

Implementing Data Masking and Data Subset with IMS Unload File Sources

Abstract

You can use PowerExchange, PowerCenter, and Test Data Management (TDM) to perform data masking and data subset with IMS unload file sources on z/OS. This article describes the procedures for installing and configuring PowerExchange and performing data masking and data subset with these sources.

Supported Versions

- PowerCenter 9.6.0
- PowerExchange 9.6.0
- Test Data Management 9.6.0

Table of Contents

Overview.	3
Preserving Relationships Among IMS Record Types.	3
Multiple-Record Data Maps and Group Source Processing.	3
Sequencing Support in Source and Target Definitions.	4
Multiple-Record Writes to IMS Unload Files.	4
Implementation Scenario.	4
Task Flow.	5
Step 1. Install and Configure PowerExchange on z/OS.	5
Install PowerExchange on z/OS.	5
Edit the DBMOVER Configuration Member on the z/OS System.	6
Configure the PowerExchange Listener JCL on z/OS.	6
Start the PowerExchange Listener.	7
Step 2. Install and Configure PowerExchange on the PowerCenter Services Machine.	7
Install PowerExchange on the PowerCenter Services Machine.	7
Edit the DBMOVER Configuration File on the PowerCenter Services Machine.	7
Step 3. Install and Configure PowerExchange on the PowerCenter Client Machine.	7
Install PowerExchange on the PowerCenter Client Machine.	8
Edit the DBMOVER Configuration File on the PowerCenter Client Machine.	8
Step 4. Define PowerExchange Data Maps for z/OS Sources and Targets.	8
Step 5. Import PowerExchange Source Metadata into PowerCenter.	10
Step 6. Log In to Test Data Manager.	13
Step 7. Define PowerExchange NRDB Batch Connections.	14
PowerExchange NRDB Batch Connections.	14
Creating a Connection.	16
Step 8. Define Data Masking Rules and Policies.	17
Creating a Standard Masking Rule.	17
Creating an Advanced Masking Rule.	18
Creating a Data Domain.	18

Creating a Policy.	19
Step 9. Create a Project and Import Data Sources.	19
Creating a Project.	19
Importing Data Sources.	20
Step 10 (Optional). Run Profiles for Data Discovery.	20
Step 11. Define Groups in Data Subset.	20
Step 12. Assign Data Masking Rules or Policies to the Source Data.	21
Step 13. Create a Data Masking and Data Subset Plan.	21
Step 14. Generate a Workflow.	22
Generating a Workflow.	22
Step 15. Run the Workflow and Monitor the Job.	23

Overview

You can perform data masking and data subset operations on IMS unload file sources.

To implement data masking and data subset for IMS unload file sources, perform the following tasks in PowerExchange, PowerCenter, and TDM:

1. In the PowerExchange Navigator, define a PowerExchange data map for the source.
2. Import the PowerExchange data map into PowerCenter to create a source definition.
3. In Test Data Manager, perform the following tasks:
 - Define PowerExchange NRDB batch connections.
 - Include the source in a project.
 - Create a plan for data masking and data subset.
 - Generate and run a workflow to implement the plan.

Preserving Relationships Among IMS Record Types

If the IMS unload source file includes multiple record types, data masking and data subset operations must preserve the relationships among these record types.

To preserve these relationships, the workflow performs the following tasks:

1. Reads the input file in a single pass.
2. Generates temporary primary key and foreign key columns and sequencing metadata.
3. Writes the output file in a single pass, preserving the relationships among the different record types.

Multiple-Record Data Maps and Group Source Processing

If a data source has multiple record types, the PowerExchange data map defines a record and a table for each unique record type. The table represents the relational view of the related record. To maintain the relationships among IMS record types, PowerExchange uses the segment information that you import from an IMS DBD into the multiple-record data map.

In PowerCenter Designer, you can import a data map with multiple record types to create a single PowerCenter source definition. To include all record types in the source definition, import the data map as a multiple-record data map.

To import the data map as a multiple-record data map, select **Multi-Record Datamaps** in the **Import from PowerExchange** dialog box. If you import a multiple-record data map, the source definition has a section for each table in the data map.

PowerExchange uses group source processing to process multiple-record data maps. When PowerExchange performs group source processing, it reads data from the same physical source in a single pass.

Sequencing Support in Source and Target Definitions

If you enable multiple-record write operations with sequencing when you import a multiple-record data map into PowerCenter, the resulting source or target definition includes special features to support record sequencing.

The following PowerCenter features support sequencing:

- The generated primary key and foreign key columns in the table definitions. The keys contain sequence numbers that uniquely identify each row and its parent.

The source definition includes the following generated columns:

- `DTL_PK_tablename`. Generated for each table. This sequence number serves as a primary key.
- `DTL_FK_tablename`. Generated for each table that corresponds to a non-root record. This sequence number serves as a foreign key.

- The following metadata extensions on the **Metadata Extensions** tab of the **Edit Tables** dialog box:

- Use Sequence Fields. This metadata extension indicates that you selected the **Use Sequence Fields** option in the **Import from PowerExchange** dialog box.
- Hierarchy Structure. This metadata extension is a segment of XML code that defines the hierarchy structure of the tables in a source or target definition.

Multiple-Record Writes to IMS Unload Files

When you generate and run a workflow in Test Data Manager with an IMS unload file source, PowerExchange performs a multiple-record write to the IMS unload file target.

To enable multiple-group writes, select the **Multi-Record Datamaps** and **Use Sequence Fields** options in the **Import from PowerExchange** dialog box for the source definition. When you generate the workflow, the generated target definition uses the same data map as the source definition.

When multiple-record writes are enabled, PowerExchange uses group source processing to read source data and group target processing to write to the target. In addition, PowerExchange uses generated sequence fields to pass metadata about the source hierarchical data structure to the target. The target unload file maintains hierarchical relationships in the same physical sequence as in the source database.

Implementation Scenario

In this scenario, design a data masking and data subset application that uses PowerExchange to read IMS unload file sources with multiple record types on z/OS and write to IMS unload file targets on z/OS.

You have already installed and configured PowerCenter and TDM on the appropriate machines.

The implementation includes the following host and client machines:

- A z/OS machine with IMS unload file data sources.

On this machine, you will install PowerExchange 9.6.0. You will also write to the IMS unload file targets on this machine.

- A Linux, UNIX, or Windows machine on which the Informatica domain and the PowerCenter Integration Service are installed. Optionally, this machine also has Informatica Data Explorer components installed.

On this machine, you will install PowerExchange 9.6.0.

- Windows machine with the PowerCenter Client installed.

On this machine, you will install PowerExchange 9.6.0, including the PowerExchange Navigator. You will run the PowerExchange Navigator on this system to create data maps. You will use the PowerCenter Designer to import these data maps to create source definitions, and you will run the PowerCenter Workflow Manager to define PowerExchange connections and data sources and edit session properties.

- Windows or UNIX machine with the PowerCenter repository 9.6.0 and TDM Server 9.6.0 installed.
- Web browser host from which you log in to Test Data Manager.

Note: You can install PowerCenter on an LPAR running zLinux and use HiperSockets for communications between PowerCenter on one LPAR and the PowerExchange Listener on the LPAR that contains the data source. In this configuration, you must install PowerExchange both on the LPAR with PowerCenter and on the LPAR with the PowerExchange Listener.

Task Flow

To use TDM with IMS unload file sources, complete the following steps:

1. Install and configure PowerExchange on z/OS.
2. Install and configure PowerExchange on the PowerCenter services machine.
3. Install and configure PowerExchange on the PowerCenter Client machine.
4. Define data maps for IMS sources and targets on z/OS.
5. Import PowerExchange source definitions into PowerCenter.
6. Log in to Test Data Manager.
7. Define PowerExchange NRDB Batch connections.
8. Define data masking rules and policies.
9. Create a project to mask specific data sources and to define a subset of the data.
10. Optionally, create and run data masking profiles and import them into the TDM repository.
11. Define groups in data subset.
12. Assign policies and rules to the source data for data masking.
13. Create a plan to run the data subset and masking operations.
14. Generate a workflow from the plan.
15. Run the workflow and monitor the job.

Step 1. Install and Configure PowerExchange on z/OS

Install PowerExchange on the z/OS system where the IMS unload data file source is located. Then edit the DBMOVER configuration file, configure the PowerExchange Listener JCL, and start the PowerExchange Listener.

Install PowerExchange on z/OS

To install PowerExchange on the z/OS system where the source is located, run the z/OS Installation Assistant. For detailed instructions, see the *PowerExchange Installation and Upgrade Guide*. This guide also gives details about access to the z/OS databases and files.

Edit the DBMOVER Configuration Member on the z/OS System

Edit the DBMOVER configuration member on the z/OS machine.

Ensure that the DBMOVER member contains the following statements:

- LISTENER statement. The LISTENER statement defines the TCP/IP port on which a named PowerExchange Listener process listens for work requests. The default DBMOVER member includes the required LISTENER statement.
- IMSID statement. This statement defines the IMS subsystem information that PowerExchange uses for IMS bulk data movement operations.

For more information about the DBMOVER configuration member, see the *PowerExchange Reference Manual*.

Configure the PowerExchange Listener JCL on z/OS

Configure the PowerExchange Listener JCL before starting the PowerExchange Listener the first time.

PowerExchange provides PowerExchange Listener JCL in the following members of the RUNLIB library:

- PWXLSTNR runs PowerExchange as a started task. When you ran the XIZZZ998 job during installation, the PWXLSTNR member was copied to the PROCLIB library.
- STARTLST runs PowerExchange as a batch job.

To configure the PowerExchange Listener JCL:

1. Verify that the STEPLIB DD concatenation in the PowerExchange Listener JCL includes the load libraries for your source DBMSs.

Usually, the PowerExchange LOADLIB library occurs before the data-source load libraries in the concatenation.

The STEPLIB concatenation in the STARTLST member includes the load libraries for all data sources in a proper order.

2. Verify that the PARM statement specifies the node name that is defined in the LISTENER statement of the DBMOVER configuration file.
3. In the REGION parameter of the EXEC statement, enter a region size. The region size indicates the amount of storage to allocate for running the PowerExchange Listener.

The sample JCL specifies a region size of 384 MB in the EXEC statement.

Informatica recommends a region size of 0M, which indicates that z/OS gives the job all the storage available below and above 16 megabytes. For important considerations about specifying a region size of 0M, see the *z/OS MVS JCL Reference*.

If you omit the REGION parameter, the system default applies.

4. Verify that the TIME parameter in the EXEC statement is set to NOLIMIT. The NOLIMIT setting enables the PowerExchange Listener job to run without a time limit and prevents timeout-related abends that might occur with abend code S322.
5. If the SECURITY statement in the DBMOVER configuration file is set to 1 or 2, APF-authorize the PowerExchange LOADLIB library and any other load libraries in the STEPLIB DD concatenation of the PowerExchange Listener JCL.

For more information, see the *PowerExchange Bulk Data Movement Guide*.

Start the PowerExchange Listener

To start the PowerExchange Listener, issue the MVS START command with the name of the started task.

Use the following syntax:

```
START listener_task_name
```

You can also run the PowerExchange Listener as a batch job. However, because it is a long-running task, using an MVS started task is more appropriate.

Step 2. Install and Configure PowerExchange on the PowerCenter Services Machine

Install and configure PowerExchange on the machine where the PowerCenter Integration Service runs. In this scenario, all PowerCenter services, including the PowerCenter Integration Service, reside on the same machine.

Install PowerExchange on the PowerCenter Services Machine

Install PowerExchange on the PowerCenter services machine.

Use the PowerExchange installation instructions for the operating system type and bit-level. If PowerCenter services run on a 64-bit machine, use the PowerExchange 64-bit installation executable or tar file. If PowerCenter services run on a 32-bit machine, use the PowerExchange 32-bit executable or tar file.

For detailed installation instructions, see the *PowerExchange Installation and Upgrade Guide*.

Edit the DBMOVER Configuration File on the PowerCenter Services Machine

Edit the DBMOVER configuration file on the PowerCenter services machine.

Include a NODE statement to define the TCP/IP host name and port that PowerExchange uses to contact the PowerExchange Listener on the z/OS system. Use the following format:

```
NODE=({node_name|node1}
      ,TCPIP
      ,host_name
      ,{port|2480}
      [, {send_bufsize|65536}]
      [, {receive_bufsize|65536}]
      [, {send_size|4096}]
      [, {receive_size|4096}]
      [, receive_timeout]
      [, {SSL|ZOSSL}]
      [, service_name]
      )
```

For more information, see the *PowerExchange Reference Manual*.

Step 3. Install and Configure PowerExchange on the PowerCenter Client Machine

Install and configure PowerExchange on the machine where you installed the PowerCenter Client. You will run the PowerExchange Navigator on this machine.

Install PowerExchange on the PowerCenter Client Machine

Install PowerExchange on the machine where the PowerCenter Client is installed.

Run the PowerExchange 32-bit executable file. You must use the Windows 32-bit executable to install the PowerExchange Navigator, even on a Windows 64-bit machine.

For detailed installation instructions, see the *PowerExchange Installation and Upgrade Guide*.

Edit the DBMOVER Configuration File on the PowerCenter Client Machine

Edit the DBMOVER configuration file on the PowerCenter Client machine.

Include a NODE statement to define the TCP/IP host name and port that PowerExchange uses to contact the PowerExchange Listener on the z/OS system. Use the following format:

```
NODE=({node_name|node1}
      ,TCPIP
      ,host_name
      ,{port|2480}
      [, {send_bufsize|65536}]
      [, {receive_bufsize|65536}]
      [, {send_size|4096}]
      [, {receive_size|4096}]
      [, receive_timeout]
      [, {SSL|ZOSSL}]
      [, service_name]
      )
```

For more information, see the *PowerExchange Reference Manual*.

Step 4. Define PowerExchange Data Maps for z/OS Sources and Targets

Create a PowerExchange data map for each IMS unload file source.

Each data map must be available to the source and corresponding target. If a target is on the same z/OS system as the source, you can use the same data map for the source and target. Otherwise, send the data map to both the source and target systems.

1. Start the PowerExchange Navigator.
2. On the **Resources** tab in the **Resource Explorer**, click **Add > Data Map**.

3. In the **Name** dialog box, enter the following information:

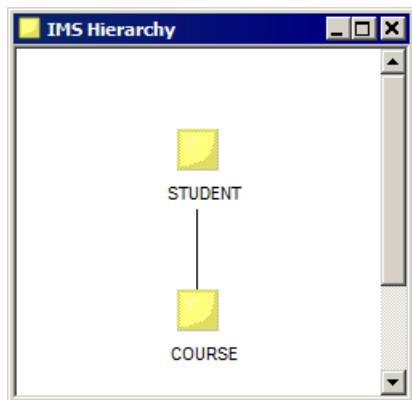
Option	Description
Schema Name	Enter a user-defined schema name. PowerExchange uses this name as the first part of the data map name, as follows: <i>schema_name.data_map_name</i> The name must consist of alphanumeric characters and begin with a letter. Maximum length is 10 characters.
Data Map Name	Enter a user-defined map name. PowerExchange uses this name as the second part of the data map name, as follows: <i>schema_name.data_map_name</i> The name must consist of alphanumeric characters and begin with a letter. Maximum length is 10 characters.
Access Method	Select one of the following access methods: - DL/1 BATCH . IMS database on z/OS. - IMS ODBA . IMS database accessed through a PowerExchange Listener on z/OS.
Import Record Definitions	Select this option to import DBD and copybook metadata into the data map. Clear this option to add only the data map. By default, this option is selected. For IMS sources, accept the default.

4. In the **Data Map Properties** dialog box for the DL/1 Batch or ODBA access method, enter the IMS SSID, DBD Name, and any additional information that is required for the access method.
5. Click **Next**, and provide the required information in the dialog boxes to import DBD metadata and then copybook metadata for each segment that is defined in the DBD.
6. For each data map, click **File > Send to Remote Node**, and complete the **Data Map Remote Node** dialog box to send the data map to the PowerExchange Listener on the source system.

If different Listeners will write to the source and target, also send each data map to the Listener on the target system.

After you define a data map, you can view the hierarchy. On the **Data Map** tab in the **Resource Explorer**, right-click a data map and click **Display IMS Hierarchy**.

The IMS hierarchy for the IMS data map appears. The following image shows an example IMS hierarchy:



For more information about creating data maps, see the *PowerExchange Navigator User Guide*.

Step 5. Import PowerExchange Source Metadata into PowerCenter

When you import PowerExchange data maps into PowerCenter to create source definitions, the source definitions become available for import into Test Data Manager.

After you connect to a source database from PowerCenter, you can import source metadata into the PowerCenter repository. You must create a folder in the PowerCenter repository to store source definitions and metadata that you import into the PowerCenter repository. When you import metadata from PowerCenter into the TDM repository, Test Data Manager creates a copy of the PowerCenter folder in the TDM repository.

You do not need to import PowerExchange target metadata into PowerCenter. When you generate a workflow from the TDM Workbench, the required target definitions are generated automatically.

1. In the PowerCenter Designer, connect to a PowerCenter repository.
2. Create a folder for the source that you are importing.
3. Click **Sources > Import from PowerExchange**.

The **Import from PowerExchange** dialog box appears. The following figure shows this dialog box:

The screenshot shows the 'Import from PowerExchange...' dialog box. The title bar includes a close button (X). The main area is divided into several sections:

- Connection Information:** Location: MERCURYZOS (dropdown), User name: DEMO_USER (text field), Password: ***** (text field), and a Connect button.
- Options:** Multi-Record Datamaps (checked), Source Type: IMS (dropdown), Use Sequence Fields (checked), Network Timeout: 180 (text field), and CDC Datamaps (unchecked).
- Metadata Filter:** Schema: (text field), Map name: (text field), and a Respect Case checkbox (unchecked).
- Select Datamaps:** A large empty list box with Select all and Select none buttons.

On the right side of the dialog, there are three buttons: OK, Cancel, and Help.

4. Enter the following information in the dialog box:

Attribute	Required/Optional	Description
Location	Required	Name of the PowerExchange Listener (node name from PowerExchange DBMOVER member) on the system on which the database resides.
User Name	Required	User name for the database to connect to.
Password	Required	Password associated with the user name.
Multi-Record Datamaps	Required	Lists multiple-record data maps. When you use data masking or data subset with IMS unload file sources, you must select this option.
Use Sequence Fields	Required	Specifies whether to generate sequence fields. When you use data masking or data subset with IMS unload file sources, you must select this option.
Source Type	Required	Data source type. Select IMS .
CDC Datamaps	n/a	Displays CDC extraction maps only. Clear this check box.
Network Timeout	Optional	Number of seconds that elapse during a network send or receive operation before PowerExchange ends the connection and issues a timeout error message.
Schema	Optional	Schema name to use to filter the resulting data maps. You can enter name patterns by using the following wildcard characters: - * (asterisk). Represents one or more characters. - ? (question mark). Represents one character.
Map name	Optional	Data map name to use to filter the resulting data maps. You can enter name patterns by using the following wildcard characters: - * (asterisk). Represents one or more characters. - ? (question mark). Represents one character.
Respect Case	Optional	If this attribute is selected, PowerExchange filters data maps by using the exact case of the Schema and Map name fields. Otherwise, PowerExchange ignores the case when searching for data maps.
Selected Datamaps	n/a	Lists the available data maps for the connection, source type, and filter details that you entered.

5. Click **Connect**.

The available schemas appear in the **Selected Datamaps** list.

6. Expand the schema to display all the tables for all the data maps in the schema, and select the tables that you want to import. You can import multiple tables from a single data map or from multiple data maps.

7. Click **OK**.

The PowerCenter source definitions are listed and have names in the format *mapname_tablename*.

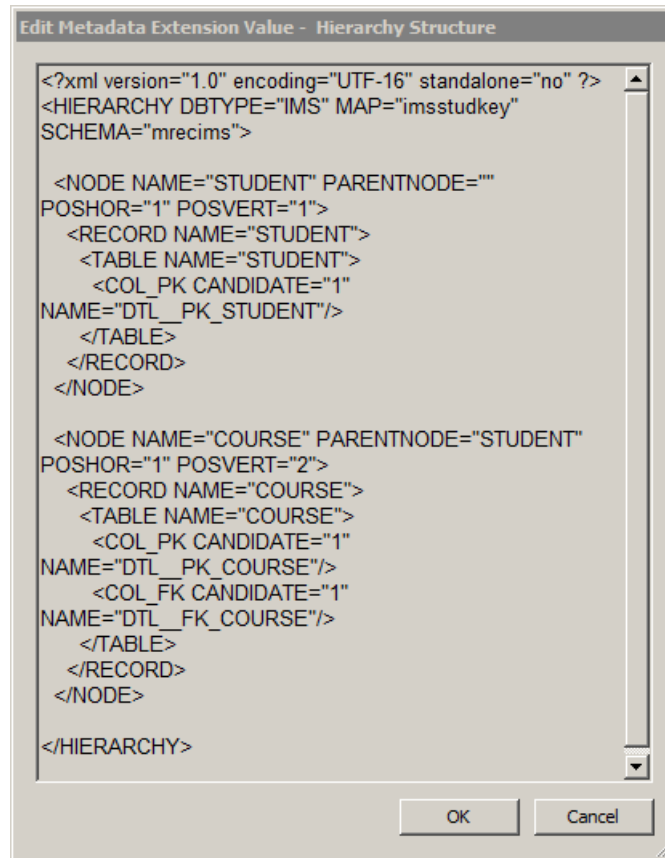
The following example shows the source definition for the COURSE-STUDENT hierarchy:

Key Types	Name	Datatype	Length/Precision
COURSE			
NOT A KEY	ST_COURSE_CODE	UZONED	5
NOT A KEY	ST_COURSE_HOURS	UZONED	2
NOT A KEY	ST_COURSE_TIME	CHAR	5
NOT A KEY	ST_COURSE_DAY	CHAR	5
NOT A KEY	ST_COURSE_INSTR	CHAR	15
NOT A KEY	ST_COURSE_BLDG	UZONED	2
NOT A KEY	FILLER_1	CHAR	16
PRIMARY KEY	DTL_PK_COURSE	NUM64	19
FOREIGN KEY	DTL_FK_COURSE	NUM64	19
STUDENT			
NOT A KEY	CHAR_KEY	CHAR	10
NOT A KEY	ST_NAME	CHAR	40
NOT A KEY	ST_ADDRESS_1	CHAR	40
NOT A KEY	ST_ADDRESS_2	CHAR	40
NOT A KEY	ST_CITY	CHAR	30
NOT A KEY	ST_STATE	CHAR	2
NOT A KEY	ST_ZIP	CHAR	9
NOT A KEY	ST_NUMBER	UZONED	12
NOT A KEY	ST_BIRTH_MM	UZONED	2
NOT A KEY	ST_BIRTH_DD	UZONED	2
NOT A KEY	ST_BIRTH_YY	UZONED	2
NOT A KEY	ST_SEX	CHAR	1
NOT A KEY	ST_HEIGHT	UZONED	2
NOT A KEY	ST_WEIGHT	UZONED	3
NOT A KEY	ST_HAIR	CHAR	5
NOT A KEY	ST_EYES	CHAR	5
NOT A KEY	ST_DATE_ENROLL_MM	UZONED	2
NOT A KEY	ST_DATE_ENROLL_YY	UZONED	2
NOT A KEY	ST_DATE_GRAD_MM	UZONED	2
NOT A KEY	ST_DATE_GRAD_YY	UZONED	2
NOT A KEY	ST_TUITION_FEES	NUM32	9
NOT A KEY	ST_C_CNT	UZONED	3
PRIMARY KEY	DTL_PK_STUDENT	NUM64	19

When you select **Multi-Record Datamaps**, the source definition includes multiple groups. When you select **Use Sequence Fields**, the source definition includes generated primary key and foreign key fields.

- Optionally, to view the XML that describes the hierarchy structure, double-click the source definition. On the **Metadata Extensions** tab of the **Edit Tables** dialog box, click the arrow in the **Value** column for **Hierarchy Structure**.

The Hierarchy Structure dialog box appears. The following image shows an example hierarchy structure:



When you generate a workflow, the Hierarchy Structure metadata extension provides the metadata to maintain the relationships between the segments in the source data as it passes through PowerCenter transformations.

For more information about importing PowerExchange source definitions into PowerCenter, see *PowerExchange Interfaces for PowerCenter*.

Step 6. Log In to Test Data Manager

Test Data Manager is a web-based user interface that you use to complete data discovery, data subset, and data masking operations.

- In the address bar of a web browser, enter the Test Data Manager URL.
 - Use the following format if Transport Layer Security (TLS) is enabled:
`https://hostname:portnumber/tdm/`
 - Use the following format if TLS is not enabled:
`http://hostname:portnumber/tdm/`

Where:

- *hostname* is the host name or IP address of the machine where you installed the TDM Server.
- *portnumber* is the port number. The default is 6643 if TLS is enabled. The default is 6605 if TLS is not enabled.

For example, you might enter the following URL:

```
http://TXW1779:6643/tdm/
```

The **Login** dialog box of Test Data Manager appears.

2. Enter a user name and password that are defined in Informatica Administrator.
Select the security domain. If the Informatica domain is configured to use LDAP authentication, the default security domain **Native**.
3. Click **Login**.
Test Data Manager appears.

Step 7. Define PowerExchange NRDB Batch Connections

Define PowerExchange NRDB batch connections in Test Data Manager.

Define separate connections for the source and target, even if you connect to them using the same PowerExchange Listener.

When you generate a workflow, you can select these source and target connections.

Before you define source and target connections, verify that the connection to the PowerCenter repository is defined in Test Data Manager.

Note: Instead of defining the connections in Test Data Manager, you can define the connections in PowerCenter and import them into the TDM repository.

PowerExchange NRDB Batch Connections

You can create a PowerExchange NRDB Batch connection in Test Data Manager to perform data subset and data masking operations.

Alternatively, you can create a PowerExchange NRDB connection by using PowerCenter Workflow Manager, and then import the connection into the TDM Repository.

The following table describes the database connection properties for a PowerExchange NRDB Batch connection:

Property	Description
Name	Required. Name of the connection. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters, start with a number, contain spaces, or contain the following special characters: ~`!\$%^&*()-+={[] \:;'"<, > . ? /
Connection Type	Required. The connection type. Select PWX NRDB Batch.
Description	The description of the connection. The description cannot exceed 255 characters.
User Name	Required. The database user name.
Password	Required. The password for the database user name.

Property	Description
Code Page	The code page that the Integration Service uses to extract data from the source database. Note: In Unicode mode, PWXPC sets the code page with this value overriding any code page specification in the PowerExchange configuration file.
Array Size	Storage array size, in number of records, that is used for partitioned or multithreaded sessions. For partitioned sessions, this array size is shared across the partitions. Valid values are from 1 through 100000. Default is 25. Note: To tune partitioned sessions, particularly when the Write Mode attribute specifies Confirm Write On, increase the array size.
Write Mode	Mode in which the Integration Service sends data to the PowerExchange Listener. Configure one of the following write modes: <ul style="list-style-type: none"> - Confirm Write On. Sends data to the PowerExchange Listener and waits for a response before sending more data. Select if error recovery is a priority. This option might decrease performance. - Confirm Write Off. Sends data to the PowerExchange Listener without waiting for a response. Use this option when you can reload the target table if an error occurs. - Asynchronous With Fault Tolerance. Sends data to the PowerExchange Listener without waiting for a response. This option also provides the ability to detect errors. This provides the speed of Confirm Write Off with the data integrity of Confirm Write On. Default is Confirm Write On.
Off Load Processing	Controls whether to use offload processing. Offload processing transfers PowerExchange bulk data processing from the source system to the Integration Service machine. Select one of the following values: <ul style="list-style-type: none"> - No. Do not use offload processing. - Auto. PowerExchange determines whether to use offload processing. - Filter After. Offloads the bulk data processing to the PowerCenter Integration Service machine, including the filtering of data. - Filter Before. Offloads processing to the PowerCenter Integration Service machine but continues to filter data on the source system. If you select Filter Before for an IMS unload file data source, PowerExchange changes the selection to Filter After. Default is No.
Worker Threads	The number of threads that PowerExchange uses on the Integration Service machine to process bulk data. For optimal performance, this value should not exceed the number of installed or available processors on the Integration Service machine. Valid values are from 1 through 64. Default is 0, which disables multithreading. If you use reader or writer pipeline partitioning, accept the default value of 0. You cannot use both multiple worker threads and partitioning.
Connection Retry Period	Number of seconds the Integration Service attempts to reconnect to the data source or target if the connection fails. If the Integration Service cannot connect to the database in the retry period, the operation fails. Default is 0.
Compression	Select to compress source data during the PowerCenter session. Default is disabled.
Encryption Type	The type of encryption that the Integration Service uses. Select one of the following values: <ul style="list-style-type: none"> - None. - RC2. - DES. Default is None.

Property	Description
Encryption Level	The level of encryption that the Integration Service uses. If you select RC2 or DES for Encryption Type, select one of the following values to indicate the encryption level. <ul style="list-style-type: none"> - 1. Uses a 56-bit encryption key for DES and RC2. - 2. Uses a 168-bit triple encryption key for DES. Uses a 64-bit encryption key for RC2. - 3. Uses a 168-bit triple encryption key for DES. Uses a 128-bit encryption key for RC2. Ignored if you do not select an encryption type. Default is 1.
Interpret As Rows	Interprets the pacing size as rows or kilobytes. Select to represent the pacing size in number of rows. If you clear this option, the pacing size represents kilobytes. Default is disabled.
Convert Character Data As String	Convert character fields to string fields so that embedded nulls in data are processed as spaces. Default is disabled.
PWX Override	PowerExchange connection override values, separated by semicolons. You can enter any of the following overrides: <pre>TCPIP_OP_TIMEOUT=network_operation_timeout; TCPIP_CON_TIMEOUT=connection_timeout; WRT_ERROR_HANDLING=Y</pre> If you use write partitioning and set the Write Mode attribute to Confirm Write On, you can specify <code>WRT_ERROR_HANDLIN=Y</code> to use an alternative method of error handling that is more efficient than the default method when the input data contains many errors. You can also specify this statement in the <code>dbmover.cfg</code> configuration file.
Pacing Size	Amount of data that the source system can pass to the PowerExchange Listener. Configure the pacing size if an external application, database, or the Integration Service node is a bottleneck. The lower the value, the faster the performance. Enter 0 for maximum performance. Default is 0.
Location	Required. Location of the PowerExchange Listener node. The location is defined in the first parameter of the NODE statement in the PowerExchange <code>dbmover.cfg</code> configuration file.
Get PowerExchange Log Entries	Specifies whether PWXPC writes PowerExchange messages related to a CDC session in the PowerCenter session log. Default is disabled.

Creating a Connection

You can create a source or target connection in Test Data Manager.

1. In the **Administrator | Connections** view, select **New Connection** from the **Actions** menu.
A tab displays the properties that you can set for a new connection.
2. Select the connection type and define the connection name, description, and user information.
3. Click **Next**.
4. Enter the connection properties.
5. Click **Test Connection** to test the connection.
6. Click **Finish** to save the connection.

The connection is visible in the **Administrator | Connections** view.

7. Select the connection in the **Administrator | Connections** view, and click **Actions > Test Connection** to validate the connection.

Note: Alternatively, you can create a PowerExchange NRDB connection by using PowerCenter Workflow Manager, and then import the connection into the TDM Repository. For more information, see the following guides:

- *PowerExchange Interfaces for PowerCenter*
- *Test Data Management Administrator Guide*

Step 8. Define Data Masking Rules and Policies

Use data masking to replace source data in sensitive columns with realistic test data for non-production environments. When you create data masking rules, you define the logic used to replace sensitive data. To specify the sensitive columns that you want to mask, assign data masking rules to source columns and policies.

A policy is a collection of data masking rules. When you assign rules to policies, you can assign multiple source columns to the data masking rules. You can also assign a rule directly to a source column. You can assign data masking rules based on the datatype of the source columns.

Complete the following high-level steps to define policies:

1. Create data masking rules.
2. Define data domains to describe which columns can receive the same masking rules.
3. Assign the data masking rules to the data domains.
4. Create a policy and assign the data domains to the policy.

Creating a Standard Masking Rule

Create a rule to define a masking technique, the datatype to mask, and masking parameters that define how to apply the technique.

1. To access the **Policies** view, click **Policies**.
2. Click **Actions > New > Masking Rule**.
The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.
4. Select the datatype of the column to apply the masking rule to.
5. Select the Standard masking rule.
6. To enable users to override masking parameters for a rule, select the **Override Allowed** option.
7. Click **Next**.
Note: The **Masking Parameters** dialog box changes based on the **Masking Technique** you select.
8. Enter the masking parameters.
9. Enter the exception handling options. Configure how to handle null or empty spaces. Configure whether to continue processing on error.
10. Click **Finish**.

Creating an Advanced Masking Rule

Create an advanced masking rule to combine more than one masking technique or to mask multiple columns.

1. To access the **Policies** view, click **Policies**.
2. Click **Actions > New > Masking Rule**.
The **Rule Wizard** appears.
3. Enter a name and optional description for the rule.
4. Select the Advanced masking rule.
5. Click **Next**.
6. In the **Input Columns** area, click **Add New**.
The **Add Column** dialog box appears.
7. Enter the column properties, such as the name, datatype, precision, and scale. Select whether the column is mandatory to assign to a data source in all projects.
Do not enter masking properties for input columns.
8. To enter more input columns, click **Add Input Column** in the **New Masking Rule** dialog box.
9. Add temporary variable columns
You must apply an expression, masking rule, or dependent column for each variable column you define. If you apply a masking rule, configure the input column to create the variable column from.

Creating a Data Domain

When you create a data domain, you can enter regular expressions that describe the data that you want to include in the data domain. You can also enter regular expressions that describe the names of database columns to include.

1. To access the policies, click **Policies**.
The **Policies** view shows a list of the policies, data domains, and rules in the TDM repository.
2. Click **Actions > New > Data Domain**.
3. Enter the name, sensitivity level, and description for the data domain. Click **Next**.
4. Click **Next**.
5. Optionally, enter a regular expression to filter columns by data pattern.
6. To add more expressions for data patterns, click the **+** icon.
7. Click **Next** to add regular expressions that filter columns by column name. Or, click **Finish** to skip entering any more data domain information.
You can add multiple expressions.
8. Enter regular expressions to filter columns by column name.
9. Click **Next** if you want to apply preferred masking rules to the data domain. Or, click **Finish** to finish configuring the data domain.
10. To add preferred masking rules to the data domain, click **Add Rules**.
The **Add Rules** dialog box appears.
11. Select the data masking rules to add.
12. Click **OK**.
13. Enable one rule as the default rule.
14. Click **Finish**.

Creating a Policy

Create a policy in the **Policies** view. You can create the policy before you create data domains and rules. You can add the data domains and rules to the policy at any time.

1. In the **Policies** view, click **Actions > New > Policy**.
The **New Policy** dialog box appears.
2. Enter a name and optional description for the policy, and click **Next**.
3. To add data domains to the policy, click **Add Data Domains**.
4. Select the data domains from the list.
5. Click **Finish**.
The policy appears in the **Policies** view.

Step 9. Create a Project and Import Data Sources

Create a project folder to contain the data discovery, subset, and masking components, and import data sources into the project.

Creating a Project

Create a project folder to contain the data discovery, subset, and masking components to apply to a data source.

1. In Test Data Manager, click **Projects** to access the projects.
A list of projects appears.
2. Click **Actions > New**.
3. In the **Create Project** dialog box, enter project properties. The following table describes the project properties:

Option	Description
Name	The name of the project.
Description	The description of the project.
PowerCenter Repository	The name of the PowerCenter repository to store the repository folder.
Folder	The name of the project folder in the repository. Default is the project name. You can choose another folder in the repository.
Owner	The name of the user that owns the folder. The folder owner has all permissions on the folder. The default is the name of the user that created the folder. You can select another user as the folder owner.

4. Click **OK**.
The properties for the project appear in Test Data Manager.

Importing Data Sources

Import data sources from PowerCenter into a project in Test Data Manager.

1. To view a list of the projects in Test Data Manager, click **Projects**.
2. Click a project in the list to open it.
3. Click **Actions > Import Metadata**.
The **Import Metadata** window appears.
4. To import metadata from PowerCenter, choose **PowerCenter Repository**.
5. Choose whether you want to review the metadata changes before you import the data sources. Test Data Manager displays the information before the import runs.
Test Data Manager shows you the metadata that you are importing and the impact on rules and domain assignments or entities and groups in Test Data Manager.
6. Click **Next**.
7. Select the PowerCenter folder that contains the data source you want to import. You can filter folders by the folder name or the description.
8. Click **Next**.
9. Select the tables to import. You can filter the tables by data source, table name, or table description.
10. Click **Next**.
11. Choose when to import the sources. Choose one of the following options:
 - **Import Now**. Import the data source immediately. To run the import in the background, select **Run import in the background**.
 - **Schedule Later**. Schedule the import to occur at a specific date and time. Click the calendar to select a date. Use the hour and minute sliders to set the time.
12. Click **Finish**.
If the import is running, view the progress of the import job in the **Monitor** view. After the job finishes, access the imported metadata through the **Data Sources** details view.

Step 10 (Optional). Run Profiles for Data Discovery

You cannot use Test Data Manager to run a profile for data discovery on nonrelational sources. However, you can create and run profiles in Informatica Data Explorer and import them into the TDM repository.

Complete the following high-level steps to perform data discovery:

1. Create profiles in Informatica Data Explorer.
2. Run the profiles in Informatica Data Explorer.
3. Export the results from Informatica Data Explorer.
4. Open Test Data Manager, and import the results to the TDM repository.

Step 11. Define Groups in Data Subset

Define groups for a data subset project. After you create a plan, you can add the groups to the plan.

Note: You can also add entities and templates to a data subset project. However, these objects are of limited usefulness for nonrelational sources.

To create a group, select the tables that you want to add to a data subset plan.

1. On the **Define | Data Subset** view in the project, click **Actions > New > Groups**.
2. In the **Create Group** dialog box, enter a name for the group and optional description for the group. Click **Next**.
3. To select one or more tables for the group, click **Add Tables**.
4. Optionally, you can filter the list of tables to search for.
5. Click **Save**.

Step 12. Assign Data Masking Rules or Policies to the Source Data

To implement data masking operations, assign masking rules from data domains or policies to one or more columns in the project source that you want to mask.

You can assign a rule to a column through a data domain default rule or by choosing a rule. If the column belongs to a data domain, the default rule appears in the column assignment.

1. In the project, click **Define | Data Masking**.
The **Data Masking** view appears.
2. Select a column to assign a masking rule to.
3. If the **Domain** field is blank for the column, click the **Policy** column and choose a policy that contains the data masking rule that you want to assign.
4. Click inside the **Rule** column to view the list of available rules.
The data domain preferred rules appear at the top of the list. The other rules in the policy appear at the bottom of the list.
5. Select a masking rule.
If you choose a substitution masking rule, you can specify the rule assignment parameters. If you choose a maplet or advanced rule, you can assign the rule columns to the columns in the project.
6. Click **Save** for each column that you update.

Step 13. Create a Data Masking and Data Subset Plan

Create a plan to run the data subset and data masking operations. You can combine a data subset and a data masking operation in the same plan or create separate plans.

When you create a plan, add components to it to define its operations. Add groups to complete data subset operations. Add rules and policies to plans to perform data masking operations.

For detailed information about Test Data Management plans and workflows, see the *Informatica Test Data Management User Guide*.

1. Open a project and click **Execute** to view the project plans.
2. Click **Actions > New**.
3. In the **New Plan** dialog box, enter a name and optional description for the plan.
4. To add a data masking operation to the plan, click **Add Masking Components**.
5. Select the policies and rules to add to the plan. Click **Next**.
6. To add a data subset operation to the plan, click **Add Subset Components**.
7. Select the entities, groups, and templates to add to the plan. Click **Next**.
You can assign criteria for masking columns and generating subsets.
8. To skip adding a data generation component, click **Next**.

9. To skip masking a rule, select the check box for the rule and click **Off**.
10. To filter subset components, select the component and choose to enter a basic or advanced expression. Test Data Manager shows any existing filter criteria, which you can change.
11. To limit the subset results, click **Limit** and choose to limit by percentage, by absolute value, or by defining an interval of rows to create. Click **Next**.
12. Configure the connections and other properties. Click **Next**.
Be sure to define the following connection options:
 - **Source File Name**
 - **Target File Name**
13. To override plan settings for a data source, select the data source and click **Override Plan Settings**. Then enter the override properties.
14. To override data source settings for a table, select the table and click **Override Data Source Settings**. Then enter the override properties.
15. Click **Finish**.
The plan appears in the project.

Step 14. Generate a Workflow

In Test Data Manager, generate a workflow from a plan, and then run the workflow to perform subset and masking operations.

When you generate a workflow for an IMS unload file source, the PowerCenter Repository Service generates the following mappings for the workflow:

- The first mapping reads the unload file source and populates comprehensive filter tables.
- The second mapping uses the filter tables from the first mapping as lookup tables and populates subset filter tables.
- The third mapping reads the unload file source a second time and filters segments using the set of filter tables from the second mapping. PowerExchange resequences the segment data so that it will appear in the same order in the output as it would have appeared in the input file. PowerExchange then creates the output multiple-record unload file. For IMS, this file also includes the necessary header and trailer records.

When you start a workflow, the PowerCenter Integration Service performs the data masking and data subset tasks defined in the plan.

You can view the status of generated workflows or the load configuration on the **Monitor** view.

Generating a Workflow

After you create a plan, generate a workflow. The PowerCenter Repository Service generates the mappings to include in the workflow.

1. In a project, click **Execute** to access the plans in the project.
2. Select a plan from the list.
3. Click **Actions > Generate Workflow**.
4. Choose whether to generate the workflow immediately or to schedule workflow generation for later.

Step 15. Run the Workflow and Monitor the Job

After you generate a workflow, you can execute the workflow to run sessions from the PowerCenter mappings.

You can start a workflow from the **Plan | Properties** view or the **Plan | Workflow** view.

1. If the plan you selected contains one workflow, click **Actions > Execute Workflow** in either view.
2. If the plan you selected contains multiple workflows, click the **Workflow** view.
 - a. Select the workflow that you want to start.
 - b. Click **Actions > Execute Workflow**.
3. View the workflow status in the **Workflow Execution** view.

Author

Jim Middleton