Informatica Comparison Utility

This document includes the following topics:

- Overview, 2
- Installing the Comparison Utility, 5
- Creating Repository Definitions, 7
- Creating a Baseline, 9
- Configuring Jobs, 12
- Configuring Properties, 17
- Working with Reports, 23
- Viewing Logs, 28
- Glossary, 30
Overview

Use the Informatica Comparison Utility to compare target data and repository metadata. You can make comparisons across different repository versions or within the same version. If you compare data or metadata in one repository version, you can compare within one repository or across two repositories.

Create a repository definition for each repository that you want to compare. The repository definition contains connection information to the repository.

To compare a set of data or metadata more than once or to compare data or metadata within the same repository, you can create a baseline. When you configure a baseline, you specify all the targets and metadata that you want to use as a baseline comparison. When you populate the baseline, the Informatica Comparison Utility saves the metadata and target data to a file.

Configure a job to compare data or metadata or both. Job definitions contain information about the targets and metadata you want to compare. For each job, select a base and a comparison repository. You can select a repository definition or a baseline. You also choose targets and metadata that exist in each repository.

When you run a job, the Informatica Comparison Utility runs workflows and gets metadata from the base repository first. If you use a baseline as the base repository, the Comparison Utility gets the data and metadata from the populated baseline. Then, it runs workflows and gets metadata from the comparison repository. It generates a report that shows the comparison.

After you run a job, you can view job logs, comparison reports, and session logs.
Figure 1-1 shows the Comparison Utility window:

Comparing Data and Metadata

This section provides an overview about the types of data and metadata that you can compare.

Target Data

You can compare target data for a target that exists in each repository. In addition to reporting on target data, the Comparison Utility provides information such as the number of differences in the targets, the number of rows compared, and the number of matched rows.

You can compare IBM DB2, Oracle, Teradata, Microsoft SQL Server, and flat file targets.

Metadata

You can compare metadata for mappings, sessions, and workflows. The Comparison Utility retrieves the validation state and the object attributes.

Usage

This section provides some usage examples based on the repository versions you want to compare.
Two Repositories of the Same Version
You might want to compare target data against the same version of a repository when you migrate from development to production. For example, if your development environment uses Microsoft SQL Server and your production environment uses Oracle, you can deploy a workflow to production and run a job to verify that the target results are the same.

One Repository
You might want to compare target data and metadata within one repository when want to update mapping or session configuration. For example, you want to verify that target data does not change when you enable pushdown optimization. Create a baseline against a workflow before you enable pushdown optimization. Then, enable pushdown optimization and run a job using the baseline and the updated workflow.

Two Repositories of Different Versions
You might want to compare target data and metadata when you upgrade to another version of PowerCenter. For example, you upgrade from 7.1.1 to 8.1.1, and you want to find out if the changes to Decimal and Double processing produces different results in the upgraded workflows. You might also want to compare metadata to see the changes to query parameters from version 7.1.1 to 8.1.1.
Installing the Comparison Utility

Install the Comparison Utility on a Windows machine that has access to the PowerCenter repositories, PowerCenter Servers, and Integration Services in your environment.

To install the Comparison Utility:

1. Close all applications before running the installation.
2. Click setup.exe in the installation folder to begin installation. The Informatica Comparison Utility Setup Wizard appears.
3. Click Next.
4. Click Next to accept the default location.
Optionally, click Browse or specify the location where you want to install.

5. Click Next.

6. Select the Start Menu folder to specify shortcuts.

7. Click Install.

8. Click Next when the setup is complete.

9. Click Finish to close the wizard.

Starting the Informatica Comparison Utility

To launch the Informatica Comparison Utility from the Windows Start menu, click Start > Programs > Comparison Utility > Informatica Comparison Utility.
Creating Repository Definitions

A repository definition contains connection information for a repository that you want to compare. When you configure a job or baseline, you must provide a repository definition.

To create a repository definition:
1. Launch the Comparison Utility.
2. Click Repository > New.

The Repository Connection Information dialog box appears.

Figure 1-2 shows the repository connection dialog box:

Figure 1-2. Repository Connection Information Dialog Box

3. Configure the connection information for the repository.
4. Click Test Connection to test the connection.
5. Click OK.
A summary of the repository connection information is displayed on the Information tab. Figure 1-3 shows the Information tab for creating a repository definition:

Figure 1-3. Repository Information Tab

Comparison Utility - v1.0 beta

Repository Information
name: cent012_mnu斟11
version: v18
gateway host: cap162088
Gateway port: 8001
domain: Domain_cap162088
username: Administrator
Creating a Baseline

A baseline is a set of repository metadata and target data that you want to compare. When you create a baseline, the Comparison Utility saves the target data and metadata to a flat file. Create a baseline when you want to compare data or metadata multiple times within one repository or between different repositories.

Figure 1-4 shows the Baseline Creation dialog box:

To create a baseline, complete the following tasks:

- **Create a baseline definition.** When you create a baseline definition, you select the repository definition, choose the targets and metadata you want in the baseline, and configure connection information to the target data.

- **Populate the baseline.** Populate a baseline to run the workflows and save the metadata and target data to file.

**To create a baseline definition:**

1. Click Baseline > New. The Informatica Baseline Creation dialog box appears.
2. Enter the job name, description, and baseline name.
3. Click Manage Repositories to create or edit a repository definition or to test a repository connection. For more information about creating repository definitions, see “Creating Repository Definitions” on page 7.

4. Optionally, you can select a repository from the list of repository definitions.

5. Select the workflows you want to compare from the Target Selection tab.  
   **Note:** Select the workflow to save the target data and the metadata including the workflow, session, and mapping.

6. Click Edit properties to select the Integration Service required to run the workflow, and to configure the target connection information. For more information about editing properties, see “Configuring Jobs” on page 12.

7. Click OK to save the baseline definition.

   A summary of the baseline information appears on the Information tab.

   Figure 1-5 shows the Information tab for a baseline:

   ![Figure 1-5. Baseline Information Tab](image)

**To populate a baseline:**

- Click Baseline > Populate.

   The Comparison Utility saves the target data and repository metadata and displays the log events on the main window.

   The following sample shows a section of the log events for populating a baseline:
Connecting to repository: cert622_uvu_811 . . .COMPLETED
Caching metadata . . .COMPLETED
Running workflow cert622_uvu_811.DB_ORACLE_uvu.QACUST_PREPOST_SQL . . .COMPLETED
Copying session logs . . .COMPLETED
Copying target: QASTORE_CERT_ORAUDB_TGT . . .COMPLETED
Copying target: QACUST_ORA_TGT . . .COMPLETED
Configuring Jobs

A job contains information about targets and metadata you want to compare. When you configure a job, choose a base and a comparison repository. You can use a baseline as the base or comparison repository. You also select targets and metadata that you want to compare. When you run the job, the Comparison Utility runs the workflows and generates a report comparing target data and metadata.

Run a job to complete the following tasks:

♦ **Compare targets.** Select the targets you want to compare. For more information about selecting targets, see “Target Selection” on page 12.

♦ **Compare metadata.** Select repository metadata you want to compare. You can compare the attributes and validation state for mappings, sessions, and workflows. For more information about selecting metadata, see “Metadata Selection” on page 13.

Target Selection

When you configure a job, select the targets you want to compare on the Target Selection tab. When you select the targets you want to compare, the corresponding repository folder and workflow appears in Base Workflows and Compare Workflows panes.

Use the following rules and guidelines when you select targets that you want to compare:

♦ Select targets that exist in both versions of the repositories.

♦ Use the arrows to change the order in which workflows are executed within the Base Workflow or Compare Workflow panes.

♦ Compare targets that exist in folders with the same name. If the target you want to compare exists in different repository folders, you can assign an alias to one of the folders you want to compare. To assign an alias right-click on the folder you want to rename and select Assign Alias.
Figure 1-6 shows the Target Selection tab in job creation window:

Figure 1-6. Target Selection Tab

Metadata Selection

When you configure a job, select the repository metadata you want to compare in the Metadata tab. Repository metadata include mappings, sessions, and workflows. You can compare attributes and validation state of repository metadata. The Comparison Utility displays any new or changed attributes in the report.

Metadata Selection

When you configure a job, select the repository metadata you want to compare in the Metadata tab. Repository metadata include mappings, sessions, and workflows. You can compare attributes and validation state of repository metadata. The Comparison Utility displays any new or changed attributes in the report.
Steps to Configure a Job

Use the following procedure to configure a job.

To configure a job:

1. Click Job > New > Verify Upgrade or Verify Update.
   
   The Job Definition and Editing window appears.

2. Specify name and description for the job.

3. In the Repository Selection area, select the base and comparison repositories. Select a repository definition or baseline.

4. Click Manage Repositories if you want to create or edit repository definitions. You can also test repository connections. For more information about creating or editing repository definitions, see “Creating Repository Definitions” on page 7.

5. Click Target Selection to view the folders available to both base and comparison repositories.
   
   The left pane displays the repository folder structure for the base repository.

6. Select the targets or workflows you want to compare.
The selected objects appear in the compare and base workflow panes.

**Note:** Use the arrows to change the order in which you want to run them.

7. Click the Metadata tab to select the repository metadata you want to compare.

8. Click Edit properties to select or configure the Integration Service required to run the workflow and to configure the target connection information. For more information about editing properties for a job, see “Configuring Properties” on page 17.

9. Click OK to save the job.

A summary of the job information appears on the Information tab.

Figure 1-8 shows the Information tab for a job:

**Figure 1-8. Information Tab for a Job**

![Information Tab for a Job](image)

**Running a Job**

Run a job to run workflows in the base and comparison repository, and generate a report comparing targets and repository metadata.

**To run a job:**

- Click Job > Execute to run the job.

The Comparison Utility runs the workflows and displays the log events on the main window.
The following sample shows a section of log events for running a job:

- Connecting to repository: cert622_uvu . . .COMPLETED
- Connecting to repository: cert622_uvu_811 . . .COMPLETED
- Running workflow cert622_uvu_811.DB_ORACLE_uvu.QACUST_PREPOST_SQL . . .COMPLETED
- Copying session logs . . .COMPLETED
- Comparing Metadata . . .COMPLETED
- Differing Metadata XML . . .COMPLETED
- Cleaning up files . . .COMPLETED

**Note:** The Comparison Utility also saves all the log information as a text file. To view log events, click Logs > Execute Logs. For more information about viewing log events, see “Viewing Logs” on page 28.
Configuring Properties

When you create a job or baseline, you configure the following types of information:

- Target type and connection information
- Target fields to compare
- PowerCenter Server or Integration Service to run the workflow
- Session logs to view

Configuring Target Type and Connection Information

You can choose file or database type when you configure target properties. The Comparison Utility can compare Oracle, Microsoft SQL Server, Teradata, and IBM DB2 database targets.

Flat File Targets

To compare flat file targets with the Comparison Utility, specify the location of flat files for the base repository and the comparison repository.

Figure 1-9 shows the Flat Files tab in the Edit Properties window:

![Figure 1-9. Configuring Properties for a flat file](image)

Table 1-1 describes the location properties you configure:

Table 1-1. Flat File Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Required/Optional</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folder Name</td>
<td>Required</td>
<td>Name of the folder that contains the target.</td>
</tr>
<tr>
<td>Workflow Name</td>
<td>Required</td>
<td>Name of the workflow that contains the target.</td>
</tr>
</tbody>
</table>
You can also select remote targets by configuring an FTP connection.

To access remote targets:

1. Click FTP. The FTP window appears:

Table 1-2 describes the properties that you configure for an FTP connection:

<table>
<thead>
<tr>
<th>Property</th>
<th>Required/Optional</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host</td>
<td>Required</td>
<td>Host name.</td>
</tr>
<tr>
<td>Port</td>
<td>Optional</td>
<td>Port number.</td>
</tr>
<tr>
<td>Username</td>
<td>Optional</td>
<td>User name required to access the host machine.</td>
</tr>
<tr>
<td>Password</td>
<td>Optional</td>
<td>Password for the user name.</td>
</tr>
<tr>
<td>Directory</td>
<td>Optional</td>
<td>Directory location on the FTP host.</td>
</tr>
</tbody>
</table>

2. Click Test Connection to test the connection.
3. Click OK to return to the Edit Properties window.

Relational Targets

If you are comparing relational targets, configure the connection information for targets in the base and comparison repositories.
Figure 1-10 shows Databases tab in the Edit Properties window:

**Figure 1-10. Configuring properties for a Relational Target**

Select database targets.

![Configuring target connections for base repository.]

Configure target connections for comparison repository.

Table 1-3 describes the relational target properties:

**Table 1-3. Relational Target Properties**

<table>
<thead>
<tr>
<th>Property</th>
<th>Required/Optional</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection Name</td>
<td>Required</td>
<td>Name of the connection.</td>
</tr>
<tr>
<td>Database Type</td>
<td>Required</td>
<td>Type of the target database.</td>
</tr>
<tr>
<td>Instance Name</td>
<td>Required</td>
<td>Name of the target instance.</td>
</tr>
<tr>
<td>Host</td>
<td>Required</td>
<td>Name of the host.</td>
</tr>
<tr>
<td>Port</td>
<td>Required</td>
<td>Port number.</td>
</tr>
<tr>
<td>User Name</td>
<td>Required</td>
<td>Name of a user who has the appropriate database permissions to read from and write to the database.</td>
</tr>
<tr>
<td>Password</td>
<td>Required</td>
<td>Database password.</td>
</tr>
</tbody>
</table>

**Selecting the Target Fields**

When you configure a job, choose target fields you want to compare. Enter a filter condition to filter results. Click Generate Filter to view the filter condition you entered.
Figure 1-11 shows the Targets tab in the Edit Properties window:

**Figure 1-11. Selecting Target Fields for a Job**

Select target fields.

Generate filter condition.

Configuring Server Connections

Select the Integration Service or PowerCenter Server to run the workflows. You can also add or edit connections.

Figure 1-12 shows the PowerCenter Servers tab in the Edit Properties window:

**Figure 1-12. Configuring Properties for the PowerCenter Server**

Select PowerCenter Server.

Configure PowerCenter Server for base repository.

Configure PowerCenter Server for comparison repository.
To add connections:

1. Click New. The New PMServer dialog box appears:

Table 1-4 describes the properties that you configure for an Integration Service or PowerCenter Server connection:

<table>
<thead>
<tr>
<th>Property</th>
<th>Required/Optional</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMServer Name</td>
<td>Required</td>
<td>Name of the Integration Service or PowerCenter Server.</td>
</tr>
<tr>
<td>Host name</td>
<td>Required</td>
<td>Name of the host.</td>
</tr>
<tr>
<td>Port Number</td>
<td>Required</td>
<td>Number for the port.</td>
</tr>
<tr>
<td>User name</td>
<td>Required</td>
<td>Name of the user.</td>
</tr>
<tr>
<td>Password</td>
<td>Required</td>
<td>Password for the user name.</td>
</tr>
</tbody>
</table>

2. Click OK to return to the Edit Properties window.

Configuring Session Log Information

To select the session logs you want to view, specify the location of session logs for the workflows you compare.
Figure 1-13 shows the Session Logs tab in the Edit properties window:

**Figure 1-13. Configuring Session Log Information**

You can write session logs to a remote location by configuring an FTP connection.

**Steps to Edit Properties**

Use the following procedure to configure target and server properties for a job or baseline.

**To edit the properties for a job or baseline:**

1. Select the job or baseline you want to edit.
2. Click Job or Baseline > Edit Properties.
3. Optionally, Right-click a job or baseline and select Edit Properties. The Edit Properties dialog box appears.
4. Specify the location of flat files on the Flat File tab.
   - **Note:** Click FTP to access a target in a remote location.
5. Click Set Path to configure the path to the target.
6. Configure the connection information for relational targets on the Databases tab.
7. Click Test to test the connection.
8. Choose the Integration Service or PowerCenter Server required to run the workflow on the PMServers tab. You can also add or edit connections.
9. Select the target fields you want to compare on the Targets tab.
   - **Note:** The Targets tab is only available when you configure properties for a job. You cannot select target fields when you configure baselines.
10. Click Close.
When you run a job, the Comparison Utility runs workflows in different repositories and generates a report comparing target data and repository metadata. You can customize the report to display information depending on the options you choose. You can also export or save part or whole of the report as an HTML file.

After you run a job, you can review the following reports:

- **Target comparison.** Displays differences in target data between repositories or a repository and a baseline.
- **Metadata comparison.** Displays the validation state of repository metadata like mappings, sessions, and workflows.
- **Session logs.** Displays the sessions logs for the workflows you compare.

### Viewing Target Data

View differences in target data between repositories, or between a repository and a baseline on the Target Comparison.

Figure 1-14 shows the Target Comparison Tab:

![Figure 1-14. Target Comparison Tab](image)

Select the following options to customize the report:

- **Masked.** Hide target data in a report. Use Masked when you need to hide confidential data.
- **Key.** Sort differences by a target field.
The Comparison Utility displays the following information in a report:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base differences</td>
<td>Number of rows the base target that are not in comparison target.</td>
</tr>
<tr>
<td>Compare differences</td>
<td>Number of rows in comparison target that are not in the base target.</td>
</tr>
<tr>
<td>Total differences</td>
<td>Total number of differences.</td>
</tr>
<tr>
<td>Total base rows</td>
<td>Number of rows in the base target.</td>
</tr>
<tr>
<td>Total compare rows</td>
<td>Number of rows in the comparison target.</td>
</tr>
<tr>
<td>Total rows</td>
<td>Total number of rows.</td>
</tr>
<tr>
<td>Rows matched by key</td>
<td>Total number of rows that have same values for a field.</td>
</tr>
<tr>
<td>Value difference percent</td>
<td>Percentage difference between number of rows that have same value for a field and the total number of rows.</td>
</tr>
</tbody>
</table>

**Viewing Metadata**

Verify and compare attribute values and validation state of repository metadata like mappings, sessions, and workflows on the Metadata Comparison tab.

The Comparison Utility displays the data in different colors. Each color represents new or changed attribute in a repository object. The rows in blue indicate that the objects in base or compare repository have additional attributes. The rows in yellow indicate that the objects in base or compare repository have different values for the same attributes.

Figure 1-15 shows the metadata comparison tab:

**Figure 1-15. Metadata Comparison Tab**

View metadata in XML format.
The Metadata Comparison tab displays the following information in the report:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object type</td>
<td>Type of repository object. The object can be a mapping, session, or workflow.</td>
</tr>
<tr>
<td>Base name</td>
<td>Fully-qualified name of the repository object in the base. For example, a fully-qualified mapping name is &lt;repository folder&gt;.&lt;workflow name&gt;.&lt;session name&gt;.&lt;mapping name&gt;.</td>
</tr>
<tr>
<td>Base valid</td>
<td>Validation state of the object in the base. Return value is TRUE or FALSE. A valid object returns TRUE.</td>
</tr>
<tr>
<td>Compare name</td>
<td>Name of the repository object being compared. When displaying object name, Comparison Utility prepends all the parent object names to the object. For example, if the object is a mapping, the name would be displayed as follows: repository folder.workflow.session.mapping.</td>
</tr>
<tr>
<td>Compare valid</td>
<td>Validation state of the object in the comparison repository. Return value is TRUE or FALSE. A valid object returns TRUE.</td>
</tr>
<tr>
<td>View Metadata</td>
<td>View metadata in XML format.</td>
</tr>
</tbody>
</table>

Viewing Attribute Values
For any attribute, you can choose to view metadata. The Comparison Utility displays the data in different colors. Each color represents new or changed attribute in a repository object. To view the attributes, click View Metadata on the Metadata Comparison tab.

Figure 1-16 shows the XML file:

Figure 1-16. Viewing Metadata
Viewing Session Logs

View the session logs for workflows you are comparing. The Comparison Utility displays the session logs for each workflow you compare. You can also create a watchlist to search for text in the session logs.

Use the watchlist to search for specific errors or messages in the session logs. Click Edit Watchlist to select or delete phrases in the watchlist.

Figure 1-17 shows the session log tab:

Steps to View a Report

You can view a report to compare the target data and repository metadata.

To view a report:

1. Select the job.
2. Click Job > View Reports. Optionally, right-click on a job, and click View Reports.
The Input window appears.

3. Select the report you want to view. Reports are listed by date that job runs.
4. Click OK. The Report window appears.
5. Click Target Comparison to view the report that compares targets.
6. Click Metadata Comparison tab to view report that compares repository metadata.
7. Click Session Logs tab to view the session logs.
Viewing Logs

Logs include information about all tasks that the Comparison Utility performs. Tasks include creating repository definitions, configuring baselines or jobs, running jobs, and populating baselines. When you perform a task, a summary of the logs is displayed on the information pane. Logs are also archived as text files in the logs directory and contain information about log events including run time, task name, message code, and task description.

The Comparison Utility generates the following log files:

- **Execute log.** View log events created when creating and executing a job or baseline.
- **Tool log.** View tracing and debug messages.

Execute Log

You can view the logs generated when creating or running a baseline or a job. Use the information printed in the execute log to view and fix errors.

The following sample shows a section of an execute log file:

```
11:42:24.811 EXECUTE CompareTask:710 - Starting Comparison Job: Repo_Upgrade_8x_job
11:42:24.811 EXECUTE CompareTask:720 - Connecting to repository: cert622_uvu
11:42:24.936 EXECUTE CompareTask:744 - Finished Connection To Repository
11:42:37.155 EXECUTE CompareTask:624 - Finished running workflow
11:42:38.858 EXECUTE CompareTask:790 - Finished copying target
11:42:38.874 EXECUTE CompareTask:804 - Connecting to repository: cert622_uvu_811
11:42:38.890 EXECUTE CompareTask:830 - Finished connecting to repository
11:43:09.061 EXECUTE CompareTask:624 - Finished running workflow
11:43:09.265 EXECUTE CompareTask:878 - Finished copying target
11:43:09.280 EXECUTE CompareTask:886 - Comparing Metadata
11:43:09.577 EXECUTE CompareTask:923 - End of Comparison Job: Repo_Upgrade_8x_job
```
Tool Log

You can view tracing and debug information in the tool log.

The following sample shows a section of a tool log file:

09:00:09,336  INFO Application:256 - Binding 6.2.2 repository connector
09:00:20,554  INFO Application:260 - Finished binding 6.2.2 repository connector
09:00:20,554  INFO Application:264 - Binding 8.1.1 repository connector
09:00:32,664  INFO Application:278 - Finished binding 8.1.1 repository connector
09:00:32,664  INFO Application:264 - Binding 7.1.5 repository connector
09:00:36,726  INFO Application:278 - Finished binding 7.1.5 repository connector
09:01:00,633  INFO Application:386 - Comparison Utility Started Successfully
09:01:23,101 ERROR ComparisonResultsManager:132 - Potentially Fatal Deserialization Operation.
java.io.InvalidClassException: Overriding serialized class version mismatch: local serialVersionUID = 2 stream serialVersionUID = 1
at model.objects.ComparisonResultsManager$ComparisonResultManagerInputStream.readClassDescriptor(ComparisonResultsManager.java:131)
at java.io.ObjectInputStream.readNonProxyDesc(ObjectInputStream.java:1534)
at java.io.ObjectInputStream.readClassDesc(ObjectInputStream.java:1466)
Glossary

Baseline Definition
Set of repository metadata and target data that you want to compare. When you create a baseline, the Comparison Utility saves the target data and metadata to a flat file.

base repository
The repository or baseline you want to compare. If you use a baseline as the base repository, the Comparison Utility gets the data and metadata from the populated baseline.

Execute Log
The logs generated when creating or running a baseline or a job.

Job
The information about targets and metadata you want to compare. When you run the job, the Comparison Utility runs the workflows and generates a report comparing target data and metadata.

Repository Definition
The connection information for a repository that you want to compare. When you configure a job or baseline, you must provide a repository definition.

Tool Log
The tracing and debug information generated when you run the Comparison Utility.

watchlist
List of words or phrases you want to search in the session logs. Use the watchlist to look for specific errors or messages in the session logs.