



Define Initial Data Models



This document supports Pentaho Business Analytics Suite 5.0 GA and Pentaho Data Integration 5.0 GA, documentation revision August 28, 2013, copyright © 2013 Pentaho Corporation. No part may be reprinted without written permission from Pentaho Corporation. All trademarks are the property of their respective owners.

Help and Support Resources

If you do not find answers to your questions here, please contact your Pentaho technical support representative.

Support-related questions should be submitted through the Pentaho Customer Support Portal at <http://support.pentaho.com>.

For information about how to purchase support or enable an additional named support contact, please contact your sales representative, or send an email to sales@pentaho.com.

For information about instructor-led training, visit <http://www.pentaho.com/training>.

Liability Limits and Warranty Disclaimer

The author(s) of this document have used their best efforts in preparing the content and the programs contained in it. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, express or implied, with regard to these programs or the documentation contained in this book.

The author(s) and Pentaho shall not be liable in the event of incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of the programs, associated instructions, and/or claims.

Trademarks

Pentaho (TM) and the Pentaho logo are registered trademarks of Pentaho Corporation. All other trademarks are the property of their respective owners. Trademarked names may appear throughout this document. Rather than list the names and entities that own the trademarks or insert a trademark symbol with each mention of the trademarked name, Pentaho states that it is using the names for editorial purposes only and to the benefit of the trademark owner, with no intention of infringing upon that trademark.

Third-Party Open Source Software

For a listing of open source software used by each Pentaho component, navigate to the folder that contains the Pentaho component. Within that folder, locate a folder named licenses. The licenses folder contains HTML files that list the names of open source software, their licenses, and required attributions.

Contact Us

Global Headquarters Pentaho Corporation
Citadel International, Suite 340
5950 Hazelhine National Drive
Orlando, FL 32822
Phone: +1 407 812-OPEN (6736)
Fax: +1 407 517-4575
<http://www.pentaho.com>
Sales Inquiries: sales@pentaho.com

Contents

| | |
|---|----|
| Introduction..... | 4 |
| Get Started with Data Models..... | 5 |
| Tour the Data Source Wizard..... | 6 |
| Choose A Data Source Type..... | 8 |
| Create CSV Data Sources..... | 10 |
| Create SQL Query Data Sources..... | 13 |
| Create Database Sources..... | 15 |
| Create a Database Table Data Source for Reporting Only..... | 15 |
| Create a Database Table Data Source for Reporting and Analysis..... | 18 |
| Tour the Data Source Model Editor..... | 21 |
| Edit Data Source Models..... | 23 |
| Edit a Data Source Model..... | 23 |
| Delete a Data Source..... | 23 |
| Add Member Properties to a Level..... | 24 |
| Edit Multidimensional Data Source Models..... | 24 |
| Assign Geographic Properties..... | 24 |
| Assign Time Dimension Properties..... | 25 |
| Use Ordinal Columns..... | 27 |
| Properties of Time Dimension Levels..... | 27 |
| Next Steps..... | 29 |

Introduction

Before you can create Pentaho reports and analysis, the BA Server needs to know where to find your data and how to connect to it, what data you want to use, how the data is structured, and what the data is like. The Data Source Wizard helps you define a data source that contains this information and guides you through the creation of your first [relational](#) or [multidimensional](#) data models for use in creating reports and analysis.

After you define a data source, you can make it available to people who create reports and analysis so they can simply pick the data source they want to use from a list and begin creating their reports or analysis. One data source can be used by any number of reports or analysis.

Prerequisites

Before you can work with the Data Source Wizard, you must have [installed](#) the Pentaho software and [configured](#) the BA Server.

Expertise

The topics in this section are written for IT administrators who know how and where data is stored, how to connect to it, and details about the computing environment.

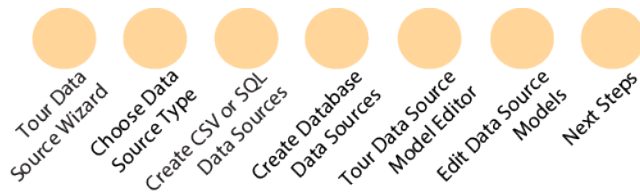
Tools

We provide a web application, the User Console, which you use to perform most data source and data modelling tasks.

Login Credentials

All of the tasks in this section require that you [log on to the User Console](#) with the Pentaho administrator user name and password.

Get Started with Data Models



When you configured your BA Server, you defined the connection information, including where the data is stored and what protocol or driver to use to connect to it. [Configure Pentaho Systems](#) shows you how to change or add connection information.

To create a data source, just pick one of the connections that you have already defined, then use the wizard's interactive data modeling tool to select tables and columns to drag onto the model canvas. You can then use the Data Source Model Editor to refine the model further.

Eventually, you might want to add more security, localization, or make other advanced modifications to your model. To do that, export the initial model created with the Data Source wizard, and import it into either [Metadata Editor](#) for relational data models and [Schema Workbench](#) for multidimensional data models.



Note: Relational and multidimensional models that have been edited in Metadata Editor and Schema Workbench can no longer be edited in the Data Source Model Editor.


It is time to take a look at the Data Source Wizard and see what all you can do with it. Just follow the *Guide Post* graphics as you work with the tutorials to create your first data models with the wizard.

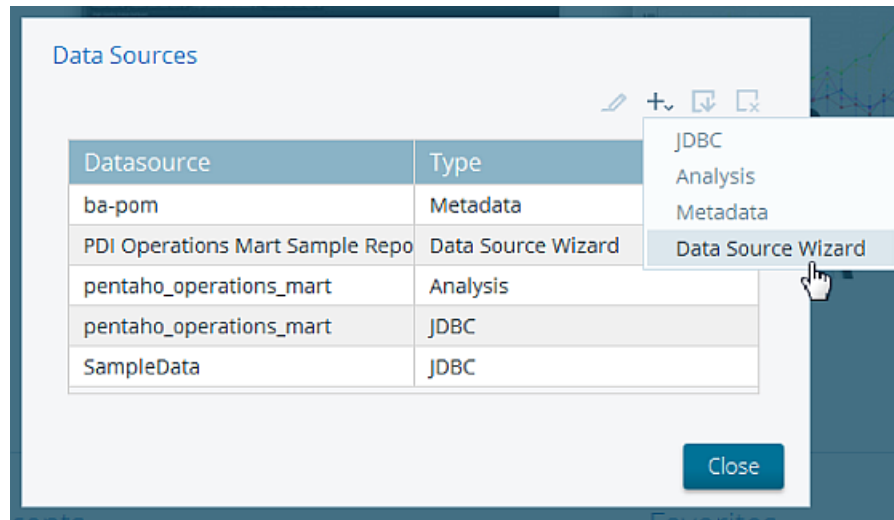
- [Tour the Data Source Wizard](#)
- [Choose a Data Source Type](#)
- [Create CSV Data Sources](#)
- [Create SQL Query Data Sources](#)
- [Create Database Sources](#)
- [Tour the Data Source Model Editor](#)
- [Refine Data Models](#)
- [Next Steps](#)

Tour the Data Source Wizard



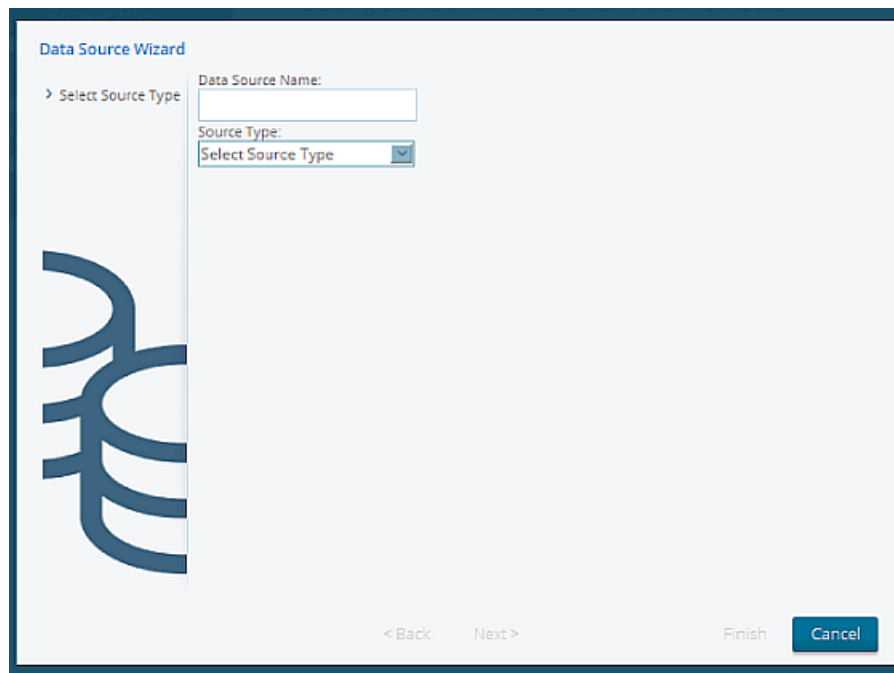
The User Console has a **Data Source Wizard** that guides you through setting up CSV files, as well as *relational* and *multidimensional* data models, as data sources for building Interactive and Analysis Reports. The initial data models you create with the wizard enable you to immediately see how your data looks in Interactive Reporting and Analyzer. Your results appear in Interactive Reporting and Analyzer as you change model structures, adding tables and columns with the wizard.

 **Note:** If you are not logged in with permissions to create, edit, and delete data sources, you are limited to view-only permissions and do not see any icons associated with adding, editing, or deleting data sources.



There are several ways to access the **Data Source Wizard** from within the **User Console**, so you do not have to backtrack to create a data source.

| If You Are In | Then Follow These Steps |
|--|---|
| Home page | Click Manage Data Sources > Add icon > Data Source Wizard |
| Home page | Click Create New > Analysis Report > Add icon > Data Source Wizard |
| Home page | Click Create New > Interactive Report > Add icon > Data Source Wizard |
| Any page, from the menu bar | Click File > New > Data Source |
| Any page, from the menu bar | Click File > Manage Data Sources > Add icon > Data Source Wizard |
| Dashboard pane, creating a chart or data table | Click the Add icon in the Select Data Source dialog box |



The image shows a 'Data Source Wizard' dialog box. On the left is a sidebar with a tree view icon and the text '> Select Source Type'. The main area contains two input fields: 'Data Source Name:' with a text box, and 'Source Type:' with a dropdown menu currently showing 'Select Source Type'. At the bottom, there are navigation buttons: '< Back', 'Next >', 'Finish', and a blue 'Cancel' button.

Data Source Wizard

> Select Source Type

Data Source Name:

Source Type:

Select Source Type

< Back Next > Finish Cancel

After you have accessed the Data Source Wizard, you are ready to begin creating your first data sources.

Choose A Data Source Type



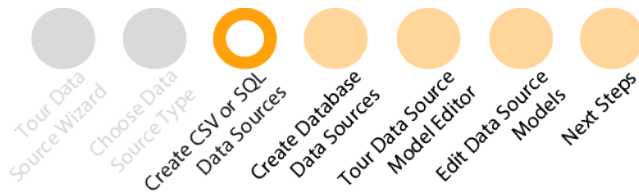
The **Data Source Wizard** supports several data source types for maximum flexibility. Each data source type is defined on where or how you store data. This table helps you decide which data source type is best for you.

Table 1: Data Source Type Decision Table

| Explore Considerations | Choose Options | | |
|------------------------|---|--|---|
| | <i>CSV Files</i> | <i>SQL Queries</i> | <i>Database Table Data Sources</i> |
| Summary | <p>A file with data that is separated by commas or some other specified delimiter character.</p> <p>You can save an Excel spreadsheet as a CSV file and use this file to stage a database table in the BA Repository. If the data in the CSV file changes, you have to upload the updated file through the wizard again.</p> <p>CSV files can be uploaded directly to the BA Repository without a data connection.</p> <p>This option creates a multidimensional model for Analyzer only and a relational model for Interactive Reports, Dashboard reports, and Analyzer.</p> | <p>An SQL query written against a relational database defines the data you want to use in Interactive Reports or Analyzer.</p> <p>You can customize how columns are presented to users who are building queries against the new data source; for example, you define column names and select options that indicate how data is aggregated (sum, min., max., etc.), and more.</p> <p>This option creates a multidimensional model for Analyzer only and a relational model for Interactive Reports, Dashboard reports, and Analyzer.</p> <p>SQL Queries are used for prototyping, not for production.</p> | <p>Database Table Data Sources allow you to create relational database tables containing the data you want to use in Interactive Reports or Analyzer.</p> <p>This data source type supports a broad range of use cases including relational reporting directly against an operational data store (ODS) or creating a data source for multidimensional analysis against a star schema.</p> <p>Database Table data sources are used in production environments.</p> |
| Expertise | Knowledge of basic database structures. | Knowledge of basic database structures and SQL. | Knowledge of relational and/or multidimensional database structures. |
| Time | Approximately 10 minutes | Approximately 20 minutes | Approximately 30 minutes to 1 hour |
| Recommendation | This data source makes it easier to evaluate or create development prototypes. | <p>This data source type works well when you are evaluating or creating development prototypes.</p> <p>SQL Queries are used for prototyping, not for production.</p> | <p>Use this data source for production.</p> <p>To use it for evaluating or creating development prototypes, limit the data in the tables to sample data that is representative of your production data.</p> |

| Explore Considerations | Choose Options | | |
|---------------------------|------------------|--|--|
| | <i>CSV Files</i> | <i>SQL Queries</i> | <i>Database Table Data Sources</i> |
| | | To use this data source in production, create the query using <i>Metadata Editor</i> . | |

Create CSV Data Sources



Before you begin working with CSV data sources, there are a few key terms that you should know.

- **Delimiter** - a character, such as a comma, used to specify a boundary between separate regions in a data stream.
- **Enclosure** - a container that holds a collection of other data objects.
- **Length** - indicates the maximum number of characters allowed in a field.
- **Precision** - the number of digits after a decimal point.

1. Login to the **User Console** as an administrator. Click **Manage Data Sources**.

The **Data Sources** window appears.

2. Click the **Add** icon, then choose **Data Source Wizard** from the drop-down menu.

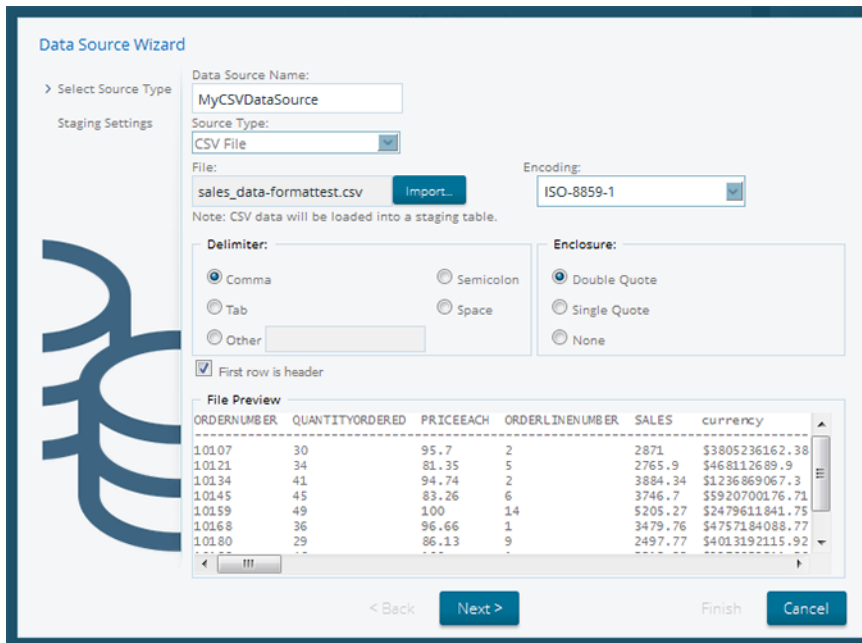
3. Enter a name that identifies your new data source in the **Data Source Name** field.

The following characters are not allowed in **Data Source Wizard** source names:

\$ < > ? & ; # % ^ * () ! ~ : ; [] { } |

4. Select **CSV File** from the **Source Type** drop-down menu.
5. Click **Import...** to browse for your CSV file. Double-click to select the CSV file you want to upload.
6. Choose your delimiter and enclosure types.

If you want to use the first row of your CSV file as headings for columns in the file, leave **First row is header** check box selected. If you want to use the first row as data, disable the **First row is header** check box.



Data Source Wizard

Select Source Type

Staging Settings

Data Source Name: MyCSVDataSource

Source Type: CSV File

File: sales_data-formattest.csv [Import...](#)

Encoding: ISO-8859-1

Note: CSV data will be loaded into a staging table.

Delimiter: ☒ Comma ☐ Semicolon ☐ Tab ☐ Space ☐ Other

Enclosure: ☒ Double Quote ☐ Single Quote ☐ None

☒ First row is header

File Preview

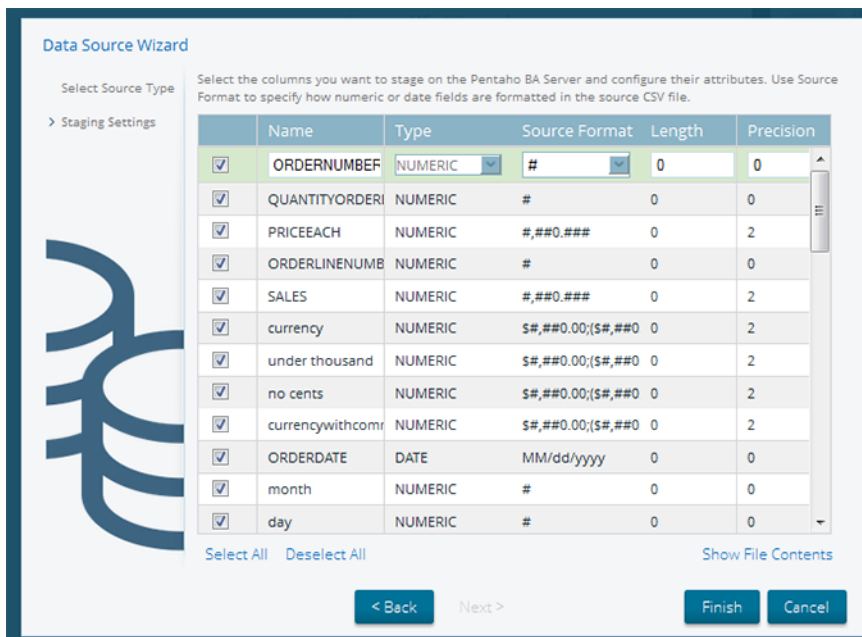
| ORDERNUMBER | QUANTITYORDERED | PRICEEACH | ORDERLINENUMBER | SALES | currency |
|-------------|-----------------|-----------|-----------------|---------|-----------------|
| 1.01.07 | 30 | 95.7 | 2 | 2871 | \$3805236162.38 |
| 1.01.21 | 34 | 81.35 | 5 | 2765.9 | \$468112689.9 |
| 1.01.34 | 41 | 94.74 | 2 | 3884.34 | \$1236869067.3 |
| 1.01.45 | 45 | 83.26 | 6 | 3746.7 | \$5920700176.71 |
| 1.01.59 | 49 | 100 | 14 | 5205.27 | \$2479611841.75 |
| 1.01.68 | 36 | 96.66 | 1 | 3479.76 | \$4757184088.77 |
| 1.01.80 | 29 | 86.13 | 9 | 2497.77 | \$4013192115.92 |

< Back [Next >](#) Finish [Cancel](#)

The **File Preview** window displays the first few lines of your CSV file based on the selections you made for the delimiter, enclosure, and header. Once the columns align correctly in the preview, the delimiter and enclosure have been set correctly.

7. Click **Next**.

The **Staging Settings** screen displays a list of columns from your CSV source file. All columns are enabled.



Data Source Wizard

Select Source Type

Staging Settings

Select the columns you want to stage on the Pentaho BA Server and configure their attributes. Use Source Format to specify how numeric or date fields are formatted in the source CSV file.

| | Name | Type | Source Format | Length | Precision |
|-------------------------------------|-------------------|---------|-------------------------|--------|-----------|
| <input checked="" type="checkbox"/> | ORDERNUMBER | NUMERIC | # | 0 | 0 |
| <input checked="" type="checkbox"/> | QUANTITYORDERED | NUMERIC | # | 0 | 0 |
| <input checked="" type="checkbox"/> | PRICEEACH | NUMERIC | #,##0.### | 0 | 2 |
| <input checked="" type="checkbox"/> | ORDERLINENUMBER | NUMERIC | # | 0 | 0 |
| <input checked="" type="checkbox"/> | SALES | NUMERIC | #,##0.### | 0 | 2 |
| <input checked="" type="checkbox"/> | currency | NUMERIC | \$#,##0.00;(\$#,##0.00) | 0 | 2 |
| <input checked="" type="checkbox"/> | under thousand | NUMERIC | \$#,##0.00;(\$#,##0.00) | 0 | 2 |
| <input checked="" type="checkbox"/> | no cents | NUMERIC | \$#,##0.00;(\$#,##0.00) | 0 | 2 |
| <input checked="" type="checkbox"/> | currencywithcomma | NUMERIC | \$#,##0.00;(\$#,##0.00) | 0 | 2 |
| <input checked="" type="checkbox"/> | ORDERDATE | DATE | MM/dd/yyyy | 0 | 0 |
| <input checked="" type="checkbox"/> | month | NUMERIC | # | 0 | 0 |
| <input checked="" type="checkbox"/> | day | NUMERIC | # | 0 | 0 |

[Select All](#) [Deselect All](#) [Show File Contents](#)

< Back [Next >](#) Finish [Cancel](#)

8. Choose the columns that you want to use in your data source, either individually or by clicking **Select All**.

You can deselect all columns by clicking **Deselect All**.

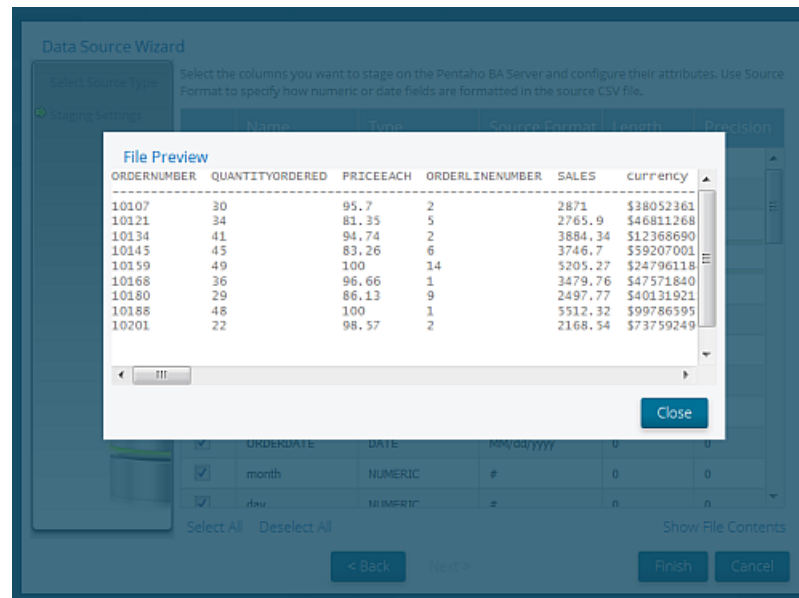
9. Change the **Name** and **Type** values, if applicable.

- Choose the options that you want to use from the drop-down menu for dates and numeric values.
- You can enter a value manually in the **Source Format** text box.



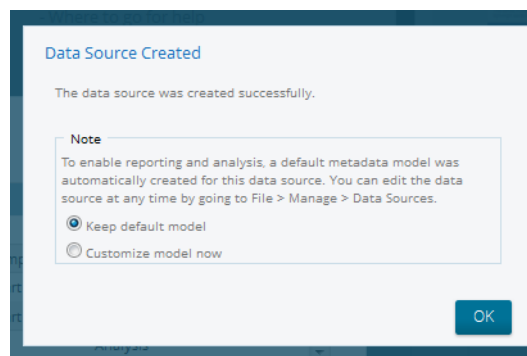
Note: Drop-down lists are not enabled for certain data types such as the String data type. Boolean values are rendered as "true" or "false."

10. Click **Show File Contents** to look at a sample of the data in your source file. Click **Close** to return to the Staging Settings screen.



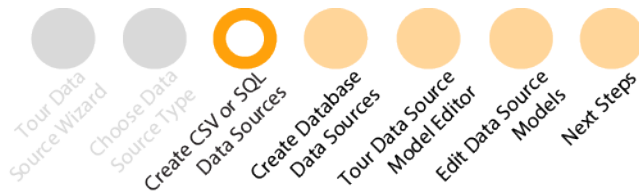
11. Continue to work with your CSV data settings or click **Finish**.

The **Data Source Created** window appears. You can choose to **Keep default model** or click **Customize model now** to launch the Data Source Model Editor and refine the model. Click **OK**.



Your new data source is now available for use in Analyzer, Interactive Reports, and Dashboard reports, or the **Data Source Model Editor** appears..

Create SQL Query Data Sources



Once you create the data source, it is available to users of Interactive Reports, Analyzer, and Dashboard reports.

1. Login to the **User Console** as an administrator. Click **Manage Data Sources**.

The **Data Sources** window appears.

2. Click the **Add** icon, then choose **Data Source Wizard** from the drop-down menu.
3. Enter a name that identifies your new data source in the **Data Source Name** field.

The following characters are not allowed in **Data Source Wizard** source names:

\$ < > ? & ; # % ^ * () ! ~ : ; [] { } |

Data Source Wizard

> Select Source Type

Data Source Name: MyFirstDataSource

Source Type: SQL Query

Connection:

SampleData

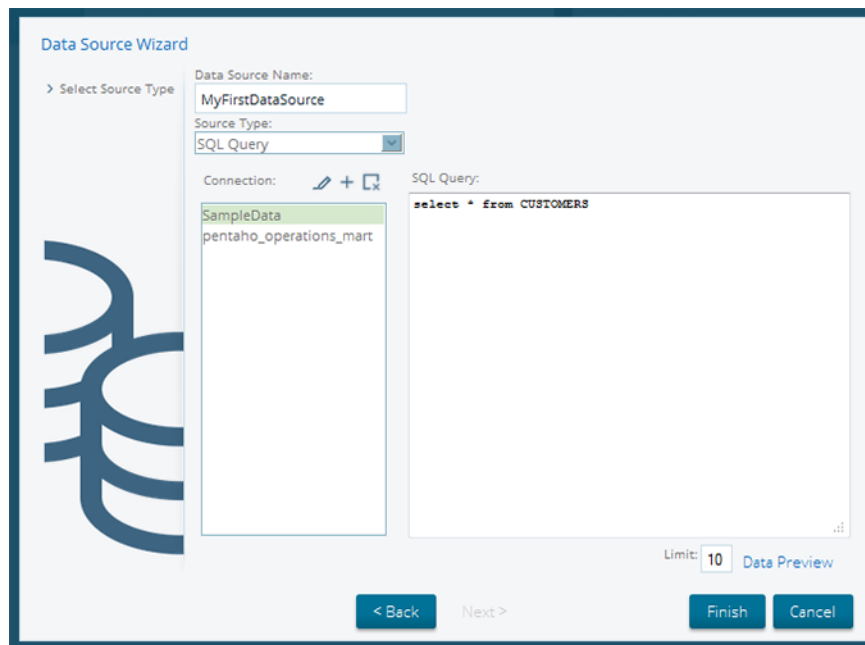
pentaho_operations_mart

SQL Query:

Limit: 10 Data Preview

< Back Next > Finish Cancel

4. Select **SQL Query** from the **Source Type** drop-down menu. Click **Next**.
5. Select a database connection from the list, under **Data Connection**.
6. Enter your SQL Query in the **SQL Query** text field. Click **Data Preview** to make sure that your query returns data. The SQL query `select * from CUSTOMERS` is shown in this example.

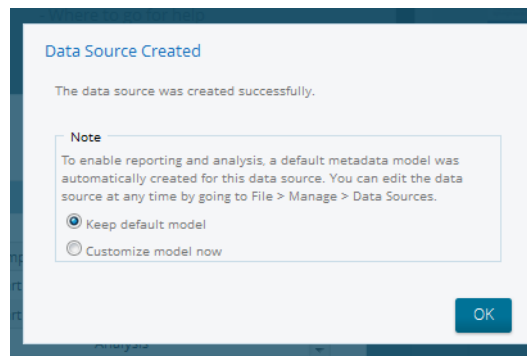


The **Data Source Wizard** dialog box is shown. It has a sidebar on the left with a tree view containing 'SampleData' and 'pentaho_operations_mart'. The main area contains the following fields:

- Data Source Name:** MyFirstDataSource
- Source Type:** SQL Query
- Connection:** A list box with 'SampleData' and 'pentaho_operations_mart'. 'SampleData' is selected.
- SQL Query:** A text area containing the query: `select * from CUSTOMERS`
- Limit:** 10
- Data Preview:** A button next to the limit field.
- Navigation:** '< Back' and 'Next >' buttons.
- Buttons:** 'Finish' and 'Cancel' buttons.

7. Click **Close** to exit the Preview window. Click **Finish**.

The **Data Source Created** window appears. You can choose to **Keep default model** or click *Customize model now* now to launch the Data Source Model Editor and refine the model. Click **OK**.

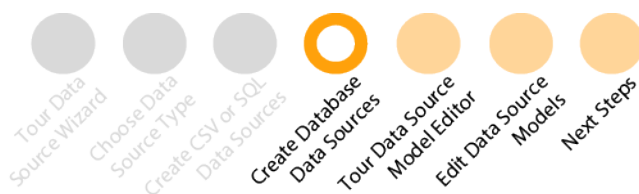


The **Data Source Created** dialog box is shown. It contains the following elements:

- Title:** Data Source Created
- Message:** The data source was created successfully.
- Note:** To enable reporting and analysis, a default metadata model was automatically created for this data source. You can edit the data source at any time by going to File > Manage > Data Sources.
- Options:**
 - ☒ Keep default model
 - ☐ Customize model now
- Button:** OK

A relational model is generated based on the SQL query for use in Analyzer, Interactive Reports, and Dashboard reports, or the **Data Source Model Editor** appears.

Create Database Sources



When you create a Database Table data source, you are presented with two options: Reporting Only or Reporting and Analysis. The choice you make depends on the structure of the database tables you are accessing and the User Console tools being used.

Before you begin working with Database Table data sources, there are a few key terms that you should know.

- **Fact table** - records measurements or metrics for a specific event, and usually consists of numbers.
- **Star schema** - has one or more fact tables that reference any number of dimension tables. Star schemas are good for simpler queries.
- **Snowflake schema** - has centralized fact tables that are connected to multiple dimensions. Snowflake schemas are good for more complex queries.

Table 2: Database Data Sources Decision Table

| Explore Considerations | Choose Options | |
|------------------------|--|---|
| | <i>Reporting Only</i> | <i>Reporting and Analysis</i> |
| Summary | <p>Choose the Reporting Only option if you are accessing data in a relational database that is operational or transactional in nature.</p> <p>When you choose the Reporting Only option, a relational model is automatically generated and can be refined in the Data Source Model Editor or further extended using Metadata Editor.</p> | <p>Choose the Reporting and Analysis option if you are accessing a multidimensional database that contains tables arranged in a star schema (tables are separated into dimensions related to a single fact table, such as a data warehouse).</p> <p>When you choose the Reporting and Analysis option, a multidimensional model is automatically generated and can be refined in the Data Source Model Editor or further extended using Schema Workbench.</p> <p>Additionally, you can create a Reporting and Analysis data source using a single table where the table acts as the single fact table and also contains dimensional data.</p> |
| Expertise | Knowledge of relational database structures. | Knowledge of multidimensional database structures. |
| Time | Approximately 15 minutes. | Approximately 30 minutes. |
| Recommendation | Select this if you want to use for Report Designer, Interactive Reports, and Metadata Editor. | Select this if you want to use for Report Designer, Analyzer (PUC), Dashboard Reports, Schema Workbench, and Aggregation Designer. |

Create a Database Table Data Source for Reporting Only

Your database must be up and running in order for you to complete the steps required in the Data Source Wizard.

1. Login to the **User Console** as an administrator. Click **Manage Data Sources**.

The **Data Sources** window appears.

2. Click the **Add** icon, then choose **Data Source Wizard** from the drop-down menu.
3. Enter a name that identifies your new data source in the **Data Source Name** field.

The following characters are not allowed in **Data Source Wizard** source names:

\$ < > ? & ; # % ^ * () ! ~ : ; [] { } |

Data Source Wizard

Select Source Type

Data Source Name: DataSourceTable

Source Type: Database Table(s)

Select a database connection and click Next to choose from a list of the available database tables.

Connection:

- SampleData
- pentaho_operations_mart

Create data source for:

☒ Reporting only

☐ Reporting and Analysis (Requires Star Schema)

< Back Next > Finish Cancel

4. Select **Database Table(s)** from the **Source Type** drop-down menu. Under **Connection**, click to choose a database connection.
 5. Click to select the **Reporting Only** radio button. Click **Next**.
 6. Choose a schema from the **Schema** drop-down menu.
- A list of **Available Tables** appears.

Data Source Wizard

Select one table to finish or select multiple tables and click Next to define their joins.

Schema: PUBLIC

Available Tables:

- "PUBLIC"."CUSTOMER_V_W_TER"
- "PUBLIC"."DEPARTMENT_MANAGERS"
- "PUBLIC"."DIM_TIME"
- "PUBLIC"."EMPLOYEES"
- "PUBLIC"."OFFICES"
- "PUBLIC"."ORDERDETAILS"
- "PUBLIC"."PAYMENTS"
- "PUBLIC"."PRODUCTS"
- "PUBLIC"."QUADRANT_ACTUALS"
- "PUBLIC"."TRIAL_BALANCE"

Selected Tables:

- "PUBLIC"."ORDERFACT"
- "PUBLIC"."ORDERS"
- "PUBLIC"."CUSTOMERS"

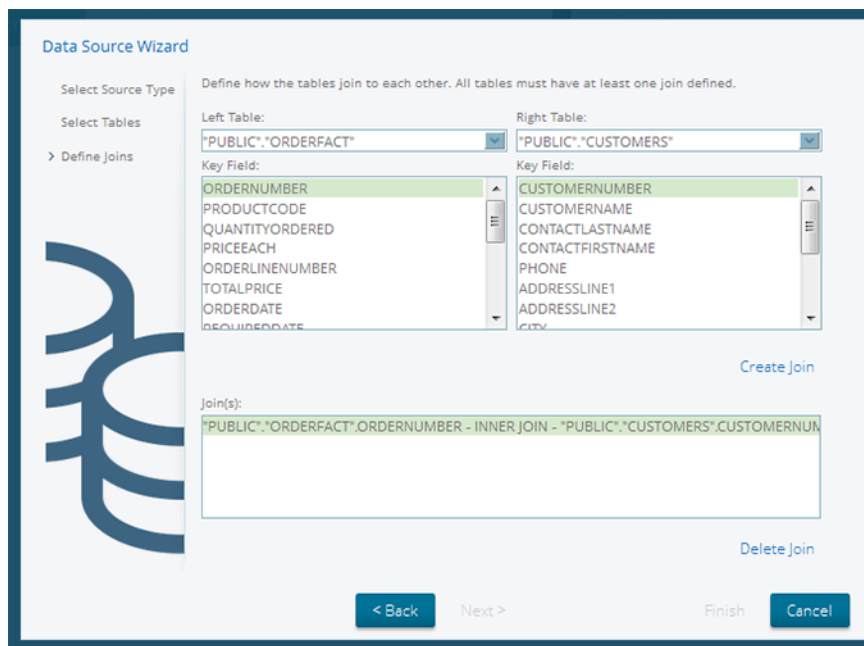
< Back Next > Finish Cancel

7. Choose a table from **Available Tables** and click the right arrow (Add) to move the table into the **Selected Tables** field.

If you add a table you decide not to use, highlight the table and click the left arrow (Remove) to remove it from the list of **Selected Tables**. To choose multiple tables, press the CTRL key down as you make your selections.

8. Click **Next**.

The **Define Joins** window appears. While you can create a data source from a single table, the more common scenario is that you will choose multiple tables which must then be joined.



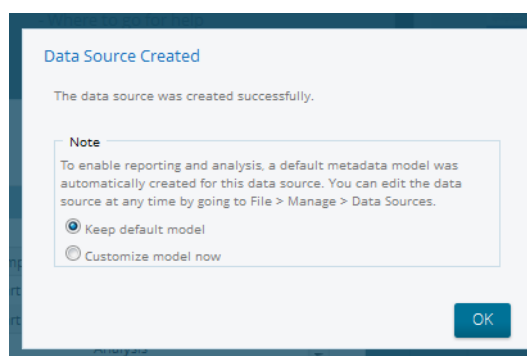
9. Choose the fact table from the **Left Table** drop-down menu, then choose a table to join to it from the **Right Table** drop-down menu. Click **Create Join** to define the inner join between each table.

- a) Create join conditions for each entry in the **Left Table** and **Right Table** drop-down menus.

The join relationship for each table is based on the field selected and is displayed in the lower portion of the dialog box. To delete a join, select the join and click **Delete Join**.

10. Click **Finish**.

The **Data Source Created** window appears. You can choose to **Keep default model** or click [Customize model now](#) now to launch the Data Source Model Editor and refine the model. Click **OK**.



A relational model is generated for use in Interactive Reports and Dashboard reports, or the **Data Source Model Editor** appears.

Create a Database Table Data Source for Reporting and Analysis

Your database must be up and running in order for you to complete the steps required in the Data Source Wizard.

1. Login to the **User Console** as an administrator. Click **Manage Data Sources**.

The **Data Sources** window appears.

2. Click the **Add** icon, then choose **Data Source Wizard** from the drop-down menu.
3. Enter a name that identifies your new data source in the **Data Source Name** field.

The following characters are not allowed in **Data Source Wizard** source names:

\$ < > ? & ; # % ^ * () ! ~ : ; [] { } |

Data Source Wizard

> Select Source Type
Select Tables
Define Joins

Data Source Name:
ReportingAndAnalysis

Source Type:
Database Table(s)

Select a database connection and click Next to choose from a list of the available database tables.

Connection:

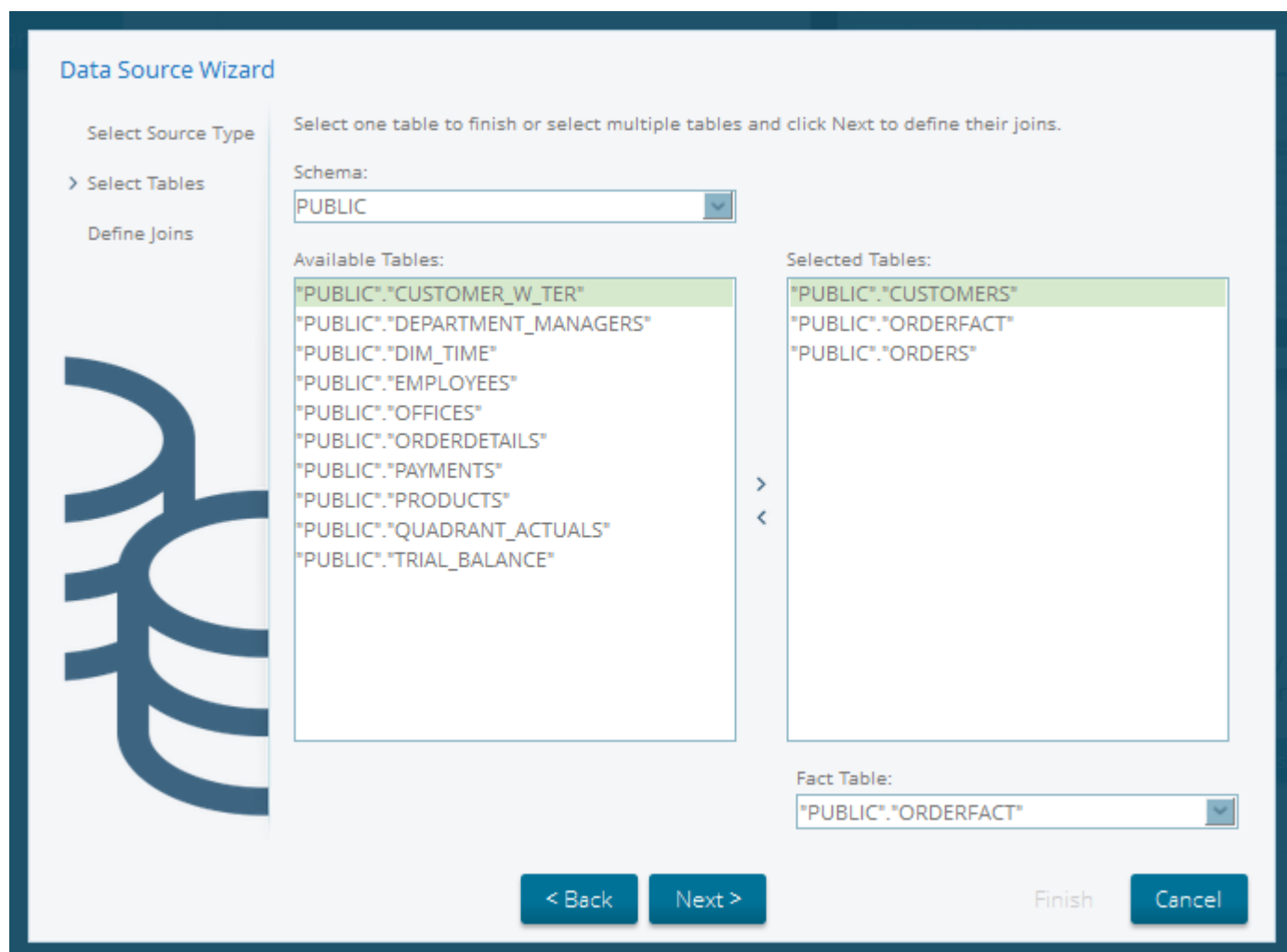
SampleData
pentaho_operations_mart

Create data source for:

☐ Reporting only
☒ Reporting and Analysis (Requires Star Schema)

< Back Next > Finish Cancel

4. Select **Database Table(s)** from the **Source Type** drop-down menu. Under **Connection**, click to choose a database connection.
5. Click to select **Reporting and Analysis**. Click **Next**.
6. Choose a schema from the **Schema** drop-down menu.
A list of **Available Tables** appears.



The screenshot shows the 'Data Source Wizard' window. On the left, a sidebar contains three steps: 'Select Source Type', '> Select Tables' (which is highlighted), and 'Define Joins'. The main area has a title bar and a subtitle: 'Select one table to finish or select multiple tables and click Next to define their joins.' Below this, there is a 'Schema:' dropdown menu set to 'PUBLIC'. Two list boxes are present: 'Available Tables' on the left and 'Selected Tables' on the right. The 'Available Tables' list includes: 'PUBLIC"."CUSTOMER_W_TER', 'PUBLIC"."DEPARTMENT_MANAGERS', 'PUBLIC"."DIM_TIME', 'PUBLIC"."EMPLOYEES', 'PUBLIC"."OFFICES', 'PUBLIC"."ORDERDETAILS', 'PUBLIC"."PAYMENTS', 'PUBLIC"."PRODUCTS', 'PUBLIC"."QUADRANT_ACTUALS', and 'PUBLIC"."TRIAL_BALANCE'. The 'Selected Tables' list includes: 'PUBLIC"."CUSTOMERS', 'PUBLIC"."ORDERFACT', and 'PUBLIC"."ORDERS'. Between these two lists are right and left arrow buttons. Below the 'Selected Tables' list is a 'Fact Table:' dropdown menu set to 'PUBLIC"."ORDERFACT'. At the bottom, there are four buttons: '< Back', 'Next >', 'Finish' (disabled), and 'Cancel'.

7. Choose a table from **Available Tables** and click the right arrow (Add) to move the table into the **Selected Tables** field.

If you add a table you decide not to use, highlight the table and click the left arrow (Remove) to remove it from the list of **Selected Tables**. To choose multiple tables, press the CTRL key down as you make your selections.

8. Choose a **Fact Table** from the drop-down list and click **Next**.

While you can create a data source from a single table, the more common scenario is that you will choose multiple tables which must then be joined.

Data Source Wizard

Select Source Type

Select Tables

> Define Joins

Define how the tables join to each other. All tables must have at least one join defined.

Left Table: "PUBLIC".ORDERFACT

Right Table: "PUBLIC".ORDERS

Key Field: ORDERNUMBER

Key Field: ORDERNUMBER

ORDERDATE
REQUIREDDATE
SHIPPEDDATE
STATUS
COMMENTS
CUSTOMERNUMBER

Create Join

Join(s):

"PUBLIC".ORDERFACT.ORDERNUMBER - INNER JOIN - "PUBLIC".CUSTOMERS.CUSTOMERNUM

"PUBLIC".ORDERFACT.ORDERNUMBER - INNER JOIN - "PUBLIC".ORDERS.ORDERNUMBER

Delete Join

< Back Next >

Finish Cancel

9. Choose the fact table from the **Left Table** drop-down menu, then choose a table to join to it from the **Right Table** drop-down menu. Click **Create Join** to define the inner join between each table.

a) Create join conditions for each entry in the **Left Table** and **Right Table** drop-down menus.

The join relationship for each table is based on the field selected and is displayed in the lower portion of the dialog box. To delete a join, select the join and click **Delete Join**.

10. Click **Finish**.

The **Data Source Created** window appears. You can choose to **Keep default model** or click **Customize model now** now to launch the Data Source Model Editor and refine the model. Click **OK**.

Data Source Created

The data source was created successfully.

Note

To enable reporting and analysis, a default metadata model was automatically created for this data source. You can edit the data source at any time by going to File > Manage > Data Sources.

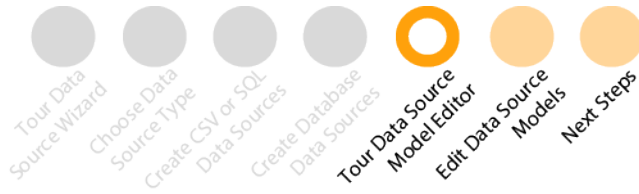
☒ Keep default model

☐ Customize model now

OK

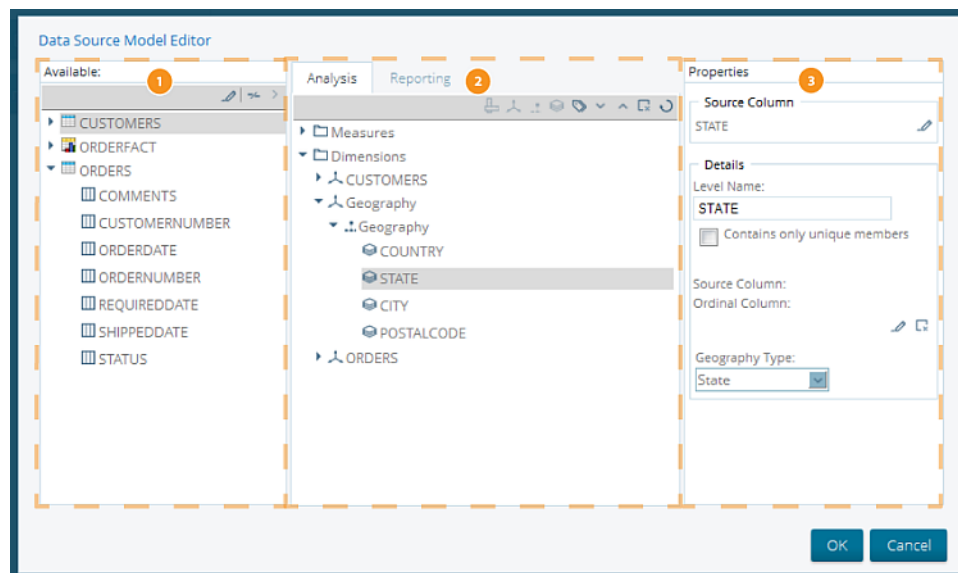
A multidimensional model is generated for use in Analyzer, Interactive Reports, and Dashboard reports, or the **Data Source Model Editor** appears.

Tour the Data Source Model Editor



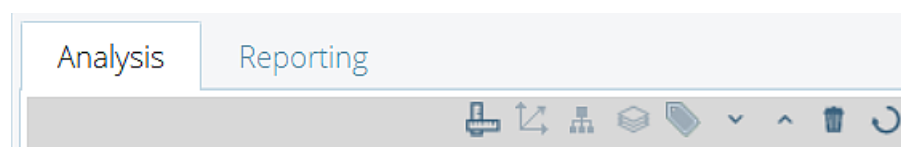
The **Data Source Model Editor** allows you to change relational and multidimensional data models to make your data more accessible. The model editor displays if you choose to customize your model in the last step of the **Data Source Wizard** or if you edit the data source later by clicking **Manage Data Sources**, selecting a data source, and clicking the **Edit** icon.

When you first open the model editor, the center pane of the dialog box displays the active model. This model is organized into categories and fields that represent the tables and columns in your data.



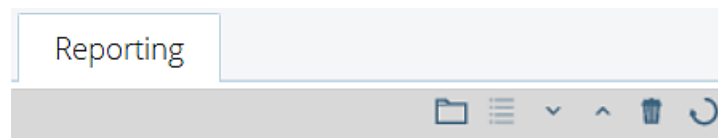
| Item | Name | Function |
|------|--|---|
| 1 | Available pane | Displays a list of the available tables and columns of the data source. |
| 2 | Analysis / Reporting model pane | Displays the active data source model. |
| 3 | Properties pane | Displays the properties associated with a selected category or field. |

Analysis Toolbar Tasks



| Icon Name | Function | Definition |
|----------------------------|--|--|
| Add Measure | Add a measure to the model. | A <i>measure</i> is a property on which calculations can be made, such as a sum, count, or average. |
| Add Dimension | Add a dimension to the model. | A <i>dimension</i> is used to provide filtering, grouping, or labeling capabilities. A dimension is related to fact data, and is typically something like a product model, color, size, geographic location, salespeople's names, or a unit of time. |
| Add Hierarchy | Add hierarchies to a dimension. | <i>Hierarchies</i> organize data by strict ranks. |
| Add Level | Add levels to a hierarchy. | A <i>level</i> is a set of objects within a hierarchy that have the same rank or importance. |
| Add Member Property | Add custom member properties to a level. | <i>Member properties</i> are the attributes of a level. |
| Move Up / Down | Move levels up and down in a hierarchy. | Moving <i>levels up</i> and <i>down</i> allow you to change their ranks within the hierarchy. |
| Remove field | Removes a field from the model. | n/a |
| Clear Model | Clears the entire model. | n/a |

Reporting Toolbar Tasks



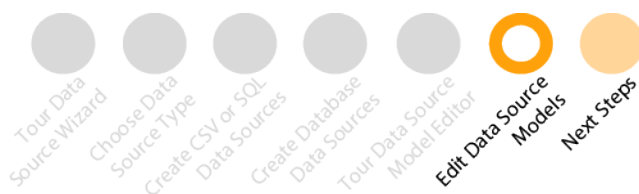
| Icon Name | Function | Definition |
|----------------------------|--|---|
| Add Category | Add a category to the model. | A <i>category</i> is a structure that contains objects that are linked. |
| Add Field | Add a field to a category. | A <i>field</i> is the place where data is stored within a category. |
| Move Up / Down | Move levels up and down in a hierarchy. | Moving <i>levels up</i> and <i>down</i> allow you to change their ranks within the hierarchy. |
| Remove field | Remove a field. | n/a |
| Add Member Property | Add custom member properties to a level. | <i>Member properties</i> are the attributes of a level. |
| Clear Model | Clears the entire model. | n/a |

Data Source Model Editor Usage Tips

Creating a data source from a CSV file treats all columns as if they are in a single table, with no limitations on where measures come from or which fields can be grouped in a hierarchy.

Creating a data source from a database table for Reporting and Analysis requires you to specify a fact table. Only numeric columns from the fact table can be used as measures. Also, all levels in a single hierarchy must come from the same dimension table.

Edit Data Source Models



The **Data Source Model Editor** guides you through editing relational and multidimensional models created using the **Data Source Wizard**. The **Data Source Model Editor** gives you the ability to move fields by dragging them to the appropriate location, mix and match fields from different tables, add fields to more than one category, or remove a field altogether.

You can edit the models with the **Data Source Model Editor** or use [Schema Workbench](#) or [Metadata Editor](#) for more advanced modifications.



Note: Relational and multidimensional models that have been edited in Metadata Editor and Schema Workbench can no longer be edited in the Data Source Model Editor.

Here are some things you can do to refine your relational and multidimensional data models with the **Data Source Model Editor**.

- [Edit a Data Source Model](#)
- [Delete a Data Source](#)
- [Add Member Properties](#)

Here are some things that you can do specifically to refine your multidimensional data models.

- [Assign Geographic Properties](#)
- [Assign Time Dimension Properties](#)
- [Use Ordinal Columns](#)
- [Properties of Time Dimension Levels](#)

Edit a Data Source Model

Use these steps to edit a data source model in the **User Console**.

1. Open the [Data Source Model Editor](#) by clicking **Manage Data Sources**, selecting a data source, and clicking the **Edit** icon.

The **Data Source Model Editor** appears.

2. Select the **Reporting** tab to edit a Reporting relational model, or the **Analysis** tab to edit the Analysis multidimensional model.
3. You can now edit the model or clear the model and start building a model from scratch.
For instance, you can choose a field and drag it to the appropriate folder on the **Analysis/Reporting** model pane. This causes that field to be displayed and used when designing reports.
4. Click **OK** when you are done editing your model.

The model builds and appears in the **Data Source** list.

Delete a Data Source

When you delete a data source, all reports or charts that end-users created using that data source no longer render.

1. Open the Data Sources window by clicking **Manage Data Sources** on the console **Home** page, selecting a data source, and clicking the **Edit** icon.
2. Click on a data source in the list to highlight it.
3. Click on the **Remove** icon in the Data Sources tool bar.

The **Remove Data Source** window appears.

4. Click **Remove** to delete the data source.

If you choose not to delete the data source, you can click **Cancel** to exit the window.

The data source is removed and is no longer available to your users. All reports or charts using that data source no longer work.

Add Member Properties to a Level

Within the **Data Source Model Editor**, you can add member properties to any level of a data source model for use in generating Interactive and Analyzer Reports.

1. Click the **Analysis** tab.
2. Select a field from the **Available** pane to the **Analysis** pane.
3. With the field selected, click the **Add Member Property** icon.
The **New Member Property** dialog appears.
4. Enter a name for the new member property. Click **OK**.
5. Set the source column for the new member property. Expand the level with new member property and choose the member property.
A yellow caution icon appears next to the parent dimension, hierarchy, and level of the new child member property. The icon indicates you need to specify the source column for the new member properties before proceeding.
6. Specify the source column for the new member property. Click **Fix Missing Column**.
The **Select Source Column** dialog appears.
7. Chose a column from the available list. Click **OK**.

Edit Multidimensional Data Source Models

For multidimensional models, the **Data Source Model Editor** helps you add custom properties to a level, such as geographic or time dimensions. You can also use the Data Source Model Editor to perform tasks such as moving fields by dragging them, setting properties to a measure, adding levels to a hierarchy, moving levels up and down within the hierarchy, and removing fields.

- [Assign Geographic Properties](#)
- [Assign Time Dimension Properties](#)
- [Use Ordinal Columns](#)
- [Properties of Time Dimension Levels](#)

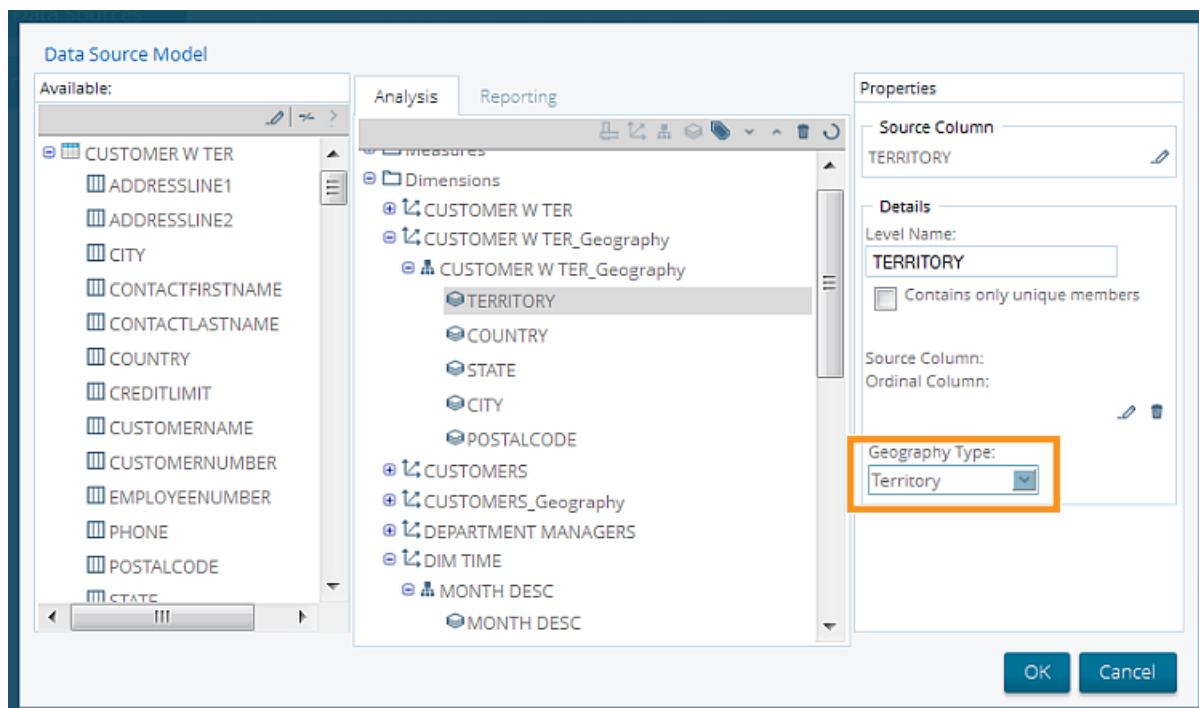
Assign Geographic Properties

Analyzer lets you visualize data on a geographic map. For common geography types, such as Country, State, City, or Postal Code, the [Data Source Wizard](#) automatically populates geographic parent-dimensions and subsequent child-levels to use in Analyzer.

To refine the use of Geography Map visualizations in Analyzer, you can associate custom data fields with a particular geography type. After you identify the fields that represent a geography type, you can use Analyzer to visualize this geographic information on a map.

These steps show you how to assign geographic properties to a level with the **User Console**.

1. Within the Data Source Model Editor, click the **Analysis** tab.
2. Select each of the geographical levels individually and assign each level a **Geography Type** in the **Properties** pane. This classifies the data with the correct geography type.



The geography type is now defined for that level.

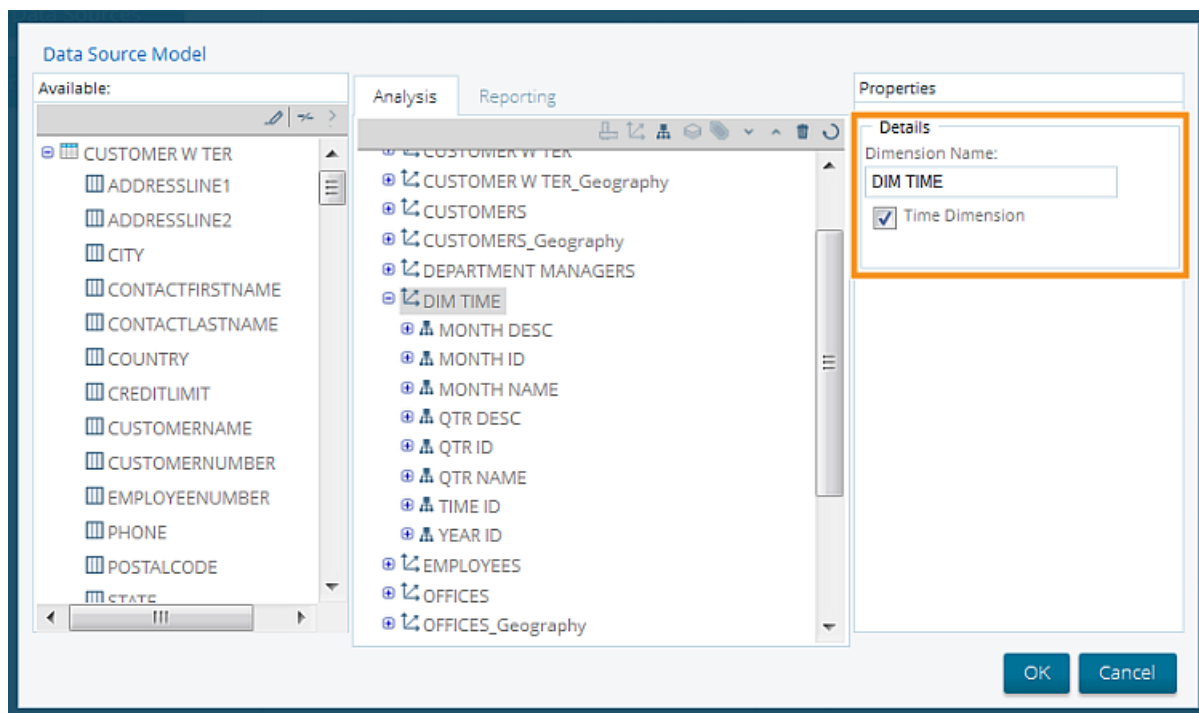
Assign Time Dimension Properties

Typically, you might sort your data by year or month. However, by designating a time dimension in the **Data Source Model Editor**, you can filter on a variety of date ranges and relative date options, such as previous periods, before [a user-defined period], after [a user-defined period], current period, or next period.

Analyzer supports many types of relative date filters. In order to apply them to a level of a time hierarchy, you need to define the time-specific properties for that level. This is because each data warehouse implementation may have a different date format and set of time hierarchy levels.

These steps show you how to assign a time property to a dimension in a multidimensional model.

1. Within the **Data Source Model Editor**, click the **Analysis** tab.
2. Select the *dimension* you want to assign as a time dimension.
3. Within the **Properties** pane on the right, click the **Time Dimension** box to assign the dimension as a time dimension.

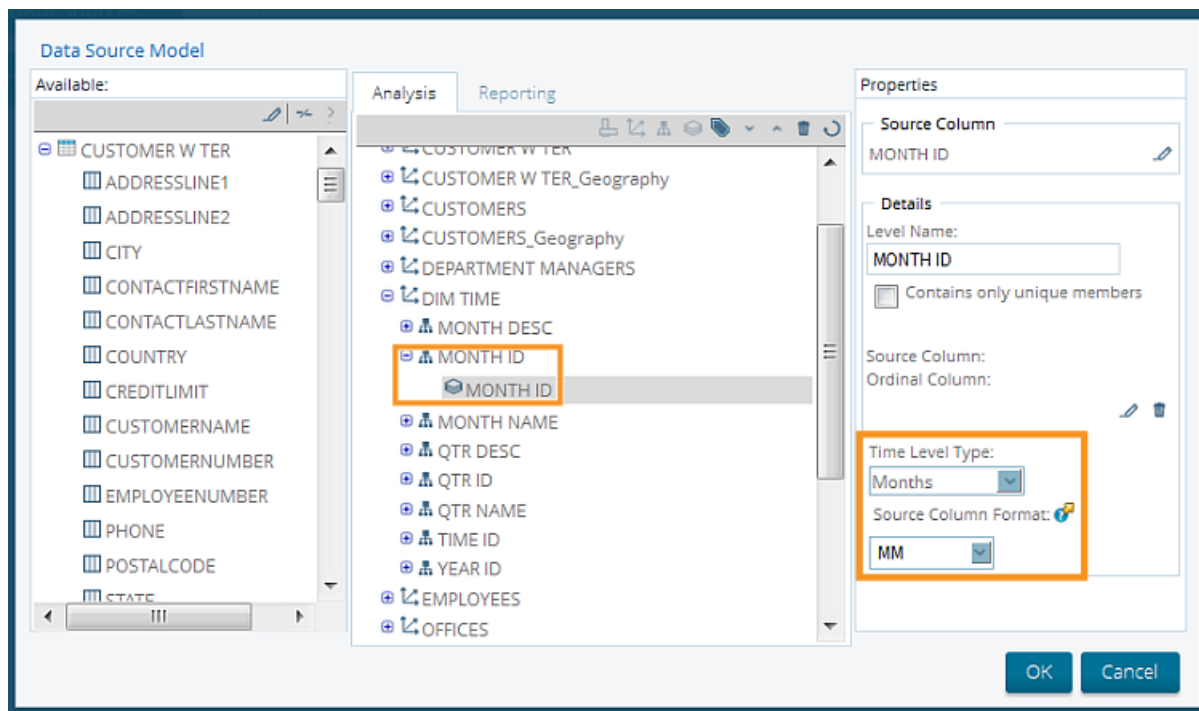


- Expand the dimension to display its hierarchy, then expand the hierarchy to display its levels. Choose the level for which you want to assign *time dimension properties*. The properties of the level display in the **Properties** pane.

The options for **Time Level Type** and **Source Column Format** do not display for a child-level if the time dimension property is not set for its parent-dimension.

- Set the **Time Level Type** and **Source Column Format** to match the format of your data.

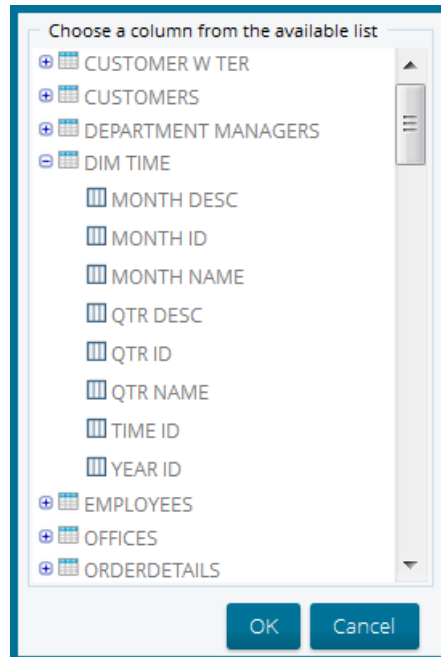
The **Time Level Type** and **Source Column Format** must match the format of the data stored in your data source for relative date filtering to work properly in Analyzer. This does not effect how your values display. For more information regarding the time level types and formats, see the section *Properties of Time Dimension Levels*.



Use Ordinal Columns

Ordinal columns are used to alter the natural order of a level when fetching members from a database.

For example, the natural alphabetical order of the level `month_name` begins with April and ends with September, when represented by a text string. Alternatively, setting the dimension `month_name` so it is sorted by an ordinal column, such as `month_id`, enables you to sort the values in the level `month_name` in chronological order.



Select the appropriate column to use as the ordinal column, then click **OK**.

To set an ordinal column, in the Data Source Model Editor, choose the appropriate level, then click **Edit** in the **Properties** pane under **Ordinal Column**. The **Select Ordinal Column** dialog box appears.

Properties of Time Dimension Levels

The Time Level Type and Source Column Format drop-down menus in the **Properties** pane of the Data Source Model Editor allow you to specify how time-specific data is formatted in your data source.

Time Level Type specifies the role of a level and the increment of the date it represents. Time Level Type is used by a number of time-period functions in Analyzer.

Source Column Format field lets you specify how a value in the source column is formatted and represents a date increment belonging to that level. This information is used when filtering in Analyzer to the levels of a time dimension.

Assigning the Source Column Format in the Data Source Model Editor does not change how the values in your data display, rather it enables you to describe the format of your data.

Here are several commonly used formats for time measurements.

- Years: "YY", "YYYY"
- Quarters: "Q", "QQ", "QQQ"
- Months: "M", "MM", "MMM"
- Weeks: "w", "ww", "W"
- Days: "d", "dd", "D", "DDD", "YYYY-MM-dd"
- Hours: "k", "kk", "H", "HH", "K", "KK"
- Minutes: "m", "mm"
- Seconds: "s", "ss"

If the options provided in the drop-down menu do not reflect the format of your data, you can override the formats by entering directly into the Source Column Format field. However, the format of your data must adhere to the [ICU Simple Date Format](#) specification.

For instance, a level may have `Quarters` as Time Level Type, indicating the level corresponds to quarter date increments. If that value is represented by the numbers 1 through 4, you would set the Source Column Format as `Q`.

Alternatively, the same increment could also be represented as strings `Q1`, `Q2`, `Q3`, and `Q4`. In this case you specify the Source Column Format with the value `'Q'Q`. The first `'Q'` indicates a string, while the second, unquoted `Q` represents a numerical value, 1 through 4.

Or, your source column could contain values like `2001-Q1`, `2001-Q2`, `2001-Q3`, `2001-Q4`, in which case you would input `YYYY-'Q'Q` into the Source Column Format field.

Next Steps



After you have finished creating your first data models, you are ready to learn how to create [web-based interactive and analysis reports](#).

Learn More

- The *Create and Refine Advanced Data Models* section of the [Pentaho InfoCenter](#) provides information on using [Metadata Editor](#), [Schema Workbench](#), and [Aggregation Designer](#) to work with relational and multi-dimensional models for production environments.
- [Create Analysis, Interactive Reports, and Dashboards](#) provides in-depth details about creating eye-catching business intelligence deliverables for your user community.
- The *Build Full-Featured Solutions* section of the [Pentaho InfoCenter](#) contains a large number of topics on working with the BA Suite. Here you can learn about defining data models, fine-tuning your reports, how to create print-quality reports, or work with Big Data.
- [Work with Big Data](#) tells you how to use big data as a data source.