IBM InfoSphere DataStage and QualityStage
Version 11 Release 3

Globalization Guide

IBM

IBM InfoSphere DataStage and QualityStage
Version 11 Release 3

Globalization Guide

IBM

Contents

Chapter 1. National Language Support

When you install InfoSphere® DataStage® With National Language Support (NLS) mode enabled, you can use InfoSphere DataStage in various languages and countries.

With NLS enabled, you can do the following:

- Use InfoSphere DataStage in various languages. This includes languages that use multi-byte characters, such as Japanese.
- Read and write data in multi-byte character sets and process the data within InfoSphere DataStage. This is regardless of the language of InfoSphere DataStage itself. For example, you can process Japanese data in an English version of InfoSphere DataStage, or process English data in a Japanese version of InfoSphere DataStage.
- Use locales to change things like collating sequence, monetary conventions, date/time format from outside a job design.

You must enable NLS when you install InfoSphere DataStage. If you choose to install a non-English language version of InfoSphere DataStage, NLS is enabled automatically. If you choose to install an English version of InfoSphere DataStage, you specify separately whether NLS is enabled or not.

Chapter 2. Server Jobs and NLS

You can create new maps and new locales for use in InfoSphere DataStage server jobs.

Creating New Maps

If the maps supplied with InfoSphere DataStage do not meet your needs, you can create new ones and use these in your jobs.

You are most likely to want to produce a variant of an existing map rather than add an entirely new one, InfoSphere DataStage allows you to base a new map on an existing one and just add or alter the required mappings. You do this by creating a table and adding it to a map to make a new map.

A map is defined by a Description, which in turn calls upon a Table to define the actual mappings. To create a new map, you need to define a Description and a Table.

When you want to produce a variant of an existing map it is important that you create a new map based on the existing one. Under no circumstances should you edit one of the maps supplied with InfoSphere DataStage.

Maps are created using the NLS administration tool. This is run in an engine shell. You need to have InfoSphere DataStage Administrator status in order to be able to run this.

Running NLS Administration Tool on a Windows engine tier host

You must follow a specific procedure to run the NLS Administration Tool on a Windows engine tier host.

Procedure

1. Start a telnet session and connect to your engine. The "Welcome to DataStage Telnet Server message" appears and you are prompted for a log in name and password.
2. Enter your InfoSphere DataStage user name and password. You are then prompted for an account name or path.
3. Enter `uv` as the account name. You are now connected to the engine.
4. At the prompt type `NLS.ADMIN` (note that case is important). The NLS Administration window appears.

Running NLS Administration Tool on a UNIX engine tier host

You must follow a specific procedure to run the NLS Administration Tool on a UNIX engine tier host.

Procedure

1. Start a telnet session and connect to your engine.
2. CD to the engine directory (`$DSHOME/DSEngine`).
3. Type `bin/uvsh` .

4. At the prompt type `NLS.ADMIN` (note that case is important). The NLS Administration window appears.

Base Maps

A map can be based on another map and this map can be based on yet another map. To understand the complete map you must follow the chain of base maps.

For more information about the construction of a map, choose **Mappings** > **Descriptions** > **Xref** and **Mappings** > **Tables** > **Xref** from the NLS Administration menu. Choose the map or table whose lineage you want to see.

For example, the map C0-CONTROLS is a single-byte character set map using the C0-CONTROLS table. It maps the set of 7-bit control characters. The description report will tell you that just about every other map has C0-CONTROLS in its lineage, while it is the base map for C1-CONTROLS and ASCII.

Creating a New Map

If the maps supplied with InfoSphere DataStage do not meet your needs, you can create new ones and use these in your jobs.

Procedure

1. Find an existing map that most closely matches the required map.
2. Identify the characters that need to be mapped differently in the new map.
3. Create a new table contains only these new mappings.
4. Create the new map by adding a new description based on the existing map but adding the new table.

Results

The following example creates a map called MY.ASCII. This map is identical to the existing ASCII map, except the input character 0x23 is mapped to the UK pound sign (£) instead of the number symbol (#).

Your first action is to create a table called MY.POUND that performs this mapping:

Example

1. In the NLS administration tool, choose **Mappings** > **Tables** > **Create** .
2. Specify MY.POUND as the table name:
   ```
   DATASTAGE NLS CHARACTER MAP/LOCALE MAINTENANCE PROGRAM

   Enter map table record ID (create), Q to quit: MY.POUND
   ```
3. The NLS Administrator editor opens, enter I to insert new lines and add lines 1 and 2 as shown below. At line 3, just press return to exit insert mode.
   ```
   DATASTAGE NLS CHARACTER MAP/LOCALE MAINTENANCE PROGRAM

   EDITING FILE: NLS.MAP.TABLES / MY.POUND

   New record.

    ----: I
   0001= * Map 0x23 to Unicode 00A3
   0002= 23 00A3
   0003=
   Bottom at line 2.
   ----: FILE
   ```
4. Type FILE to write the file and leave the table editor.

Next you need to create a description.

1. In the NLS administration tool, choose **Mappings** > **Descriptions** > **Create** .
2. Specify MY.ASCII as the description name:

   ```
   DATASTAGE NLS CHARACTER MAP/LOCALE MAINTENANCE PROGRAM

    Enter map table record ID (create), Q to quit: MY.POUND
   ```

3. The NLS Administration tool asks you if you want to base the new description on an existing one. As you only require a short description, it is easier just to enter it directly, so type Q.
4. As the administration tool prompts for each field, enter the information as shown:

   ```
   NLS CHARACTER MAP DESCRIPTION REVISE.1 Mon May 19 10:22:20 2003

   Map ID=MY.ASCII
   New Record
   Map Description= * Modified ASCII with UK pound
   Based Map ID = ASCII
   Map type = SBCS
   Table ID- MY.POUND
   Display length-
   Prefix string=
   Offset value=
   Unknown char seq.=
   Compose seq.=
   Input Table ID=
   ```

5. The NLS administration tool shows you the description and gives you the opportunity to change any fields you are not happy with.

Now that you've defined your new map you can use the InfoSphere DataStage administrator client to make it available within your projects.

Map Description Fields

The following table shows the fields of a map description

Table 1. Map Description Fields

Field	Name	Description
0	Map ID	The name used to specify the map in commands and programs.
1	Map Description	A description of the map.
2	Base Map ID	The name of a map to base this one on. This value must be the record ID of another description.
3	Map type	The value of this field must be either SBCS for a single-byte character set, or DBCS for a double-byte or multibyte character set. The default value is SBCS.
4	Table ID	The record ID of the map table that this map description refers to. You do not need to specify a value if the map table has the same ID as the map description.
5	Display length	The display length of all characters in the mapping table specified in field 4. Most double-byte character sets have some characters that print as two display positions on a screen (for example, Hangul characters or CJK ideographs). However, the same map will usually require that ASCII characters are printed as one display position. This field does not pick up a value from any base map description. The default value is 1.

Table 1. Map Description Fields (continued)

Field	Name	Description
6	Unknown char seq.	This field specifies the character sequence to substitute for unknown characters that do not form part of the character set. The value, which is a byte sequence in the external character set, should be a hexadecimal number from one to four bytes. The default value is 3F, the ASCII question mark character. The default is used if neither this map nor any underlying base map has a value in this field.
7	Compose seq.	This field contains the character sequence to compose hexadecimal Unicode values from one to four bytes. If InfoSphere DataStage detects the sequence on input, the next four bytes entered are checked to see if they are hexadecimal values. If so, the Unicode character with that value is entered directly. If neither this map nor any base map has a value in this field, you cannot input Unicode characters by this means. A value of NONE overrides a compose sequence set by an underlying map.
8	Input Table ID	The name of a map table to be used for inputting deadkey sequences.
9	Prefix string	A string in hexadecimal numbers to be prefixed to all external character mappings in the table referenced by field 4. Used mainly for mapping Japanese character sets.
10	Offset value	A value in hexadecimal numbers to be added to each external mapping in the table referenced by field 4. If prefixed by a minus sign, the value is subtracted. Used mainly for mapping Japanese character sets.

How Locales Work

Locales, categories, and conventions have important distinctions that are important to know about.

Before you attempt to create new locales, you need to know a bit more about how InfoSphere DataStage defines Locales.

It is important to distinguish between a locale, a category, and a convention.
- A locale comprises a set of categories.
- A category comprises a set of conventions.
- A convention is a rule describing how data values are input or displayed.

In NLS each locale comprises five categories:
- Time
- Numeric
- Monetary
- Ctype
- Collate

Each category comprises various conventions specific to the type of data in each category.

For example, conventions in the Time category include the names of the days of the week, the strings used to indicate AM or PM, the character that separates the hours, minutes, and seconds, and so forth. You can view this information using the NLS Administration tool:

Examining Locale Conventions

You examine the conventions defined for a locale using the NLS Administration tool.

Before you begin

You must have InfoSphere DataStage Administrator status to perform this task.

About this task

You can run the NLS Administration tool in an engine shell as described in "Running NLS Administration Tool on a Windows engine tier host" on page 3 and "Running NLS Administration Tool on a UNIX engine tier host" on page 3.

Procedure

1. Choose **Locales** > **View**.
2. When prompted for a Locale ID, enter one of the Locale IDs (as listed in the InfoSphere DataStage Administrator). You can then examine the categories from which the Locale is built:
   a. Choose **Categories** > *category_type* > **List all** where *category_type* is the type of category you want to examine. This gives a list of all the categories defined for this type.
   b. Choose **Categories** > *category_type* > **View** where *category_type* is the type of category you want to examine.
   c. When prompted for a Category ID, enter one of the Category IDs (as listed by the List all command).

Example

The following example shows the record for the US-ENGLISH locale as displayed by the NLS Administration tool:

```
Locale name..... USA
Description..... Territory=USA, Language=English
Time/Date....... US-ENGLISH
Numeric......... DEFAULT
Monetary........ USA
Ctype........... DEFAULT
Collate......... DEFAULT
.
.
.
```

A locale can be built from existing conventions without duplication. Different locales can share conventions, and one convention can be based on another.

For example, Canada uses the locales CA-FRENCH and CA-ENGLISH. The two locales are not completely different; they share the same Monetary convention. The records for the CA-FRENCH and CA-ENGLISH locales look like this:

```
Locale name..... CA-FRENCH
Description..... Country=Canada, Language=French
Time/Date....... CA-FRENCH
```

```
                    Numeric.........  CA-FRENCH
                    Monetary........  CANADA
                    Ctype...........  DEFAULT
                    Collate.........  DEFAULT+ACCENT+CASE
                    .
                    .
                    .
                    Locale name.....  CA-ENGLISH
                    Description.....  Country=Canada, Language=English
                    Time/Date.......  CA-ENGLISH
                    Numeric.........  CA-ENGLISH
                    Monetary........  CANADA
                    Ctype...........  DEFAULT
                    Collate.........  DEFAULT
                    .
                    .
                    .
```

Notice that for both locales the Monetary field points to a monetary convention called CANADA. The other fields contain the appropriate value for the language concerned.

A detailed description of the format of the conventions in each category is given in Chapter 5, "NLS and Server Jobs - Supplementary Information," on page 39.

Creating New Locales

If the locales supplied with InfoSphere DataStage do not meet your needs, you can create new ones and use these in your jobs.

You are most likely to want to produce a variant of an existing locale rather than add an entirely new one, InfoSphere DataStage allows you to base a new locale on an existing one and just add or alter the required details.

Note: When you want to produce a variant of an existing locale it is important that you create a new locale based on the existing one. Under no circumstances should you edit one of the locales supplied with InfoSphere DataStage.

Locales are created using the NLS administration tool. This is run in an engine shell as described in "Running NLS Administration Tool on a Windows Server" and "Running NLS Administration Tool on a UNIX Server" . You need to have DataStage Administrator status in order to be able to run this.

The instructions take you through an example which creates a new Locale called GB-ENGLISH-EURO. Such a locale will be needed if and when the UK joins the Euro zone. It is a copy of the GB-ENGLISH locale except that it uses a different monetary category which gives a Euro sign rather than a pound sign.

You will be following these steps:
1. Create a new monetary category (based on an existing one) with a Euro sign as the money symbol.
2. Create a new locale, based on the GB-ENGLISH one, that uses the Euro monetary category.

Creating a New Convention

You can create new conventions if you do not find those that you need with InfoSphere DataStage.

About this task

It is assumed that the UK will keep its existing monetary conventions, that is, decimal separator of . (full stop) and thousands separator of , (comma). You are therefore going to base the UK-EURO category on the existing UK category:

Procedure

1. Choose **Categories** > **Monetary** > **Create** .
2. When prompted enter UK-EURO as the record ID for the new category.
3. When prompted, enter UK as the existing record you want to copy:
4. The NLS Administration tool displays the current UK category and allows you to edit it. Type the number of the line you want to change. InfoSphere DataStage displays the convention heading and you can type in the new data. For the UK-EURO category, you are changing the Currency Symbol and International currency string conventions:

Creating a New Locale

You can create new locales if you do not find the ones you want within InfoSphere DataStage.

About this task

You are going to create the GB-ENGLISH-EURO locale based on the GB-ENGLISH locale. The only difference is that it uses the UK-EURO monetary category.

Procedure

1. Choose **Locales** > **Create**.
2. When prompted, enter GB-ENGLISH-EURO as the id of the record to create.
3. When prompted, enter GB-ENGLISH as the id of the record you are going to base the new locale on:
4. The NLS Administration tool displays the current GB-ENGLISH locale and allows you to edit it. Type the number of the line you want to change. InfoSphere DataStage displays the line heading and you can type in the new data. For the GB-ENGLISH-EURO category, change the MONETARY category to UK-EURO.

Results

Now that you've defined your new locale you can use the Administrator client to make it available within your projects.

Chapter 3. Parallel Jobs and NLS

You can create new maps and new locales for use in InfoSphere DataStage parallel jobs.

Creating New Maps

If the maps supplied with InfoSphere DataStage do not meet your needs, you can create new ones and use these in your jobs.

About this task

You are most likely to want to produce a variant of an existing map rather than add an entirely new one. The system will not allow you to overwrite an existing map, so any maps you create must have a unique name. Note that map names are case insensitive, and ignore underscores, dashes, and spaces, so the map name "cso_iso_latin_1" would be taken as identical to "CSOISOLATIN1".

Source files for all the ASCL_ maps are provided with InfoSphere DataStage (these are the parallel job equivalents of most of the server job maps). You can copy these files and base new ones on them, you should not edit the original ASCL_ files.

The procedure for setting up a new map is:

Procedure
1. Configure your environment to allow map building.
2. Produce a new map source file.
3. Use the supplied tool to build the map.

Setting the Environment

You need to ensure you have the correct environment settings before you create and build new maps.

Windows
Ensure you have the correct environment settings for a Windows system before you create and build new maps.

Typical settings for a Windows system are:

```
set APT_ORCHHOME=/IBM/InformationServer/Server/PXEngine
set APT_CONFIG_FILE=/IBM/InformationServer/Server/Configurations/default.apt
set PATH=%PATH%;C:\IBM\InformationServer\Server\PXEngine\bin;
    C:\IBM\InformationServer\Server\PXEngine\etc
set ICU_DATA=C:\IBM\InformationServer\Server\PXEngine\nls\charmaps
```

Solaris
Ensure you have the correct environment settings for a Solaris system before you create and build new maps.

Typical settings for a Solaris system are:

```
APT_ORCHHOME=/opt/IBM/InformationServer/Server/PXEngine ; export APT_ORCHHOME
PATH=$PATH:$APT_ORCHHOME/bin:$APT_ORCHHOME/etc; export PATH
LD_LIBRARY_PATH=$APT_ORCHHOME/lib ; export LD_LIBRARY_PATH
```

```
APT_CONFIG_FILE=/opt/IBM/InformationServer/Server/Configurations/default.apt ;
export
APT_CONFIG_FILE
ICU_DATA=$APT_ORCHHOME/nls/charmaps
```

HP-UX

Ensure that you have the correct environment settings for an HP-UX system before you create and build new maps.

Typical settings for an HP-UX system are:

```
APT_ORCHHOME=/opt/IBM/InformationServer/Server/PXEngine ; export APT_ORCHHOME
PATH=$PATH:$APT_ORCHHOME/bin:$APT_ORCHHOME/etc; export PATH
SHLIB_PATH=$APT_ORCHHOME/lib ; export LD_LIBRARY_PATH
APT_CONFIG_FILE=/opt/IBM/InformationServer/Server/Configurations/default.apt ;
export
APT_CONFIG_FILE
ICU_DATA=$APT_ORCHHOME/nls/charmaps
```

AIX

Ensure you have the correct environment settings for AIX® system before you create and build new maps.

Typical settings for an AIX system are:

```
APT_ORCHHOME=/opt/IBM/InformationServer/Server/PXEngine ; export APT_ORCHHOME
PATH=$PATH:$APT_ORCHHOME/bin:$APT_ORCHHOME/etc ; export PATH
LIBPATH=$APT_ORCHHOME/lib ; export LIBPATH
APT_CONFIG_FILE=/opt/IBM/InformationServer/Server/Configurations/default.apt ;
export
APT_CONFIG_FILE
ICU_DATA=$APT_ORCHHOME/nls/charmaps
```

LINUX

Ensure you have the correct environment settings for a LINUX system before you create and build new maps.

Typical settings for a LINUX system are:

```
APT_ORCHHOME=/opt/IBM/InformationServer/Server//PXEngine ; export APT_ORCHHOME
PATH=$PATH:$APT_ORCHHOME/bin:$APT_ORCHHOME/etc; export PATH
LD_LIBRARY_PATH=$APT_ORCHHOME/lib ; export LD_LIBRARY_PATH
APT_CONFIG_FILE=/opt/IBM/InformationServer/Server//Configurations/default.apt ;
export
APT_CONFIG_FILE
ICU_DATA=$APT_ORCHHOME/nls/charmaps
```

Map Source Files

Use map source files (.ucm files) to help create new maps.

Map source files end in *.ucm*. They are located in:

```
$APT_ORCHHOME/nls/charmaps
```

and must be built from this location.

As an example, you will create a new map called MY_ASCII which is based on the ASCL_ASCII map, except the input character 0x23 is mapped to the UK pound sign (£) instead of the number symbol (#). To create this new map:

1. In the `$APT_ORCHHOME/nls/charmaps` directory, copy `ASCL_ASCII.ucm` to `MY_ASCII.ucm`.

2. Edit the `MY_ASCII.ucm` file. The format is fairly self-explanatory. The header information identifies the character set. The map itself is described between "CHARMAP" and "END CHARMAP". The string *<UNNNN>* gives the Unicode character in hexadecimal. The string *\xNN* gives the map character in hexadecimal.

Find the line that for the input character 0x23:

```
<U0023> \x23 |0
```

Replace the value for the Unicode character:

```
<U00A3> \x23 |0
```

3. Write the file. It is now ready to be built.

Building a New Map

If the maps supplied with InfoSphere® DataStage® do not meet your needs, you can create a new one and use it in your job.

About this task

The example map is built in the `$APT_ORCHHOME/nls/charmaps` directory using the following command:

```
addCustomMaps.sh MY_ASCII.ucm
```

Once the build is complete, the map is visible in your parallel jobs and is ready to use.

If the map is not visible in your parallel jobs, you can add it with the following procedure.

Procedure

1. Set the environment variables for your system.

Option	Description
Windows	Ensure PATH contains `$APT_ORCHHOME\bin`.
AIX	Ensure PATH contains `$APT_ORCHHOME/bin`. Ensure LIBPATH contains `$APT_ORCHHOME/lib`
HPUX (not on Intel Itanium)	Ensure PATH contains `$APT_ORCHHOME/bin` Ensure SHLIB_PATH contains `$APT_ORCHHOME/lib`
Other systems	Ensure PATH contains `$APT_ORCHHOME/bin` Ensure LD_LIBRARY_PATH contains `$APT_ORCHHOME/lib`

2. Edit the `addCustomMaps.sh` file so that it creates a symlink to the `cnvalias.icu` file. Replace the last line that says `echo "DONE"`.

Option	Description
| Big-endian systems | ```
# ICU 3.2 workaround:
# create a symlink for the compiled alias file
# so the map list is populated correctly.
binAliasFile=cnvalias\.icu
icuPlatformPrefix=icudt32b_
ln -s $ICU_DATA$binAliasFile
  $ICU_DATA$icuPlatformPrefix$binAliasFile
echo "\nD O N E\n"
``` |
| Little-endian systems | ```
# ICU 3.2 workaround:
# create a symlink for the compiled alias file
# so the map list is populated correctly.
binAliasFile=cnvalias\.icu
icuPlatformPrefix=icudt32l_
ln -s $ICU_DATA$binAliasFile
  $ICU_DATA$icuPlatformPrefix$binAliasFile
echo "\nD O N E\n"
``` |

3. Edit the `convrtrs.txt` file. Delete the entry for your new map, `MY_ASCII.ucm`, if the entry is in the file.
4. Issue the following command to create the `cnvalias.icu` file, and to create a symlink to the following file:`addCustomMaps.sh MY_ASCII.ucm`
5. Ensure that the ICU_DATA environment variable is set in the DataStage environment.
6. Stop and restart the DataStage engine.

Results

The map is visible in your parallel jobs and is ready to use.

Deleting a Custom Map

If the maps supplied with InfoSphere DataStage do not meet your needs, you can create new ones and use these in your jobs. You can delete these custom maps if needed.

About this task

If you subsequently want to delete a custom map:

Procedure

1. Edit the file `$APT_ORCHHOME/nls/charmaps/convrtrs.txt`.
2. Go to the last section in the file, headed "User added custom map" and delete the name of the offending map.
3. From the `$APT_ORCHHOME/nls/charmaps` directory, execute the following command:

   `gencnval convrtrs.txt`

Overriding collate conventions

InfoSphere DataStage allows you to tailor existing collate conventions by adding rules to them. The rules that you add override what is set by the current locale. You specify the new rules in a text file which you can reference at project, job, or stage level.

Text File Basic Format

You specify new rules that tailor existing collate conventions in a text file that you can reference at project, job, or stage level.

The text file comprises a set of one or more rules, each on a separate line. Each rule contains a string of ordered characters that starts with an anchor point This is an absolute point that determines the order of other characters. It has the format *&character.* For example &a means the character "a" is the anchor point, all other rules on that line are relative to that letter. The following table gives the other symbols you can use:

Symbol	Example	Description
<	a < b	Identifies a primary (base letter) difference between "a" and "b"
<<	a<<ä	Signifies a secondary (accent) difference between "a" and "ä"
<<<	a<<<A	Identifies a tertiary difference between "a" and "A"
=	x =y	Signifies no difference between "x" and "y"

For example, the rule &a < g has the following sorting consequences:

Without Rule	With Rule
apple	apple
Abernathy	Abernathy
bird	green
Boston	bird
green	Boston
Graham	Graham

Add the rule &A<<<G and the sorting would be as follows:

With Additional Rule
apple
Abernathy
green
Graham
bird
Boston

There are also options that you can specify in the file, and more advanced syntactical elements that you can use. These are described in full at:

http://oss.software.ibm.com/icu/userguide/Collate_Customization.html

For details of the UCA rules see:

http://www.unicode.org/unicode/reports/tr10/

Using an Override File

Once you have set up an override file you can reference it at project level, job level or stage level.

Using an Override File at Project Level

Once you have set up an override file you can reference it at project level.

Procedure

1. Open the InfoSphere DataStage Administrator.
2. Click the **Projects** tab to go to the Projects page.
3. Select the project for which you want to set a default map and click the **NLS...** button to open the Project NLS Settings dialog box for that project. Click the **Parallel Locales** tab to go to the Parallel Locales page.
4. Click the browse button next to the **Collate** list box.
5. Browse for the file containing the override rules.

Using an Override File at Job Level

Once you have set up an override file you can reference it at job level.

Procedure

1. Open the job for which you want to set the locale in the Designer client.
2. Open the Job Properties dialog box for that job (choose **Edit Job Properties**).
3. Click the **NLS** tab to go to the NLS page.
4. Click the browse button next to the **Default collation locale for stages** list box.
5. Browse for the file containing the override rules.

Results

Using an Override File at Project Level

Using an Override File at Stage Level from the Stage Tab

Once you have set up an override file you can reference it at stage level from the Stage tab.

Procedure

1. Open the stage editor and go to the NLS Locale tab of the Stage page.
2. Click the arrow button next to the **Collate** list box and choose **Browse for file...** from the shortcut menu..
3. Browse for the file containing the override rules and select it.

Using an Override File at Stage Level from the Partition Tab

Once you have set up an override file you can reference it at stage level from the Partition tab.

Procedure

1. Open the stage editor and go to the Partitioning tab of the Inputs page.
2. Click the properties button in the **Sorting** area. The Sort Properties dialog box opens.
3. Click the arrow button next to the **Collate** list box and choose **Browse for file...** from the shortcut menu

4. Browse for the file containing the override rules and select it.

Chapter 4. Maps and Locales Supplied with InfoSphere DataStage

This appendix provides lists of the character set maps and locales that are supplied with InfoSphere DataStage.

Server Job Character Set Maps

A number of character set maps for server jobs is supplied with InfoSphere DataStage.

To obtain a list of the character set maps for server jobs that are supplied with InfoSphere DataStage, issue these commands:

1. cd $DSHOME
2. $DSHOME/bin/uvsh "LIST.MAPS ALL"

To obtain a list of the character set maps for server jobs that are loaded and ready for use, issue these commands:

1. cd $DSHOME
2. $DSHOME/bin/uvsh "LIST.MAPS"

The following list shows all the maps for major character sets used worldwide that are supplied with InfoSphere DataStage for use with server jobs. The left column contains the name of the map, the middle column contains the name of the map table used by the map (in NLS.MAP.TABLES), and the right column contains a description of the map.

Note: The maps that are supported by your installation of IBM® InfoSphere Information Server might vary from this list. Use the commands shown above to see the maps that are supported by your system.

Character Set	Table Name	Description
ASCII	ASCII	Standard ASCII 7-bit set
ASCII+C1	ASCII	ASCII 7-bit + C1 control chars
ASCII+MARKS	UV-MARKS	Std ASCII 7-bit set for type 1&19 files w/ marks
BIG5	BIG5	AIWAN: "Big 5" standard
C0-CONTROLS	C0-CONTROLS	Standard ISO2022 C0 control set, chars 00-1F+7F
C1-CONTROLS	C1-CONTROLS	Standard 8-bit ISO control set, 80-9F
EBCDIC	EBCDIC	IBM EBCDIC as implemented by standard uniVerse - control chars only
EBCDIC-037	EBCDIC-037	IBM EBCDIC variant 037
EBCDIC-1026	EBCDIC-1026	IBM EBCDIC variant 1026 (Turkish)

Character Set	Table Name	Description
EBCDIC-500V1	EBCDIC-500V1	IBM EBCDIC variant 500V1
EBCDIC-875	EBCDIC-875	IBM EBCDIC variant 875 (Greek)
EBCDIC-CTRLS	EBCDIC-CTRLS	IBM EBCDIC as implemented by standard uniVerse - control chars only
GB2312	GB2312-80	CHINESE: EUC as described by GB 2312
ISO8859-1	ISO8859-1	Standard ISO8859 part 1: Latin-1
ISO8859-1+MARKS	ISO8859-1+MARKS	Standard ISO8859 part 1: Latin-1 for type 1& 19 files with marks
ISO8859-10	ISO8859-10	Standard ISO8859 part 10: Latin-6
ISO8859-2	ISO8859-2	Standard ISO8859 part 2: Latin-2
ISO8859-3	ISO8859-3	Standard ISO8859 part 3: Latin-3
ISO8859-4	ISO8859-4	Standard ISO8859 part 4: Latin-4
ISO8859-5	ISO8859-5	Standard ISO8859 part 5: Latin-Cyrillic
ISO8859-6	ISO8859-6	Standard ISO8859 part 6: Latin-Arabic
ISO8859-7	ISO8859-7	Standard ISO8859 part 7: Latin-Greek
ISO8859-8	ISO8859-8	Standard ISO8859 part 8: Latin-Hebrew
ISO8859-9	ISO8859-9	Standard ISO8859 part 5: Latin-5
JIS-EUC	JISX0208	JAPANESE: EUC excluding JIS X 0212 Kanji
JIS-EUC+	JISX0212	JAPANESE: EUC including JIS X 0212 Kanji
JIS-EUC-HWK	JISX0201-K	JAPANESE: 1/2 width katakana for JIS-EUC
JIS-EUC2	JISX0208	JAPANESE: EUC fixed width excluding JIS X 02 12 kanji
JIS-EUC2-C0	C0-CONTROLS	JAPANESE: EUC2 fixed width C0 control chars
JIS-EUC2-C1	C1-CONTROLS	JAPANESE: EUC fixed width C1 control chars
JIS-EUC2-HWK	JISX0201-K	JAPANESE: EUC fixed width representation of 1 /2 width katakana
JIS-EUC2-MARKS	JIS-EUC2-MARKS	JAPANESE: EUC2 fixed width mark characters (external form

Character Set	Table Name	Description
JIS-EUC2-ROMAN	JISX0201-A	JAPANESE: Variant of 7-bit ASCII
JISX0201	JISX0201-K	JAPANESE: Single-byte set, 1/2 width katakana + ASCII
KOI8-R	KOI8-R	KOI8-R Russian/Cyrillic set
KSC5601	KSC5601	#KOREAN: Wansung code as described by KS C 5601-1987
MAC-GREEK	MAC-GREEK	Apple Macintosh Greek Repertoire (like ISO8859-7)
MAC-GREEK2	MAC-GREEK2	Apple Macintosh Greek Repertoire based on APPLE II
MAC-ROMAN	MAC-ROMAN	Apple Macintosh Roman character set, based on ASCII
MNEMONICS		ASCII mnemonics for many Unicodes, based on UTF8
MNEMONICS-1	ISO8859-1	As for MNEMONICS, but ISO8859-1 capable
MS1250	MS1250	MS Windows code page 1250 (Latin 2)
MS1251	MS1251	MS Windows code page 1251 (Cyrillic)
MS1252	MS1252	MS Windows code page 1252 (Latin 1)
MS1253	MS1253	MS Windows code page 1253 (Greek)
MS1254	MS1254	MS Windows code page 1254 (Turkish)
MS1255	MS1255	MS Windows code page 1255 (Hebrew)
MS1256	MS1256	MS Windows code page 1256 (Arabic)
PC1040	PC1040	PC DOS code page 1040 (Korean)
PC1041	PC1041	PC DOS code page 1041 (Japanese)
PC437	PC437	PC DOS code page 437 (US)
PC850	PC850	PC DOS code page 850 (Latin 1)
PC852	PC852	PC DOS code page 852 (Latin 2)
PC855	PC855	PC DOS code page 855 (Cyrillic)
PC857	PC857	PC DOS code page 857 (Turkish)
PC860	PC860	PC DOS code page 860 (Portuguese)

Character Set	Table Name	Description
PC861	PC861	PC DOS code page 861 (Icelandic)
PC863	PC863	PC DOS code page 863 (Canada-Fr)
PC864	PC864	PC DOS code page 864 (Arabic)
PC865	PC865	PC DOS code page 865 (Nordic)
PC866	PC866	PC DOS code page 866 (Cyrillic)
PC869	PC869	PC DOS code page 869 (Greek)
PIECS	PIECS	PI and PI/open Extended Character Set
PRIME-SHIFT-JIS	PJISX0208	JAPANESE: Shift-JIS main map (Prime variant)
SHIFT-JIS	SJISX0208	JAPANESE: Shift-JIS main map
TAU-SHIFT-JIS	TJISX0208	JAPANESE: Shift-JIS main map (Tau variant)
TIS620	TIS620-A	THAI: standard TIS 620 ("Thai ASCII")
TIS620-B	TIS620-B	Non-spacing characters part of TIS620 (Thai)

Server Job Locales

A number of locales for use with server jobs is supplied withInfoSphere DataStage.

To obtain a list of the locales for server jobs that are supplied with InfoSphere DataStage, issue these commands:

1. cd $DSHOME
2. $DSHOME/bin/uvsh "LIST.LOCALES ALL"

To obtain a list of the locales for server jobs that are loaded and ready for use, issue these commands:

1. cd $DSHOME
2. $DSHOME/bin/uvsh "LIST.LOCALES"

The following list shows the locales supplied with InfoSphere DataStage for use with server jobs, the territory that uses each locale, and the relevant language.

Note: The locales that are supported by your installation of IBM InfoSphere Information Server might vary from this list. Use the commands shown above to see the locales that are supported by your system.

Locale	Description
AR-SPANISH	Territory=Argentina, Language=Spanish
AT-GERMAN	Territory=Austria, Language=German

Locale	Description
AU-ENGLISH	Territory=Australia, Language=English
BE-DUTCH	Territory=Belgium, Language=Dutch
BE-FRENCH	Territory=Belgium, Language=French
BE-GERMAN	Territory=Belgium, Language=German
BG-BULGARIAN	Territory=Bulgaria, Language=Bulgarian
BO-SPANISH	Territory=Bolivia, Language=Spanish
BR-PORTUGUESE	Territory=Brazil, Language=Portuguese
CA-ENGLISH	Territory=Canada, Language=English
CA-FRENCH	Territory=Canada, Language=French
CH-FRENCH	Territory=Switzerland, Language=French
CH-GERMAN	Territory=Switzerland, Language=German
CH-ITALIAN	Territory=Switzerland, Language=Italian
CL-SPANISH	Territory=Chile, Language=Spanish
CN-CHINESE	Territory=China (PRC), Language=Chinese
CO-SPANISH	Territory=Colombia, Language=Spanish
CR-SPANISH	Territory=Costa Rica, Language=Spanish
CZ-CZECH	Territory=Czech Republic, Language=Czech
DE-GERMAN	Territory=Germany, Language=German
DK-DANISH	Territory=Denmark, Language=Danish
DO-SPANISH	Territory=Dominican Republic, Language=Spanish
EC-SPANISH	Territory=Ecuador, Language=Spanish
EV-SPANISH	Territory=El Salvador, Language=Spanish
FI-FINNISH	Territory=Finland, Language=Finnish
FO-FAEROESE	Territory=Faeroe Islands, Language=Faeroese
FR-FRENCH	Territory=France, Language=French
GB-ENGLISH	Territory=UK, Language=English
GL-GREENLANDIC	Territory=Greenland, Language=Greenlandic
GR-GREEK	Territory=Greece, Language=Greek
GT-SPANISH	Territory=Guatemala, Language=Spanish
HN-SPANISH	Territory=Honduras, Language=Spanish
HR-CROATIAN	Territory=Croatia, Language=Croatian
HU-HUNGARIAN	Territory=Hungary, Language=Hungarian
IE-ENGLISH	Territory=Ireland, Language=English
IL-ENGLISH	Territory=Israel, Language=English
IL-HEBREW	Territory=Israel, Language=Hebrew
IS-ICELANDIC	Territory=Iceland, Language=Icelandic
IT-ITALIAN	Territory=Italy, Language=Italian
JP-JAPANESE	Territory=Japan, Language=Japanese
KP-KOREAN	Territory=Democratic People's Republic of Korea (NORTH), Language=Korean

Locale	Description
KR-KOREAN	Territory=Republic of Korea (SOUTH), Language=Korean
LT-LITHUANIAN	Territory=Lithuania, Language=Lithuanian
LV-LATVIAN	Territory=Latvia, Language=Latvian
MX-SPANISH	Territory=Mexico, Language=Spanish
NL-DUTCH	Territory=Netherlands, Language=Dutch
NO-NORWEGIAN	Territory=Norway, Language=Norwegian
NZ-ENGLISH	Territory=New Zealand, Language=English
PA-SPANISH	Territory=Panama, Language=Spanish
PE-SPANISH	Territory=Peru, Language=Spanish
PL-POLISH	Territory=Poland, Language=Polish
PT-PORTUGUESE	Territory=Portugal, Language=Portuguese
RO-ROMANIAN	Territory=Romania, Language=Romanian
RU-RUSSIAN	Territory=Russia, Language=Russian
SE-SWEDISH	Territory=Sweden, Language=Swedish
SI-SLOVENIAN	Territory=Slovenia, Language=Slovenian
TR-TURKISH	Territory=Turkey, Language=Turkish
TW-CHINESE	Territory=Taiwan, Language=Chinese
US-ENGLISH	Territory=USA, Language=English
UY-SPANISH	Territory=Uruguay, Language=Spanish
VE-SPANISH	Territory=Venezuela, Language=Spanish
ZA-ENGLISH	Territory=South Africa, Language=English

Parallel Job Character Set Maps

A number of character set maps for parallel jobs are supplied with InfoSphere DataStage.

To obtain a list of the character set maps for parallel jobs that are supplied with InfoSphere DataStage, issue this command:

```
$APT_ORCHHOME/bin/icu_utils listmaps
```

The following table lists the character set maps available for parallel maps. The maps whose names start with ASCL_ are the equivalents of the server job maps - see "Server Job Character Set Maps" on page 19. (Parallel job versions of most of the server job maps are supplied).

Note: The maps that are supported by your installation of IBM InfoSphere Information Server might vary from this list. Use the command shown above to see the maps that are supported by your system.

Character Set	Description
Big5	Chinese for Taiwan Multi-byte set
BOCU-1	Compressed UTF-8 (http://www.unicode.org/notes/tn6)

Character Set	Description
CESU-8	8-bit Compatibility Encoding Scheme for UTF-16 (http://www.unicode.org/unicode/reports/tr26
EUC-KR	Korean for Internet messages
Extended_UNIX_Code_Packed_Format_for_Japanese	Extended UNIX Code Packed Format for Japanese
ebcdic-xml-us	EBCDIC for XML (US)
GB_2312-80	Chinese (1980)
GB2312	Chinese (1999)
GBK	Chinese (1995)
gb18030	Chinese (2000)
HZ-GB-2312	Chinese (HZ)
hp-roman8	http://www.faqs.org/rfcs/rfc1345.htm
IBM00858	IBM codepage 850 (multilingual) with Euro symbol
IBM01140	EBCDIC US with Euro symbol
IBM01141	EBCDIC German with Euro symbol
IBM01142	EBCDIC Danish/Norwegian with Euro symbol
IBM01143	EBCDIC Finnish/Swedish with Euro symbol
IBM01144	EBCDIC Italian with Euro symbol
IBM01145	EBCDIC Spanish with Euro symbol
IBM01146	EBCDIC GB with Euro symbol
IBM01147	EBCDIC French with Euro symbol
IBM01148	EBCDIC international with Euro symbol
IBM01149	EBCDIC Icelandic with Euro symbol
IBM037	EPCDIC CP US
IBM1026	EBCDIC Latin-5 Turkey
IBM273	EBCDIC Austria, Germany
IBM277	EBCDIC Denmark, Norway
IBM278	EBCDIC Sweden, Finland
IBM280	EBCDIC Italy
IBM284	EBCDIC Spanish
IBM285	EBCDIC GB
IBM290	EBCDIC Japanese (kana)
IBM297	EBCDIC France
IBM367	ASCII
IBM420	EBCDIC Arabic
IBM424	EBCDIC Hebrew
IBM500	EBCDIC International
IBM850	MS-DOS Latin-1
IBM851	MS-DOS Greek

Character Set	Description
IBM852	MS-DOS Latin-2
IBM855	EBCDIC Cyrillic
IBM857	EBCDIC Turkey
IBM860	MS-DOS Portuguese
IBM861	MS-DOS Icelandic
IBM862	PC Hebrew
IBM863	MS-DOS Canadian French
IBM864	PC Arabic
IBM865	MS-DOS Nordic
IBM868	MS-DOS Pakistan
IBM869	EBCDIC Modern Greek
IBM870	EBCDIC Multilingual Latin-2
IBM871	EBCDIC Iceland
IBM918	EBCDIC Pakistan(Urdu)
ISCII,version=0	Indian Standard Code for Information Interchange, version 0
ISCII,version=1	Indian Standard Code for Information Interchange, version 1
ISCII,version=2	Indian Standard Code for Information Interchange, version 2
ISCII,version=3	Indian Standard Code for Information Interchange, version 3
ISCII,version=4	Indian Standard Code for Information Interchange, version 4
ISCII,version=5	Indian Standard Code for Information Interchange, version 5
ISCII,version=6	Indian Standard Code for Information Interchange, version 6
ISCII,version=7	Indian Standard Code for Information Interchange, version 7
ISCII,version=8	Indian Standard Code for Information Interchange, version 8
ISO-2022-CN	Chinese
ISO-2022-CN-EXT	Chinese extended
ISO-2022-JP	Japanese (JIS)
ISO-2022-JP-2	Japanese (JIS) extension
ISO-2022-KR	Korean
ISO_2022,locale=ja,version=3	(JIS7)
ISO_2022,locale=ja,version=4	(JIS8)
ISO_2022,locale=zh,version=2	Chinese extended
ISO_2022,locale=ko,version=1	Korean TCP
ISO_8859-1:1987	Latin alphabet No. 1
ISO_8859-2:1987	Latin alphabet No. 2

Character Set	Description
ISO_8859-3:1988	Latin alphabet No. 3
ISO_8859-4:1988	Latin alphabet No. 4
ISO_8859-5:1988	Latin/Cyrillic alphabet
ISO_8859-6:1987	Latin/Arabic alphabet
ISO_8859-7:1987	Latin/Greek alphabet
ISO_8859-8:1988	Latin/Hebrew alphabet
ISO_8859-9:1989	Latin alphabet No. 5
ISO-8859-10	Latin alphabet No. 6
iso-8859_11-2001	Latin/Thai character set
ISO-8859-13	Latin alphabet No. 7
ISO-8859-14	Latin alphabet No. 8 (Celtic)
ISO-8859-15	Latin alphabet No. 9
ibm-1006_P100-1995	ISO Urdu
ibm-1025_P100-1995	EBCDIC Cyrillic
ibm-1047_P100-1995	EBCDIC Open Edition
ibm-1047_P100-1995,swaplfnl	EBCDIC Open Edition with LF/NL swapped
ibm-1097_P100-1995	EBCDIC Farsi
ibm-1098_P100-1995	ISO Farsi
ibm-1112_P100-1995	EBCDIC Baltic
ibm-1122_P100-1999	EBCDIC Estonia
ibm-1123_P100-1995	EBCDIC Ukraine
ibm-1124_P100-1996	PC Ukraine
ibm-1125_P100-1997	PC Cyrillic Ukraine
ibm-1129_P100-1997	ISO Vietnamese
ibm-1130_P100-1997	EBCDIC Vietnamese
ibm-1131_P100-1997	PC Cyrillic Belarus
ibm-1132_P100-1998	EBCDIC Lao
ibm-1133_P100-1997	ISO Lao
ibm-1137_P100-1999	EBCDIC Devanagari with LF/NL swapped
ibm-1140_P100-1997,swaplfnl	EBCDIC United States with LF/NL swapped
ibm-1141_P100-1997,swaplfnl	
ibm-1142_P100-1997	EBCDIC Denmark, Norway
ibm-1142_P100-1997,swaplfnl	EBCDIC Denmark, Norway with LF/NL swapped
ibm-1143_P100-1997,swaplfnl	EBCDIC Finland, Sweden with LF/NL swapped
ibm-1144_P100-1997,swaplfnl	EBCDIC Italy with LF/NL swapped
ibm-1145_P100-1997,swaplfnl	EBCDIC Spain with LF/NL swapped
ibm-1146_P100-1997,swaplfnl	EBCDIC UK, Ireland with LF/NL swapped
ibm-1147_P100-1997,swaplfnl	EBCDIC France with LF/NL swapped
ibm-1148_P100-1997,swaplfnl	EBCDIC Multilingual with LF/NL swapped

Character Set	Description
ibm-1149_P100-1997,swaplfnl	EBCDIC Iceland with LF/NL swapped
ibm-1153_P100-1999	As ibm-1153
ibm-1153_P100-1999,swaplfnl	As ibm-1153 with LF/NL swapped
ibm-1154_P100-1999	EBCDIC Cyrillic Multilingual
ibm-1156_P100-1999	ibm-1157
ibm-1157_P100-1999	EBCDIC Estonia
ibm-1158_P100-1999	EBCDIC Cyrillic Ukraine
ibm-1160_P100-1999	EBCDIC Thailand
ibm-1162_P100-1999	
ibm-1164_P100-1999	EBCDIC Vietnam
ibm-1168_P100-2002	KOI8-U
ibm-1250_P100-1995	Windows Latin 2
ibm-1251_P100-1995	Windows Cyrillic
ibm-1252_P100-2000	Windows Latin 1
ibm-1253_P100-1995	Windows Greek
ibm-1254_P100-1995	Windows Latin 5 (Turkey)
ibm-1255_P100-1995	Windows Hebrew
ibm-1256_P110-1997	Windows Arabic
ibm-1257_P100-1995	Windows Latin 4 (Baltic)
ibm-1258_P100-1997	Windows Vietnamese
ibm-12712_P100-1998	EBCDIC Hebrew
ibm-12712_P100-1998,swapflnl	EBCDIC Hebrew with LF/NL swapped
ibm-1276_P100-1995	Adobe Standard Encoding
ibm-1277_P100-1995	Adobe Latin1 Encoding
ibm-1363_P110-1997	PC Korea KS extended
ibm-1363_P11B-1998	PC Korea KS extended
ibm-1364_P110-1997	EBCDIC Korea KS extended
ibm-1364_P110-2007	EBCDIC Korea KS extended
ibm-1371_P100-1999	EBCDIC Taiwan (euro)
ibm-1375_P100-2003	Big5-HKSCS
ibm-1375_P100-2007	Big5-HKSCS
ibm-1381_P110-1999	PC China GB
ibm-1386_P100-2001	
ibm-1386_P100-2002	
ibm-1388_P103-2001	EBCDIC China GBK
ibm-1390_P110-2003	EBCDIC Japan Katakana (euro)
ibm-1399_P110-2003	EBCDIC Japan Latin (euro)
ibm-16804_X110-1999	
ibm-16804_X110-1999,swaplfnl	
ibm-16684_P110-2003	DBCS Jis + Roman Jis Host
ibm-33722_P120-1999	EUC Japan

Character Set	Description
ibm-33722_P12A-1999	
ibm-367_P100-1995	
ibm-37_P100-1995,swaplfnl	EBCDIC United States with LF/NL swapped
ibm-1373_P100-2002	
ibm-4517_P100-2005	
ibm-437_P100-1995	IBM437
ibm-4899_P100-1998	Old EBCDIC Hebrew
ibm-4909_P100-1999	
ibm-4971_P100-1999	EBCDIC Greek
ibm-5123_P100-1999	Host Roman Jis
ibm-5351_P100-1998	
ibm-5352_P100-1998	
ibm-5353_P100-1998	
ibm-5471_P100-2006	
ibm-720_P100-1997	
ibm-737_P100-1997	
ibm-775_P100-1996	
ibm-803_P100-1999	
ibm-813_P100-1995	ISO Greek
ibm-838_P100-1995	IBM-Thai
ibm-8482_P100-1999	host SBCS (Katakana)
ibm-856_P100-1995	PC Hebrew (old)
ibm-866_P100-1995	PC Russia
ibm-867_P100-1998	PC Israel
ibm-874_P100-1995	PC Thai
ibm-875_P100-1995	EBCDIC Greek
ibm-878_P100-1996	KOI8-R
ibm-897_P100-1995	
ibm-901_P100-1999	PC Baltic
ibm-902_P100-1999	PC Estonian
ibm-9067_X100-2005	
ibm-916_P100-1995	
ibm-918_P100-1995	
ibm-921_P100-1995	PC Baltic
ibm-922_P100-1999	PC Estonian
ibm-930_P120-1999	EBCDIC Japan DBCS
ibm-933_P110-1995	EBCDIC Korea DBCS
ibm-935_P110-1999	EBCDIC China DBCS
ibm-937_P110-1999	EBCDIC Taiwan DBCS
ibm-939_P120-1999	EBCDIC Japan Extended DBCS
ibm-942_P12A-1999	PC Japan SJIS-78 syntax

Character Set	Description
ibm-943_P130-1999	PC Japan SJIS-90
ibm-949_P110-1999	PC DBCS-only Taiwan
ibm-949_P11A-1999	
ibm-950_P110-1999	PC Taiwan
Ibm-954_P101-2007	
ibm-954_P101-2000	
ibm-964_P110-1999	EUC Taiwan
ibm-970_P110-1995	
ibm-971_P100-1995	
IMAP-mailbox-name	
JIS_Encoding	
LMBCS-1	Lotus multi-byte character set – Latin 1
macintosh	Macintosh
macos-2566-10.2	
macos-29-10.2	x-mac-centarleurroman
macos-35-10.2	x-mac-turkish
macos-6-10.2	
macos-6_2-10.4	x-mac-greek
macos-7_3-10.2	x-mac-cyrillic
SCSU	http://www.iana.org/assignments/charset-reg/SCSU
Shift_JIS	Shift-JIS, Japanese
US-ASCII	ANSI_X3.4-1968
TIS-620	TIS-620, Thai
UTF-16	UTF-16 Unicode
UTF-16BE	UTF-16 Unicode Big Endian
UTF-16LE	UTF-16 Unicode Little Endian
UTF-32	UTF-32 Unicode
UTF-32BE	UTF-32 Unicode Big Endian
UTF-32LE	UTF-32 Unicode Little Endian
UTF-7	UTF-7 Unicode
UTF-8	UTF-8 Unicode
windows-874-2000	
windows-949-2000	
windows-1250	Windows Latin 2
windows-1251	Windows Cyrillic
windows-1252	Windows Latin 1
windows-1253	Windows Greek
windows-1254	Windows Latin 5 (Turkey)
windows-1255	Windows Hebrew
windows-1256	Windows Arabic

Character Set	Description
windows-1256-2000	
windows-1257	Windows Latin 4 (Baltic)
windows-1258	Windows Vietnamese
X11-compound-text	Compound text strings

Parallel Job Locales

A number of locales for parallel jobs are supplied with InfoSphere DataStage.

To obtain a list of the locales for parallel jobs that are supplied with InfoSphere DataStage, issue this command:

```
$APT_ORCHHOME/bin/icu_utils listlocales
```

The following list shows the locales supplied with InfoSphere DataStage for use with parallel jobs for collation purposes, the territory that uses each locale, and the relevant language.

Note: The locales that are supported by your installation of IBM InfoSphere Information Server might vary from this list. Use the command shown above to see the locales that are supported by your system.

Locale	Description
af	Language=Afrikaans
af_ZA	Language=Afrikaans, Territory=South Africa
am	Language=Amharic
am_ET	Language=Amharic, Territory=Ethiopia
ar	Language=Arabic
ar_AE	Language=Arabic, Territory=United Arab Emirates
ar_BH	Language=Arabic, Territory=Bahrain
ar_DZ	Language=Arabic, Territory=Algeria
ar_EG	Language=Arabic, Territory=Egypt
ar_IN	Language=Arabic, Territory=India
ar_IQ	Language=Arabic, Territory=Iraq
ar_JO	Language=Arabic, Territory=Jordan
ar_KW	Language=Arabic, Territory=Kuwait
ar_LB	Language=Arabic, Territory=Lebanon
ar_LY	Language=Arabic, Territory=Libya
ar_MA	Language=Arabic, Territory=Morocco
ar_OM	Language=Arabic, Territory=Oman
ar_QA	Language=Arabic, Territory=Qatar
ar_SA	Language=Arabic, Territory=Saudi Arabia
ar_SD	Language=Arabic, Territory=Sudan
ar_SY	Language=Arabic, Territory=Syria
ar_TN	Language=Arabic, Territory=Tunisia

Locale	Description
ar_YE	Language=Arabic, Territory=Yemen
be	Language=Belarusian
be_BY	Language=Belarusian, Territory=Belarus
bg	Language=Bulgarian
bg_BG	Language=Bulgarian, Territory=Bulgaria
bn	Language=Bengali
bn_IN	Language=Bengali, Territory=India
ca	Language=Catalan
ca_ES	Language=Catalan, Territory=Spain
ca_ES_PREEURO	Language=Catalan
cs	Language=Czech
cs_CZ	Language=Czech
da	Language=Danish
da_DK	Language=Danish, Territory=Denmark
de	Language=German
de_PHONEBOOK	Language=German, Territory=Phonebook order
de_AT	Language=German, Territory=Austria
de_AT_PREEURO	Language=German, Territory=Austria
de_BE	Language=German, Territory=Belgium
de_CH	Language=German, Territory=Switzerland
de_DE	Language=German, Territory=Germany
de_DE_PREEURO	Language=German, Territory=Germany
de_LU	Language=German, Territory=Luxembourg
de_LU_PREEURO	Language=German, Territory=Luxembourg
el	Language=Greek
el_GR	Language=Greek, Territory=Greece
el_GR_PREEURO	Language=Greek, Territory=Greece
en	Language=English
en_AU	Language=English, Territory=Australia
en_BE	Language=English, Territory=Belgium
en_BE_PREEURO	Language=English, Territory=Belgium
en_BW	Language=English, Territory=Botswana
en_CA	Language=English, Territory=Canada
en_GB	Language=English, Territory=Great Britain
en_GB_EURO	Language=English, Territory=Great Britain
en_HK	Language=English, Territory=Hong Kong S.A.R of China
en_IE	Language=English, Territory=Ireland
en_IE_PREEURO	Language=English, Territory=Ireland
en_IN	Language=English, Territory=India

Locale	Description
en_MT	Language=English, Territory=Malta
en_NZ	Language=English, Territory=New Zealand
en_PH	Language=English, Territory=Philippines
en_SG	Language=English, Territory=Singapore
en_US	Language=English, Territory=United States
en_US_POSIX	Language=English, Territory=United States
en_VI	Language=English, Territory=U.S. Virgin Islands
en_ZA	Language=English, Territory=South Africa
en_ZW	Language=English, Territory=Zimbabwe
eo	Language=Esperanto
es	Language=Spanish
es_TRADITIONAL	Language=Spanish
es_AR	Language=Spanish, Territory=Argentina
es_BO	Language=Spanish, Territory=Bolivia
es_CL	Language=Spanish, Territory=Chile
es_CO	Language=Spanish, Territory=Colombia
es_CR	Language=Spanish, Territory=Costa Rica
es_DO	Language=Spanish, Territory=Dominican Republic
es_EC	Language=Spanish, Territory=Ecuador
es_ES	Language=Spanish, Territory=Spain
es_ES_PREEURO	Language=Spanish, Territory=Spain
es_GT	Language=Spanish, Territory=Guatemala
es_HN	Language=Spanish, Territory=Honduras
es_MX	Language=Spanish, Territory=Mexico
es_NI	Language=Spanish, Territory=Nicaragua
es_PA	Language=Spanish, Territory=Panama
es_PE	Language=Spanish, Territory=Peru
es_PR	Language=Spanish, Territory=Puerto Rico
es_PY	Language=Spanish, Territory=Paraguay
es_SV	Language=Spanish, Territory=El Salvador
es_US	Language=Spanish, Territory=United States
es_UY	Language=Spanish, Territory=Uruguay
es_VE	Language=Spanish, Territory=Venezuela
et	Language=Estonian
et_EE	Language=Estonian, Territory=Estonia
eu	Language=Basque
eu_ES	Language=Basque, Territory=Spain
eu_ES_PREEURO	Language=Basque, Territory=Spain
fa	Language=Persian

Locale	Description
fa_IN	Language=Persian, Territory=India
fa_IR	Language=Persian, Territory=Iran
fi	Language=Finnish
fi_FI	Language=Finnish, Territory=Finland
fi_FI_PREEURO	Language=Finnish, Territory=Finland
fo	Language=Faeroese
fo_FO	Language=Faeroese, Territory=Faeroe Islands
fr	Language=French
fr_BE	Language=French, Territory=Belgium
fr_BE_PREEURO	Language=French, Territory=Belgium
fr_CA	Language=French, Territory=Canada
fr_CH	Language=French, Territory=Switzerland
fr_FR	Language=French, Territory=France
fr_FR_PREEURO	Language=French, Territory=France
fr_LU	Language=French, Territory=Luxembourg
fr_LU_PREEURO	Language=French, Territory=Luxembourg
ga	Language=Irish
ga_IE	Language=Irish, Territory=Ireland
ga_IE_PREEURO	Language=Irish, Territory=Ireland
gl	Language=Gallegan
gl_ES	Language=Gallegan, Territory=Spain
gl_ES_PREEURO	Language=Gallegan, Territory=Spain
gu	Language=Gujarati
gu_IN	Language=Gujarati, Territory=India
gv	Language=Manx
gv_GB	Language=Manx, Territory=Great Britain
he_	Language=Hebrew
he_IL	Language=Hebrew, Territory=Israel
hi	Language=Hindi
hi_DIRECT	Language=Hindi
hi_IN	Language=Hindi, Territory=India
hr	Language=Croatian
hr_HR	Language=Croatian, Territory=Croatia
hu	Language=Hungarian
hu_HU	Language=Hungarian, Territory=Hungary
hy	Language=Armenian
hy_AM	Language=Armenian, Territory=Armenia
hy_AM_REVISED	Language=Armenian, Territory=Armenia
id	Language=Indonesian
id_ID	Language=Indonesian, Territory=Indonesia
is	Language=Icelandic

Locale	Description
is_IS	Language=Icelandic, Territory=Iceland
it	Language=Italian
it_CH	Language=Italian, Territory=Switzerland
it_IT	Language=Italian, Territory=Italy
it_IT_PREEURO	Language=Italian, Territory=Italy
ja	Language=Japanese
ja_JP	Language=Japanese, Territory=Japan
kl	Language=Kalaallisut
kl_GL	Language=Kalaallisut, Territory=Greenland
kn	Language=Kannada
kn_IN	Language=Kannada, Territory=India
ko	Language=Korean
ko_KR	Language=Korean, Territory=South Korea
kok	Language=Konkani
kok_IN	Language=Konkani, Territory=India
kw	Language=Cornish
kw_GB	Language=Cornish, Territory=Great Britain
lt	Language=Lithuanian
lt_LT	Language=Lithuanian, Territory=Lithuania
lv	Language=Latvian
lv_LV	Language=Latvian, Territory=Latvia
mk	Language=Macedonian
mk_MK	Language=Macedonian, Territory=Macedonia
mr	Language=Marathi
mr_IN	Language=Marathi, Territory=India
mt	Language=Maltese
mt_MT	Language=Maltese, Territory=Malta
nb	Language=Norwegian Bokmå
nb_NO	Language=Norwegian Bokmå, Territory=Norway
nl	Language=Dutch
nl_BE	Language=Dutch, Territory=Belgium
nl_BE_PREEURO	Language=Dutch, Territory=Belgium
nl_NL	Language=Dutch, Territory=Netherlands
nl_NL_PREEURO	Language=Dutch, Territory=Netherlands
nn	Language=Norwegian Nynorsk
nn_NO	Language=Norwegian Nynorsk, Territory=Norway
om	Language=Oromo
om_ET	Language=Oromo, Territory=Ethiopia
om_KE	Language=Oromo, Territory=Kenya

Locale	Description
pl	Language=Polish
pl_PL	Language=Polish, Territory=Poland
pt	Language=Portugese
pt_BR	Language=Portugese, Territory=Brazil
pt_PT	Language=Portugese, Territory=Portugal
pt_PT_PREEURO	Language=Portugese, Territory=Portugal
ro	Language=Romanian, Territory=
ro_RO	Language=Romanian, Territory=Romania
ru	Language=Russian
ru_RU	Language=Russian, Territory=Russia
ru_UA	Language=Russian, Territory=Ukraine
sh	Language=Serbo-Croatian
sh_YU	Language=Serbo-Croatian, Territory=Yugoslavia
sk	Language=Slovak
sk_SK	Language=Slovak, Territory=Slovakia
sl	Language=Slovenian
sl_SI	Language=Slovenian, Territory=Slovenia
so	Language=Somali
so_DJ	Language=Somali, Territory=Djibouti
so_ET	Language=Somali, Territory=Ethiopia
so_KE	Language=Somali, Territory=Kenya
so_SO	Language=Somali, Territory=Somalia
sq	Language=Albanian
sq_AL	Language=Albanian, Territory=Albania
sr	Language=Serbian
sr_YU	Language=Serbian, Territory=Yugoslavia
sv	Language=Swedish, Territory=
sv_FI	Language=Swedish, Territory=Finland
sv_SE	Language=Swedish, Territory=Sweden
sw	Language=Swahili
sw_KE	Language=Swahili, Territory=Kenya
sw_TZ	Language=Swahili, Territory=Tanzania
ta	Language=Tamil
ta_IN	Language=Tamil, Territory=India
te	Language=Telugu
te_IN	Language=Telugu, Territory=India
th	Language=Thai
th_TH	Language=Thai, Territory=Thailand
ti	Language=Tigrinya
ti_ER	Language=Tigrinya, Territory=Eritrea

Locale	Description
ti_ET	Language=Tigrinya, Territory=Ethiopia
tr	Language=Turkish
tr_TR	Language=Turkish, Territory=Turkey
uk	Language=Ukrainian
uk_UA	Language=Ukrainian, Territory=Ukraine
vi	Language=Vietnamese
vi_VN	Language=Vietnamese, Territory=Vietnam
zh	Language=Chinese
zh_PINYIN	Language=Chinese
zh_CN	Language=Chinese, Territory=China
zh_HK	Language=Chinese, Territory=Hong Kong S.A.R of China
zh_MO	Language=Chinese, Territory=Macau S.A.R of China
zh_SG	Language=Chinese, Territory=Singapore
zh_TW	Language=Chinese, Territory=Taiwan
zh_TW_STROKE	Language=Chinese, Territory=Taiwan

Chapter 5. NLS and Server Jobs - Supplementary Information

This Appendix gives supplementary information about NLS and server jobs.

The NLS Administration Tool

Use the NLS Administration tool to perform a variety of tasks within the National Language Support feature of IBM InfoSphere DataStage.

This section gives a complete description of the NLS Administration tool menus.

You must be a DataStage Administrator in the engine account (UV) to use the menus. To display the main NLS Administration menu, use the NLS.ADMIN command. The NLS Administration menu has the following options:

* **Unicode**. This option lets you examine the Unicode character set using various search criteria.
* **Mappings**. This option lets you view, create, or modify map descriptions or map tables.
* **Locales**. This option lets you view, create, or modify locale definitions.
* **Categories**. This option lets you view, create, or modify category files and weight tables.
* **Installation**. This option lets you install maps into shared memory or edit the `uvconfig` file.

The options lead to further menus that are described in the following sections.

Unicode Menu

Use the Unicode menu to examine the Unicode character set.

The following options are available:

* **Characters**. This option leads to a further menu containing the following options:
  - **List All descriptions**. Provides a very long listing of all the Unicode characters.
  - **by Value**. Prompts you to enter a Unicode 4-digit hexadecimal value, then returns its description.
  - **by Char description**. Prompts you to enter a partial description of a character, then returns possible matches.
  - **by block Number**. Lists all characters in a given Unicode block in Unicode order.
  - **by Block descriptions**. Lists the Unicode block numbers, the official description of what each block contains, the start and end points in the Unicode set, and the number of characters in the block.
  - **Ideograph xref**. The start of further levels of menu, which are of interest to multibyte users only. These let you do the following:

    Display a listing of how the Unicode ideographic area maps to Chinese, Japanese, and Korean standards

    Search for a character in Unicode, given its external character set reference number

Convert between external encodings and standard reference numbers, for example, convert shift-JIS to row and column format
  – **Mnemonic search**. Looks up entries in the MNEMONICS input map by description.
- **Alphabetics**. This option lists the NLS.CS.ALPHAS file. This file contains records that define ranges of code points within which characters are considered to be alphabetic. Use the Ctype category to modify these ranges.
- **Digits**. This option lists the NLS.CS.TYPES file. This file contains records that describe code points normally considered to represent the digits 0 through 9 in different scripts. Use the Numeric category to modify these ranges.
- **Non-printing**. This option lists the NLS.CS.TYPES file. This file contains records that describe code points normally considered to be nonprinting characters. Use the Ctype category to modify these ranges.
- **case Rules**. This option lists the NLS.CS.CASES file. This file describes the normal rules for converting uppercase to lowercase and lowercase to uppercase for all code points in Unicode. Use the Ctype category to modify these ranges.
- **Exit**.

Mappings Menu

Use the Mappings menu to examine, create, and edit map description and map table records, and to compile maps.

The following options are available:
- **View**. Displays a listing of all map description records.
- **Descriptions**. Leads to a submenu for manipulating map descriptions, that is, records in the NLS.MAP.DESCS file. The **Xref** option produces a cross-reference listing that lets you see which maps and tables are being used as the basis for others.
- **Tables**. Leads to a submenu for manipulating map tables, that is, records in the NLS.MAP.TABLES file. From the submenu you can list, create, edit, delete, and cross-reference map tables.
- **Clients**. Administers the NLS.CLIENT.MAPS file, which provides synonyms between map names on a client and the InfoSphere DataStage NLS maps on the server. You can list, create, edit, and delete records using this option.
- **Build**. Compiles a single map.

Locales Menu

Use the Locales menu to examine, create, and edit locale definitions.

The following options are available:
- **List All**. Lists all the locales that are available in InfoSphere DataStage, that is, all the records in the NLS.LC.ALL file. You might need to build the locales in order to install them into shared memory.
- **View**. Prompts you for the name of a locale, then lists the record for that locale.
- **Create**. Creates a new locale record.
- **Edit**. Edits an existing locale record.
- **Delete**. Deletes a locale record
- **Xref**. Cross-references a locale. This lets you see the relationship between various locale definitions.

- **Clients**. Administers the NLS.CLIENT.LCS file, which provides synonyms between locale names on a client, and the InfoSphere DataStage NLS locales on the server. You can list, create, edit, and delete records using this option.
- **Report**. Lets you produce a report on records in locale categories. You can choose from All, Time/date, Numeric, Monetary, Ctype, and Collate.
- **Build**. Builds a locale.

Categories Menu

From the Categories menu you can administer the NLS category files for different types of convention.

The following options are available:
- **Time/date**
- **Numeric**
- **Monetary**
- **Ctype**
- **Collate**
- **Weight tables**
- **Language info**

The first five options call submenus that let you list, view, create, edit, delete, and cross-reference records in the specific category. The final two options have differences as described below.
- **Weight tables**. This option has two additional suboptions as follows:
  - **Accent weights**. This option lists all the records in the NLS.WT.LOOKUP file that refer to accents.
  - **Case weights**. This option lists all the records in the NLS.WT.LOOKUP file that refer to casing.
- **Language info**. This option administers the NLS.LANG.INFO file and lets you list, view, create, edit, delete, and cross-reference records in the file.

Installation Menu

Use the Installation menu to edit the system configuration file or to install maps in shared memory.

The following options are available:
- **Edit uvconfig**. This option lets you edit the configurable parameters in the uvconfig file. You can edit all the parameters, or just those referring to NLS, maps, locales, or clients.
- **Maps**. This option leads to a further menu with the following options:
  - **Configure**. Runs the NLS map configuration program.
  - **All binaries**. Lists all the built maps that are available to be installed into shared memory.
  - **In memory**. Lists the names of all maps currently installed in shared memory and available for use within InfoSphere DataStage.
  - **(re-)Build**. Compiles a single map in the same way as the **Build** option on the **Mappings** menu.
  - **Delete binary**. Removes a binary map. This takes effect when InfoSphere DataStage is restarted.
- **Locales**. This option leads to a further menu with the following options:

- **Configure**. Runs the NLS locale configuration program.
- **All binaries**. Lists all the built locales that are available to be installed into shared memory.
- **In memory**. Lists the names of all locales currently installed in shared memory and available for use within DataStage. Use this option if the SET.LOCALE command fails with the error `locale not loaded`. This option lets you identify locales that are built but not loaded.
- **(re-)Build**. Compiles a single locale.
- **Delete binary**. Removes a binary locale. This takes effect when InfoSphere DataStage is restarted.
- **By language**. This option lets you configure NLS by specifying a particular language. The configuration program selects the appropriate locales and maps to be built and an appropriate configuration for the *uvconfig* file.

The NLS Database

The NLS database contains files that enable the operation of the National Language Support feature of InfoSphere DataStage.

This section describes the files in the NLS database. You can use the NLS.ADMIN command to perform all NLS administration, but you can list and edit these tables directly if you are familiar with TCL.

The NLS database is in the `nls` subdirectory of the server engine directory. The `nls` directory contains the subdirectories `charset`, `locales`, and `maps`.

Each subdirectory of the NLS directory contains further subdirectories, such as the `listing` and `install` subdirectories. **listing** contains listing information generated when building maps and locales (if the user selects this option). `install` contains the binary files that are loaded into memory.

The VOC names for NLS files start with the prefix NLS (this prefix is absent if you view the files from the operating system). The second part of the filename indicates the logical group that the file belongs to. The logical groups are as follows:

These letters...	Indicate this file group...
CLIENT	Data received from client programs
CS	Information about Unicode character sets
LANG	Languages
LC	Locales
MAP	Character set maps
WT	Weight tables

The third part of the filename indicates the contents of the file. For example, the file called NLS.LC.COLLATE is an NLS file belonging to the locales group that contains information about collating sequences.

Table 2 on page 43 lists all the files in the NLS database.

Table 2. NLS Database Files

File	Description
NLS.CLIENT.LCS	Defines the locales to be used by client programs connecting to InfoSphere DataStage.
NLS.CLIENT.MAPS	Defines the character set used by client programs.
NLS.CS.ALPHAS	Defines which characters are defined as alphabetic in the Unicode standard. Each record ID is a hexadecimal code point value that indicates the start of a range of characters. The record itself specifies the last character in the range. These default values can be overridden by a national convention. You should not modify this file; it is for information only.
NLS.CS.BLOCKS	Defines the blocks of consecutive code point values for characters that are normally used together as a set for one or more languages. The record IDs are block numbers. This file is cross-referenced by the NLS.CS.DESCS file. You should not modify this file; it is for information only.
NLS.CS.CASES	Defines those characters that have an uppercase and lowercase version, and how they map between the two, according to the Unicode standard. These default values can be overridden by a national convention. Each record ID is the hexadecimal code point value for a character. You should not modify this file; it is for information only.
NLS.CS.DESCS	Contains descriptions of every character supported by InfoSphere DataStage NLS. Each character has its own record, using its hexadecimal code point value as the record ID. The descriptions are based on those used by the Unicode standard. You should not modify this file; it is for information only.
NLS.CS.TYPES	Defines which characters are numbers, nonprintable characters, and so on, according to the Unicode standard. These default values can be overridden by a national convention. Each record ID is the hexadecimal code point value for a character. You should not modify this file; it is for information only.
NLS.LANG.INFO	Contains information about languages. Provides possible mappings between language, locale and character set map. It is used for installing NLS and reporting on locales, and should not be modified.

Table 2. NLS Database Files (continued)

File	Description
NLS.LC.ALL	Holds records for all the locales known to InfoSphere DataStage. The record IDs are the locale names. The fields of each record are the IDs of records in other locale files. These files contain data about the categories that make up a locale (Time, Numeric, and so on). For a description of the record format for this file, see "Creating New Locales".
NLS.LC.COLLATE	Each record in this file defines a collating sequence used by a locale. The collating sequences are defined according to how they differ from the default collating sequence. For a description of the record format for this file, see "Format of Convention Records".
NLS.LC.CTYPE	Each record in this file holds character typing information used in a locale, that is, which characters are alphabetic, numeric, lowercase, uppercase, nonprinting, and so on. The character types are defined according to how they differ from the default character typing. For a description of the record format for this file, see "Format of Convention Records" .
NLS.LC.MONETARY	Each record in this file holds the monetary formatting convention used in a locale. For a description of the record format for this file, see "Format of Convention Records".
NLS.LC.NUMERIC	Each record in this file holds the numeric formatting convention used in a locale. For a description of the record format for this file, see "Format of Convention Records".
NLS.LC.TIME	Each record in this file holds the time and date formatting convention for a locale. For a description of the record format for this file, see "Format of Convention Records".
NLS.MAP.DESCS	Contains descriptions of every map known to InfoSphere DataStage. The record ID of each map is the map name used in InfoSphere DataStage commands or BASIC programs. The record IDs must comprise ASCII-7 characters only. For a description of the record format for this file, see "Creating a New Map".
NLS.MAP.TABLES	A type 19 file that contains the map tables for mapping an external character set to the DataStage internal character set. For more information about the structure of this file, see "Creating a New Map".
NLS.WT.LOOKUP	Contains weightings given to characters during a sort, based on the Unicode standard. This file should not be modified.

Table 2. NLS Database Files (continued)

File	Description
NLS.WT.TABLES	Contains specific weight information about characters used in a locale. For more information about the structure of this file, see "Editing Weight Tables".

Format of Convention Records

Locales are organized in categories which are in turn made up of a set of conventions.

The following sections describe the fields in convention records in the five categories:

- Time
- Numeric
- Monetary
- Ctype
- Collate

Time Records

Time records are one type of convention record used in InfoSphere DataStage.

The following table shows each field number, its display name, and a description for time and date information:

Field	Name	Description
0	Category Name	The name of the convention.
1	Description	A description of the convention. It usually includes the territory that the convention applies to and the language it is used with.
2	Based on	The name of another convention record that this convention is based on.
3	TIMEDATE format	A format for combined time and date used by the BASIC TIMEDATE function and the TIME command. The value should consist of an MT or TI time conversion code, and a D or DI date conversion code. The two codes can be in any order. They should be separated by a tab character, or a text or subvalue mark.
4	Full DATE format	The full combined date and time format used by the TIME command. The value should consist of an MT or TI time conversion code, and a D or DI date conversion code. The two codes can be in any order. They should be separated by a tab character, or a text or subvalue mark.
5	Date `D' format	The default date format for the D conversion code. The value should be any D or DI conversion code.

Field	Name	Description
6	Date `DI' format	The default date format for the DI conversion code. The value should be a D conversion code. The order is specified by the DMY order (field 23). The separator is specified by the date separator (field 24).
7	Time `MT' format	The default time format for the MT conversion code. The value should be an MT conversion code. In most cases, use the value TI.
8	Time `TI' format	The format for the TI conversion code. The value should be an MT conversion code that specifies separators. The default separator is a colon (:) as specified by the time separator (field 25).
9	Days of the week	A multivalued list of the full names of the days of the week. For example, Monday, Tuesday. Fields 9 and 10 are associated multi-valued fields; the same number of values must exist in each field.
10	Abbreviated	A multivalued list of abbreviated names of the days of the week. For example, Mon, Tue. See field 9.
11	Month names	A multivalued list of the full names of the months of the year. For example, January, February. Fields 11 and 12 are associated multivalued fields; the same number of values must exist in each field.
12	Abbreviated	A multivalued list of abbreviated names of the months of the year. For example, Jan, Feb. See field 11.
13	Chinese years	A multivalued list of Chinese year names (Monkey to Sheep).
14	AM string	A string used to denote times before noon in 12-hour formats.
15	PM string	A string used to denote times after noon in 12-hour formats.
16	BC string	A string to be added to dates before the date 01 Jan 0001 in the Gregorian calendar. This corresponds to -718432, the DataStage internal date.
17	Era name	A multivalued list of names of eras and their start dates, beginning with the most recent, for example, Japanese Imperial Era Heisei. This field can be used for any locale that uses a calendar with several year zeros. For example, the Thai Buddhist Era commencing 1/1/543 BC. See "Defining Era Names" on page 47.
18	Start date	Corresponding era start dates for the era names specified in InfoSphere DataStage internal date format.
19	HEADING/FOOTING D format	A D or DI conversion code used in HEADING and FOOTING statements.

Field	Name	Description
20	HEADING/FOOTING T format	An MT or TI conversion code used in HEADING and FOOTING statements.
21	Gregorian calendar day 1	The date at which the calendar changes from Julian to Gregorian, expressed as a InfoSphere DataStage internal date. The default is -140607, corresponding to 11 January 1583.
22	Number of days skipped	The number of days to skip when the calendar changes from Julian to Gregorian. The default is 10.
23	Default DMY order	The order of day, month, and year, for example, DMY.
24	Default date separator	The separator used between day, month, and year. The default is the slash (/).
25	Default time separator	The separator used between hours, minutes, and seconds. The default is the colon (:).

Defining Era Names

The values in the ERA_NAMES field can contain the format code:

Name [*%n*] [*string*]

Name is the era name.

%n is a digit from 1 through 9, or the characters +, -, or Y.

string is any text string.

The *%n* syntax allows era year numbers to be included in the era name and indicates how the era year numbers are to be calculated. If *%n* is omitted, %1 is assumed.

The rules for the *%n* syntax are as follows:
- %1 - %9: The number following the % is the number to be used for the first year *n* of this era. This is effectively an offset which is added to the era year number. This will usually be 1 or 2.
- %+: The era year numbers count backward relative to year numbers; that is, if era year number 1 corresponds to Julian year Y, year 2 corresponds to Y-1, year 3 to
  Y-2, and so on
- %- : The same as for %+, but uses negative era year numbers; that is, first year Y is -1, Y-1 is -2, Y-2 is -3, and so forth.
- %Y: Uses the Julian year numbers for the era year numbers. The year number will be displayed as a 4-digit year number.

The %+, %-, and %Y syntax should only be used in the last era name in the list of era names, that is, the first era, since the list of era names must be in descending date order.

string allows any text string to be appended to the era name. It is frequently the case that the first year or part-year of an era is followed by some qualifying

characters. Therefore, the actual era is divided into two values, each with the same era name, but one terminated by %1string and the other by %2. You must define the era names accordingly.

Example

This example shows the contents of the records named DEFAULT and US-ENGLISH. The US-ENGLISH record is based on the ENGLISH.NAMES record. An empty field specifies that its definition is derived from any category on which it is based. If there is no base category, the default category is used.

```
Time/Date Conventions for Locale DEFAULT

Category name............ DEFAULT
Description.............. System defaults
Based on................
TIMEDATE format.......... MTS
                        . D4
Full DATE format......... D4WAMADY[", ", " ", ", "]
                        . MT
Date 'D' format.......... D4 DMBY
Date 'DI' format......... D2-YMD
Time 'MT' format......... TI
Time 'TI' format......... MTS:
Days of the week.................. Abbreviated.........
Sunday                           Sun
Monday                           Mon
Tuesday                          Tue
Wednesday                        Wed
Thursday                         Thu
Friday                           Fri
Saturday                         Sat
Month names....................... Abbreviated.......
January                          Jan
February                         Feb
March                            Mar
April                            Apr
May                              May
June                             Jun
July                             Jul
August                           Aug
September                        Sep
October                          Oct
November                         Nov
December                         Dec
Chinese years............ MONKEY
                        . COCK
                        . DOG
                        . BOAR
                        . RAT
                        . OX
                        . TIGER
                        . RABBIT
                        . DRAGON
                        . SNAKE
                        . HORSE
                        . SHEEP
AM string................ am
PM string................ pm
BC string............... BC
Era name............................. Start date....
Heisi                                08 JAN 1989
Showa                                25 DEC 1926
Taisho                               30 JUL 1912
Meiji                                08 SEP 1868
HEADING/FOOTING D format. D2-
```

```
HEADING/FOOTING T format. MTS
                        . D2-
Gregorian calendar day 1.        11 JAN 1583
Number of days skipped...        10
Default DMY order........
Default date separator...
Default time separator...
Time/Date Conventions for US-ENGLISH

Category name...........    US-ENGLISH
Description.............    Territory=USA,Language=English
Based on................    .ENGLISH.NAMES
TIMEDATE format.........
Full DATE format........
Date 'D' format.........
Date 'DI' format........    D2/MDY
Time 'MT' format........
Time 'TI' format........    MTHS:
Days of the week............Abbreviated.........

Month names.................... Abbreviated........

Chinese years...........
AM string...............
PM string...............
BC string...............
Era name............................ Start date....

HEADING/FOOTING D format.
HEADING/FOOTING T format.
Gregorian calendar day 1.
Number of days skipped...
Default DMY order........ MDY
Default date separator...
Default time separator...
```

Numeric Records

Numeric records are one type of convention record used by National Language Support.

The following table shows each field number, its display name, and a description:

Field	Name	Description
0	Category Name	The name of the convention.
1	Description	A description of the convention. It usually includes the territory that the convention applies to and the language it is used with.
2	Based on	The name of another convention record that this convention is based on.
3	Decimal separator	The character used as a decimal separator (radix character). The value can be expressed as either a single character or the hexadecimal Unicode value of a character.

Field	Name	Description
4	Thousands separator	The character used as a thousands separator. The value can be expressed as either a single character or the hexadecimal Unicode value of a character. Use the value NONE to indicate that no separator is needed.
5	Suppress leading zero	Defines whether leading zeros should be suppressed for numbers in the range 1 through -1. A value of 0 or N means insert a zero; any other value suppresses the zero.
6	Alternative digits (0 first)	A multivalued field containing 10 values that can be used as alternatives to the corresponding ASCII digits 0 through 9.

This example shows the contents of the records named DEFAULT and DEC.COMMA+DOT locale (used by DE-GERMAN) in the NLS.LC.NUMERIC file. The DEC.COMMA+DOT conventions are based on DEFAULT.

```
Numeric Conventions for DEFAULT

Category name.....         DEFAULT
Description.......         System defaults: Decimal separator =
                           dot, thousands = comma
Based on..............
Decimal separator..... .   -  FULL STOP
Thousands separator... ,   -  COMMA
Suppress leading zero. 0
Alternative digits (0 first).
Numeric Conventions for DEC.COMMA+DOT

Category name......        DEC.COMMA+DOT
Description........        Decimal separator = comma,
                           thousands = dot
Based on.............. DEFAULT
Decimal separator..... ,      -  COMMA
Thousands separator... .      -  FULL STOP
Suppress leading zero.
Alternative digits (0 first).
```

Ctype Records

Ctype records are one type of convention record used in National Language Support.

The following table shows each field number, its display name, and a description for fields in the Ctype record. Many of the defaults are based directly on Unicode settings. These can be viewed by choosing the appropriate item from the Unicode menu in the NLS Administration tool.

For fields 3 onward, you can enter the values as characters or as Unicode values. You can specify a range of values separated by a dash (-)

Field	Name	Description
0	Category Name	The name of the convention.

Field	Name	Description
1	Description	A description of the convention. It usually includes the territory that the convention applies to and the language it is used with.
2	Based on	The name of another convention record that this convention is based on.
3	Lowercase	A multivalued list of lowercase values whose associated uppercase values differ from the Unicode defaults.
4	->Upper	A multivalued list of the equivalent uppercase values for the characters listed in field 3.
5	Uppercase	A multivalued list of uppercase values whose associated lowercase values differ from the Unicode defaults.
6	->Lower	A mutivalued list of the equivalent lowercase values for the characters listed in field 5.
7	Alphabetics	A multivalued list of characters that are alphabetic but are not described as such under the Unicode defaults. You can specify this value as a Unicode block value using the format BLOCK=*nn*, where *nn* is the Unicode block number.
8	Non-Alphabetics	A multivalued list of characters that are not alphabetic but are described as such under the Unicode defaults. You can specify this value as a Unicode block value using the format BLOCK=*nn*, where *nn* is the Unicode block number.
9	Numerics	A multivalued list of characters that should be considered as numeric but are not described as such under the Unicode defaults.
10	Non-Numerics	A multivalued list of characters that are not considered to be numeric but are described as such under the Unicode defaults.
11	Printables	A multivalued list of characters that are considered to be printable but are not described as such under the Unicode defaults.
12	Non-Printables	A multivalued list of characters that are not considered to be printable but are described as such under the Unicode defaults.
13	Trimmables	A multivalued list of characters that are to be removed by TRIM functions in addition to spaces and tab characters.

.

In Spanish, accented characters other than ñ drop their accents when converted to uppercase. In French, all accented characters drop their accents in uppercase.

This example shows a convention called NOACCENT.UPCASE (based on DEFAULT), which the locale FR-FRENCH uses, and a convention called SPANISH, that is based on it.

In this example, the only characters affected are those in general use in French and Spanish. There are many other accented characters in Unicode. This example displays <N?> that comes from the MNEMONICS map. This lets you easily enter non-ASCII characters rather than their Unicode values.

```
Character Type Conventions for ACCENTLESS.UPPERCASE

Category name. NOACCENT.UPCASE
Description... ISO8859-1 lowercase accented chars lose
               accents in uppercase
Based on...... DEFAULT
Lowercase............................. -> Uppercase...........................
00E0 - LATIN SMALL LETTER A WITH GRAVE  0041 - LATIN CAPITAL LETTER A
00E1 - LATIN SMALL LETTER A WITH ACUTE  0041 - LATIN CAPITAL LETTER A
00E2 - LATIN SMALL LETTER A WITH        0041 - LATIN CAPITAL LETTER A
       CIRCUMFLEX
00E3 - LATIN SMALL LETTER A WITH TILDE  0041 - LATIN CAPITAL LETTER A
00E4 - LATIN SMALL LETTER A WITH        0041 - LATIN CAPITAL LETTER A
       DIAERESIS
00E5 - LATIN SMALL LETTER A WITH RING   0041 - LATIN CAPITAL LETTER A
       ABOVE
00E7 - LATIN SMALL LETTER C WITH        0043 - LATIN CAPITAL LETTER C
       CEDILLA
00E8 - LATIN SMALL LETTER E WITH GRAVE  0045 - LATIN CAPITAL LETTER E
00E9 - LATIN SMALL LETTER E WITH ACUTE  0045 - LATIN CAPITAL LETTER E
00EA - LATIN SMALL LETTER E WITH        0045 - LATIN CAPITAL LETTER E
       CIRCUMFLEX
00EB - LATIN SMALL LETTER E WITH        0045 - LATIN CAPITAL LETTER E
       DIAERESIS
00EC - LATIN SMALL LETTER I WITH GRAVE  0049 - LATIN CAPITAL LETTER I
00ED - LATIN SMALL LETTER I WITH ACUTE  0049 - LATIN CAPITAL LETTER I
00EE - LATIN SMALL LETTER I WITH        0049 - LATIN CAPITAL LETTER I
       CIRCUMFLEX
00EF - LATIN SMALL LETTER I WITH        0049 - LATIN CAPITAL LETTER I
       DIAERESIS
00F1 - LATIN SMALL LETTER N WITH TILDE  004E - LATIN CAPITAL LETTER N
00F2 - LATIN SMALL LETTER O WITH GRAVE  004F - LATIN CAPITAL LETTER O
00F3 - LATIN SMALL LETTER O WITH ACUTE  004F - LATIN CAPITAL LETTER O
00F4 - LATIN SMALL LETTER O WITH        004F - LATIN CAPITAL LETTER O
       CIRCUMFLEX
00F5 - LATIN SMALL LETTER O WITH TILDE  004F - LATIN CAPITAL LETTER O
00F6 - LATIN SMALL LETTER O WITH        004F - LATIN CAPITAL LETTER O
       DIAERESIS
00F8 - LATIN SMALL LETTER O WITH STROKE 004F - LATIN CAPITAL LETTER O
00F9 - LATIN SMALL LETTER U WITH GRAVE  0055 - LATIN CAPITAL LETTER U
00FA - LATIN SMALL LETTER U WITH ACUTE  0055 - LATIN CAPITAL LETTER U
00FB - LATIN SMALL LETTER U WITH        0055 - LATIN CAPITAL LETTER U
       CIRCUMFLEX
00FC - LATIN SMALL LETTER U WITH        0055 - LATIN CAPITAL LETTER U
       DIAERESIS
00FD - LATIN SMALL LETTER Y WITH ACUTE  0059 - LATIN CAPITAL LETTER Y
00FF - LATIN SMALL LETTER Y WITH        0059 - LATIN CAPITAL LETTER Y
       DIAERESIS
Uppercase............................. -> Lowercase...............

Alphabetics.....
Non-Alphabetics.
Numerics........
Non-Numerics....
Printables......
Non-Printables..
Trimmables......
```

```
Character Type Conventions for SPANISH

Category name. SPANISH
Description... Language=Spanish - SMALL N WITH TILDE
               keeps tilde on uppercasing
Based on...... NOACCENT.UPCASE
Lowercase.......................... -> Uppercase..........................
<n?> - LATIN SMALL LETTER N WITH TILDE  <N?> - LATIN CAPITAL LETTER N WITH
                                               TILDE
Uppercase.......................... -> Lowercase..........................

Alphabetics.....
Non-Alphabetics.
Numerics........
Non-Numerics....
Printables......
Non-Printables..
Trimmables......
```

Collate Records

Collate records are one type of convention record used in National Language Support.

The following table shows each field number, its display name, and a description for Collate category records. Many of the fields are Boolean. An empty field or a value of 0 or N indicates false; any other value indicates true.

Field	Name	Description
0	Category Name	The name of the convention.
1	Description	A description of the convention. It usually includes the territory that the convention applies to and the language it is used with.
2	Based on	The name of another convention record that this convention is based on.
3	Accented Sort?	This field determines how accents on characters affect the collate order. A false value indicates that accents are not collated separately. A true value indicates that accents are used as tie breakers in the sort. See "Collating".
4	In reverse?	If field 3 indicates an accented collation, this field determines the direction of that collation. A false value indicates forward collation. A true value indicates reverse collation.
5	Cased Sort?	This field determines whether the case of a character is considered during collation. A false value indicates that case is not considered. A true value indicates that case is used as a tie breaker in the collation.
6	Lowercase first?	If field 5 indicates a cased collation, this field determines which case is collated first. A false value indicates that lowercase is collated first. A true value indicates that uppercase is collated first.

Field	Name	Description
7	Expand	A multivalued field containing Unicode values of characters that are expanded before collation. See "Contractions and Expansions".
8	Expanded	A multivalued field associated with field 7 that supplies the values the characters expand to. Each value might be one or more Unicode values separated by tab characters or spaces. To override an expansion inherited from a based convention named in field 2, enter the same multivalue in fields 7 and 8. (For another method, see the description of field 10.)
9	Before?	A multivalued field associated with fields 7 and 8 that determines how expanded characters collate. A false value indicates that a character is collated after expansion; a true value indicates that a character is collated before expansion.
10	Contract	A multivalued field containing a list of pairs of Unicode values of characters after contraction. The values should be separated by tab characters or spaces. To override an expansion inherited from a based convention named in field 2, enter a value in this field and a corresponding empty value in field 11. See "Contractions and Expansions".
11	Before	A multivalued field associated with field 10. It gives the Unicode value of the character that a contracted pair precedes in the collation order.
12	Weight Tables	A multivalued field supplying the weight information for characters in this locale. The values should be record IDs in the NLS.WT.TABLES file. The default is the name of the locale. The weight information is processed in the order supplied in this field.

This example shows the Collate records named DEFAULT, GERMAN, and SPANISH:

- DEFAULT uses no expansion or contraction, but does collate in a sequence other than the Unicode value.
- GERMAN uses the DEFAULT collating sequence, but introduces an expansion.
- SPANISH is also based on DEFAULT, but introduces eight contractions.

```
Collating Sequence Conventions for DEFAULT

Category name.... DEFAULT
Description...... System defaults
Based on........
Accented Sort?... N
In reverse?...... N
Cased Sort?...... N
Lowercase first?. N
Expand ------------------->..... Before? Expanded.. ........................
```

```
Contract... ---------------------->..... Before ..............................

Weight Tables.... LATIN1-DEFAULT
                . LATINX-DEFAULT
                . LATINX2-DEFAULT
                . LATINX3-DEFAULT
                . GREEK-DEFAULT
                . CYRILLIC-DEFAULT


Collating Sequence Conventions for GERMAN

Category name.... GERMAN
Description...... Language=German
Based on......... DEFAULT
Accented Sort?... Y
In reverse?...... N
Cased Sort?...... Y
Lowercase first?. N
Expand ------------------->..... Before? Expanded.. .........................
<ss>   LATIN SMALL LETTER SHARP S      N S S      LATIN CAPITAL LETTER S
                                                  LATIN CAPITAL LETTER S
Contract... ---------------------->..... Before ..............................

Weight Tables....


Collating Sequence Conventions for SPANISH

Category name.... SPANISH
Description...... Language=Spanish
Based on......... DEFAULT
Accented Sort?... Y
In reverse?...... N
Cased Sort?...... Y
Lowercase first?. N
Expand ------------------->..... Before? Expanded.. .........................

Contract... ---------------------->..... Before ..............................
C H        LATIN CAPITAL LETTER C       D     LATIN CAPITAL LETTER D
           LATIN CAPITAL LETTER H
C h        LATIN CAPITAL LETTER C       D     LATIN CAPITAL LETTER D
c h        LATIN SMALL LETTER C         d     LATIN SMALL LETTER D
           LATIN SMALL LETTER H
c H        LATIN SMALL LETTER C         d     LATIN SMALL LETTER D
           LATIN CAPITAL LETTER H
L L        LATIN CAPITAL LETTER L       M     LATIN CAPITAL LETTER M
           LATIN CAPITAL LETTER L
L l        LATIN CAPITAL LETTER L       M     LATIN CAPITAL LETTER M
           LATIN SMALL LETTER L
l l        LATIN SMALL LETTER L         m     LATIN SMALL LETTER M
           LATIN SMALL LETTER L
l L        LATIN SMALL LETTER L         m     LATIN SMALL LETTER M
           LATIN CAPITAL LETTER L
Weight Tables.... LATIN-SPANISH
```

Collating

A number of issues regarding collating must be considered when you use
InfoSphere DataStage with National Language Support (NLS) mode enabled.

Collating is a complex issue for many languages. It is not sufficient to collate a
character set in numerical order of its Unicode values. Locales that share a

character set often have different collating rules. For example, these are the main issues that affect collating in Western European languages:

- Accented characters. Should accented characters come before or after their unaccented equivalents? Or should accents only be examined if two strings being compared would otherwise be identical (that is, as a tie breaker)?
- Expanding characters. Some languages treat certain single characters as two separate characters for collating purposes.
- Contracting characters. Some languages have pairs of characters that collate as though they were a single character.
- Should case be considered? Should case be used as a tie breaker for otherwise identical strings? If so, which comes first, uppercase or lowercase?
- Should hyphens or other punctuation be considered as tie breakers?

How InfoSphere DataStage Collates

To overcome collating problems, InfoSphere DataStage allows each Unicode character to be assigned up to three weights. The weight is a numeric value to use instead of the character during collation.

The three Unicode characters weights are as follows:

Weight Type	Meaning
Shared weight	All characters that are essentially the same have the same shared weight, even though they might differ in accent or case.
Accent weight	This weight shows the order of precedence for accented characters. The Collate convention determines the direction of the collation.
Case weight	This weight differentiates between uppercase and lowercase characters. The Collate convention determines which case has precedence.

Before collation begins, InfoSphere DataStage expands or contracts any characters as defined in the Collate convention. The collation works as follows:

1. The characters are compared by shared weight.
2. If two characters have the same shared weight, they are compared by accent weight.
3. If the accent weight is the same, they are compared by case weight.

Example of Accented Collation

Words in a language that differ in their accents are collated in different ways depending on the configuration of their weight tables.

This table compares how four French words that differ only in their accents are collated in two different ways, depending on how the weight tables have been configured:

Order	Accented Collation	Unaccented Collation
1	cote	cote
2	côte	coté

Order	Accented Collation	Unaccented Collation
3	coté	côte
4	côté	côté

In the accented collation, the words are in the order they would be found in a French dictionary. (It is actually a reverse accented collation.) Each accented character has the same shared weight as it would have without the accent. The order is decided by referring to the accent weight.

In the unaccented collation, each accented character has a different shared weight unrelated to its unaccented equivalent. The order is decided by the shared weight alone.

Example of Cased Collation

Words in a language are collated in different ways depending on their case.

The three words Aaron, Aardvark, and aardvark show how case affects collation:

Order	Cased Collation	Uncased Collation
1	Aardvark	Aardvark
2	aardvark	Aaron
3	Aaron	aardvark

In the cased collation, Aaron follows aardvark because the characters `A' and `a' have the same shared weight. The case weight is only considered for the two strings that are otherwise identical, that is, Aardvark and aardvark.

In the uncased collation, Aaron precedes aardvark because the characters `A' and `a' have different shared weights.

Shared Weights and Blocks

Shared weights are assigned so that characters collate correctly within each Unicode block.

Unicode is divided into blocks of related characters. For example, Cyrillic characters form one block, while Hebrew characters form another. In most circumstances, it is unlikely that you need to collate characters from more than one block at a time.

Contractions and Expansions

Contractions and expansions are interactions between language characters either collating or acting as separate characters.

Some languages have pairs of characters that collate as though they were a single character. Other languages treat certain single characters as two separate characters for collating. These contractions and expansions are done before InfoSphere DataStage begins a collation.

For example, in Spanish, the character pairs CH and LL (in any combination of case) are treated as a single, separate character. CH comes between C and D in the

collating sequence, and LL comes between L and M. InfoSphere DataStage identifies these character pairs before collation begins. In German, the character ß is expanded to SS before collation begins.

Editing Weight Tables

You can change weight values that are assigned for numeric comparisons in sorting to customize your locale.

Collating character sets in different languages is a complex issue. Each character has an assigned weight value used for numeric comparisons in sorting, but you can change these weight values to sort in a different way when you want to customize your locale.

You can edit the weight table for a locale by choosing **Categories** > **-> Weight Tables** > **Edit** from the NLS Administration menu. Any change you make to the weight assigned to a character overrides the default weight derived from its Unicode value.

The weights are held in the NLS.WT.TABLES file, which is a type 19 file. Each record in the file can contain:
- Comment lines, introduced by a # or *
- A set of weight values for a Unicode code point

Each weight value line has the following fields, separated by at least one ASCII space or tab character:

character [*block.weight* /] *shared.weight accent.weight case.weight* [*comments*]

character is a Unicode character value. This should be four hexadecimal digits, zero-filled as necessary.

The *block.weight* / *shared.weight* value is one or two decimal integers, separated by a slash (/) if necessary. *block.weight* can be 1 through 127; *shared.weight* 1 through 32767. If *block.weight* is omitted, it is taken as the value of the Unicode block number to which *character* belongs. *shared.weight* might be given as a hyphen, in which case it is taken as the value of the most recent weight value line without a hyphen for *shared.weight*. Characters that should sort together if accents and case are disregarded should have the same *block.weight* / *shared.weight* value.

accent.weight is a decimal integer 1 through 63. It might be given as a hyphen, in which case it is taken as the value of the most recent weight value line without a hyphen for *accent.weight*. Characters that are distinguished only by accent should have the same *block.weight* / *shared.weight* value and differ in their *accent.weight* value. A list of conventional values to assign to this field can be found by listing records starting with "AW..." in the NLS.WT.LOOKUP file.

case.weight is a decimal integer 1 through 7, or the letter U or L to indicate uppercase and lowercase. *case.weight* can be given as a hyphen, in which case it is taken as the value of the most recent weight value line without a hyphen for *case.weight*. Characters that are distinguished only by case should have the same *block.weight* / *shared.weight* value and *accent.weight* value and differ only in their *case.weight* value. A list of conventional values to assign to this field can be found by listing records starting with "CW..." in the NLS.WT.LOOKUP file.

comments can contain any characters.

Calculating the Overall Weight

InfoSphere DataStage allows each Unicode character to be assigned up to three weights. The overall weight is calculated from this information.

The overall weight assigned to *character* is calculated using the following formula:

$$(block.weight \times 2^{24}) + (shared.weight \times 2^{9})$$
$$+ (accent.weight \times 2^{3}) + case.weight$$

If *character* is not mentioned in a table, the default weight is calculated as follows:

$$(BW \times 2^{24}) + (SW \times 2^{9})$$

BW is the character's Unicode block number. SW depends on its position within the block: the first character has a SW of 1, the second a SW of 2, and so on.

Example of a Weight Table

InfoSphere DataStage allows each Unicode character to be assigned up to three weights. These weights are included in a weight table for collation.

This example shows a weight table for collating Turkish characters:

```
* Sorting weight table for TURKISH characters (from ISO8859/9)
* in order on top of LATIN1/LATINX tables. These characters are:
*
* Between G and H: G BREVE
* Between H and J: I WITH DOT ABOVE (uppercase version of SMALL I 0069)
*                  DOTLESS I (lowercase version of CAPITAL I 0049)
* (Note: the sequence is H, dotless I, I dot + accented versions, J, ...)
* Between S and T: S CEDILLA
*
* SYNTAX:
* Each non-comment line gives one or more weights for a character,as
* follows (character value in hex, weights in decimal):
* Field 1 = Unicode character value
* Field 2 = Shared weight (characters that sort together if
*           accents and case were to be disregarded should
*           have the same SW)
*       Or, Block Weight/Shared Weight. This form allows
*           characters in different Unicode blocks to have
*           equal SWs. If BW is omitted, only SWs for characters in
*           the same block are equal.
* Field 3 = Accent weight, or '-' to omit or copy from previous.
*           Please use values as defined in the file NLS.WT.LOOKUP.
* Field 4 = Case weight, or 'U' for upper and 'L' for lower case chars.
*
*****************************************************************
* HEX   (BW/)SW        AW      CW
* After G:
011E    4/1092         5       U  * G WITH BREVE
011F    -              5       L
* I, dotted and undotted:
* (Note we do not use AWs here, but use SWs to differentiate
* these characters from the unaccented versions.)
0049    4/1109         -       U  * I
0131    -              -       L  * DOTLESS I
0130    4/1110         -       U  * I WITH DOT ABOVE
0069    -              -       L  * I
* S cedilla
015E    4/1232         40      U  * S WITH CEDILLA
015F    -              40      L
*
* END
```

Appendix A. Product accessibility

You can get information about the accessibility status of IBM products.

The IBM InfoSphere Information Server product modules and user interfaces are not fully accessible.

For information about the accessibility status of IBM products, see the IBM product accessibility information at http://www.ibm.com/able/product_accessibility/index.html.

Accessible documentation

Accessible documentation for InfoSphere Information Server products is provided in an information center. The information center presents the documentation in XHTML 1.0 format, which is viewable in most web browsers. Because the information center uses XHTML, you can set display preferences in your browser. This also allows you to use screen readers and other assistive technologies to access the documentation.

The documentation that is in the information center is also provided in PDF files, which are not fully accessible.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.

Appendix B. Reading command-line syntax

This documentation uses special characters to define the command-line syntax.

The following special characters define the command-line syntax:

[ ]   Identifies an optional argument. Arguments that are not enclosed in brackets are required.

...   Indicates that you can specify multiple values for the previous argument.

| Indicates mutually exclusive information. You can use the argument to the left of the separator or the argument to the right of the separator. You cannot use both arguments in a single use of the command.

{ }   Delimits a set of mutually exclusive arguments when one of the arguments is required. If the arguments are optional, they are enclosed in brackets ([ ]).

Note:
- The maximum number of characters in an argument is 256.
- Enclose argument values that have embedded spaces with either single or double quotation marks.

For example:

wsetsrc*[-S server] [-l label] [-n name] source*

The *source* argument is the only required argument for the **wsetsrc** command. The brackets around the other arguments indicate that these arguments are optional.

wlsac *[-l | -f format] [key...] profile*

In this example, the -l and -f format arguments are mutually exclusive and optional. The *profile* argument is required. The *key* argument is optional. The ellipsis (...) that follows the *key* argument indicates that you can specify multiple key names.

wrb -import *{rule_pack | rule_set}...*

In this example, the rule_pack and rule_set arguments are mutually exclusive, but one of the arguments must be specified. Also, the ellipsis marks (...) indicate that you can specify multiple rule packs or rule sets.

Appendix C. Contacting IBM

You can contact IBM for customer support, software services, product information, and general information. You also can provide feedback to IBM about products and documentation.

The following table lists resources for customer support, software services, training, and product and solutions information.

Table 3. IBM resources

Resource	Description and location
IBM Support Portal	You can customize support information by choosing the products and the topics that interest you at www.ibm.com/support/ entry/portal/Software/ Information_Management/ InfoSphere_Information_Server
Software services	You can find information about software, IT, and business consulting services, on the solutions site at www.ibm.com/ businesssolutions/
My IBM	You can manage links to IBM Web sites and information that meet your specific technical support needs by creating an account on the My IBM site at www.ibm.com/account/
Training and certification	You can learn about technical training and education services designed for individuals, companies, and public organizations to acquire, maintain, and optimize their IT skills at http://www.ibm.com/training
IBM representatives	You can contact an IBM representative to learn about solutions at www.ibm.com/connect/ibm/us/en/

Appendix D. Accessing the product documentation

Documentation is provided in a variety of formats: in the online IBM Knowledge Center, in an optional locally installed information center, and as PDF books. You can access the online or locally installed help directly from the product client interfaces.

IBM Knowledge Center is the best place to find the most up-to-date information for InfoSphere Information Server. IBM Knowledge Center contains help for most of the product interfaces, as well as complete documentation for all the product modules in the suite. You can open IBM Knowledge Center from the installed product or from a web browser.

Accessing IBM Knowledge Center

There are various ways to access the online documentation:
- Click the **Help** link in the upper right of the client interface.
- Press the F1 key. The F1 key typically opens the topic that describes the current context of the client interface.

   Note: The F1 key does not work in web clients.
- Type the address in a web browser, for example, when you are not logged in to the product.

   Enter the following address to access all versions of InfoSphere Information Server documentation:

   `http://www.ibm.com/support/knowledgecenter/SSZJPZ/`

   If you want to access a particular topic, specify the version number with the product identifier, the documentation plug-in name, and the topic path in the URL. For example, the URL for the 11.3 version of this topic is as follows. (The ⇒ symbol indicates a line continuation):

   `http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/⇒`
   `com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html`

   Tip:

   The knowledge center has a short URL as well:

   `http://ibm.biz/knowctr`

   To specify a short URL to a specific product page, version, or topic, use a hash character (#) between the short URL and the product identifier. For example, the short URL to all the InfoSphere Information Server documentation is the following URL:

   `http://ibm.biz/knowctr#SSZJPZ/`

   And, the short URL to the topic above to create a slightly shorter URL is the following URL (The ⇒ symbol indicates a line continuation):

   `http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/⇒`
   `common/accessingiidoc.html`

Changing help links to refer to locally installed documentation

IBM Knowledge Center contains the most up-to-date version of the documentation. However, you can install a local version of the documentation as an information center and configure your help links to point to it. A local information center is useful if your enterprise does not provide access to the internet.

Use the installation instructions that come with the information center installation package to install it on the computer of your choice. After you install and start the information center, you can use the **iisAdmin** command on the services tier computer to change the documentation location that the product F1 and help links refer to. (The ⇒ symbol indicates a line continuation):

Windows
```
IS_install_path\ASBServer\bin\iisAdmin.bat -set -key ⇒
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

AIX Linux
```
IS_install_path/ASBServer/bin/iisAdmin.sh -set -key ⇒
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

Where <host> is the name of the computer where the information center is installed and <port> is the port number for the information center. The default port number is 8888. For example, on a computer named server1.example.com that uses the default port, the URL value would be http://server1.example.com:8888/help/topic/.

Obtaining PDF and hardcopy documentation

- The PDF file books are available online and can be accessed from this support document: https://www.ibm.com/support/docview.wss?uid=swg27008803 &wv=1.
- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss.

Appendix E. Providing feedback on the product documentation

You can provide helpful feedback regarding IBM documentation.

Your feedback helps IBM to provide quality information. You can use any of the following methods to provide comments:

- To provide a comment about a topic in IBM Knowledge Center that is hosted on the IBM website, sign in and add a comment by clicking **Add Comment** button at the bottom of the topic. Comments submitted this way are viewable by the public.
- To send a comment about the topic in IBM Knowledge Center to IBM that is not viewable by anyone else, sign in and click the **Feedback** link at the bottom of IBM Knowledge Center.
- Send your comments by using the online readers' comment form at www.ibm.com/software/awdtools/rcf/.
- Send your comments by e-mail to comments@us.ibm.com. Include the name of the product, the version number of the product, and the name and part number of the information (if applicable). If you are commenting on specific text, include the location of the text (for example, a title, a table number, or a page number).

Notices and trademarks

This information was developed for products and services offered in the U.S.A. This material may be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

Notices

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Privacy policy considerations

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session or persistent cookies. If a product or component is not listed, that product or component does not use cookies.

Table 4. Use of cookies by InfoSphere Information Server products and components

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
Any (part of InfoSphere Information Server installation)	InfoSphere Information Server web console	• Session • Persistent	User name	• Session management • Authentication	Cannot be disabled
Any (part of InfoSphere Information Server installation)	InfoSphere Metadata Asset Manager	• Session • Persistent	No personally identifiable information	• Session management • Authentication • Enhanced user usability • Single sign-on configuration	Cannot be disabled

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
InfoSphere DataStage	Big Data File stage	• Session • Persistent	• User name • Digital signature • Session ID	• Session management • Authentication • Single sign-on configuration	Cannot be disabled
InfoSphere DataStage	XML stage	Session	Internal identifiers	• Session management • Authentication	Cannot be disabled
InfoSphere DataStage	IBM InfoSphere DataStage and QualityStage® Operations Console	Session	No personally identifiable information	• Session management • Authentication	Cannot be disabled
InfoSphere Data Click	InfoSphere Information Server web console	• Session • Persistent	User name	• Session management • Authentication	Cannot be disabled
InfoSphere Data Quality Console		Session	No personally identifiable information	• Session management • Authentication • Single sign-on configuration	Cannot be disabled
InfoSphere QualityStage Standardization Rules Designer	InfoSphere Information Server web console	• Session • Persistent	User name	• Session management • Authentication	Cannot be disabled
InfoSphere Information Governance Catalog		• Session • Persistent	• User name • Internal identifiers • State of the tree	• Session management • Authentication • Single sign-on configuration	Cannot be disabled
InfoSphere Information Analyzer	Data Rules stage in the InfoSphere DataStage and QualityStage Designer client	Session	Session ID	Session management	Cannot be disabled

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at http://www.ibm.com/privacy and IBM's Online Privacy Statement at http://www.ibm.com/privacy/details the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at http://www.ibm.com/software/info/product-privacy.

Trademarks

IBM, the IBM logo, and ibm.com® are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml.

The following terms are trademarks or registered trademarks of other companies:

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java™ and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACS^Link, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS^Link licensee of the United States Postal Service.

Other company, product or service names may be trademarks or service marks of others.

Index

M

N

O

P

S

T

U

W

Z

IBM ®

Printed in USA