Introducing the Pentaho BI Suite 3.8 Community Edition

Help and Support Resources

If you have questions that are not covered in this guide, or if you would like to report errors in the documentation, 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 on the topics covered in this guide, visit http://www.pentaho.com/training.

Limits of Liability and Disclaimer of Warranty

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.

Company Information

Pentaho Corporation
Citadel International, Suite 340
5950 Hazeltine National Drive
Orlando, FL 32822
Phone: +1 407 812-OPEN (6736)
Fax: +1 407 517-4575
http://www.pentaho.com

E-mail: communityconnection@pentaho.com

Sales Inquiries: sales@pentaho.com

Documentation Suggestions: documentation@pentaho.com

Sign-up for our newsletter: http://community.pentaho.com/newsletter/

Contents

Introduction

The Pentaho BI Suite Community Edition is an open source business intelligence package that includes ETL, analysis, metadata, and reporting capabilities. It is entirely open source software, licensed mostly under the GNU General Public License version 2, with parts under the LGPLv2, the Common Public License, and the Mozilla Public License. Pentaho optimizes, platform-tests, and guarantees certified builds of the BI Suite; this enhanced version of the software is packaged with a powerful service management tool called Enterprise Console, user and developer support, IP indemnification, and professional documentation, and sold by Pentaho as Enterprise Edition.

The purpose of this guide is to introduce new users to the Pentaho BI Suite, explain how and where to interact with the Pentaho community, and provide some basic instructions to help you get started using the software.

Community Edition or Enterprise Edition?

The BI Suite Community Edition is ideal for:

- Business intelligence aficionados
- Open source software programmers
- Early adopters
- College students

Pentaho no longer suggests using Community Edition for enterprise evaluations. If you are a business user interested in trying out the BI Suite Enterprise Edition, follow the Enterprise Edition evaluation link on the pentaho.com front page, or contact a Pentaho sales representative.

The enhancements, service, and support packaged with the BI Suite Enterprise Edition are designed to accommodate production environments, especially where downtime and time spent figuring out how to install, configure, and maintain a business intelligence solution are prohibitively expensive. If your business will save money or make more money as a result of a successful business intelligence implementation, then Enterprise Edition is the most appropriate choice.

Community Edition Support Options

As a Pentaho BI Suite Community Edition user, you will have to install, configure, and maintain the software on your own. Your only support options are the community forum (*http://forums.pentaho.org*) and the community Wiki (*http://wiki.pentaho.com*). If you do not find an answer right away, please be a good community participant and contribute a Wiki article that explains the solution after you've figured it out.

At any time, you can contact Pentaho sales and upgrade to Enterprise Edition. Enterprise Edition customers get phone support, access to Pentaho software engineers, and a Web-based knowledge base that is updated weekly with new support articles, tips, and comprehensive user guides.

The Pentaho Client Tools

The Pentaho client tools are:

- **Report Designer:** An advanced report creation tool. If you want to build a complex data-driven report, this is the right tool to use. Report Designer offers far more flexibility and functionality than the ad hoc reporting capabilities of the Pentaho User Console.
- **Design Studio:** An Eclipse-based tool that enables you to hand-edit a report or analysis view xaction file. Generally, people use Design Studio to add modifications to an existing report that cannot be added with Report Designer.
- **Aggregation Designer:** A graphical tool that helps improve Mondrian cube efficiency.
- **Metadata Editor:** Enables you to add a custom metadata layer to an existing data source. Usually you would do this for a data source that you intend to use for analysis or reporting; it's not required, but it makes it easier for business users to parse the database when building a query.
- **Pentaho Data Integration:** The Kettle extract, transform, and load (ETL) tool, which enables you to access and prepare data sources for analysis, data mining, or reporting. This is generally where you will start if you want to prepare data for analysis.
- **Schema Workbench:** A graphical tool that helps you create ROLAP schemas for analysis. This is a required step in preparing data for analysis.

After they're installed, you can find all of these tools in their own directories in the `/pentaho/design-tools/` directory. The scripts that run them should be fairly self-explanatory. If you are using Windows, there should be a Pentaho program group with icons that will initialize the BI Server and run the client tools.

Installation

Follow the instructions below to download and install the Pentaho BI Suite Community Edition.

Hardware Requirements

The Pentaho BI Suite does not have strict limits on computer or network hardware. As long as you meet the minimum software requirements (note that your operating system will have its own minimum hardware requirements), Pentaho is hardware agnostic. There is, however, a recommended set of minimum system specifications:

Server:

RAM	At least 2GB
Hard drive space	At least 2GB for the software, and more for solution and content files
Processor	Dual-core AMD64 or EM64T

Workstation:

RAM	At least 1GB
Hard drive space	At least 1GB for the software, and more for solution and content files
Processor	Dual-core AMD64 or EM64T

It's possible to use less capable machines, but in most realistic scenarios, the too-limited system resources will result in an undesirable level of performance.

Your environment does not have to be 64-bit, even if your processor architecture supports it; while all modern desktop, workstation, and server machines have 64-bit processors, they often ship by default with 32-bit operating systems. If you want to run the Pentaho BI Suite in a pure 64-bit environment, you will have to install a 64-bit operating system, ensure that your solution database and Java Runtime Environment are 64-bit, and install the BI Suite via the archive-based or manual deployment methods.

Note: A 32-bit JRE has a hard memory limit of 2GB (1.5GB on Windows), so if you have 2GB or more of RAM, you must use a 64-bit JRE on a 64-bit operating system to take full advantage of it.

Software Requirements

Note: The system requirements listed below apply to the BI Suite. The BI Suite graphical installation utility, however, will only work on Windows or Linux.

In terms of operating systems, Windows (XP SP2, 2008, 7) modern Linux distributions (SUSE Linux Enterprise Desktop and Server 10 and Red Hat Enterprise Linux 5 are officially supported, but most others should work), Solaris 10, and Mac OS X 10.5 are all officially supported.

The BI Suite installer will provide you with a Sun Java Runtime Environment (JRE) version 1.6 (sometimes referenced as version 6.0) installed. If you have other versions installed, they will be ignored.

Workstations will need to have reasonably modern Web browsers to access Pentaho's Web interface. Internet Explorer 7 or higher; Firefox 3.8 or higher (or the Mozilla or Netscape equivalent); and Safari 4 or higher will all work.

Your environment can be either 32-bit or 64-bit as long as it meets the above requirements.

The aforementioned configurations are officially supported by Pentaho. Other operating systems such as FreeBSD and OpenBSD; other Java virtual machines like Blackdown; other application servers such as Liferay and Websphere; and other Web browsers like Opera may work without any problems. However, the Pentaho support team will probably not be able to help you if you have trouble installing or using the BI Suite under these conditions.

Note: Some Pentaho client tools, such as Metadata Editor, Aggregation Designer, Pentaho Data Integration, and Design Studio, require that the Eclipse SWT JAR be in your Java classpath. This can be an issue in scenarios where standalone client tools are installed onto a machine that does not also have the BI Server installed, particularly on platforms other than Windows and Linux.

Note: The Pentaho Reporting engine requires a graphical environment in order to create charts. If you are installing the BI Server onto a headless Linux, BSD, or Solaris server and do not have X11R6 on it, you should install the **Xvfb** package on your server to satisfy the charting dependency.

Downloading and Installing the BI Suite

Follow the below process to download and install the Pentaho BI Suite Community Edition.

1. Open a Web browser and navigate to the Pentaho page on SourceForge.net.

   http://sourceforge.net/projects/pentaho/. If you cannot click on links in this document, you can simply navigate to http://sourceforge.net/projects/pentaho/

2. Click **Download**.

3. At the SourceForge download screen, click **Business Intelligence Server**.

4. In the **Latest** category at the top of the list, click either the .zip or .tar.gz file for the **biserver-ce** project.

   This is an archive package of the Pentaho BI Platform, along with a Tomcat Java application server configured to run it. There is no functional difference between the zip and tar archives; they are merely in compression formats that are generally preferred by Windows and Linux users, respectively.

5. Once the file is downloaded, create a `/pentaho/server/` directory in an accessible place in your filesystem, and unpack the files using your preferred archive utility.

   Ideally you would be unpacking this on what you expect to be your server, though there is no reason why you can't install the Pentaho client tools on the same machine.

6. Repeat this process (creating a `/pentaho/design-tools/` directory to contain them) for any or all of the following Pentaho client tool projects:

   - Report Designer
   - Pentaho Metadata
   - Design Studio
   - Data Integration
   - Schema Workbench

   You may not need all of these tools, but it can't hurt to download all of them.

You have retrieved all of the relevant Pentaho software, and are ready to configure the BI Platform.

Starting the BI Platform

In order to use and configure the Pentaho BI Platform, you must start the BI Server, then the Pentaho Administration Console.

Starting the BI Server

To start the BI Server, run the **start-pentaho.bat (or .sh)** script in the `/pentaho/server/biserver-ce/` directory.

Starting the Pentaho Administration Console

To start the Pentaho Administration Console, run the **start-pac.bat (or .sh)** script (on Windows) or **startup** script (on Linux) in the `/pentaho/server/administration-console/` directory.

Configuring the BI Platform With the Administration Console

Follow the below process to log into the Pentaho Administration Console.

1. Open a Web browser and type in the Web or IP address of the Pentaho Administration Console server, which is `http://localhost:8099/admin` by default.

2. Type in your administrator credentials, then click **Login**.

   The default credentials are **admin** for the user, and **password** for the password.

3. Click the **Administration** tab on the left side of the window.

4. Remove the default sample users and roles and create your own.
5. Click the **Data Sources** tab at the top of the window.
6. Enter the connection details for the data source you want to use for reporting and analysis.

   By default, there is a **sampledata** source listed. If you intend to follow the examples later in this guide, you must leave this data source intact.

You are now logged into the Pentaho Administrator Console and ready to finish configuring the BI Platform.

Allowing Other Computers to Access the BI Platform

To allow other computers to access the Pentaho BI Platform, the base URL must be set in the **web.xml** file.

1. Locate the web.xml file, its default location is: `pentaho-demo/jboss/server/default/deploy/` `pentaho.war/WEB-INF`.
2. Open **web.xml** in an XML or text editor.
3. Set the param-value of the context parameter base-url with the URL for the server. The URL should have the DNS Name or IP address of the machine, the port default 8080 and the context pentaho.
   The final result should look like the example below:

```
<context-param>
    <param-name>base-url</param-name>
    <param-value>http://myserver.pentaho.org:8080/pentaho</param-value>
</context-param>
```

Getting Started

Your workflow will vary depending on your BI goals. Typically, Pentaho BI Suite users will start with Pentaho Data Integration to prepare a data source, then use Metadata Editor to create a metadata layer for that data source, then potentially Schema Workbench to create a ROLAP schema. At that point, you're ready to create reports and analysis views.

If you just want to create a quick report, the ad hoc reporting component of the Pentaho User Console is the best tool for the job. If you want to create a detailed report, go directly to Report Designer instead. If you have created a ROLAP schema, then you can do some data exploration first by using an analysis view, which allows you to drill down into the smallest of details in a data source.

Ideally, everything will end up being published to the BI Platform, which enables you to display, run, and share your reports with others, or to schedule them to run at given intervals.

Once you've got some reports and/or analysis views that you like, you might create some dashboards that display them in creative and useful ways for your business users.

Follow the instructions below to log into the Pentaho User Console and familiarize yourself with its graphical interface.

How to Log Into the Pentaho User Console

Follow the below process to log into the Pentaho User Console.

1. Open a Web browser and type in the Web or IP address of the Pentaho server, which is `http://localhost:8080/pentaho/` by default.

   You'll see an introductory screen with some Pentaho-related information and a **Login** button in the center of the screen.
2. Click **Login**.

   The login dialog appears.
3. For the locally installed version of the BI Suite, select **Joe** from the user drop-down box, and type in **password** into the password field, then click **Login**. For hosted demo users, select **Guest** and type in **guest** as the password instead.

You are now logged into the Pentaho User Console and ready to start creating and running reports.

Navigating the Pentaho User Console

The first thing you will see when logging in is the quick launch screen, shown here:

If you'd like to experiment on your own before continuing on to the tutorials, click one of the four icons in the center of the screen to start a new analysis view, create a new ad hoc report, edit existing solutions, add a new data source.

The button bar near the top of the page also contains icons for creating new ad hoc reports and analysis views, along with a button to print the current report or analysis view, and to open a previously saved solution.

Different user roles have different levels of access in the Pentaho User Console. The menu above the button bar performs these same functions as the buttons, plus administrative actions if you are logged in as an administrator, and also offers access to My Workspace and external links to help and support resources.

The three buttons in the quick launch screen will appear when you log into the Pentaho User Console for the first time, and when you close all tabs in the solution browser.

You can change views between My Workspace and the solution browser at any time by clicking the rightmost icons in the top button bar, or through the **View** menu.

Tutorials

The sections that follow provide, in no specific order, basic tutorials for the three major pillars of the BI Suite: Reporting, analysis, and data integration. These tutorials assume that you are working with the included sample data source, and that you have Report Designer and Pentaho Data Integration installed, and that you are logged into the Pentaho User Console.

Report Designer Tutorials

The tutorials below are for Pentaho Report Designer.

Starting Report Designer

Follow the process below to start Report Designer.

1. Click the **Report Designer** entry in the **Pentaho** folder in the **Programs** section of your Start menu, or navigate to the **/pentaho/design-tools/report-designer/** directory and run **report-designer.bat** on Windows, or **report-designer.sh** on Linux.
2. Before the program starts, it runs a version checking utility. You can click either option in the version checker screen to start Report Designer – if you just downloaded this file, it is assuredly the most current version, so you may not need to enable this feature right now.

After the version check is complete (or skipped), Report Designer will start. Report Designer displays a Welcome screen and a default workspace at startup. The Welcome screen provides you with a brief introduction to the program, some instructions for getting started, and access to sample content and recently opened reports. For this walkthrough, you won't be following the instructions on the Welcome screen, though if you would like to experiment with Report Designer a little before continuing, feel free.

Exploring the Report Designer Interface

Report Designer's interface is similar to that of other graphic design and layout tools. A typical menu and button bar are at the top of the screen; a tool palette for adding design elements is on the left; and on the right are two panes that contain data and structure elements, and show the properties of a selected report element.

The palette on the left side of the main window is where most of your design tools are located:

The Structure, Data, Style, and Attribute panes show your report elements, data sources, and their configurable options:
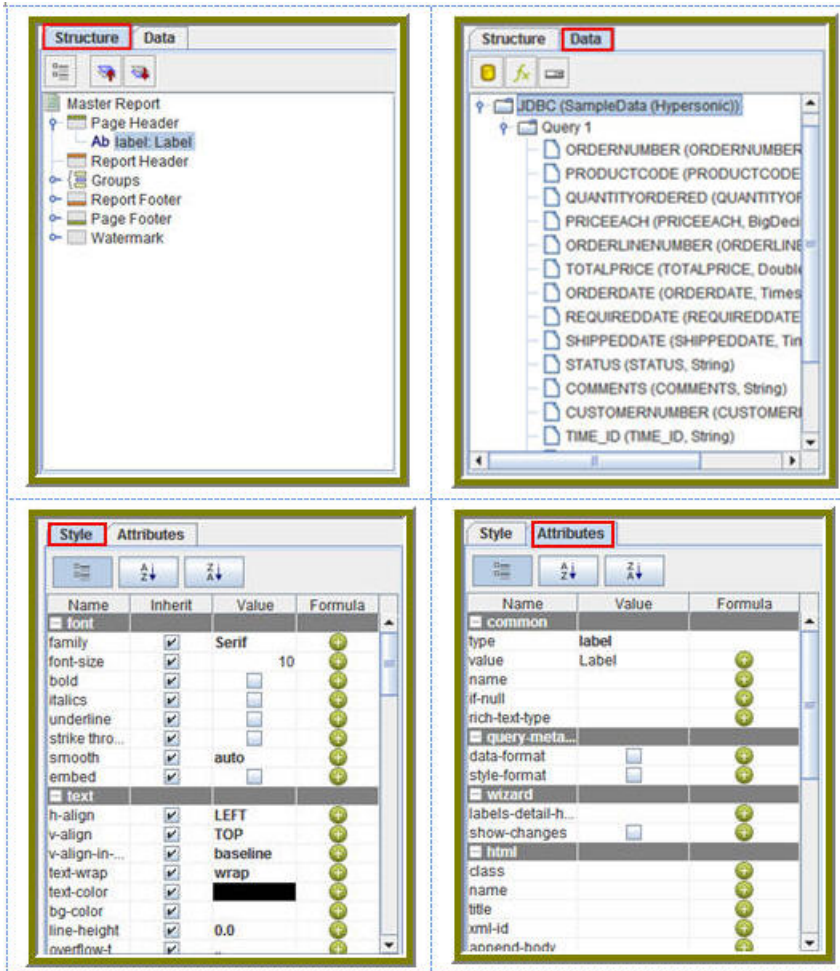
Creating a Report Using Report Designer

Pentaho Reporting provides unmatched deployment flexibility. Whether you're looking for a standalone desktop reporting tool, Web-based reporting, or comprehensive business intelligence (BI) including reporting, analysis, and dashboards, Pentaho Reporting allows you to "start small" and scale up if your reporting needs grow in the future.

The Pentaho Report Designer provides you with the following features:

- Drag-and-drop graphical designer that gives users full control of data access, layout, grouping, calculations, charting and formatting for pixel-perfect reports
- Integrated, step-by-step wizard that guides report designers through the design process
- Report templates that accelerate report creation and provide consistent look-and-feel

The Report Designer allows you to create a report by following a four step wizard; however, to show you a larger range of features, the exercises that follow walk you through the manual procedures for creating a simple report. Keep in mind that is basic tutorial and will not provide details about advanced Report Designer features.

1. Start the Report Designer. Go to **Start** -> **Programs** -> **Pentaho Enterprise Edition** -> **Design Tools** -> **Report Designer.**
    The Report Designer home page appears.
2. Click **New Report** in the **Welcome** dialog box.
    The design workspace appears.
3. In the right pane, click the **Data** tab.
4. For the purpose of this exercise, right-click **Data Sets** and choose **JDBC**. Alternatively, you can click the yellow database icon to display the JDBC dialog box.
    The **JDBC Data Source** dialog box appears.
5. Under **Connections**, select **SampleData (Memory)**.
6. Next to **Available Queries** click (Add).

    Query 1 appears under **Available Queries**. Notice that the edit icon is enabled.
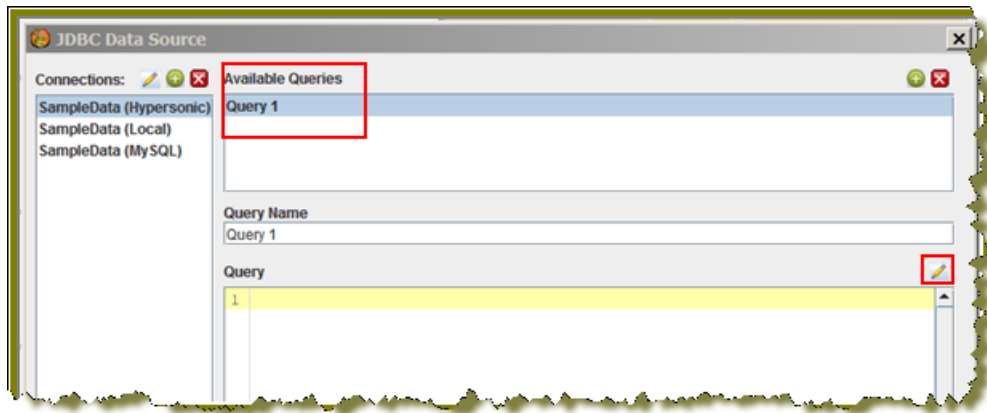
7. Click ✏ (Edit).

The Query Designer window appears. The Query Designer provides you with a graphical environment that allows you to work with the data even if you don't understand SQL, the standard programming language for retrieving content from databases.
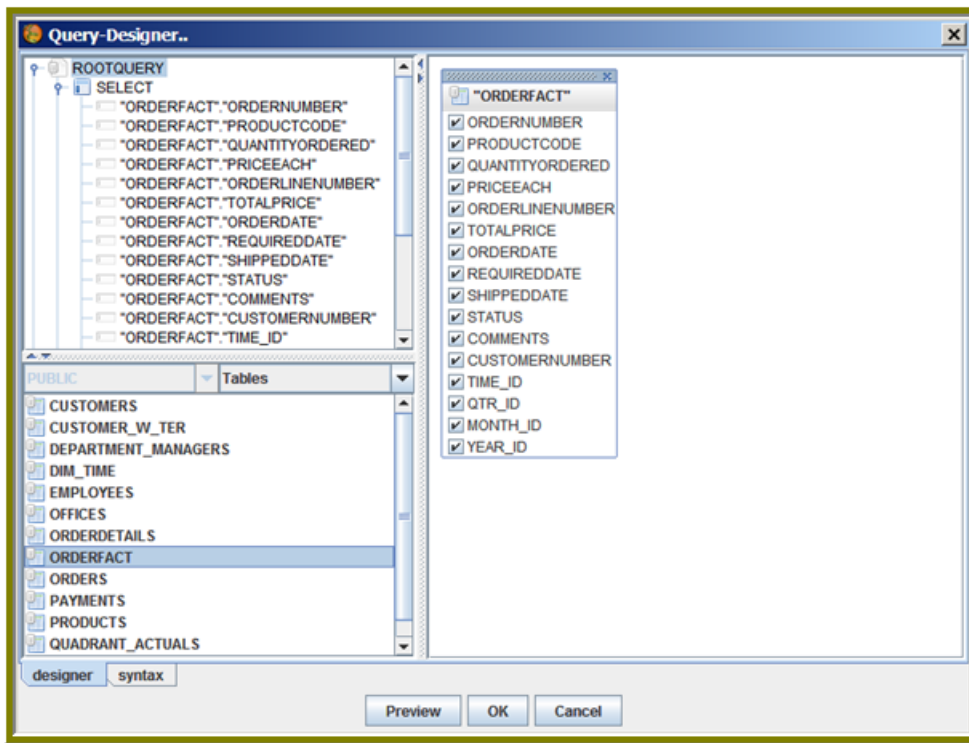
8. Double-click **ORDERFACT** so that the table appears in the workspace as shown in the image above.

9. In the Query Designer workspace, right-click **"ORDERFACT"** and choose **deselect all**.
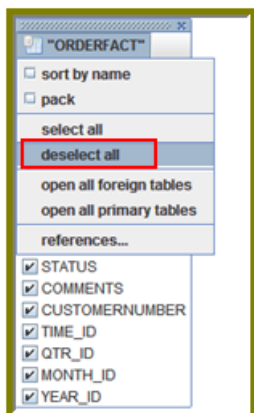
10. Now, select the following fields in the **ORDERFACT** table: **ORDERNUMBER**, **QUANTITYORDERED**, **PRICEEACH**, and **ORDERDATE**.
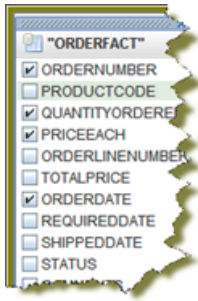
11. Double-click **PRODUCTS** so that the table appears in the workspace.

    Notice that there is a line that joins the **ORDERFACT** and **PRODUCTS** tables together.

12. Deselect all PRODUCTS table fields, except for **PRODUCTNAME** and **PRODUCTLINE**.
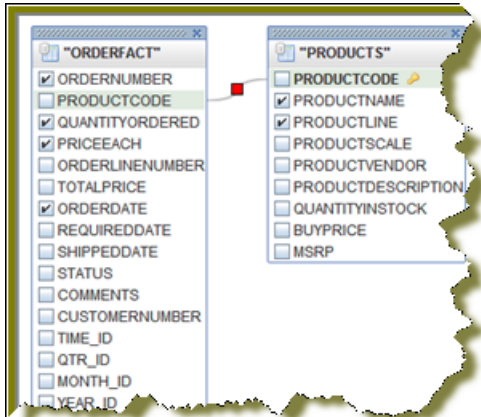
13. For the purpose of this exercise, click the **Syntax** tab in the lower left portion of the Query Designer workspace to display a simple SQL statement associated with the tables. Notice that **PRODUCTCODE** is the common field between the **ORDERFACT** and **PRODUCTS** tables.


```
SELECT
     "ORDERFACT"."ORDERNUMBER",
     "ORDERFACT"."QUANTITYORDERED",
     "ORDERFACT"."PRICEEACH",
     "ORDERFACT"."ORDERDATE",
     "PRODUCTS"."PRODUCTNAME",
     "PRODUCTS"."PRODUCTLINE"
FROM
     "PRODUCTS" INNER JOIN "ORDERFACT" ON "PRODUCTS"."PRODUCTCODE" = "ORDERFACT"."PRODUCTCODE"
```

14. Click **OK** in the syntax window to return to the JDBC Data Source dialog box. Notice that the SQL statement appears on the right under **Query**.

15. In the JDBC Data Source dialog box, click **OK** to go to the **Design** page.

    Notice that the fields associated with your tables are listed under **Query 1**. You are now ready to start *designing your report*.
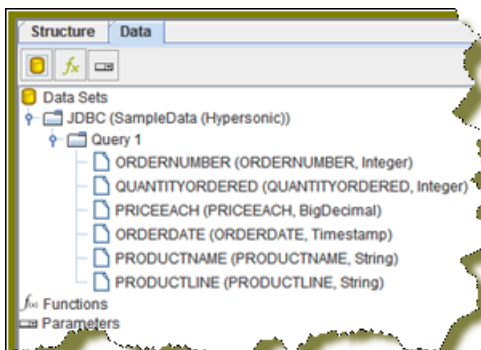
Designing Your Report
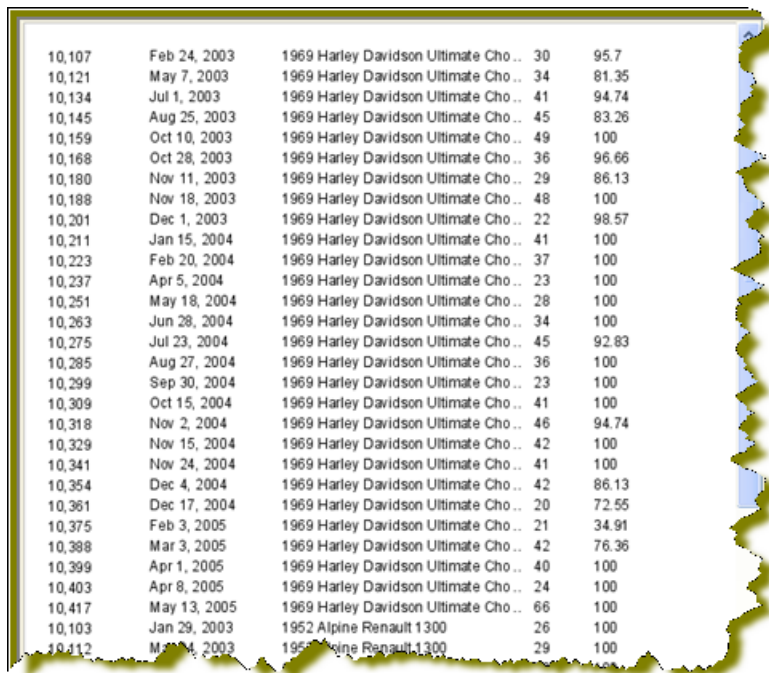
This exercise walks you through the process of designing the look-and-feel of your report.

1. Under the **View** item in the Report Designer menubar, click **Element Alignment Hints** and **Snap to Elements** to enable them. These options help you to align the elements of your report.

2. In the **Design** page, under **Query 1**, double-click and drag the **ORDERNUMBER** field into the **Details** band. Make sure that the top line of the field name and the top line of the Details band match up.

3. Place the **ORDERDATE**, **PRODUCTNAME**, **QUANTITYORDERED**, and **PRICEEACH** fields into the Details Band. Take care not to overlap the fields or your report will not display correctly.

4. Use the resizing handles to make the **PRODUCTNAME** field larger and the **QUANTITYORDERED** field smaller as shown in the example below:
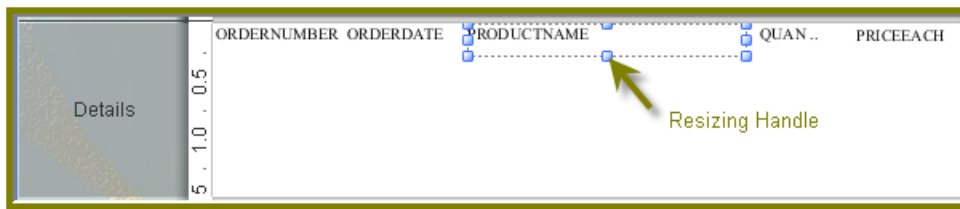
5.

   You have created your first report. Click 👁 (Preview) to examine your report. Click ✏️ (Edit) to return to the workspace view.

   👉 **Tip:** You can also click 👁 (Preview) on the left side of your workspace or select it from the **View** menu option to preview a report. Click ✏️ (Edit) to return to the workspace view.


   But, wait... There's a problem. Without headers, report users will have a hard time understanding its content. You need to continue *refining your report*.

Refining Your Report

You've created a report in the previous exercise but now you need to make the report more descriptive so that users can understand the content in the report. Follow the instructions below to refine your report.

1.
   Click ✏️ (Edit) to return to the Design page.

2. Click and drag a 🔤 (Label) from the tools palette into the middle of the **Page Header** band. Notice how Report Designer keeps track of the report structure (shown below).
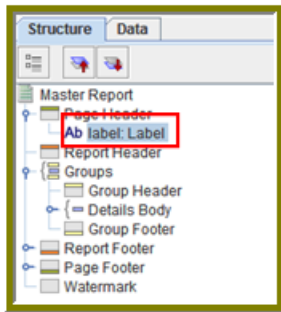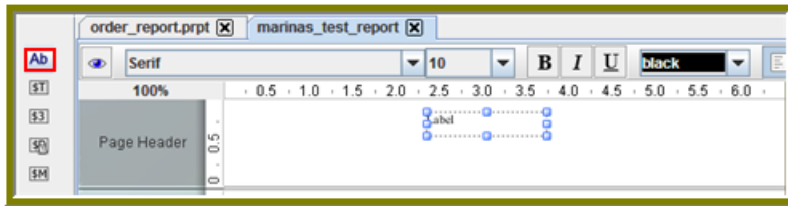
3. Click inside the **Label** item and type **Order Report**

4. Double-click inside the Order Report label to select the text, then in the toolbar, select a larger font size (18 point) and apply boldface.
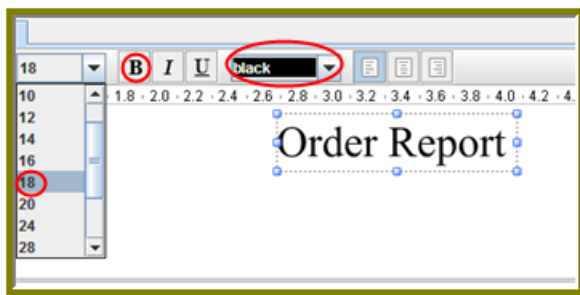
The changes are applied to the text; however, now that the text is bigger you may not see all of it, so use your resizing handles and enlarge the label until you can see all of the text. Alternatively, you can stretch the resizing handles all the way to each edge of the workspace and click the align center icon in the toolbar so that the text is automatically placed in the center of the report page.

5. With the **Order Report** label still selected, click down arrow of the font color icon in the toolbar. Select a color for your label.
   The font color changes. This page header will appear on every page of your report.

6. Now, you must create column headers. On the right side of your workspace, click **Structure** -> **Details Header**.

7. In the lower right section of your workspace, click **Attributes**.

8. Under **common**, change the **Value** of the **hide-on-canvas** option to **False**.
   Notice that the **Details Header** pane appears in your workspace.

9. In the toolbar, click (Select Objects).
   Notice that the icon changes to a cross hair as you move into the workspace.

10. Move your mouse to the far right of the **Details** pane. Now, drag your mouse to the far left over all your column objects to select them. You may have to move the mouse under the headings to make them easier to select.
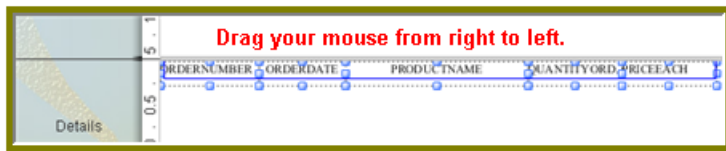
11. Click <**CTRL+C**> to copy your objects and <**CTRL+V**> to paste them into the **Details Header** pane.

   Note: Alternatively, you can choose **Copy** from the right-click menu.

12. Under **Format** in the Report Designer menubar, select **Morph**.
   The column objects are changed to labels.

13. Type the correct heading names for each of your columns: **Order No.**, **Order Date**, **Product Name**, **Quan.**, and **Price Each**.

Your headers will align correctly over your columns.

14. Click (Preview) to display your report.

The report looks good but you may want to make it even easier to read by applying some banding.

15. In the toolbar, go to **Format** -> **Row Banding**.

16. In the **Row Banding** dialog box, choose **Yellow** from the drop-down list next to **Visible Color** and click **OK**.

17. Click (Preview) to display your report.

18. In the menubar, go to **File** -> **Save** to save your report in the **...\report-designer\samples** folder. Type **Orders** in the **File Name** text box.

> **Note:** See *More about Row Banding, Data Formatting, and Alignment* for additional information about refining your report.

More about Row Banding, Data Formatting, and Alignment

Row Banding

By creating a row band element, you can select the *specific* fields in your report that will display a row band. For example, you may want to emphasize specific fields and not others on a line. You can give your row band element any name you choose. In the example below, the row band element is called **row-band-element**.
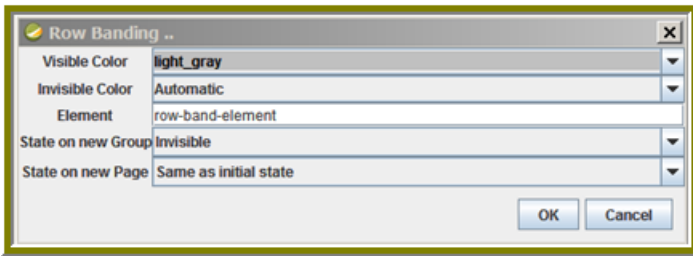
After you create your element, go back to the report and select the columns (fields) whose data will always be displayed with a row band. You must also type row-band-element in the **name** field under **Attributes**. is In the example below, the data associated with each of the columns in the report will display a row band. Notice the banding in the report preview.
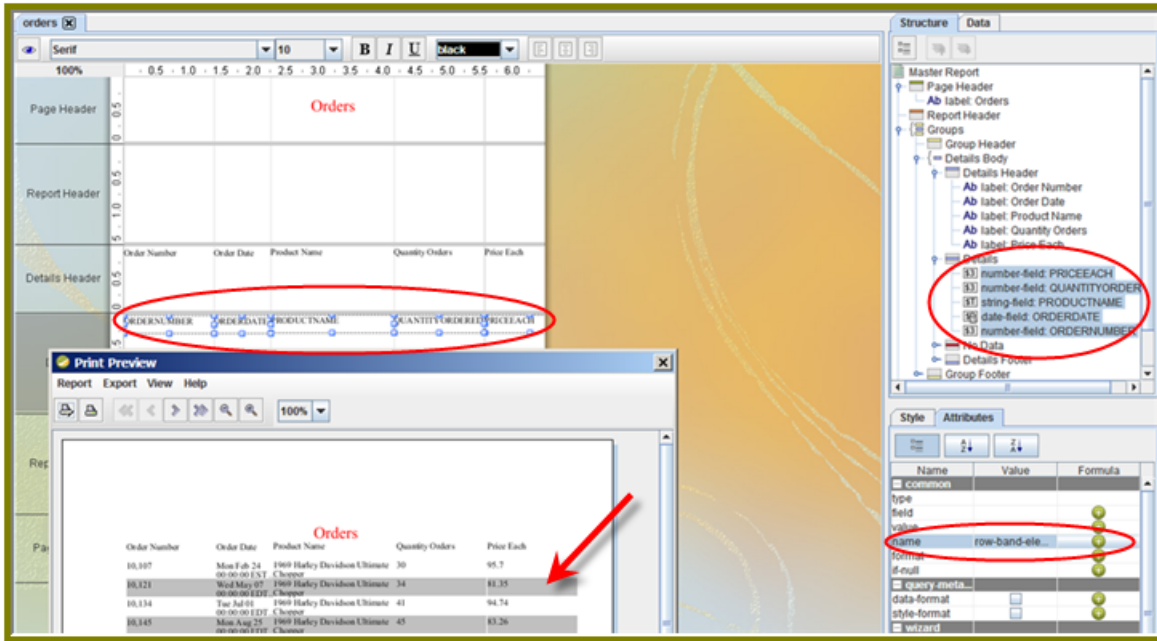
Data Formatting

Report Designer uses default formats for dates and numbers. You can change how dates and numbers display by selecting the object (field) and selecting the appropriate value for the format from the drop-down list next to **format** (under **Attributes**). In the example below, the dates associated with the **Order Date** field will display as **MM-dd-yy**.
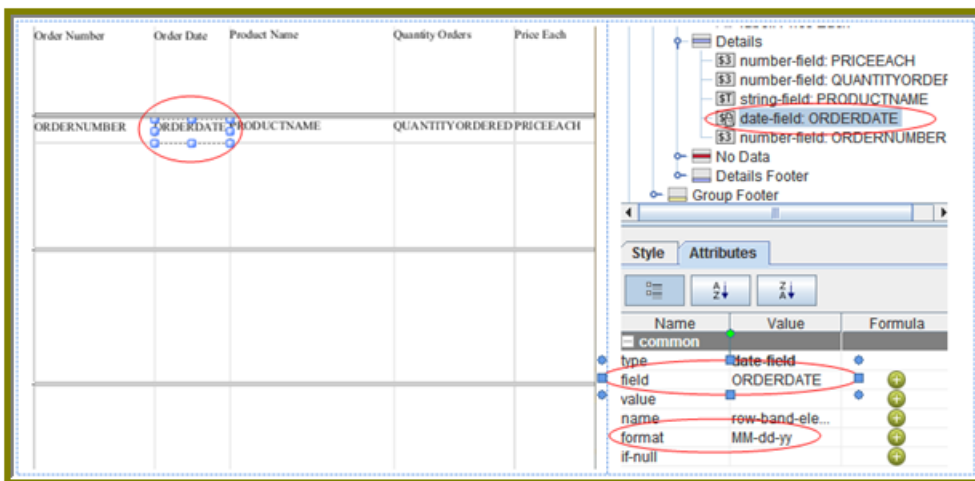
When you preview the report, notice that it displays in a cleaner format:

Note: You can type a value for your own format if you know the correct JavaScript string nomenclature.

Alignment

To align multiple objects press <**SHIFT+ CLICK**> to select each object. Then, choose an alignment option from the **Format** menu. Alternatively, you can click ⬚ (Select Objects) and drag your mouse over the objects you want to select and then choose an alignment option.

In the example below, the selected objects will be aligned in the middle of the band.
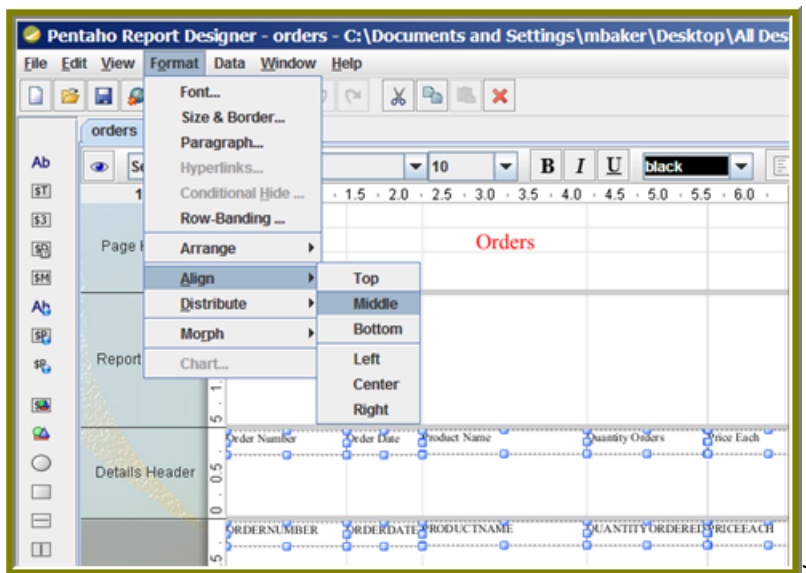
Adding Parameters to Your Report

When you set parameters, users are prompted for a value or values when they run the report. The ability to provide parameters is an important part of creating a report.

1. In the Report Designer, click **File** - > **Open** to select the report you created.

2. In the menubar go to **Data** -> **Add Parameter**. Alternatively, you can click ▥ (Master Report Parameter) under the **Data Tab** in the Report Designer workspace.
The **Add Parameter** dialog box appears.

3. In the Add Parameter dialog box, type **enter_prodline** in the **Name** text field.

4. Type **Select Line** in the **Label** text field.

5. Next to **Display Type**, select **Drop Down** so users can select a product line.

6. Click 🖊 (Edit) to add a query that supplies the values, (motorcycles, cars, ships, and so on), from which users of the report must choose.

> 👉 **Note:** Click on **JDBC (SampleData - Memory)** under **Data Sources** if the Edit icon is disabled.

The JDBC Data Sources dialog box appears.

7. Under **Connections**, select **SampleData (Memory)**.

8. Next to **Available Queries** click ➕ (Add).
A new query placeholder is added to the list (Query 2).

9. In the Query Name text field, type **prodlineList**.

10. Enter your SQL query in the **Query** box.

You can copy and paste the required lines, (shown below) directly into the SQL statement or you can use the alternate steps in the table below.

> 👉 **Important:** Make sure to use curly brackets, (not parentheses), before and after **{**enter_prodline**}** or the report will not display correctly.

```
SELECT DISTINCT "PRODUCTS"."PRODUCTLINE"
                          FROM "PRODUCTS"
```
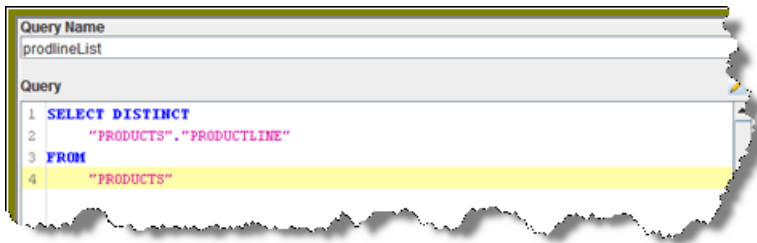

By entering these lines, report users see a prompt when they open the report in the Pentaho User Console that allows them to enter a product line. That way, they can examine orders by product line. If you do not add the lines, the report displays orders for all product lines.

Alternatively, you can use the **Query Designer** to build your query:

Step	Description
1	In the JDBC Data Source dialog box, click 🖊 (the Edit icon on the right).
2	Under **Choose Schema**, select **Public**, and click **OK** to open the **Query Designer.**
3	In the Query Designer, select the **PRODUCTS** table on the left.
4	On the right, click **PRODUCTS** and choose **Deselect All**.
5	Right-click **SELECT** on the left and choose **Distinct**.
6	On the right, select **PRODUCTLINE** and click **Preview**. The product line list appears. Click **Close**.
7	Click **OK** to exit the Query Designer and go to Step 11.

11. Click **OK** to exit **Data Source** dialog box.

12. In the **Add Parameter** dialog box, under Data Sources, double-click **JDBC SampleData (Memory)** and select **prodlineList**.

13. Next to **Value Type**, select **String**.

14. Type a default value, for example, "Motorcycles," in the **Default Value** text box. (Optional)
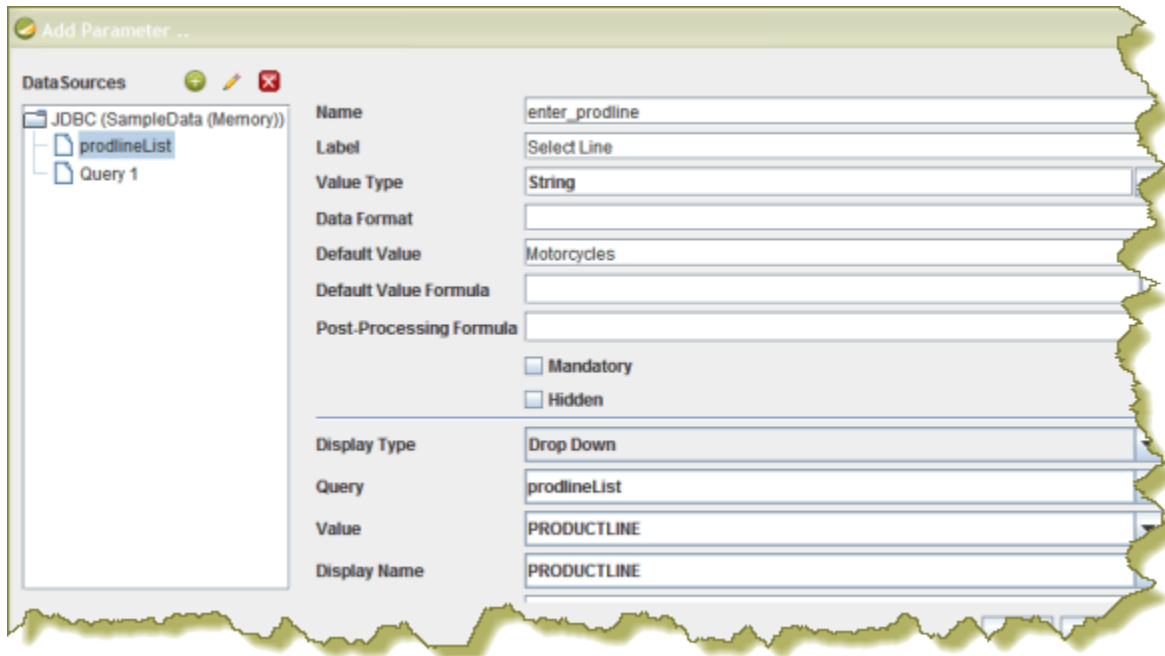
15. Click **OK** to exit the **Add Parameter** dialog box.

16. Now that you've created a product line parameter, you must map it back to your query (Query 1). Under **Data**, double-click **Query 1**.
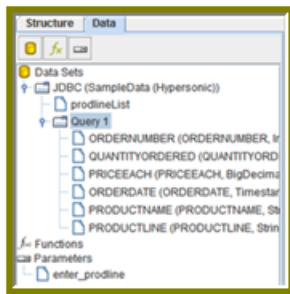
17. Right-click **PRODUCTLINE** and select **add where condition**.
The **condition.edit** dialog box appears.

18. Type `${enter_prodline}` in the edit area and click **OK**.
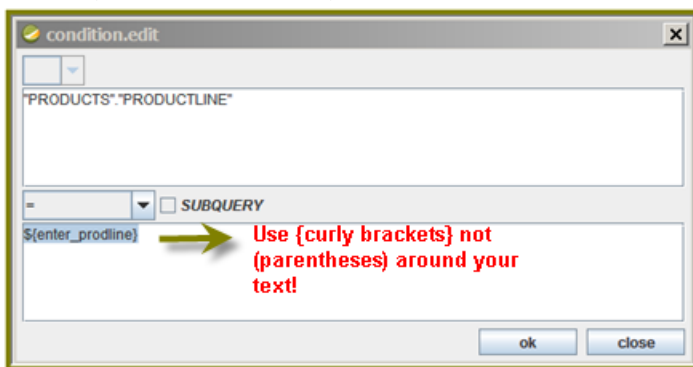
19. Click **OK** to exit the Query Designer.

20. Click **OK** to exit the Data Source dialog box.

21. Click (Preview).
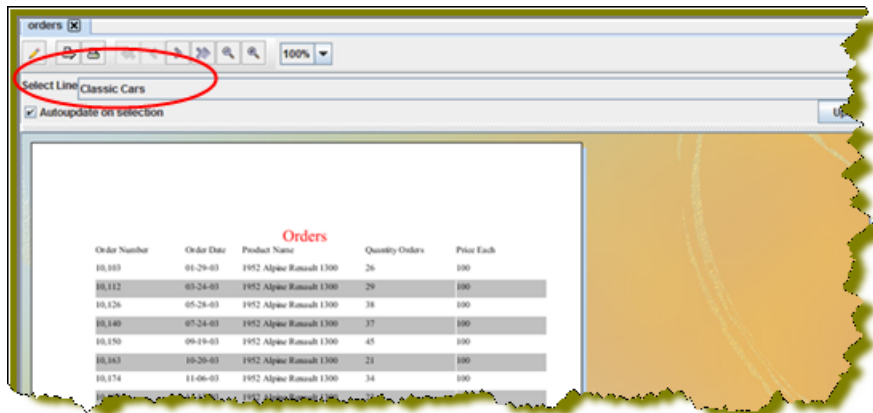
You should see your product line drop-down list.

22. You are now ready to *publish your report*.

Publishing Your Report

You have created and formatted a simple report, added a chart, and now you are ready to share the report with your users.

1. In the Report Designer, click **File** -> **Open** to open the report you just created.

2. Click **File** -> **Publish**. Alternatively, click .

   👉 **Note:** If you are publishing against the open source server you must enter the server's publish password only, which is located in the `publisher_config.xml` file in `pentaho-solutions/system`. You do not need to enter user ID or password when publishing to an open source server as there is no additional security. When publishing to the BI server you must supply a valid user ID, password, and system publish password. See *http://wiki.pentaho.com/display/Reporting/8.+Publishing+a+Report* for more details.

   If you haven't saved the report, a warning message reminds you to save it. The Login dialog box appears, pre-populated with credentials valid for the evaluation.

   👉 **Important:** Make sure that the Server URL is set to `http://localhost:8080/pentaho/` for the open source server.
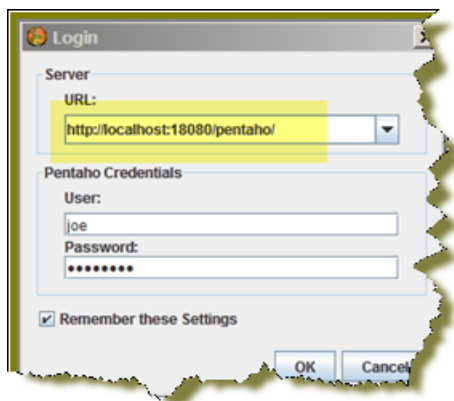
3. Click **OK**.
   The **Publish to Server** dialog box appears.

4. In the Publish to Server dialog box, type in a report name and description into the appropriate fields.

5. Under **Location**, save the report in the `.../steel-wheels/reports` folder.

6. Select **html** as the **Output Type**.

7. In the Publish Password text field, type the publish password, **password**, and click **OK**.
   A success message appears.

8. Click **Yes** to go directly to the Pentaho User Console to view the report you just published.

   If you want to access the report later, log into the BI Server by going to **http:// localhost:8080** in your Web browser, then navigate to the **Reporting Examples** directory in the **Solution Browser**. You should see your published report in the list. If not, click **Tools** -> **Refresh Repository**.

9. Log in as **Joe**. Joe's password is **password**.

10. Your report displays in the Pentaho User Console.
You now have a report that users can view at any time.

11. In the Pentaho User Console select your product line parameter from the drop-down list. Accept the default under **Output Type**.

Using the Data Sources Feature in the Pentaho User Console

The Data Sources feature makes it easy for you to access your own data when you are evaluating the capabilities of the Pentaho User Console.

If you are running a test environment, you can quickly add, edit, and delete a relational or CSV (flat file) data source in the Pentaho User Console. You can restrict access to a data source by individual user or user role; by default, authenticated users are limited to view-only permissions. (See *Assigning Data Source Permissions for the Pentaho User Console* for more information.) This means that they are limited to selecting a data source from a list of available options.

👉 **Note:** Non-administrative users of the Pentaho User Console do not see icons or buttons associated with adding, editing, or deleting a data source

When selecting a relational data source, users can query against the data source within the scope of data defined in query you set up. If users choose a CSV-based data source, they can query the flat file data to create a chart or data table.

Creating a Relational Data Source in the Pentaho User Console

The **New Data Source** dialog box, which allows you to set up access to a relational data source in the Pentaho User Console, is available in the following instances:

- When you click the **New Data Source** button on the Pentaho User Console Quick Launch page
- When you click **Add** in the **Select Data Source** step in Ad Hoc Reports

In addition to allowing you to connect to your database (using JDBC), you must also enter a SQL statement that defines the "scope, "(similar to a database view), of your data source.

👉 **Note:** In the context of the Pentaho User console, a relational "data source" is a combination of a connection plus a query. Pentaho recommends that you use the Pentaho Metadata Editor to create relational data sources for a production environment.

You can customize how columns are presented to users who are building queries against the new data source; for example, you can define "friendly" column names and select options that define how data is aggregated (sum, min., max., etc.), and more.

Users who do not have administrative permissions do not see New Data Source dialog box; however, they do see a list of available data sources that have been previously created by an administrator; they can query against the data sources within the scope of data defined in queries created by an administrator.

Once you create the data source, it is available to users of the Dashboard Designer, Chart Designer, and Ad Hoc Reports.

You must be logged on to the Pentaho User Console as an administrator to follow the instructions below:

1. In the **New Data Source** dialog box, click **Database**.
2. Type a name for the new data source in the **Name** field.

   👉 **Note:** The name of the data source is important because it becomes the name of the data source displayed in Ad Hoc Reports and Chart Designer.

3. If not already configured, click (+) to add a connection. The **Database Connection** dialog box appears.
4. In the **Connection Name** text box, enter an easy-to-remember name that identifies the data you are accessing.
5. Select a **Connection Type** from the list.

   Text fields for required settings associated with your connection type appear under **Settings** on the right.
6. Enter the appropriate connection information for your database type.

   Consult the *A List of JDBC Drivers* for more information.
7. In the Connection dialog box, click **Test** to test your connection and click **OK**.
   A success message appears if your settings are correct.
8. Under **SQL Query**, enter your SQL Query. Click **Preview** to ensure that your query returns data.
9. Click **Close** to exit the Preview window.
10. Click **Finish**.
    In the background, after you click **Finish**, a default metadata model is generated based on the query for use in Ad Hoc Reporting and Chart Designer.

Creating a CSV Data Source in the Pentaho User Console

The **New Data Source** dialog box allows you to set up a CSV (flat file) as a data source. Once you define a CSV file as a data source, end-users can create reports, charts, and data tables based on data in the CSV file. The New Data Source dialog box is available in the following instances:

- When you click **Add** in the **Select Data Source** step in Ad Hoc Reports
- When you click **Add** in the **Select Data Source** step in Ad Hoc Reports

Follow the instructions below to create a CSV-based data source:

1. In the **Data Source Wizard**, type a name that identifies your new data source in the **Data Source Name** field.
2. Under **Source Type** select **CSV** from the list to define a CSV file as your data source and click **Next**.
3. Click **Import** to open the **Import File** dialog box.
4. Click **Browse** to search for your CSV file; click **Import** when you locate the file.
   The Import File dialog box closes and the CSV file is uploaded to the BI Server. You can click **Cancel**, if necessary, and select different file.

   👉 **Note:** Notice that a preview of the contents of your file appears in the lower portion of the wizard page. If items are not aligning correctly, it is because you must specify your delimiter and enclosure types. Go on to the next step.

5. Specify the delimiter and enclosure types. Enable **First row is header** if the first row of your CSV file is a heading for columns in the file.

   👉 **Important:** You must make selections under **Delimiter** and **Enclosure** for the Metadata table to build correctly. In the example line, `1,4,5,02/12/77,"String Value"` the commas are delimiters and the quotation marks around "String Value" is the enclosure.

   The **File Preview** window displays the first few lines of your CSV file based on the selections you made for specifying the delimiter, enclosure, and header. Once the columns align correctly in the preview you will know that the delimiter and enclosure have been set correctly.
6. Click **Next** to display more details about the columns in your source file and to edit staging-related settings.

> **Note:** Notice that all columns are enabled in this step. Click **Deselect All** to disable all columns. You can then select the columns that require formatting individually. Click **Select All** to enable all columns again, if necessary.

7. If applicable, change the **Name** and **Type** values. Use the drop-down list to display format options for dates and numeric values. Formats are auto-detected based on the contents of your data source; however, if you see a format option that is not correct, enter it manually in the appropriate **Source Format** text box.

> **Note:** Refer to *Sample Date Formats* for additional help formatting dates. Refer to *Decimal Formats* for additional help formatting numbers.

Length indicates the maximum number of characters allowed in a field. **Precision** is the number of digits after a decimal point. Drop-down lists are not enabled for certain data types such as the String data type.

> **Important:** Boolean values are rendered as "true" or "false."

8. Click **Show File Contents** to see a sample of the data in your source file. Click **OK** to return to staging and make changes if necessary.
9. Click **Finish** when you have completed your configuration and formatting tasks.
   A success message appears indicating the number of rows being loaded into the database and that a default metadata model was created. Enable **Keep default model** to use the default model or click **Customize model now** to customize the model.

The data in your CSV file is extracted and staged temporarily in a database table. Your new data source is now available to users. To refine the data and make it more accessible to users of reports, dashboards, and data tables, you must use the Data Source Modeler. See *Editing a Data Source in the Pentaho User Console*.

Increasing the CSV File Upload Limit

Follow the instructions below if you need to increase the size of the upload limit associated with CSV files.

1. Go to `...\biserver-ce\pentaho-solutions\system` and open the **pentaho.xml** file
2. Edit the XML as needed; (sizes are measured in bytes):

```
<file-upload-defaults>
       <relative-path>/system/metadata/csvfiles/</relative-path>

       <!-- max-file-limit is the maximum file size, in bytes, to allow to be uploaded
 to the server -->
       <max-file-limit>10000000</max-file-limit>

       <!-- max-folder-limit is the maximum combined size of all files in the upload
 folder, in bytes. -->
       <max-folder-limit>500000000</max-folder-limit>

</file-upload-defaults>
```

3. Save your changes to the file.
4. In the **Pentaho User Console**, go to **Tools** -> **Refresh System Settings** to ensure that the change is implemented.

Using the Modeler to Edit a Data Source

You can edit a data source in Ad Hoc Reports, Pentaho Analyzer, and in the Dashboard. If you are editing a data source while using Ad Hoc Reports, select the data source you want to edit from the list of available data sources and

click **Edit**. If you are creating a chart or data table, select the data source you want to edit from the list and click 🖊

(Edit). If you have created an Analyzer report, click 🖊 (Edit) in the upper right corner next to **Available Fields**. You can also edit a data source by navigating to **File** -> **Manage** -> **Data Sources** in the Pentaho User Console menu bar and opening the **Select Data Source** dialog box.
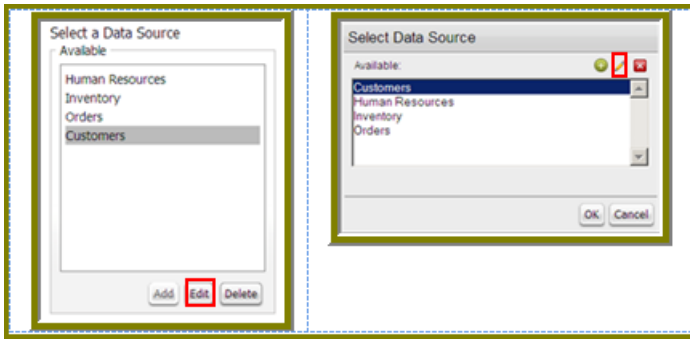
Note: If you want to edit the connection information, see *Editing a Data Source Connection in the Pentaho User Console*.

When you edit a data source you are using the Data Source Modeler. The modeler allows you to design reporting and OLAP metadata models on the fly. The modeler also allows you to refine and organize your data so that it is more accessible to report users. For example, you can change the field display names, set an aggregation type, and change formatting,

Follow the instructions below to edit a data source:

1. In the **Data Source Modeler**, examine your data.
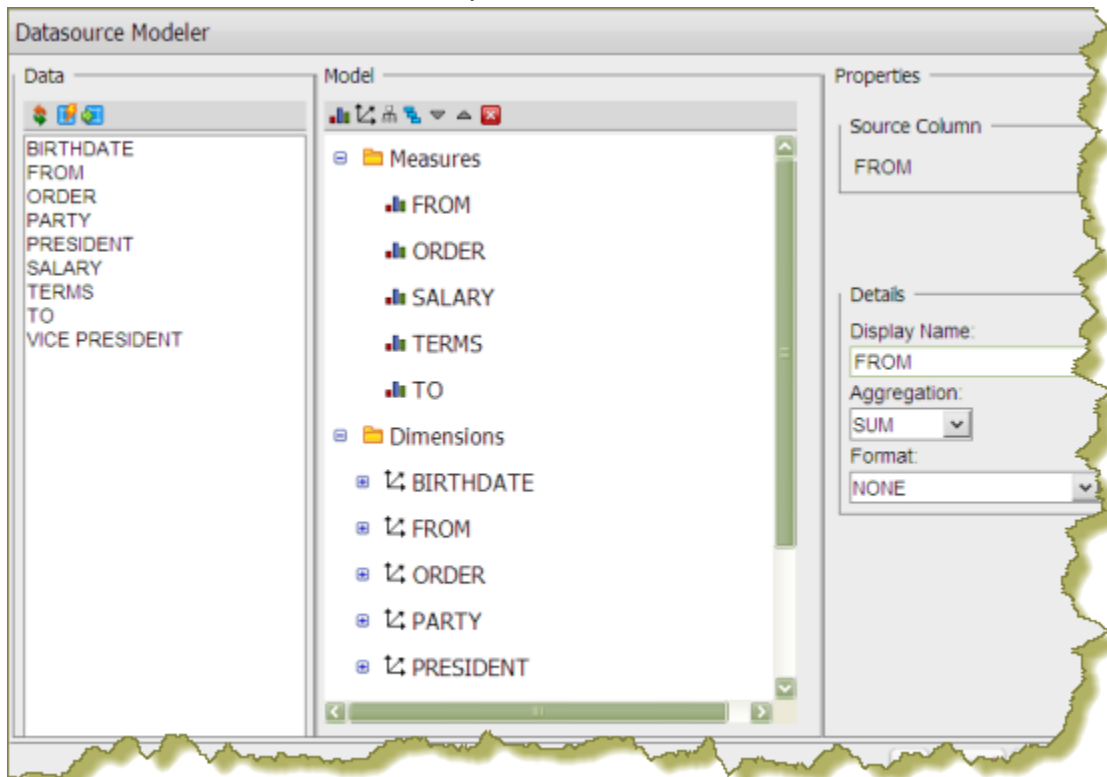
Measures are numeric values such as sales price, cost, and profit. **Dimensions** are the context that allow users to understand the significance of the measures. For a sales database, the dimensions could include Product, Store, Salesperson and so on. In other words, users can examine the sales figures by store, by product, and by sales.

Dimensions can have one or more hierarchies (⛛); (hierarchies are not shown in the example). Hierarchies are levels that organize data into a logical structure. Click (+) next to fields under Measures and Hierarchies to display their contents. Moving up and down between the levels of a hierarchy is called *drilling*.

You can also edit the source data to define the properties of your CSV file or re-enter a SQL query. In the example above, the FROM, TO, TERMS, and ORDER fields must be redefined as string types, not numeric. The only

applicable measure in this example, is SALARY. Unnecessary fields can be removed clicking (Delete) in the modeler toolbar.

You can also move a field under Measures or Dimensions by clicking and dragging the field into its appropriate location. A green check mark indicates that you can place your field without error.

 Note: Click (Clear Model, in the upper right corner) to clear the existing model and start over.
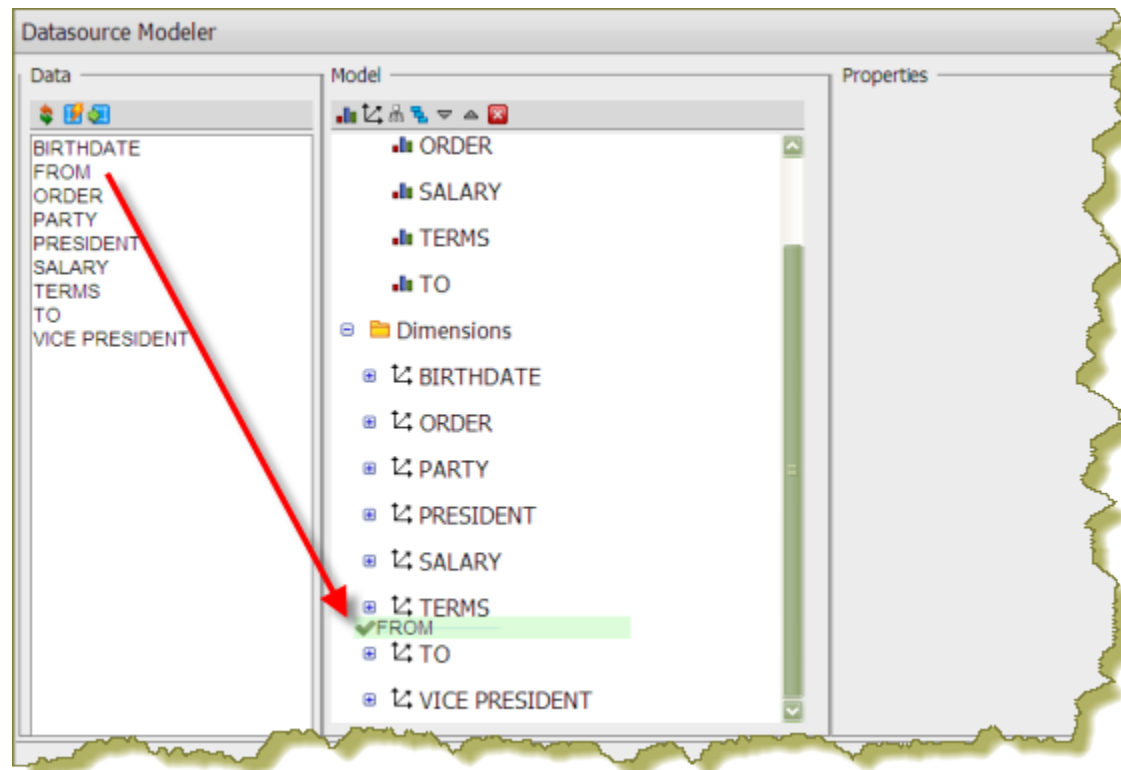
2. If you have clicked **Edit Source** in the modeler to make changes to a SQL query or CSV flat file (delimiter, enclosure, or header), click **Finish** to update the metadata table.

   If you disable columns that were previously enabled, a prompt appears in the modeler indicating that the data has been deleted. You must adjust the data in the modeler before proceeding.

3. Click **OK** to update the data source.

The updated data source is now available to users.

Editing a Data Source Connection in the Pentaho User Console

Follow the instructions below to edit a data source connection in the Pentaho User Console:

1. In the **New Data Source** dialog box, under **Connection**, select the connection you want to edit.
2. Click (Edit).
   The Database Connection dialog box appears.
3. Update the connection information as needed. You can click **Test** to make sure your connection is working correctly.
4. Click **OK** to exit the Database Connection dialog box.

Deleting a Data Source Connection in the Pentaho User Console

Follow the instructions below to delete a data source connection associated with the Pentaho User Console.

 Important: If you delete a data source connection, all reports, charts, and other content users have created using the connection will no longer render.

1. Under **Connection** in the New Data Source dialog box, select the connection you want to delete.
2. Click (Delete).
   A warning dialog box appears.

3. Click **OK** to delete the connection.

Deleting a Data Source in the Pentaho User Console

If you have administrative permissions, you can delete data sources associated with ad hoc reports and dashboards in the Pentaho User Console.

👉 **Important:** When you delete a data source all reports or charts that end-users have created based on that data source will no longer render. Deleting a data source also deletes the model associated with that data source.

The delete Data Source option is available in the following instances:

- When you select a data source and click (Delete) when in the **Select Data Source** dialog box available from **File** -> **Manage** -> **Data Sources**
- When you select a data source and click **Delete** in the **Select Data Source** step in Ad Hoc Reports
- When you select a data source and click (Delete) when creating a chart
- When you select a data source and click (Delete) when creating a data table

Assigning Data Source View Permissions

By default, authenticated users have view-only permissions to data sources in the Pentaho User Console. Users with administrative permissions can create, delete, and view data sources. If you want to fine tune permissions associated with data sources, you must edit the appropriate **settings.xml** file. Follow the instructions below to edit the file.

1. Go to `...\biserver-ee\pentaho-solutions\system\data-access-plugin` and open **settings.xml**.
2. Edit the settings.xml file as needed. You can assign permissions by individual user or by user role. If you are using LDAP, you can define the correct ACLs value for view permissions; the default value is "31."
3. Save your changes to the **settings.xml** file.
4. Refresh the Admin Console.

Creating an Ad Hoc Report

Ad Hoc Reporting allows you and your users to create a basic data-driven report quickly and efficiently. With Ad Hoc Reporting, users do not need to know the structure of the database, nor do they need to know any SQL. A metadata model created with the Pentaho Metadata Editor, (a graphical user interface for creating user friendly metadata models), acts as a buffer between users and the complexities of relational data sources. The Pentaho BI Suite comes with three pre-built business models (datasets).

👉 **Note:** What is metadata? Quite simply, metadata is "data about data." For example, a library catalog is considered "metadata" because it contains information about books and other publications. A metadata model, is a collection of related categories of data.

You can quickly access data in your own relational database or CSV flat file for evaluation or testing purposes. To find out more, see *Using the Data Sources Feature in the Pentaho User Console* .
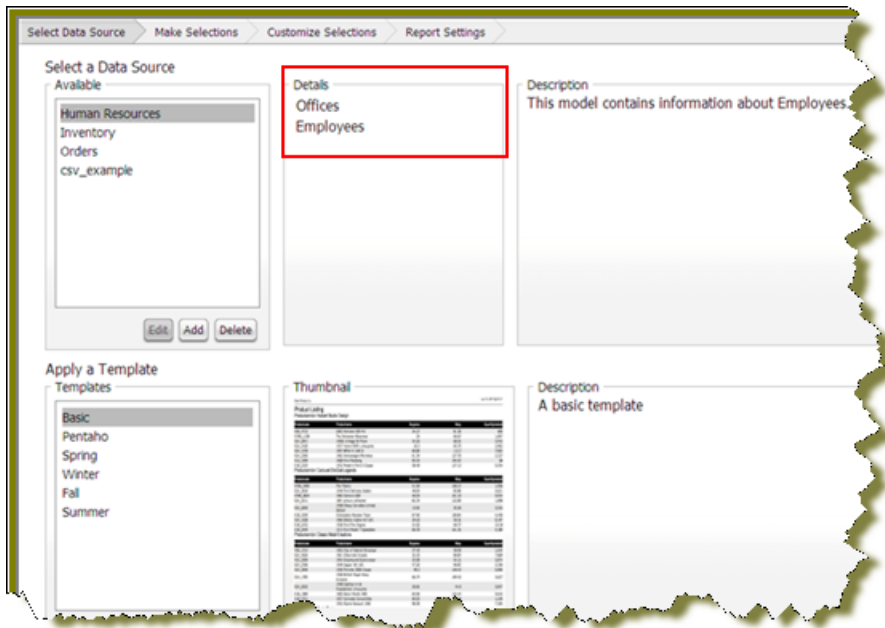
Ad Hoc Reporting Interface provides...

- Easy-to-create connections to your relational or CSV (flat file) data sources for evaluation and testing purposes
- Interactive "drag and drop" Web interface for business user self-service report creation
- Wizard-driven authoring supporting report templates, metadata-based query creation, sorting, and filtering
- Interoperability with Pentaho Report Designer allowing ad hoc reports to be "promoted" for fine-tuning by IT professionals
- Single reporting engine for ad hoc and pixel-perfect reports for lowest possible total cost of ownership

To create an ad hoc report, you must logged onto the Pentaho User Console.

1. In the Pentaho User Console click **Create New Report**.

   The ad hoc query wizard starts.

2. In the first step of the wizard, select **Orders** from the list under **Select a Data Source**.

   As you become more familiar with Ad Hoc Reports features, you can add your own relational or flat file (CSV) data sources to ad hoc reports. See *Using the Data Sources Feature in the Pentaho User Console* for more information.

   The tables associated with the **Orders** data source are listed in the **Details** pane.

3. In the **Apply a Template** field, select a predefined report template.

   A thumbnail preview of the template appears in the **Template Details** field. A template specifies a variety of properties in the report that affect its appearance, like font size and background colors for various report elements.
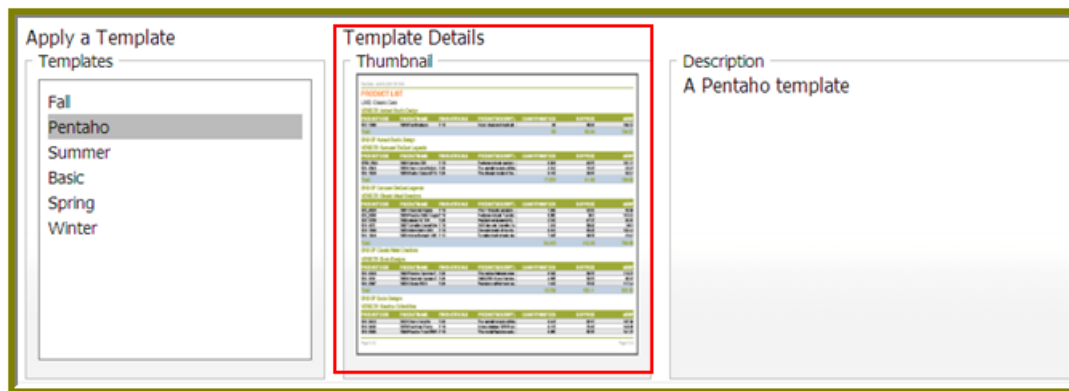
4. Click **Next**.

5. In the **Available Items** list, click the **Territory** business column and drag it to the upper right into the **Level 1 box**.

   This determines how the data is grouped.

6. Drag and drop the **Amount** and **Buy Price** into the **Details** box on the right.

   This determines which fields to display for the given groups.

7. Click **Go** to preview how these new items have affected the report, then close the preview tab when you're done.

8. Click **Next**.

9. Click the **Territory** item in the **Groups** list.

    A list of general options appear on the right.

10. Click **Center**.

    This centers the territory name above each table, making it easier to read.

11. Click **Amount**, then click **Add** in the **Sort Detail Columns** area on the right.

    This sorts the sales amounts from lowest to highest.

12. Click **Go** to test the new change, or **Next** to continue to the next part of the wizard.

13. To set the header, footer, description, paper type, and page orientation, change the on-screen values for these elements accordingly.

    PDF is the only output type that has a concept of a page, so the **Page** portion of the **Header** and **Footer** sections only applies to PDFs.

14. Click the blue **Save** button in the top toolbar to save your report. In the ensuing file dialog, navigate to the location you want to save the report to, and type in a filename for the report.

    You can continue to modify your report after it's been saved; just click **Save** to update the report file after you've made changes.

You now have a report that shows how much revenue is coming from each sales territory, and the itemized price of each purchased product. As you can see, ad hoc reporting provides business users with a quick and simple solution for building basic reports. For report designers who need access to more advanced features like pixel perfect layout, conditional formatting, or parameterization, Pentaho provides the full-client Report Designer.

Analysis View Tutorial

Analysis views are similar to reports, except they're designed to be totally interactive and dynamic, whereas reports tend to be static or minimally interactive after they're created. Analysis views allow you to dynamically explore your data and drill down into it to discover previously hidden details. In this example, you'll try to find out which product line is responsible for the highest number of cancelled orders.

1. In the **File** menu, select the **New** sub-menu, then click **New Analysis View...**.

    This is one of several ways to create a new analysis view; all methods lead to the same screen in the Pentaho User Console.

2. Select the **SteelWheels** schema and **SteelWheels Sales** cube from the drop-down lists.

3. Click **OK** to continue.

   An analysis view will open in a new tab.

4. Open the **OLAP Navigator** by clicking the cube icon in the top button bar – it's the third from the left.

   A new table will appear above the analysis view fields. The default basis for comparison is **Measures**, though that is not very useful for finding returned products.

5. In the **Rows** section, click the two-tone square next to **Product**. This will move it up to the **Columns** section.

6. In the **Columns** section, click the funnel icon next to **Measures**. This will move it down to the **Filters** section.

7. Click **OK** to modify the analysis view.

8. Click the **+** next to **All Products** to drill down into it.

9. Click the **+** next to **All Status Types** to drill down into it.

   Clearly the **ships** category has the most cancelled orders.

10. Drill down further into ships by opening the OLAP Navigator again.

11. Click **Product** in the **Columns** section.

12. Un-check all products except **Ships**.

13. Click **OK**.

14. Click **Order Status**.

15. Un-check all statuses except **Cancelled**.

16. Click **OK**, then **OK** again to close the OLAP Navigator and refine the analysis view.

17. Click the **+** next to **Ships** to show its constituent product lines.

18. Notice that the **Carousel DieCast Legends** series has the highest number of cancelled orders.

You now know which product line has the most cancelled orders.

Building a simple input-output transformation

You must have a database connection to complete this exercise.

This exercise walks you through the process of building a transformation that uses input from a database table (a list of customers), applies an SQL statement, and outputs the data stream to a plain text file. This exercise is useful for testing and tuning transformations to see how the fields are passed in the data stream; it also shows you how to examine the results of each step of a transformation, as well as the final output.

1. In the left pane, click **Design**, then click **Input** to expand the node and view the input step options.

2. Click **Table Input** and drag the cursor to the right pane.

   A **Table Input** step is added to the transformation. This step reads information from the database using the database connection and SQL.

3. Double-click **Table Input** to display a dialog box that allows you to view and edit the SQL statement.

   As a general rule, it is a good idea to maximize this window. This is a relatively simple example, but for longer statements, maximizing the window ensures that you see all parts of the statement. You must edit the following parameters to specify the table to use as input, the values get from the table, and (optionally) the conditions:

(SELECT) Values:	Specific value(s) or an asterisk * to indicate all values.
(FROM) Table Name:	Name of the database table, for example **customers**.
(WHERE) Conditions:	Specific condition(s), or may be deleted as in this example.

4. Click **Preview** to view the data stream flowing from the **Table Input** step. A dialog box appears that allows you to specify the number of rows you want to preview.

   Preview shows you what is in the data stream coming into the step. If there are errors, a log file appears that includes a record of the errors to assist in troubleshooting.

5. In the left pane, click the **Output** folder to expand it and view the output step options. Click **Text File Output** and drag the cursor to the right pane.

   A **Text File Output** step is added to the transformation. This step sends the fields in the incoming data stream to a text file.

6. Click **Table Input**, then press <**Shift**> and drag the cursor to **Text File Output**.

   This adds a hop (data connection) between the two steps and displays as an arrow connecting the steps. The hop represents the data stream flowing from the **Table Input** step to the **Text File Output** step.

7. Double-click the **Text File Output** step and type a path/file name in the **File name** box.

   The **Content** tab allows you to specify optional settings such as the field separator and the format (DOS or UNIX).

8. Click the Fields tab, then click **Get Fields**.

   Note: The file extension (in this case, .txt) is filled in for you.

   The Fields tab displays all the fields in the data stream that is flowing through the hop from **Table Input** step to the **Text File Output** step, and this data will in turn flow into the text file you defined. You can also type fields into this tab to add them to the data stream.

9. Right-click **Text File Output** to display the context menu, then click **Show Input Fields**.

   This dialog box displays all the fields in the data input stream and their transformations. In this simple example, there are no intermediate transformation steps, for example filters, but in a transformation with intermediate steps, the details display here.

10. On the context menu, click **Show Output Fields**.

   This dialog box displays all the fields in the data output stream and their transformations. In this simple example, there are no intermediate transformation steps, for example filters, but in a transformation with intermediate steps, the details display here. Together, the **Show Input Fields** and **Show Output Fields** dialog boxes display all the detail information on every field that is entering and leaving a step. This is important when you are designing, tuning, and troubleshooting a transformation.

11. Click **File** > **Save**, then click ▶ (Run) .

   Note: You must save the transformation before you can run it.

12. Click **Launch**.

   The transformation runs and the results display in a pane in the lower portion of the window.

13. As an alternate to the Text File Output, you can use a Table Output, which inserts fields into a database table.

   You can double-click **Table Output** to display the Table Output dialog box, where you can specify settings for the output.