

Importing Flat File Sources in Test Data Management

Abstract

In Test Data Manager, you can import flat file metadata sources from the PowerCenter repository to perform Test Data Management (TDM) operations. This article describes how to import a flat file metadata into the PowerCenter repository, import metadata from the flat file sources in Test Data Manager, and configure flat file settings in a plan. This article assumes that you have knowledge of TDM and are familiar with TDM operations.

Supported Versions

- Test Data Management 9.5.1 to 9.6.0
- Test Data Management 9.6.1
- Test Data Management 9.6.1 HotFix 1
- Test Data Management 9.7.0
- Test Data Management 9.7.1
- Test Data Management 9.7.1 HotFix 1
- Test Data Management 10.1.0
- Test Data Management 10.2.0

Table of Contents

Overview.	2
Create a Folder in the PowerCenter Repository.	3
Import a Flat File into PowerCenter.	4
Importing a Fixed-Width Flat File.	4
Importing a Delimited Flat File.	8
Formatting Datetime Data Types.	9
Create a Project and Import Flat File Sources in Test Data Manager.	11
Perform TDM Operations.	12

Overview

You can import metadata from a flat file in Test Data Manager through the PowerCenter repository.

You must import flat files into the PowerCenter repository from the Designer.

You can import the following types of flat files:

Delimited

Delimited flat files are character-oriented and line sequential. The column precision is measured in characters for string columns and in digits for numeric columns. Each row ends with a newline character.

Fixed Width

In fixed-width flat files, the field lengths are measured in bytes. Each row ends with a newline character.

You can use the flat file sources in a project to perform data subset, data masking, and data generation operations.

Create a Folder in the PowerCenter Repository

Create a folder where you can define the data sources and targets, build mappings, and run workflows. In this folder, you can import the flat file metadata that you want to use for TDM operations.

1. In the Repository Manager, click **Folder > Create**.
2. Enter a name for the folder.

By default, the user account that logged in is the owner of the folder and has full permissions on the folder. The following image shows the **Create Folder** dialog box in the Repository Manager:

The image shows a 'Create Folder' dialog box with the following fields and options:

- Name:** FlatFile_Credit
- Descriptions:** (Empty text area)
- Owner:** Administrator
- OS Profile:** <None>
- Options:** Allow Shortcut
- Status:** Active

Buttons: OK, Cancel, Help

3. Click **OK**.
The Repository Manager displays a message that the folder has been successfully created.
4. Click **OK**.
The folder appears in the repository.
5. Exit the Repository Manager.

Import a Flat File into PowerCenter

You can import fixed-width and delimited flat file definitions that do not contain binary data.

When you import the definitions, the file must be in a directory local to the client machine. In addition, the PowerCenter Integration Service must be able to access all source files during the session. Copy the file to the source file location on the machine that hosts the PowerCenter Integration Service. The default location is `<Informatica installation directory>/server/infa_shared/srcfiles`.

When you create a file source, target, or lookup definition, you must define the properties of the file. The Flat File Import Wizard prompts you for the following file properties:

- File name and location
- File code page
- File type
- Column names and data types
- Number of header rows in the file
- Column size and null characters for fixed-width files
- Delimiter types, quote character, and escape character for delimited files

Importing a Fixed-Width Flat File

Import a fixed-width flat file into the folder that you created in the Repository Manager. When you import a fixed-width file, you can create, move, or delete column breaks.

1. In the Designer, connect to the repository and open the folder that you created in the Repository Manager.
2. To import a flat file source definition, open the Source Analyzer and click **Sources > Import from File**.
The **Open Flat File** dialog box appears.
3. Select the flat file that you want to use.
4. Select a code page.

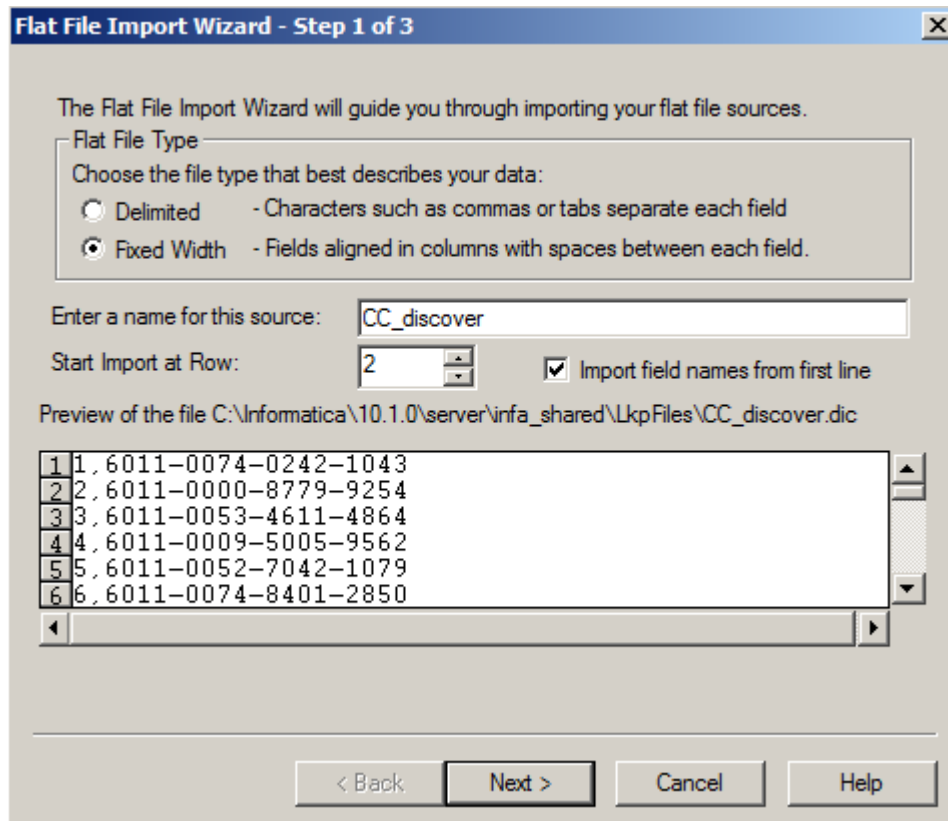
When you import a flat file source definition, select a code page that matches the code page of the data in the file.

5. Click **OK**.

The Flat File Import Wizard opens and the contents of the file appear at the bottom of the wizard.

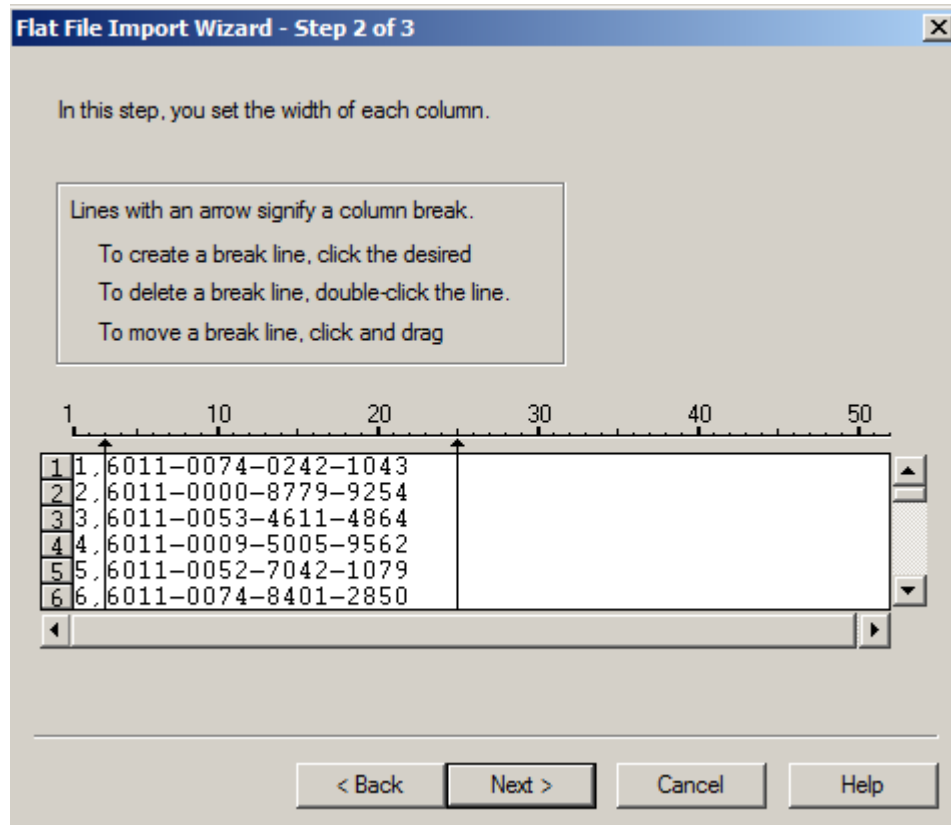
6. In the wizard, select the **Fixed Width** file type.
7. Enter a name for the source. Default is the name of the flat file.
8. Enter the row number at which the wizard starts reading when it imports the file. Default is 1.
9. If the first row contains the column names, select **Import field names from first line**. When you select this option the Start Import at Row field value changes to 2.

The following image shows the wizard in which you can configure the fixed-width flat file properties:



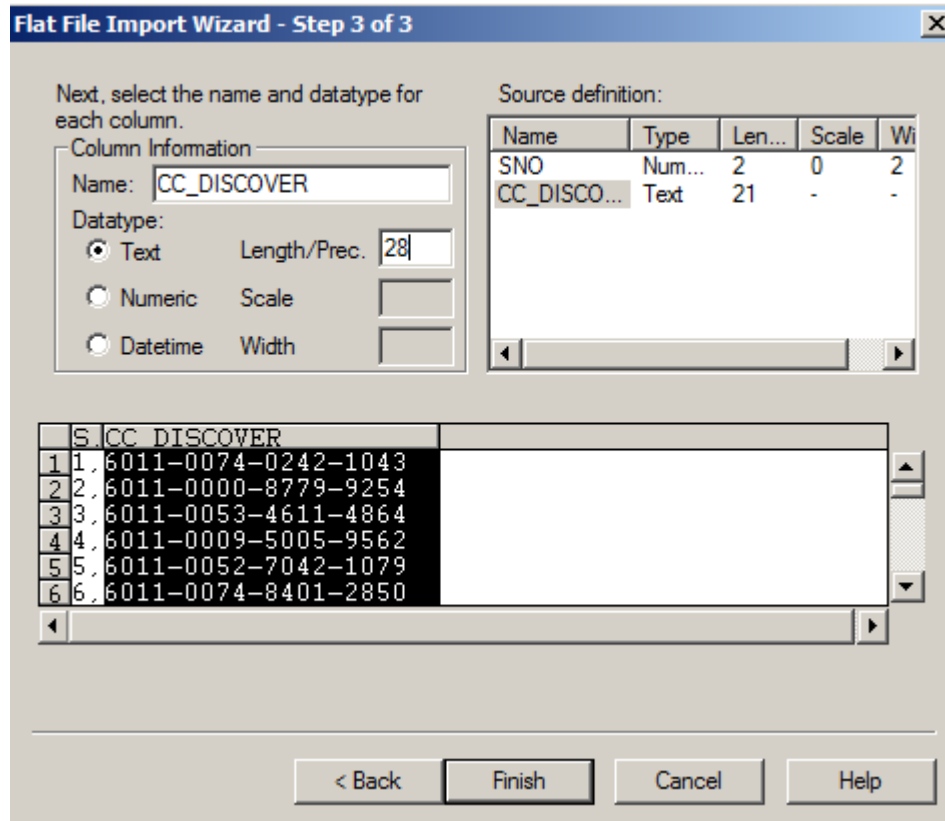
10. Click **Next**.
11. To set the width of the columns, drag the existing column breaks. Follow the directions in the wizard to manipulate the column breaks in the file preview window.

The following image shows the wizard in which you can set the width of the columns:



12. Click **Next**.
13. Enter the following column information for each column in the file:
 - Name. Default is the column name in the file. Optionally, you can enter another column name.
 - Datatype. Select Text, Numeric, or Datetime. Based on the data type, enter the appropriate length, precision, scale, and width. Or you can use the default values that the wizard contains.
14. To switch between columns, select a column in the **Source Definition**, or click the column heading in the file preview window.

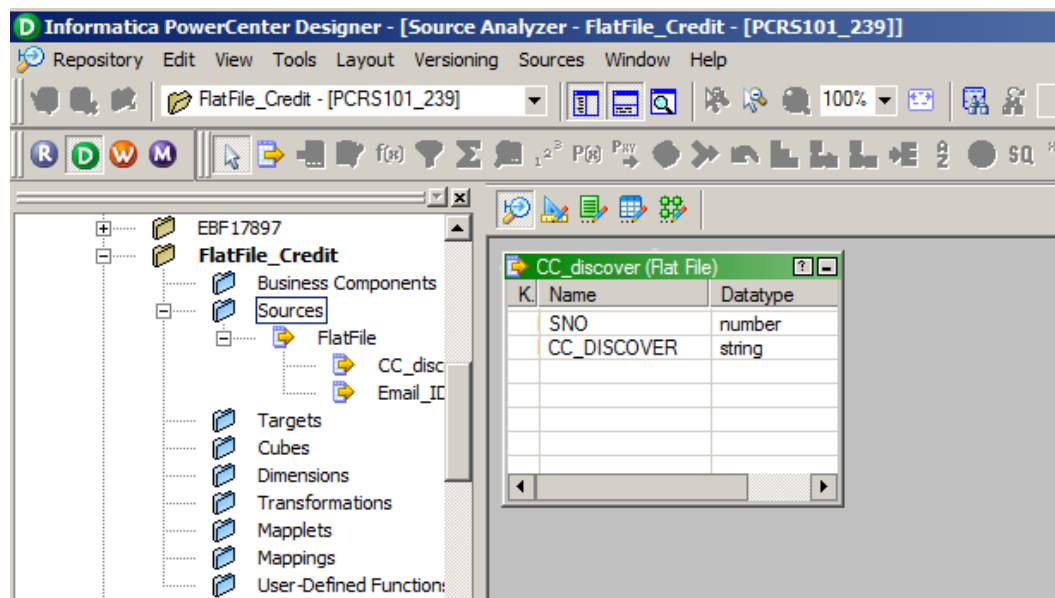
The following image shows the column details that you can configure:



- Click **Finish**.

The flat file appears on the Source Analyzer window.

The following image shows the fixed-width flat file that you imported in the task:



Importing a Delimited Flat File

Import a delimited flat file into the folder that you created in the Repository Manager. You can specify delimiters, optional quotes, and escape characters that are contained in the source file code page.

1. In the Designer, connect to the repository and open the folder that you created in the Repository Manager.

2. Open the Source Analyzer, click **Sources > Import from File**.

The **Open Flat File** dialog box appears.

3. Select the flat file that you want to use.

4. Select a code page.

When you import a flat file source definition, select a code page that matches the code page of the data in the file.

5. Click **OK**.

The Flat File Import Wizard opens and the contents of the file appear at the bottom of the wizard.

6. In the wizard, select the **Delimited** file type.

7. Enter a name for the source. Default is the name of the flat file.

8. Enter the row number at which the wizard starts reading when it imports the file. Default is 1.

9. If the first row contains the column names, select **Import field names from first line**. When you select this option the Start Import at Row field value changes to 2.

10. Click **Next**.

Any incorrectly parsed fields display in red text in the file preview window at the bottom of this screen.

11. To separate columns of data, select the delimiter present in the flat file that you imported.

You can use the Other field to enter a different delimiter.

12. If you want the wizard to read one or more consecutive column delimiters as one, select **Treat consecutive delimiters as one**. If you want the wizard to read a specified set of delimiters as one, select **Treat multiple delimiters as AND**.

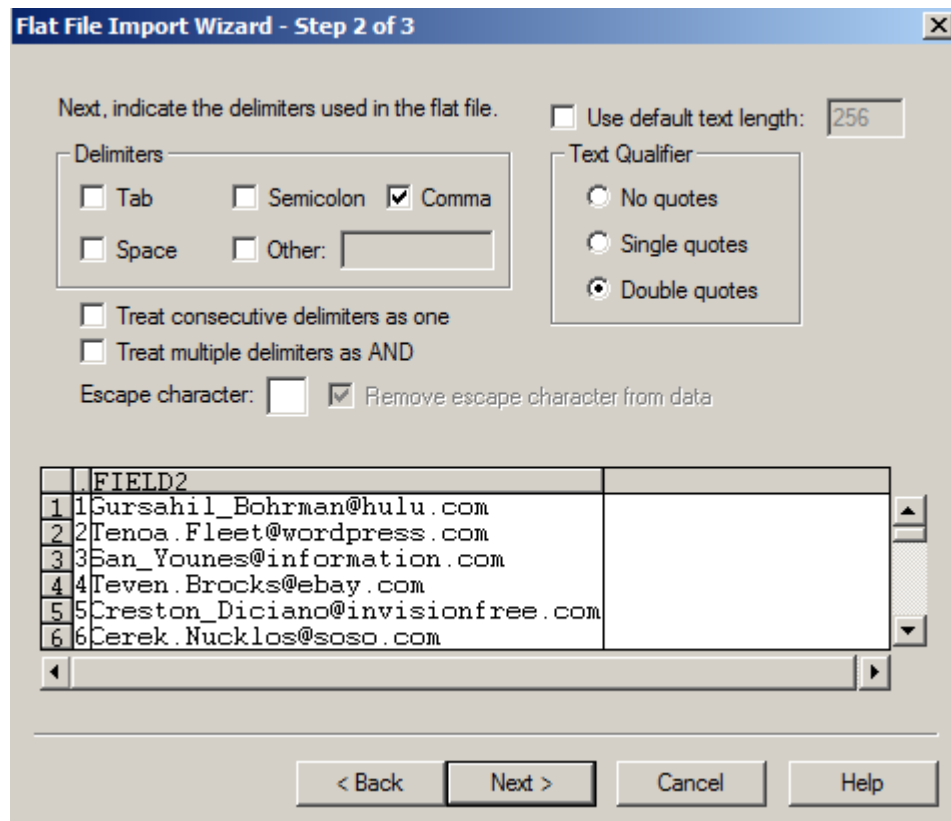
13. If you want to escape a delimiter, specify an escape character.

When you specify an escape character, the PowerCenter Integration Service reads the delimiter character as a regular character.

14. In the Text Qualifier field, choose No Quote, Single Quote, or Double Quotes.

If you select a quote character, the Flat File Wizard ignores delimiters within pairs of quotes.

The following image shows the delimiters that you can configure for a delimited flat file:



15. Click **Next**.

16. Enter the following column information for each column in the file:

- Name. Default is the column name in the file. Optionally, you can enter another column name.
- Datatype. Select Text, Numeric, or Datetime. Based on the data type, enter the appropriate length, precision, scale, and width. Or you can use the default values that the wizard contains.

To switch between columns, select a column in the **Source Definition**, or click the column heading in the file preview window.

17. Click **Finish**.

Formatting Datetime Data Types

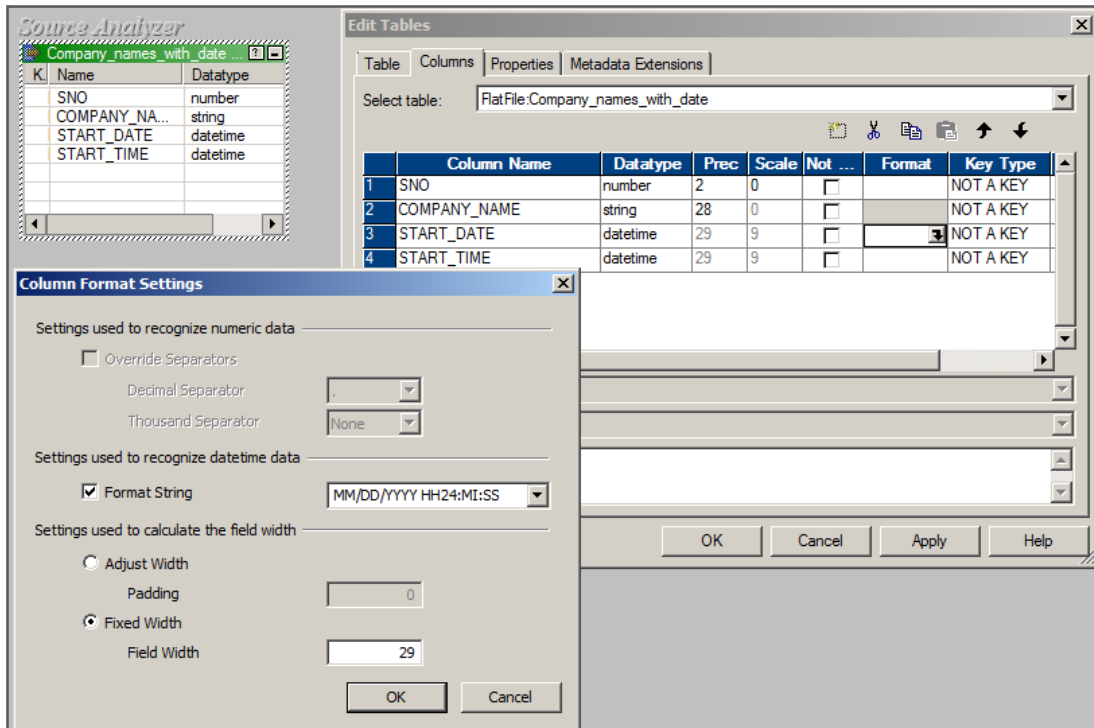
You can define formats for datetime values after you import a flat file into PowerCenter.

To configure the flat file properties for datetime data type, you must double-click the flat file that you imported into PowerCenter.

Use the **Format** column of a flat file definition to define the format for each column with datetime data type. You can define the following formatting options for datetime values in the **Column Format Settings** dialog box:

- Format string
- Field width

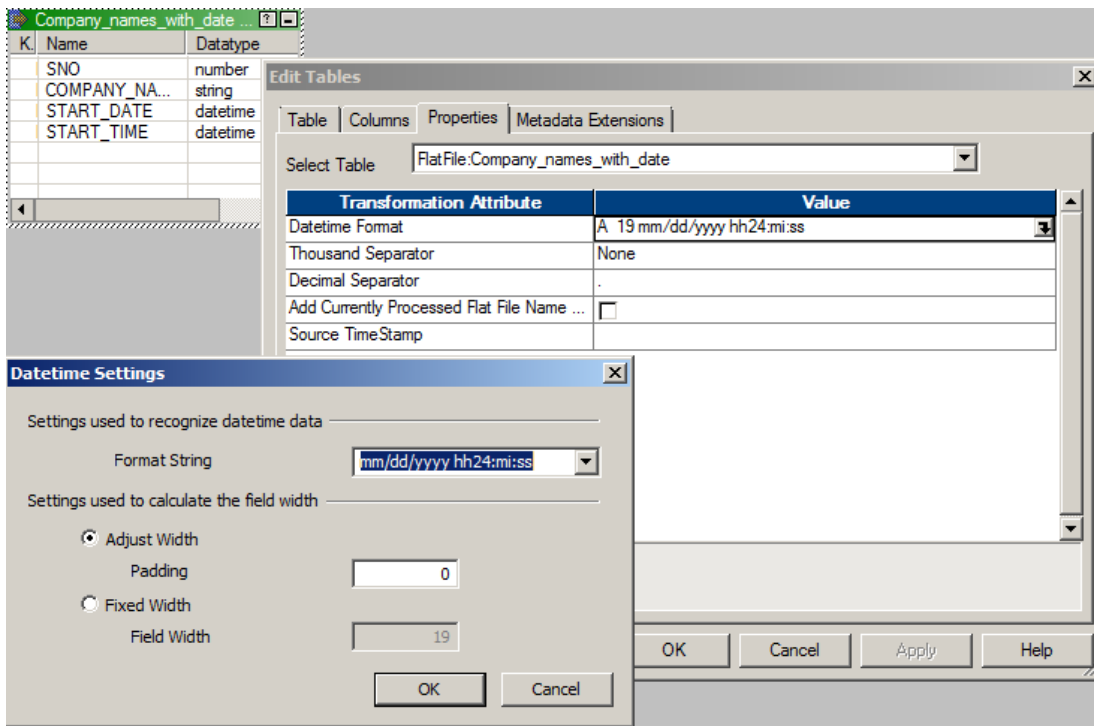
The following image shows the format settings that you can configure for a flat file column with the datetime data type:



To specify the datetime format, choose **Format String** and enter the format in the **Format String** field. You can choose a format from the list, or you can enter a format using the keyboard. The default format is MM/DD/YYYY HH24:MI:SS, which has a field width of 19.

To configure datetime values for a flat file, click the **Properties** tab on the **Edit Tables** dialog box.

The following image shows the format settings that you can configure for a flat file for the datetime data type:



Create a Project and Import Flat File Sources in Test Data Manager

In Test Data Manager, create a project to store the TDM components. Import flat file metadata from the PowerCenter repository to the project to perform TDM operations on the source data.

1. In Test Data Manager, click **Projects**.

A list of projects appears.

2. Click **Actions > New**.

3. In the **Create Project** dialog box, enter the following project properties:

Name

Enter a name for the project.

Description

Optional. Enter a description.

PowerCenter Repository

TDM populates the name of the PowerCenter repository to store the repository folder.

Folder

Default is the project name. You can choose another folder in the repository.

Owner

Default is the name of the user that created the project. You can select another user as the project owner.

4. Click **OK**.

The project opens in a separate tab.

5. In the project, click **Actions > Import Metadata**.

The **Import Metadata** window appears.

6. To import metadata from a flat file, select **PowerCenter Repository**.

7. Choose to review the metadata changes before you import the metadata.

Test Data Manager displays the information before the import runs.

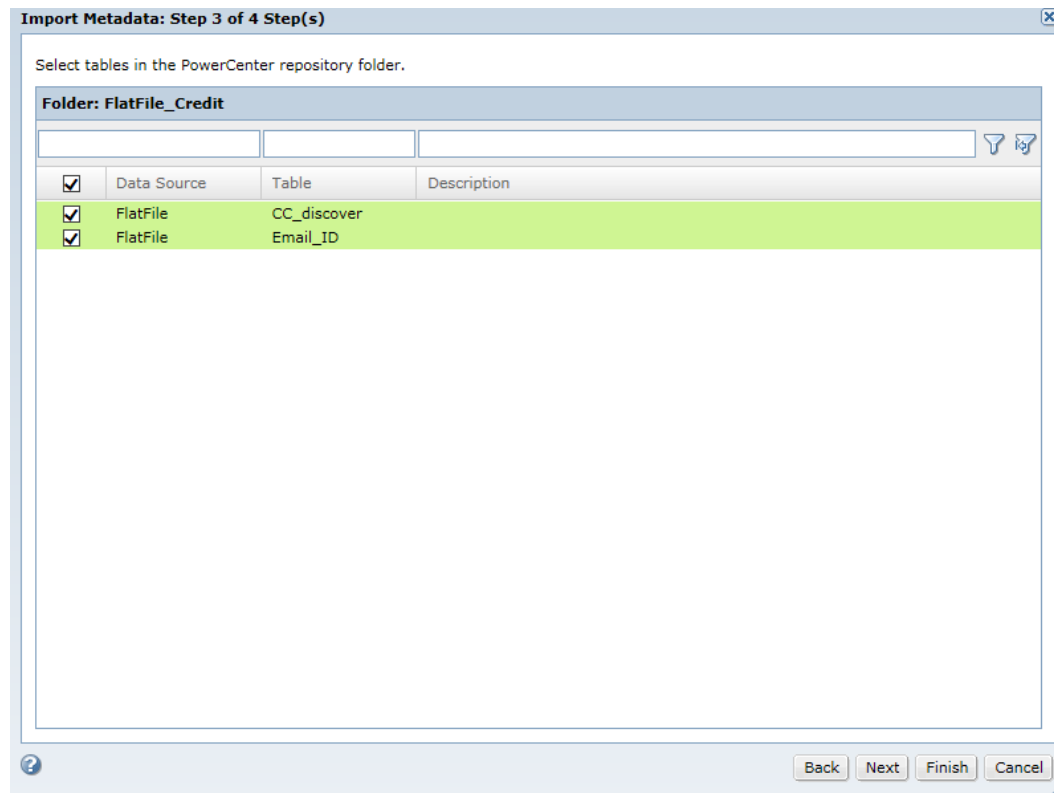
8. Click **Next**.

9. Select the PowerCenter folder that contains the flat file source that you created in the Repository Manager.

10. Click **Next**.

11. Select the delimited and fixed-width flat files that you created in the PowerCenter repository.

The following image shows the Import Metadata dialog box from where you can select the flat files:



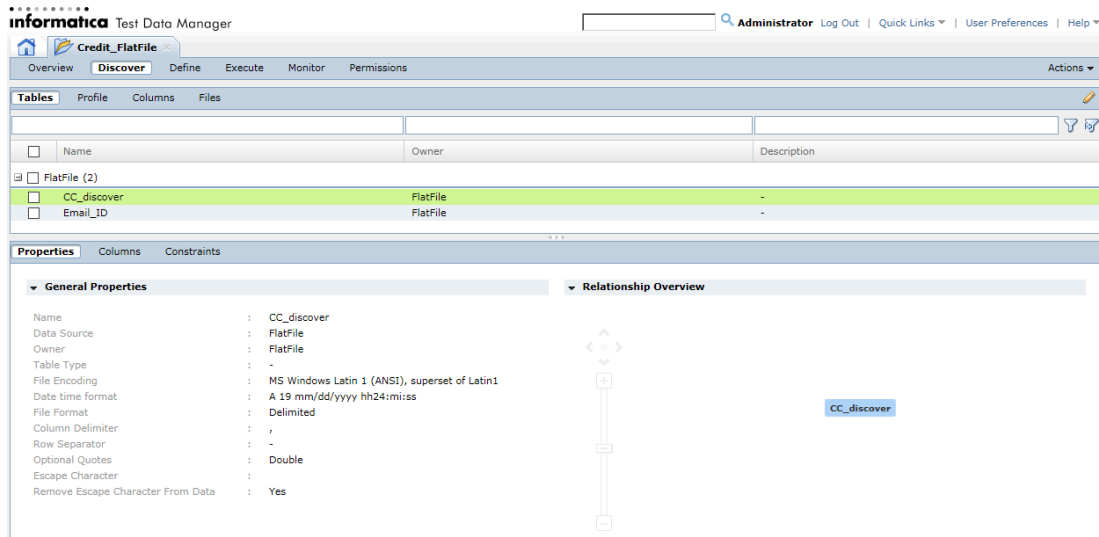
12. Click **Next**.
13. To import the metadata immediately, choose **Import Now**.
14. Click **Finish**.

Perform TDM Operations

You can perform data subset, data masking, and data generation operations on the flat file data.

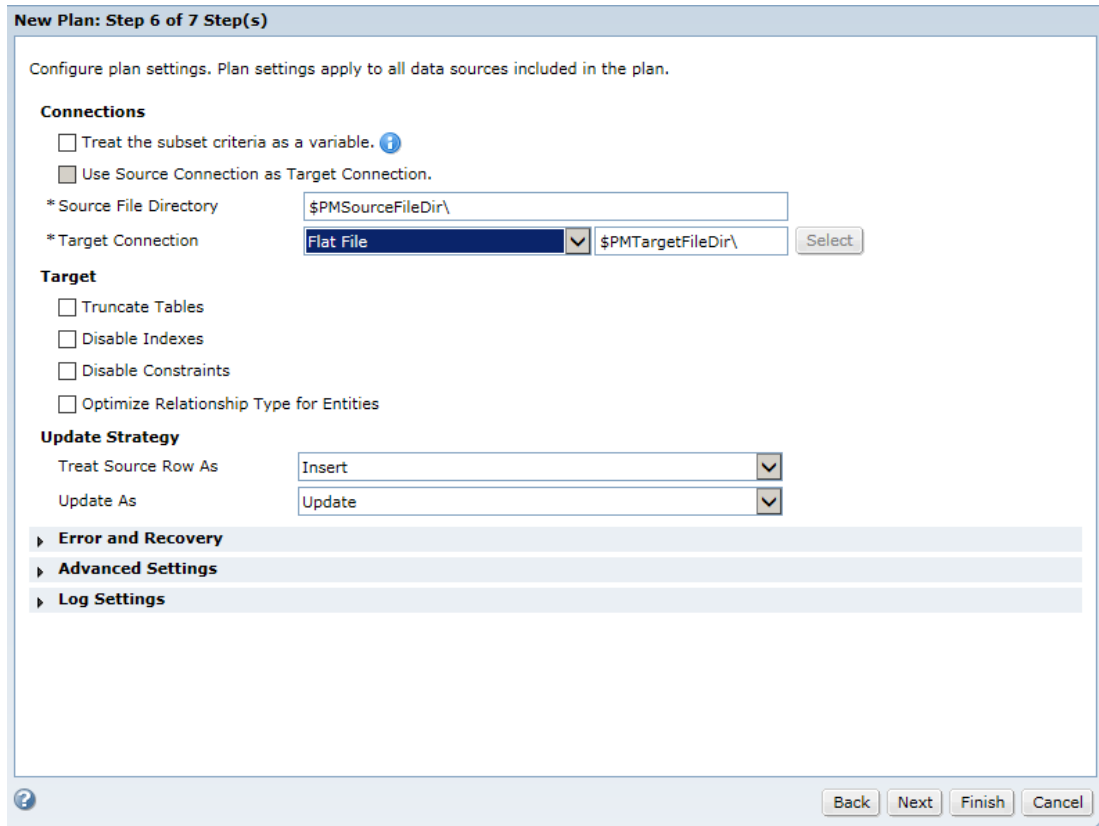
After you import flat file metadata into the project, you can verify the properties that you configured for a flat file in the **Properties** tab from the **Discover | Tables** view. To view the properties that you configured for each column of the flat file, you can click the **Columns** tab.

The following image shows the flat file properties that you can view in a project:



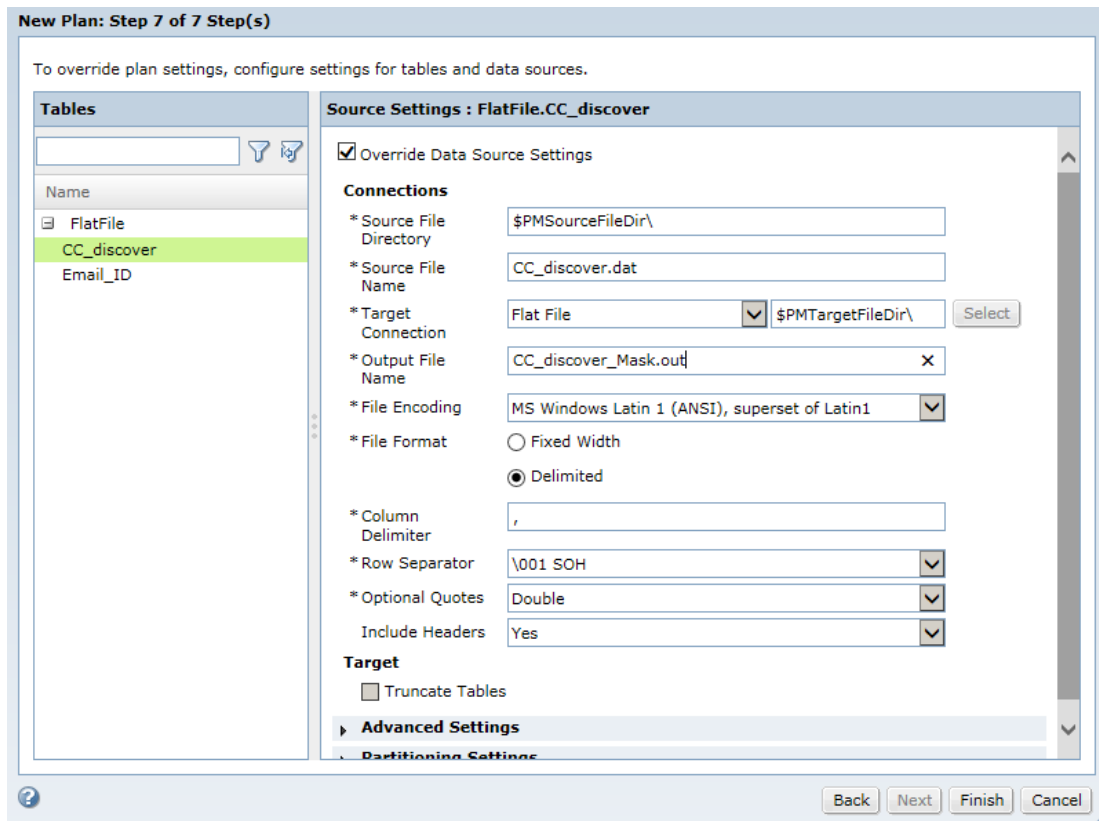
To perform a TDM operation, you must create a plan. Add the components to the plan and specify the source and target connections. You can specify the datetime format settings in the advanced plan settings.

The following image shows the plan properties that you can configure for a flat file source:



If you specify a flat file target, you can configure the flat file properties when you override the source settings. You can specify the datetime format settings in the advanced source settings of the plan. If the source is a relational database and the target is a flat file, TDM produces datetime output in the format that you specify in the **Advanced Settings** properties in step 6. You can override the settings in the **Source Settings** properties in step 7. If both the source and target are flat files, TDM produces datetime output in the format that you configured in PowerCenter.

The following image shows the properties that you can configure for a flat file:



The following table describes the flat file properties that you can configure in a plan:

Connection Options	Description
Output Filename	The name of the target file. Required if the target is a flat file. Default is <name of selected source>.out.
File Encoding	The file encoding type. Required if the target is a flat file. If the source is a flat file, the default type is what the flat file contains. If the source is relational, default is MS Windows Latin 1 (ANSI), superset of Latin1. You can select the type of flat file encoding that you want in the target.
File Format	The format of a flat file. Required if the target is a flat file. You can choose one of the following options: - Fixed Width. The width of the columns is fixed. You cannot specify a column delimiter. - Delimited. You can limit the width of the columns. Specify a column delimiter. Default is Delimited.
Column Delimiter	A character that separates columns of data. Required if the source is a flat file and if the file format is delimited. Default is a comma (,).
Row Separator	A character that separates rows of data. Required if the source is a flat file. Default is \012 LF (\n).

Connection Options	Description
Optional Quotes	<p>Select No Quotes, Single Quote, or Double Quotes. If you select a quote character, the PowerCenter Integration Service ignores delimiter characters within the quote characters. Therefore, the Integration Service uses quote characters to escape the delimiter.</p> <p>For example, a source file uses a comma as a delimiter and contains the following row:</p> <pre>342-3849, 'Smith, Jenna', 'Rockville, MD', 6.</pre> <p>If you select the optional single quote character, the PowerCenter Integration Service ignores the commas within the quotes and reads the row as four fields.</p> <p>If you do not select the optional single quote, the PowerCenter Integration Service reads six separate fields.</p> <p>When the PowerCenter Integration Service reads two optional quote characters within a quoted string, it treats them as one quote character. For example, the Integration Service reads the following quoted string as</p> <pre>I'm going tomorrow:</pre> <pre>2353, 'I'm going tomorrow', MD</pre> <p>Additionally, if you select an optional quote character, the PowerCenter Integration Service reads a string as a quoted string if the quote character is the first character of the field.</p> <p>Note: You can improve session performance if the source file does not contain quotes or escape characters.</p>
Include Headers	Optional. You can choose to include the headers in the target flat file. Default is Yes.

Author

Vinita Arun Kumar

Acknowledgements

The author would like to acknowledge Soumi Mondal, Software Engineer and Shawn Mendes, QA Engineer, for their technical assistance.