



## Introducing the Pentaho BI Suite Community Edition



This document is copyright © 2009 Pentaho Corporation. No part may be reprinted without written permission from Pentaho Corporation. All trademarks are the property of their respective owners.

## About This Document

If you have questions that are not covered in this guide, or if you find errors in the instructions or language, please contact the Pentaho Technical Publications team at [documentation@pentaho.com](mailto:documentation@pentaho.com). The Publications team cannot help you resolve technical issues with products.

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](mailto: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](mailto:communityconnection@pentaho.com)  
Sales Inquiries: [sales@pentaho.com](mailto:sales@pentaho.com)  
Documentation Suggestions: [documentation@pentaho.com](mailto:documentation@pentaho.com)

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

# Contents

<b>Introduction</b> .....	2
Community Edition or Enterprise Edition? .....	2
Community Edition Support Options .....	2
The Pentaho Client Tools .....	2
<b>Installation</b> .....	4
Hardware Requirements .....	4
Software Requirements .....	4
Downloading and Installing the BI Suite .....	5
Starting the BI Platform .....	6
Configuring the BI Platform With the Administration Console .....	6
<b>Getting Started</b> .....	7
How to Log Into the Pentaho User Console .....	7
Navigating the Pentaho User Console .....	7
<b>Tutorials</b> .....	9
Report Designer Tutorials .....	9
Starting Report Designer .....	9
Exploring the Report Designer Interface .....	9
Creating a Report Using Report Designer .....	11
Designing Your Report .....	15
Refining Your Report .....	16
Adding Parameters to Your Report .....	18
Publishing Your Report .....	19
Using the Data Sources Feature in the Pentaho User Console .....	20
Creating a Relational Data Source in the Pentaho User Console .....	21
Creating a CSV Data Source in the Pentaho User Console .....	22
Assigning Data Source View Permissions .....	23
Creating an Ad Hoc Report .....	23
Analysis View Tutorial .....	26
Building a simple input/output transformation .....	27

# 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 software 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 system specifications:

RAM	At least 2GB
Hard drive space	At least 1GB
Processor	Dual-core AMD64 or EM64T

It's possible to use a less capable system, 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 typically 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.

## 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 with Service Pack 2, 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 10.4 are all officially supported.

No matter which operating system you use, you must have the Sun Java Runtime Environment (JRE) version 1.5 (sometimes referenced as version 5.0) installed. 1.4.2 will not work, and while 1.6 (6.0) will probably work in most cases, Pentaho does not yet officially support it.

**Note:** The GNU Compiler for Java, or GCJ for short, interferes with the way many native Java programs work on Linux, including some of the components of the Pentaho BI Suite. If you are using a Linux distribution that installs GCJ by default (which includes all of the most popular distros), then before you begin installation you must remove, disable, or circumvent GCJ. If you cannot remove it, you can simply ensure that your JAVA\_HOME variable is properly set, and add the Java Runtime Environment's /bin/ directory to the beginning of your PATH variable in ~/.bashrc or /etc/environment, then relog before continuing.

Workstations will need to have reasonably modern Web browsers to access Pentaho's Web interface. Internet Explorer 6 or higher; Firefox 3.0 or higher (or the Mozilla or Netscape equivalent); and Safari 2.0.3 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 Windows Vista, 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 may 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 Platform 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 Platform 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** script in the `/pentaho/server/biserver-ce/` directory.

### Starting the Pentaho Administration Console

To start the Pentaho Administration Console, run the **start** 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.



## 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**.

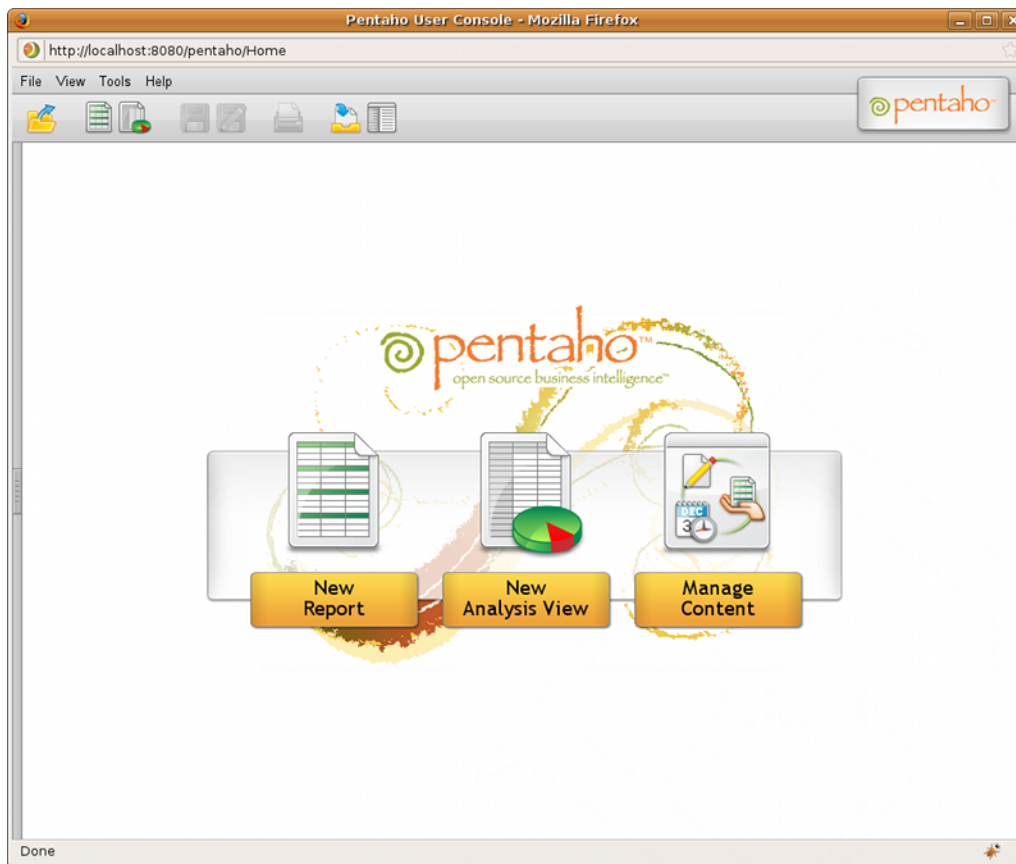
The login dialog will appear.

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 three icons in the center of the screen to create a new ad hoc report, start a new analysis view, or edit existing solutions.

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

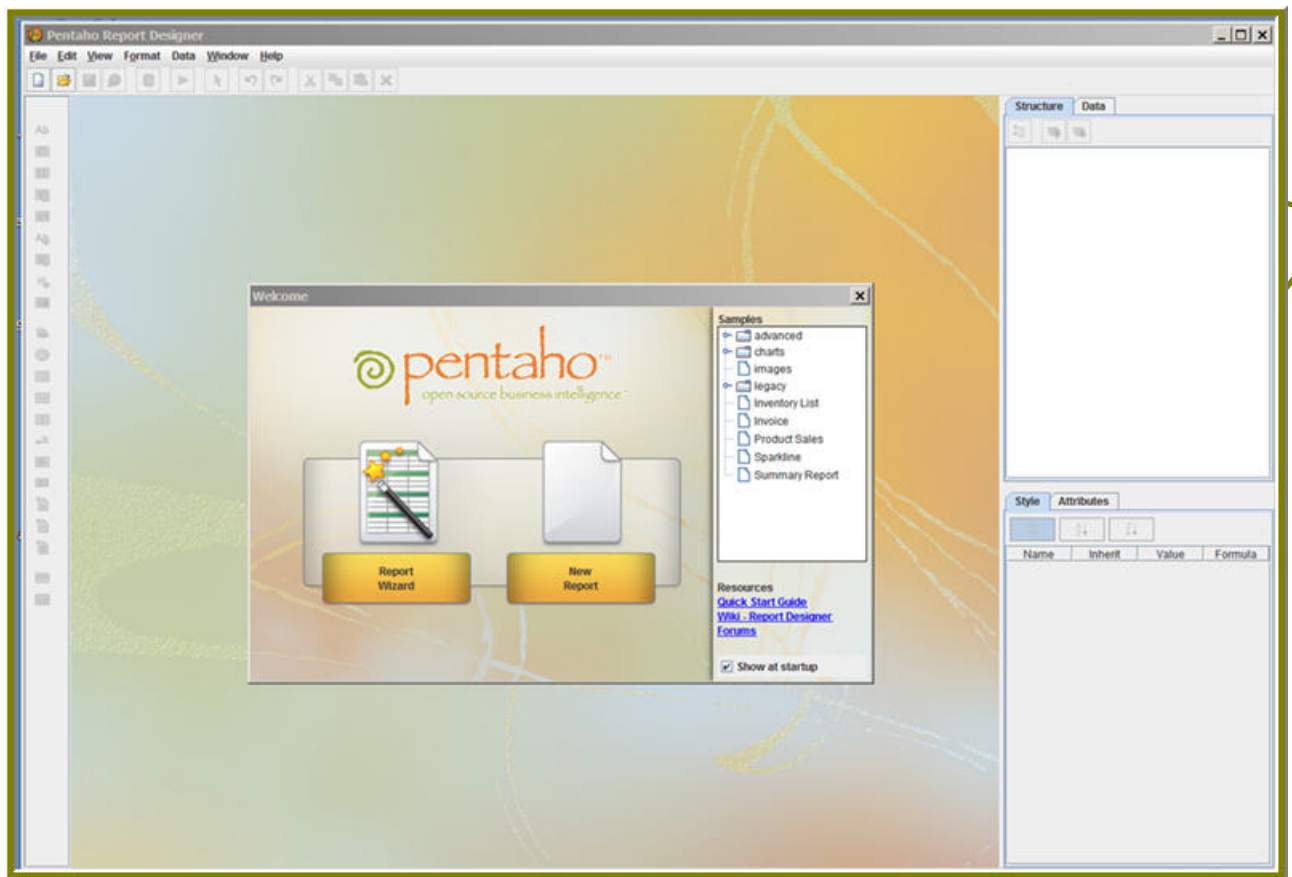
Follow the below process to start Report Designer.

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 **ReportDesigner.exe** on Windows, or `startdesigner_linux.sh` on Linux.
2. Before the program starts, it will run 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

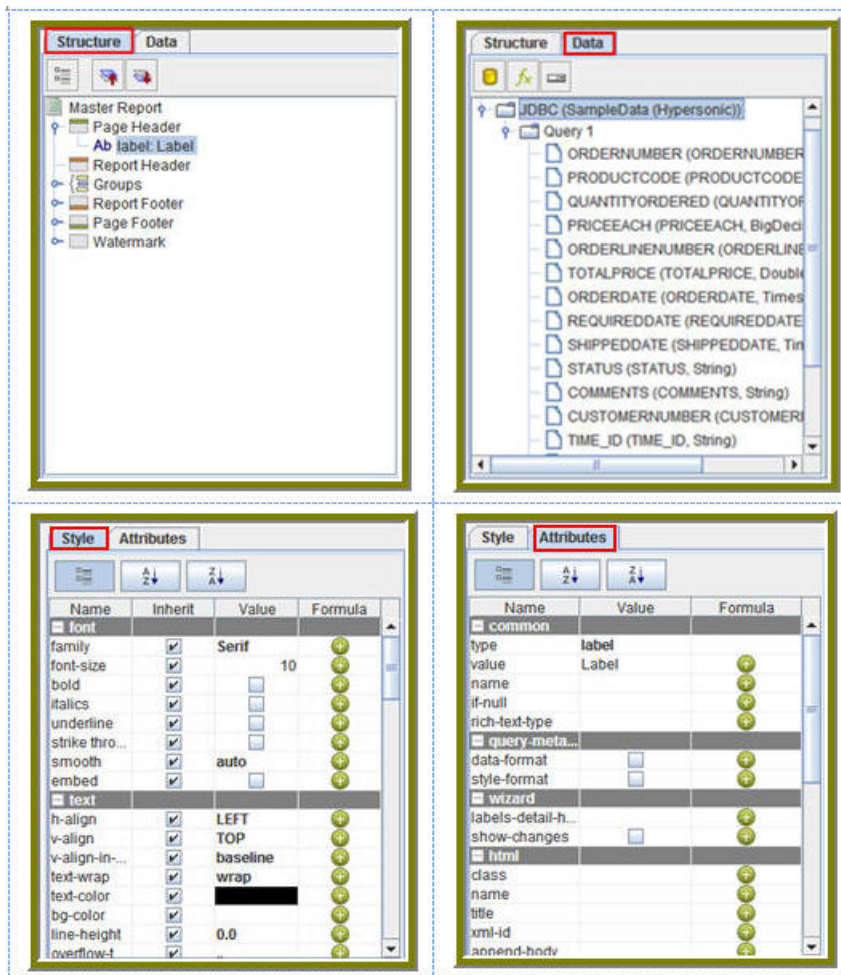
Report Designer's interface is similar to that of other graphic design and layout tools. A typical menu and button bar are at the top of the screen; a tool palette for adding design elements is on the left; and on the right are two panes that contain data and structure elements, and show the properties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

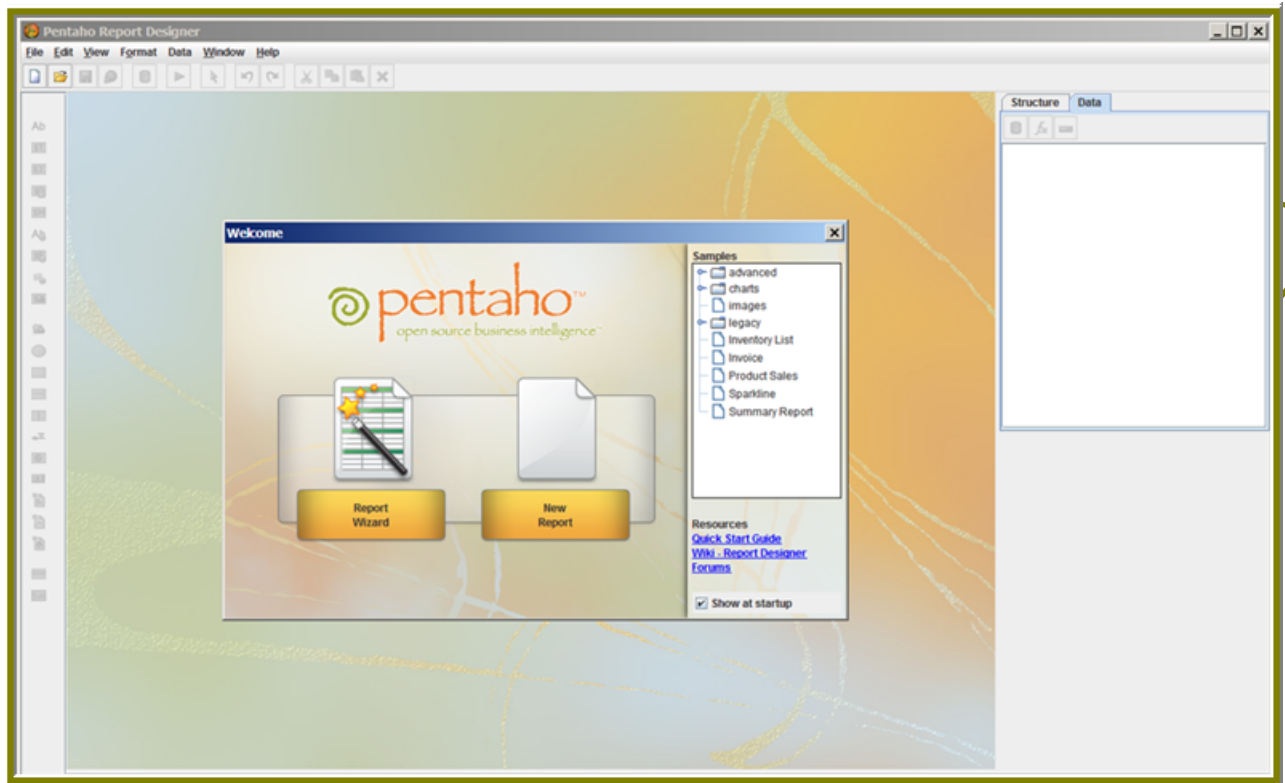


## Creating a Report Using Report Designer

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**.

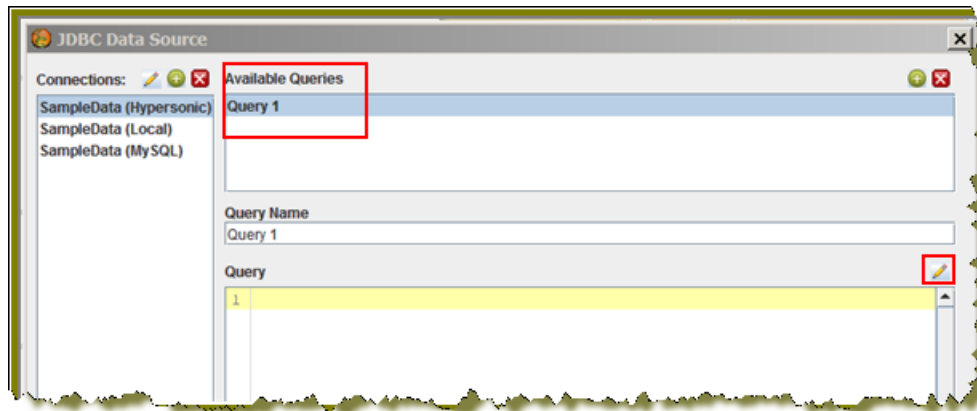
The **JDBC Data Source** dialog box appears.

5. Under **Connections**, select **SampleData (Hypersonic)**.

6. Next to **Available Queries** click  (Add).

**Query 1** appears under **Available Queries**. Notice that the edit icon  is enabled.

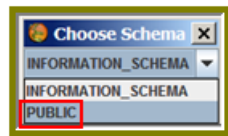




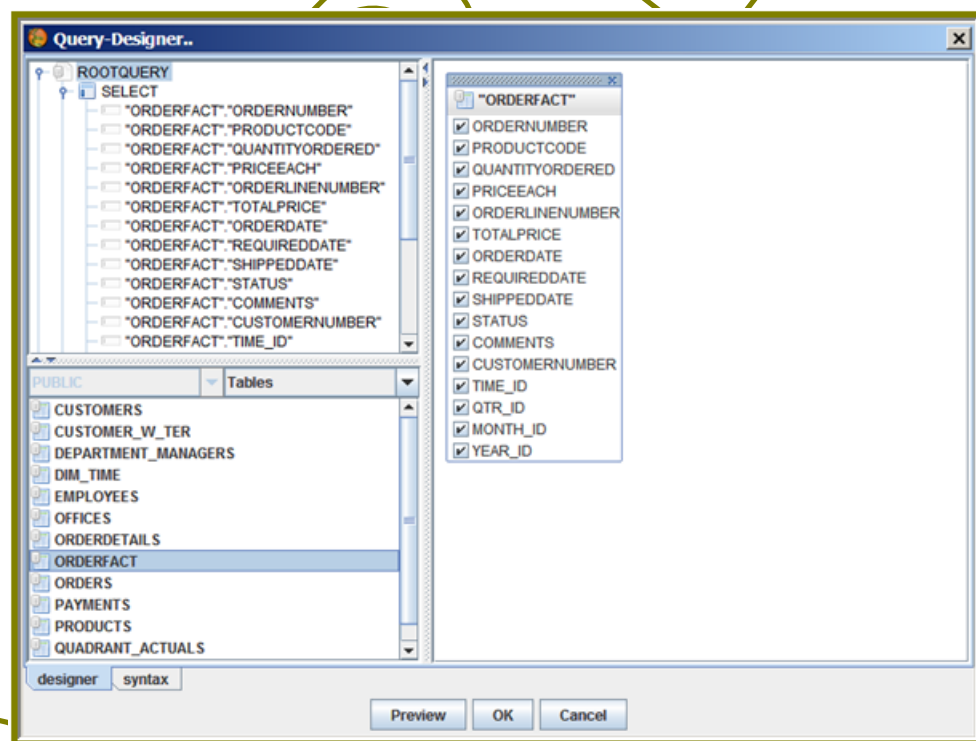
7. Click  (Edit).

The Choose Schema dialog box appears.

8. In the **Choose Schema** dialog box select **Public** from the drop-down list.

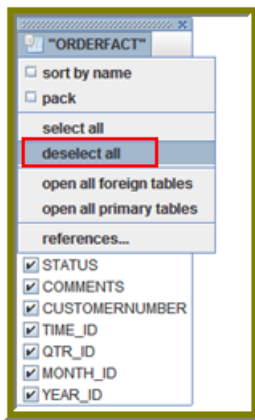


The Query Designer window appears. The Query Designer provides you with a graphical environment that allows you to work with the data even if you don't understand SQL, the standard programming language for retrieving content from databases.

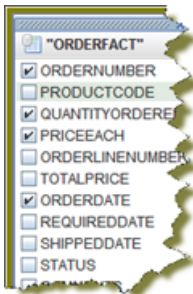


9. Double-click **ORDERFACT** so that the table appears in the workspace as shown in the image above.
10. In the Query Designer workspace, right-click "**ORDERFACT**" and choose **deselect all**.





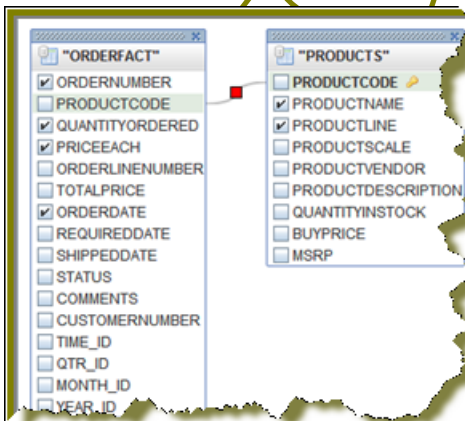
11. Now, select the following fields in the **ORDERFACT** table: **ORDERNUMBER**, **QUANTITYORDERED**, **PRICEEACH**, and **ORDERDATE**.



12. 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.

13. Deselect all **PRODUCTS** table fields, except for **PRODUCTNAME** and **PRODUCTLINE**.

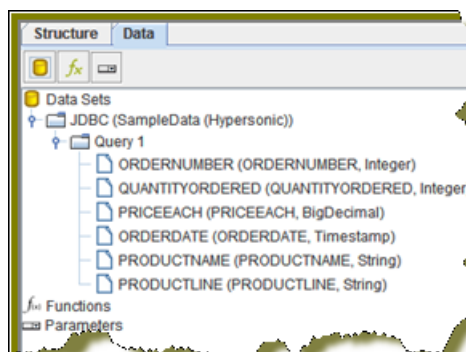


14. For the purpose of this exercise, click **Syntax** in the lower left portion of the Query Builder workspace to display a simple SQL statement associated with the tables. Notice that **PRODUCTCODE** is the common field between the **ORDERFACT** and **PRODUCTS** tables.



15. Click **OK** in the Syntax window to return to the **Configure** page. Notice that the SQL statement appears on the right under **Query Details**.
16. In the JDBC Data Source dialog box, click **OK** to return to the **Design** page.

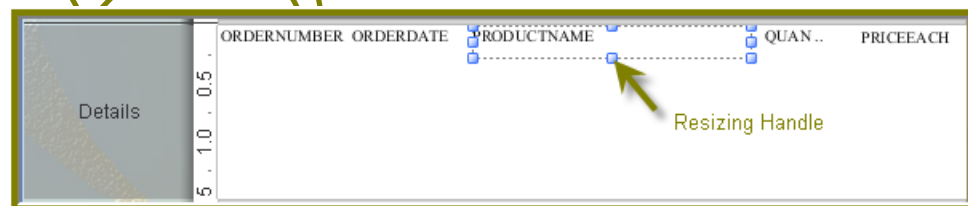
Notice that the fields associated with your tables are listed under **Query 1**. You are now ready to start **designing your report** on page 15 .



## Designing Your Report

This exercise walks you through the process of designing the look-and-feel of your report.

1. 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.
2. 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.
3. Use the resizing handles to make the **PRODUCTNAME** field larger and the **QUANTITYORDERED** field smaller as shown in the example below:



4. You have created your first report. Click  (Preview) to examine your report.

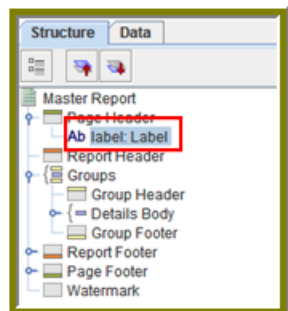
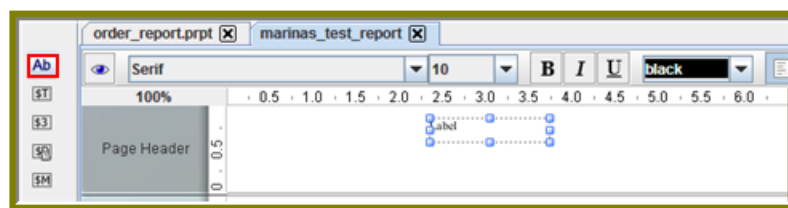
10,107	Feb 24, 2003	1969 Harley Davidson Ultimate Cho..	30	95.7
10,121	May 7, 2003	1969 Harley Davidson Ultimate Cho..	34	81.35
10,134	Jul 1, 2003	1969 Harley Davidson Ultimate Cho..	41	94.74
10,145	Aug 25, 2003	1969 Harley Davidson Ultimate Cho..	45	83.26
10,159	Oct 10, 2003	1969 Harley Davidson Ultimate Cho..	49	100
10,168	Oct 28, 2003	1969 Harley Davidson Ultimate Cho..	36	96.66
10,180	Nov 11, 2003	1969 Harley Davidson Ultimate Cho..	29	86.13
10,188	Nov 18, 2003	1969 Harley Davidson Ultimate Cho..	48	100
10,201	Dec 1, 2003	1969 Harley Davidson Ultimate Cho..	22	98.57
10,211	Jan 15, 2004	1969 Harley Davidson Ultimate Cho..	41	100
10,223	Feb 20, 2004	1969 Harley Davidson Ultimate Cho..	37	100
10,237	Apr 5, 2004	1969 Harley Davidson Ultimate Cho..	23	100
10,251	May 18, 2004	1969 Harley Davidson Ultimate Cho..	28	100
10,263	Jun 28, 2004	1969 Harley Davidson Ultimate Cho..	34	100
10,275	Jul 23, 2004	1969 Harley Davidson Ultimate Cho..	45	92.83
10,285	Aug 27, 2004	1969 Harley Davidson Ultimate Cho..	36	100
10,299	Sep 30, 2004	1969 Harley Davidson Ultimate Cho..	23	100
10,309	Oct 15, 2004	1969 Harley Davidson Ultimate Cho..	41	100
10,318	Nov 2, 2004	1969 Harley Davidson Ultimate Cho..	46	94.74
10,329	Nov 15, 2004	1969 Harley Davidson Ultimate Cho..	42	100
10,341	Nov 24, 2004	1969 Harley Davidson Ultimate Cho..	41	100
10,354	Dec 4, 2004	1969 Harley Davidson Ultimate Cho..	42	86.13
10,361	Dec 17, 2004	1969 Harley Davidson Ultimate Cho..	20	72.55
10,375	Feb 3, 2005	1969 Harley Davidson Ultimate Cho..	21	34.91
10,388	Mar 3, 2005	1969 Harley Davidson Ultimate Cho..	42	76.36
10,399	Apr 1, 2005	1969 Harley Davidson Ultimate Cho..	40	100
10,403	Apr 8, 2005	1969 Harley Davidson Ultimate Cho..	24	100
10,417	May 13, 2005	1969 Harley Davidson Ultimate Cho..	66	100
10,103	Jan 29, 2003	1952 Alpine Renault 1300	26	100
10,112	Mar 24, 2003	1952 Alpine Renault 1300	29	100

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](#) on page 16 .

## Refining Your Report

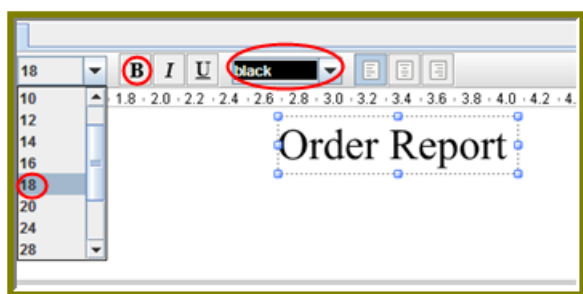
You've created a report in the previous exercise but now you need to make the report more descriptive so that users can understand the content in the report. Follow the instructions below to refine your report.

1. In the Design page, click and drag a **Ab** (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).



2. Click inside the **Label** item and type **Order Report**

- 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.


- With the **Order Report** label still selected, click down arrow of the font color icon in the toolbar. Select a color for your label.

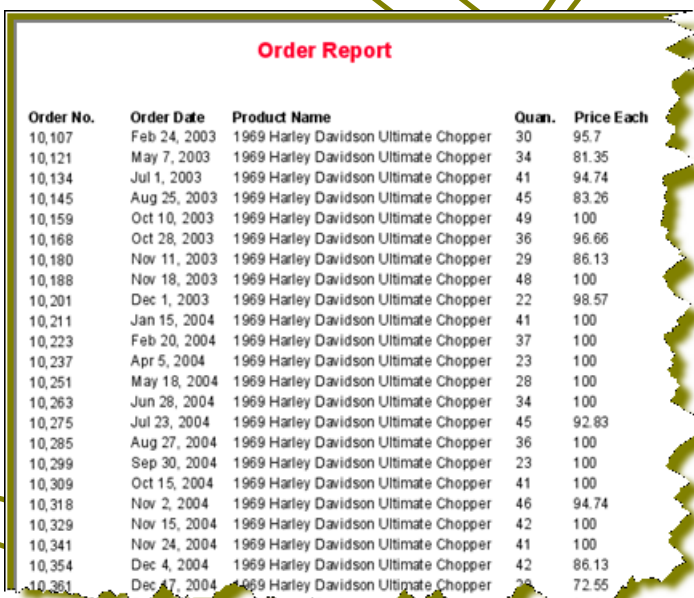
The font color changes. This page header will appear on every page of your report.

- Now, you must create column headers. Click and drag labels for the following columns into the **Page Header** band: **Order No.**, **Order Date**, **Product Name**, **Quan.**, and **Price Each** as shown in the example below. Apply boldface to your column header names if you choose.




**Note:** To align a header and field name perfectly, select the field name then use **SHIFT +CTRL+CLICK** to select the header. With both selected, right-click and select **Align -> Align Left**. The items will align perfectly. You may need to play with alignment and resizing to ensure that your report displays correctly.

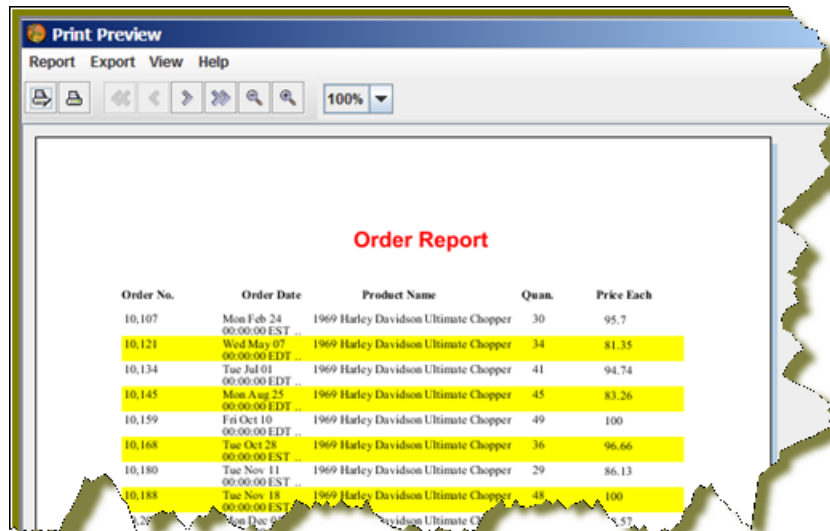
- Click  (Preview) to display your report.



Order No.	Order Date	Product Name	Quan.	Price Each
10,107	Feb 24, 2003	1969 Harley Davidson Ultimate Chopper	30	95.7
10,121	May 7, 2003	1969 Harley Davidson Ultimate Chopper	34	81.35
10,134	Jul 1, 2003	1969 Harley Davidson Ultimate Chopper	41	94.74
10,145	Aug 25, 2003	1969 Harley Davidson Ultimate Chopper	45	83.26
10,159	Oct 10, 2003	1969 Harley Davidson Ultimate Chopper	49	100
10,168	Oct 28, 2003	1969 Harley Davidson Ultimate Chopper	36	96.66
10,180	Nov 11, 2003	1969 Harley Davidson Ultimate Chopper	29	86.13
10,188	Nov 18, 2003	1969 Harley Davidson Ultimate Chopper	48	100
10,201	Dec 1, 2003	1969 Harley Davidson Ultimate Chopper	22	98.57
10,211	Jan 15, 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,223	Feb 20, 2004	1969 Harley Davidson Ultimate Chopper	37	100
10,237	Apr 5, 2004	1969 Harley Davidson Ultimate Chopper	23	100
10,251	May 18, 2004	1969 Harley Davidson Ultimate Chopper	28	100
10,263	Jun 28, 2004	1969 Harley Davidson Ultimate Chopper	34	100
10,275	Jul 23, 2004	1969 Harley Davidson Ultimate Chopper	45	92.83
10,285	Aug 27, 2004	1969 Harley Davidson Ultimate Chopper	36	100
10,299	Sep 30, 2004	1969 Harley Davidson Ultimate Chopper	23	100
10,309	Oct 15, 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,318	Nov 2, 2004	1969 Harley Davidson Ultimate Chopper	46	94.74
10,329	Nov 15, 2004	1969 Harley Davidson Ultimate Chopper	42	100
10,341	Nov 24, 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,354	Dec 4, 2004	1969 Harley Davidson Ultimate Chopper	42	86.13
10,361	Dec 17, 2004	1969 Harley Davidson Ultimate Chopper	29	72.55

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

7. In the toolbar, go to **Format -> Row Banding**.
8. In the Row Banding dialog box, choose **Yellow** from the drop-down list next to **Visible Color** and click **OK**.
9. Click  (Preview) to display your report.



**Print Preview**

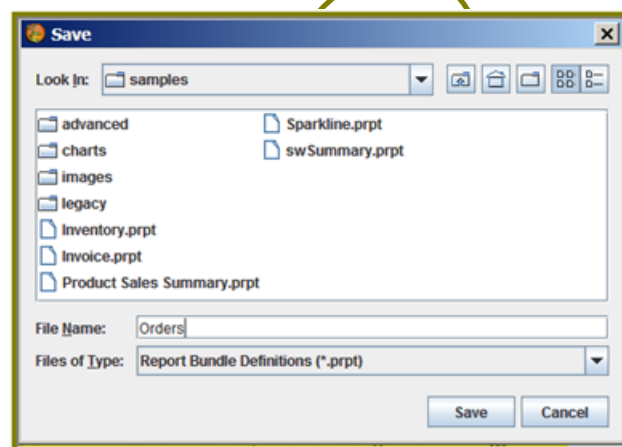
Report Export View Help

100%

**Order Report**

Order No.	Order Date	Product Name	Quan.	Price Each
10,107	Mon Feb 24 00:00:00 EST	1969 Harley Davidson Ultimate Chopper	30	95.7
10,121	Wed May 07 00:00:00 EDT	1969 Harley Davidson Ultimate Chopper	34	81.35
10,134	Tue Jul 01 00:00:00 EDT	1969 Harley Davidson Ultimate Chopper	41	94.74
10,145	Mon Aug 25 00:00:00 EDT	1969 Harley Davidson Ultimate Chopper	45	83.26
10,159	Fri Oct 10 00:00:00 EDT	1969 Harley Davidson Ultimate Chopper	49	100
10,168	Tue Oct 28 00:00:00 EST	1969 Harley Davidson Ultimate Chopper	36	96.66
10,180	Tue Nov 11 00:00:00 EST	1969 Harley Davidson Ultimate Chopper	29	86.13
10,188	Tue Nov 18 00:00:00 EST	1969 Harley Davidson Ultimate Chopper	48	100
10,200	Mon Dec 04 00:00:00 EST	1969 Harley Davidson Ultimate Chopper	35	95.7

10. Go to **File -> Save** to save your report in the **...report-designer/configuration-template/report-designer/samples** folder. Type **Orders** in the **File Name** text box.




## 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**.

The **Add Parameter** dialog box appears.

3. In the Add Parameter dialog box, type **Line** in the **Name** text field.
4. Next to **Type**, select **Text** so users can select a product line.

5. Click  to add a query that supplies the values, (motorcycles, cars, ships, and so on), from which users must choose.
6. Click **JDBC**.

The JDBC Data Sources dialog box appears.

7. Under **Connections**, select **SampleData (Hypersonic)**.
8. Next to **Available Queries** click .
9. In the Query Name text field, use your backspace key to delete **Query 1**. Type **List\_of\_ProductLines** in the text field.
10. Enter your SQL query in the **Query** box.

You can copy and paste the required lines, (shown below) directly into the SQL statement.

**Important:** Make sure to use curly brackets, (not parentheses), before and after {enter\_procline} 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 as shown in the image below. If you do not add the lines, the report displays orders for all product lines.

11. You are now ready to **publish your report** on page 19.

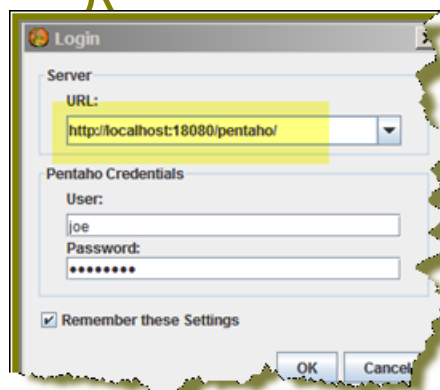
## Publishing Your Report

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 .

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:18080/pentaho/`.



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. Leave the **Location** as is.
6. Select **html** as the **Report Output Type**.
7. In the Publish Password text field, type the publish password, **password**, and click **OK**.

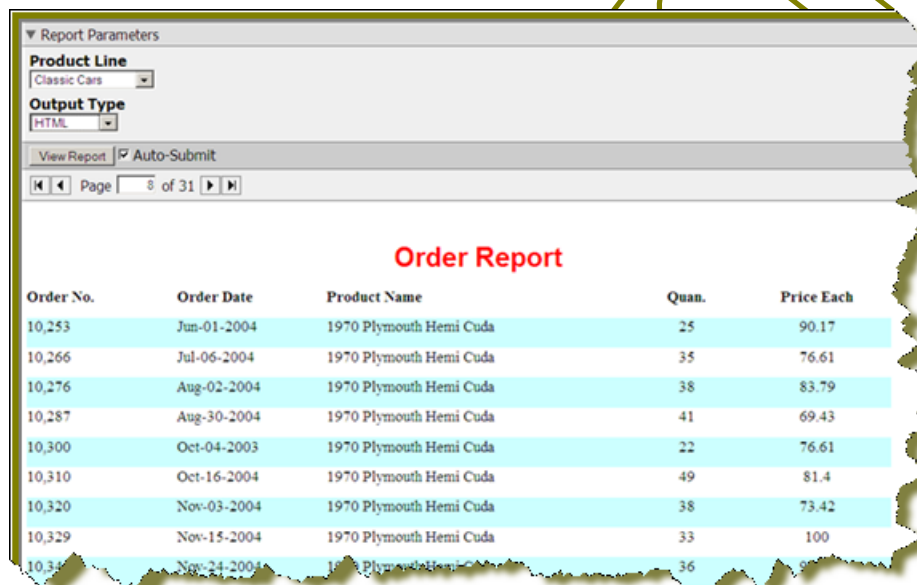
A success message appears.

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:18080** 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.



The screenshot shows the Pentaho User Console interface. At the top, there's a 'Report Parameters' section with a 'Product Line' dropdown set to 'Classic Cars' and an 'Output Type' dropdown set to 'HTML'. Below this are buttons for 'View Report' and 'Auto-Submit' (checked). A pagination bar shows 'Page 8 of 31'. The main content area displays a table titled 'Order Report' with the following data:

Order No.	Order Date	Product Name	Quan.	Price Each
10,253	Jun-01-2004	1970 Plymouth Hemi Cuda	25	90.17
10,266	Jul-06-2004	1970 Plymouth Hemi Cuda	35	76.61
10,276	Aug-02-2004	1970 Plymouth Hemi Cuda	38	83.79
10,287	Aug-30-2004	1970 Plymouth Hemi Cuda	41	69.43
10,300	Oct-04-2003	1970 Plymouth Hemi Cuda	22	76.61
10,310	Oct-16-2004	1970 Plymouth Hemi Cuda	49	81.4
10,320	Nov-03-2004	1970 Plymouth Hemi Cuda	38	73.42
10,329	Nov-15-2004	1970 Plymouth Hemi Cuda	33	100
10,341	Nov-24-2004	1970 Plymouth Hemi Cuda	36	90.17

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. Pentaho does not recommend this feature for use in an enterprise or production environment; however, it can be useful for testing and proof-of-concept scenarios.

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](#) on page 23 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 **Add** in the **Select Data Source** step in Ad Hoc Reports
- When you click **Create Data** in charts in the Dashboard
- When you click (+) as a first step of the create Data Table process

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. In the **New Data Source** dialog box, enter your SQL Query. Click **Preview** to ensure that your query is correct.
9. Click **Apply**.

10. If applicable, edit the display names of the columns, the Type (string, numeric), and the Aggregation Type (a default aggregation type for a column is displayed). A sample of each data type displays under Sample. Aggregation types are defined below:

Aggregate Type	Description
None	No aggregation is assigned
Sum	Sums a specific columns values determined by grouping
Average	Averages a specific columns values determined by grouping
Count	Returns the number of distinct values associated with the specified column determined by grouping
Distinct Count	Returns the number of distinct values of the specified column determined by grouping
Minimum	Selects the minimum column value determined by grouping
Maximum	Selects the maximum column value determined by grouping

11. Click **OK**.

In the background, after you click **OK**, a default metadata data source is generated based on the query. When you convert the query to a metadata data source, the query can be reused in Ad Hoc Reports, Report Designer, 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 **Create Data** in charts in the Dashboard
- When you click (+) as a first step of the create Data Table process

**Important:** Pentaho does not recommend the use of CSV-based data sources in a production environment. See [Using the Data Sources Feature in the Pentaho User Console](#) on page 20 for more information.

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

1. In the New Data Source dialog box, click **CSV** to define a CSV file as your data source.
2. In the **Name** field, type a name that identifies your new data source. (This is a free-form text field.)
3. Under **File**, click **Browse** to locate the CSV file you want to use as you data source.
4. If your delimiter and enclosure types are different than the defaults, comma

**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.

5. Click **Apply**.

The Metadata table updates with CSV data. By default, the first row of your CSV file set as the header.

6. If applicable, change the Display Name and Type values. The samples column displays the values for the first row. Click the ellipsis (...) to see additional details.
7. If you choose to change a default aggregate click the ellipsis (...) next to the aggregate type to launch a dialog box that allows you to enable and disable aggregate selections. Aggregate types are defined below:

Aggregate Type	Description
None	No aggregation is assigned
Sum	Sums a specific columns values determined by grouping
Average	Averages a specific columns values determined by grouping
Count	Returns the number of distinct values associated with the specified column
Distinct Count	Returns the number of distinct values of the specified column
Minimum	Selects the minimum column value determined by grouping
Maximum	Selects the maximum column value determined by grouping

8. Click **OK** to create the new data source.

**Note:** If you select the wrong CSV file, click **Browse** and repeat steps 3 through 7.

## 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. The default values are shown in the sample below. 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.

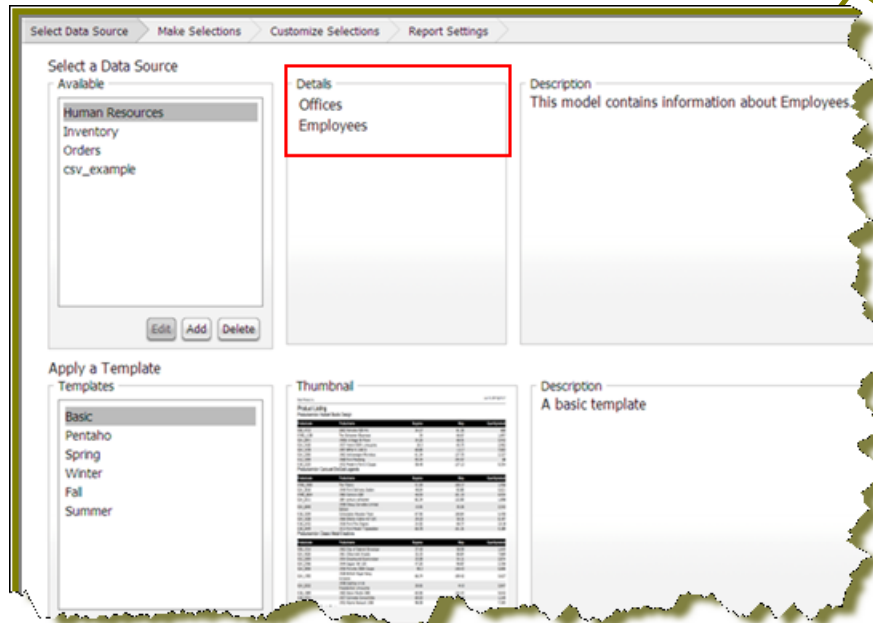
## Creating an Ad Hoc Report

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.

New in version 3.5, is the ability to 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](#) on page .



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.

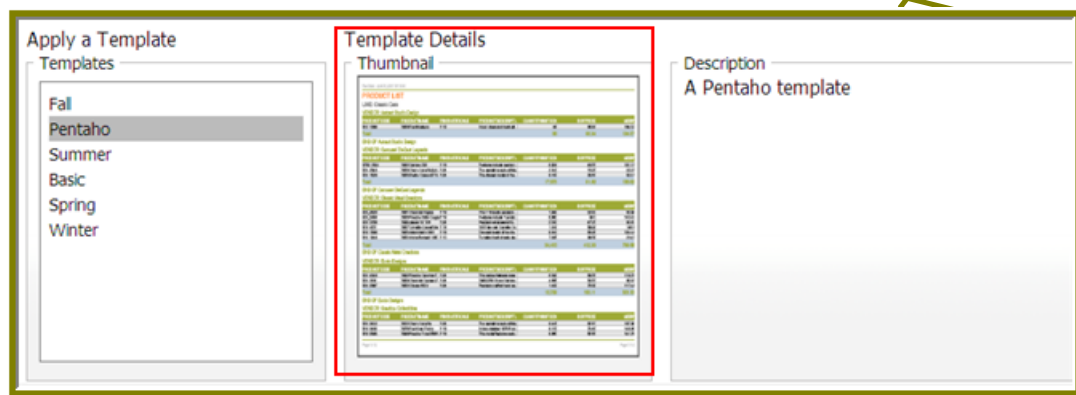
2. In the first step of the wizard, select **Orders** from the list under **Select a Data Source**.

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 \_\_\_\_\_ 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.

Territory: APAC	
Amount	
\$2,476.8	
\$9,738.18	
\$10,640.29	
\$12,334.82	
\$21,730.03	
\$23,294.8	
\$24,219.59	
\$26,421.24	
\$27,098.8	
\$28,397.26	
\$29,852.17	
\$34,332.29	
\$35,160.25	
\$35,911.81	
\$36,409.31	
\$37,754.33	
\$37,878.55	
\$37,905.15	
\$39,267.74	
\$40,034.67	
\$41,297.14	
\$41,418.55	
\$41,791.95	
\$42,902.84	
\$43,930.62	
\$5,098.72	

## 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.

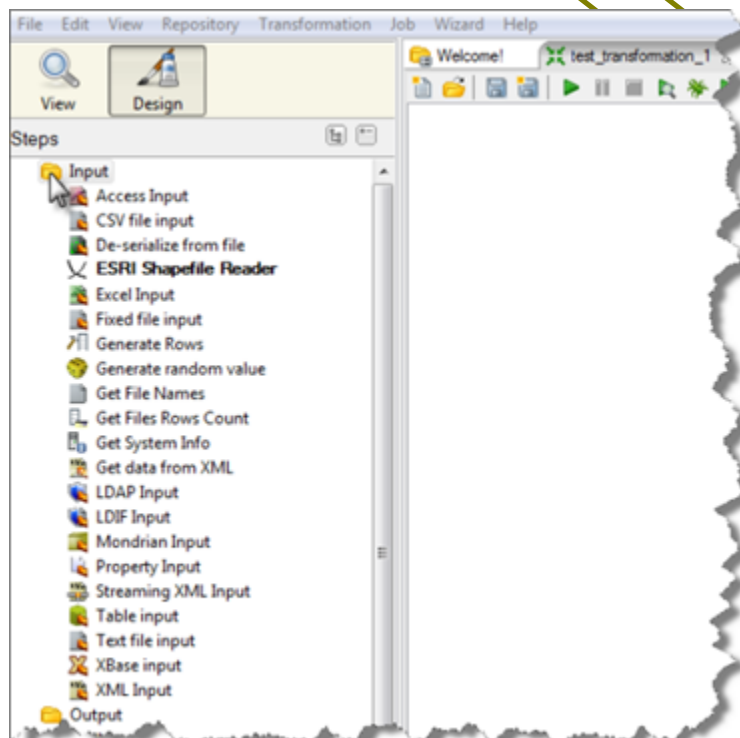
You now know which product line has the most cancelled orders.

## Building a simple input-output transformation

You must have a database connection to complete this exercise.

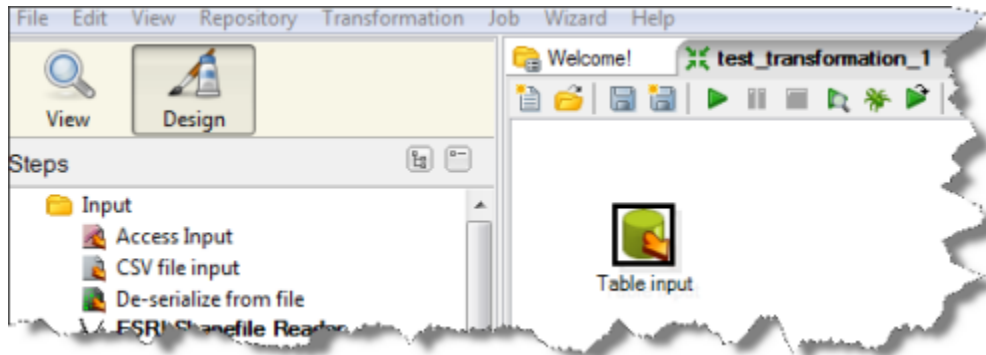
This exercise walks you through the process of building a transformation that uses input from a database table (a list of customers), applies an SQL statement, and outputs the data stream to a plain text file. This exercise is useful for testing and tuning transformations to see how the fields are passed 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 folder and view the input step options.



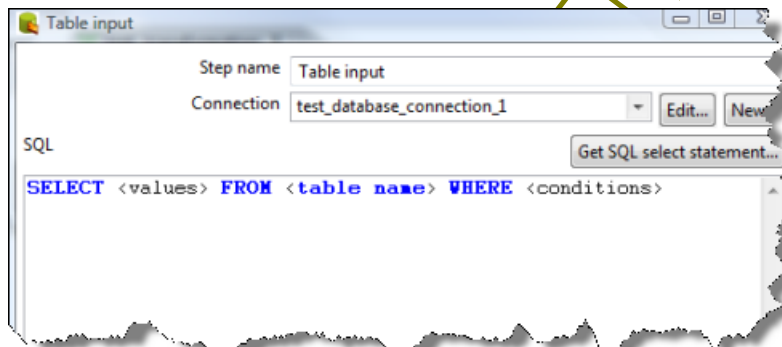
2. Click **Table Input** and drag the cursor to the right pane.





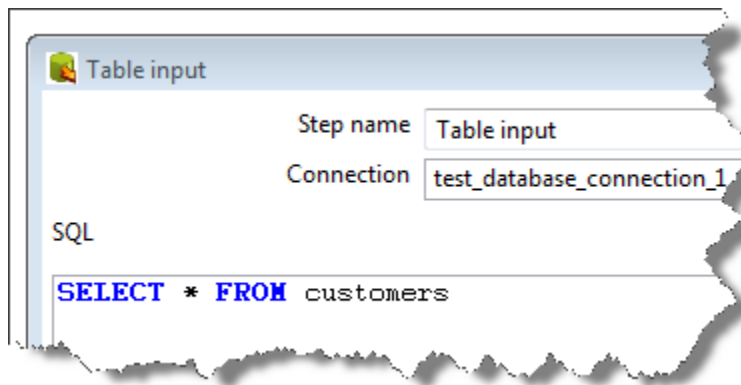
A **Table Input** step is added to the transformation. This step reads information from the database using the database connection and SQL.

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:

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



- 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.

Examine preview data

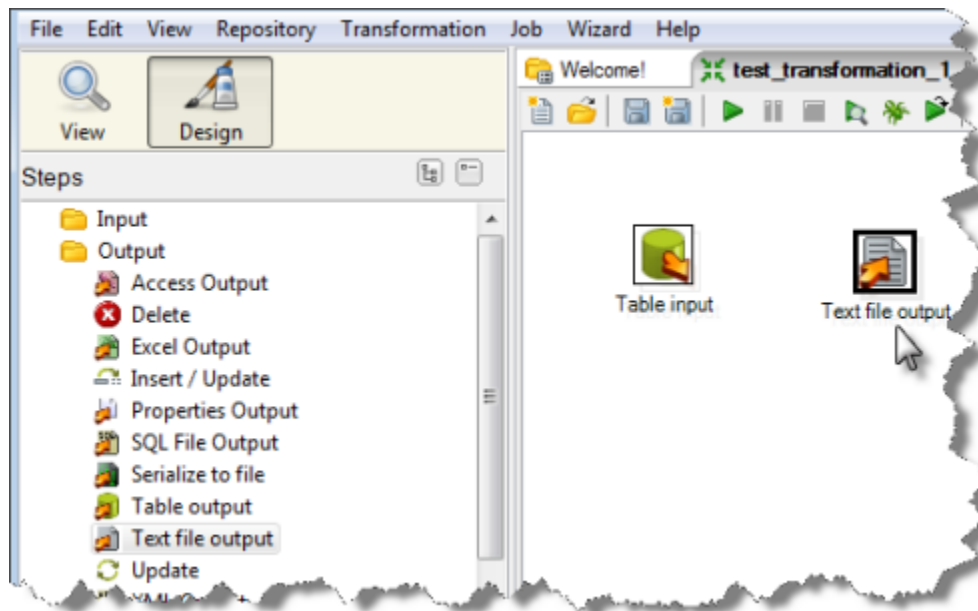
Rows of step: Table input

#	CUSTOMERNUMBER	CUSTOMERNAME	CONTACTLASTNAME	CONTACTFIRSTNAME
1	103	Atelier graphique	Schmitt	Carine
2	112	Signal Gift Stores	King	Sue
3	114	Australian Collectors, Co.	Ferguson	Peter
4	119	La Rochelle Gifts	Labrun	Janine
5	121	Baane Mini Imports	Bergulfsen	Jonas
6	124	Mini Gifts Distributors Ltd.	Nelson	Valarie
7	125	Havel & Zbyszek Co	Piestrzeniewicz	Zbyszek
8	128	Blauer See Auto, Co.	Keitel	Roland
9	129	Mini Wheels Co.	Murphy	Julie

Close Show Log

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.

- 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.



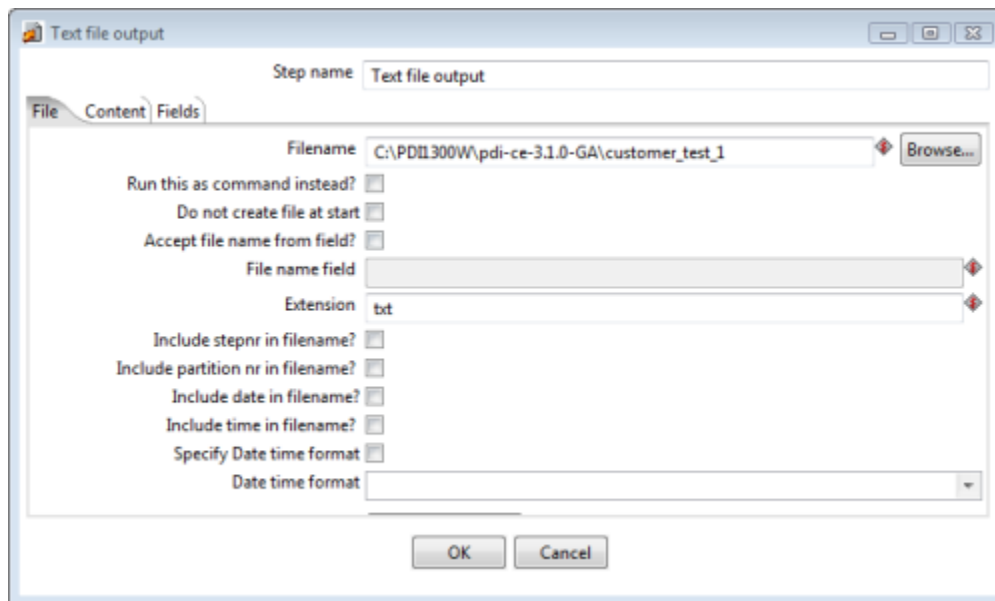
A **Text File Output** step is added to the transformation. This step sends the fields in the incoming data stream to a text file.

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

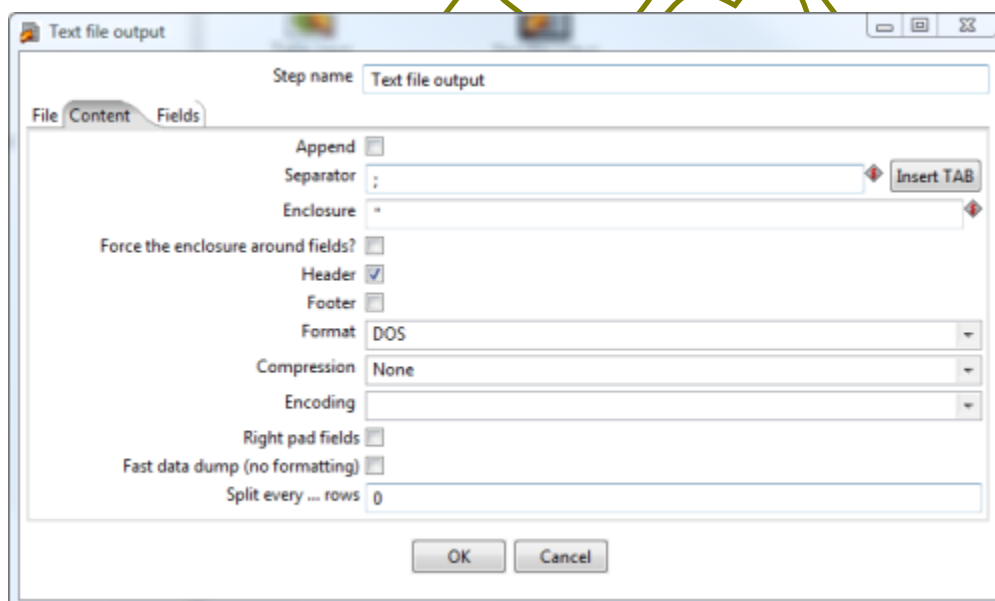


This adds a hop (data connection) between the two steps and displays as an arrow connecting the steps. The hop represents the data stream flowing from the **Table Input** step to the **Text File Output** step.

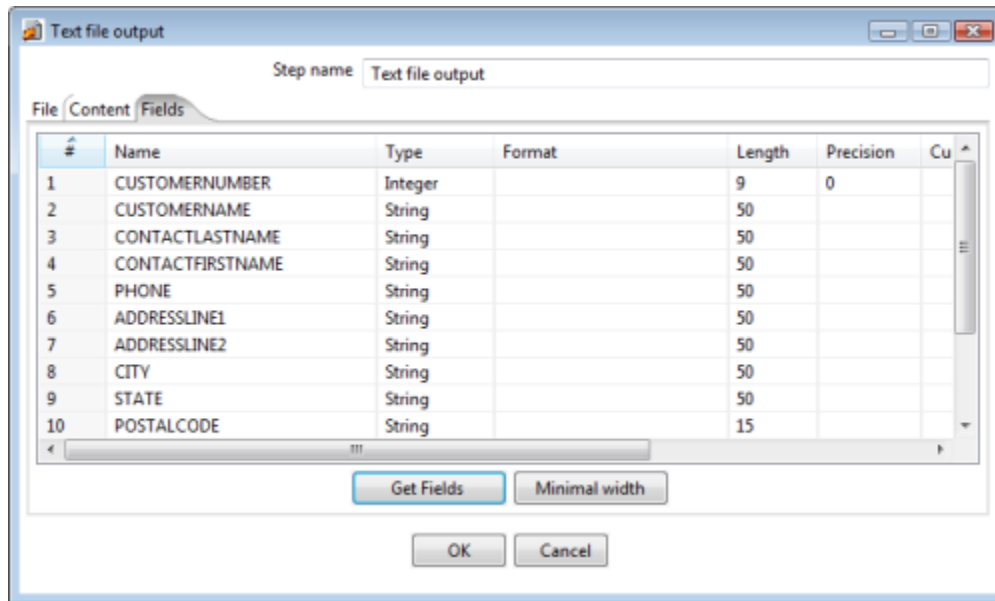
7. Double-click the **Text File Output** step and type a path/file name in the **File name** box.



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



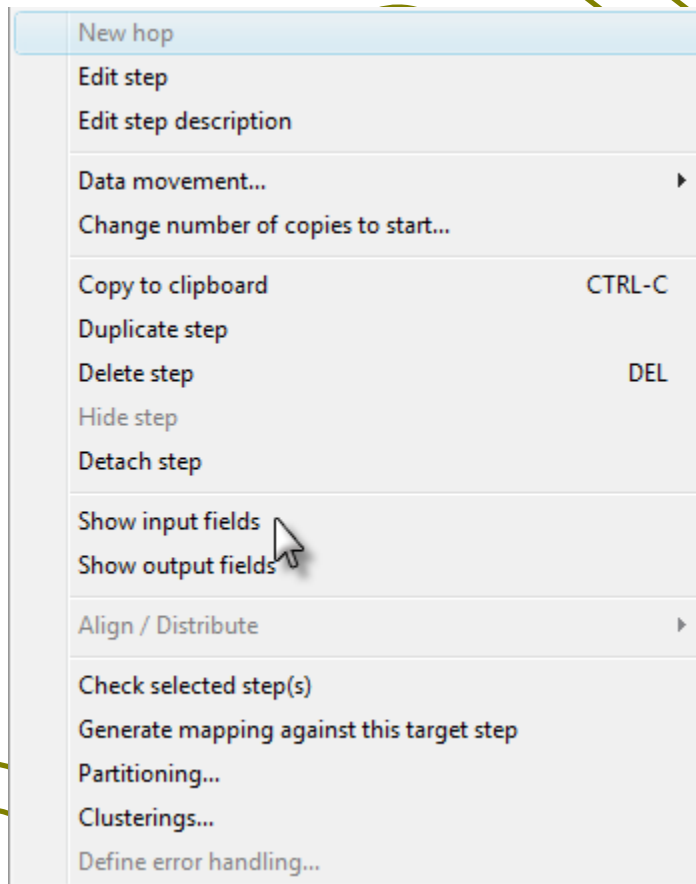
8. Click the Fields tab, then click **Get Fields**.



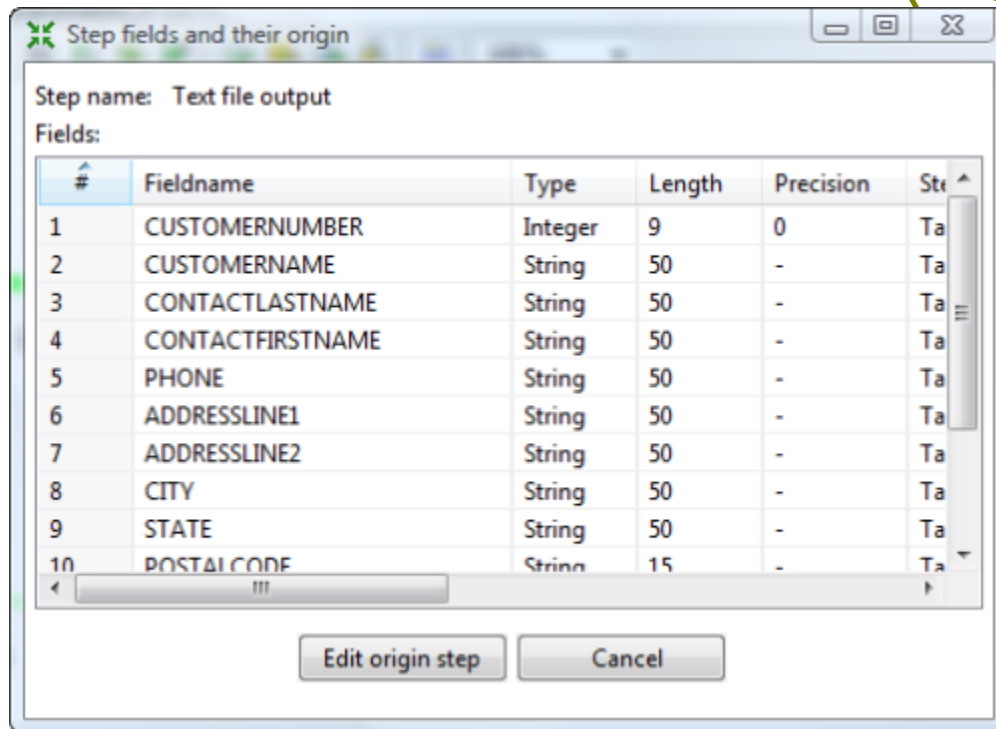
**Note:** The file extension (in this case, .txt) is filled in for you.

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.

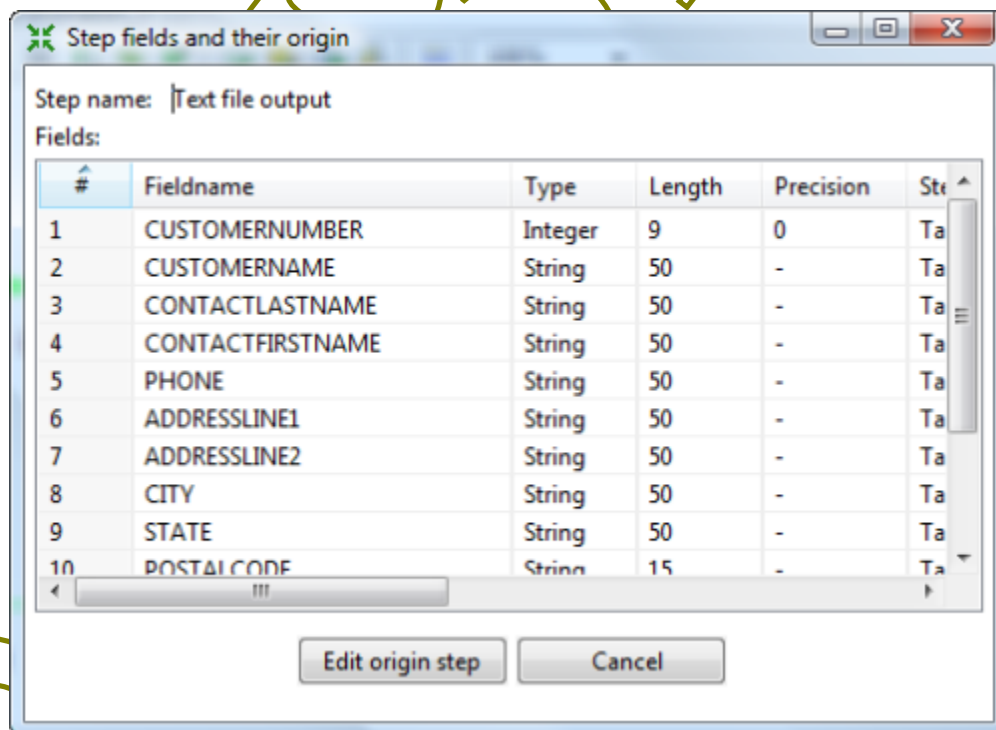
9. Right-click **Text File Output** to display the context menu, then click **Show Input Fields**.



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**.



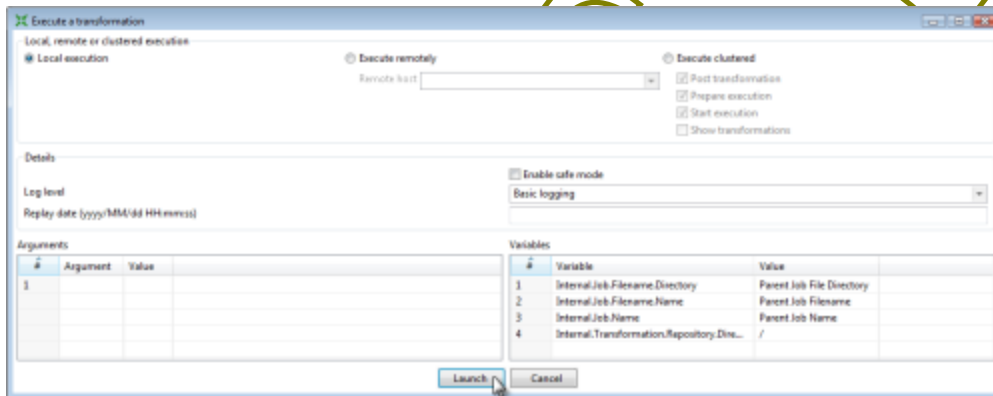
This dialog box displays all the fields in the data output stream and their transformations. In this simple example, there are no intermediate transformation steps, for example filters, but in a transformation with intermediate steps, the details display here. Together, the **Show Input Fields** and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

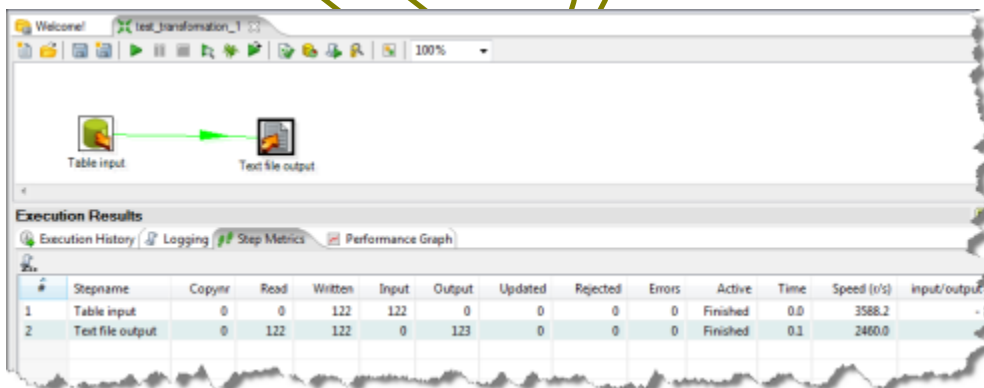


You must save the transformation before you can run it.

12. Click **Launch**.

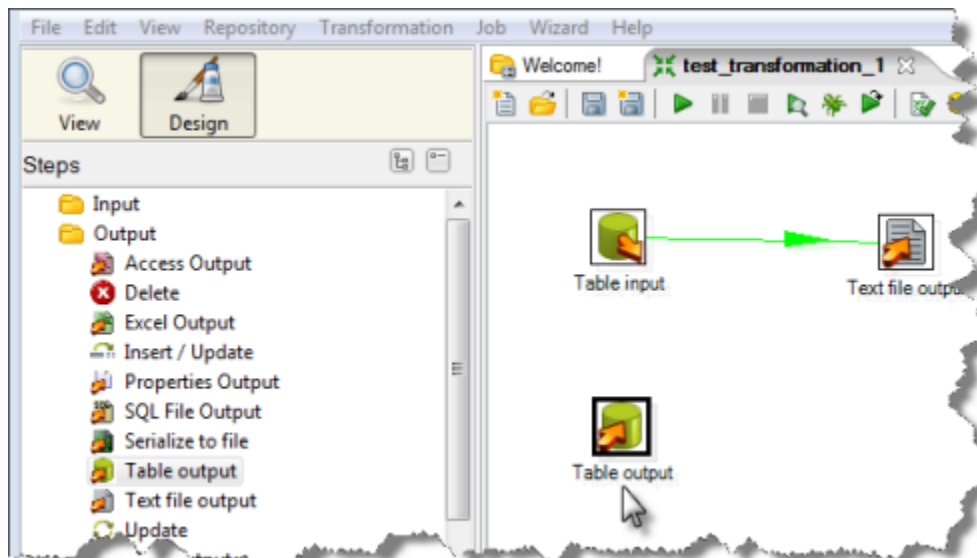


The transformation runs and the results display in a pane in the lower portion of the window.



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.

