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI	U+1F9F	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
U+1FA0	GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI	U+1FA8	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI
U+1FA1	GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI	U+1FA9	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI
U+1FA2	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI	U+1FAA	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI
U+1FA3	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI	U+1FAB	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI
U+1FA4	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI	U+1FAC	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI
U+1FA5	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI	U+1FAD	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI
U+1FA6	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI	U+1FAE	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI
U+1FA7	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI	U+1FAF	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI
U+1FB0	GREEK SMALL LETTER ALPHA WITH VRACHY	U+1FB8	GREEK CAPITAL LETTER ALPHA WITH VRACHY
U+1FB1	GREEK SMALL LETTER ALPHA WITH MACRON	U+1FB9	GREEK CAPITAL LETTER ALPHA WITH MACRON

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1FB3	GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI	U+1FBC	GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI
U+1FBE	GREEK PROSGEGRAMMENI	U+0399	GREEK CAPITAL IOTA
U+1FC3	GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI	U+1FCC	GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI
U+1FD0	GREEK SMALL LETTER IOTA WITH VRACHY	U+1FD8	GREEK CAPITAL LETTER IOTA WITH VRACHY
U+1FD1	GREEK SMALL LETTER IOTA WITH MACRON	U+1FD9	GREEK CAPITAL LETTER IOTA WITH MACRON
U+1FE0	GREEK SMALL LETTER UPSILON WITH VRACHY	U+1FE8	GREEK CAPITAL LETTER UPSILON WITH VRACHY
U+1FE1	GREEK SMALL LETTER UPSILON WITH MACRON	U+1FE9	GREEK CAPITAL LETTER UPSILON WITH MACRON
U+1FE5	GREEK SMALL LETTER RHO WITH DASIA	U+1FEC	GREEK CAPITAL LETTER RHO WITH DASIA
U+1FF3	GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI	U+1FFC	GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI
U+214E	TURNED SMALL F	U+2132	TURNED CAPITAL F
U+2170	SMALL ROMAN NUMERAL ONE	U+2160	ROMAN NUMERAL ONE
U+2171	SMALL ROMAN NUMERAL TWO	U+2161	ROMAN NUMERAL TWO
U+2172	SMALL ROMAN NUMERAL THREE	U+2162	ROMAN NUMERAL THREE
U+2173	SMALL ROMAN NUMERAL FOUR	U+2163	ROMAN NUMERAL FOUR
U+2174	SMALL ROMAN NUMERAL FIVE	U+2164	ROMAN NUMERAL FIVE
U+2175	SMALL ROMAN NUMERAL SIX	U+2165	ROMAN NUMERAL SIX
U+2176	SMALL ROMAN NUMERAL SEVEN	U+2166	ROMAN NUMERAL SEVEN
U+2177	SMALL ROMAN NUMERAL EIGHT	U+2167	ROMAN NUMERAL EIGHT
U+2178	SMALL ROMAN NUMERAL NINE	U+2168	ROMAN NUMERAL NINE
U+2179	SMALL ROMAN NUMERAL TEN	U+2169	ROMAN NUMERAL TEN
U+217A	SMALL ROMAN NUMERAL ELEVEN	U+216A	ROMAN NUMERAL ELEVEN
U+217B	SMALL ROMAN NUMERAL TWELVE	U+216B	ROMAN NUMERAL TWELVE
U+217C	SMALL ROMAN NUMERAL FIFTY	U+216C	ROMAN NUMERAL FIFTY

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+217D	SMALL ROMAN NUMERAL ONE HUNDRED	U+216D	ROMAN NUMERAL ONE HUNDRED
U+217E	SMALL ROMAN NUMERAL FIVE HUNDRED	U+216E	ROMAN NUMERAL FIVE HUNDRED
U+217F	SMALL ROMAN NUMERAL ONE THOUSAND	U+216F	ROMAN NUMERAL ONE THOUSAND
U+2184	LATIN SMALL LETTER REVERSED C	U+2183	ROMAN NUMERAL REVERSED ONE HUNDRED
U+24D0	CIRCLED LATIN SMALL LETTER A	U+24B6	CIRCLED LATIN CAPITAL LETTER A
U+24D1	CIRCLED LATIN SMALL LETTER B	U+24B7	CIRCLED LATIN CAPITAL LETTER B
U+24D2	CIRCLED LATIN SMALL LETTER C	U+24B8	CIRCLED LATIN CAPITAL LETTER C
U+24D3	CIRCLED LATIN SMALL LETTER D	U+24B9	CIRCLED LATIN CAPITAL LETTER D
U+24D4	CIRCLED LATIN SMALL LETTER E	U+24BA	CIRCLED LATIN CAPITAL LETTER E
U+24D5	CIRCLED LATIN SMALL LETTER F	U+24BB	CIRCLED LATIN CAPITAL LETTER F
U+24D6	CIRCLED LATIN SMALL LETTER G	U+24BC	CIRCLED LATIN CAPITAL LETTER G
U+24D7	CIRCLED LATIN SMALL LETTER H	U+24BD	CIRCLED LATIN CAPITAL LETTER H
U+24D8	CIRCLED LATIN SMALL LETTER I	U+24BE	CIRCLED LATIN CAPITAL LETTER I
U+24D9	CIRCLED LATIN SMALL LETTER J	U+24BF	CIRCLED LATIN CAPITAL LETTER J
U+24DA	CIRCLED LATIN SMALL LETTER K	U+24C0	CIRCLED LATIN CAPITAL LETTER K
U+24DB	CIRCLED LATIN SMALL LETTER L	U+24C1	CIRCLED LATIN CAPITAL LETTER L
U+24DC	CIRCLED LATIN SMALL LETTER M	U+24C2	CIRCLED LATIN CAPITAL LETTER M
U+24DD	CIRCLED LATIN SMALL LETTER N	U+24C3	CIRCLED LATIN CAPITAL LETTER N
U+24DE	CIRCLED LATIN SMALL LETTER O	U+24C4	CIRCLED LATIN CAPITAL LETTER O
U+24DF	CIRCLED LATIN SMALL LETTER P	U+24C5	CIRCLED LATIN CAPITAL LETTER P
U+24E0	CIRCLED LATIN SMALL LETTER Q	U+24C6	CIRCLED LATIN CAPITAL LETTER Q
U+24E1	CIRCLED LATIN SMALL LETTER R	U+24C7	CIRCLED LATIN CAPITAL LETTER R
U+24E2	CIRCLED LATIN SMALL LETTER S	U+24C8	CIRCLED LATIN CAPITAL LETTER S
U+24E3	CIRCLED LATIN SMALL LETTER T	U+24C9	CIRCLED LATIN CAPITAL LETTER T
U+24E4	CIRCLED LATIN SMALL LETTER U	U+24CA	CIRCLED LATIN CAPITAL LETTER U
U+24E5	CIRCLED LATIN SMALL LETTER V	U+24CB	CIRCLED LATIN CAPITAL LETTER V

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+24E6	CIRCLED LATIN SMALL LETTER W	U+24CC	CIRCLED LATIN CAPITAL LETTER W
U+24E7	CIRCLED LATIN SMALL LETTER X	U+24CD	CIRCLED LATIN CAPITAL LETTER X
U+24E8	CIRCLED LATIN SMALL LETTER Y	U+24CE	CIRCLED LATIN CAPITAL LETTER Y
U+24E9	CIRCLED LATIN SMALL LETTER Z	U+24CF	CIRCLED LATIN CAPITAL LETTER Z
U+2C30	# GLAGOLITIC SMALL LETTER AZU	U+2C00	GLAGOLITIC CAPITAL LETTER AZU
U+2C31	# GLAGOLITIC SMALL LETTER BUKY	U+2C01	GLAGOLITIC CAPITAL LETTER BUKY
U+2C32	# GLAGOLITIC SMALL LETTER VEDE	U+2C02	GLAGOLITIC CAPITAL LETTER VEDE
U+2C33	# GLAGOLITIC SMALL LETTER GLAGOLI	U+2C03	GLAGOLITIC CAPITAL LETTER GLAGOLI
U+2C34	# GLAGOLITIC SMALL LETTER DOBRO	U+2C04	GLAGOLITIC CAPITAL LETTER DOBRO
U+2C35	# GLAGOLITIC SMALL LETTER YESTU	U+2C05	GLAGOLITIC CAPITAL LETTER YESTU
U+2C36	# GLAGOLITIC SMALL LETTER ZHIVETE	U+2C06	GLAGOLITIC CAPITAL LETTER ZHIVETE
U+2C37	# GLAGOLITIC SMALL LETTER DZELO	U+2C07	GLAGOLITIC CAPITAL LETTER DZELO
U+2C38	# GLAGOLITIC SMALL LETTER ZEMLJA	U+2C08	GLAGOLITIC CAPITAL LETTER ZEMLJA
U+2C39	# GLAGOLITIC SMALL LETTER IZHE	U+2C09	GLAGOLITIC CAPITAL LETTER IZHE
U+2C3A	# GLAGOLITIC SMALL LETTER INITIAL IZHE	U+2C0A	GLAGOLITIC CAPITAL LETTER INITIAL IZHE
U+2C3B	# GLAGOLITIC SMALL LETTER I	U+2C0B	GLAGOLITIC CAPITAL LETTER I
U+2C3C	# GLAGOLITIC SMALL LETTER DJERVI	U+2C0C	GLAGOLITIC CAPITAL LETTER DJERVI
U+2C3D	# GLAGOLITIC SMALL LETTER KAKO	U+2C0D	GLAGOLITIC CAPITAL LETTER KAKO
U+2C3E	# GLAGOLITIC SMALL LETTER LJUDIJE	U+2C0E	GLAGOLITIC CAPITAL LETTER LJUDIJE
U+2C3F	# GLAGOLITIC SMALL LETTER MYSLITE	U+2C0F	GLAGOLITIC CAPITAL LETTER MYSLITE
U+2C40	# GLAGOLITIC SMALL LETTER NASHI	U+2C10	GLAGOLITIC CAPITAL LETTER NASHI
U+2C41	# GLAGOLITIC SMALL LETTER ONU	U+2C11	GLAGOLITIC CAPITAL LETTER ONU

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C42	# GLAGOLITIC SMALL LETTER POKOJI	U+2C12	GLAGOLITIC CAPITAL LETTER POKOJI
U+2C43	# GLAGOLITIC SMALL LETTER RITSI	U+2C13	GLAGOLITIC CAPITAL LETTER RITSI
U+2C44	# GLAGOLITIC SMALL LETTER SLOVO	U+2C14	GLAGOLITIC CAPITAL LETTER SLOVO
U+2C45	# GLAGOLITIC SMALL LETTER TVRIDO	U+2C15	GLAGOLITIC CAPITAL LETTER TVRIDO
U+2C46	# GLAGOLITIC SMALL LETTER UKU	U+2C16	GLAGOLITIC CAPITAL LETTER UKU
U+2C47	# GLAGOLITIC SMALL LETTER FRITU	U+2C17	GLAGOLITIC CAPITAL LETTER FRITU
U+2C48	# GLAGOLITIC SMALL LETTER HERU	U+2C18	GLAGOLITIC CAPITAL LETTER HERU
U+2C49	# GLAGOLITIC SMALL LETTER OTU	U+2C19	GLAGOLITIC CAPITAL LETTER OTU
U+2C4A	GLAGOLITIC SMALL LETTER PE	U+2C1A	GLAGOLITIC CAPITAL LETTER PE
U+2C4B	GLAGOLITIC SMALL LETTER SHTA	U+2C1B	GLAGOLITIC CAPITAL LETTER SHTA
U+2C4C	GLAGOLITIC SMALL LETTER TSI	U+2C1C	GLAGOLITIC CAPITAL LETTER TSI
U+2C4D	GLAGOLITIC SMALL LETTER CHRIVI	U+2C1D	GLAGOLITIC CAPITAL LETTER CHRIVI
U+2C4E	GLAGOLITIC SMALL LETTER SHA	U+2C1E	GLAGOLITIC CAPITAL LETTER SHA
U+2C4F	GLAGOLITIC SMALL LETTER YERU	U+2C1F	GLAGOLITIC CAPITAL LETTER YERU
U+2C50	GLAGOLITIC SMALL LETTER YERI	U+2C20	GLAGOLITIC CAPITAL LETTER YERI
U+2C51	GLAGOLITIC SMALL LETTER YATI	U+2C21	GLAGOLITIC CAPITAL LETTER YATI
U+2C52	GLAGOLITIC SMALL LETTER SPIDERY HA	U+2C22	GLAGOLITIC CAPITAL LETTER SPIDERY HA
U+2C53	GLAGOLITIC SMALL LETTER YU	U+2C23	GLAGOLITIC CAPITAL LETTER YU
U+2C54	GLAGOLITIC SMALL LETTER SMALL YUS	U+2C24	GLAGOLITIC CAPITAL LETTER SMALL YUS
U+2C55	GLAGOLITIC SMALL LETTER SMALL YUS WITH TAIL	U+2C25	GLAGOLITIC CAPITAL LETTER SMALL YUS WITH TAIL
U+2C56	GLAGOLITIC SMALL LETTER YO	U+2C26	GLAGOLITIC CAPITAL LETTER YO
U+2C57	GLAGOLITIC SMALL LETTER IOTATED SMALL YUS	U+2C27	GLAGOLITIC CAPITAL LETTER IOTATED SMALL YU

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C58	GLAGOLITIC SMALL LETTER BIG YUS	U+2C28	GLAGOLITIC CAPITAL LETTER BIG YUS
U+2C59	GLAGOLITIC SMALL LETTER IOTATED BIG YUS	U+2C29	GLAGOLITIC CAPITAL LETTER IOTATED BIG YUS
U+2C5A	GLAGOLITIC SMALL LETTER FITA	U+2C2A	GLAGOLITIC CAPITAL LETTER FITA
U+2C5B	GLAGOLITIC SMALL LETTER IZHITSA	U+2C2B	GLAGOLITIC CAPITAL LETTER IZHITSA
U+2C5C	GLAGOLITIC SMALL LETTER SHTAPIC	U+2C2C	GLAGOLITIC CAPITAL LETTER SHTAPIC
U+2C5D	GLAGOLITIC SMALL LETTER TROKUTASTI A	U+2C2D	GLAGOLITIC CAPITAL LETTER TROKUTASTI A
U+2C5E	GLAGOLITIC SMALL LETTER LATINATE MYSLITE	U+2C2E	GLAGOLITIC CAPITAL LETTER LATINATE MYSLITE
U+2C61	LATIN SMALL LETTER L WITH DOUBLE BAR	U+2C60	LATIN CAPITAL LETTER L WITH DOUBLE BAR
U+2C65	LATIN SMALL LETTER A WITH STROKE	U+023A	LATIN CAPITAL LETTER A WITH STROKE
U+2C66	LATIN SMALL LETTER T WITH DIAGONAL STROKE	U+023E	LATIN CAPITAL LETTER T WITH DIAGONAL STROKE
U+2C68	LATIN SMALL LETTER H WITH DESCENDER	U+2C67	LATIN CAPITAL LETTER H WITH DESCENDER
U+2C6A	LATIN SMALL LETTER K WITH DESCENDER	U+2C69	LATIN CAPITAL LETTER K WITH DESCENDER
U+2C6C	LATIN SMALL LETTER Z WITH DESCENDER	U+2C6B	LATIN CAPITAL LETTER Z WITH DESCENDER
U+2C73	LATIN SMALL LETTER W WITH HOOK	U+2C72	LATIN CAPITAL LETTER W WITH HOOK
U+2C76	LATIN SMALL LETTER HALF H	U+2C75	LATIN CAPITAL LETTER HALF H
U+2C81	COPTIC SMALL LETTER ALFA	U+2C80	COPTIC CAPITAL LETTER ALFA
U+2C83	COPTIC SMALL LETTER VIDA	U+2C82	COPTIC CAPITAL LETTER VIDA
U+2C85	COPTIC SMALL LETTER GAMMA	U+2C84	COPTIC CAPITAL LETTER GAMMA
U+2C87	COPTIC SMALL LETTER DALDA	U+2C86	COPTIC CAPITAL LETTER DALDA
U+2C89	COPTIC SMALL LETTER EIE	U+2C88	COPTIC CAPITAL LETTER EIE
U+2C8B	COPTIC SMALL LETTER SOU	U+2C8A	COPTIC CAPITAL LETTER SOU

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C8D	COPTIC SMALL LETTER ZATA	U+2C8C	COPTIC CAPITAL LETTER ZATA
U+2C8F	COPTIC SMALL LETTER HATE	U+2C8E	COPTIC CAPITAL LETTER HATE
U+2C91	COPTIC SMALL LETTER THETHE	U+2C90	COPTIC CAPITAL LETTER THETHE
U+2C93	COPTIC SMALL LETTER IAUDA	U+2C92	COPTIC CAPITAL LETTER IAUDA
U+2C95	COPTIC SMALL LETTER KAPA	U+2C94	COPTIC CAPITAL LETTER KAPA
U+2C97	COPTIC SMALL LETTER LAULA	U+2C96	COPTIC CAPITAL LETTER LAULA
U+2C99	COPTIC SMALL LETTER MI	U+2C98	COPTIC CAPITAL LETTER MI
U+2C9B	COPTIC SMALL LETTER NI	U+2C9A	COPTIC CAPITAL LETTER NI
U+2C9D	COPTIC SMALL LETTER KSI	U+2C9C	COPTIC CAPITAL LETTER KSI
U+2C9F	COPTIC SMALL LETTER O	U+2C9E	COPTIC CAPITAL LETTER O
U+2CA1	COPTIC SMALL LETTER PI	U+2CA0	COPTIC CAPITAL LETTER PI
U+2CA3	COPTIC SMALL LETTER RO	U+2CA2	COPTIC CAPITAL LETTER RO
U+2CA5	COPTIC SMALL LETTER SIMA	U+2CA4	COPTIC CAPITAL LETTER SIMA
U+2CA7	COPTIC SMALL LETTER TAU	U+2CA6	COPTIC CAPITAL LETTER TAU
U+2CA9	COPTIC SMALL LETTER UA	U+2CA8	COPTIC CAPITAL LETTER UA
U+2CAB	COPTIC SMALL LETTER FI	U+2CAA	COPTIC CAPITAL LETTER FI
U+2CAD	COPTIC SMALL LETTER KHI	U+2CAC	COPTIC CAPITAL LETTER KHI
U+2CAF	COPTIC SMALL LETTER PSI	U+2CAE	COPTIC CAPITAL LETTER PSI
U+2CB1	COPTIC SMALL LETTER OOU	U+2CB0	COPTIC CAPITAL LETTER OOU
U+2CB3	COPTIC SMALL LETTER DIALECT- P ALEF	U+2CB2	COPTIC CAPITAL LETTER DIALECT- P ALEF
U+2CB5	COPTIC SMALL LETTER OLD COPTIC AIN	U+2CB4	COPTIC CAPITAL LETTER OLD COPTIC AIN
U+2CB7	COPTIC SMALL LETTER CRYPTOGRAMMIC EIE	U+2CB6	COPTIC CAPITAL LETTER CRYPTOGRAMMIC EIE
U+2CB9	COPTIC SMALL LETTER DIALECT- P KAPA	U+2CB8	COPTIC CAPITAL LETTER DIALECT- P KAPA
U+2CBB	COPTIC SMALL LETTER DIALECT- P NI	U+2CBA	COPTIC CAPITAL LETTER DIALECT- P NI
U+2CBD	COPTIC SMALL LETTER CRYPTOGRAMMIC NI	U+2CBC	COPTIC CAPITAL LETTER CRYPTOGRAMMIC NI

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2CBF	COPTIC SMALL LETTER OLD COPTIC OOU	U+2CBE	COPTIC CAPITAL LETTER OLD COPTIC OOU
U+2CC1	COPTIC SMALL LETTER SAMPI	U+2CC0	COPTIC CAPITAL LETTER SAMPI
U+2CC3	COPTIC SMALL LETTER CROSSED SHEI	U+2CC2	COPTIC CAPITAL LETTER CROSSED SHEI
U+2CC5	COPTIC SMALL LETTER OLD COPTIC SHEI	U+2CC4	COPTIC CAPITAL LETTER OLD COPTIC SHEI
U+2CC7	COPTIC SMALL LETTER OLD COPTIC ESH	U+2CC6	COPTIC CAPITAL LETTER OLD COPTIC ESH
U+2CC9	COPTIC SMALL LETTER AKHMIMIC KHEI	U+2CC8	COPTIC CAPITAL LETTER AKHMIMIC KHEI
U+2CCB	COPTIC SMALL LETTER DIALECT-P HORI	U+2CCA	COPTIC CAPITAL LETTER DIALECT-P HORI
U+2CCD	COPTIC SMALL LETTER OLD COPTIC HORI	U+2CCC	COPTIC CAPITAL LETTER OLD COPTIC HORI
U+2CCF	COPTIC SMALL LETTER OLD COPTIC HA	U+2CCE	COPTIC CAPITAL LETTER OLD COPTIC HA
U+2CD1	COPTIC SMALL LETTER L-SHAPED HA	U+2CD0	COPTIC CAPITAL LETTER L-SHAPED HA
U+2CD3	COPTIC SMALL LETTER OLD COPTIC HEI	U+2CD2	COPTIC CAPITAL LETTER OLD COPTIC HEI
U+2CD5	COPTIC SMALL LETTER OLD COPTIC HAT	U+2CD4	COPTIC CAPITAL LETTER OLD COPTIC HAT
U+2CD7	COPTIC SMALL LETTER OLD COPTIC GANGIA	U+2CD6	COPTIC CAPITAL LETTER OLD COPTIC GANGIA
U+2CD9	COPTIC SMALL LETTER OLD COPTIC DJA	U+2CD8	COPTIC CAPITAL LETTER OLD COPTIC DJA
U+2CDB	COPTIC SMALL LETTER OLD COPTIC SHIMA	U+2CDA	COPTIC CAPITAL LETTER OLD COPTIC SHIMA
U+2CDD	COPTIC SMALL LETTER OLD NUBIAN SHIMA	U+2CDC	COPTIC CAPITAL LETTER OLD NUBIAN SHIMA
U+2CDF	COPTIC SMALL LETTER OLD NUBIAN NGI	U+2CDE	COPTIC CAPITAL LETTER OLD NUBIAN NGI
U+2CE1	COPTIC SMALL LETTER OLD NUBIAN NYI	U+2CE0	COPTIC CAPITAL LETTER OLD NUBIAN NYI

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2CE3	COPTIC SMALL LETTER OLD NUBIAN WAU	U+2CE2	COPTIC CAPITAL LETTER OLD NUBIAN WAU
U+2CEC	COPTIC SMALL LETTER CRYPTOGRAMMIC SHEI	U+2CEB	COPTIC CAPITAL LETTER CRYPTOGRAMMIC SHEI
U+2CEE	COPTIC SMALL LETTER CRYPTOGRAMMIC GANGIA	U+2CED	COPTIC CAPITAL LETTER CRYPTOGRAMMIC GANGIA
U+2D00	GEORGIAN SMALL LETTER AN	U+10A0	GEORGIAN CAPITAL LETTER AN
U+2D01	GEORGIAN SMALL LETTER BAN	U+10A1	GEORGIAN CAPITAL LETTER BAN
U+2D02	GEORGIAN SMALL LETTER GAN	U+10A2	GEORGIAN CAPITAL LETTER GAN
U+2D03	GEORGIAN SMALL LETTER DON	U+10A3	GEORGIAN CAPITAL LETTER DON
U+2D04	GEORGIAN SMALL LETTER EN	U+10A4	GEORGIAN CAPITAL LETTER EN
U+2D05	GEORGIAN SMALL LETTER VIN	U+10A5	GEORGIAN CAPITAL LETTER VIN
U+2D06	GEORGIAN SMALL LETTER ZEN	U+10A6	GEORGIAN CAPITAL LETTER ZEN
U+2D07	GEORGIAN SMALL LETTER TAN	U+10A7	GEORGIAN CAPITAL LETTER TAN
U+2D08	GEORGIAN SMALL LETTER IN	U+10A8	GEORGIAN CAPITAL LETTER IN
U+2D09	GEORGIAN SMALL LETTER KAN	U+10A9	GEORGIAN CAPITAL LETTER KAN
U+2D0A	GEORGIAN SMALL LETTER LAS	U+10AA	GEORGIAN CAPITAL LETTER LAS
U+2D0B	GEORGIAN SMALL LETTER MAN	U+10AB	GEORGIAN CAPITAL LETTER MAN
U+2D0C	GEORGIAN SMALL LETTER NAR	U+10AC	GEORGIAN CAPITAL LETTER NAR
U+2D0D	GEORGIAN SMALL LETTER ON	U+10AD	GEORGIAN CAPITAL LETTER ON
U+2D0E	GEORGIAN SMALL LETTER PAR	U+10AE	GEORGIAN CAPITAL LETTER PAR
U+2D0F	GEORGIAN SMALL LETTER ZHAR	U+10AF	GEORGIAN CAPITAL LETTER ZHAR
U+2D10	GEORGIAN SMALL LETTER RAE	U+10B0	GEORGIAN CAPITAL LETTER RAE
U+2D11	GEORGIAN SMALL LETTER SAN	U+10B1	GEORGIAN CAPITAL LETTER SAN
U+2D12	GEORGIAN SMALL LETTER TAR	U+10B2	GEORGIAN CAPITAL LETTER TAR
U+2D13	GEORGIAN SMALL LETTER UN	U+10B3	GEORGIAN CAPITAL LETTER UN
U+2D14	GEORGIAN SMALL LETTER PHAR	U+10B4	GEORGIAN CAPITAL LETTER PHAR
U+2D15	GEORGIAN SMALL LETTER KHAR	0x10B5	GEORGIAN CAPITAL LETTER KHAR
U+2D16	GEORGIAN SMALL LETTER GHAN	U+10B6	GEORGIAN CAPITAL LETTER GHAN

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2D17	GEORGIAN SMALL LETTER QAR	U+10B7	GEORGIAN CAPITAL LETTER QAR
U+2D18	GEORGIAN SMALL LETTER SHIN	U+10B8	GEORGIAN CAPITAL LETTER SHIN
U+2D19	GEORGIAN SMALL LETTER CHIN	U+10B9	GEORGIAN CAPITAL LETTER CHIN
U+2D1A	GEORGIAN SMALL LETTER CAN	U+10BA	GEORGIAN CAPITAL LETTER CAN
U+2D1B	GEORGIAN SMALL LETTER JIL	U+10BB	GEORGIAN CAPITAL LETTER JIL
U+2D1C	GEORGIAN SMALL LETTER CIL	U+10BC	GEORGIAN CAPITAL LETTER CIL
U+2D1D	GEORGIAN SMALL LETTER CHAR	U+10BD	GEORGIAN CAPITAL LETTER CHAR
U+2D1E	GEORGIAN SMALL LETTER XAN	U+10BE	GEORGIAN CAPITAL LETTER XAN
U+2D1F	GEORGIAN SMALL LETTER JHAN	U+10BF	GEORGIAN CAPITAL LETTER JHAN
U+2D20	GEORGIAN SMALL LETTER HAE	U+10C0	GEORGIAN CAPITAL LETTER HAE
U+2D21	GEORGIAN SMALL LETTER HE	U+10C1	GEORGIAN CAPITAL LETTER HE
U+2D22	GEORGIAN SMALL LETTER HIE	U+10C2	GEORGIAN CAPITAL LETTER HIE
U+2D23	GEORGIAN SMALL LETTER WE	U+10C3	GEORGIAN CAPITAL LETTER WE
U+2D24	GEORGIAN SMALL LETTER HAR	U+10C4	GEORGIAN CAPITAL LETTER HAR
U+2D25	GEORGIAN SMALL LETTER HOE	U+10C5	GEORGIAN CAPITAL LETTER HOE
U+A641	CYRILLIC SMALL LETTER ZEMLYA	U+A640	CYRILLIC CAPITAL LETTER ZEMLYA
U+A643	CYRILLIC SMALL LETTER DZELO	U+A642	CYRILLIC CAPITAL LETTER DZELO
U+A645	CYRILLIC SMALL LETTER REVERSED DZE	U+A644	CYRILLIC CAPITAL LETTER REVERSED DZE
U+A647	CYRILLIC SMALL LETTER IOTA	U+A646	CYRILLIC CAPITAL LETTER IOTA
U+A649	CYRILLIC SMALL LETTER DJERV	U+A648	CYRILLIC CAPITAL LETTER DJERV
U+A64B	CYRILLIC SMALL LETTER MONOGRAPH UK	U+A64A	CYRILLIC CAPITAL LETTER MONOGRAPH UK
U+A64D	CYRILLIC SMALL LETTER BROAD OMEGA	U+A64C	CYRILLIC CAPITAL LETTER BROAD OMEGA
U+A64F	CYRILLIC SMALL LETTER NEUTRAL YER	U+A64E	CYRILLIC CAPITAL LETTER NEUTRAL YER
U+A651	CYRILLIC SMALL LETTER YERU WITH BACK YER	U+A650	CYRILLIC CAPITAL LETTER YERU WITH BACK YER
U+A653	CYRILLIC SMALL LETTER IOTIFIED YAT	U+A652	CYRILLIC CAPITAL LETTER IOTIFIED YAT

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A655	CYRILLIC SMALL LETTER REVERSED YU	U+A654	CYRILLIC CAPITAL LETTER REVERSED YU
U+A657	CYRILLIC SMALL LETTER IOTIFIED A	U+A656	CYRILLIC CAPITAL LETTER IOTIFIED A
U+A659	CYRILLIC SMALL LETTER CLOSED LITTLE YUS	U+A658	CYRILLIC CAPITAL LETTER CLOSED LITTLE YUS
U+A65B	CYRILLIC SMALL LETTER BLENDED YUS	U+A65A	CYRILLIC CAPITAL LETTER BLENDED YUS
U+A65D	CYRILLIC SMALL LETTER IOTIFIED CLOSED LITTLE YUS	U+A65C	CYRILLIC CAPITAL LETTER IOTIFIED CLOSED LITTLE YUS
U+A65F	CYRILLIC SMALL LETTER YN &	0xA65E	CYRILLIC CAPITAL LETTER YN
U+A661	CYRILLIC SMALL LETTER REVERSED TSE	0xA660	CYRILLIC CAPITAL LETTER REVERSED TSE
U+A663	CYRILLIC SMALL LETTER SOFT DE &	0xA662	CYRILLIC CAPITAL LETTER SOFT DE
U+A665	CYRILLIC SMALL LETTER SOFT EL &	0xA664	CYRILLIC CAPITAL LETTER SOFT EL
U+A667	CYRILLIC SMALL LETTER SOFT EM &	0xA666	CYRILLIC CAPITAL LETTER SOFT EM
U+A669	CYRILLIC SMALL LETTER MONOCULAR O &	0xA668	CYRILLIC CAPITAL LETTER MONOCULAR O
U+A66B	CYRILLIC SMALL LETTER BINOCULAR O	0xA66A	CYRILLIC CAPITAL LETTER BINOCULAR O
U+A66D	CYRILLIC SMALL LETTER DOUBLE MONOCULAR O	0xA66C	CYRILLIC CAPITAL LETTER DOUBLE MONOCULAR O
U+A681	CYRILLIC SMALL LETTER DWE &	0xA680	CYRILLIC CAPITAL LETTER DWE
U+A683	CYRILLIC SMALL LETTER DZWE &	0xA682	CYRILLIC CAPITAL LETTER DZWE
U+A685	CYRILLIC SMALL LETTER ZHWE &	0xA684	CYRILLIC CAPITAL LETTER ZHWE
U+A687	CYRILLIC SMALL LETTER CCHE &	0xA686	CYRILLIC CAPITAL LETTER CCHE
U+A689	CYRILLIC SMALL LETTER DZZE &	0xA688	CYRILLIC CAPITAL LETTER DZZE
U+A68B	CYRILLIC SMALL LETTER TE WITH MIDDLE HOOK &	0xA68A	CYRILLIC CAPITAL LETTER TE WITH MIDDLE HOOK
U+A68D	CYRILLIC SMALL LETTER TWE &	0xA68C	CYRILLIC CAPITAL LETTER TWE
U+A68F	CYRILLIC SMALL LETTER TSWE &	0xA68E	CYRILLIC CAPITAL LETTER TSWE
U+A691	CYRILLIC SMALL LETTER TSSE	0xA690	CYRILLIC CAPITAL LETTER TSSE

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A693	CYRILLIC SMALL LETTER TCHE	0xA692	CYRILLIC CAPITAL LETTER TCHE
U+A695	CYRILLIC SMALL LETTER HWE	0xA694	CYRILLIC CAPITAL LETTER HWE
U+A697	CYRILLIC SMALL LETTER SHWE	0xA696	CYRILLIC CAPITAL LETTER SHWE
U+A723	LATIN SMALL LETTER EGYPTOLOGICAL ALEF	0xA722	LATIN CAPITAL LETTER EGYPTOLOGICAL ALEF
U+A725	LATIN SMALL LETTER EGYPTOLOGICAL AIN	0xA724	LATIN CAPITAL LETTER EGYPTOLOGICAL AIN
U+A727	LATIN SMALL LETTER HENG	0xA726	LATIN CAPITAL LETTER HENG
U+A729	LATIN SMALL LETTER TZ	0xA728	LATIN CAPITAL LETTER TZ
U+A72B	LATIN SMALL LETTER TRESILLO	0xA72A	LATIN CAPITAL LETTER TRESILLO
U+A72D	LATIN SMALL LETTER CUATRILLO	0xA72C	LATIN CAPITAL LETTER CUATRILLO
U+A72F	LATIN SMALL LETTER CUATRILLO WITH COMMA	0xA72E	LATIN CAPITAL LETTER CUATRILLO WITH COMMA
U+A733	LATIN SMALL LETTER AA	0xA732	LATIN CAPITAL LETTER AA
U+A735	LATIN SMALL LETTER AO	0xA734	LATIN CAPITAL LETTER AO
U+A737	LATIN SMALL LETTER AU	0xA736	LATIN CAPITAL LETTER AU
U+A739	LATIN SMALL LETTER AV	0xA738	LATIN CAPITAL LETTER AV
U+A73B	LATIN SMALL LETTER AV WITH HORIZONTAL BAR	0xA73A	LATIN CAPITAL LETTER AV WITH HORIZONTAL BAR
U+A73D	LATIN SMALL LETTER AY	0xA73C	LATIN CAPITAL LETTER AY
U+A73F	LATIN SMALL LETTER REVERSED C WITH DOT	0xA73E	LATIN CAPITAL LETTER REVERSED C WITH DOT
U+A741	LATIN SMALL LETTER K WITH STROKE &	0xA740	LATIN CAPITAL LETTER K WITH STROKE
U+A743	LATIN SMALL LETTER K WITH DIAGONAL STROKE	0xA742	LATIN CAPITAL LETTER K WITH DIAGONAL STROKE
U+A745	LATIN SMALL LETTER K WITH STROKE AND DIAGONAL STROKE	0xA744	LATIN CAPITAL LETTER K WITH STROKE AND DIAGONAL STROKE
U+A747	LATIN SMALL LETTER BROKEN L	0xA746	LATIN CAPITAL LETTER BROKEN L
U+A749	LATIN SMALL LETTER L WITH HIGH STROKE	0xA748	LATIN CAPITAL LETTER L WITH HIGH STROKE
U+A74B	LATIN SMALL LETTER O WITH LONG STROKE OVERLAY	0xA74A	LATIN CAPITAL LETTER O WITH LONG STROKE OVERLAY

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A74D	LATIN SMALL LETTER O WITH LOOP	0xA74C	LATIN CAPITAL LETTER O WITH LOOP
U+A74F	LATIN SMALL LETTER OO	0xA74E	LATIN CAPITAL LETTER OO
U+A751	LATIN SMALL LETTER P WITH STROKE THROUGH DESCENDER	0xA750	LATIN CAPITAL LETTER P WITH STROKE THROUGH DESCENDER
U+A753	LATIN SMALL LETTER P WITH FLOURISH	0xA752	LATIN CAPITAL LETTER P WITH FLOURISH
U+A755	LATIN SMALL LETTER P WITH SQUIRREL TAIL	0xA754	LATIN CAPITAL LETTER P WITH SQUIRREL TAIL
U+A757	LATIN SMALL LETTER Q WITH STROKE THROUGH DESCENDER	0xA756	LATIN CAPITAL LETTER Q WITH STROKE THROUGH DESCENDER
U+A759	LATIN SMALL LETTER Q WITH DIAGONAL STROKE	0xA758	LATIN CAPITAL LETTER Q WITH DIAGONAL STROKE
U+A75B	LATIN SMALL LETTER R ROTUNDA	0xA75A	LATIN CAPITAL LETTER R ROTUNDA
U+A75D	LATIN SMALL LETTER RUM ROTUNDA	0xA75C	LATIN CAPITAL LETTER RUM ROTUNDA
U+A75F	LATIN SMALL LETTER V WITH DIAGONAL STROKE	0xA75E	LATIN CAPITAL LETTER V WITH DIAGONAL STROKE
U+A761	LATIN SMALL LETTER VY	0xA760	LATIN CAPITAL LETTER VY
U+A763	LATIN SMALL LETTER VISIGOTHIC Z &	0xA762	LATIN CAPITAL LETTER VISIGOTHIC Z
U+A765	LATIN SMALL LETTER THORN WITH STROKE	0xA764	LATIN CAPITAL LETTER THORN WITH STROKE
U+A767	LATIN SMALL LETTER THORN WITH STROKE THROUGH DESCENDER	0xA766	LATIN CAPITAL LETTER THORN WITH STROKE THROUGH DESCENDER
U+A769	LATIN SMALL LETTER VEND	0xA768	LATIN CAPITAL LETTER VEND
U+A76B	LATIN SMALL LETTER ET	0xA76A	LATIN CAPITAL LETTER ET
U+A76D	LATIN SMALL LETTER IS &	0xA76C	LATIN CAPITAL LETTER IS
U+A76F	LATIN SMALL LETTER CON &	0xA76E	LATIN CAPITAL LETTER CON
U+A77A	LATIN SMALL LETTER INSULAR D	0xA779	LATIN CAPITAL LETTER INSULAR D
U+A77C	LATIN SMALL LETTER INSULAR F	0xA77B	LATIN CAPITAL LETTER INSULAR F
U+A77F	LATIN SMALL LETTER TURNED INSULAR G	0xA77E	LATIN CAPITAL LETTER TURNED INSULAR G

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A781	LATIN SMALL LETTER TURNED L	0xA780	LATIN CAPITAL LETTER TURNED L
U+A783	LATIN SMALL LETTER INSULAR R	0xA782	LATIN CAPITAL LETTER INSULAR R
U+A785	LATIN SMALL LETTER INSULAR S	0xA784	LATIN CAPITAL LETTER INSULAR S
U+A787	LATIN SMALL LETTER INSULAR T	0xA786	LATIN CAPITAL LETTER INSULAR T
U+A78C	LATIN SMALL LETTER SALTILLO	0xA78B	LATIN CAPITAL LETTER SALTILLO
U+A791	LATIN SMALL LETTER N WITH DESCENDER	0xA790	LATIN CAPITAL LETTER N WITH DESCENDER
U+A7A1	LATIN SMALL LETTER G WITH OBLIQUE STROKE	0xA7A0	LATIN CAPITAL LETTER G WITH OBLIQUE STROKE
U+A7A3	LATIN SMALL LETTER K WITH OBLIQUE STROKE	U+A7A2	LATIN CAPITAL LETTER K WITH OBLIQUE STROKE
U+A7A5	LATIN SMALL LETTER N WITH OBLIQUE STROKE	U+A7A4	LATIN CAPITAL LETTER N WITH OBLIQUE STROKE
U+A7A7	LATIN SMALL LETTER R WITH OBLIQUE STROKE	U+A7A6	LATIN CAPITAL LETTER R WITH OBLIQUE STROKE
U+A7A9	LATIN SMALL LETTER S WITH OBLIQUE STROKE	U+A7A8	LATIN CAPITAL LETTER S WITH OBLIQUE STROKE
U+FF41	FULLWIDTH LATIN SMALL LETTER A	U+FF21	FULLWIDTH LATIN CAPITAL LETTER A
U+FF42	FULLWIDTH LATIN SMALL LETTER B	U+FF22	FULLWIDTH LATIN CAPITAL LETTER B
U+FF43	FULLWIDTH LATIN SMALL LETTER C	U+FF23	FULLWIDTH LATIN CAPITAL LETTER C
U+FF44	FULLWIDTH LATIN SMALL LETTER D	U+FF24	FULLWIDTH LATIN CAPITAL LETTER D
U+FF45	FULLWIDTH LATIN SMALL LETTER E	U+FF25	FULLWIDTH LATIN CAPITAL LETTER E
U+FF46	FULLWIDTH LATIN SMALL LETTER F	U+FF26	FULLWIDTH LATIN CAPITAL LETTER F
U+FF47	FULLWIDTH LATIN SMALL LETTER G	U+FF27	FULLWIDTH LATIN CAPITAL LETTER G
U+FF48	FULLWIDTH LATIN SMALL LETTER H	U+FF28	FULLWIDTH LATIN CAPITAL LETTER H
U+FF49	FULLWIDTH LATIN SMALL LETTER I	U+FF29	FULLWIDTH LATIN CAPITAL LETTER I

UNICODE Lowercase Input		UNICODE Uppercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF4A	FULLWIDTH LATIN SMALL LETTER J	U+FF2A	FULLWIDTH LATIN CAPITAL LETTER J
U+FF4B	FULLWIDTH LATIN SMALL LETTER K	U+FF2B	FULLWIDTH LATIN CAPITAL LETTER K
U+FF4C	FULLWIDTH LATIN SMALL LETTER L	U+FF2C	FULLWIDTH LATIN CAPITAL LETTER L
U+FF4D	FULLWIDTH LATIN SMALL LETTER M	U+FF2D	FULLWIDTH LATIN CAPITAL LETTER M
U+FF4E	FULLWIDTH LATIN SMALL LETTER N	U+FF2E	FULLWIDTH LATIN CAPITAL LETTER N
U+FF4F	FULLWIDTH LATIN SMALL LETTER O	U+FF2F	FULLWIDTH LATIN CAPITAL LETTER O
U+FF50	FULLWIDTH LATIN SMALL LETTER P	U+FF30	FULLWIDTH LATIN CAPITAL LETTER P
U+FF51	FULLWIDTH LATIN SMALL LETTER Q	U+FF31	FULLWIDTH LATIN CAPITAL LETTER Q
U+FF52	FULLWIDTH LATIN SMALL LETTER R	U+FF32	FULLWIDTH LATIN CAPITAL LETTER R
U+FF53	FULLWIDTH LATIN SMALL LETTER S	U+FF33	FULLWIDTH LATIN CAPITAL LETTER S
U+FF54	FULLWIDTH LATIN SMALL LETTER T	U+FF34	FULLWIDTH LATIN CAPITAL LETTER T
U+FF55	FULLWIDTH LATIN SMALL LETTER U	U+FF35	FULLWIDTH LATIN CAPITAL LETTER U
U+FF56	FULLWIDTH LATIN SMALL LETTER V	U+FF36	FULLWIDTH LATIN CAPITAL LETTER V
U+FF57	FULLWIDTH LATIN SMALL LETTER W	U+FF37	FULLWIDTH LATIN CAPITAL LETTER W
U+FF58	FULLWIDTH LATIN SMALL LETTER X	U+FF38	FULLWIDTH LATIN CAPITAL LETTER X
U+FF59	FULLWIDTH LATIN SMALL LETTER Y	U+FF39	FULLWIDTH LATIN CAPITAL LETTER Y
U+FF5A	FULLWIDTH LATIN SMALL LETTER Z	U+FF3A	FULLWIDTH LATIN CAPITAL LETTER Z

UNICODE Uppercase to Lowercase

The following table maps UNICODE Uppercase characters to their Lowercase equivalents. Teradata Database uses this mapping when you call the LOWER function on character strings that are defined with the UNICODE server character set.

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0041	LATIN CAPITAL LETTER A	U+0061	LATIN SMALL LETTER A
U+0042	LATIN CAPITAL LETTER B	U+0062	LATIN SMALL LETTER B
U+0043	LATIN CAPITAL LETTER C	U+0063	LATIN SMALL LETTER C
U+0044	LATIN CAPITAL LETTER D	U+0064	LATIN SMALL LETTER D
U+0045	LATIN CAPITAL LETTER E	U+0065	LATIN SMALL LETTER E
U+0046	LATIN CAPITAL LETTER F	U+0066	LATIN SMALL LETTER F
U+0047	LATIN CAPITAL LETTER G	U+0067	LATIN SMALL LETTER G
U+0048	LATIN CAPITAL LETTER H	U+0068	LATIN SMALL LETTER H
U+0049	LATIN CAPITAL LETTER I	U+0069	LATIN SMALL LETTER I
U+004A	LATIN CAPITAL LETTER J	U+006A	LATIN SMALL LETTER J
U+004B	LATIN CAPITAL LETTER K	U+006B	LATIN SMALL LETTER K
U+004C	LATIN CAPITAL LETTER L	U+006C	LATIN SMALL LETTER L
U+004D	LATIN CAPITAL LETTER M	U+006D	LATIN SMALL LETTER M
U+004E	LATIN CAPITAL LETTER N	U+006E	LATIN SMALL LETTER N
U+004F	LATIN CAPITAL LETTER O	U+006F	LATIN SMALL LETTER O
U+0050	LATIN CAPITAL LETTER P	U+0070	LATIN SMALL LETTER P
U+0051	LATIN CAPITAL LETTER Q	U+0071	LATIN SMALL LETTER Q
U+0052	LATIN CAPITAL LETTER R	U+0072	LATIN SMALL LETTER R
U+0053	LATIN CAPITAL LETTER S	U+0073	LATIN SMALL LETTER S
U+0054	LATIN CAPITAL LETTER T	U+0074	LATIN SMALL LETTER T
U+0055	LATIN CAPITAL LETTER U	U+0075	LATIN SMALL LETTER U
U+0056	LATIN CAPITAL LETTER V	U+0076	LATIN SMALL LETTER V
U+0057	LATIN CAPITAL LETTER W	U+0077	LATIN SMALL LETTER W
U+0058	LATIN CAPITAL LETTER X	U+0078	LATIN SMALL LETTER X

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0059	LATIN CAPITAL LETTER Y	U+0079	LATIN SMALL LETTER Y
U+005A	LATIN CAPITAL LETTER Z	U+007A	LATIN SMALL LETTER Z
U+00C0	LATIN CAPITAL LETTER A WITH GRAVE	U+00E0	LATIN SMALL LETTER A WITH GRAVE
U+00C1	LATIN CAPITAL LETTER A WITH ACUTE	U+00E1	LATIN SMALL LETTER A WITH ACUTE
U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX	U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX
U+00C3	LATIN CAPITAL LETTER A WITH TILDE	U+00E3	LATIN SMALL LETTER A WITH TILDE
U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE	U+00E5	LATIN SMALL LETTER A WITH RING ABOVE
U+00C6	LATIN CAPITAL LETTER AE	U+00E6	LATIN SMALL LETTER AE
U+00C7	LATIN CAPITAL LETTER C WITH CEDILLA	U+00E7	LATIN SMALL LETTER C WITH CEDILLA
U+00C8	LATIN CAPITAL LETTER E WITH GRAVE	U+00E8	LATIN SMALL LETTER E WITH GRAVE
U+00C9	LATIN CAPITAL LETTER E WITH ACUTE	U+00E9	LATIN SMALL LETTER E WITH ACUTE
U+00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX	U+00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX
U+00CB	LATIN CAPITAL LETTER E WITH DIAERESIS	U+00EB	LATIN SMALL LETTER E WITH DIAERESIS
U+00CC	LATIN CAPITAL LETTER I WITH GRAVE	U+00EC	LATIN SMALL LETTER I WITH GRAVE
U+00CD	LATIN CAPITAL LETTER I WITH ACUTE	U+00ED	LATIN SMALL LETTER I WITH ACUTE
U+00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX	U+00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX
U+00CF	LATIN CAPITAL LETTER I WITH DIAERESIS	U+00EF	LATIN SMALL LETTER I WITH DIAERESIS
U+00D0	LATIN CAPITAL LETTER ETH	U+00F0	LATIN SMALL LETTER ETH
U+00D1	LATIN CAPITAL LETTER N WITH TILDE	U+00F1	LATIN SMALL LETTER N WITH TILDE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+00D2	LATIN CAPITAL LETTER O WITH GRAVE	U+00F2	LATIN SMALL LETTER O WITH GRAVE
U+00D3	LATIN CAPITAL LETTER O WITH ACUTE	U+00F3	LATIN SMALL LETTER O WITH ACUTE
U+00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX	U+00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX
U+00D5	LATIN CAPITAL LETTER O WITH TILDE	U+00F5	LATIN SMALL LETTER O WITH TILDE
U+00D6	LATIN CAPITAL LETTER O WITH DIAERESIS	U+00F6	LATIN SMALL LETTER O WITH DIAERESIS
U+00D8	LATIN CAPITAL LETTER O WITH STROKE	U+00F8	LATIN SMALL LETTER O WITH STROKE
U+00D9	LATIN CAPITAL LETTER U WITH GRAVE	U+00F9	LATIN SMALL LETTER U WITH GRAVE
U+00DA	LATIN CAPITAL LETTER U WITH ACUTE	U+00FA	LATIN SMALL LETTER U WITH ACUTE
U+00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX	U+00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX
U+00DC	LATIN CAPITAL LETTER U WITH DIAERESIS	U+00FC	LATIN SMALL LETTER U WITH DIAERESIS
U+00DD	LATIN CAPITAL LETTER Y WITH ACUTE	U+00FD	LATIN SMALL LETTER Y WITH ACUTE
U+00DE	LATIN CAPITAL LETTER THORN	U+00FE	LATIN SMALL LETTER THORN
U+0100	LATIN CAPITAL LETTER A WITH MACRON	U+0101	LATIN SMALL LETTER A WITH MACRON
U+0102	LATIN CAPITAL LETTER A WITH BREVE	U+0103	LATIN SMALL LETTER A WITH BREVE
U+0104	LATIN CAPITAL LETTER A WITH OGONEK	U+0105	LATIN SMALL LETTER A WITH OGONEK
U+0106	LATIN CAPITAL LETTER C WITH ACUTE	U+0107	LATIN SMALL LETTER C WITH ACUTE
U+0108	LATIN CAPITAL LETTER C WITH CIRCUMFLEX	U+0109	LATIN SMALL LETTER C WITH CIRCUMFLEX
U+010A	LATIN CAPITAL LETTER C WITH DOT ABOVE	U+010B	LATIN SMALL LETTER C WITH DOT ABOVE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+010C	LATIN CAPITAL LETTER C WITH CARON	U+010D	LATIN SMALL LETTER C WITH CARON
U+010E	LATIN CAPITAL LETTER D WITH CARON	U+010F	LATIN SMALL LETTER D WITH CARON
U+0110	LATIN CAPITAL LETTER D WITH STROKE	U+0111	LATIN SMALL LETTER D WITH STROKE
U+0112	LATIN CAPITAL LETTER E WITH MACRON	U+0113	LATIN SMALL LETTER E WITH MACRON
U+0114	LATIN CAPITAL LETTER E WITH BREVE	U+0115	LATIN SMALL LETTER E WITH BREVE
U+0116	LATIN CAPITAL LETTER E WITH DOT ABOVE	U+0117	LATIN SMALL LETTER E WITH DOT ABOVE
U+0118	LATIN CAPITAL LETTER E WITH OGONEK	U+0119	LATIN SMALL LETTER E WITH OGONEK
U+011A	LATIN CAPITAL LETTER E WITH CARON	U+011B	LATIN SMALL LETTER E WITH CARON
U+011C	LATIN CAPITAL LETTER G WITH CIRCUMFLEX	U+011D	LATIN SMALL LETTER G WITH CIRCUMFLEX
U+011E	LATIN CAPITAL LETTER G WITH BREVE	U+011F	LATIN SMALL LETTER G WITH BREVE
U+0120	LATIN CAPITAL LETTER G WITH DOT ABOVE	U+0121	LATIN SMALL LETTER G WITH DOT ABOVE
U+0122	LATIN CAPITAL LETTER G WITH CEDILLA	U+0123	LATIN SMALL LETTER G WITH CEDILLA
U+0124	LATIN CAPITAL LETTER H WITH CIRCUMFLEX	U+0125	LATIN SMALL LETTER H WITH CIRCUMFLEX
U+0126	LATIN CAPITAL LETTER H WITH STROKE	U+0127	LATIN SMALL LETTER H WITH STROKE
U+0128	LATIN CAPITAL LETTER I WITH TILDE	U+0129	LATIN SMALL LETTER I WITH TILDE
U+012A	LATIN CAPITAL LETTER I WITH MACRON	U+012B	LATIN SMALL LETTER I WITH MACRON
U+012C	LATIN CAPITAL LETTER I WITH BREVE	U+012D	LATIN SMALL LETTER I WITH BREVE
U+012E	LATIN CAPITAL LETTER I WITH OGONEK	U+012F	LATIN SMALL LETTER I WITH OGONEK

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0130	LATIN CAPITAL LETTER I WITH DOT ABOVE	U+0069	LATIN SMALL LETTER I
U+0132	LATIN CAPITAL LIGATURE IJ	U+0133	LATIN SMALL LIGATURE IJ
U+0134	LATIN CAPITAL LETTER J WITH CIRCUMFLEX	U+0135	LATIN SMALL LETTER J WITH CIRCUMFLEX
U+0136	LATIN CAPITAL LETTER K WITH CEDILLA	U+0137	LATIN SMALL LETTER K WITH CEDILLA
U+0139	LATIN CAPITAL LETTER L WITH ACUTE	U+013A	LATIN SMALL LETTER L WITH ACUTE
U+013B	LATIN CAPITAL LETTER L WITH CEDILLA	U+013C	LATIN SMALL LETTER L WITH CEDILLA
U+013D	LATIN CAPITAL LETTER L WITH CARON	U+013E	LATIN SMALL LETTER L WITH CARON
U+013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT	U+0140	LATIN SMALL LETTER L WITH MIDDLE DOT
U+0141	LATIN CAPITAL LETTER L WITH STROKE	U+0142	LATIN SMALL LETTER L WITH STROKE
U+0143	LATIN CAPITAL LETTER N WITH ACUTE	U+0144	LATIN SMALL LETTER N WITH ACUTE
U+0145	LATIN CAPITAL LETTER N WITH CEDILLA	U+0146	LATIN SMALL LETTER N WITH CEDILLA
U+0147	LATIN CAPITAL LETTER N WITH CARON	U+0148	LATIN SMALL LETTER N WITH CARON
U+014A	LATIN CAPITAL LETTER ENG	U+014B	LATIN SMALL LETTER ENG
U+014C	LATIN CAPITAL LETTER O WITH MACRON	U+014D	LATIN SMALL LETTER O WITH MACRON
U+014E	LATIN CAPITAL LETTER O WITH BREVE	U+014F	LATIN SMALL LETTER O WITH BREVE
U+0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE	U+0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE
U+0152	LATIN CAPITAL LIGATURE OE	U+0153	LATIN SMALL LIGATURE OE
U+0154	LATIN CAPITAL LETTER R WITH ACUTE	U+0155	LATIN SMALL LETTER R WITH ACUTE
U+0156	LATIN CAPITAL LETTER R WITH CEDILLA	U+0157	LATIN SMALL LETTER R WITH CEDILLA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0158	LATIN CAPITAL LETTER R WITH CARON	U+0159	LATIN SMALL LETTER R WITH CARON
U+015A	LATIN CAPITAL LETTER S WITH ACUTE	U+015B	LATIN SMALL LETTER S WITH ACUTE
U+015C	LATIN CAPITAL LETTER S WITH CIRCUMFLEX	U+015D	LATIN SMALL LETTER S WITH CIRCUMFLEX
U+015E	LATIN CAPITAL LETTER S WITH CEDILLA	U+015F	LATIN SMALL LETTER S WITH CEDILLA
U+0160	LATIN CAPITAL LETTER S WITH CARON	U+0161	LATIN SMALL LETTER S WITH CARON
U+0162	LATIN CAPITAL LETTER T WITH CEDILLA	U+0163	LATIN SMALL LETTER T WITH CEDILLA
U+0164	LATIN CAPITAL LETTER T WITH CARON	U+0165	LATIN SMALL LETTER T WITH CARON
U+0166	LATIN CAPITAL LETTER T WITH STROKE	U+0167	LATIN SMALL LETTER T WITH STROKE
U+0168	LATIN CAPITAL LETTER U WITH TILDE	U+0169	LATIN SMALL LETTER U WITH TILDE
U+016A	LATIN CAPITAL LETTER U WITH MACRON	U+016B	LATIN SMALL LETTER U WITH MACRON
U+016C	LATIN CAPITAL LETTER U WITH BREVE	U+016D	LATIN SMALL LETTER U WITH BREVE
U+016E	LATIN CAPITAL LETTER U WITH RING ABOVE	U+016F	LATIN SMALL LETTER U WITH RING ABOVE
U+0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE	U+0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE
U+0172	LATIN CAPITAL LETTER U WITH OGONEK	U+0173	LATIN SMALL LETTER U WITH OGONEK
U+0174	LATIN CAPITAL LETTER W WITH CIRCUMFLEX	U+0175	LATIN SMALL LETTER W WITH CIRCUMFLEX
U+0176	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX	U+0177	LATIN SMALL LETTER Y WITH CIRCUMFLEX
U+0178	LATIN CAPITAL LETTER Y WITH DIAERESIS	U+00FF	LATIN SMALL LETTER Y WITH DIAERESIS
U+0179	LATIN CAPITAL LETTER Z WITH ACUTE	U+017A	LATIN SMALL LETTER Z WITH ACUTE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+017B	LATIN CAPITAL LETTER Z WITH DOT ABOVE	U+017C	LATIN SMALL LETTER Z WITH DOT ABOVE
U+017D	LATIN CAPITAL LETTER Z WITH CARON	U+017E	LATIN SMALL LETTER Z WITH CARON
U+0181	LATIN CAPITAL LETTER B WITH HOOK	U+0253	LATIN SMALL LETTER B WITH HOOK
U+0182	LATIN CAPITAL LETTER B WITH TOPBAR	U+0183	LATIN SMALL LETTER B WITH TOPBAR
U+0184	LATIN CAPITAL LETTER TONE SIX	U+0185	LATIN SMALL LETTER TONE SIX
U+0186	LATIN CAPITAL LETTER OPEN O	U+0254	LATIN SMALL LETTER OPEN O
U+0187	LATIN CAPITAL LETTER C WITH HOOK	U+0188	LATIN SMALL LETTER C WITH HOOK
U+0189	LATIN CAPITAL LETTER AFRICAN D	U+0256	LATIN SMALL LETTER D WITH TAIL
U+018A	LATIN CAPITAL LETTER D WITH HOOK	U+0257	LATIN SMALL LETTER D WITH HOOK
U+018B	LATIN CAPITAL LETTER D WITH TOPBAR	U+018C	LATIN SMALL LETTER D WITH TOPBAR
U+018E	LATIN CAPITAL LETTER REVERSED E	U+01DD	LATIN SMALL LETTER TURNED E
U+018F	LATIN CAPITAL LETTER SCHWA	U+0259	LATIN SMALL LETTER SCHWA
U+0190	LATIN CAPITAL LETTER OPEN E	U+025B	LATIN SMALL LETTER OPEN E
U+0191	LATIN CAPITAL LETTER F WITH HOOK	U+0192	LATIN SMALL LETTER F WITH HOOK
U+0193	LATIN CAPITAL LETTER G WITH HOOK	U+0260	LATIN SMALL LETTER G WITH HOOK
U+0194	LATIN CAPITAL LETTER GAMMA	U+0263	LATIN SMALL LETTER GAMMA
U+0196	LATIN CAPITAL LETTER IOTA	U+0269	LATIN SMALL LETTER IOTA
U+0197	LATIN CAPITAL LETTER I WITH STROKE	U+0268	LATIN SMALL LETTER I WITH STROKE
U+0198	LATIN CAPITAL LETTER K WITH HOOK	U+0199	LATIN SMALL LETTER K WITH HOOK
U+019C	LATIN CAPITAL LETTER TURNED M	U+026F	LATIN SMALL LETTER TURNED M

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+019D	LATIN CAPITAL LETTER N WITH LEFT HOOK	U+0272	LATIN SMALL LETTER N WITH LEFT HOOK
U+019F	LATIN CAPITAL LETTER O WITH MIDDLE TILDE	U+0275	LATIN SMALL LETTER BARRED O
U+01A0	LATIN CAPITAL LETTER O WITH HORN	U+01A1	LATIN SMALL LETTER O WITH HORN
U+01A2	LATIN CAPITAL LETTER OI	U+01A3	LATIN SMALL LETTER OI
U+01A4	LATIN CAPITAL LETTER P WITH HOOK	U+01A5	LATIN SMALL LETTER P WITH HOOK
U+01A6	LATIN LETTER YR	U+0280	LATIN LETTER SMALL CAPITAL R
U+01A7	LATIN CAPITAL LETTER TONE TWO	U+01A8	LATIN SMALL LETTER TONE TWO
U+01A9	LATIN CAPITAL LETTER ESH	U+0283	LATIN SMALL LETTER ESH
U+01AC	LATIN CAPITAL LETTER T WITH HOOK	U+01AD	LATIN SMALL LETTER T WITH HOOK
U+01AE	LATIN CAPITAL LETTER T WITH RETROFLEX HOOK	U+0288	LATIN SMALL LETTER T WITH RETROFLEX HOOK
U+01AF	LATIN CAPITAL LETTER U WITH HORN	U+01B0	LATIN SMALL LETTER U WITH HORN
U+01B1	LATIN CAPITAL LETTER UPSILON	U+028A	LATIN SMALL LETTER UPSILON
U+01B2	LATIN CAPITAL LETTER V WITH HOOK	U+028B	LATIN SMALL LETTER V WITH HOOK
U+01B3	LATIN CAPITAL LETTER Y WITH HOOK	U+01B4	LATIN SMALL LETTER Y WITH HOOK
U+01B5	LATIN CAPITAL LETTER Z WITH STROKE	U+01B6	LATIN SMALL LETTER Z WITH STROKE
U+01B7	LATIN CAPITAL LETTER EZH	U+0292	LATIN SMALL LETTER EZH
U+01B8	LATIN CAPITAL LETTER EZH REVERSED	U+01B9	LATIN SMALL LETTER EZH REVERSED
U+01BC	LATIN CAPITAL LETTER TONE FIVE	U+01BD	LATIN SMALL LETTER TONE FIVE
U+01C4	LATIN CAPITAL LETTER DZ WITH CARON	U+01C6	LATIN SMALL LETTER DZ WITH CARON
U+01C5	LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON	U+01C6	LATIN SMALL LETTER DZ WITH CARON

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+01C7	LATIN CAPITAL LETTER LJ	U+01C9	LATIN SMALL LETTER LJ
U+01C8	LATIN CAPITAL LETTER L WITH SMALL LETTER J	U+01C9	LATIN SMALL LETTER LJ
U+01CA	LATIN CAPITAL LETTER NJ	U+01CC	LATIN SMALL LETTER NJ
U+01CB	LATIN CAPITAL LETTER N WITH SMALL LETTER J	U+01CC	LATIN SMALL LETTER NJ
U+01CD	LATIN CAPITAL LETTER A WITH CARON	U+01CE	LATIN SMALL LETTER A WITH CARON
U+01CF	LATIN CAPITAL LETTER I WITH CARON	U+01D0	LATIN SMALL LETTER I WITH CARON
U+01D1	LATIN CAPITAL LETTER O WITH CARON	U+01D2	LATIN SMALL LETTER O WITH CARON
U+01D3	LATIN CAPITAL LETTER U WITH CARON	U+01D4	LATIN SMALL LETTER U WITH CARON
U+01D5	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON	U+01D6	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON
U+01D7	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE	U+01D8	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE
U+01D9	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON	U+01DA	LATIN SMALL LETTER U WITH DIAERESIS AND CARON
U+01DB	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE	U+01DC	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
U+01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON	U+01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
U+01E0	LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON	U+01E1	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
U+01E2	LATIN CAPITAL LETTER AE WITH MACRON	U+01E3	LATIN SMALL LETTER AE WITH MACRON
U+01E4	LATIN CAPITAL LETTER G WITH STROKE	U+01E5	LATIN SMALL LETTER G WITH STROKE
U+01E6	LATIN CAPITAL LETTER G WITH CARON	U+01E7	LATIN SMALL LETTER G WITH CARON
U+01E8	LATIN CAPITAL LETTER K WITH CARON	U+01E9	LATIN SMALL LETTER K WITH CARON
U+01EA	LATIN CAPITAL LETTER O WITH OGONEK	U+01EB	LATIN SMALL LETTER O WITH OGONEK

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+01EC	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON	U+01ED	LATIN SMALL LETTER O WITH OGONEK AND MACRON
U+01EE	LATIN CAPITAL LETTER EZH WITH CARON	U+01EF	LATIN SMALL LETTER EZH WITH CARON
U+01F1	LATIN CAPITAL LETTER DZ	U+01F3	LATIN SMALL LETTER DZ
U+01F2	LATIN CAPITAL LETTER D WITH SMALL LETTER Z	U+01F3	LATIN SMALL LETTER DZ
U+01F4	LATIN CAPITAL LETTER G WITH ACUTE	U+01F5	LATIN SMALL LETTER G WITH ACUTE
U+01F6	LATIN CAPITAL LETTER HWAIR	U+0195	LATIN SMALL LETTER HV
U+01F7	LATIN CAPITAL LETTER WYNN	U+01BF	LATIN LETTER WYNN
U+01F8	LATIN CAPITAL LETTER N WITH GRAVE	U+01F9	LATIN SMALL LETTER N WITH GRAVE
U+01FA	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE	U+01FB	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
U+01FC	LATIN CAPITAL LETTER AE WITH ACUTE	U+01FD	LATIN SMALL LETTER AE WITH ACUTE
U+01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE	U+01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE
U+0200	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE	U+0201	LATIN SMALL LETTER A WITH DOUBLE GRAVE
U+0202	LATIN CAPITAL LETTER A WITH INVERTED BREVE	U+0203	LATIN SMALL LETTER A WITH INVERTED BREVE
U+0204	LATIN CAPITAL LETTER E WITH DOUBLE GRAVE	U+0205	LATIN SMALL LETTER E WITH DOUBLE GRAVE
U+0206	LATIN CAPITAL LETTER E WITH INVERTED BREVE	U+0207	LATIN SMALL LETTER E WITH INVERTED BREVE
U+0208	LATIN CAPITAL LETTER I WITH DOUBLE GRAVE	U+0209	LATIN SMALL LETTER I WITH DOUBLE GRAVE
U+020A	LATIN CAPITAL LETTER I WITH INVERTED BREVE	U+020B	LATIN SMALL LETTER I WITH INVERTED BREVE
U+020C	LATIN CAPITAL LETTER O WITH DOUBLE GRAVE	U+020D	LATIN SMALL LETTER O WITH DOUBLE GRAVE
U+020E	LATIN CAPITAL LETTER O WITH INVERTED BREVE	U+020F	LATIN SMALL LETTER O WITH INVERTED BREVE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0210	LATIN CAPITAL LETTER R WITH DOUBLE GRAVE	U+0211	LATIN SMALL LETTER R WITH DOUBLE GRAVE
U+0212	LATIN CAPITAL LETTER R WITH INVERTED BREVE	U+0213	LATIN SMALL LETTER R WITH INVERTED BREVE
U+0214	LATIN CAPITAL LETTER U WITH DOUBLE GRAVE	U+0215	LATIN SMALL LETTER U WITH DOUBLE GRAVE
U+0216	LATIN CAPITAL LETTER U WITH INVERTED BREVE	U+0217	LATIN SMALL LETTER U WITH INVERTED BREVE
U+0218	LATIN CAPITAL LETTER S WITH COMMA BELOW	U+0219	LATIN SMALL LETTER S WITH COMMA BELOW
U+021A	LATIN CAPITAL LETTER T WITH COMMA BELOW	U+021B	LATIN SMALL LETTER T WITH COMMA BELOW
U+021C	LATIN CAPITAL LETTER YOGH	U+021D	LATIN SMALL LETTER YOGH
U+021E	LATIN CAPITAL LETTER H WITH CARON	U+021F	LATIN SMALL LETTER H WITH CARON
U+0220	LATIN CAPITAL LETTER N WITH LONG RIGHT LEG	U+019E	LATIN SMALL LETTER N WITH LONG RIGHT LEG
U+0222	LATIN CAPITAL LETTER OU	U+0223	LATIN SMALL LETTER OU
U+0224	LATIN CAPITAL LETTER Z WITH HOOK	U+0225	LATIN SMALL LETTER Z WITH HOOK
U+0226	LATIN CAPITAL LETTER A WITH DOT ABOVE	U+0227	LATIN SMALL LETTER A WITH DOT ABOVE
U+0228	LATIN CAPITAL LETTER E WITH CEDILLA	U+0229	LATIN SMALL LETTER E WITH CEDILLA
U+022A	LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON	U+022B	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON
U+022C	LATIN CAPITAL LETTER O WITH TILDE AND MACRON	U+022D	LATIN SMALL LETTER O WITH TILDE AND MACRON
U+022E	LATIN CAPITAL LETTER O WITH DOT ABOVE	U+022F	LATIN SMALL LETTER O WITH DOT ABOVE
U+0230	LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON	U+0231	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON
U+0232	LATIN CAPITAL LETTER Y WITH MACRON	U+0233	LATIN SMALL LETTER Y WITH MACRON
U+023A	LATIN CAPITAL LETTER A WITH STROKE	U+2C65	LATIN SMALL LETTER A WITH STROKE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+023B	LATIN CAPITAL LETTER C WITH STROKE	U+023C	LATIN SMALL LETTER C WITH STROKE
U+023D	LATIN CAPITAL LETTER L WITH BAR	U+019A	LATIN SMALL LETTER L WITH BAR
U+023E	LATIN CAPITAL LETTER T WITH DIAGONAL STROKE	U+2C66	LATIN SMALL LETTER T WITH DIAGONAL STROKE
U+0241	LATIN CAPITAL LETTER GLOTTAL STOP	U+0242	LATIN SMALL LETTER GLOTTAL STOP
U+0243	LATIN CAPITAL LETTER B WITH STROKE	U+0180	LATIN SMALL LETTER B WITH STROKE
U+0244	LATIN CAPITAL LETTER U BAR	U+0289	LATIN SMALL LETTER U BAR
U+0245	LATIN CAPITAL LETTER TURNED V	U+028C	LATIN SMALL LETTER TURNED V
U+0246	LATIN CAPITAL LETTER E WITH STROKE	U+0247	LATIN SMALL LETTER E WITH STROKE
U+0248	LATIN CAPITAL LETTER J WITH STROKE	U+0249	LATIN SMALL LETTER J WITH STROKE
U+024A	LATIN CAPITAL LETTER SMALL Q WITH HOOK TAIL	U+024B	LATIN SMALL LETTER Q WITH HOOK TAIL
U+024C	LATIN CAPITAL LETTER R WITH STROKE	U+024D	LATIN SMALL LETTER R WITH STROKE
U+024E	LATIN CAPITAL LETTER Y WITH STROKE	U+024F	LATIN SMALL LETTER Y WITH STROKE
U+0370	GREEK CAPITAL LETTER HETA	U+0371	GREEK SMALL LETTER HETA
U+0372	GREEK CAPITAL LETTER ARCHAIC SAMPI	U+0373	GREEK SMALL LETTER ARCHAIC SAMPI
U+0376	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA	U+0377	GREEK SMALL LETTER PAMPHYLIAN DIGAMMA
U+0386	GREEK CAPITAL LETTER ALPHA WITH TONOS	U+03AC	GREEK SMALL LETTER ALPHA WITH TONOS
U+0388	GREEK CAPITAL LETTER EPSILON WITH TONOS	U+03AD	GREEK SMALL LETTER EPSILON WITH TONOS
U+0389	GREEK CAPITAL LETTER ETA WITH TONOS	U+03AE	GREEK SMALL LETTER ETA WITH TONOS
U+038A	GREEK CAPITAL LETTER IOTA WITH TONOS	U+03AF	GREEK SMALL LETTER IOTA WITH TONOS

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+038C	GREEK CAPITAL LETTER OMICRON WITH TONOS	U+03CC	GREEK SMALL LETTER OMICRON WITH TONOS
U+038E	GREEK CAPITAL LETTER UPSILON WITH TONOS	U+03CD	GREEK SMALL LETTER UPSILON WITH TONOS
U+038F	GREEK CAPITAL LETTER OMEGA WITH TONOS	U+03CE	GREEK SMALL LETTER OMEGA WITH TONOS
U+0391	GREEK CAPITAL LETTER ALPHA	U+03B1	GREEK SMALL LETTER ALPHA
U+0392	GREEK CAPITAL LETTER BETA	U+03B2	GREEK SMALL LETTER BETA
U+0393	GREEK CAPITAL LETTER GAMMA	U+03B3	GREEK SMALL LETTER GAMMA
U+0394	GREEK CAPITAL LETTER DELTA	U+03B4	GREEK SMALL LETTER DELTA
U+0395	GREEK CAPITAL LETTER EPSILON	U+03B5	GREEK SMALL LETTER EPSILON
U+0396	GREEK CAPITAL LETTER ZETA	U+03B6	GREEK SMALL LETTER ZETA
U+0397	GREEK CAPITAL LETTER ETA	U+03B7	GREEK SMALL LETTER ETA
U+0398	GREEK CAPITAL LETTER THETA	U+03B8	GREEK SMALL LETTER THETA
U+0399	GREEK CAPITAL LETTER IOTA	U+03B9	GREEK SMALL LETTER IOTA
U+039A	GREEK CAPITAL LETTER KAPPA	U+03BA	GREEK SMALL LETTER KAPPA
U+039B	GREEK CAPITAL LETTER LAMDA	U+03BB	GREEK SMALL LETTER LAMDA
U+039C	GREEK CAPITAL LETTER MU	U+03BC	GREEK SMALL LETTER MU
U+039D	GREEK CAPITAL LETTER NU	U+03BD	GREEK SMALL LETTER NU
U+039E	GREEK CAPITAL LETTER XI	U+03BE	GREEK SMALL LETTER XI
U+039F	GREEK CAPITAL LETTER OMICRON	U+03BF	GREEK SMALL LETTER OMICRON
U+03A0	GREEK CAPITAL LETTER PI	U+03C0	GREEK SMALL LETTER PI
U+03A1	GREEK CAPITAL LETTER RHO	U+03C1	GREEK SMALL LETTER RHO
U+03A3	GREEK CAPITAL LETTER SIGMA	U+03C3	GREEK SMALL LETTER SIGMA
U+03A4	GREEK CAPITAL LETTER TAU	U+03C4	GREEK SMALL LETTER TAU
U+03A5	GREEK CAPITAL LETTER UPSILON	U+03C5	GREEK SMALL LETTER UPSILON
U+03A6	GREEK CAPITAL LETTER PHI	U+03C6	GREEK SMALL LETTER PHI
U+03A7	GREEK CAPITAL LETTER CHI	U+03C7	GREEK SMALL LETTER CHI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+03A8	GREEK CAPITAL LETTER PSI	U+03C8	GREEK SMALL LETTER PSI
U+03A9	GREEK CAPITAL LETTER OMEGA	U+03C9	GREEK SMALL LETTER OMEGA
U+03AA	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA	U+03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA
U+03AB	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA	U+03CB	GREEK SMALL LETTER UPSILON WITH DIALYTIKA
U+03CF	GREEK CAPITAL KAI SYMBOL	U+03D7	GREEK KAI SYMBOL
U+03D8	GREEK LETTER ARCHAIC KOPPA	U+03D9	GREEK SMALL LETTER ARCHAIC KOPPA
U+03DA	GREEK LETTER STIGMA	U+03DB	GREEK SMALL LETTER STIGMA
U+03DC	GREEK LETTER DIGAMMA	U+03DD	GREEK SMALL LETTER DIGAMMA
U+03DE	GREEK LETTER KOPPA	U+03DF	GREEK SMALL LETTER KOPPA
U+03E0	GREEK LETTER SAMPI	U+03E1	GREEK SMALL LETTER SAMPI
U+03E2	COPTIC CAPITAL LETTER SHEI	U+03E3	COPTIC SMALL LETTER SHEI
U+03E4	COPTIC CAPITAL LETTER FEI	U+03E5	COPTIC SMALL LETTER FEI
U+03E6	COPTIC CAPITAL LETTER KHEI	U+03E7	COPTIC SMALL LETTER KHEI
U+03E8	COPTIC CAPITAL LETTER HORI	U+03E9	COPTIC SMALL LETTER HORI
U+03EA	COPTIC CAPITAL LETTER GANGIA	U+03EB	COPTIC SMALL LETTER GANGIA
U+03EC	COPTIC CAPITAL LETTER SHIMA	U+03ED	COPTIC SMALL LETTER SHIMA
U+03EE	COPTIC CAPITAL LETTER DEI	U+03EF	COPTIC SMALL LETTER DEI
U+03F4	GREEK CAPITAL THETA SYMBOL	U+03B8	GREEK SMALL LETTER THETA
U+03F7	GREEK CAPITAL LETTER SHO	U+03F8	GREEK SMALL LETTER SHO
U+03F9	GREEK CAPITAL LUNATE SIGMA SYMBOL	U+03F2	GREEK LUNATE SIGMA SYMBOL
U+03FA	GREEK CAPITAL LETTER SAN	U+03FB	GREEK SMALL LETTER SAN
U+03FD	GREEK CAPITAL REVERSED LUNATE SIGMA SYMBOL	U+037B	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL
U+03FE	GREEK CAPITAL DOTTED LUNATE SIGMA SYMBOL	U+037C	GREEK SMALL DOTTED LUNATE SIGMA SYMBOL

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+03FF	GREEK CAPITAL REVERSED DOTTED LUNATE SIGMA SYMBOL	U+037D	GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL
U+0400	CYRILLIC CAPITAL LETTER IE WITH GRAVE	U+0450	CYRILLIC SMALL LETTER IE WITH GRAVE
U+0401	CYRILLIC CAPITAL LETTER IO	U+0451	CYRILLIC SMALL LETTER IO
U+0402	CYRILLIC CAPITAL LETTER DJE	U+0452	CYRILLIC SMALL LETTER DJE
U+0403	CYRILLIC CAPITAL LETTER GJE	U+0453	CYRILLIC SMALL LETTER GJE
U+0404	CYRILLIC CAPITAL LETTER UKRAINIAN IE	U+0454	CYRILLIC SMALL LETTER UKRAINIAN IE
U+0405	CYRILLIC CAPITAL LETTER DZE	U+0455	CYRILLIC SMALL LETTER DZE
U+0406	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I	U+0456	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
U+0407	CYRILLIC CAPITAL LETTER YI	U+0457	CYRILLIC SMALL LETTER YI
U+0408	CYRILLIC CAPITAL LETTER JE	U+0458	CYRILLIC SMALL LETTER JE
U+0409	CYRILLIC CAPITAL LETTER LJE	U+0459	CYRILLIC SMALL LETTER LJE
U+040A	CYRILLIC CAPITAL LETTER NJE	U+045A	CYRILLIC SMALL LETTER NJE
U+040B	CYRILLIC CAPITAL LETTER TSHE	U+045B	CYRILLIC SMALL LETTER TSHE
U+040C	CYRILLIC CAPITAL LETTER KJE	U+045C	CYRILLIC SMALL LETTER KJE
U+040D	CYRILLIC CAPITAL LETTER I WITH GRAVE	U+045D	CYRILLIC SMALL LETTER I WITH GRAVE
U+040E	CYRILLIC CAPITAL LETTER SHORT U	U+045E	CYRILLIC SMALL LETTER SHORT U
U+040F	CYRILLIC CAPITAL LETTER DZHE	U+045F	CYRILLIC SMALL LETTER DZHE
U+0410	CYRILLIC CAPITAL LETTER A	U+0430	CYRILLIC SMALL LETTER A
U+0411	CYRILLIC CAPITAL LETTER BE	U+0431	CYRILLIC SMALL LETTER BE
U+0412	CYRILLIC CAPITAL LETTER VE	U+0432	CYRILLIC SMALL LETTER VE
U+0413	CYRILLIC CAPITAL LETTER GHE	U+0433	CYRILLIC SMALL LETTER GHE
U+0414	CYRILLIC CAPITAL LETTER DE	U+0434	CYRILLIC SMALL LETTER DE
U+0415	CYRILLIC CAPITAL LETTER IE	U+0435	CYRILLIC SMALL LETTER IE
U+0416	CYRILLIC CAPITAL LETTER ZHE	U+0436	CYRILLIC SMALL LETTER ZHE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0417	CYRILLIC CAPITAL LETTER ZE	U+0437	CYRILLIC SMALL LETTER ZE
U+0418	CYRILLIC CAPITAL LETTER I	U+0438	CYRILLIC SMALL LETTER I
U+0419	CYRILLIC CAPITAL LETTER SHORT I	U+0439	CYRILLIC SMALL LETTER SHORT I
U+041A	CYRILLIC CAPITAL LETTER KA	U+043A	CYRILLIC SMALL LETTER KA
U+041B	CYRILLIC CAPITAL LETTER EL	U+043B	CYRILLIC SMALL LETTER EL
U+041C	CYRILLIC CAPITAL LETTER EM	U+043C	CYRILLIC SMALL LETTER EM
U+041D	CYRILLIC CAPITAL LETTER EN	U+043D	CYRILLIC SMALL LETTER EN
U+041E	CYRILLIC CAPITAL LETTER O	U+043E	CYRILLIC SMALL LETTER O
U+041F	CYRILLIC CAPITAL LETTER PE	U+043F	CYRILLIC SMALL LETTER PE
U+0420	CYRILLIC CAPITAL LETTER ER	U+0440	CYRILLIC SMALL LETTER ER
U+0421	CYRILLIC CAPITAL LETTER ES	U+0441	CYRILLIC SMALL LETTER ES
U+0422	CYRILLIC CAPITAL LETTER TE	U+0442	CYRILLIC SMALL LETTER TE
U+0423	CYRILLIC CAPITAL LETTER U	U+0443	CYRILLIC SMALL LETTER U
U+0424	CYRILLIC CAPITAL LETTER EF	U+0444	CYRILLIC SMALL LETTER EF
U+0425	CYRILLIC CAPITAL LETTER HA	U+0445	CYRILLIC SMALL LETTER HA
U+0426	CYRILLIC CAPITAL LETTER TSE	U+0446	CYRILLIC SMALL LETTER TSE
U+0427	CYRILLIC CAPITAL LETTER CHE	U+0447	CYRILLIC SMALL LETTER CHE
U+0428	CYRILLIC CAPITAL LETTER SHA	U+0448	CYRILLIC SMALL LETTER SHA
U+0429	CYRILLIC CAPITAL LETTER SHCHA	U+0449	CYRILLIC SMALL LETTER SHCHA
U+042A	CYRILLIC CAPITAL LETTER HARD SIGN	U+044A	CYRILLIC SMALL LETTER HARD SIGN
U+042B	CYRILLIC CAPITAL LETTER YERU	U+044B	CYRILLIC SMALL LETTER YERU
U+042C	CYRILLIC CAPITAL LETTER SOFT SIGN	U+044C	CYRILLIC SMALL LETTER SOFT SIGN
U+042D	CYRILLIC CAPITAL LETTER E	U+044D	CYRILLIC SMALL LETTER E
U+042E	CYRILLIC CAPITAL LETTER YU	U+044E	CYRILLIC SMALL LETTER YU
U+042F	CYRILLIC CAPITAL LETTER YA	U+044F	CYRILLIC SMALL LETTER YA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0460	CYRILLIC CAPITAL LETTER OMEGA	U+0461	CYRILLIC SMALL LETTER OMEGA
U+0462	CYRILLIC CAPITAL LETTER YAT	U+0463	CYRILLIC SMALL LETTER YAT
U+0464	CYRILLIC CAPITAL LETTER IOTIFIED E	U+0465	CYRILLIC SMALL LETTER IOTIFIED E
U+0466	CYRILLIC CAPITAL LETTER LITTLE YUS	U+0467	CYRILLIC SMALL LETTER LITTLE YUS
U+0468	CYRILLIC CAPITAL LETTER IOTIFIED LITTLE YUS	U+0469	CYRILLIC SMALL LETTER IOTIFIED LITTLE YUS
U+046A	CYRILLIC CAPITAL LETTER BIG YUS	U+046B	CYRILLIC SMALL LETTER BIG YUS
U+046C	CYRILLIC CAPITAL LETTER IOTIFIED BIG YUS	U+046D	CYRILLIC SMALL LETTER IOTIFIED BIG YUS
U+046E	CYRILLIC CAPITAL LETTER KSI	U+046F	CYRILLIC SMALL LETTER KSI
U+0470	CYRILLIC CAPITAL LETTER PSI	U+0471	CYRILLIC SMALL LETTER PSI
U+0472	CYRILLIC CAPITAL LETTER FITA	U+0473	CYRILLIC SMALL LETTER FITA
U+0474	CYRILLIC CAPITAL LETTER IZHITSA	U+0475	CYRILLIC SMALL LETTER IZHITSA
U+0476	CYRILLIC CAPITAL LETTER IZHITSA WITH DOUBLE GRAVE ACCENT	U+0477	CYRILLIC SMALL LETTER IZHITSA WITH DOUBLE GRAVE ACCENT
U+0478	CYRILLIC CAPITAL LETTER UK	U+0479	CYRILLIC SMALL LETTER UK
U+047A	CYRILLIC CAPITAL LETTER ROUND OMEGA	U+047B	CYRILLIC SMALL LETTER ROUND OMEGA
U+047C	CYRILLIC CAPITAL LETTER OMEGA WITH TITLO	U+047D	CYRILLIC SMALL LETTER OMEGA WITH TITLO
U+047E	CYRILLIC CAPITAL LETTER OT	U+047F	CYRILLIC SMALL LETTER OT
U+0480	CYRILLIC CAPITAL LETTER KOPPA	U+0481	CYRILLIC SMALL LETTER KOPPA
U+048A	CYRILLIC CAPITAL LETTER SHORT I WITH TAIL	U+048B	CYRILLIC SMALL LETTER SHORT I WITH TAIL
U+048C	CYRILLIC CAPITAL LETTER SEMISOFT SIGN	U+048D	CYRILLIC SMALL LETTER SEMISOFT SIGN
U+048E	CYRILLIC CAPITAL LETTER ER WITH TICK	U+048F	CYRILLIC SMALL LETTER ER WITH TICK

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0490	CYRILLIC CAPITAL LETTER GHE WITH UPTURN	U+0491	CYRILLIC SMALL LETTER GHE WITH UPTURN
U+0492	CYRILLIC CAPITAL LETTER GHE WITH STROKE	U+0493	CYRILLIC SMALL LETTER GHE WITH STROKE
U+0494	CYRILLIC CAPITAL LETTER GHE WITH MIDDLE HOOK	U+0495	CYRILLIC SMALL LETTER GHE WITH MIDDLE HOOK
U+0496	CYRILLIC CAPITAL LETTER ZHE WITH DESCENDER	U+0497	CYRILLIC SMALL LETTER ZHE WITH DESCENDER
U+0498	CYRILLIC CAPITAL LETTER ZE WITH DESCENDER	U+0499	CYRILLIC SMALL LETTER ZE WITH DESCENDER
U+049A	CYRILLIC CAPITAL LETTER KA WITH DESCENDER	U+049B	CYRILLIC SMALL LETTER KA WITH DESCENDER
U+049C	CYRILLIC CAPITAL LETTER KA WITH VERTICAL STROKE	U+049D	CYRILLIC SMALL LETTER KA WITH VERTICAL STROKE
U+049E	CYRILLIC CAPITAL LETTER KA WITH STROKE	U+049F	CYRILLIC SMALL LETTER KA WITH STROKE
U+04A0	CYRILLIC CAPITAL LETTER BASHKIR KA	U+04A1	CYRILLIC SMALL LETTER BASHKIR KA
U+04A2	CYRILLIC CAPITAL LETTER EN WITH DESCENDER	U+04A3	CYRILLIC SMALL LETTER EN WITH DESCENDER
U+04A4	CYRILLIC CAPITAL LIGATURE EN GHE	U+04A5	CYRILLIC SMALL LIGATURE EN GHE
U+04A6	CYRILLIC CAPITAL LETTER PE WITH MIDDLE HOOK	U+04A7	CYRILLIC SMALL LETTER PE WITH MIDDLE HOOK
U+04A8	CYRILLIC CAPITAL LETTER ABKHASIAN HA	U+04A9	CYRILLIC SMALL LETTER ABKHASIAN HA
U+04AA	CYRILLIC CAPITAL LETTER ES WITH DESCENDER	U+04AB	CYRILLIC SMALL LETTER ES WITH DESCENDER
U+04AC	CYRILLIC CAPITAL LETTER TE WITH DESCENDER	U+04AD	CYRILLIC SMALL LETTER TE WITH DESCENDER
U+04AE	CYRILLIC CAPITAL LETTER STRAIGHT U	U+04AF	CYRILLIC SMALL LETTER STRAIGHT U
U+04B0	CYRILLIC CAPITAL LETTER STRAIGHT U WITH STROKE	U+04B1	CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE
U+04B2	CYRILLIC CAPITAL LETTER HA WITH DESCENDER	U+04B3	CYRILLIC SMALL LETTER HA WITH DESCENDER

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+04B4	CYRILLIC CAPITAL LIGATURE TE TSE	U+04B5	CYRILLIC SMALL LIGATURE TE TSE
U+04B6	CYRILLIC CAPITAL LETTER CHE WITH DESCENDER	U+04B7	CYRILLIC SMALL LETTER CHE WITH DESCENDER
U+04B8	CYRILLIC CAPITAL LETTER CHE WITH VERTICAL STROKE	U+04B9	CYRILLIC SMALL LETTER CHE WITH VERTICAL STROKE
U+04BA	CYRILLIC CAPITAL LETTER SHHA	U+04BB	CYRILLIC SMALL LETTER SHHA
U+04BC	CYRILLIC CAPITAL LETTER ABKHASIAN CHE	U+04BD	CYRILLIC SMALL LETTER ABKHASIAN CHE
U+04BE	CYRILLIC CAPITAL LETTER ABKHASIAN CHE WITH DESCENDER	U+04BF	CYRILLIC SMALL LETTER ABKHASIAN CHE WITH DESCENDER
U+04C0	CYRILLIC LETTER PALOCHKA	U+04CF	CYRILLIC SMALL LETTER PALOCHKA
U+04C1	CYRILLIC CAPITAL LETTER ZHE WITH BREVE	U+04C2	CYRILLIC SMALL LETTER ZHE WITH BREVE
U+04C3	CYRILLIC CAPITAL LETTER KA WITH HOOK	U+04C4	CYRILLIC SMALL LETTER KA WITH HOOK
U+04C5	CYRILLIC CAPITAL LETTER EL WITH TAIL	U+04C6	CYRILLIC SMALL LETTER EL WITH TAIL
U+04C7	CYRILLIC CAPITAL LETTER EN WITH HOOK	U+04C8	CYRILLIC SMALL LETTER EN WITH HOOK
U+04C9	CYRILLIC CAPITAL LETTER EN WITH TAIL	U+04CA	CYRILLIC SMALL LETTER EN WITH TAIL
U+04CB	CYRILLIC CAPITAL LETTER KHAKASSIAN CHE	U+04CC	CYRILLIC SMALL LETTER KHAKASSIAN CHE
U+04CD	CYRILLIC CAPITAL LETTER EM WITH TAIL	U+04CE	CYRILLIC SMALL LETTER EM WITH TAIL
U+04D0	CYRILLIC CAPITAL LETTER A WITH BREVE	U+04D1	CYRILLIC SMALL LETTER A WITH BREVE
U+04D2	CYRILLIC CAPITAL LETTER A WITH DIAERESIS	U+04D3	CYRILLIC SMALL LETTER A WITH DIAERESIS
U+04D4	CYRILLIC CAPITAL LIGATURE A IE	U+04D5	CYRILLIC SMALL LIGATURE A IE
U+04D6	CYRILLIC CAPITAL LETTER IE WITH BREVE	U+04D7	CYRILLIC SMALL LETTER IE WITH BREVE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+04D8	CYRILLIC CAPITAL LETTER SCHWA	U+04D9	CYRILLIC SMALL LETTER SCHWA
U+04DA	CYRILLIC CAPITAL LETTER SCHWA WITH DIAERESIS	U+04DB	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS
U+04DC	CYRILLIC CAPITAL LETTER ZHE WITH DIAERESIS	U+04DD	CYRILLIC SMALL LETTER ZHE WITH DIAERESIS
U+04DE	CYRILLIC CAPITAL LETTER ZE WITH DIAERESIS	U+04DF	CYRILLIC SMALL LETTER ZE WITH DIAERESIS
U+04E0	CYRILLIC CAPITAL LETTER ABKHASIAN DZE	U+04E1	CYRILLIC SMALL LETTER ABKHASIAN DZE
U+04E2	CYRILLIC CAPITAL LETTER I WITH MACRON	U+04E3	CYRILLIC SMALL LETTER I WITH MACRON
U+04E4	CYRILLIC CAPITAL LETTER I WITH DIAERESIS	U+04E5	CYRILLIC SMALL LETTER I WITH DIAERESIS
U+04E6	CYRILLIC CAPITAL LETTER O WITH DIAERESIS	U+04E7	CYRILLIC SMALL LETTER O WITH DIAERESIS
U+04E8	CYRILLIC CAPITAL LETTER BARRED O	U+04E9	CYRILLIC SMALL LETTER BARRED O
U+04EA	CYRILLIC CAPITAL LETTER BARRED O WITH DIAERESIS	U+04EB	CYRILLIC SMALL LETTER BARRED O WITH DIAERESIS
U+04EC	CYRILLIC CAPITAL LETTER E WITH DIAERESIS	U+04ED	CYRILLIC SMALL LETTER E WITH DIAERESIS
U+04EE	CYRILLIC CAPITAL LETTER U WITH MACRON	U+04EF	CYRILLIC SMALL LETTER U WITH MACRON
U+04F0	CYRILLIC CAPITAL LETTER U WITH DIAERESIS	U+04F1	CYRILLIC SMALL LETTER U WITH DIAERESIS
U+04F2	CYRILLIC CAPITAL LETTER U WITH DOUBLE ACUTE	U+04F3	CYRILLIC SMALL LETTER U WITH DOUBLE ACUTE
U+04F4	CYRILLIC CAPITAL LETTER CHE WITH DIAERESIS	U+04F5	CYRILLIC SMALL LETTER CHE WITH DIAERESIS
U+04F6	CYRILLIC CAPITAL LETTER GHE WITH DESCENDER	U+04F7	CYRILLIC SMALL LETTER GHE WITH DESCENDER
U+04F8	CYRILLIC CAPITAL LETTER YERU WITH DIAERESIS	U+04F9	CYRILLIC SMALL LETTER YERU WITH DIAERESIS
U+04FA	CYRILLIC CAPITAL LETTER GHE WITH STROKE AND HOOK	U+04FB	CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+04FC	CYRILLIC CAPITAL LETTER HA WITH HOOK	U+04FD	CYRILLIC SMALL LETTER HA WITH HOOK
U+04FE	CYRILLIC CAPITAL LETTER HA WITH STROKE	U+04FF	CYRILLIC SMALL LETTER HA WITH STROKE
U+0500	CYRILLIC CAPITAL LETTER KOMI DE	U+0501	CYRILLIC SMALL LETTER KOMI DE
U+0502	CYRILLIC CAPITAL LETTER KOMI DJE	U+0503	CYRILLIC SMALL LETTER KOMI DJE
U+0504	CYRILLIC CAPITAL LETTER KOMI ZJE	U+0505	CYRILLIC SMALL LETTER KOMI ZJE
U+0506	CYRILLIC CAPITAL LETTER KOMI DZJE	U+0507	CYRILLIC SMALL LETTER KOMI DZJE
U+0508	CYRILLIC CAPITAL LETTER KOMI LJE	U+0509	CYRILLIC SMALL LETTER KOMI LJE
U+050A	CYRILLIC CAPITAL LETTER KOMI NJE	U+050B	CYRILLIC SMALL LETTER KOMI NJE
U+050C	CYRILLIC CAPITAL LETTER KOMI SJE	U+050D	CYRILLIC SMALL LETTER KOMI SJE
U+050E	CYRILLIC CAPITAL LETTER KOMI TJE	U+050F	CYRILLIC SMALL LETTER KOMI TJE
U+0510	CYRILLIC CAPITAL LETTER REVERSED ZE	U+0511	CYRILLIC SMALL LETTER REVERSED ZE
U+0512	CYRILLIC CAPITAL LETTER EL WITH HOOK	U+0513	CYRILLIC SMALL LETTER EL WITH HOOK
U+0514	CYRILLIC CAPITAL LETTER LHA	U+0515	CYRILLIC SMALL LETTER LHA
U+0516	CYRILLIC CAPITAL LETTER RHA	U+0517	CYRILLIC SMALL LETTER RHA
U+0518	CYRILLIC CAPITAL LETTER YAE	U+0519	CYRILLIC SMALL LETTER YAE
U+051A	CYRILLIC CAPITAL LETTER QA	U+051B	CYRILLIC SMALL LETTER QA
U+051C	CYRILLIC CAPITAL LETTER WE	U+051D	CYRILLIC SMALL LETTER WE
U+051E	CYRILLIC CAPITAL LETTER ALEUT KA	U+051F	CYRILLIC SMALL LETTER ALEUT KA
U+0520	CYRILLIC CAPITAL LETTER EL WITH MIDDLE HOOK	U+0521	CYRILLIC SMALL LETTER EL WITH MIDDLE HOOK
U+0522	CYRILLIC CAPITAL LETTER EN WITH MIDDLE HOOK	U+0523	CYRILLIC SMALL LETTER EN WITH MIDDLE HOOK

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0524	CYRILLIC CAPITAL LETTER PE WITH DESCENDER	U+0525	CYRILLIC SMALL LETTER PE WITH DESCENDER
U+0526	CYRILLIC CAPITAL LETTER SHHA WITH DESCENDER	U+0527	CYRILLIC SMALL LETTER SHHA WITH DESCENDER
U+0531	ARMENIAN CAPITAL LETTER AYB	U+0561	ARMENIAN SMALL LETTER AYB
U+0532	ARMENIAN CAPITAL LETTER BEN	U+0562	ARMENIAN SMALL LETTER BEN
U+0533	ARMENIAN CAPITAL LETTER GIM	U+0563	ARMENIAN SMALL LETTER GIM
U+0534	ARMENIAN CAPITAL LETTER DA	U+0564	ARMENIAN SMALL LETTER DA
U+0535	ARMENIAN CAPITAL LETTER ECH	U+0565	ARMENIAN SMALL LETTER ECH
U+0536	ARMENIAN CAPITAL LETTER ZA	U+0566	ARMENIAN SMALL LETTER ZA
U+0537	ARMENIAN CAPITAL LETTER EH	U+0567	ARMENIAN SMALL LETTER EH
U+0538	ARMENIAN CAPITAL LETTER ET	U+0568	ARMENIAN SMALL LETTER ET
U+0539	ARMENIAN CAPITAL LETTER TO	U+0569	ARMENIAN SMALL LETTER TO
U+053A	ARMENIAN CAPITAL LETTER ZHE	U+056A	ARMENIAN SMALL LETTER ZHE
U+053B	ARMENIAN CAPITAL LETTER INI	U+056B	ARMENIAN SMALL LETTER INI
U+053C	ARMENIAN CAPITAL LETTER LIWN	U+056C	ARMENIAN SMALL LETTER LIWN
U+053D	ARMENIAN CAPITAL LETTER XEH	U+056D	ARMENIAN SMALL LETTER XEH
U+053E	ARMENIAN CAPITAL LETTER CA	U+056E	ARMENIAN SMALL LETTER CA
U+053F	ARMENIAN CAPITAL LETTER KEN	U+056F	ARMENIAN SMALL LETTER KEN
U+0540	ARMENIAN CAPITAL LETTER HO	U+0570	ARMENIAN SMALL LETTER HO
U+0541	ARMENIAN CAPITAL LETTER JA	U+0571	ARMENIAN SMALL LETTER JA
U+0542	ARMENIAN CAPITAL LETTER GHAD	U+0572	ARMENIAN SMALL LETTER GHAD
U+0543	ARMENIAN CAPITAL LETTER CHEH	U+0573	ARMENIAN SMALL LETTER CHEH
U+0544	ARMENIAN CAPITAL LETTER MEN	U+0574	ARMENIAN SMALL LETTER MEN
U+0545	ARMENIAN CAPITAL LETTER YI	U+0575	ARMENIAN SMALL LETTER YI
U+0546	ARMENIAN CAPITAL LETTER NOW	U+0576	ARMENIAN SMALL LETTER NOW
U+0547	ARMENIAN CAPITAL LETTER SHA	U+0577	ARMENIAN SMALL LETTER SHA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+0548	ARMENIAN CAPITAL LETTER VO	U+0578	ARMENIAN SMALL LETTER VO
U+0549	ARMENIAN CAPITAL LETTER CHA	U+0579	ARMENIAN SMALL LETTER CHA
U+054A	ARMENIAN CAPITAL LETTER PEH	U+057A	ARMENIAN SMALL LETTER PEH
U+054B	ARMENIAN CAPITAL LETTER JHEH	U+057B	ARMENIAN SMALL LETTER JHEH
U+054C	ARMENIAN CAPITAL LETTER RA	U+057C	ARMENIAN SMALL LETTER RA
U+054D	ARMENIAN CAPITAL LETTER SEH	U+057D	ARMENIAN SMALL LETTER SEH
U+054E	ARMENIAN CAPITAL LETTER VEW	U+057E	ARMENIAN SMALL LETTER VEW
U+054F	ARMENIAN CAPITAL LETTER TIWN	U+057F	ARMENIAN SMALL LETTER TIWN
U+0550	ARMENIAN CAPITAL LETTER REH	U+0580	ARMENIAN SMALL LETTER REH
U+0551	ARMENIAN CAPITAL LETTER CO	U+0581	ARMENIAN SMALL LETTER CO
U+0552	ARMENIAN CAPITAL LETTER YIWN	U+0582	ARMENIAN SMALL LETTER YIWN
U+0553	ARMENIAN CAPITAL LETTER PIWR	U+0583	ARMENIAN SMALL LETTER PIWR
U+0554	ARMENIAN CAPITAL LETTER KEH	U+0584	ARMENIAN SMALL LETTER KEH
U+0555	ARMENIAN CAPITAL LETTER OH	U+0585	ARMENIAN SMALL LETTER OH
U+0556	ARMENIAN CAPITAL LETTER FEH	U+0586	ARMENIAN SMALL LETTER FEH
U+10A0	GEORGIAN CAPITAL LETTER AN	U+2D00	GEORGIAN SMALL LETTER AN
U+10A1	GEORGIAN CAPITAL LETTER BAN	U+2D01	GEORGIAN SMALL LETTER BAN
U+10A2	GEORGIAN CAPITAL LETTER GAN	U+2D02	GEORGIAN SMALL LETTER GAN
U+10A3	GEORGIAN CAPITAL LETTER DON	U+2D03	GEORGIAN SMALL LETTER DON
U+10A4	GEORGIAN CAPITAL LETTER EN	U+2D04	GEORGIAN SMALL LETTER EN
U+10A5	GEORGIAN CAPITAL LETTER VIN	U+2D05	GEORGIAN SMALL LETTER VIN
U+10A6	GEORGIAN CAPITAL LETTER ZEN	U+2D06	GEORGIAN SMALL LETTER ZEN
U+10A7	GEORGIAN CAPITAL LETTER TAN	U+2D07	GEORGIAN SMALL LETTER TAN
U+10A8	GEORGIAN CAPITAL LETTER IN	U+2D08	GEORGIAN SMALL LETTER IN

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+10A9	GEORGIAN CAPITAL LETTER KAN	U+2D09	GEORGIAN SMALL LETTER KAN
U+10AA	GEORGIAN CAPITAL LETTER LAS	U+2D0A	GEORGIAN SMALL LETTER LAS
U+10AB	GEORGIAN CAPITAL LETTER MAN	U+2D0B	GEORGIAN SMALL LETTER MAN
U+10AC	GEORGIAN CAPITAL LETTER NAR	U+2D0C	GEORGIAN SMALL LETTER NAR
U+10AD	GEORGIAN CAPITAL LETTER ON	U+2D0D	GEORGIAN SMALL LETTER ON
U+10AE	GEORGIAN CAPITAL LETTER PAR	U+2D0E	GEORGIAN SMALL LETTER PAR
U+10AF	GEORGIAN CAPITAL LETTER ZHAR	U+2D0F	GEORGIAN SMALL LETTER ZHAR
U+10B0	GEORGIAN CAPITAL LETTER RAE	U+2D10	GEORGIAN SMALL LETTER RAE
U+10B1	GEORGIAN CAPITAL LETTER SAN	U+2D11	GEORGIAN SMALL LETTER SAN
U+10B2	GEORGIAN CAPITAL LETTER TAR	U+2D12	GEORGIAN SMALL LETTER TAR
U+10B3	GEORGIAN CAPITAL LETTER UN	U+2D13	GEORGIAN SMALL LETTER UN
U+10B4	GEORGIAN CAPITAL LETTER PHAR	U+2D14	GEORGIAN SMALL LETTER PHAR
U+10B5	GEORGIAN CAPITAL LETTER KHAR	U+2D15	GEORGIAN SMALL LETTER KHAR
U+10B6	GEORGIAN CAPITAL LETTER GHAN	U+2D16	GEORGIAN SMALL LETTER GHAN
U+10B7	GEORGIAN CAPITAL LETTER QAR	U+2D17	GEORGIAN SMALL LETTER QAR
U+10B8	GEORGIAN CAPITAL LETTER SHIN	U+2D18	GEORGIAN SMALL LETTER SHIN
U+10B9	GEORGIAN CAPITAL LETTER CHIN	U+2D19	GEORGIAN SMALL LETTER CHIN
U+10BA	GEORGIAN CAPITAL LETTER CAN	U+2D1A	GEORGIAN SMALL LETTER CAN
U+10BB	GEORGIAN CAPITAL LETTER JIL	U+2D1B	GEORGIAN SMALL LETTER JIL
U+10BC	GEORGIAN CAPITAL LETTER CIL	U+2D1C	GEORGIAN SMALL LETTER CIL
U+10BD	GEORGIAN CAPITAL LETTER CHAR	U+2D1D	GEORGIAN SMALL LETTER CHAR
U+10BE	GEORGIAN CAPITAL LETTER XAN	U+2D1E	GEORGIAN SMALL LETTER XAN
U+10BF	GEORGIAN CAPITAL LETTER JHAN	U+2D1F	GEORGIAN SMALL LETTER JHAN

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+10C0	GEORGIAN CAPITAL LETTER HAE	U+2D20	GEORGIAN SMALL LETTER HAE
U+10C1	GEORGIAN CAPITAL LETTER HE	U+2D21	GEORGIAN SMALL LETTER HE
U+10C2	GEORGIAN CAPITAL LETTER HIE	U+2D22	GEORGIAN SMALL LETTER HIE
U+10C3	GEORGIAN CAPITAL LETTER WE	U+2D23	GEORGIAN SMALL LETTER WE
U+10C4	GEORGIAN CAPITAL LETTER HAR	U+2D24	GEORGIAN SMALL LETTER HAR
U+10C5	GEORGIAN CAPITAL LETTER HOE	U+2D25	GEORGIAN SMALL LETTER HOE
U+1E00	LATIN CAPITAL LETTER A WITH RING BELOW	U+1E01	LATIN SMALL LETTER A WITH RING BELOW
U+1E02	LATIN CAPITAL LETTER B WITH DOT ABOVE	U+1E03	LATIN SMALL LETTER B WITH DOT ABOVE
U+1E04	LATIN CAPITAL LETTER B WITH DOT BELOW	U+1E05	LATIN SMALL LETTER B WITH DOT BELOW
U+1E06	LATIN CAPITAL LETTER B WITH LINE BELOW	U+1E07	LATIN SMALL LETTER B WITH LINE BELOW
U+1E08	LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE	U+1E09	LATIN SMALL LETTER C WITH CEDILLA AND ACUTE
U+1E0A	LATIN CAPITAL LETTER D WITH DOT ABOVE	U+1E0B	LATIN SMALL LETTER D WITH DOT ABOVE
U+1E0C	LATIN CAPITAL LETTER D WITH DOT BELOW	U+1E0D	LATIN SMALL LETTER D WITH DOT BELOW
U+1E0E	LATIN CAPITAL LETTER D WITH LINE BELOW	U+1E0F	LATIN SMALL LETTER D WITH LINE BELOW
U+1E10	LATIN CAPITAL LETTER D WITH CEDILLA	U+1E11	LATIN SMALL LETTER D WITH CEDILLA
U+1E12	LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW	U+1E13	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW
U+1E14	LATIN CAPITAL LETTER E WITH MACRON AND GRAVE	U+1E15	LATIN SMALL LETTER E WITH MACRON AND GRAVE
U+1E16	LATIN CAPITAL LETTER E WITH MACRON AND ACUTE	U+1E17	LATIN SMALL LETTER E WITH MACRON AND ACUTE
U+1E18	LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW	U+1E19	LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW
U+1E1A	LATIN CAPITAL LETTER E WITH TILDE BELOW	U+1E1B	LATIN SMALL LETTER E WITH TILDE BELOW

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E1C	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE	U+1E1D	LATIN SMALL LETTER E WITH CEDILLA AND BREVE
U+1E1E	LATIN CAPITAL LETTER F WITH DOT ABOVE	U+1E1F	LATIN SMALL LETTER F WITH DOT ABOVE
U+1E20	LATIN CAPITAL LETTER G WITH MACRON	U+1E21	LATIN SMALL LETTER G WITH MACRON
U+1E22	LATIN CAPITAL LETTER H WITH DOT ABOVE	U+1E23	LATIN SMALL LETTER H WITH DOT ABOVE
U+1E24	LATIN CAPITAL LETTER H WITH DOT BELOW	U+1E25	LATIN SMALL LETTER H WITH DOT BELOW
U+1E26	LATIN CAPITAL LETTER H WITH DIAERESIS	U+1E27	LATIN SMALL LETTER H WITH DIAERESIS
U+1E28	LATIN CAPITAL LETTER H WITH CEDILLA	U+1E29	LATIN SMALL LETTER H WITH CEDILLA
U+1E2A	LATIN CAPITAL LETTER H WITH BREVE BELOW	U+1E2B	LATIN SMALL LETTER H WITH BREVE BELOW
U+1E2C	LATIN CAPITAL LETTER I WITH TILDE BELOW	U+1E2D	LATIN SMALL LETTER I WITH TILDE BELOW
U+1E2E	LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE	U+1E2F	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE
U+1E30	LATIN CAPITAL LETTER K WITH ACUTE	U+1E31	LATIN SMALL LETTER K WITH ACUTE
U+1E32	LATIN CAPITAL LETTER K WITH DOT BELOW	U+1E33	LATIN SMALL LETTER K WITH DOT BELOW
U+1E34	LATIN CAPITAL LETTER K WITH LINE BELOW	U+1E35	LATIN SMALL LETTER K WITH LINE BELOW
U+1E36	LATIN CAPITAL LETTER L WITH DOT BELOW	U+1E37	LATIN SMALL LETTER L WITH DOT BELOW
U+1E38	LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON	U+1E39	LATIN SMALL LETTER L WITH DOT BELOW AND MACRON
U+1E3A	LATIN CAPITAL LETTER L WITH LINE BELOW	U+1E3B	LATIN SMALL LETTER L WITH LINE BELOW
U+1E3C	LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW	U+1E3D	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW
U+1E3E	LATIN CAPITAL LETTER M WITH ACUTE	U+1E3F	LATIN SMALL LETTER M WITH ACUTE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E40	LATIN CAPITAL LETTER M WITH DOT ABOVE	U+1E41	LATIN SMALL LETTER M WITH DOT ABOVE
U+1E42	LATIN CAPITAL LETTER M WITH DOT BELOW	U+1E43	LATIN SMALL LETTER M WITH DOT BELOW
U+1E44	LATIN CAPITAL LETTER N WITH DOT ABOVE	U+1E45	LATIN SMALL LETTER N WITH DOT ABOVE
U+1E46	LATIN CAPITAL LETTER N WITH DOT BELOW	U+1E47	LATIN SMALL LETTER N WITH DOT BELOW
U+1E48	LATIN CAPITAL LETTER N WITH LINE BELOW	U+1E49	LATIN SMALL LETTER N WITH LINE BELOW
U+1E4A	LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW	U+1E4B	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW
U+1E4C	LATIN CAPITAL LETTER O WITH TILDE AND ACUTE	U+1E4D	LATIN SMALL LETTER O WITH TILDE AND ACUTE
U+1E4E	LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS	U+1E4F	LATIN SMALL LETTER O WITH TILDE AND DIAERESIS
U+1E50	LATIN CAPITAL LETTER O WITH MACRON AND GRAVE	U+1E51	LATIN SMALL LETTER O WITH MACRON AND GRAVE
U+1E52	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE	U+1E53	LATIN SMALL LETTER O WITH MACRON AND ACUTE
U+1E54	LATIN CAPITAL LETTER P WITH ACUTE	U+1E55	LATIN SMALL LETTER P WITH ACUTE
U+1E56	LATIN CAPITAL LETTER P WITH DOT ABOVE	U+1E57	LATIN SMALL LETTER P WITH DOT ABOVE
U+1E58	LATIN CAPITAL LETTER R WITH DOT ABOVE	U+1E59	LATIN SMALL LETTER R WITH DOT ABOVE
U+1E5A	LATIN CAPITAL LETTER R WITH DOT BELOW	U+1E5B	LATIN SMALL LETTER R WITH DOT BELOW
U+1E5C	LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON	U+1E5D	LATIN SMALL LETTER R WITH DOT BELOW AND MACRON
U+1E5E	LATIN CAPITAL LETTER R WITH LINE BELOW	U+1E5F	LATIN SMALL LETTER R WITH LINE BELOW
U+1E60	LATIN CAPITAL LETTER S WITH DOT ABOVE	U+1E61	LATIN SMALL LETTER S WITH DOT ABOVE
U+1E62	LATIN CAPITAL LETTER S WITH DOT BELOW	U+1E63	LATIN SMALL LETTER S WITH DOT BELOW

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E64	LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE	U+1E65	LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE
U+1E66	LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE	U+1E67	LATIN SMALL LETTER S WITH CARON AND DOT ABOVE
U+1E68	LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE	U+1E69	LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE
U+1E6A	LATIN CAPITAL LETTER T WITH DOT ABOVE	U+1E6B	LATIN SMALL LETTER T WITH DOT ABOVE
U+1E6C	LATIN CAPITAL LETTER T WITH DOT BELOW	U+1E6D	LATIN SMALL LETTER T WITH DOT BELOW
U+1E6E	LATIN CAPITAL LETTER T WITH LINE BELOW	U+1E6F	LATIN SMALL LETTER T WITH LINE BELOW
U+1E70	LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW	U+1E71	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW
U+1E72	LATIN CAPITAL LETTER U WITH DIAERESIS BELOW	U+1E73	LATIN SMALL LETTER U WITH DIAERESIS BELOW
U+1E74	LATIN CAPITAL LETTER U WITH TILDE BELOW	U+1E75	LATIN SMALL LETTER U WITH TILDE BELOW
U+1E76	LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW	U+1E77	LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW
U+1E78	LATIN CAPITAL LETTER U WITH TILDE AND ACUTE	U+1E79	LATIN SMALL LETTER U WITH TILDE AND ACUTE
U+1E7A	LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS	U+1E7B	LATIN SMALL LETTER U WITH MACRON AND DIAERESIS
U+1E7C	LATIN CAPITAL LETTER V WITH TILDE	U+1E7D	LATIN SMALL LETTER V WITH TILDE
U+1E7E	LATIN CAPITAL LETTER V WITH DOT BELOW	U+1E7F	LATIN SMALL LETTER V WITH DOT BELOW
U+1E80	LATIN CAPITAL LETTER W WITH GRAVE	U+1E81	LATIN SMALL LETTER W WITH GRAVE
U+1E82	LATIN CAPITAL LETTER W WITH ACUTE	U+1E83	LATIN SMALL LETTER W WITH ACUTE
U+1E84	LATIN CAPITAL LETTER W WITH DIAERESIS	U+1E85	LATIN SMALL LETTER W WITH DIAERESIS
U+1E86	LATIN CAPITAL LETTER W WITH DOT ABOVE	U+1E87	LATIN SMALL LETTER W WITH DOT ABOVE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1E88	LATIN CAPITAL LETTER W WITH DOT BELOW	U+1E89	LATIN SMALL LETTER W WITH DOT BELOW
U+1E8A	LATIN CAPITAL LETTER X WITH DOT ABOVE	U+1E8B	LATIN SMALL LETTER X WITH DOT ABOVE
U+1E8C	LATIN CAPITAL LETTER X WITH DIAERESIS	U+1E8D	LATIN SMALL LETTER X WITH DIAERESIS
U+1E8E	LATIN CAPITAL LETTER Y WITH DOT ABOVE	U+1E8F	LATIN SMALL LETTER Y WITH DOT ABOVE
U+1E90	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX	U+1E91	LATIN SMALL LETTER Z WITH CIRCUMFLEX
U+1E92	LATIN CAPITAL LETTER Z WITH DOT BELOW	U+1E93	LATIN SMALL LETTER Z WITH DOT BELOW
U+1E94	LATIN CAPITAL LETTER Z WITH LINE BELOW	U+1E95	LATIN SMALL LETTER Z WITH LINE BELOW
U+1E9E	LATIN CAPITAL LETTER SHARP S	U+00DF	LATIN SMALL LETTER SHARP S
U+1EA0	LATIN CAPITAL LETTER A WITH DOT BELOW	U+1EA1	LATIN SMALL LETTER A WITH DOT BELOW
U+1EA2	LATIN CAPITAL LETTER A WITH HOOK ABOVE	U+1EA3	LATIN SMALL LETTER A WITH HOOK ABOVE
U+1EA4	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE	U+1EA5	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE
U+1EA6	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE	U+1EA7	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE
U+1EA8	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	U+1EA9	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
U+1EAA	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE	U+1EAB	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
U+1EAC	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW	U+1EAD	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW
U+1EAE	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE	U+1EAF	LATIN SMALL LETTER A WITH BREVE AND ACUTE
U+1EB0	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE	U+1EB1	LATIN SMALL LETTER A WITH BREVE AND GRAVE
U+1EB2	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE	U+1EB3	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1EB4	LATIN CAPITAL LETTER A WITH BREVE AND TILDE	U+1EB5	LATIN SMALL LETTER A WITH BREVE AND TILDE
U+1EB6	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW	U+1EB7	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW
U+1EB8	LATIN CAPITAL LETTER E WITH DOT BELOW	U+1EB9	LATIN SMALL LETTER E WITH DOT BELOW
U+1EBA	LATIN CAPITAL LETTER E WITH HOOK ABOVE	U+1EBB	LATIN SMALL LETTER E WITH HOOK ABOVE
U+1EBC	LATIN CAPITAL LETTER E WITH TILDE	U+1EBD	LATIN SMALL LETTER E WITH TILDE
U+1EBE	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE	U+1EBF	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE
U+1EC0	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE	U+1EC1	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE
U+1EC2	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	U+1EC3	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
U+1EC4	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE	U+1EC5	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
U+1EC6	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW	U+1EC7	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW
U+1EC8	LATIN CAPITAL LETTER I WITH HOOK ABOVE	U+1EC9	LATIN SMALL LETTER I WITH HOOK ABOVE
U+1ECA	LATIN CAPITAL LETTER I WITH DOT BELOW	U+1ECB	LATIN SMALL LETTER I WITH DOT BELOW
U+1ECC	LATIN CAPITAL LETTER O WITH DOT BELOW	U+1ECD	LATIN SMALL LETTER O WITH DOT BELOW
U+1ECE	LATIN CAPITAL LETTER O WITH HOOK ABOVE	U+1ECF	LATIN SMALL LETTER O WITH HOOK ABOVE
U+1ED0	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE	U+1ED1	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE
U+1ED2	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE	U+1ED3	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE
U+1ED4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	U+1ED5	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1ED6	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE	U+1ED7	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
U+1ED8	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW	U+1ED9	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW
U+1EDA	LATIN CAPITAL LETTER O WITH HORN AND ACUTE	U+1EDB	LATIN SMALL LETTER O WITH HORN AND ACUTE
U+1EDC	LATIN CAPITAL LETTER O WITH HORN AND GRAVE	U+1EDD	LATIN SMALL LETTER O WITH HORN AND GRAVE
U+1EDE	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE	U+1EDF	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
U+1EE0	LATIN CAPITAL LETTER O WITH HORN AND TILDE	U+1EE1	LATIN SMALL LETTER O WITH HORN AND TILDE
U+1EE2	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW	U+1EE3	LATIN SMALL LETTER O WITH HORN AND DOT BELOW
U+1EE4	LATIN CAPITAL LETTER U WITH DOT BELOW	U+1EE5	LATIN SMALL LETTER U WITH DOT BELOW
U+1EE6	LATIN CAPITAL LETTER U WITH HOOK ABOVE	U+1EE7	LATIN SMALL LETTER U WITH HOOK ABOVE
U+1EE8	LATIN CAPITAL LETTER U WITH HORN AND ACUTE	U+1EE9	LATIN SMALL LETTER U WITH HORN AND ACUTE
U+1EEA	LATIN CAPITAL LETTER U WITH HORN AND GRAVE	U+1EEB	LATIN SMALL LETTER U WITH HORN AND GRAVE
U+1EEC	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE	U+1EED	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
U+1EEE	LATIN CAPITAL LETTER U WITH HORN AND TILDE	U+1EEF	LATIN SMALL LETTER U WITH HORN AND TILDE
U+1EF0	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW	U+1EF1	LATIN SMALL LETTER U WITH HORN AND DOT BELOW
U+1EF2	LATIN CAPITAL LETTER Y WITH GRAVE	U+1EF3	LATIN SMALL LETTER Y WITH GRAVE
U+1EF4	LATIN CAPITAL LETTER Y WITH DOT BELOW	U+1EF5	LATIN SMALL LETTER Y WITH DOT BELOW
U+1EF6	LATIN CAPITAL LETTER Y WITH HOOK ABOVE	U+1EF7	LATIN SMALL LETTER Y WITH HOOK ABOVE
U+1EF8	LATIN CAPITAL LETTER Y WITH TILDE	U+1EF9	LATIN SMALL LETTER Y WITH TILDE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1EFA	LATIN CAPITAL LETTER MIDDLE-WELSH LL	U+1EFB	LATIN SMALL LETTER MIDDLE-WELSH LL
U+1EFC	LATIN CAPITAL LETTER MIDDLE-WELSH V	U+1EFD	LATIN SMALL LETTER MIDDLE-WELSH V
U+1EFE	LATIN CAPITAL LETTER Y WITH LOOP	U+1EFF	LATIN SMALL LETTER Y WITH LOOP
U+1F08	GREEK CAPITAL LETTER ALPHA WITH PSILI	U+1F00	GREEK SMALL LETTER ALPHA WITH PSILI
U+1F09	GREEK CAPITAL LETTER ALPHA WITH DASIA	U+1F01	GREEK SMALL LETTER ALPHA WITH DASIA
U+1F0A	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA	U+1F02	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA
U+1F0B	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA	U+1F03	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA
U+1F0C	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA	U+1F04	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA
U+1F0D	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA	U+1F05	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA
U+1F0E	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI	U+1F06	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI
U+1F0F	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI	U+1F07	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI
U+1F18	GREEK CAPITAL LETTER EPSILON WITH PSILI	U+1F10	GREEK SMALL LETTER EPSILON WITH PSILI
U+1F19	GREEK CAPITAL LETTER EPSILON WITH DASIA	U+1F11	GREEK SMALL LETTER EPSILON WITH DASIA
U+1F1A	GREEK CAPITAL LETTER EPSILON WITH PSILI AND VARIA	U+1F12	GREEK SMALL LETTER EPSILON WITH PSILI AND VARIA
U+1F1B	GREEK CAPITAL LETTER EPSILON WITH DASIA AND VARIA	U+1F13	GREEK SMALL LETTER EPSILON WITH DASIA AND VARIA
U+1F1C	GREEK CAPITAL LETTER EPSILON WITH PSILI AND OXIA	U+1F14	GREEK SMALL LETTER EPSILON WITH PSILI AND OXIA
U+1F1D	GREEK CAPITAL LETTER EPSILON WITH DASIA AND OXIA	U+1F15	GREEK SMALL LETTER EPSILON WITH DASIA AND OXIA
U+1F28	GREEK CAPITAL LETTER ETA WITH PSILI	U+1F20	GREEK SMALL LETTER ETA WITH PSILI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F29	GREEK CAPITAL LETTER ETA WITH DASIA	U+1F21	GREEK SMALL LETTER ETA WITH DASIA
U+1F2A	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA	U+1F22	GREEK SMALL LETTER ETA WITH PSILI AND VARIA
U+1F2B	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA	U+1F23	GREEK SMALL LETTER ETA WITH DASIA AND VARIA
U+1F2C	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA	U+1F24	GREEK SMALL LETTER ETA WITH PSILI AND OXIA
U+1F2D	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA	U+1F25	GREEK SMALL LETTER ETA WITH DASIA AND OXIA
U+1F2E	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI	U+1F26	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI
U+1F2F	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI	U+1F27	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI
U+1F38	GREEK CAPITAL LETTER IOTA WITH PSILI	U+1F30	GREEK SMALL LETTER IOTA WITH PSILI
U+1F39	GREEK CAPITAL LETTER IOTA WITH DASIA	U+1F31	GREEK SMALL LETTER IOTA WITH DASIA
U+1F3A	GREEK CAPITAL LETTER IOTA WITH PSILI AND VARIA	U+1F32	GREEK SMALL LETTER IOTA WITH PSILI AND VARIA
U+1F3B	GREEK CAPITAL LETTER IOTA WITH DASIA AND VARIA	U+1F33	GREEK SMALL LETTER IOTA WITH DASIA AND VARIA
U+1F3C	GREEK CAPITAL LETTER IOTA WITH PSILI AND OXIA	U+1F34	GREEK SMALL LETTER IOTA WITH PSILI AND OXIA
U+1F3D	GREEK CAPITAL LETTER IOTA WITH DASIA AND OXIA	U+1F35	GREEK SMALL LETTER IOTA WITH DASIA AND OXIA
U+1F3E	GREEK CAPITAL LETTER IOTA WITH PSILI AND PERISPOMENI	U+1F36	GREEK SMALL LETTER IOTA WITH PSILI AND PERISPOMENI
U+1F3F	GREEK CAPITAL LETTER IOTA WITH DASIA AND PERISPOMENI	U+1F37	GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI
U+1F48	GREEK CAPITAL LETTER OMICRON WITH PSILI	U+1F40	GREEK SMALL LETTER OMICRON WITH PSILI
U+1F49	GREEK CAPITAL LETTER OMICRON WITH DASIA	U+1F41	GREEK SMALL LETTER OMICRON WITH DASIA
U+1F4A	GREEK CAPITAL LETTER OMICRON WITH PSILI AND VARIA	U+1F42	GREEK SMALL LETTER OMICRON WITH PSILI AND VARIA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F4B	GREEK CAPITAL LETTER OMICRON WITH DASIA AND VARIA	U+1F43	GREEK SMALL LETTER OMICRON WITH DASIA AND VARIA
U+1F4C	GREEK CAPITAL LETTER OMICRON WITH PSILI AND OXIA	U+1F44	GREEK SMALL LETTER OMICRON WITH PSILI AND OXIA
U+1F4D	GREEK CAPITAL LETTER OMICRON WITH DASIA AND OXIA	U+1F45	GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA
U+1F59	GREEK CAPITAL LETTER UPSILON WITH DASIA	U+1F51	GREEK SMALL LETTER UPSILON WITH DASIA
U+1F5B	GREEK CAPITAL LETTER UPSILON WITH DASIA AND VARIA	U+1F53	GREEK SMALL LETTER UPSILON WITH DASIA AND VARIA
U+1F5D	GREEK CAPITAL LETTER UPSILON WITH DASIA AND OXIA	U+1F55	GREEK SMALL LETTER UPSILON WITH DASIA AND OXIA
U+1F5F	GREEK CAPITAL LETTER UPSILON WITH DASIA AND PERISPOMENI	U+1F57	GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI
U+1F68	GREEK CAPITAL LETTER OMEGA WITH PSILI	U+1F60	GREEK SMALL LETTER OMEGA WITH PSILI
U+1F69	GREEK CAPITAL LETTER OMEGA WITH DASIA	U+1F61	GREEK SMALL LETTER OMEGA WITH DASIA
U+1F6A	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA	U+1F62	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA
U+1F6B	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA	U+1F63	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA
U+1F6C	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA	U+1F64	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA
U+1F6D	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA	U+1F65	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA
U+1F6E	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI	U+1F66	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI
U+1F6F	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI	U+1F67	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI
U+1F88	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PROSGEGRAMMENIGREEK	U+1F80	SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F89	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PROSGEGRAMMENI	U+1F81	GREEK SMALL LETTER ALPHA WITH DASIA AND YPOGEGRAMMENI
U+1F8A	GREEK CAPITAL LETTER ALPHA WITH PSILI AND VARIA AND PROSGEGRAMMENI	U+1F82	GREEK SMALL LETTER ALPHA WITH PSILI AND VARIA AND YPOGEGRAMMENI
U+1F8B	GREEK CAPITAL LETTER ALPHA WITH DASIA AND VARIA AND PROSGEGRAMMENI	U+1F83	GREEK SMALL LETTER ALPHA WITH DASIA AND VARIA AND YPOGEGRAMMENI
U+1F8C	GREEK CAPITAL LETTER ALPHA WITH PSILI AND OXIA AND PROSGEGRAMMENI	U+1F84	GREEK SMALL LETTER ALPHA WITH PSILI AND OXIA AND YPOGEGRAMMENI
U+1F8D	GREEK CAPITAL LETTER ALPHA WITH DASIA AND OXIA AND PROSGEGRAMMENI	U+1F85	GREEK SMALL LETTER ALPHA WITH DASIA AND OXIA AND YPOGEGRAMMENI
U+1F8E	GREEK CAPITAL LETTER ALPHA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI	U+1F86	GREEK SMALL LETTER ALPHA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
U+1F8F	GREEK CAPITAL LETTER ALPHA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI	U+1F87	GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
U+1F98	GREEK CAPITAL LETTER ETA WITH PSILI AND PROSGEGRAMMENI	U+1F90	GREEK SMALL LETTER ETA WITH PSILI AND YPOGEGRAMMENI
U+1F99	GREEK CAPITAL LETTER ETA WITH DASIA AND PROSGEGRAMMENI	U+1F91	GREEK SMALL LETTER ETA WITH DASIA AND YPOGEGRAMMENI
U+1F9A	GREEK CAPITAL LETTER ETA WITH PSILI AND VARIA AND PROSGEGRAMMENI	U+1F92	GREEK SMALL LETTER ETA WITH PSILI AND VARIA AND YPOGEGRAMMENI
U+1F9B	GREEK CAPITAL LETTER ETA WITH DASIA AND VARIA AND PROSGEGRAMMENI	U+1F93	GREEK SMALL LETTER ETA WITH DASIA AND VARIA AND YPOGEGRAMMENI
U+1F9C	GREEK CAPITAL LETTER ETA WITH PSILI AND OXIA AND PROSGEGRAMMENI	U+1F94	GREEK SMALL LETTER ETA WITH PSILI AND OXIA AND YPOGEGRAMMENI
U+1F9D	GREEK CAPITAL LETTER ETA WITH DASIA AND OXIA AND PROSGEGRAMMENI	U+1F95	GREEK SMALL LETTER ETA WITH DASIA AND OXIA AND YPOGEGRAMMENI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1F9E	GREEK CAPITAL LETTER ETA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI	U+1F96	GREEK SMALL LETTER ETA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
U+1F9F	GREEK CAPITAL LETTER ETA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI	U+1F97	GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
U+1FA8	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PROSGEGRAMMENI	U+1FA0	GREEK SMALL LETTER OMEGA WITH PSILI AND YPOGEGRAMMENI
U+1FA9	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PROSGEGRAMMENI	U+1FA1	GREEK SMALL LETTER OMEGA WITH DASIA AND YPOGEGRAMMENI
U+1FAA	GREEK CAPITAL LETTER OMEGA WITH PSILI AND VARIA AND PROSGEGRAMMENI	U+1FA2	GREEK SMALL LETTER OMEGA WITH PSILI AND VARIA AND YPOGEGRAMMENI
U+1FAB	GREEK CAPITAL LETTER OMEGA WITH DASIA AND VARIA AND PROSGEGRAMMENI	U+1FA3	GREEK SMALL LETTER OMEGA WITH DASIA AND VARIA AND YPOGEGRAMMENI
U+1FAC	GREEK CAPITAL LETTER OMEGA WITH PSILI AND OXIA AND PROSGEGRAMMENI	U+1FA4	GREEK SMALL LETTER OMEGA WITH PSILI AND OXIA AND YPOGEGRAMMENI
U+1FAD	GREEK CAPITAL LETTER OMEGA WITH DASIA AND OXIA AND PROSGEGRAMMENI	U+1FA5	GREEK SMALL LETTER OMEGA WITH DASIA AND OXIA AND YPOGEGRAMMENI
U+1FAE	GREEK CAPITAL LETTER OMEGA WITH PSILI AND PERISPOMENI AND PROSGEGRAMMENI	U+1FA6	GREEK SMALL LETTER OMEGA WITH PSILI AND PERISPOMENI AND YPOGEGRAMMENI
U+1FAF	GREEK CAPITAL LETTER OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI	U+1FA7	GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI AND YPOGEGRAMMENI
U+1FB8	GREEK CAPITAL LETTER ALPHA WITH VRACHY	U+1FB0	GREEK SMALL LETTER ALPHA WITH VRACHY
U+1FB9	GREEK CAPITAL LETTER ALPHA WITH MACRON	U+1FB1	GREEK SMALL LETTER ALPHA WITH MACRON
U+1FBA	GREEK CAPITAL LETTER ALPHA WITH VARIA	U+1F70	GREEK SMALL LETTER ALPHA WITH VARIA
U+1FBB	GREEK CAPITAL LETTER ALPHA WITH OXIA	U+1F71	GREEK SMALL LETTER ALPHA WITH OXIA
U+1FBC	GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI	U+1FB3	GREEK SMALL LETTER ALPHA WITH YPOGEGRAMMENI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1FC8	GREEK CAPITAL LETTER EPSILON WITH VARIA	U+1F72	GREEK SMALL LETTER EPSILON WITH VARIA
U+1FC9	GREEK CAPITAL LETTER EPSILON WITH OXIA	U+1F73	GREEK SMALL LETTER EPSILON WITH OXIA
U+1FCA	GREEK CAPITAL LETTER ETA WITH VARIA	U+1F74	GREEK SMALL LETTER ETA WITH VARIA
U+1FCB	GREEK CAPITAL LETTER ETA WITH OXIA	U+1F75	GREEK SMALL LETTER ETA WITH OXIA
U+1FCC	GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI	U+1FC3	GREEK SMALL LETTER ETA WITH YPOGEGRAMMENI
U+1FD8	GREEK CAPITAL LETTER IOTA WITH VRACHY	U+1FD0	GREEK SMALL LETTER IOTA WITH VRACHY
U+1FD9	GREEK CAPITAL LETTER IOTA WITH MACRON	U+1FD1	GREEK SMALL LETTER IOTA WITH MACRON
U+1FDA	GREEK CAPITAL LETTER IOTA WITH VARIA	U+1F76	GREEK SMALL LETTER IOTA WITH VARIA
U+1FDB	GREEK CAPITAL LETTER IOTA WITH OXIA	U+1F77	GREEK SMALL LETTER IOTA WITH OXIA
U+1FE8	GREEK CAPITAL LETTER UPSILON WITH VRACHY	U+1FE0	GREEK SMALL LETTER UPSILON WITH VRACHY
U+1FE9	GREEK CAPITAL LETTER UPSILON WITH MACRON	U+1FE1	GREEK SMALL LETTER UPSILON WITH MACRON
U+1FEA	GREEK CAPITAL LETTER UPSILON WITH VARIA	U+1F7A	GREEK SMALL LETTER UPSILON WITH VARIA
U+1FEB	GREEK CAPITAL LETTER UPSILON WITH OXIA	U+1F7B	GREEK SMALL LETTER UPSILON WITH OXIA
U+1FEC	GREEK CAPITAL LETTER RHO WITH DASIA	U+1FE5	GREEK SMALL LETTER RHO WITH DASIA
U+1FF8	GREEK CAPITAL LETTER OMICRON WITH VARIA	U+1F78	GREEK SMALL LETTER OMICRON WITH VARIA
U+1FF9	GREEK CAPITAL LETTER OMICRON WITH OXIA	U+1F79	GREEK SMALL LETTER OMICRON WITH OXIA
U+1FFA	GREEK CAPITAL LETTER OMEGA WITH VARIA	U+1F7C	GREEK SMALL LETTER OMEGA WITH VARIA
U+1FFB	GREEK CAPITAL LETTER OMEGA WITH OXIA	U+1F7D	GREEK SMALL LETTER OMEGA WITH OXIA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+1FFC	GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI	U+1FF3	GREEK SMALL LETTER OMEGA WITH YPOGEGRAMMENI
U+2126	OHM SIGN	U+03C9	GREEK SMALL LETTER OMEGA
U+212A	KELVIN SIGN	U+006B	LATIN SMALL LETTER K
U+212B	ANGSTROM SIGN	U+00E5	LATIN SMALL LETTER A WITH RING ABOVE
U+2132	TURNED CAPITAL F	U+214E	TURNED SMALL F
U+2160	ROMAN NUMERAL ONE	U+2170	SMALL ROMAN NUMERAL ONE
U+2161	ROMAN NUMERAL TWO	U+2171	SMALL ROMAN NUMERAL TWO
U+2162	ROMAN NUMERAL THREE	U+2172	SMALL ROMAN NUMERAL THREE
U+2163	ROMAN NUMERAL FOUR	U+2173	SMALL ROMAN NUMERAL FOUR
U+2164	ROMAN NUMERAL FIVE	U+2174	SMALL ROMAN NUMERAL FIVE
U+2165	ROMAN NUMERAL SIX	U+2175	SMALL ROMAN NUMERAL SIX
U+2166	ROMAN NUMERAL SEVEN	U+2176	SMALL ROMAN NUMERAL SEVEN
U+2167	ROMAN NUMERAL EIGHT	U+2177	SMALL ROMAN NUMERAL EIGHT
U+2168	ROMAN NUMERAL NINE	U+2178	SMALL ROMAN NUMERAL NINE
U+2169	ROMAN NUMERAL TEN	U+2179	SMALL ROMAN NUMERAL TEN
U+216A	ROMAN NUMERAL ELEVEN	U+217A	SMALL ROMAN NUMERAL ELEVEN
U+216B	ROMAN NUMERAL TWELVE	U+217B	SMALL ROMAN NUMERAL TWELVE
U+216C	ROMAN NUMERAL FIFTY	U+217C	SMALL ROMAN NUMERAL FIFTY
U+216D	ROMAN NUMERAL ONE HUNDRED	U+217D	SMALL ROMAN NUMERAL ONE HUNDRED
U+216E	ROMAN NUMERAL FIVE HUNDRED	U+217E	SMALL ROMAN NUMERAL FIVE HUNDRED
U+216F	ROMAN NUMERAL ONE THOUSAND	U+217F	SMALL ROMAN NUMERAL ONE THOUSAND

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2183	ROMAN NUMERAL REVERSED ONE HUNDRED	U+2184	LATIN SMALL LETTER REVERSED C
U+24B6	CIRCLED LATIN CAPITAL LETTER A	U+24D0	CIRCLED LATIN SMALL LETTER A
U+24B7	CIRCLED LATIN CAPITAL LETTER B	U+24D1	CIRCLED LATIN SMALL LETTER B
U+24B8	CIRCLED LATIN CAPITAL LETTER C	U+24D2	CIRCLED LATIN SMALL LETTER C
U+24B9	CIRCLED LATIN CAPITAL LETTER D	U+24D3	CIRCLED LATIN SMALL LETTER D
U+24BA	CIRCLED LATIN CAPITAL LETTER E	U+24D4	CIRCLED LATIN SMALL LETTER E
U+24BB	CIRCLED LATIN CAPITAL LETTER F	U+24D5	CIRCLED LATIN SMALL LETTER F
U+24BC	CIRCLED LATIN CAPITAL LETTER G	U+24D6	CIRCLED LATIN SMALL LETTER G
U+24BD	CIRCLED LATIN CAPITAL LETTER H	U+24D7	CIRCLED LATIN SMALL LETTER H
U+24BE	CIRCLED LATIN CAPITAL LETTER I	U+24D8	CIRCLED LATIN SMALL LETTER I
U+24BF	CIRCLED LATIN CAPITAL LETTER J	U+24D9	CIRCLED LATIN SMALL LETTER J
U+24C0	CIRCLED LATIN CAPITAL LETTER K	U+24DA	CIRCLED LATIN SMALL LETTER K
U+24C1	CIRCLED LATIN CAPITAL LETTER L	U+24DB	CIRCLED LATIN SMALL LETTER L
U+24C2	CIRCLED LATIN CAPITAL LETTER M	U+24DC	CIRCLED LATIN SMALL LETTER M
U+24C3	CIRCLED LATIN CAPITAL LETTER N	U+24DD	CIRCLED LATIN SMALL LETTER N
U+24C4	CIRCLED LATIN CAPITAL LETTER O	U+24DE	CIRCLED LATIN SMALL LETTER O
U+24C5	CIRCLED LATIN CAPITAL LETTER P	U+24DF	CIRCLED LATIN SMALL LETTER P
U+24C6	CIRCLED LATIN CAPITAL LETTER Q	U+24E0	CIRCLED LATIN SMALL LETTER Q

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+24C7	CIRCLED LATIN CAPITAL LETTER R	U+24E1	CIRCLED LATIN SMALL LETTER R
U+24C8	CIRCLED LATIN CAPITAL LETTER S	U+24E2	CIRCLED LATIN SMALL LETTER S
U+24C9	CIRCLED LATIN CAPITAL LETTER T	U+24E3	CIRCLED LATIN SMALL LETTER T
U+24CA	CIRCLED LATIN CAPITAL LETTER U	U+24E4	CIRCLED LATIN SMALL LETTER U
U+24CB	CIRCLED LATIN CAPITAL LETTER V	U+24E5	CIRCLED LATIN SMALL LETTER V
U+24CC	CIRCLED LATIN CAPITAL LETTER W	U+24E6	CIRCLED LATIN SMALL LETTER W
U+24CD	CIRCLED LATIN CAPITAL LETTER X	U+24E7	CIRCLED LATIN SMALL LETTER X
U+24CE	CIRCLED LATIN CAPITAL LETTER Y	U+24E8	CIRCLED LATIN SMALL LETTER Y
U+24CF	CIRCLED LATIN CAPITAL LETTER Z	U+24E9	CIRCLED LATIN SMALL LETTER Z
U+2C00	GLAGOLITIC CAPITAL LETTER AZU	U+2C30	GLAGOLITIC SMALL LETTER AZU
U+2C01	GLAGOLITIC CAPITAL LETTER BUKY	U+2C31	GLAGOLITIC SMALL LETTER BUKY
U+2C02	GLAGOLITIC CAPITAL LETTER VEDE	U+2C32	GLAGOLITIC SMALL LETTER VEDE
U+2C03	GLAGOLITIC CAPITAL LETTER GLAGOLI	U+2C33	GLAGOLITIC SMALL LETTER GLAGOLI
U+2C04	GLAGOLITIC CAPITAL LETTER DOBRO	U+2C34	GLAGOLITIC SMALL LETTER DOBRO
U+2C05	GLAGOLITIC CAPITAL LETTER YESTU	U+2C35	GLAGOLITIC SMALL LETTER YESTU
U+2C06	GLAGOLITIC CAPITAL LETTER ZHIVETE	U+2C36	GLAGOLITIC SMALL LETTER ZHIVETE
U+2C07	GLAGOLITIC CAPITAL LETTER DZELO	U+2C37	GLAGOLITIC SMALL LETTER DZELO
U+2C08	GLAGOLITIC CAPITAL LETTER ZEMLJA	U+2C38	GLAGOLITIC SMALL LETTER ZEMLJA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C09	GLAGOLITIC CAPITAL LETTER IZHE	U+2C39	GLAGOLITIC SMALL LETTER IZHE
U+2C0A	GLAGOLITIC CAPITAL LETTER INITIAL IZHE	U+2C3A	GLAGOLITIC SMALL LETTER INITIAL IZHE
U+2C0B	GLAGOLITIC CAPITAL LETTER I	U+2C3B	GLAGOLITIC SMALL LETTER I
U+2C0C	GLAGOLITIC CAPITAL LETTER DJERVI	U+2C3C	GLAGOLITIC SMALL LETTER DJERVI
U+2C0D	GLAGOLITIC CAPITAL LETTER KAKO	U+2C3D	GLAGOLITIC SMALL LETTER KAKO
U+2C0E	GLAGOLITIC CAPITAL LETTER LJUDIJE	U+2C3E	GLAGOLITIC SMALL LETTER LJUDIJE
U+2C0F	GLAGOLITIC CAPITAL LETTER MYSLITE	U+2C3F	GLAGOLITIC SMALL LETTER MYSLITE
U+2C10	GLAGOLITIC CAPITAL LETTER NASHI	U+2C40	GLAGOLITIC SMALL LETTER NASHI
U+2C11	GLAGOLITIC CAPITAL LETTER ONU	U+2C41	GLAGOLITIC SMALL LETTER ONU
U+2C12	GLAGOLITIC CAPITAL LETTER POKOJI	U+2C42	GLAGOLITIC SMALL LETTER POKOJI
U+2C13	GLAGOLITIC CAPITAL LETTER RITSI	U+2C43	GLAGOLITIC SMALL LETTER RITSI
U+2C14	GLAGOLITIC CAPITAL LETTER SLOVO	U+2C44	GLAGOLITIC SMALL LETTER SLOVO
U+2C15	GLAGOLITIC CAPITAL LETTER TVRIDO	U+2C45	GLAGOLITIC SMALL LETTER TVRIDO
U+2C16	GLAGOLITIC CAPITAL LETTER UKU	U+2C46	GLAGOLITIC SMALL LETTER UKU
U+2C17	GLAGOLITIC CAPITAL LETTER FRITU	U+2C47	GLAGOLITIC SMALL LETTER FRITU
U+2C18	GLAGOLITIC CAPITAL LETTER HERU	U+2C48	GLAGOLITIC SMALL LETTER HERU
U+2C19	GLAGOLITIC CAPITAL LETTER OTU	U+2C49	GLAGOLITIC SMALL LETTER OTU
U+2C1A	GLAGOLITIC CAPITAL LETTER PE	U+2C4A	GLAGOLITIC SMALL LETTER PE
U+2C1B	GLAGOLITIC CAPITAL LETTER SHTA	U+2C4B	GLAGOLITIC SMALL LETTER SHTA

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C1C	GLAGOLITIC CAPITAL LETTER TSI	U+2C4C	GLAGOLITIC SMALL LETTER TSI
U+2C1D	GLAGOLITIC CAPITAL LETTER CHRIVI	U+2C4D	GLAGOLITIC SMALL LETTER CHRIVI
U+2C1E	GLAGOLITIC CAPITAL LETTER SHA	U+2C4E	GLAGOLITIC SMALL LETTER SHA
U+2C1F	GLAGOLITIC CAPITAL LETTER YERU	U+2C4F	GLAGOLITIC SMALL LETTER YERU
U+2C20	GLAGOLITIC CAPITAL LETTER YERI	U+2C50	GLAGOLITIC SMALL LETTER YERI
U+2C21	GLAGOLITIC CAPITAL LETTER YATI	U+2C51	GLAGOLITIC SMALL LETTER YATI
U+2C22	GLAGOLITIC CAPITAL LETTER SPIDERY HA	U+2C52	GLAGOLITIC SMALL LETTER SPIDERY HA
U+2C23	GLAGOLITIC CAPITAL LETTER YU	U+2C53	GLAGOLITIC SMALL LETTER YU
U+2C24	GLAGOLITIC CAPITAL LETTER SMALL YUS	U+2C54	GLAGOLITIC SMALL LETTER SMALL YUS
U+2C25	GLAGOLITIC CAPITAL LETTER SMALL YUS WITH TAIL	U+2C55	GLAGOLITIC SMALL LETTER SMALL YUS WITH TAIL
U+2C26	GLAGOLITIC CAPITAL LETTER YO	U+2C56	GLAGOLITIC SMALL LETTER YO
U+2C27	GLAGOLITIC CAPITAL LETTER IOTATED SMALL YUS	U+2C57	GLAGOLITIC SMALL LETTER IOTATED SMALL YUS
U+2C28	GLAGOLITIC CAPITAL LETTER BIG YUS	U+2C58	GLAGOLITIC SMALL LETTER BIG YUS
U+2C29	GLAGOLITIC CAPITAL LETTER IOTATED BIG YUS	U+2C59	GLAGOLITIC SMALL LETTER IOTATED BIG YUS
U+2C2A	GLAGOLITIC CAPITAL LETTER FITA	U+2C5A	GLAGOLITIC SMALL LETTER FITA
U+2C2B	GLAGOLITIC CAPITAL LETTER IZHITSA	U+2C5B	GLAGOLITIC SMALL LETTER IZHITSA
U+2C2C	GLAGOLITIC CAPITAL LETTER SHTAPIC	U+2C5C	GLAGOLITIC SMALL LETTER SHTAPIC
U+2C2D	GLAGOLITIC CAPITAL LETTER TROKUTASTI A	U+2C5D	GLAGOLITIC SMALL LETTER TROKUTASTI A
U+2C2E	GLAGOLITIC CAPITAL LETTER LATINATE MYSLITE	U+2C5E	GLAGOLITIC SMALL LETTER LATINATE MYSLITE

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C60	LATIN CAPITAL LETTER L WITH DOUBLE BAR	U+2C61	LATIN SMALL LETTER L WITH DOUBLE BAR
U+2C62	LATIN CAPITAL LETTER L WITH MIDDLE TILDE	U+026B	LATIN SMALL LETTER L WITH MIDDLE TILDE
U+2C63	LATIN CAPITAL LETTER P WITH STROKE	U+1D7D	LATIN SMALL LETTER P WITH STROKE
U+2C64	LATIN CAPITAL LETTER R WITH TAIL	U+027D	LATIN SMALL LETTER R WITH TAIL
U+2C67	LATIN CAPITAL LETTER H WITH DESCENDER	U+2C68	LATIN SMALL LETTER H WITH DESCENDER
U+2C69	LATIN CAPITAL LETTER K WITH DESCENDER	U+2C6A	LATIN SMALL LETTER K WITH DESCENDER
U+2C6B	LATIN CAPITAL LETTER Z WITH DESCENDER	U+2C6C	LATIN SMALL LETTER Z WITH DESCENDER
U+2C6D	LATIN CAPITAL LETTER ALPHA	U+0251	LATIN SMALL LETTER ALPHA
U+2C6E	LATIN CAPITAL LETTER M WITH HOOK	U+0271	LATIN SMALL LETTER M WITH HOOK
U+2C6F	LATIN CAPITAL LETTER TURNED A	U+0250	LATIN SMALL LETTER TURNED A
U+2C70	LATIN CAPITAL LETTER TURNED ALPHA	U+0252	LATIN SMALL LETTER TURNED ALPHA
U+2C72	LATIN CAPITAL LETTER W WITH HOOK	U+2C73	LATIN SMALL LETTER W WITH HOOK
U+2C75	LATIN CAPITAL LETTER HALF H	U+2C76	LATIN SMALL LETTER HALF H
U+2C7E	LATIN CAPITAL LETTER S WITH SWASH TAIL	U+023F	LATIN SMALL LETTER S WITH SWASH TAIL
U+2C7F	LATIN CAPITAL LETTER Z WITH SWASH TAIL	U+0240	LATIN SMALL LETTER Z WITH SWASH TAIL
U+2C80	COPTIC CAPITAL LETTER ALFA	U+2C81	COPTIC SMALL LETTER ALFA
U+2C82	COPTIC CAPITAL LETTER VIDA	U+2C83	COPTIC SMALL LETTER VIDA
U+2C84	COPTIC CAPITAL LETTER GAMMA	U+2C85	COPTIC SMALL LETTER GAMMA
U+2C86	COPTIC CAPITAL LETTER DALDA	U+2C87	COPTIC SMALL LETTER DALDA
U+2C88	COPTIC CAPITAL LETTER EIE	U+2C89	COPTIC SMALL LETTER EIE
U+2C8A	COPTIC CAPITAL LETTER SOU	U+2C8B	COPTIC SMALL LETTER SOU

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2C8C	COPTIC CAPITAL LETTER ZATA	U+2C8D	COPTIC SMALL LETTER ZATA
U+2C8E	COPTIC CAPITAL LETTER HATE	U+2C8F	COPTIC SMALL LETTER HATE
U+2C90	COPTIC CAPITAL LETTER THETHE	U+2C91	COPTIC SMALL LETTER THETHE
U+2C92	COPTIC CAPITAL LETTER IAUDA	U+2C93	COPTIC SMALL LETTER IAUDA
U+2C94	COPTIC CAPITAL LETTER KAPA	U+2C95	COPTIC SMALL LETTER KAPA
U+2C96	COPTIC CAPITAL LETTER LAULA	U+2C97	COPTIC SMALL LETTER LAULA
U+2C98	COPTIC CAPITAL LETTER MI	U+2C99	COPTIC SMALL LETTER MI
U+2C9A	COPTIC CAPITAL LETTER NI	U+2C9B	COPTIC SMALL LETTER NI
U+2C9C	COPTIC CAPITAL LETTER KSI	U+2C9D	COPTIC SMALL LETTER KSI
U+2C9E	COPTIC CAPITAL LETTER O	U+2C9F	COPTIC SMALL LETTER O
U+2CA0	COPTIC CAPITAL LETTER PI	U+2CA1	COPTIC SMALL LETTER PI
U+2CA2	COPTIC CAPITAL LETTER RO	U+2CA3	COPTIC SMALL LETTER RO
U+2CA4	COPTIC CAPITAL LETTER SIMA	U+2CA5	COPTIC SMALL LETTER SIMA
U+2CA6	COPTIC CAPITAL LETTER TAU	U+2CA7	COPTIC SMALL LETTER TAU
U+2CA8	COPTIC CAPITAL LETTER UA	U+2CA9	COPTIC SMALL LETTER UA
U+2CAA	COPTIC CAPITAL LETTER FI	U+2CAB	COPTIC SMALL LETTER FI
U+2CAC	COPTIC CAPITAL LETTER KHI	U+2CAD	COPTIC SMALL LETTER KHI
U+2CAE	COPTIC CAPITAL LETTER PSI	U+2CAF	COPTIC SMALL LETTER PSI
U+2CB0	COPTIC CAPITAL LETTER OOU	U+2CB1	COPTIC SMALL LETTER OOU
U+2CB2	COPTIC CAPITAL LETTER DIALECT-P ALEF	U+2CB3	COPTIC SMALL LETTER DIALECT-P ALEF
U+2CB4	COPTIC CAPITAL LETTER OLD COPTIC AIN	U+2CB5	COPTIC SMALL LETTER OLD COPTIC AIN
U+2CB6	COPTIC CAPITAL LETTER CRYPTOGRAMMIC EIE	U+2CB7	COPTIC SMALL LETTER CRYPTOGRAMMIC EIE
U+2CB8	COPTIC CAPITAL LETTER DIALECT-P KAPA	U+2CB9	COPTIC SMALL LETTER DIALECT-P KAPA
U+2CBA	COPTIC CAPITAL LETTER DIALECT-P NI	U+2CBB	COPTIC SMALL LETTER DIALECT-P NI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2CBC	COPTIC CAPITAL LETTER CRYPTOGRAMMIC NI	U+2CBD	COPTIC SMALL LETTER CRYPTOGRAMMIC NI
U+2CBE	COPTIC CAPITAL LETTER OLD COPTIC OOU	U+2CBF	COPTIC SMALL LETTER OLD COPTIC OOU
U+2CC0	COPTIC CAPITAL LETTER SAMPI	U+2CC1	COPTIC SMALL LETTER SAMPI
U+2CC2	COPTIC CAPITAL LETTER CROSSED SHEI	U+2CC3	COPTIC SMALL LETTER CROSSED SHEI
U+2CC4	COPTIC CAPITAL LETTER OLD COPTIC SHEI	U+2CC5	COPTIC SMALL LETTER OLD COPTIC SHEI
U+2CC6	COPTIC CAPITAL LETTER OLD COPTIC ESH	U+2CC7	COPTIC SMALL LETTER OLD COPTIC ESH
U+2CC8	COPTIC CAPITAL LETTER AKHMIMIC KHEI	U+2CC9	COPTIC SMALL LETTER AKHMIMIC KHEI
U+2CCA	COPTIC CAPITAL LETTER DIALECT-P HORI	U+2CCB	COPTIC SMALL LETTER DIALECT-P HORI
U+2CCC	COPTIC CAPITAL LETTER OLD COPTIC HORI	U+2CCD	COPTIC SMALL LETTER OLD COPTIC HORI
U+2CCE	COPTIC CAPITAL LETTER OLD COPTIC HA	U+2CCF	COPTIC SMALL LETTER OLD COPTIC HA
U+2CD0	COPTIC CAPITAL LETTER L-SHAPED HA	U+2CD1	COPTIC SMALL LETTER L-SHAPED HA
U+2CD2	COPTIC CAPITAL LETTER OLD COPTIC HEI	U+2CD3	COPTIC SMALL LETTER OLD COPTIC HEI
U+2CD4	COPTIC CAPITAL LETTER OLD COPTIC HAT	U+2CD5	COPTIC SMALL LETTER OLD COPTIC HAT
U+2CD6	COPTIC CAPITAL LETTER OLD COPTIC GANGIA	U+2CD7	COPTIC SMALL LETTER OLD COPTIC GANGIA
U+2CD8	COPTIC CAPITAL LETTER OLD COPTIC DJA	U+2CD9	COPTIC SMALL LETTER OLD COPTIC DJA
U+2CDA	COPTIC CAPITAL LETTER OLD COPTIC SHIMA	U+2CDB	COPTIC SMALL LETTER OLD COPTIC SHIMA
U+2CDC	COPTIC CAPITAL LETTER OLD NUBIAN SHIMA	U+2CDD	COPTIC SMALL LETTER OLD NUBIAN SHIMA
U+2CDE	COPTIC CAPITAL LETTER OLD NUBIAN NGI	U+2CDF	COPTIC SMALL LETTER OLD NUBIAN NGI

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+2CE0	COPTIC CAPITAL LETTER OLD NUBIAN NYI	U+2CE1	COPTIC SMALL LETTER OLD NUBIAN NYI
U+2CE2	COPTIC CAPITAL LETTER OLD NUBIAN WAU	U+2CE3	COPTIC SMALL LETTER OLD NUBIAN WAU
U+2CEB	COPTIC CAPITAL LETTER CRYPTOGRAMMIC SHEI	U+2CEC	COPTIC SMALL LETTER CRYPTOGRAMMIC SHEI
U+2CED	COPTIC CAPITAL LETTER CRYPTOGRAMMIC GANGIA	U+2CEE	COPTIC SMALL LETTER CRYPTOGRAMMIC GANGIA
U+A640	CYRILLIC CAPITAL LETTER ZEMLYA	U+A641	CYRILLIC SMALL LETTER ZEMLYA
U+A642	CYRILLIC CAPITAL LETTER DZELO	U+A643	CYRILLIC SMALL LETTER DZELO
U+A644	CYRILLIC CAPITAL LETTER REVERSED DZE	U+A645	CYRILLIC SMALL LETTER REVERSED DZE
U+A646	CYRILLIC CAPITAL LETTER IOTA	U+A647	CYRILLIC SMALL LETTER IOTA
U+A648	CYRILLIC CAPITAL LETTER DJERV	U+A649	CYRILLIC SMALL LETTER DJERV
U+A64A	CYRILLIC CAPITAL LETTER MONOGRAPH UK	U+A64B	CYRILLIC SMALL LETTER MONOGRAPH UK
U+A64C	CYRILLIC CAPITAL LETTER BROAD OMEGA	U+A64D	CYRILLIC SMALL LETTER BROAD OMEGA
U+A64E	CYRILLIC CAPITAL LETTER NEUTRAL YER	U+A64F	CYRILLIC SMALL LETTER NEUTRAL YER
U+A650	CYRILLIC CAPITAL LETTER YERU WITH BACK YER	U+A651	CYRILLIC SMALL LETTER YERU WITH BACK YER
U+A652	CYRILLIC CAPITAL LETTER IOTIFIED YAT	U+A653	CYRILLIC SMALL LETTER IOTIFIED YAT
U+A654	CYRILLIC CAPITAL LETTER REVERSED YU	U+A655	CYRILLIC SMALL LETTER REVERSED YU
U+A656	CYRILLIC CAPITAL LETTER IOTIFIED A	U+A657	CYRILLIC SMALL LETTER IOTIFIED A
U+A658	CYRILLIC CAPITAL LETTER CLOSED LITTLE YUS	U+A659	CYRILLIC SMALL LETTER CLOSED LITTLE YUS
U+A65A	CYRILLIC CAPITAL LETTER BLENDED YUS	U+A65B	CYRILLIC SMALL LETTER BLENDED YUS

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A65C	CYRILLIC CAPITAL LETTER IOTIFIED CLOSED LITTLE YUS	U+A65D	CYRILLIC SMALL LETTER IOTIFIED CLOSED LITTLE YUS
U+A65E	CYRILLIC CAPITAL LETTER YN	U+A65F	CYRILLIC SMALL LETTER YN
U+A660	CYRILLIC CAPITAL LETTER REVERSED TSE	U+A661	CYRILLIC SMALL LETTER REVERSED TSE
U+A662	CYRILLIC CAPITAL LETTER SOFT DE	U+A663	CYRILLIC SMALL LETTER SOFT DE
U+A664	CYRILLIC CAPITAL LETTER SOFT EL	U+A665	CYRILLIC SMALL LETTER SOFT EL
U+A666	CYRILLIC CAPITAL LETTER SOFT EM	U+A667	CYRILLIC SMALL LETTER SOFT EM
U+A668	CYRILLIC CAPITAL LETTER MONOCULAR O	U+A669	CYRILLIC SMALL LETTER MONOCULAR O
U+A66A	CYRILLIC CAPITAL LETTER BINOCULAR O	U+A66B	CYRILLIC SMALL LETTER BINOCULAR O
U+A66C	CYRILLIC CAPITAL LETTER DOUBLE MONOCULAR O	U+A66D	CYRILLIC SMALL LETTER DOUBLE MONOCULAR O
U+A680	CYRILLIC CAPITAL LETTER DWE	U+A681	CYRILLIC SMALL LETTER DWE
U+A682	CYRILLIC CAPITAL LETTER DZWE	U+A683	CYRILLIC SMALL LETTER DZWE
U+A684	CYRILLIC CAPITAL LETTER ZHWE	U+A685	CYRILLIC SMALL LETTER ZHWE
U+A686	CYRILLIC CAPITAL LETTER CCHE	U+A687	CYRILLIC SMALL LETTER CCHE
U+A688	CYRILLIC CAPITAL LETTER DZZE	U+A689	CYRILLIC SMALL LETTER DZZE
U+A68A	CYRILLIC CAPITAL LETTER TE WITH MIDDLE HOOK	U+A68B	CYRILLIC SMALL LETTER TE WITH MIDDLE HOOK
U+A68C	CYRILLIC CAPITAL LETTER TWE	U+A68D	CYRILLIC SMALL LETTER TWE
U+A68E	CYRILLIC CAPITAL LETTER TSWE	U+A68F	CYRILLIC SMALL LETTER TSWE
U+A690	CYRILLIC CAPITAL LETTER TSSE	U+A691	CYRILLIC SMALL LETTER TSSE
U+A692	CYRILLIC CAPITAL LETTER TCHE	U+A693	CYRILLIC SMALL LETTER TCHE
U+A694	CYRILLIC CAPITAL LETTER HWE	U+A695	CYRILLIC SMALL LETTER HWE
U+A696	CYRILLIC CAPITAL LETTER SHWE	U+A697	CYRILLIC SMALL LETTER SHWE
U+A722	LATIN CAPITAL LETTER EGYPTOLOGICAL ALEF	U+A723	LATIN SMALL LETTER EGYPTOLOGICAL ALEF

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A724	LATIN CAPITAL LETTER EGYPTOLOGICAL AIN	U+A725	LATIN SMALL LETTER EGYPTOLOGICAL AIN
U+A726	LATIN CAPITAL LETTER HENG	U+A727	LATIN SMALL LETTER HENG
U+A728	LATIN CAPITAL LETTER TZ	U+A729	LATIN SMALL LETTER TZ
U+A72A	LATIN CAPITAL LETTER TRESILLO	U+A72B	LATIN SMALL LETTER TRESILLO
U+A72C	LATIN CAPITAL LETTER CUATRILLO	U+A72D	LATIN SMALL LETTER CUATRILLO
U+A72E	LATIN CAPITAL LETTER CUATRILLO WITH COMMA	U+A72F	LATIN SMALL LETTER CUATRILLO WITH COMMA
U+A732	LATIN CAPITAL LETTER AA	U+A733	LATIN SMALL LETTER AA
U+A734	LATIN CAPITAL LETTER AO	U+A735	LATIN SMALL LETTER AO
U+A736	LATIN CAPITAL LETTER AU	U+A737	LATIN SMALL LETTER AU
U+A738	LATIN CAPITAL LETTER AV	U+A739	LATIN SMALL LETTER AV
U+A73A	LATIN CAPITAL LETTER AV WITH HORIZONTAL BAR	U+A73B	LATIN SMALL LETTER AV WITH HORIZONTAL BAR
U+A73C	LATIN CAPITAL LETTER AY	U+A73D	LATIN SMALL LETTER AY
U+A73E	LATIN CAPITAL LETTER REVERSED C WITH DOT	U+A73F	LATIN SMALL LETTER REVERSED C WITH DOT
U+A740	LATIN CAPITAL LETTER K WITH STROKE	U+A741	LATIN SMALL LETTER K WITH STROKE
U+A742	LATIN CAPITAL LETTER K WITH DIAGONAL STROKE	U+A743	LATIN SMALL LETTER K WITH DIAGONAL STROKE
U+A744	LATIN CAPITAL LETTER K WITH STROKE AND DIAGONAL STROKE	U+A745	LATIN SMALL LETTER K WITH STROKE AND DIAGONAL STROKE
U+A746	LATIN CAPITAL LETTER BROKEN L	U+A747	LATIN SMALL LETTER BROKEN L
U+A748	LATIN CAPITAL LETTER L WITH HIGH STROKE	U+A749	LATIN SMALL LETTER L WITH HIGH STROKE
U+A74A	LATIN CAPITAL LETTER O WITH LONG STROKE OVERLAY	U+A74B	LATIN SMALL LETTER O WITH LONG STROKE OVERLAY
U+A74C	LATIN CAPITAL LETTER O WITH LOOP	U+A74D	LATIN SMALL LETTER O WITH LOOP
U+A74E	LATIN CAPITAL LETTER OO	U+A74F	LATIN SMALL LETTER OO

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A750	LATIN CAPITAL LETTER P WITH STROKE THROUGH DESCENDER	U+A751	LATIN SMALL LETTER P WITH STROKE THROUGH DESCENDER
U+A752	LATIN CAPITAL LETTER P WITH FLOURISH	U+A753	LATIN SMALL LETTER P WITH FLOURISH
U+A754	LATIN CAPITAL LETTER P WITH SQUIRREL TAIL	U+A755	LATIN SMALL LETTER P WITH SQUIRREL TAIL
U+A756	LATIN CAPITAL LETTER Q WITH STROKE THROUGH DESCENDER	U+A757	LATIN SMALL LETTER Q WITH STROKE THROUGH DESCENDER
U+A758	LATIN CAPITAL LETTER Q WITH DIAGONAL STROKE	U+A759	LATIN SMALL LETTER Q WITH DIAGONAL STROKE
U+A75A	LATIN CAPITAL LETTER R ROTUNDA	U+A75B	LATIN SMALL LETTER R ROTUNDA
U+A75C	LATIN CAPITAL LETTER RUM ROTUNDA	U+A75D	LATIN SMALL LETTER RUM ROTUNDA
U+A75E	LATIN CAPITAL LETTER V WITH DIAGONAL STROKE	U+A75F	LATIN SMALL LETTER V WITH DIAGONAL STROKE
U+A760	LATIN CAPITAL LETTER VY	U+A761	LATIN SMALL LETTER VY
U+A762	LATIN CAPITAL LETTER VISIGOTHIC Z	U+A763	LATIN SMALL LETTER VISIGOTHIC Z
U+A764	LATIN CAPITAL LETTER THORN WITH STROKE	U+A765	LATIN SMALL LETTER THORN WITH STROKE
U+A766	LATIN CAPITAL LETTER THORN WITH STROKE THROUGH DESCENDER	U+A767	LATIN SMALL LETTER THORN WITH STROKE THROUGH DESCENDER
U+A768	LATIN CAPITAL LETTER VEND	U+A769	LATIN SMALL LETTER VEND
U+A76A	LATIN CAPITAL LETTER ET	U+A76B	LATIN SMALL LETTER ET
U+A76C	LATIN CAPITAL LETTER IS	U+A76D	LATIN SMALL LETTER IS
U+A76E	LATIN CAPITAL LETTER CON	U+A76F	LATIN SMALL LETTER CON
U+A779	LATIN CAPITAL LETTER INSULAR D	U+A77A	LATIN SMALL LETTER INSULAR D
U+A77B	LATIN CAPITAL LETTER INSULAR F	U+A77C	LATIN SMALL LETTER INSULAR F
U+A77D	LATIN CAPITAL LETTER INSULAR G	U+1D79	LATIN SMALL LETTER INSULAR G

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+A77E	LATIN CAPITAL LETTER TURNED INSULAR G	U+A77F	LATIN SMALL LETTER TURNED INSULAR G
U+A780	LATIN CAPITAL LETTER TURNED L	U+A781	LATIN SMALL LETTER TURNED L
U+A782	LATIN CAPITAL LETTER INSULAR R	U+A783	LATIN SMALL LETTER INSULAR R
U+A784	LATIN CAPITAL LETTER INSULAR S	U+A785	LATIN SMALL LETTER INSULAR S
U+A786	LATIN CAPITAL LETTER INSULAR T	U+A787	LATIN SMALL LETTER INSULAR T
U+A78B	LATIN CAPITAL LETTER SALTILLO	U+A78C	LATIN SMALL LETTER SALTILLO
U+A78D	LATIN CAPITAL LETTER TURNED H	U+0265	LATIN SMALL LETTER TURNED H
U+A790	LATIN CAPITAL LETTER N WITH DESCENDER	U+A791	LATIN SMALL LETTER N WITH DESCENDER
U+A7A0	LATIN CAPITAL LETTER G WITH OBLIQUE STROKE	U+A7A1	LATIN SMALL LETTER G WITH OBLIQUE STROKE
U+A7A2	LATIN CAPITAL LETTER K WITH OBLIQUE STROKE	U+A7A3	LATIN SMALL LETTER K WITH OBLIQUE STROKE
U+A7A4	LATIN CAPITAL LETTER N WITH OBLIQUE STROKE	U+A7A5	LATIN SMALL LETTER N WITH OBLIQUE STROKE
U+A7A6	LATIN CAPITAL LETTER R WITH OBLIQUE STROKE	U+A7A7	LATIN SMALL LETTER R WITH OBLIQUE STROKE
U+A7A8	LATIN CAPITAL LETTER S WITH OBLIQUE STROKE	U+A7A9	LATIN SMALL LETTER S WITH OBLIQUE STROKE
U+FF21	FULLWIDTH LATIN CAPITAL LETTER A	U+FF41	FULLWIDTH LATIN SMALL LETTER A
U+FF22	FULLWIDTH LATIN CAPITAL LETTER B	U+FF42	FULLWIDTH LATIN SMALL LETTER B
U+FF23	FULLWIDTH LATIN CAPITAL LETTER C	U+FF43	FULLWIDTH LATIN SMALL LETTER C
U+FF24	FULLWIDTH LATIN CAPITAL LETTER D	U+FF44	FULLWIDTH LATIN SMALL LETTER D
U+FF25	FULLWIDTH LATIN CAPITAL LETTER E	U+FF45	FULLWIDTH LATIN SMALL LETTER E

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF26	FULLWIDTH LATIN CAPITAL LETTER F	U+FF46	FULLWIDTH LATIN SMALL LETTER F
U+FF27	FULLWIDTH LATIN CAPITAL LETTER G	U+FF47	FULLWIDTH LATIN SMALL LETTER G
U+FF28	FULLWIDTH LATIN CAPITAL LETTER H	U+FF48	FULLWIDTH LATIN SMALL LETTER H
U+FF29	FULLWIDTH LATIN CAPITAL LETTER I	U+FF49	FULLWIDTH LATIN SMALL LETTER I
U+FF2A	FULLWIDTH LATIN CAPITAL LETTER J	U+FF4A	FULLWIDTH LATIN SMALL LETTER J
U+FF2B	FULLWIDTH LATIN CAPITAL LETTER K	U+FF4B	FULLWIDTH LATIN SMALL LETTER K
U+FF2C	FULLWIDTH LATIN CAPITAL LETTER L	U+FF4C	FULLWIDTH LATIN SMALL LETTER L
U+FF2D	FULLWIDTH LATIN CAPITAL LETTER M	U+FF4D	FULLWIDTH LATIN SMALL LETTER M
U+FF2E	FULLWIDTH LATIN CAPITAL LETTER N	U+FF4E	FULLWIDTH LATIN SMALL LETTER N
U+FF2F	FULLWIDTH LATIN CAPITAL LETTER O	U+FF4F	FULLWIDTH LATIN SMALL LETTER O
U+FF30	FULLWIDTH LATIN CAPITAL LETTER P	U+FF50	FULLWIDTH LATIN SMALL LETTER P
U+FF31	FULLWIDTH LATIN CAPITAL LETTER Q	U+FF51	FULLWIDTH LATIN SMALL LETTER Q
U+FF32	FULLWIDTH LATIN CAPITAL LETTER R	U+FF52	FULLWIDTH LATIN SMALL LETTER R
U+FF33	FULLWIDTH LATIN CAPITAL LETTER S	U+FF53	FULLWIDTH LATIN SMALL LETTER S
U+FF34	FULLWIDTH LATIN CAPITAL LETTER T	U+FF54	FULLWIDTH LATIN SMALL LETTER T
U+FF35	FULLWIDTH LATIN CAPITAL LETTER U	U+FF55	FULLWIDTH LATIN SMALL LETTER U
U+FF36	FULLWIDTH LATIN CAPITAL LETTER V	U+FF56	FULLWIDTH LATIN SMALL LETTER V
U+FF37	FULLWIDTH LATIN CAPITAL LETTER W	U+FF57	FULLWIDTH LATIN SMALL LETTER W

UNICODE Uppercase Output		UNICODE Lowercase Output	
Code	Unicode Character Name	Code	Unicode Character Name
U+FF38	FULLWIDTH LATIN CAPITAL LETTER X	U+FF58	FULLWIDTH LATIN SMALL LETTER X
U+FF39	FULLWIDTH LATIN CAPITAL LETTER Y	U+FF59	FULLWIDTH LATIN SMALL LETTER Y
U+FF3A	FULLWIDTH LATIN CAPITAL LETTER Z	U+FF5A	FULLWIDTH LATIN SMALL LETTER Z

KanjiShiftJIS Mappings

Overview

The following lists KANJISJIS-related mappings.

Client KANJISJIS_0S Single-Byte to Server KANJI1

The following table shows how Teradata Database maps single-byte characters in the KANJISJIS_0S client character set (JIS-x0201) to the KANJI1 server character set.

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0x00	0x00	<control> NULL
0x01	0x01	<control> START OF HEADING
0x02	0x02	<control> START OF TEXT
0x03	0x03	<control> END OF TEXT
0x04	0x04	<control> END OF TRANSMISSION
0x05	0x05	<control> ENQUIRY
0x06	0x06	<control> ACKNOWLEDGE
0x07	0x07	<control> BELL
0x08	0x08	<control> BACKSPACE
0x09	0x09	<control> CHARACTER TABULATION (horizontal tabulation)
0x0A	0x0A	<control> LINE FEED
0x0B	0x0B	<control> LINE TABULATION (vertical tabulation)
0x0C	0x0C	<control> FORM FEED
0x0D	0x0D	<control> CARRIAGE RETURN
0x0E	0x0E	<control> SHIFT OUT
0x0F	0x0F	<control> SHIFT IN
0x10	0x10	<control> DATA LINK ESCAPE
0x11	0x11	<control> DEVICE CONTROL ONE
0x12	0x12	<control> DEVICE CONTROL TWO

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0x13	0x13	<control> DEVICE CONTROL THREE
0x14	0x14	<control> DEVICE CONTROL FOUR
0x15	0x15	<control> NEGATIVE ACKNOWLEDGE
0x16	0x16	<control> SYNCHRONOUS IDLE
0x17	0x17	<control> END OF TRANSMISSION BLOCK
0x18	0x18	<control> CANCEL
0x19	0x19	<control> END OF MEDIUM
0x1A	0x1A	<control> SUBSTITUTE
0x1B	0x1B	<control> ESCAPE
0x1C	0x1C	<control> INFORMATION SEPARATOR FOUR (file separator)
0x1D	0x1D	<control> INFORMATION SEPARATOR THREE (group separator)
0x1E	0x1E	<control> INFORMATION SEPARATOR TWO (record separator)
0x1F	0x1F	<control> INFORMATION SEPARATOR ONE (unit separator)
0x20	0x20	SPACE
0x21	0x21	EXCLAMATION MARK
0x22	0x22	QUOTATION MARK
0x23	0x23	NUMBER SIGN
0x24	0x24	DOLLAR SIGN
0x25	0x25	PERCENT SIGN
0x26	0x26	AMPERSAND
0x27	0x27	APOSTROPHE
0x28	0x28	LEFT PARENTHESIS
0x29	0x29	RIGHT PARENTHESIS
0x2A	0x2A	ASTERISK
0x2B	0x2B	PLUS SIGN
0x2C	0x2C	COMMA
0x2D	0x2D	HYPHEN-MINUS
0x2E	0x2E	FULL STOP

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0x2F	0x2F	SOLIDUS
0x30	0x30	DIGIT ZERO
0x31	0x31	DIGIT ONE
0x32	0x32	DIGIT TWO
0x33	0x33	DIGIT THREE
0x34	0x34	DIGIT FOUR
0x35	0x35	DIGIT FIVE
0x36	0x36	DIGIT SIX
0x37	0x37	DIGIT SEVEN
0x38	0x38	DIGIT EIGHT
0x39	0x39	DIGIT NINE
0x3A	0x3A	COLON
0x3B	0x3B	SEMICOLON
0x3C	0x3C	LESS-THAN SIGN
0x3D	0x3D	EQUALS SIGN
0x3E	0x3E	GREATER-THAN SIGN
0x3F	0x3F	QUESTION MARK
0x40	0x40	COMMERCIAL AT
0x41	0x41	LATIN CAPITAL LETTER A
0x42	0x42	LATIN CAPITAL LETTER B
0x43	0x43	LATIN CAPITAL LETTER C
0x44	0x44	LATIN CAPITAL LETTER D
0x45	0x45	LATIN CAPITAL LETTER E
0x46	0x46	LATIN CAPITAL LETTER F
0x47	0x47	LATIN CAPITAL LETTER G
0x48	0x48	LATIN CAPITAL LETTER H
0x49	0x49	LATIN CAPITAL LETTER I
0x4A	0x4A	LATIN CAPITAL LETTER J

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0x4B	0x4B	LATIN CAPITAL LETTER K
0x4C	0x4C	LATIN CAPITAL LETTER L
0x4D	0x4D	LATIN CAPITAL LETTER M
0x4E	0x4E	LATIN CAPITAL LETTER N
0x4F	0x4F	LATIN CAPITAL LETTER O
0x50	0x50	LATIN CAPITAL LETTER P
0x51	0x51	LATIN CAPITAL LETTER Q
0x52	0x52	LATIN CAPITAL LETTER R
0x53	0x53	LATIN CAPITAL LETTER S
0x54	0x54	LATIN CAPITAL LETTER T
0x55	0x55	LATIN CAPITAL LETTER U
0x56	0x56	LATIN CAPITAL LETTER V
0x57	0x57	LATIN CAPITAL LETTER W
0x58	0x58	LATIN CAPITAL LETTER X
0x59	0x59	LATIN CAPITAL LETTER Y
0x5A	0x5A	LATIN CAPITAL LETTER Z
0x5B	0x5B	LEFT SQUARE BRACKET
0x5C	0x5C	YEN SIGN
0x5D	0x5D	RIGHT SQUARE BRACKET
0x5E	0x5E	CIRCUMFLEX ACCENT
0x5F	0x5F	LOW LINE
0x60	0x60	GRAVE ACCENT
0x61	0x61	LATIN SMALL LETTER A
0x62	0x62	LATIN SMALL LETTER B
0x63	0x63	LATIN SMALL LETTER C
0x64	0x64	LATIN SMALL LETTER D
0x65	0x65	LATIN SMALL LETTER E
0x66	0x66	LATIN SMALL LETTER F

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0x67	0x67	LATIN SMALL LETTER G
0x68	0x68	LATIN SMALL LETTER H
0x69	0x69	LATIN SMALL LETTER I
0x6A	0x6A	LATIN SMALL LETTER J
0x6B	0x6B	LATIN SMALL LETTER K
0x6C	0x6C	LATIN SMALL LETTER L
0x6D	0x6D	LATIN SMALL LETTER M
0x6E	0x6E	LATIN SMALL LETTER N
0x6F	0x6F	LATIN SMALL LETTER O
0x70	0x70	LATIN SMALL LETTER P
0x71	0x71	LATIN SMALL LETTER Q
0x72	0x72	LATIN SMALL LETTER R
0x73	0x73	LATIN SMALL LETTER S
0x74	0x74	LATIN SMALL LETTER T
0x75	0x75	LATIN SMALL LETTER U
0x76	0x76	LATIN SMALL LETTER V
0x77	0x77	LATIN SMALL LETTER W
0x78	0x78	LATIN SMALL LETTER X
0x79	0x79	LATIN SMALL LETTER Y
0x7A	0x7A	LATIN SMALL LETTER Z
0x7B	0x7B	LEFT CURLY BRACKET
0x7C	0x7C	VERTICAL LINE
0x7D	0x7D	RIGHT CURLY BRACKET
0x7E	0x7E	OVERLINE
0x7F	0x7F	DELETE, RUBOUT
0xA1	0xA1	HALFWIDTH IDEOGRAPHIC FULL STOP
0xA2	0xA2	HALFWIDTH LEFT CORNER BRACKET
0xA3	0xA3	HALFWIDTH RIGHT CORNER BRACKET

KANJISJIS _0S Single-Byte	KANJI1	Unicode Character Name
0xA4	0xA4	HALFWIDTH IDEOGRAPHIC COMMA
0xA5	0xA5	HALFWIDTH KATAKANA MIDDLE DOT
0xA6	0xA6	HALFWIDTH KATAKANA LETTER WO
0xA7	0xA7	HALFWIDTH KATAKANA LETTER SMALL A
0xA8	0xA8	HALFWIDTH KATAKANA LETTER SMALL I
0xA9	0xA9	HALFWIDTH KATAKANA LETTER SMALL U
0xAA	0xAA	HALFWIDTH KATAKANA LETTER SMALL E
0xAB	0xAB	HALFWIDTH KATAKANA LETTER SMALL O
0xAC	0xAC	HALFWIDTH KATAKANA LETTER SMALL YA
0xAD	0xAD	HALFWIDTH KATAKANA LETTER SMALL YU
0xAE	0xAE	HALFWIDTH KATAKANA LETTER SMALL YO
0xAF	0xAF	HALFWIDTH KATAKANA LETTER SMALL TU
0xB0	0xB0	HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK
0xB1	0xB1	HALFWIDTH KATAKANA LETTER A
0xB2	0xB2	HALFWIDTH KATAKANA LETTER I
0xB3	0xB3	HALFWIDTH KATAKANA LETTER U
0xB4	0xB4	HALFWIDTH KATAKANA LETTER E
0xB5	0xB5	HALFWIDTH KATAKANA LETTER O
0xB6	0xB6	HALFWIDTH KATAKANA LETTER KA
0xB7	0xB7	HALFWIDTH KATAKANA LETTER KI
0xB8	0xB8	HALFWIDTH KATAKANA LETTER KU
0xB9	0xB9	HALFWIDTH KATAKANA LETTER KE
0xBA	0xBA	HALFWIDTH KATAKANA LETTER KO
0xBB	0xBB	HALFWIDTH KATAKANA LETTER SA
0xBC	0xBC	HALFWIDTH KATAKANA LETTER SI
0xBD	0xBD	HALFWIDTH KATAKANA LETTER SU
0xBE	0xBE	HALFWIDTH KATAKANA LETTER SE
0xBF	0xBF	HALFWIDTH KATAKANA LETTER SO

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0xC0	0xC0	HALFWIDTH KATAKANA LETTER TA
0xC1	0xC1	HALFWIDTH KATAKANA LETTER TI
0xC2	0xC2	HALFWIDTH KATAKANA LETTER TU
0xC3	0xC3	HALFWIDTH KATAKANA LETTER TE
0xC4	0xC4	HALFWIDTH KATAKANA LETTER TO
0xC5	0xC5	HALFWIDTH KATAKANA LETTER NA
0xC6	0xC6	HALFWIDTH KATAKANA LETTER NI
0xC7	0xC7	HALFWIDTH KATAKANA LETTER NU
0xC8	0xC8	HALFWIDTH KATAKANA LETTER NE
0xC9	0xC9	HALFWIDTH KATAKANA LETTER NO
0xCA	0xCA	HALFWIDTH KATAKANA LETTER HA
0xCB	0xCB	HALFWIDTH KATAKANA LETTER HI
0xCC	0xCC	HALFWIDTH KATAKANA LETTER HU
0xCD	0xCD	HALFWIDTH KATAKANA LETTER HE
0xCE	0xCE	HALFWIDTH KATAKANA LETTER HO
0xCF	0xCF	HALFWIDTH KATAKANA LETTER MA
0xD0	0xD0	HALFWIDTH KATAKANA LETTER MI
0xD1	0xD1	HALFWIDTH KATAKANA LETTER MU
0xD2	0xD2	HALFWIDTH KATAKANA LETTER ME
0xD3	0xD3	HALFWIDTH KATAKANA LETTER MO
0xD4	0xD4	HALFWIDTH KATAKANA LETTER YA
0xD5	0xD5	HALFWIDTH KATAKANA LETTER YU
0xD6	0xD6	HALFWIDTH KATAKANA LETTER YO
0xD7	0xD7	HALFWIDTH KATAKANA LETTER RA
0xD8	0xD8	HALFWIDTH KATAKANA LETTER RI
0xD9	0xD9	HALFWIDTH KATAKANA LETTER RU
0xDA	0xDA	HALFWIDTH KATAKANA LETTER RE
0xDB	0xDB	HALFWIDTH KATAKANA LETTER RO

KANJISJIS_0S Single-Byte	KANJI1	Unicode Character Name
0xDC	0xDC	HALFWIDTH KATAKANA LETTER WA
0xDD	0xDD	HALFWIDTH KATAKANA LETTER N
0xDE	0xDE	HALFWIDTH KATAKANA VOICED SOUND MARK
0xDF	0xDF	HALFWIDTH KATAKANA SEMI-VOICED SOUND MARK

Client KANJISJIS_0S Single-Byte Lowercase to Server KANJI1 Uppercase

The following table shows the mapping of single-byte characters in the KANJISJIS_0S client character set to KANJI1 uppercase characters. An example of when Teradata Database uses this mapping is when the client character set is KANJISJIS_0S and single-byte character data is imported from the client to a character column that is CHARACTER SET KANJI1 UPPERCASE.

KANJISJIS_0S Single-Byte Lowercase		KANJI1 Uppercase	
Code	Character Name	Code	Character Name
0x61	LATIN SMALL LETTER A	0x41	LATIN CAPITAL LETTER A
0x62	LATIN SMALL LETTER B	0x42	LATIN CAPITAL LETTER B
0x63	LATIN SMALL LETTER C	0x43	LATIN CAPITAL LETTER C
0x64	LATIN SMALL LETTER D	0x44	LATIN CAPITAL LETTER D
0x65	LATIN SMALL LETTER E	0x45	LATIN CAPITAL LETTER E
0x66	LATIN SMALL LETTER F	0x46	LATIN CAPITAL LETTER F
0x67	LATIN SMALL LETTER G	0x47	LATIN CAPITAL LETTER G
0x68	LATIN SMALL LETTER H	0x48	LATIN CAPITAL LETTER H
0x69	LATIN SMALL LETTER I	0x49	LATIN CAPITAL LETTER I
0x6A	LATIN SMALL LETTER J	0x4A	LATIN CAPITAL LETTER J
0x6B	LATIN SMALL LETTER K	0x4B	LATIN CAPITAL LETTER K
0x6C	LATIN SMALL LETTER L	0x4C	LATIN CAPITAL LETTER L
0x6D	LATIN SMALL LETTER M	0x4D	LATIN CAPITAL LETTER M
0x6E	LATIN SMALL LETTER N	0x4E	LATIN CAPITAL LETTER N
0x6F	LATIN SMALL LETTER O	0x4F	LATIN CAPITAL LETTER O
0x70	LATIN SMALL LETTER P	0x50	LATIN CAPITAL LETTER P

KANJI SJIS_0S Single-Byte Lowercase		KANJI1 Uppercase	
Code	Character Name	Code	Character Name
0x71	LATIN SMALL LETTER Q	0x51	LATIN CAPITAL LETTER Q
0x72	LATIN SMALL LETTER R	0x52	LATIN CAPITAL LETTER R
0x73	LATIN SMALL LETTER S	0x53	LATIN CAPITAL LETTER S
0x74	LATIN SMALL LETTER T	0x54	LATIN CAPITAL LETTER T
0x75	LATIN SMALL LETTER U	0x55	LATIN CAPITAL LETTER U
0x76	LATIN SMALL LETTER V	0x56	LATIN CAPITAL LETTER V
0x77	LATIN SMALL LETTER W	0x57	LATIN CAPITAL LETTER W
0x78	LATIN SMALL LETTER X	0x58	LATIN CAPITAL LETTER X
0x79	LATIN SMALL LETTER Y	0x59	LATIN CAPITAL LETTER Y
0x7A	LATIN SMALL LETTER Z	0x5A	LATIN CAPITAL LETTER Z

KANJISJIS to KANJISJIS Single-Byte

Teradata Database uses the mapping in the following table for validation of internal KANJISJIS single-byte data (JIS-x0201).

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0x00	0x00	<control> NULL
0x01	0x01	<control> START OF HEADING
0x02	0x02	<control> START OF TEXT
0x03	0x03	<control> END OF TEXT
0x04	0x04	<control> END OF TRANSMISSION
0x05	0x05	<control> ENQUIRY
0x06	0x06	<control> ACKNOWLEDGE
0x07	0x07	<control> BELL
0x08	0x08	<control> BACKSPACE
0x09	0x09	<control> CHARACTER TABULATION (horizontal tabulation)
0x0A	0x0A	<control> LINE FEED
0x0B	0x0B	<control> LINE TABULATION (vertical tabulation)

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0x0C	0x0C	<control> FORM FEED
0x0D	0x0D	<control> CARRIAGE RETURN
0x0E	0x0E	<control> SHIFT OUT
0x0F	0x0F	<control> SHIFT IN
0x10	0x10	<control> DATA LINK ESCAPE
0x11	0x11	<control> DEVICE CONTROL ONE
0x12	0x12	<control> DEVICE CONTROL TWO
0x13	0x13	<control> DEVICE CONTROL THREE
0x14	0x14	<control> DEVICE CONTROL FOUR
0x15	0x15	<control> NEGATIVE ACKNOWLEDGE
0x16	0x16	<control> SYNCHRONOUS IDLE
0x17	0x17	<control> END OF TRANSMISSION BLOCK
0x18	0x18	<control> CANCEL
0x19	0x19	<control> END OF MEDIUM
0x1A	0x1A	<control> SUBSTITUTE
0x1B	0x1B	<control> ESCAPE
0x1C	0x1C	<control> INFORMATION SEPARATOR FOUR (file separator)
0x1D	0x1D	<control> INFORMATION SEPARATOR THREE (group separator)
0x1E	0x1E	<control> INFORMATION SEPARATOR TWO (record separator)
0x1F	0x1F	<control> INFORMATION SEPARATOR ONE (unit separator)
0x20	0x20	SPACE
0x21	0x21	EXCLAMATION MARK
0x22	0x22	QUOTATION MARK
0x23	0x23	NUMBER SIGN
0x24	0x24	DOLLAR SIGN
0x25	0x25	PERCENT SIGN

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0x26	0x26	AMPERSAND
0x27	0x27	APOSTROPHE
0x28	0x28	LEFT PARENTHESIS
0x29	0x29	RIGHT PARENTHESIS
0x2A	0x2A	ASTERISK
0x2B	0x2B	PLUS SIGN
0x2C	0x2C	COMMA
0x2D	0x2D	HYPHEN-MINUS
0x2E	0x2E	FULL STOP
0x2F	0x2F	SOLIDUS
0x30	0x30	DIGIT ZERO
0x31	0x31	DIGIT ONE
0x32	0x32	DIGIT TWO
0x33	0x33	DIGIT THREE
0x34	0x34	DIGIT FOUR
0x35	0x35	DIGIT FIVE
0x36	0x36	DIGIT SIX
0x37	0x37	DIGIT SEVEN
0x38	0x38	DIGIT EIGHT
0x39	0x39	DIGIT NINE
0x3A	0x3A	COLON
0x3B	0x3B	SEMICOLON
0x3C	0x3C	LESS-THAN SIGN
0x3D	0x3D	EQUALS SIGN
0x3E	0x3E	GREATER-THAN SIGN
0x3F	0x3F	QUESTION MARK
0x40	0x40	COMMERCIAL AT
0x41	0x41	LATIN CAPITAL LETTER A

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0x42	0x42	LATIN CAPITAL LETTER B
0x43	0x43	LATIN CAPITAL LETTER C
0x44	0x44	LATIN CAPITAL LETTER D
0x45	0x45	LATIN CAPITAL LETTER E
0x46	0x46	LATIN CAPITAL LETTER F
0x47	0x47	LATIN CAPITAL LETTER G
0x48	0x48	LATIN CAPITAL LETTER H
0x49	0x49	LATIN CAPITAL LETTER I
0x4A	0x4A	LATIN CAPITAL LETTER J
0x4B	0x4B	LATIN CAPITAL LETTER K
0x4C	0x4C	LATIN CAPITAL LETTER L
0x4D	0x4D	LATIN CAPITAL LETTER M
0x4E	0x4E	LATIN CAPITAL LETTER N
0x4F	0x4F	LATIN CAPITAL LETTER O
0x50	0x50	LATIN CAPITAL LETTER P
0x51	0x51	LATIN CAPITAL LETTER Q
0x52	0x52	LATIN CAPITAL LETTER R
0x53	0x53	LATIN CAPITAL LETTER S
0x54	0x54	LATIN CAPITAL LETTER T
0x55	0x55	LATIN CAPITAL LETTER U
0x56	0x56	LATIN CAPITAL LETTER V
0x57	0x57	LATIN CAPITAL LETTER W
0x58	0x58	LATIN CAPITAL LETTER X
0x59	0x59	LATIN CAPITAL LETTER Y
0x5A	0x5A	LATIN CAPITAL LETTER Z
0x5B	0x5B	LEFT SQUARE BRACKET
0x5C	0x5C	YEN SIGN
0x5D	0x5D	RIGHT SQUARE BRACKET

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0x5E	0x5E	CIRCUMFLEX ACCENT
0x5F	0x5F	LOW LINE
0x60	0x60	GRAVE ACCENT
0x61	0x61	LATIN SMALL LETTER A
0x62	0x62	LATIN SMALL LETTER B
0x63	0x63	LATIN SMALL LETTER C
0x64	0x64	LATIN SMALL LETTER D
0x65	0x65	LATIN SMALL LETTER E
0x66	0x66	LATIN SMALL LETTER F
0x67	0x67	LATIN SMALL LETTER G
0x68	0x68	LATIN SMALL LETTER H
0x69	0x69	LATIN SMALL LETTER I
0x6A	0x6A	LATIN SMALL LETTER J
0x6B	0x6B	LATIN SMALL LETTER K
0x6C	0x6C	LATIN SMALL LETTER L
0x6D	0x6D	LATIN SMALL LETTER M
0x6E	0x6E	LATIN SMALL LETTER N
0x6F	0x6F	LATIN SMALL LETTER O
0x70	0x70	LATIN SMALL LETTER P
0x71	0x71	LATIN SMALL LETTER Q
0x72	0x72	LATIN SMALL LETTER R
0x73	0x73	LATIN SMALL LETTER S
0x74	0x74	LATIN SMALL LETTER T
0x75	0x75	LATIN SMALL LETTER U
0x76	0x76	LATIN SMALL LETTER V
0x77	0x77	LATIN SMALL LETTER W
0x78	0x78	LATIN SMALL LETTER X
0x79	0x79	LATIN SMALL LETTER Y

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0x7A	0x7A	LATIN SMALL LETTER Z
0x7B	0x7B	LEFT CURLY BRACKET
0x7C	0x7C	VERTICAL LINE
0x7D	0x7D	RIGHT CURLY BRACKET
0x7E	0x7E	OVERLINE
0xA1	0xA1	HALFWIDTH IDEOGRAPHIC FULL STOP
0xA2	0xA2	HALFWIDTH LEFT CORNER BRACKET
0xA3	0xA3	HALFWIDTH RIGHT CORNER BRACKET
0xA4	0xA4	HALFWIDTH IDEOGRAPHIC COMMA
0xA5	0xA5	HALFWIDTH KATAKANA MIDDLE DOT
0xA6	0xA6	HALFWIDTH KATAKANA LETTER WO
0xA7	0xA7	HALFWIDTH KATAKANA LETTER SMALL A
0xA8	0xA8	HALFWIDTH KATAKANA LETTER SMALL I
0xA9	0xA9	HALFWIDTH KATAKANA LETTER SMALL U
0xAA	0xAA	HALFWIDTH KATAKANA LETTER SMALL E
0xAB	0xAB	HALFWIDTH KATAKANA LETTER SMALL O
0xAC	0xAC	HALFWIDTH KATAKANA LETTER SMALL YA
0xAD	0xAD	HALFWIDTH KATAKANA LETTER SMALL YU
0xAE	0xAE	HALFWIDTH KATAKANA LETTER SMALL YO
0xAF	0xAF	HALFWIDTH KATAKANA LETTER SMALL TU
0xB0	0xB0	HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK
0xB1	0xB1	HALFWIDTH KATAKANA LETTER A
0xB2	0xB2	HALFWIDTH KATAKANA LETTER I
0xB3	0xB3	HALFWIDTH KATAKANA LETTER U
0xB4	0xB4	HALFWIDTH KATAKANA LETTER E
0xB5	0xB5	HALFWIDTH KATAKANA LETTER O
0xB6	0xB6	HALFWIDTH KATAKANA LETTER KA
0xB7	0xB7	HALFWIDTH KATAKANA LETTER KI

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0xB8	0xB8	HALFWIDTH KATAKANA LETTER KU
0xB9	0xB9	HALFWIDTH KATAKANA LETTER KE
0xBA	0xBA	HALFWIDTH KATAKANA LETTER KO
0xBB	0xBB	HALFWIDTH KATAKANA LETTER SA
0xBC	0xBC	HALFWIDTH KATAKANA LETTER SI
0xBD	0xBD	HALFWIDTH KATAKANA LETTER SU
0xBE	0xBE	HALFWIDTH KATAKANA LETTER SE
0xBF	0xBF	HALFWIDTH KATAKANA LETTER SO
0xC0	0xC0	HALFWIDTH KATAKANA LETTER TA
0xC1	0xC1	HALFWIDTH KATAKANA LETTER TI
0xC2	0xC2	HALFWIDTH KATAKANA LETTER TU
0xC3	0xC3	HALFWIDTH KATAKANA LETTER TE
0xC4	0xC4	HALFWIDTH KATAKANA LETTER TO
0xC5	0xC5	HALFWIDTH KATAKANA LETTER NA
0xC6	0xC6	HALFWIDTH KATAKANA LETTER NI
0xC7	0xC7	HALFWIDTH KATAKANA LETTER NU
0xC8	0xC8	HALFWIDTH KATAKANA LETTER NE
0xC9	0xC9	HALFWIDTH KATAKANA LETTER NO
0xCA	0xCA	HALFWIDTH KATAKANA LETTER HA
0xCB	0xCB	HALFWIDTH KATAKANA LETTER HI
0xCC	0xCC	HALFWIDTH KATAKANA LETTER HU
0xCD	0xCD	HALFWIDTH KATAKANA LETTER HE
0xCE	0xCE	HALFWIDTH KATAKANA LETTER HO
0xCF	0xCF	HALFWIDTH KATAKANA LETTER MA
0xD0	0xD0	HALFWIDTH KATAKANA LETTER MI
0xD1	0xD1	HALFWIDTH KATAKANA LETTER MU
0xD2	0xD2	HALFWIDTH KATAKANA LETTER ME
0xD3	0xD3	HALFWIDTH KATAKANA LETTER MO

KanjiShiftJIS	KanjiShiftJIS Single-Byte	Unicode Character Name
0xD4	0xD4	HALFWIDTH KATAKANA LETTER YA
0xD5	0xD5	HALFWIDTH KATAKANA LETTER YU
0xD6	0xD6	HALFWIDTH KATAKANA LETTER YO
0xD7	0xD7	HALFWIDTH KATAKANA LETTER RA
0xD8	0xD8	HALFWIDTH KATAKANA LETTER RI
0xD9	0xD9	HALFWIDTH KATAKANA LETTER RU
0xDA	0xDA	HALFWIDTH KATAKANA LETTER RE
0xDB	0xDB	HALFWIDTH KATAKANA LETTER RO
0xDC	0xDC	HALFWIDTH KATAKANA LETTER WA
0xDD	0xDD	HALFWIDTH KATAKANA LETTER N
0xDE	0xDE	HALFWIDTH KATAKANA VOICED SOUND MARK
0xDF	0xDF	HALFWIDTH KATAKANA SEMI-VOICED SOUND MARK

UNICODE to KanjiShiftJIS Single-Byte

Teradata Database uses the mapping in the following table for the following conversions:

- Internal to internal conversion of the UNICODE server character set to the single-byte portion of the KANJISJIS server character set.
- Internal to external conversion of the UNICODE server character set to the single-byte portion of the KANJISJIS_OS client character set

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+0000	0x00	<control> NULL
U+0001	0x01	<control> START OF HEADING
U+0002	0x02	<control> START OF TEXT
U+0003	0x03	<control> END OF TEXT
U+0004	0x04	<control> END OF TRANSMISSION
U+0005	0x05	<control> ENQUIRY
U+0006	0x06	<control> ACKNOWLEDGE
U+0007	0x07	<control> BELL

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+0008	0x08	<control> BACKSPACE
U+0009	0x09	<control> CHARACTER TABULATION (horizontal tabulation)
U+000A	0x0A	<control> LINE FEED
U+000B	0x0B	<control> LINE TABULATION (vertical tabulation)
U+000C	0x0C	<control> FORM FEED
U+000D	0x0D	<control> CARRIAGE RETURN
U+000E	0x0E	<control> SHIFT OUT
U+000F	0x0F	<control> SHIFT IN
U+0010	0x10	<control> DATA LINK ESCAPE
U+0011	0x11	<control> DEVICE CONTROL ONE
U+0012	0x12	<control> DEVICE CONTROL TWO
U+0013	0x13	<control> DEVICE CONTROL THREE
U+0014	0x14	<control> DEVICE CONTROL FOUR
U+0015	0x15	<control> NEGATIVE ACKNOWLEDGE
U+0016	0x16	<control> SYNCHRONOUS IDLE
U+0017	0x17	<control> END OF TRANSMISSION BLOCK
U+0018	0x18	<control> CANCEL
U+0019	0x19	<control> END OF MEDIUM
U+001A	0x1A	<control> SUBSTITUTE
U+001B	0x1B	<control> ESCAPE
U+001C	0x1C	<control> INFORMATION SEPARATOR FOUR (file separator)
U+001D	0x1D	<control> INFORMATION SEPARATOR THREE (group separator)
U+001E	0x1E	<control> INFORMATION SEPARATOR TWO (record separator)
U+001F	0x1F	<control> INFORMATION SEPARATOR ONE (unit separator)
U+0020	0x20	SPACE
U+0021	0x21	EXCLAMATION MARK
U+0022	0x22	QUOTATION MARK

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+0023	0x23	NUMBER SIGN
U+0024	0x24	DOLLAR SIGN
U+0025	0x25	PERCENT SIGN
U+0026	0x26	AMPERSAND
U+0027	0x27	APOSTROPHE
U+0028	0x28	LEFT PARENTHESIS
U+0029	0x29	RIGHT PARENTHESIS
U+002A	0x2A	ASTERISK
U+002B	0x2B	PLUS SIGN
U+002C	0x2C	COMMA
U+002D	0x2D	HYPHEN-MINUS
U+002E	0x2E	FULL STOP
U+002F	0x2F	SOLIDUS
U+0030	0x30	DIGIT ZERO
U+0031	0x31	DIGIT ONE
U+0032	0x32	DIGIT TWO
U+0033	0x33	DIGIT THREE
U+0034	0x34	DIGIT FOUR
U+0035	0x35	DIGIT FIVE
U+0036	0x36	DIGIT SIX
U+0037	0x37	DIGIT SEVEN
U+0038	0x38	DIGIT EIGHT
U+0039	0x39	DIGIT NINE
U+003A	0x3A	COLON
U+003B	0x3B	SEMICOLON
U+003C	0x3C	LESS-THAN SIGN
U+003D	0x3D	EQUALS SIGN
U+003E	0x3E	GREATER-THAN SIGN

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+003F	0x3F	QUESTION MARK
U+0040	0x40	COMMERCIAL AT
U+0041	0x41	LATIN CAPITAL LETTER A
U+0042	0x42	LATIN CAPITAL LETTER B
U+0043	0x43	LATIN CAPITAL LETTER C
U+0044	0x44	LATIN CAPITAL LETTER D
U+0045	0x45	LATIN CAPITAL LETTER E
U+0046	0x46	LATIN CAPITAL LETTER F
U+0047	0x47	LATIN CAPITAL LETTER G
U+0048	0x48	LATIN CAPITAL LETTER H
U+0049	0x49	LATIN CAPITAL LETTER I
U+004A	0x4A	LATIN CAPITAL LETTER J
U+004B	0x4B	LATIN CAPITAL LETTER K
U+004C	0x4C	LATIN CAPITAL LETTER L
U+004D	0x4D	LATIN CAPITAL LETTER M
U+004E	0x4E	LATIN CAPITAL LETTER N
U+004F	0x4F	LATIN CAPITAL LETTER O
U+0050	0x50	LATIN CAPITAL LETTER P
U+0051	0x51	LATIN CAPITAL LETTER Q
U+0052	0x52	LATIN CAPITAL LETTER R
U+0053	0x53	LATIN CAPITAL LETTER S
U+0054	0x54	LATIN CAPITAL LETTER T
U+0055	0x55	LATIN CAPITAL LETTER U
U+0056	0x56	LATIN CAPITAL LETTER V
U+0057	0x57	LATIN CAPITAL LETTER W
U+0058	0x58	LATIN CAPITAL LETTER X
U+0059	0x59	LATIN CAPITAL LETTER Y
U+005A	0x5A	LATIN CAPITAL LETTER Z

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+005B	0x5B	LEFT SQUARE BRACKET
U+005D	0x5D	RIGHT SQUARE BRACKET
U+005E	0x5E	CIRCUMFLEX ACCENT
U+005F	0x5F	LOW LINE
U+0060	0x60	GRAVE ACCENT
U+0061	0x61	LATIN SMALL LETTER A
U+0062	0x62	LATIN SMALL LETTER B
U+0063	0x63	LATIN SMALL LETTER C
U+0064	0x64	LATIN SMALL LETTER D
U+0065	0x65	LATIN SMALL LETTER E
U+0066	0x66	LATIN SMALL LETTER F
U+0067	0x67	LATIN SMALL LETTER G
U+0068	0x68	LATIN SMALL LETTER H
U+0069	0x69	LATIN SMALL LETTER I
U+006A	0x6A	LATIN SMALL LETTER J
U+006B	0x6B	LATIN SMALL LETTER K
U+006C	0x6C	LATIN SMALL LETTER L
U+006D	0x6D	LATIN SMALL LETTER M
U+006E	0x6E	LATIN SMALL LETTER N
U+006F	0x6F	LATIN SMALL LETTER O
U+0070	0x70	LATIN SMALL LETTER P
U+0071	0x71	LATIN SMALL LETTER Q
U+0072	0x72	LATIN SMALL LETTER R
U+0073	0x73	LATIN SMALL LETTER S
U+0074	0x74	LATIN SMALL LETTER T
U+0075	0x75	LATIN SMALL LETTER U
U+0076	0x76	LATIN SMALL LETTER V
U+0077	0x77	LATIN SMALL LETTER W

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+0078	0x78	LATIN SMALL LETTER X
U+0079	0x79	LATIN SMALL LETTER Y
U+007A	0x7A	LATIN SMALL LETTER Z
U+007B	0x7B	LEFT CURLY BRACKET
U+007C	0x7C	VERTICAL LINE
U+007D	0x7D	RIGHT CURLY BRACKET
U+00A5	0x5C	YEN SIGN
U+203E	0x7E	OVERLINE
U+FF61	0xA1	HALFWIDTH IDEOGRAPHIC FULL STOP
U+FF62	0xA2	HALFWIDTH LEFT CORNER BRACKET
U+FF63	0xA3	HALFWIDTH RIGHT CORNER BRACKET
U+FF64	0xA4	HALFWIDTH IDEOGRAPHIC COMMA
U+FF65	0xA5	HALFWIDTH KATAKANA MIDDLE DOT
U+FF66	0xA6	HALFWIDTH KATAKANA LETTER WO
U+FF67	0xA7	HALFWIDTH KATAKANA LETTER SMALL A
U+FF68	0xA8	HALFWIDTH KATAKANA LETTER SMALL I
U+FF69	0xA9	HALFWIDTH KATAKANA LETTER SMALL U
U+FF6A	0xAA	HALFWIDTH KATAKANA LETTER SMALL E
U+FF6B	0xAB	HALFWIDTH KATAKANA LETTER SMALL O
U+FF6C	0xAC	HALFWIDTH KATAKANA LETTER SMALL YA
U+FF6D	0xAD	HALFWIDTH KATAKANA LETTER SMALL YU
U+FF6E	0xAE	HALFWIDTH KATAKANA LETTER SMALL YO
U+FF6F	0xAF	HALFWIDTH KATAKANA LETTER SMALL TU
U+FF70	0xB0	HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK
U+FF71	0xB1	HALFWIDTH KATAKANA LETTER A
U+FF72	0xB2	HALFWIDTH KATAKANA LETTER I
U+FF73	0xB3	HALFWIDTH KATAKANA LETTER U
U+FF74	0xB4	HALFWIDTH KATAKANA LETTER E

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+FF75	0xB5	HALFWIDTH KATAKANA LETTER O
U+FF76	0xB6	HALFWIDTH KATAKANA LETTER KA
U+FF77	0xB7	HALFWIDTH KATAKANA LETTER KI
U+FF78	0xB8	HALFWIDTH KATAKANA LETTER KU
U+FF79	0xB9	HALFWIDTH KATAKANA LETTER KE
U+FF7A	0xBA	HALFWIDTH KATAKANA LETTER KO
U+FF7B	0xBB	HALFWIDTH KATAKANA LETTER SA
U+FF7C	0xBC	HALFWIDTH KATAKANA LETTER SI
U+FF7D	0xBD	HALFWIDTH KATAKANA LETTER SU
U+FF7E	0xBE	HALFWIDTH KATAKANA LETTER SE
U+FF7F	0xBF	HALFWIDTH KATAKANA LETTER SO
U+FF80	0xC0	HALFWIDTH KATAKANA LETTER TA
U+FF81	0xC1	HALFWIDTH KATAKANA LETTER TI
U+FF82	0xC2	HALFWIDTH KATAKANA LETTER TU
U+FF83	0xC3	HALFWIDTH KATAKANA LETTER TE
U+FF84	0xC4	HALFWIDTH KATAKANA LETTER TO
U+FF85	0xC5	HALFWIDTH KATAKANA LETTER NA
U+FF86	0xC6	HALFWIDTH KATAKANA LETTER NI
U+FF87	0xC7	HALFWIDTH KATAKANA LETTER NU
U+FF88	0xC8	HALFWIDTH KATAKANA LETTER NE
U+FF89	0xC9	HALFWIDTH KATAKANA LETTER NO
U+FF8A	0xCA	HALFWIDTH KATAKANA LETTER HA
U+FF8B	0xCB	HALFWIDTH KATAKANA LETTER HI
U+FF8C	0xCC	HALFWIDTH KATAKANA LETTER HU
U+FF8D	0xCD	HALFWIDTH KATAKANA LETTER HE
U+FF8E	0xCE	HALFWIDTH KATAKANA LETTER HO
U+FF8F	0xCF	HALFWIDTH KATAKANA LETTER MA
U+FF90	0xD0	HALFWIDTH KATAKANA LETTER MI

UNICODE	KanjiShiftJIS Single-Byte	Unicode Character Name
U+FF91	0xD1	HALFWIDTH KATAKANA LETTER MU
U+FF92	0xD2	HALFWIDTH KATAKANA LETTER ME
U+FF93	0xD3	HALFWIDTH KATAKANA LETTER MO
U+FF94	0xD4	HALFWIDTH KATAKANA LETTER YA
U+FF95	0xD5	HALFWIDTH KATAKANA LETTER YU
U+FF96	0xD6	HALFWIDTH KATAKANA LETTER YO
U+FF97	0xD7	HALFWIDTH KATAKANA LETTER RA
U+FF98	0xD8	HALFWIDTH KATAKANA LETTER RI
U+FF99	0xD9	HALFWIDTH KATAKANA LETTER RU
U+FF9A	0xDA	HALFWIDTH KATAKANA LETTER RE
U+FF9B	0xDB	HALFWIDTH KATAKANA LETTER RO
U+FF9C	0xDC	HALFWIDTH KATAKANA LETTER WA
U+FF9D	0xDD	HALFWIDTH KATAKANA LETTER N
U+FF9E	0xDE	HALFWIDTH KATAKANA VOICED SOUND MARK
U+FF9F	0xDF	HALFWIDTH KATAKANA SEMI-VOICED SOUND MARK
U+FFFD	0x1A	REPLACEMENT CHARACTER & <control>

Character Set Files

The text files listed in the tables provide the following information:

- Valid characters that you can use in object names
- Mappings between character sets
- Character set collations

Download these files as a zip file:

1. Access *Teradata Vantage™ - Advanced SQL Engine International Character Set Support*, B035-1125 at <https://docs.teradata.com/>.
2. In the left pane, download the zip file from the download tab ↓.

Valid Characters in Object Names

Title	File Name	Publication ID
Unicode Characters Allowed in Object Names on Systems with Extended Object Naming	UOBJNEXT.txt	B035-1200
Unicode in Object Names on Japanese Language Support Systems	UOBJNJAP.txt	B035-1177
Unicode in Object Names on Standard Language Support Systems	UOBJNSTD.txt	B035-1176

Character Set Mappings and Collations

Title	Filename	Description	Publication ID
ARABIC1256_6A0 to Unicode	A6A0SUCD.txt	ARABIC1256 to Unicode	B035-1165
CYRILLIC1251_2A0 to Unicode	C2A0SUCD.txt	CYRILLIC1251 to Unicode	B035-1166
HANGUL949_7R0 Multibyte to Unicode	H7R0MUCD.txt	HANGUL949 (multibyte character portion) to Unicode	B035-1170
HANGUL949_7R0 Single Byte to Unicode	H7R0SUCD.txt	HANGUL949 (single-byte character portion) to Unicode	B035-1169
HANGULEBCDIC933_1II Multibyte to Unicode	H1IMUNCD.txt	Hangul EBCDIC (IBM CCSID 933) multibyte (IBM CP 834) to Unicode	B035-1135
HANGULEBCDIC933_1II Single Byte to Unicode	H1ISUNCD.txt	Hangul EBCDIC (IBM CCSID 933) single-byte (IBM CP 833) to Unicode	B035-1134
HANGULKSC5601_2R4 Multibyte to Unicode	H2RMUNCD.txt	Hangul (mixed KS Roman/KS C 5601) multibyte (KS C 5601) to Unicode	B035-1137

Title	Filename	Description	Publication ID
HANGULKSC5601_2R4 Single Byte to Unicode	H2RSUNCD.txt	Hangul (mixed KS Roman/KS C 5601) single-byte (KS Roman) to Unicode	B035-1136
HEBREW1255_5A0 to Unicode	H5A0SUCD.txt	HEBREW1255 to Unicode	B035-1164
JIS_COLL Case-Blind Collation	JISCOLBL.txt	JIS_COLL Case-Blind collation	B035-1061
JIS_COLL Case-Specific Collation	JIS_COLL.txt	JIS_COLL Case-Specific collation	B035-1060
KANJI932_1S0 Multibyte to Unicode	K1S0MUCD.txt	KANJI932 (multibyte character portion) to Unicode	B035-1175
KANJI932_1S0 Single Byte to Unicode	K1S0SUCD.txt	KANJI932 (single-byte character portion) to Unicode	B035-1174
KanjiEBCDIC Multibyte (SO/SI) to Unicode	SOSIUNCD.txt	KanjiEBCDIC Multibyte (Shift-Out /Shift-In) to Unicode	B035-1055
KanjiEUC Code Set 1 to Unicode	EUC1UNCD.txt	KanjiEUC Code Set 1 (JIS-x0208) to Unicode	B035-1115
KanjiEUC Code Set 2 to Unicode	EUC2UNCD.txt	KanjiEUC Code Set 2 (JIS-x0201 Katakana) to Unicode	B035-1139
KanjiEUC Code Set 3 to Unicode	EUC3UNCD.txt	KanjiEUC Code Set 3 (JIS-x0212) to Unicode	B035-1116
KanjiShiftJIS to KanjiShiftJIS Multibyte	SJISSJIS.txt	KanjiSJIS to KanjiSJIS multibyte characters	B035-1053
KanjiShiftJIS to Unicode Multibyte	SJISUNCD.txt	KanjiSJIS to multibyte Unicode	B035-1054
LATIN1250_1A0 to Unicode	L1A0SUCD.txt	LATIN1250 to Unicode	B035-1168
LATIN1252_3A0 to Unicode	L3A0SUCD.txt	LATIN1252 to Unicode	B035-1163
LATIN1254 to Unicode	L7A0SUCD.txt	LATIN1254 to Unicode	B035-1171
LATIN1258_8A0 to Unicode	L8A0SUCD.txt	LATIN1258 to Unicode	B035-1173
LATIN Server Character Set	latin_server.txt	Supported characters in the NewSQL Engine LATIN server character set	B035-1207
Multinational Case-Blind Default Collation	blinddef.txt	Default for Multinational Case-Blind collation	B035-1050
Multinational Case-Specific Default Collation	multnatl.txt	Default for Multinational Case-Specific collation	B035-1062
SCHEBCDIC935_2IJ Multibyte to Unicode	C2IMUNCD.txt	Simplified Chinese EBCDIC (IBM CCSID 935) multibyte (IBM CP 837) to Unicode	B035-1131

Title	Filename	Description	Publication ID
SCHEBCDIC935_2IJ Single Byte to Unicode	C2ISUNCD.txt	Simplified Chinese EBCDIC (IBM CCSID 935) single-byte (IBM CP 836) to Unicode	B035-1130
SCHGB2312_1T0 Code Set 0 to Unicode	C1T0UNCD.txt	Simplified Chinese (mixed ASCII /GB 2312-1980) Code Set 0 (ASCII) to Unicode	B035-1126
SCHGB2312_1T0 Code Set 1 to Unicode	C1T1UNCD.txt	Simplified Chinese (mixed ASCII /GB 2312-1980) Code Set 1 (GB 2312-1980) to Unicode	B035-1127
SCHINESE936_6R0 Multibyte to Unicode	S6R0MUCD.txt	SCHINESE936 (multibyte character portion) to Unicode	B035-1162
SCHINESE936_6R0 Single Byte to Unicode	S6R0SUCD.txt	SCHINESE936 (single-byte character portion) to Unicode	B035-1161
TCHBIG5_1R0 Multibyte to Unicode	C1RMUNCD.txt	Traditional Chinese (Big5) multibyte to Unicode	B035-1129
TCHBIG5_1R0 Single Byte to Unicode	C1RSUNCD.txt	Traditional Chinese (Big5) single-byte (ASCII) to Unicode	B035-1128
TCHEBCDIC937_3IB Multibyte to Unicode	C3IMUNCD.txt	Traditional Chinese EBCDIC (IBM CCSID 937) multibyte (IBM CP 835) to Unicode	B035-1133
TCHEBCDIC937_3IB Single Byte to Unicode	C3ISUNCD.txt	Traditional Chinese EBCDIC (IBM CCSID 937) single-byte (IBM CP 037) to Unicode	B035-1132
TCHINESE950_8R0 Multibyte to Unicode	T8R0MUCD.txt	TCHINESE950 (multibyte character portion) to Unicode	B035-1178
TCHINESE950_8R0 Single Byte to Unicode	T8R0SUCD.txt	TCHINESE950 (single-byte character portion) to Unicode	B035-1172
THAI874_4A0 Single Byte to Unicode	T4A0SUCD.txt	THAI874 to Unicode	B035-1167
UNICODE Server Character Set	UNCDUNCD.txt	Supported characters in the NewSQL Engine UNICODE server character set	B035-1056
Unicode to KanjiEBCDIC Multibyte (SO/SI)	UNCDSOSI.txt	Unicode to KanjiEBCDIC multibyte (Shift-Out/Shift-In)	B035-1104
Unicode to KanjiEUC Code Sets 1, 2, and 3	UNCDE123.txt	Unicode to KanjiEUC Code Set 1, 2,3 (JIS-x0208) as Unix Process Code (UPC)	B035-1117
Unicode to KanjiSJIS Multibyte	UNCDSJIS.txt	Unicode to KanjiSJIS multibyte	B035-1058

Title	Filename	Description	Publication ID
UNICODE to UNICODE_Fullwidth	UNCDH2F.txt	Halfwidth Unicode to fullwidth Unicode	B035-1202
UNICODE to UNICODE_Halfwidth	UNCDF2H.txt	Fullwidth Unicode to halfwidth Unicode	B035-1201
UNICODE to VARGRAPHIC	UNCDVARG.txt	<ul style="list-style-type: none"> • Halfwidth letters of Unicode to the fullwidth letters of Unicode • SPACE (0x0020) to the IDEOGRAPHIC SPACE (0x3000) • Valid characters of graphic 	B035-1057

Additional Information

Teradata Links

Link	Description
https://docs.teradata.com/	Search Teradata Documentation, customize content to your needs, and download PDFs. Customers: Log in to access Orange Books.
https://support.teradata.com	One-stop source for Teradata community support, software downloads, and product information. Log in for customer access to: <ul style="list-style-type: none">• Community support• Software updates• Knowledge articles
https://www.teradata.com/University/Overview	Teradata education network
https://support.teradata.com/community	Link to Teradata community