IBM InfoSphere DataStage and QualityStage
Version 8 Release 7

# Connectivity Guide for IBM Informix Databases

**IBM**

IBM InfoSphere DataStage and QualityStage
Version 8 Release 7

*Connectivity Guide for IBM Informix Databases*

IBM

# Contents

# Chapter 1. Introduction

IBM® InfoSphere® DataStage® provides several ways to access Informix® databases. The following sections provide an introduction to the Informix connectivity stages. The second half of the chapter covers installation instructions and configuration information.

Although this book contains several stages specifically designed to access Informix databases, you can also access them by using the Dynamic Relational Stage (DRS). Use this stage to create a job that requires one relational database at design time and another at runtime. For more information about this stage, see the *InfoSphere DataStageConnectivity Guide for the Dynamic Relational Stage*.

InfoSphere DataStage allows you to read data from or write data to Informix databases. This guide describes how to use the Informix CLI, Informix enterprise, Informix Load, and Informix XPS Load stages in InfoSphere DataStage and QualityStage Designer jobs.

You have the option to install these connectivity stages automatically when you install InfoSphere DataStage. They appear in the **Database** category on the server job palette.

This guide is intended for
- InfoSphere DataStage designers who create or modify jobs that use Informix CLI, Informix enterprise, Informix Load, and Informix XPS Load stages
- InfoSphere DataStage administrators who install or upgrade InfoSphere DataStage.

## Introduction to Informix CLI Stages

The Informix CLI stage is the Informix implementation of the Microsoft Open Database Connectivity (ODBC) standard. It enables InfoSphere DataStage to read from Informix databases and write data to Informix databases. It uses the Informix CLI native programming interface. The Informix CLI stage offers the following advantages over the ODBC stage:
- Increased processing speed
- Simplified configuration on UNIX platforms
- Support for Version 9 of the Informix server software
- Advanced support for target table DDL (Create and Drop Table)
- Native metadata import

The Informix CLI stage lets you connect and process SQL statements in the native Informix environment. Because the CLI programming API is a native Informix implementation of the ODBC standard, Windows users must still use an ODBC driver manager to access CLI and use the ODBC Data Source Administrator to configure data sources.

**Note:** The term Windows represents the following:
- Windows 2000
- Windows Server 2003

In summary, the Informix CLI stage lets you do the following for a target Informix database:

- Read and write data
- Create and drop tables
- Import table and column definitions
- Browse native data with the custom user interface

Each Informix CLI stage is a passive stage that can have any number of links.

## Input Links

Input links specify the data you are writing, which is a stream of rows to be loaded into an Informix database. You can specify the data on an input link by using an SQL statement generated by the stage or a user-constructed statement.

## Output Links

Output link specify the data you are extracting, which is a stream of rows to be read from an Informix database. You can specify the data on an output link by using an SQL SELECT statement generated by the stage or a user-constructed statement.

## Reference Output Links

Each link represents rows that are key read from an Informix database (using the key columns in a WHERE clause of the SELECT statement that is constructed by the stage or user-specified).

# Introduction to Informix enterprise stages

The Informix enterprise stage is a database stage. It allows you to read data from and write data to an Informix 7.x, 8.x or 9.x database. Informix Load Stages are available for parallel jobs only.

# Introduction to Informix Load Stages

The Informix Load stage enables you to rapidly and efficiently prepare and load streams of tabular data from any IBM InfoSphere DataStage stage (for example, the ODBC stage, the Sequential File stage, and so forth) into tables of the target Informix database.

Informix Load is a passive stage that uses the Informix DBLOAD utility.

This stage supports one or more stream input links (but no output links). Each input link corresponds to a different bulk loading session within an InfoSphere DataStage job.

# Introduction to Informix XPS Load Stages

The Informix XPS Load stage is a passive stage that loads data into an Informix XPS table. This stage has one input link and no output links. The input link provides the rows of data to be loaded into the Informix XPS table.

# Configuring the Environment for Informix CLI, Informix Load, and Informix XPS Load

These Informix stages require the setting of environment variables in order to work correctly on a UNIX platform. To add or change an environment variable, include any environment variables in the *dsenv* file. Table 1 identifies the specific environment variables that are required.

*Table 1. Required environment variables for UNIX*

| For... | Set the following environment variable on the IBM InfoSphere DataStage server... |
|---|---|
| Informix CLI | INFORMIXDIR LD_LIBRARY_PATH PATH |
| Informix Load | INFORMIXDIR LD_LIBRARY_PATH PATH |
| Informix XPS Load | INFORMIXDIR LD_LIBRARY_PATH PATH |

The name of one particular environment variable, referred to as LD_LIBRARY_PATH above, differs depending on the platform. See the following to determine the correct name to use for your environment.

- If the platform is AIX®, use LIBPATH.
- If the platform is HP_UX, use SHLIB_PATH.
- If the platform is LINUX or Solaris use LD_LIBRARY_PATH.

INFORMIXDIR points to the installation directory of the Informix server. The library path must contain $INFORMIXDIR/lib:$INFORMIXDIR/lib/esql/ $INFORMIXDIR/lib/cli. PATH must contain $INFORMIXDIR/bin.

For example, the settings for an AIX system might be as follows:

```
INFORMIXDIR=/opt/informix/IDS
LIBPATH=/opt/informix/IDS/lib:/opt/informix/IDS/lib/esql:/opt/informix/IDS/lib/cli
PATH=/opt/informix/IDS/bin
```

## Additional UNIX Notes for Informix CLI

If the data source uses a translation DLL, you must add INFORMIXDIR/lib/esql to the shared library search path in *dsenv*. If you do not, and your data source requires a translation DLL, the message "Unable to load translation DLL" appears in the InfoSphere DataStage job log.

Informix CLI looks for an *.odbc.ini* file in the dshome directory, the InfoSphere DataStage server engine home directory that is stored in */.dshome*. (To see the location of the InfoSphere DataStage installation directory, enter `$cat `cat /.dshome``.)

The driver field in the data source entry for the *.odbc.ini* file *must* reference the DataDirect 4.0 non-wired ODBC driver that is included on the installation CD. Other drivers are *not* supported.

Tru64 requires the Informix CLI ODBC driver.

### Informix Connection Examples

The following example shows sample DSN entries for AIX and HP platforms for IBM InfoSphere DataStage:

```
[Informix]
Driver=/u4/dsadm/IBM/WDIS/Server/branded_odbc/lib/VMinf19.so
Description=DataDirect Informix
Database=<stores_demo>
LogonID=<userid>
Password=<password>
ServerName=<informixserver>
HostName=<informixhost>
Service=<online>
Protocol=ontlitcp
EnableInsertCursors=0
GetDBListFromInformix=0
CursorBehavior=0
CancelDetectInterval=0
TrimBlankFromIndexName=1
ApplicationUsingThreads=1
```

The following example shows sample DSN entries for Solaris platforms for
InfoSphere DataStage:

```
[Informix]
Driver=/u4/dsadm/IBM/InformationServer/Server/branded_odbc/lib/VMinf19.so
Description=DataDirect Informix
Database=<stores_demo>
LogonID=<userid>
Password=<password>
ServerName=<informixserver>
HostName=<informixhost>
Service=<online>
;Protocol=ontlitcp
EnableInsertCursors=0
GetDBListFromInformix=0
CursorBehavior=0
CancelDetectInterval=0
TrimBlankFromIndexName=1
ApplicationUsingThreads=1
```

The following example shows sample DSN entries for Tru64 and Linux platforms
for InfoSphere DataStage7.5:

```
[hpds_stores]
Driver=/u1/informix/lib/cli/iclis09b.so
Description=INFORMIX 3.3 32-BIT
Database=<stores7>
LogonID=<userid>
pwd=<password>
Servername=<hpds.1>
CursorBehavior=0
CLIENT_LOCALE=en_us.8859-1
DB_LOCALE=en_us.8859-1
TRANSLATIONDLL=/u1/informix/lib/esql/igo4a304.so
```

Every Informix data source to which your InfoSphere DataStage jobs connect must
have an entry in the *.odbc.ini* file. The only required fields in the data source
specification are the **Database** and **Server name**. If you choose to include the login
ID (UID) and/or password (PWD), you can leave the **User Name** and **Password**
properties blank. If you enter values for these properties, the values in *.odbc.ini* are
ignored.

You can also use this *.odbc.ini* file for other ODBC applications including
InfoSphere DataStage jobs using the ODBC stage.

# Configuring the environment for Informix enterprise stages

### About this task

You must have the correct privileges and settings in order to use the Informix enterprise stage. You must have a valid account and appropriate privileges on the databases to which you connect.

You require read and write privileges on any table to which you connect, and Resource privileges for by using the Partition Table property on an output link or by using create and replace modes on an input link.

To configure access to Informix:

### Procedure

1. Make sure that IBM Informix is running.
2. Make sure the INFORMIXSERVER is set in your environment. This corresponds to a server name in sqlhosts and is set to the coserver name of coserver 1. The coserver must be accessible from the node on which you invoke your IBM InfoSphere DataStage job.
3. Make sure that INFORMIXDIR points to the installation directory of your INFORMIX server.
4. Make sure that INFORMIXSQLHOSTS points to the sql hosts path (for example, */disk6/informix/informix_runtime/etc/sqlhosts*).

## Considerations for Using the High Performance Loader (HPL)

You can read and write data to an Informix IDS 10X or XPS 8.5 database by using the Informix High Performance Loader by specifying a connection method of HPL in the input or output properties.

Note the following when reading or writing by using the High Performance Loader:
- The INFORMIX onpload database must exist and be set up. You do this by running the INFORMIX ipload utility once and exiting it. An appropriate warning appears if the database is not set up properly.
- The High Performance Loader uses more shared memory, and therefore more semaphores, than INFORMIX does in general. If the HPL is unable to allocate enough shared memory or semaphores, the IBM InfoSphere DataStage read or write may not work. For more information about shared memory limits, contact your system administrator.
- You use the environment variable HPLINFORMIXDIR to specify the install directory of the HPL server.

### Reading Data on a Remote Machine using HPL

You can use read data on a remote machine by using the High Performance Loader without having INFORMIX installed on your local machine. This uses the HPL connect method in the Output properties. The machines must be cross-mounted in order to make a remote connection.

### Establishing a remote connection to an Informix enterprise stage
**About this task**

These instructions assume that IBM InfoSphere DataStage has already been installed on your local machine and that the parallel engine is available on the remote machine.

To establish a remote connection to an Informix enterprise stage:

**Procedure**

1. Verify that the INFORMIX *sqlhosts* file on the remote machine has a TCP interface. A TCP interface is necessary to use the remote connection functionality.

2. Copy the INFORMIX *etc/sqlhosts* file from the remote machine to a directory on your local machine. Set the INFORMIX HPLINFORMIXDIR environment variable to this directory.

   For example, if the directory on the local machine is */apt/informix*, the *sqlhosts* file should be in the directory */apt/informix/etc*, and the HPLINFORMIXDIR variable should be set to */apt/informix*.

3. Set the INFORMIXSERVER environment variable to the name of the remote INFORMIX server.

4. Add the remote INFORMIX server nodes to your PX node configuration file located in *$APT_ORCHHOME/../../config*; and use a nodepool resource constraint to limit the execution of the Informix enterprise stage to these nodes.

   In the example configuration file below, the local machine is fastname local_machine, and the INFORMIX remote server machine is fastname remote_machine. The nodepool for the remote nodes is arbitrarily named "InformixServer". The configuration file must contain at least two nodes, one for the local machine and one for the remote machine.

5. Go to the **Stage** page **Advanced** tab of the Informix enterprise stage. Select **Node pool and resource constraints** and **Nodepool** along with the name of the node pool constraint (i.e., "InformixServer" in the example configuration file above).

6. Set up environment variables. Remote access to an INFORMIX database requires the use of two HPLINFORMIXDIR environment variable settings, one for the local InfoSphere DataStage workstation which is set up in step 2 above, and one for the workstation with the remote INFORMIX database. The remote variable needs to be set in a startup script which you must create on the local workstation. This startup script is executed automatically by the parallel engine.

   Here is a sample *startup.apt* file with HPLINFORMIXDIR being set to */usr/informix/9.4*, the INFORMIX directory on the remote machine:

   ```
   #! /bin/sh
   HPLINFORMIXDIR=/usr/informix/9.4
   export INFORMIXDIR
   INFORMIXSQLHOSTS=$HPLINFORMIXDIR/etc/sqlhosts
   export INFORMIXSQLHOSTS
   shift 2
   exec $*
   ```

7. Set the environment variable APT_STARTUP_SCRIPT to the full path name of the *startup.apt* file.

**Results**

You are now ready to run an InfoSphere DataStage job which uses the Informix enterprise stage HPL read method to connect to a remote INFORMIX server. If you are unable to connect to the remote server, try making either one or both of the following changes to your *sqlhosts* file on the local workstation:

- In the fourth column in the row corresponding to the remote INFORMIX server name, replace the INFORMIX server name with the INFORMIX server port number found in the */etc/services* file on the remote workstation.
- The third column contains the host name of the remote workstation. Change this to the IP address of the remote workstation.

## Sample Configuration File

Here is the IBM InfoSphere DataStage example configuration file before any changes have been made:

```
{
node "node0"
    {
    fastname "local_machine"
    pools "" "node0" "local_machine"resource disk
    "/orch/s0" {}
    resource disk "/orch/s1" {}
    resource scratchdisk "/scratch" {}
    }
node "node1"
    {
    fastname "local_machine"
    pools "" "node1" "local_machine"
    resource disk "/orch/s0" {}
    resource disk "/orch/s1" {}
    resource scratchdisk "/scratch" {}
    }
}
```

Here is the InfoSphere DataStage example configuration file with changes made for the Informix enterprise stage:

```
{
node "node0"
    {
    fastname "local_machine"
    pools "" "local_machine"
    resource disk "/orch/s0" {}
    resource disk "/orch/s1" {}
    resource scratchdisk "/scratch" {}
    }
node "node1"
    {
    fastname "local_machine"
    pools "" "local_machine"
    resource disk "/orch/s0" {}
    resource disk "/orch/s1" {}
    resource scratchdisk "/scratch" {}
    }
node "node2"
    {
    fastname "remote_machine"
    pools "InformixServer" "remote_machine"
    resource disk "/orch/s0" {}
    resource disk "/orch/s1" {}
    resource scratchdisk "/scratch" {}
    }
```

```
node "node3"
  {
  fastname "remote_machine"
  pools "InformixServer" "remote_machine"
  resource disk "/orch/s0" {}
  resource disk "/orch/s1" {}
  resource scratchdisk "/scratch" {}
  }
```

## Using Informix XPS Stages on AIX Systems

In order to run jobs containing Informix XPS stages on AIX systems, you need to have the Informix client sdk 2.9X version installed along with the Informix XPS server. The LIBPATH order should be set as follows:

LIBPATH=$APT_ORCHHOME/lib:$INFORMIXDIR/lib:`dirname $DSHOME`/branded_odbc/lib:$DSHOME/lib:$DSHOME/uvdlls:$DSHOME/ java/jre/bin/classic:$DSHOME/java/jre/bin:$INFORMIXDIR/ lib:$INFORMIXDIR/lib/cli:$INFORMIXDIR/lib/esql

# Informix Stages and the Parallel Canvas

The Informix enterprise stage is available only with parallel jobs.

The other Informix connectivity stages can run on the parallel canvas. The default for *all* stages is Sequential. "In Parallel" means you can set it to run in parallel, but this is *not* the default. Use Table 2 to determine which Informix connectivity stages are available on the parallel canvas.

*Table 2. Availability on the parallel canvas*

| Connectivity stage | Available on the parallel canvas in Windows | Available on the parallel canvas in UNIX | Used as a source or a target or for processing | Runs sequentially or in parallel |
|---|---|---|---|---|
| Informix CLI | | Yes | Source or Target | Source: Sequential; Target: In Parallel |
| Informix Load | | Yes | Target | Sequential |
| Informix XPS Load | | Yes | Target | Sequential |

## Mapping String Data

The purpose of the NONE map on the Server canvas is to turn off mapping of string data in any stage in which the map is set, that is, to pass the data through verbatim. This feature is handled differently on the parallel canvas. When you define string data (for example, char, varchar), there is an additional field in the stage's Columns grid called Extended. This can be set to blank or Unicode. If this option is set to blank, no mapping occurs (for example, "NONE"); the map specified on the NLS tab is ignored. If this option is set to Unicode, the NLS map is applied. In order to read or write Japanese data, for example, set Extended to Unicode. When the job compiler detects this combination (char, varchar, and Unicode), it generates the appropriate runtime code.

# Chapter 2. Informix CLI Stages

The Informix CLI stage reads data to any IBM InfoSphere DataStage stage from an Informix database and writes data from any InfoSphere DataStage stage to an Informix database. It also provides native data browsing and metadata import from the Informix database to InfoSphere DataStage.

## Functionality of Informix CLI Stages

Informix CLI has the following functionality:
- Stream input, stream output, and reference output links.
- The ability to use the **Derivation** field to specify fully qualified column names for output links. (If the ODBC stage precedes the Informix CLI stage in a job flow, you can also use the **Derivation** field to specify fully qualified column names for input links.)
- The ability to import table and column definitions from the target Informix database and store them in the IBM InfoSphere DataStage repository.
- NLS (National Language Support).
- Reject row handling.
- File names to contain your SQL statements.
- Stored procedures.
- Data browsing, which is the ability to use the custom GUI for the stage to view sample native table data residing on the target Informix database.

The following functionality is not supported:
- Bulk loading of Informix tables from stream input. Continue to use either the Informix or Informix XPS loaders for bulk loading into an Informix database.
- Replacing the ODBC stage. The Informix CLI stage does not replace the ODBC stage. Users who created jobs by using the ODBC stage to access an Informix database might continue to run these jobs.
- Non-ANSI SQL constructs in stage-generated SQL statements.
- Version-specific SQL constructs in stage-generated SQL statements.
- Text and byte data types.

## Defining the Informix CLI Connection

When you use the GUI to edit an Informix CLI stage, the Informix CLI Stage dialog box opens. This dialog box has the **Stage**, **Input**, and **Output** pages (depending on whether there are inputs to and outputs from the stage):
- **Stage.** This page displays the name of the stage you are editing. The **General** tab defines the Informix data source, login information, and transaction isolation level information for concurrency control in jobs. You can enter text to describe the purpose of the stage in the **Description** field. The properties on this tab define the connection to the Informix data source. The **NLS** tab defines a character set map to be used with the stage. This tab appears only if you have installed NLS for InfoSphere DataStage. For details, see "Defining Character Set Mapping" on page 11.

  **Note:** You cannot change the name of the stage from this dialog box.

- **Input.** This page is displayed only if you have an input link to this stage. It specifies the SQL table to use and the associated column definitions for each data input link. It also specifies how data is written and contains the SQL statement or call syntax used to write data to an Informix table. It also specifies how to create the target table if desired and how to drop it if necessary.
- **Output.** This page is displayed only if you have an output or reference link to this stage. It specifies the SQL tables to use and the associated column definitions for each data output link. It contains the SQL SELECT statement or call syntax used to read data from one or more Informix tables or views.

# Steps in Defining an Informix CLI Stage

### About this task

The main phases in defining an Informix CLI stage from the InfmxCLI Stage dialog box are as follows:

### Procedure

1. Connect to an Informix data source (see below).
2. Optional. Define a character set map (see "Defining Character Set Mapping" on page 11).
3. Define the data on the input links (see "Defining Informix CLI Input Data" on page 11).
4. Define the data on the output links (see "Defining Informix CLI Output Data" on page 16).

## Connecting to an Informix Data Source
### About this task

The Informix connection parameters are set on the **General** tab on the Stage page. To connect to an Informix data source:

### Procedure

1. Enter the name of the Informix data source to access in the **Data source name** field. There is no default. You can enter any valid data source although unpredictable results may occur if the data source is not an Informix data source. This field is required.

   **Windows.** Define data sources by using the ODBC Administrator.

   **HP-UX 11.0 UNIX platform.** Define data sources (DSN) for Informix databases in the .odbc.ini *file.*
2. Enter the name to use to connect to the Informix data source in the **User ID** field. The Informix CLI connection also uses operating system authentication or the .odbc.ini data source definition on UNIX platforms for connection to the database.

   **Note:** You cannot use an ODBC stage that uses an Informix database as a source or target in the same job with another stage that uses Informix CLI.

   Without a user name in the **User ID** field, the Plug-in tries to use a user name provided in the data source definition. Without this user name, the operating system (OS) authentication uses the OS user name of the person running the IBM InfoSphere DataStage job. However, if the Plug-in is running on an HP platform, it does not default to the OS user name of the person running the InfoSphere DataStage job.

This user must have sufficient privileges to access the specified database and source and target tables.

3. Enter the password that is associated with the specified user name to use in the **Password** field. For security, it displays asterisks instead of the value you enter. There is no default.

   For platforms other than HP, if no user name is provided in the **User ID** field, leave **Password** blank in order to use default authentication (data source definition or OS authentication).

4. Choose an appropriate transaction isolation level to use from the **Transaction Isolation** drop-down list box. This level provides the necessary concurrency control between transactions in the job and other transactions.

   Use one of the following transaction isolation levels:

   - **Read Uncommitted.** Takes exclusive locks on modified data. These locks are held until a commit or rollback is executed. However, other transactions can still read but not modify the uncommitted changes. No other locks are taken.
   - **Read Committed.** Takes exclusive locks on modified data and sharable locks on all other data. Exclusive locks are held until a commit or rollback is executed. Uncommitted changes are not readable by other transactions. Shared locks are released immediately after the data has been processed, allowing other transactions to modify it. This is the default.
   - **Serializable.** Takes exclusive locks on modified data and sharable locks on all other data. All locks are held until a commit or rollback is executed, preventing other transactions from modifying any data that has been referenced during the transaction.

     **Note:** Transaction isolation levels apply only to databases that support logging. Otherwise, they are ignored.

5. Optional. Describe the purpose of the Informix CLI stage in the **Description** field.

## Defining Character Set Mapping

You can define a character set map for a stage. Do this from the **NLS** tab that appears on the Stage page. The **NLS** tab appears only if you have installed NLS.

Specify information by using the following button and fields:

- **Map name to use with stage.** The default character set map is defined for the project or the job. You can change the map by selecting a map name from the list.
- **Use Job Parameter... .** Specifies parameter values for the job. Use the format #*Param*#, where *Param* is the name of the job parameter. The string #*Param*# is replaced by the job parameter when the job is run.
- **Show all maps.** Lists all the maps that are shipped with IBM InfoSphere DataStage.
- **Loaded maps only.** Lists only the maps that are currently loaded.

## Defining Informix CLI Input Data

When you write data to a table in an Informix database, the Informix CLI stage has an input link. Define the properties of this link and the column definitions of the data on the Input page in the InformixCLI Stage dialog box of the GUI.

The Input page has an **Input name** field, the **General Columns** and **SQL** tabs, and the **Table Properties** (at the right of the **Drop table action** list box), **Columns...**, and **View Data...** buttons:

- **Input name.** The name of the input link. Choose the link you want to edit from the **Input name** drop-down list box. This list box displays all the input links to the Informix CLI stage.
- Click the **Columns...** button to display a brief list of the columns designated on the input link. As you enter detailed metadata in the **Columns** tab, you can leave this list displayed.
- Click the **View Data...** button to open the Data Browser. This lets you look at the data associated with the input link.

## General Tab

Use this tab to specify details of a write operation.

This tab contains the following fields:

- **Table name.** The name of the target table to update. You must specify **Table name** if you do not specify **User-defined SQL**. There is no default. This field is editable when the update action is *not* **User-defined SQL** (otherwise, it is read-only). You can also click the **...** button at the right of the **Table name** field to browse the repository to select the table.
- **Array size.** The number of rows written to the database at one time. The default is 50 rows that are cached before being written to the database.
- **Transaction size.** The number of rows that the stage processes before committing a transaction to the database. The transaction size should always be a multiple of the array size. The default is 100. This field is ignored for nonlogging databases.
- **Update action.** Specifies which stage-generated SQL statements are used to update the target table. Some update actions require key columns to update or delete rows. The default is to insert rows without clearing. Choose one of the following options:
  - **Insert rows without clearing.** Inserts the new rows in the table.
  - **Clear the table, then insert rows.** Deletes the contents of the table before inserting the new rows, with slower performance because of transaction logging.
  - **Delete existing rows only.** Deletes existing rows in the target file that have identical keys in the input rows.
  - **Replace existing rows completely.** Deletes the existing rows in the target table that have the same key as the new rows, then adds the new rows to the table.
  - **Update existing rows only.** Updates the existing data rows. Any rows in the data that do not exist in the table are ignored.
  - **Update existing or insert new rows.** Updates the existing data rows before inserting new rows. Performance depends on the contents of the target table and the rows being processed in the job. If most rows exist in the target table, it is faster to update first.
  - **Insert new or update existing rows.** Inserts the new rows before updating existing rows. Performance depends on the contents of the target table and the rows being processed in the job. If most rows do not exist in the target table, it is faster to insert first.

- **User-defined SQL.** Writes the data using a user-defined SQL statement. When you select this option, it overrides the default SQL statement generated by the stage. If you choose this option, you enter data on the **SQL** tab.
- **Create table action.** Choose one of the following options to create the target table in the specified database:
  - **Do not create target table.** Specifies that the target table is not created, and the **Drop table action** field and the **Table Properties** button (at the right of the field) are disabled.
  - **Generate DDL.** Specifies that the stage generates the CREATE TABLE statement using information obtained from "Target Table," the column definitions grid, and the advanced table properties (see the **Table Properties** button later in this section).
  - **User-defined DDL.** Specifies that you enter the appropriate CREATE TABLE statement on the **SQL** tab.
- **Drop table action.** Lets you control the dropping of the target table before it is created by the stage. If you choose not to create the target table, this field is disabled. The list box displays the same items as the **Create table action** list box except that they apply to the DROP TABLE statement.
- **Table Properties button.** Click the button at the right of the **Drop table action** list box to display the Create Table Properties dialog box. You can then specify the following advanced table properties from this dialog box.
  - **Informix Version.** Specifies the version of the target Informix database. This property lets the interface display interface elements that are relevant to the database version. For example, **Table type** applies only for Version 8, so it is disabled for Versions 7 or 9.
  - **Table type.** Choose one of the following uneditable options: **RAW, OPERATIONAL, STANDARD, or STATIC.** This field is available only for Informix Version 8.
  - **Lock Mode.** Choose one of the following non-editable options: **PAGE, TABLE, or ROW.** Row locking is available only for Informix Version 8.
  - **Extent size and Next size.** These buttons accept an integer. The GUI restrains the lower limit of these values.
  - **Storage expression.** Lets you enter a storage expression that is appropriate for the target database. It corresponds to the Storage Options clause described in *Informix Guide to SQL Syntax*. No validation is performed.
- **Description.** Optionally, enter text to describe the purpose of the link.

## Columns Tab

This tab contains the column definitions for the data written to the table or file. The **Columns** tab behaves the same way as the **Columns** tab in the ODBC stage.

The column definitions are used in the same way and order they appear in the Columns metadata definitions grid except for the **Description** field. The Informix CLI stage uses the text in this field to define the column precision for Informix DATETIME columns when you want the target table to be created. The **Description** field must be empty for a DATETIME column with default precision.

Columns defined as InfoSphere DataStage Time are always created in Informix as "DATETIME hour to second." InfoSphere DataStage Timestamp columns are created in Informix as DATETIME. Use the **Description** field to specify the Informix DATETIME precision if you do not want to use the default for the database.

For example, suppose call_dtime is a column name for the SQL Timestamp data type. Because Informix uses the DATETIME data type to represent dates and times, a conversion must be made between the InfoSphere DataStage representation and the Informix representation. Informix allows DATETIME columns to be defined with a precision that represents a full or contiguous subset of the components of a timestamp. This DATETIME precision specification is given when the table is created.

When the Informix CLI stage generates the CREATE TABLE DDL statement, it uses the value specified in the **Description** column to define the DATETIME precision. Specify **year to day** in the **Description** column to cause call_dtime to be created as a DATETIME column that accepts dates in the 1997-04-22 format. Specify **year to fraction(3)** to cause res_dtime to be created as a DATETIME column accepting full ANSI timestamps.

## SQL Tab

Use this tab to view details of a generated query or to define your own query.

This tab contains the **Generated**, **User-defined**, **Before**, **After**, **Generated DDL**, and **User-defined DDL** tabs. Use these tabs to display the stage-generated SQL statement and the SQL statement that you can enter.

- **Generated.** Displays the SQL statements constructed by IBM InfoSphere DataStage that are used to write data to Informix. The statements represent the uneditable result of the selection made in the **Update action field on the General** tab. You can use **Copy** to copy them to the Clipboard for use elsewhere. This tab is displayed by default.

- **User-defined.** Select **User-defined SQL** from the **Update action** field on the **General** tab to enable this tab. The GUI displays the stage-generated SQL statement on this tab as a starting point. However, you can enter any valid, appropriate SQL statement. The box size changes proportionately when you resize the main window to display long SQL statements.

- **Before.** This tab contains the SQL statements executed before the stage processes any job data rows. The elements on this tab correspond to the "Before SQL" and "Continue if Before SQL failed" grid properties. The **Before** and **After** tabs look alike. The "**Continue if Before SQL fails" and** "**Continue if After SQL fails"** properties are represented by check boxes and the SQL statement is entered in a resizable edit box.

- **After.** This tab contains the SQL statements executed after the stage processes job data rows. The elements on this tab correspond to the "After SQL" and "Continue if After SQL failed" grid properties. The **Before** and **After** tabs look alike. The "**Continue if Before SQL fails" and** "**Continue if After SQL fails"** properties are represented by check boxes and the SQL statement is entered in a resizable edit box.

  - **Generated DDL.** Select **Generate DDL** or **User-defined DDL** from the **Create table action** field on the **General** tab to enable this tab. The **CREATE statement** field displays the CREATE TABLE statement that is generated from the column metadata definitions and the information provided on the Create Table Properties dialog box. If you select an option other than **Do not drop target table** from the **Drop table action** list, the **DROP statement** field displays the generated DROP TABLE statement for dropping the target table.

- **User-defined DDL.** Select **User-defined DDL** from the **Create table action** or **Drop table action** field on the **General** tab to enable this tab. The generated DDL statement is displayed as a starting point to define a CREATE TABLE and a DROP TABLE statement.

The **DROP statement** field is disabled if **User-defined DDL** is not selected from the **Drop table action** field. If **Do not drop target** is selected, the **DROP statement** field is empty in the **Generated DDL** and **User-defined DDL** tabs.

**Note:** Once you modify the user-defined DDL statement from the original generated DDL statement, changes made to other table-related properties do not affect the user-defined DDL statement. If, for example, you add a new column in the column grid after modifying the user-defined DDL statement, the new column appears in the generated DDL statement but does not appear in the user-defined DDL statement.

# Writing Data to Informix

The following sections describe the differences when you use stage-generated or user-defined SQL INSERT, DELETE, or UPDATE statements to write data from IBM InfoSphere DataStage to an Informix database.

## Using Generated SQL Statements
### About this task

By default, IBM InfoSphere DataStage writes data to an Informix table using an SQL INSERT, DELETE, or UPDATE statement that it constructs. The generated SQL statement is automatically constructed using the InfoSphere DataStage table and column definitions that you specify in the input properties for this stage. The **Generated** tab on the **SQL** tab displays the SQL statement used to write the data.

To use a generated statement:

### Procedure
1. Enter a table name in **Table name** on the **General** tab on the Input page.
2. Specify how you want the data to be written by choosing an option from the **Update action** drop-down list box.
3. Refer "Defining Informix CLI Input Data" on page 11 for a description of the update actions.
4. Enter an optional description of the input link in the **Description** field.
5. Click the **Columns** tab on the Input page.
6. Edit the Columns grid to specify column definitions for the columns you want to write.
   The SQL statement is automatically constructed using your chosen update action and the columns you have specified. You can now optionally view this SQL statement.
7. Click the **SQL** tab on the Input page, then the **Generated** tab to view this SQL statement. You cannot edit the statement here, but you can access this tab at any time to select and copy parts of the generated statement to paste into the user-defined SQL statement.
8. Click **OK** to close this dialog box. Changes are saved when you save your job design.

## Using User-Defined SQL Statements
### About this task

Instead of writing data by using an SQL statement constructed by IBM InfoSphere DataStage, you can enter your own SQL INSERT, DELETE, or UPDATE statement

for each Informix CLI input link. Ensure that the SQL statement contains the table name, the type of update action you want to perform, and the columns you want to write.

To enter an SQL statement:

### Procedure

1. Choose **User-defined SQL** from the **Update action** drop-down list box from the **General** tab of the Input page.
2. Click **User-defined** tab on the SQL page. By default you see the stage-generated SQL statement. You can edit this statement or enter the SQL statement you want to use to write data to the target Informix tables. This statement must contain the table name, the type of update action you want to perform, and the columns you want to write.

   If the property value begins with {FILE}, the remaining text is interpreted as a path name, and the contents of the file supplies the property value.

   When writing data, the INSERT statements must contain a VALUES clause with parameter markers ( ? ) for each stage input column. UPDATE statements must contain a SET clause with parameter markers for each stage input column. UPDATE and DELETE statements must contain a WHERE clause with parameter markers for the primary key columns. The parameter markers must be in the same order as the associated columns listed in the stage properties. For example:

   ```
   INSERT emp (emp_no, emp_name) VALUES (?, ?)
   ```

   If you specify multiple SQL statements, each is executed as a separate transaction. Terminate individual SQL statements with a semicolon ( ; ).

   Like the **Storage expression** edit box on the Create Table Properties dialog box, the size of this box changes proportionately when the main window is resized in order to allow the convenient display of very long or complex SQL statements.

   Unless you specify a user-defined SQL statement, the stage automatically generates an SQL statement.
3. Click **OK** to close the Informix CLI Stage dialog box. Changes are saved when you save your job design.

## Defining Informix CLI Output Data

When you read data from an Informix data source, the Informix CLI stage has an output link. The properties of this link and the column definitions of the data are defined on the Output page in the Informix CLI Stage dialog box.

The Output page has an **Output name** field and the **General Columns**, **Selection** and **SQL** tabs. The tabs displayed depend on how you specify the SQL statement to output the data. The **Columns...** and **View Data...** buttons function identically to those on the Input page.

- **Output name.** The name of the output link. Choose the link you want to edit from the **Output name** drop-down list box. This list box displays all the output links.
- Click **Columns...** to display a brief list of the columns designated on the input link. As you enter detailed metadata in the **Columns** tab, you can leave this list displayed.

- Click **View Data...** to invoke the Data Browser. This enables you to look at the data associated with the output link.

## General Tab

Use this tab to specify details of a read operation.

This tab is displayed by default. It provides the interface for entering table names, the number of prefetch rows returned from Informix, and the type of query. The **General** tab contains the following fields:

- **Table names.** This field appears only when you select **Generated SQL query**. It contains the names of the Informix source tables or files being accessed. These tables must exist or be created and populated by Before SQL statements. There is no default.

  Separate multiple table names by a comma ( , ). You must have select privileges on each table. If you specify **User-defined SQL query**, **Table names** is ignored. You must specify a table name in **Table names** if you do not define **User-defined SQL query**.

  Additionally, you can use a job parameter to specify the table name.

  You can also click the **...** button at the right of the **Table name** field to browse the repository to select the table.

- **Prefetch rows.** The number of rows that Informix returns when InfoSphere DataStage fetches data from the source tables. Specifying a value greater than 1 improves performance (memory usage increases to accommodate buffering multiple rows).

- **Query type.** Displays the Generated SQL query and **User-defined SQL query** options. **Query type** combines the "**Generate SQL**" and "**User Defined SQL**" grid properties.

  - **Generated SQL query.** This is the default setting, which specifies that the data is extracted using an SQL statement constructed by the stage. When this option is selected, the **Generated** tab appears. You cannot edit this statement.

  - **User-defined SQL query.** Specifies that the data is extracted using a user-defined SQL query. When this option is selected, the **User-defined** tab appears allowing you to edit SQL statements.

- **Description.** Contains an optional description of the output link.

## Columns Tab

This tab contains the column definitions for the data being output on the chosen link.

**Note:** The **Derivation** field on the **Columns** tab contains a fully qualified representation of column names when table definitions are loaded from the repository.

The column definitions for reference links require a key field. Key fields join reference inputs to a Transformer stage. The Informix CLI stage key reads the data by using a WHERE clause in the SQL SELECT statement.

## Selection Tab

This tab is used primarily with generated SQL queries. It contains optional SQL SELECT clauses for the conditional extraction of data. These clauses are appended to the generated SQL statements.

# Reading Data from Informix

The following sections describe the differences when you use generated queries or user-defined queries to read data from an Informix database into IBM InfoSphere DataStage.

## Using Generated Queries

By default, IBM InfoSphere DataStage extracts data from an Informix data source by using an SQL SELECT statement that it constructs. The SQL statement is automatically constructed by using the table and column definitions that you entered in the stage output properties.

When you select **Generated SQL query**, data is extracted from an Informix database by using an SQL SELECT statement constructed by the stage. SQL SELECT statements have the following syntax:

```
SELECT clause FROM clause
[WHERE clause]
[GROUP BY clause]
[HAVING clause]
[ORDER BY clause];
```

When you specify the tables to use and the columns to be output from the Informix CLI stage, the SQL SELECT statement is automatically constructed and can be viewed by clicking the **SQL** tab on the Output page.

For example, if you extract the **Name**, **Address**, and **Phone** columns from a table called Table1, the SQL statement displayed on the **SQL** tab is:

```
SELECT Name, Address, Phone FROM Table1;
```

The SELECT and FROM clauses are the minimum required and are automatically generated by the stage. However, you can use any of these SQL SELECT clauses:

- **SELECT clause.** Specifies the columns to select from the database.
- **FROM clause.** Specifies the tables containing the selected columns.
- **WHERE clause.** Specifies the criteria that rows must meet to be selected.
- **GROUP BY clause.** Groups rows to summarize results.
- **HAVING clause.** Specifies the criteria that grouped rows must meet to be selected.
- **ORDER BY clause.** Sorts selected rows.

If you want to use the additional SQL SELECT clauses, you must enter them on the **Selection** tab on the Output page. These clauses are appended to the SQL statement that is generated by the stage. If this link is a reference link, only the WHERE clause is enabled.

The **Selection** tab is divided into two areas (panes). You can resize an area by dragging the split bar for displaying long SQL clauses.

- **WHERE clause.** This text box allows you to insert an SQL WHERE clause to specify criteria that the data must meet before being selected.
- **Other clauses.** This text box allows you to insert a GROUP BY, HAVING, or ORDER BY clause.

## Using User-Defined Queries

### About this task

Instead of using the SQL statement constructed by the stage, you can enter your own SQL statement for each output link.

### Procedure

1. Select **User-defined SQL query** from the **Query type** drop-down list box on the **General** tab of the Output page. The **User-defined** tab on the **SQL** tab is enabled.

2. You can edit the statements or drag-and-drop the selected columns into your user-defined SQL statement. You must ensure that the table definitions for the output link are correct and represent the columns that are expected.

   If your entry begins with {FILE}, the remaining text is interpreted as a path name, and the contents of the file supplies the text for the query.

3. Click **OK** to close the dialog box. Changes are saved when you save your job design.

# Mapping Data Types

You can map IBM InfoSphere DataStage data types to Informix data types. When "Create Table" is set to Yes for input links, the target table is created by using the input link column definitions and the specific input link properties that define the properties for the target table.

There is no exact translation between an Informix data type and an InfoSphere DataStage data type such as MONEY. The following table shows the Informix data types that are generated from the corresponding InfoSphere DataStage data types:

*Table 3. Data types*

| InfoSphere DataStage Data Type | Informix Data Type |
|---|---|
| SQL_BIGINT | INT8 |
| SQL_BINARY | BYTE (Not supported) |
| SQL_BIT | BOOLEAN (Supported only with Informix Universal Server 9.*n*) |
| SQL_CHAR | CHAR(*n*) |
| SQL_DATE | DATE |
| SQL_DECIMAL | DECIMAL(*p, s*) |
| SQL_DOUBLE | DOUBLE PRECISION |
| SQL_FLOAT | FLOAT(*n*) |
| SQL_INTEGER | INTEGER |
| SQL_LONGVARBINARY | BYTE (Not supported) |
| SQL_LONGVARCHAR | TEXT |
| SQL_NUMERIC | DECIMAL |
| SQL_REAL | REAL |
| SQL_SMALLINT | SMALLINT |
| SQL_TIME | DATETIME hour to second |
| SQL_TIMESTAMP | DATETIME |

*Table 3. Data types  (continued)*

| InfoSphere DataStage Data Type | Informix Data Type |
|---|---|
| SQL_TINYINT | Not supported |
| SQL_VARBINARY | BYTE (Not supported) |
| SQL_VARCHAR | VARCHAR($m,r$) |

# Stored Procedure Support

You can call stored procedures from the server Informix CLI stage. The following restrictions apply:

- Output parameters are not supported.
- You can call stored procedures as part of the Before SQL and After SQL statements. Any result sets generated by the procedure are discarded.
- You can also call stored procedures as part of the "User Defined SQL." For input links, the stored procedure must have at least the same number of input parameters as stage input columns, and the data type for the column should match the parameter type for the procedure. If there are more input parameters than columns, they must be specified as literals or job parameters.
- For output and reference links, the stored procedure must generate a row result set that matches the stage output column definitions. Additionally, a stored procedure for a reference link must produce one row based on the value of the primary key columns for the link.

# Chapter 3. Informix enterprise stage

The Informix enterprise stage is a database stage. By using the Informix enterprise stage, you can read data from or write data to an IBM Informix Dynamic Server.

## Working with Informix enterprise stage

To use an Informix enterprise stage in a job, you must set the properties of the stage. The stage properties that you must set depend on whether you want to write data to an Informix database or read data from an Informix database.

### Writing data to an Informix database
#### Before you begin

You must specify properties of the stage by using the properties tab before writing data to an Informix database.

#### About this task

These properties are necessary for writing data to an Informix database.
- On the **Properties Tab**, under the Target category:
  - Specify the table that you want to write data to.
  - Specify the write mode. By default, the input link of the Informix enterprise stage appends data to existing tables. You can choose to create a table, replace an existing table, or keep existing table details but replace all the rows.

    Under the Connection category:
  - Specify the connection method. You can choose Extended Parallel Server Fast (XPS Fast) for connecting to the XPS framework, High Performance Load (HPL) for connecting to HPL servers, or Native for connecting to any version release 7.x and above.
  - Optionally specify the name of the database that you want to connect to.
  - If you specified the XPS Fast or HPL Connection Method, specify the name of the server hosting Informix XPS. The name of the server specified will be taken from the INFORMIXSERVER environment variable.
- Ensure that column metadata is specified for the write operation.

### Reading data from an Informix database
#### Before you begin

You must specify properties of the stage by using the properties tab to read data from an Informix database.

#### About this task

These properties are necessary for reading data from an Informix database.
- In the **Properties Tab**, under the Source category:
  - Choose a read method. The default value is table, which reads data directly from a table. You can choose to read data by using auto-generated SQL or user-generated SQL.

- Specify the table to read the data from.
- If you specify a read method of user-generated SQL, specify the SELECT SQL statement to use. The output link provides an auto-generated statement that you can edit as required.

  Under the Connection category:

- Specify the connection method. You can choose XPS Fast (for connecting to the XPS framework), HPL (for connecting to HPL servers), or Native (for connecting to Informix version 8.5).
- Optionally specify the name of the database that you are connecting to.
- If you specified the XPS Fast or HPL connection method, specify the name of the server where Informix XPS is installed.

- Ensure that column metadata is specified for the read operation.

# Stage page

The Stage page has a maximum of three tabs - General, Advanced, and NLS. Using these tabs, you can specify the properties of the Informix enterprise stage.

## General tab

On the General tab, you can specify an optional description of the stage.

## Advanced tab

On the **Advanced** tab, you must specify properties. These properties determine how the stage should be run.

- **Execution Mode**. The execution mode depends on the type of operation that the stage is performing.

*Table 4. Execution modes based on type of operation*

| Type of operation | Connection method | Execution mode |
|---|---|---|
| Writing data to an Informix database | XPS Fast | Parallel |
| Writing data to an Informix database | HPL | Sequential |
| Writing data to an Informix database | Native | Sequential |
| Reading data from an Informix database | HPL | Sequential |

- **Combinability mode**. Combinability mode is Auto by default. In auto mode, operators for the parallel stages are combined to run in the same process to improve performance.
- **Preserve partitioning**. You can select **Set** or **Clear**. If you select **Set**, a read operation sends a request to the next stage to preserve partitioning.
- **Node pool and resource constraints**. Select this option to constrain parallel execution to the node pools or resource pools that are specified in the grid. Use the grid to select node pools and resource pools. The selections are populated from the configuration file.
- **Node map constraint**. Select this option to constrain parallel execution to the nodes in a defined node map. You can define a node map by typing node numbers into the text box or by clicking the browse button to select nodes from the **Available Nodes** window**.**

**NLS Map tab**

You can enter a character set map for the Informix enterprise stage by using the NLS Map tab. NLS Map tab overrides the default character set map set for the project or the job. You can specify that the map may be supplied as a job parameter.

# Input page

On the Input page, you can specify details about how the Informix enterprise stage writes data to an Informix database. To specify these details, you can use the General, Properties, Partitioning, Columns, and Advanced tabs.

The Informix stage can only have one input link that writes data to one table.

## General tab

On the General tab, you can specify an optional description of the stage.

## Columns tab

On the Columns tab, you can specify the column definitions of the incoming data.

## Advanced tab

On the **Advanced** tab, you can change the default buffering settings for the input link.

## Properties tab

On the **Properties** tab, you can specify details of exactly what the link does.

The properties determine how incoming data is written and where the data is written. Some of the properties are required, although many have default settings. Properties without default settings are displayed in red in this tab and turn black when you supply a value for them.

The following table summarizes the properties and their attributes. A more detailed description of each property follows.

*Table 5. Input link properties and values*

| Category and Property | Values | Default | Required? |
|---|---|---|---|
| Target | | | |
| Target Category | • Append<br>• Create<br>• Replace<br>• Truncate | Append | Yes |
| Table | Table Name | N/A | Yes |
| Connection | | | |

*Table 5. Input link properties and values (continued)*

| Category and Property | Values | Default | Required? |
|---|---|---|---|
| Connection method | • XPS Fast<br>• HPL<br>• Native | XPS Fast | Yes |
| Remote Server | • True<br>• False | False | Yes if Connection Method = Native |
| User | User ID | N/A | Yes if Connection Method = Native and Remote Server = True |
| Password | Password | N/A | Yes if Connection Method = Native and Remote Server = True |
| Database | Database name | N/A | Yes |
| Server | Server name | N/A | No |
| Options | | | |
| Close Command | Close command SQL statement | N/A | No |
| Open Command | Open command SQL statement | N/A | No |
| Silently Drop Columns Not in Table | • True<br>• False | False | Yes |
| Default String Length | String length | 32 | Yes |

## Target category

Under Target, you must specify the write mode and destination table.

**Write mode**

Select from the available values:
- **Append.** Appends new records to the table. The database user who writes in this mode must have Resource privileges. This is the mode is the default mode.
- Create. Creates a new table. The database user who writes in this mode must have Resource privileges. The stage returns an error if an existing table has the same name as the table name that you specified.
- **Replace.** Deletes the existing table and creates a new one in its place. The database user who writes in this mode must have Resource privileges.

  You cannot create or replace a table that has primary keys. Do not specify primary keys in your metadata.
- **Truncate**. Retains the table attributes but discards existing records and appends new ones. The stage will run more slowly in this mode if the user does not have Resource privileges.

## Table

Specify the name of the Informix table to write data to. The following property is dependent on the Table property:

- **Select List**

  Specifies a list that determines which columns are written. If you do not supply the list, the stage writes data to all columns.

## Connection category

Under connection category, you must specify the connection method, remote server, user, password, database and server properties.

### Connection method

Specify the method to use to connect to the Informix database:

- **XPS fast**. Use this option to connect to an Informix XPS (version 8.5) database. IBM InfoSphere DataStage connects directly to the XPS framework.
- **HPL**. Use this option to connect to Informix servers (versions 9.4, 10.0) by using the High Performance Loader (HPL).
- **Native**. Use this option to connect to any version of Informix (versions 9.4, 10.0) by using native interfaces.

### Remote Server

This option is available if you select the Native connection method. The default value is False. If you select True, you must specify authentication details for the remote server.

### User

This option is available only for a connection method of Native with the Remote Server option set to True. Specify the user ID for connecting to the remote database.

### Password

This option is available only for a connection method of native with the remote server option set to true. Specify the password for connecting to the remote database with the user ID specified by the User option. The password is encrypted.

### Database

Optionally specify the name of the Informix database that contains the table that is specified by the Table Property.

### Server

This option is available only with a connection method of XPS Fast or HPL. Specify the name of the Informix XPS server.

## Option category

### Close Command

Specify an Informix SQL statement to be parsed and run by the Informix server on all processing nodes after the table is populated.

**Open Command**

Specify an Informix SQL statement to be parsed and run by the Informix server on all processing nodes before the table is opened.

**Silently Drop Columns Not in Table**

Use this property to drop, with a warning, all input columns that do not correspond to the columns of an existing table. If you do not specify drop, an unmatched column generates an error, and the associated step terminates.

**Default String Length**

Set the default length of string columns. If you do not specify a length, the default is 32 bytes. You can specify a length up to 255 bytes.

# Partitioning tab

On the **Partitioning** tab you can specify details about how the incoming data is partitioned or collected before it is written to the Informix database. You can also specify that the data should be sorted before it is written to the Informix database.

By default, Informix enterprise stage partitions data in Auto mode. Auto mode is the best partitioning method based on the execution modes specified for the current and preceding stages and the number of nodes specified in the configuration file.

If the stage is operating in sequential mode, it first collects data and then writes the data to the file by using the default Auto collection method.

You can override this default behavior by using the **Partitioning** tab. The options that you set on this tab behave differently according to whether the job in the preceding or the current stage is set to run in parallel or sequential mode.

If the stage is set to run in parallel, then you can set a partitioning method by selecting from the **Partition type** drop-down list. The partition type that you select will override any current partitioning.

If the stage is set to run in sequential mode, but the preceding stage is run in parallel, then you can set a collection method from the **Collector type** drop-down list. The collection method that you select will override the default collection method.

The following partitioning methods are available:
- **(Auto)**. By default, Informix enterprise stage partitions data in Auto mode. Auto mode provides the best partitioning method, depending on the execution modes of current and preceding stages and how many nodes are specified in the configuration file.
- **Entire**. Each file that is written to receives the entire data set.
- Hash. The records are divided into partitions based on the value of key columns that are selected from the **Available** list.

- **Modulus**. The records are divided into partitions using a modulus function on the key column you select from the **Available** list. This method is commonly used to create partitions on tag columns.
- **Random**. The records are divided into partitions randomly, based on the output of a random number generator.
- **Round Robin**. The records are partitioned on a round robin basis as they enter the stage. The Round Robin partitioning method reads a record from the first input partition, then from the second partition, and so on. After reaching the last partition, the read operation starts with the first record again
- **Same**. Preserves the current partitions.
- **Range**. The records in a data set are divided into approximately equal- size partitions based on one or more partitioning keys. Range partitioning is often a preprocessing step that is performed before a total sort on a data set. Additional properties must be set. Access these properties by clicking the properties button.

The following collection methods are available:
- **Auto**. The default data collection method for Informix enterprise stage is Auto. You can use Auto mode if you want to read any row from any input partition as the row becomes available
- **Ordered**. Reads all records from the first partition, then all records from the second partition, and so on.
- **Round Robin**. Reads a record from the first input partition, then from the second partition, and so on. After reaching the last partition, the read operation starts with the first record again.
- **Sort Merge**. Reads records in an order that is based on one or more columns of the record. This method requires you to select a collecting key column from the Available list.

You can specify that data that arrives on the input link should be sorted before it is written to the database. The sort is always performed within data partitions. If the stage divides incoming data into partitions, the sort operation occurs after the data is partitioned. If the stage is collecting data, the sort operation occurs before the data is collected. The availability of sorting depends on the selected partitioning or collecting method. The following options are available:
- **Perform Sort**. Sorts the data that is arriving on the link. Select the columns to sort on from the **Available** list.
- **Stable**. Preserves previously sorted data sets. This option is default.
- **Unique**. Retains only one record if multiple records have identical sorting key values. If the stable sort option is also set, the first record is retained.

If NLS is enabled, you can select a locale that specifies a collate convention for the sort operation. Click the NLS button to select the locale.

You can also specify sort direction, case sensitivity, whether to sort data as ASCII or EBCDIC, and whether null columns appear first or last for each column. If you are using a keyed partitioning method, you can also specify whether the column is used as a key for sorting, for partitioning, or for both. Select the column in the **Selected** list and right-click to open the shortcut menu.

# Output page

On the Output page, you can specify details about how the Informix enterprise stage reads data from an Informix database. The stage can have only one output link.

To specify these details, you can use the General, Properties, Columns, and Advanced tabs.

### General tab

On the General tab, you can specify an optional description of the stage.

### Columns tab

On the Columns tab, you can specify the column definitions of the incoming data.

### Advanced tab

On the **Advanced** tab, you can change the default buffering settings for the output link.

# Properties tab

On the **Properties** tab, you can specify details of exactly what the link does.

The properties that you set for the output link define how incoming data is read and from what table. Some of the properties are required, although many have default settings. Properties without default settings appear as red in this tab and turn black when you supply a value for them.

The following table lists the properties and their attributes. A more detailed description of each property follows.

*Table 6. Output link properties and values*

| Category and Property | Values | Default | Required? | Dependent of |
|---|---|---|---|---|
| Source | | | | |
| Read Method | • Table<br>• Auto-generated SQL<br>• User-defined SQL | Table | Yes | N/A |
| Table | Table name | Yes | Yes if Read Method = Table or Auto-generated SQL | N/A |
| Select List | List | N/A | No | Table |
| Where Clause | Filter | N/A | No | Table |
| Partition Table | Table | N/A | No | Table |

*Table 6. Output link properties and values  (continued)*

| Category and Property | Values | Default | Required? | Dependent of |
|---|---|---|---|---|
| Query | SQL query | N/A | Yes if Read Method = User-defined SQL or Auto-generated SQL | N/A |
| Connection | | | | |
| Connection method | • XPS Fast<br>• HPL<br>• Native | XPS Fast | Yes | N/A |
| Remote server | • True<br>• False | False | Yes if Connection Method = Native | N/A |
| User | User ID | N/A | Yes if Connection Method = Native and Remote Server = True | N/A |
| Password | Password | N/A | Yes if Connection Method = Native and Remote Server = True) | N/A |
| Database | Database name | N/A | No | N/A |
| Server | Server name | N/A | No | N/A |

## Source category

Under source, you have to specify the read method, table, and query properties.

**Read Method**

The default read method is Table. Select **Table** to use the Table property to specify the read method. Select **Auto-generated SQL** to automatically generate an SQL query based on the columns that you have defined and the table that you have specified in the Table property. Select **User-defined SQL** to define your own SQL query.

**Table**

Specify the name of the Informix table to read data from. You can prefix the table name with a table owner in the form *table_owner.table_name*.

• **Where Clause**

  Specify selection criteria to be used as part of an SQL statement WHERE clause, to specify the rows of the table to include in or exclude from the data set.

• **Select List**

  Specify a list that determines which columns are read. If you do not supply the list, the stage reads all columns. Do not include formatting characters in the list.

• **Partition Table**

Specify this property if the table is fragmented to improve performance by creating one instance of the stage per table fragment. If the table is fragmented across nodes, this property creates one instance of the stage per fragment per node. If the table is fragmented and you do not specify this option, the stage functions successfully, but more slowly. You must have Resource privilege to specify this property. This property is available only for the connection methods of XPS Fast and Native.

These dependent properties are available only when you have specify a read method of Table rather than Auto-generated SQL.

**Query**

This property is used to contain the SQL query when you choose a read method of User-defined SQL or Auto-generated SQL. If you are using Auto-generated SQL, you must select a table and specify some column definitions to have generate the query.

## Connection category

Under connection category, you must specify the connection method, remote server, user, password, database and server properties.

**Connection method**

Specify the method to use to connect to the Informix database:
- **XPS fast**. Use this option to connect to an Informix XPS (version 8.5) database. IBM InfoSphere DataStage connects directly to the XPS framework.
- **HPL**. Use this option to connect to Informix servers (versions 9.4, 10.0) by using the High Performance Loader (HPL).
- **Native**. Use this option to connect to any version of Informix (versions 9.4, 10.0) by using native interfaces.

**Remote Server**

This option is available if you select the Native connection method. The default value is False. If you select True, you must specify authentication details for the remote server.

**User**

This option is available only for a connection method of Native with the Remote Server option set to True. Specify the user ID for connecting to the remote database.

**Password**

This option is available only for a connection method of native with the remote server option set to true. Specify the password for connecting to the remote database with the user ID specified by the User option. The password is encrypted.

**Database**

Optionally specify the name of the Informix database that contains the table that is specified by the Table Property.

**Server**

This option is available only with a connection method of XPS Fast or HPL. Specify the name of the Informix XPS server.

## Options category

Under options category, you must specify close command and open command.

**Close Command**

Specify an Informix SQL statement to be parsed and run by the Informix server on all processing nodes after the table selection or query is completed.

**Open Command**

Optionally specify an Informix SQL statement to be parsed and run by the database on all processing nodes before the read query is prepared and run.

# Chapter 4. Informix Load Stage

Informix Load is a passive stage that loads streams of tabular data into tables of a target Informix database.

## Functionality of Informix Load Stages

The Informix Load stage has the following functionality:
- Support for data files that exceed the 2-GB file size limit for 64-bit file systems.
- An option to use the Informix DBLOAD utility for automatic bulk loading. This is useful when you run an IBM InfoSphere DataStage job from an administrative account on the same machine as the Informix server.
- Multiple bulk-loading operations through a single Informix Load stage.
- Before- and after-link external user routines that can use arguments for link properties that you specify. The values for these link properties, such as Table Name, Database Name, Control File Name, and Data File Name, are substituted at run time.

  You can use these external routines for before-link operations (for example, initiating customized table cleanup) or after-link operations (for example, initiating custom or remote bulk loading, deleting temporary files, initiating file transfer and remote processing, recreating table indexes that were dropped before bulk loading, and so forth).
- Support for NLS (National Language Support).

The following functionality is not supported:
- Bulk loading multiple tables from a single link (you can do this only by providing a customized control file).
- Generation of fixed-record data file format.
- Multiple data files for a single bulk loading session (you can do this only by using multiple links through the Informix Load stage).
- Automatic bulk loading option works only when you run a InfoSphere DataStage job from an administrative account on the same machine as the Windows Informix server. Otherwise, you can also use automatic bulk loading by providing an external after-link routine.

The bulk loading process does not clean up the table before loading. Use an external before-link routine to start the cleanup of the table.

## Terminology Used in Informix Load Stages

The Informix terms used in this document are explained below.

**before-link and after-link routines**
> A user-specified external routine or process that performs an action before or after all link-related activities for a specific link, for example, cleaning tables before bulk loading.

**Bulk Load stage**
> A passive stage whose role in an IBM InfoSphere DataStage job is to take streams of tabular data and load them into tables of a target database.

**33**

**control file**
A file of commands for bulk loading one or more tables from a single link.

**data file**
An ASCII file of row/column data from an input link that is to be loaded.

**local bulk loading**
Bulk loading that occurs when both the InfoSphere DataStage job and Informix server are on the same physical machine.

**remote bulk loading**
Bulk loading that occurs when both the InfoSphere DataStage job and Informix server are on different physical machines.

**Server** The Informix database to which you connect in order to use the Informix Load stage.

:

## Local and Remote Processing

Informix Load supports both local bulk loading and remote bulk loading.

To achieve maximum bulk-loading performance, use the Informix Load stage locally on the same machine where the Informix database server is located. The default settings in the Informix Load stage assume the local bulk-loading processing.

In the case of the local processing, the only required link-level settings for the Informix Load stage are the following:

- Database Server Name
- Database Name
- Table Name (if the custom control file is not present)

Additional link-level properties, however, may be used to customize or increase the scope of the Informix Load stage, for example, allowing for remote bulk loading.

## Defining Data Types

The following sections describe DATETIME and DATE definitions.

### DATETIME Considerations

In Informix, a valid DATETIME data type can be either *yyyy-mo-dd hh:mi:ss:ff*... or any contiguous subset of the previous string, for example, *mo-dd hh*. Any column that is defined to contain such a subset does not accept a full ANSI DATETIME.

Therefore, the DATETIME type generated within the data file corresponds to the DATETIME type defined for the actual column in a table. To be consistent with the Informix DDL, you should define the following in the derivation for every DATETIME column (except the default):

**Expression**
**Description**

**Hour to Fraction(5)**
> A partial ANSI DATETIME type, starting with hours and ending with fractions, with a precision of 5.

**Hour to Second**
> A partial ANSI DATETIME type, starting with hours and ending with seconds.

**Month to Minute**
> A partial ANSI DATETIME type, starting with months and ending with minutes.

**Year to Fraction(3)**
> The default value. Full ANSI DATETIME, with a precision of 3.

## DATE Considerations using Informix Load Stages

In Informix, the interpretation of a DATE data type by a DBLOAD bulk loader depends on the settings of the DBDATE or GL_DATE environment variables. For example, if DBDATE is set to DMY4, then the *xx-yy-zzzz* string is interpreted as the *xx* day of the *yy* month for the *zzzz* year.

Informix Load cannot read the DBDATE or GL_DATE environment variables automatically for date formatting for the DBLOAD utility, since the InfoSphere DataStage job can be run on a different physical machine than the one where the Informix server is installed.

The default DBDATE Informix setting is MDY4. DBLOAD converts the date *xxxx-yy-zz* format only if the DBDATE setting is Y4MD; otherwise, an error occurs converting the string to a date.

Regardless of the DBDATE setting, the DBLOAD utility correctly interprets the following:
- Date separator as - or /
- Two-digit rather than four-digit year (MDY4 is equivalent to MDY2 for dates, for example, with the *xx-yy-zz* format)

## Loading an Informix Database in Bulk

### About this task

Using the IBM InfoSphere DataStage Designer client, you can load an Informix database in bulk.

### Procedure

1. Add an Informix Load stage to your InfoSphere DataStage job.
2. Link the Informix Load stage to its data source.
3. Specify column definitions using the **Columns** tab.
4. Add the appropriate property values on the Inputs tab.
5. Compile the job.
6. If the job compiles correctly, you can choose one of the following:
   - Run the job from within Designer.
   - Run or schedule the job using the InfoSphere DataStage Director.
7. If the job does not compile correctly, correct the errors and recompile.

# Link Properties

Informix Load supports the following link properties that are visible from the IBM InfoSphere DataStage Designer client. The following table includes these column heads:

- **Property** is the program-visible name of the property.
- **Prompt** is the text that you see in the stage editor user interface.
- **Default** is the text used if you do not supply any value.
- **Help Text** is added to the table to describe the properties (this text is not visible by using the Informix Load stage).

| Prompt | Type | Default | Description |
|---|---|---|---|
| Table Name | String | None | (Required if Control File is not specified) Table name to load into. |
| Database Name | String | None | (Required) The name of the database on which the target table resides. |
| Server Name | String | None | (Required) The name of the Informix server on which the target table resides. |
| Run DBLOAD | String List | Yes | (Yes, No) Specifies whether to run a local DBLOAD utility automatically or not. If a Informix Load stage job is on a different physical machine than the Informix database, this setting should be set to No. |
| Hex Escapes | String List | Yes | (Yes, No) Recognize hexadecimal escapes in character fields. |
| Table Locking | String List | Yes | (Yes, No) Specifies whether to provide for regular table locking during bulk loading. |
| Exclusive locking | String List | Yes | (Yes, No) Specifies whether to provide for exclusive table locking during bulk loading. |

| Prompt | Type | Default | Description |
|---|---|---|---|
| Field Delimiter | String | None | Field delimiter character. To represent the delimiter itself, precede it by a backslash (\). To represent a backslash, double it. If you use a backslash in your data, whether NLS is off or on, you must populate Field Delimiter in order for a proper conversion to occur. |
| Before-link Routine | String | None | Any valid name of the external process or routine with arguments referring to link property values (for example, *xxx.bat %CTRFILE% %DATAFILE% %ERRLOG%*). The arguments within the percent (%) delimiters correspond to program-visible property names. This process executes before the link activity. |
| After-link Routine | String | None | Any valid name of the external process or routine with arguments, containing link property values (for example, *Postlink.bat %CTRFILE% %DATAFILE% %ERRLOGFILE%*). The arguments within the percent (%) delimiters correspond to program-visible property names. This process executes after completion of all link activity. |

| Prompt | Type | Default | Description |
|---|---|---|---|
| Control File | String | None | If this property contains a valid custom control file name, this file is used, and the Table Name property is ignored. Otherwise, a file *database_table.ctl* is generated (the file name combines the database name and table name property values). |
| Data File | String | None | If this property contains a valid file name, this file is generated. Otherwise, a file *database_table.dat* is generated (the file name combines the database name and table name property values). |
| Errorlog File | String | None | If this property contains a valid file name, this file is generated. Otherwise, a unique file *database_table.log* is generated (the file name combines the database name and table name property values). |
| Directory Name | String | None | Full directory path to place all generated files. If empty, the current directory is assumed. |
| Datetime Format | String | None | Informix DDL-based default DATETIME data type format for all the columns in the table. Individual columns can override this setting in their description attribute. |

| Prompt | Type | Default | Description |
|---|---|---|---|
| Date Format | Long | None | Valid date formats are: 1 - *mm-dd-yyyy* 2 - *dd-mm-yyyy* 3 - *yyyy-mm-dd*<br><br>The DATE type should correspond to the type defined through the DBDATE or GL_DATE environment variables. |
| Maximum Bad Rows | Long | 10 | Maximum errant rows before aborting bulk loading. |
| Rows Batch Size | Long | 100 | Number of rows to process before commit. |
| Rows to Ignore | Long | 0 | Starting rows to ignore. |

# Chapter 5. Informix XPS Load Stage

The Informix XPS Load connectivity stage loads data into an Informix XPS table.

## Functionality of Informix XPS Load Stages

Informix XPS Load has the following functionality:
- Support for data files that exceed the 2-GB file size limit for 64-bit file systems.
- NLS (National Language Support).
- Bulk loading from a stream input link to provide rows of data into the target table.
- Bulk loading by using sequential files and named pipes.
- Immediate and delayed loading.
- Target table creation.
- Express and deluxe load modes.

The following functionality is not supported:
- Output links
- metadata import

## Load Methods

There are two methods of loading data into an Informix XPS table:
- Sequential file method
- Named pipe method

The value for the **Load Method** property determines which method to use to load the data. The value for the sequential file method is **DISK**, and the value for the named pipe method is **PIPE**.

### Sequential File Method

When you load data using the sequential file method, rows from the input link are written in delimited format into an intermediate data file. The data file is loaded immediately or it can be deferred, depending on the value of the **Automatic Load** property.

**Immediate.** When you set **Automatic Load** to **Yes**, the IBM InfoSphere DataStage job automatically loads the table. With automatic loading by using a sequential file, a data file consisting of rows from the input link is constructed and is loaded into the target table after all rows have been read from the input link. Automatic loading by using the sequential file method is slower than the named pipe method because the data cannot be loaded until this data file is constructed.

**Deferred.** When you set **Automatic Load** to **No**, a data file consisting of rows from the input link is constructed, but the data is not loaded into the target table. Additionally, a file containing SQL statements is generated. You can manually edit the data file and the SQL file for advanced customization of the load process and filtering of input data before loading. You can then move these files to any Informix XPS host system for execution.

The data file can be loaded manually later by using the SQL file as input to the *dbaccess* utility. Use this method if the InfoSphere DataStage server does not reside on a system hosting an Informix XPS server or coserver.

## Named Pipe Method

When you load data by using the named pipe method, the rows from the input link are streamed continuously into the named pipe for loading the data until the end of the data. Use this method when you need to load the rows from the input link immediately as they are streamed into the pipe. There is no delay between the time during which the rows are received from the input link and the start of the actual load process.

## Loading a Target Table

Informix XPS Load supports the following three values of the **Load Action** property for loading a target table. For an explanation of how IBM InfoSphere DataStage columns are translated into Informix columns, see "Creating the Target Table" on page 43.

- **TRUNCATE.** This default value deletes all existing rows from the table before starting the load.
- **CREATE.** Informix XPS Load creates the table on the target database. It supports table creation using properties that specify table type, fragmentation strategies, and storage allocation.
- **APPEND.** This value loads its data regardless of whether the target table already contains data. Constraint violations may result. However, any rejected rows are written to a configurable reject file.

For more information about the Load Action property, see "Properties" on page 45.

## Load Modes

The **Load Modes** property allows tables to be loaded in express or deluxe mode. If this property is not specified, the default load mode is based on the table type.

# Data Types

The DATETIME and DATE data types need special consideration.

## DATETIME Considerations using Informix Load Stages

In Informix, a valid DATETIME format can be either *yyyy-mo-dd hh:mi:ss.ff* or any contiguous subset of this string, for example, *mo-dd hh*. Any column defined to contain such a subset would not accept a full ANSI DATETIME.

Therefore, the DATETIME type generated within the data file should correspond to the DATETIME type defined for the actual column in a table. To be consistent with the Informix DDL, you should use the Datetime Format property to specify a default format for all DATETIME columns. Then specify a DATETIME format in the Description field of the column definition grid for every DATETIME column that differs from the default. The following table shows some examples for specifying this property:

Table 7. Descriptions of DATETIME expressions

| DATETIME Column Description Expression | Description |
|---|---|
| Year to Fraction(3) | The default value (does not have to be entered). Full ANSI DATETIME with a precision of 3. |
| Hour to Second | A partial ANSI DATETIME type, starting from hours and ending at seconds. |
| Month to Minute | A partial ANSI DATETIME type, starting from months and ending at minutes. |
| Hour to Fraction(5) | A partial ANSI DATETIME type, starting from hours and ending at fractions with a precision of 5. |

## DATE Considerations using Informix XPS Load Stages

Informix uses the DBDATE or GL_DATE environment variables for interpreting and displaying the string representation of the DATE data type. For example, if DBDATE is set to DMY4, then the *dd-mm-yyyy* string is interpreted as *dd* day of *mm* month of *yyyy* year.

Reading DBDATE or GL_DATE environment variables automatically for date formatting by Informix XPS Load is not an option, since you can run IBM InfoSphere DataStage jobs on a physical machine different from the one where the Informix XPS server is installed.

The default DBDATE Informix setting is MDY4. Informix understands the date *yyyy-mm-dd correctly only if the DBDATE setting is Y4MD. Otherwise, an error occurs when converting a string to a date.* Independent of the DBDATE setting, Informix correctly interprets the following:

- Date separator as - or /
- Two-digit as compared to four-digit year (MDY4 is equivalent to MDY2 dates such as *mm-dd-yy*)

Use the Date Format property to specify the format of date strings that the Informix XPS Load stage sends to Informix. The value of this property should correspond to the value of DBDATE or GLDATE for the target database.

## Creating the Target Table
### About this task

If the value of the **Load Action** property is **CREATE**, Informix XPS Load creates the table on the target database if it does not already exist.

**Note:** Informix XPS Load does not drop the table if it already exists. The job will terminate.

If you do not want automatic loading, the CREATE TABLE SQL statement is written to the SQL file.

The table is created by using the table name specified in the **Table Name** property and the column names specified for the column definitions. If the table has a

primary key, it is created by using the column order indicated by the column definitions. Manually add any additional table attributes such as constraints following the load.

The following table describes how IBM InfoSphere DataStage column data types are translated to Informix column data types:

*Table 8. Data types*

| InfoSphere DataStage Data Types | Informix Data Types |
| --- | --- |
| BigInt | INTEGER |
| Binary | BYTE |
| Bit | Not supported |
| Char | CHAR |
| Date | DATE |
| Decimal | DECIMAL |
| Double | FLOAT |
| Float | FLOAT |
| Integer | INTEGER |
| LongVarBinary | BYTE |
| LongVarChar | VARCHAR |
| Numeric | DECIMAL |
| Real | REAL |
| SmallInt | SMALLINT |
| Time | DATETIME hour to second |
| Timestamp | DATETIME year to fraction (3) |
| TinyInt | SMALLINT |
| VarBinary | BYTE |
| VarChar | VARCHAR |

# Loading an Informix XPS Table

## About this task

By using the InfoSphere DataStage Designer client, you can load an Informix XPS table.

## Procedure

1. Add an Informix XPS Load stage to your InfoSphere DataStage job.
2. Link the Informix XPS stage to its data source.
3. Specify column definitions by using the **Columns** tab.
4. Add the appropriate property values on the **Inputs** tab, as documented in "Properties" on page 45.
5. Compile the job.
6. If the job compiles correctly, you can choose one of the following:
   - Run the job from within Designer.
   - Run or schedule the job by using the InfoSphere DataStage Director.

7. If the job does not compile correctly, correct the errors and recompile.

## Properties

The following table includes these column heads:
* **Prompt** is the text that the job designer sees in the stage editor user interface.
* **Type** is the set of values that are valid for the property.
* **Default** is the text used if the job designer does not supply any value.
* **Description** describes the properties.

Each property is described in the order in which you might use it. You can also access this information from the stage property page in the InfoSphere DataStage Designer.

*Table 9. Informix XPS properties*

| Prompt | Type | Default | Description |
|--------|------|---------|-------------|
| Load Mode | String List | DEFAULT | The method used to load the data into the target file.<br><br>If set to **DEFAULT**, **Create Table Type** determines the loading mode when **Load Action** is set to **CREATE**. Otherwise, it is determined by the type of existing table. If **Load Action** is set to **CREATE**, set **Create Table Type** to a value compatible with the desired loading mode.<br><br>Use express-mode loading for faster loads. Express-mode loading is only possible when **Create Table Type** is set to **RAW** or **OPERATIONAL** and has no indexes.<br><br>Use deluxe-mode loading when **Create Table Type** is set to **OPERATIONAL** or **STANDARD**.<br><br>(DEFAULT/ EXPRESS/DELUXE) |
| Server Name | String | | The name of the Informix XPS server on which the target table resides. |

*Table 9. Informix XPS properties (continued)*

| Prompt | Type | Default | Description |
|---|---|---|---|
| Database Name | String | | The name of the database on which the target table resides. |
| User Name | String | | The user name if you connect to the database as a user other than the user who executes the InfoSphere DataStage server process. If **User Name** is not set, the user name for the InfoSphere DataStage server process is used. |
| Password | String | | If **User Name** is set, the password for this user name. If not set, **Password** is ignored. |
| Table Name | String | | The name of the Informix table to load. |
| Automatic Load | String List | Yes | Specifies whether tables are automatically loaded by the InfoSphere DataStage job. This occurs only when the InfoSphere DataStage server resides on the same system as an Informix coserver.

If set to **No**, the stage generates an external SQL file containing a sequence of SQL commands that can be manually edited and executed on any Informix host system.

(Yes/No) |

*Table 9. Informix XPS properties  (continued)*

| Prompt | Type | Default | Description |
|---|---|---|---|
| Load Method | String List | PIPE | The method of loading data into an Informix XPS table. Named pipe loading (set to **PIPE**) can be faster because data is loaded as it is written to the output stream. Set to **PIPE** only if the InfoSphere DataStage server is running on the same system as an Informix coserver.<br><br>Otherwise, set to **DISK** to retain a copy of the loaded data in a sequential file or if you do not want automatic loading.<br><br>(PIPE/DISK) |
| Load Action | String List | TRUNCATE | Specifies how data should be loaded to the target table. Set to **CREATE** if the target table does not exist. The load fails if the target table already exists. The XPS Load stage will not drop any existing tables on the target database.<br><br>Set to **TRUNCATE** to delete all existing rows before starting the load.<br><br>Set to **APPEND** to add rows to the table. Unique or primary key constraint violations can result.<br><br>(CREATE/ TRUNCATE/ APPEND) |

*Table 9. Informix XPS properties  (continued)*

| Prompt | Type | Default | Description |
|---|---|---|---|
| Create Table Type | String List | STANDARD | Specifies the type of table to create if **Load Action** is set to **CREATE**. The default load mode depends on the type of table being loaded.<br><br>Express mode loads can occur only if set to **RAW** or **OPERATIONAL**.<br><br>This property is ignored if **Load Action** is not set to **CREATE**.<br><br>(STANDARD/RAW/ OPERATIONAL/ STATIC) |
| Fragmentation Strategy | String List | NONE | The fragmentation strategy to use if **Load Mode** is set to **CREATE**. **NONE** specifies that no strategy is used.<br><br>(NONE/ ROUNDROBIN/ EXPRESSION/ HASH/HYBRIDS) |
| Fragmentation Rule | String | | The clause that corresponds to the value for **Fragmentation Strategy**. For example, for a value of EXPRESSION, enter a clause like the following:<br><br>*col_id* < 100 IN dbsp1, *col_id* >= 100 IN dbsp2<br><br>If **Fragmentation Strategy** is set to **NONE**, you can still specify a value for this property to direct table creation in the specified *dbspace* or *dbslice*. |

*Table 9. Informix XPS properties  (continued)*

| Prompt | Type | Default | Description |
|---|---|---|---|
| Extent Size | Long | 16 | The minimum length is four times the disk page size for your system. The maximum length equals the chunk size. |
| Table Lock Mode | String List | EXCLUSIVE | The type of locking for user access to the table during loading. Set to **EXCLUSIVE** so no users can access the table while it is being loaded. Set to **SHARED** so other users can continue to read from the table while it is being loaded.<br><br>(EXCLUSIVE/ SHARED) |
| Directory Path | String | | The path name of the directory where the data file, reject file, and SQL file will be created. If not set, it defaults to the current InfoSphere DataStage project directory. |
| SQL File | String | | The file generated when **Automatic Load** is set to **No**. It contains the sequence of SQL commands required to load an Informix AD/XP table by using high-performance loading from external tables. You can modify and execute this file on any Informix AD/XP host system.<br><br>If not set, a value is generated as<br><br>*<server_name>_ <database_name>_ <table_name>.sql*. It is created in the directory specified by **Directory Path**. |

*Table 9. Informix XPS properties  (continued)*

| Prompt | Type | Default | Description |
|---|---|---|---|
| Data File | String | | Identifies the sequential file created when **Load Mode** is set to **DISK**, or when automatic loading is not desired.<br><br>If not set, a value is generated as for **SQL File** with an extension of *.dat*. |
| Reject File | String | | Identities the reject file created if any rows are rejected during the loading process. These rows are written to this file.<br><br>If not set, a value is generated the same way as for **SQL File** with an extension of *.rej*. |
| Delimiter | String | \| (vertical bar) | The character to be used to delimit fields in the loader input data. This value is also used in the CREATE EXTERNAL TABLE statement. |
| Date Format | Long | 1 | Use one of the following date formats:<br><br>1 - *mm-dd-yyyy* (default)<br><br>2 - *dd-mm-yyyy*<br><br>3 - *yyyy-mm-dd*<br><br>The Date type should correspond to the type defined in the DBDATE environment variable. |

*Table 9. Informix XPS properties (continued)*

| Prompt | Type | Default | Description |
|---|---|---|---|
| Datetime Format | String | Year to Fraction(3) | Valid DATETIME formats mirror Informix DDL syntax for a DATETIME type, for example, Hour to Fraction(2). The DATETIME type should correspond to the DATETIME defined for the actual column in a table. You can override the value defined in the present stage property within a description column option of a column. The default is Year to Fraction(3). |
| Maximum Bad Rows | Long | 10 | Sets the number of errors allowed per coserver before the database server stops the load. |
| Hex Escapes | String List | No | Directs the database server to recognize ASCII special characters embedded in ASCII text-based data files. |
| Before-link Routine | String | | A command to be executed by the host system. It can be an executable, a UNIX shell script, or a DOS batch file. Its format must be understood by the host operating system. The stage does not perform any parsing or syntax checks. |
| After-link Routine | String | | A command to be executed by the host system. It can be an executable, a UNIX shell script, or a DOS batch file. Its format must be understood by the host operating system. The stage does not perform any parsing or syntax checks. |

# 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 10. 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/software/sw-training/ |
| IBM representatives | You can contact an IBM representative to learn about solutions at www.ibm.com/connect/ibm/us/en/ |

## Providing feedback

The following table describes how to provide feedback to IBM about products and product documentation.

Table 11. Providing feedback to IBM

| Type of feedback | Action |
|---|---|
| Product feedback | You can provide general product feedback through the Consumability Survey at www.ibm.com/software/data/info/ consumability-survey |

*Table 11. Providing feedback to IBM  (continued)*

| Type of feedback | Action |
|---|---|
| Documentation feedback | To comment on the information center, click the Feedback link on the top right side of any topic in the information center. You can also send comments about PDF file books, the information center, or any other documentation in the following ways:<br><br>• Online reader comment form: www.ibm.com/software/data/rcf/<br><br>• E-mail: comments@us.ibm.com |

# Accessing product documentation

Documentation is provided in a variety of locations and formats, including in help that is opened directly from the product client interfaces, in a suite-wide information center, and in PDF file books.

The information center is installed as a common service with IBM InfoSphere Information Server. The information 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 the information center from the installed product or from a Web browser.

## Accessing the information center

You can use the following methods to open the installed information center.

- Click the **Help** link in the upper right of the client interface.

  **Note:** From IBM InfoSphere FastTrack and IBM InfoSphere Information Server Manager, the main Help item opens a local help system. Choose **Help > Open Info Center** to open the full suite information center.

- 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.

- Use a Web browser to access the installed information center even when you are not logged in to the product. Enter the following address in a Web browser: http://host_name:port_number/infocenter/topic/ com.ibm.swg.im.iis.productization.iisinfsv.home.doc/ic-homepage.html. The host_name is the name of the services tier computer where the information center is installed, and port_number is the port number for InfoSphere Information Server. The default port number is 9080. For example, on a Microsoft® Windows® Server computer named iisdocs2, the Web address is in the following format: http://iisdocs2:9080/infocenter/topic/ com.ibm.swg.im.iis.productization.iisinfsv.nav.doc/dochome/ iisinfsrv_home.html.

A subset of the information center is also available on the IBM Web site and periodically refreshed at http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r7/ index.jsp.

## Obtaining PDF and hardcopy documentation

- A subset of the PDF file books are available through the InfoSphere Information Server software installer and the distribution media. The other 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.

## Providing feedback about the documentation

You can send your comments about documentation in the following ways:

- Online reader comment form: www.ibm.com/software/data/rcf/
- E-mail: comments@us.ibm.com

# How to read syntax diagrams

The following rules apply to the syntax diagrams that are used in this information:

- Read the syntax diagrams from left to right, from top to bottom, following the path of the line. The following conventions are used:
  - The >>--- symbol indicates the beginning of a syntax diagram.
  - The ---> symbol indicates that the syntax diagram is continued on the next line.
  - The >--- symbol indicates that a syntax diagram is continued from the previous line.
  - The --->< symbol indicates the end of a syntax diagram.
- Required items appear on the horizontal line (the main path).

  ►►—*required_item*———————————————————————————————►◄

- Optional items appear below the main path.

  ►►—*required_item*————————————————————————————————►◄
          └—*optional_item*—┘
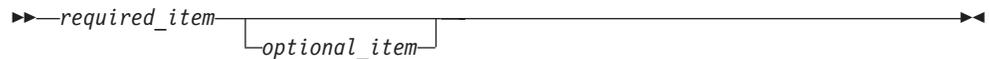
  If an optional item appears above the main path, that item has no effect on the execution of the syntax element and is used only for readability.

          ┌—*optional_item*—┐
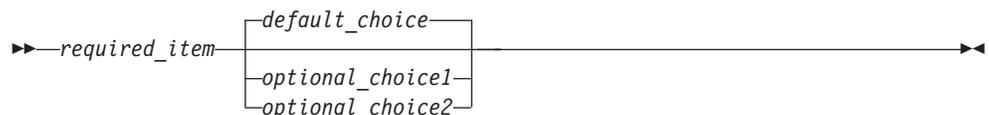  ►►—*required_item*—┴————————┴————————————————————►◄

- If you can choose from two or more items, they appear vertically, in a stack.

  If you must choose one of the items, one item of the stack appears on the main path.

  ►►—*required_item*——┬—*required_choice1*—┬—————————————►◄
                 └—*required_choice2*—┘
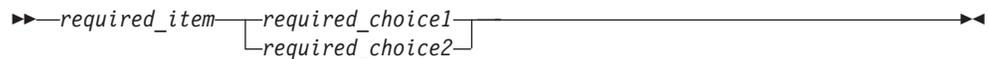
  If choosing one of the items is optional, the entire stack appears below the main path.

  ►►—*required_item*—————————————————————————————————►◄
          ├—*optional_choice1*—┤
          └—*optional_choice2*—┘
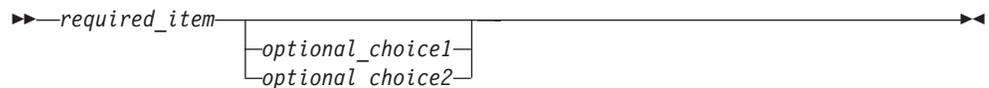
  If one of the items is the default, it appears above the main path, and the remaining choices are shown below.

          ┌—*default_choice*—┐
  ►►—*required_item*—┼————————————————┼——————————————————►◄
          ├—*optional_choice1*—┤
          └—*optional_choice2*—┘

- An arrow returning to the left, above the main line, indicates an item that can be repeated.

      **57**

```
                      ┌───────────────┐
►►──required_item──────┴─repeatable_item─┴──────────────────────────────►◄
```

If the repeat arrow contains a comma, you must separate repeated items with a comma.

```
                      ┌──,─────────────┐
►►──required_item──────┴─repeatable_item─┴──────────────────────────────►◄
```

A repeat arrow above a stack indicates that you can repeat the items in the stack.

- Sometimes a diagram must be split into fragments. The syntax fragment is shown separately from the main syntax diagram, but the contents of the fragment should be read as if they are on the main path of the diagram.

```
►►──required_item──┤ fragment-name ├────────────────────────────────────►◄
```

**Fragment-name:**

```
├──required_item─────────────────────┤
    └─optional_item─┘
```
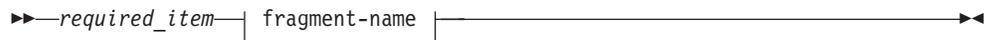
- Keywords, and their minimum abbreviations if applicable, appear in uppercase. They must be spelled exactly as shown.
- Variables appear in all lowercase italic letters (for example, `column-name`). They represent user-supplied names or values.
- Separate keywords and parameters by at least one space if no intervening punctuation is shown in the diagram.
- Enter punctuation marks, parentheses, arithmetic operators, and other symbols, exactly as shown in the diagram.
- Footnotes are shown by a number in parentheses, for example (1).

# 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. The installation program installs the following product modules and components:

- IBM InfoSphere Business Glossary
- IBM InfoSphere Business Glossary Anywhere
- IBM InfoSphere DataStage
- IBM InfoSphere FastTrack
- IBM InfoSphere Information Analyzer
- IBM InfoSphere Information Services Director
- IBM InfoSphere Metadata Workbench
- IBM InfoSphere QualityStage™

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. XHTML allows you to set display preferences in your browser. It also allows you to use screen readers and other assistive technologies to access the documentation.

## IBM and accessibility

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

# Notices and trademarks

This information was developed for products and services offered in the U.S.A.

## 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.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 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.

## Trademarks

IBM, the IBM logo, and ibm.com are 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.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium 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, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office

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

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

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

## C

customer support
   contacting  53

## D

Data Browser
   Informix CLI stages  12, 17
DBLOAD utility  2
dynamic access  1

## I

Informix CLI stages
   functionality  9
   generated queries  18
   generated SQL statements  15
   Informix data source, connecting
    to  10
   input links  2
   Input page  10, 11
    Columns tab  13, 14
    General tab  12
    SQL tab  14
   mapping data types  19
   output links  2
   Output page  10, 16
    Columns tab  17
    General tab  17
    Selection tab  17
   parallel canvas  8
   reading data  18
   reference output links  2
   Stage page  9, 11
    General tab  9
   stored procedure support  20
   user-defined queries  19
   user-defined SQL statements  15
   writing data  15
Informix enterprise stages
   Advanced tab
    Stage page  22, 23, 28
   General tab
    Stage page  22, 23, 28
   Input page  23
   input properties  23, 28
   introduction  2
   NLS Map tab
    Stage page  23
   Output page  28
   parallel canvas  8
   Partitioning tab
    Input page  26
   reading databases  21
   Stage page  22
   writing to databases  21
Informix Load stages
   data types  34
   DATE data type  35
   DATETIME data type  34

Informix Load stages *(continued)*
   DBLOAD utility  2
   functionality  33
   local bulk loading  34
   parallel canvas  8
   properties  36
   remote bulk loading  34
   terminology  33
Informix XPS Load stages
   creating a table  43
   creating a target table  43
   data types  42, 44
   DATE data type  43
   DATETIME data type  42
   functionality  41
   load methods  41
   loading a table  42
   named pipe load method  42
   parallel canvas  8
   properties  45
   sequential file load method  41

## L

legal notices  61

## M

mapping string data  8

## P

parallel canvas  8
product accessibility
   accessibility  59
product documentation
   accessing  55
properties
   Informix enterprise stages input  23,
    28
   Informix Load stages  36

## S

software services
   contacting  53
support
   customer  53

## T

trademarks
   list of  61

## W

web sites
   non-IBM  57

**IBM** ®

Printed in USA