

# Get Started with Pentaho Data Integration and Analytics

9.4.x

---

Pentaho Data Integration and Analytics

MK-95PDIA000-05

Last updated: 2024-02-28

Generated from docs.hitachivantara.com



# Get started with Pentaho Data Integration and Analytics

Pentaho Data Integration and Analytics is a comprehensive platform used to access, integrate, manipulate, visualize, and analyze your data. Whether data is stored in a flat file, relational database, Hadoop cluster, NoSQL database, analytic database, social media streams, operational stores, or in the cloud, Pentaho products can help you discover, analyze, and visualize data to find the answers you need, even if you have no coding experience. Advanced users with programming experience can use our extensive API to customize reports, queries, transformations to extend functionality.

## Data Integration and Analytics components and tools

Pentaho Data Integration and Analytics applications include both web-based components and design tools. The components and tools you use depend on your workflow and what your environment supports.

## Data Integration and Analytics web-based components

Use the Pentaho web-based components to share business intelligence solutions by analyzing data, creating reports, and building integrated dashboards.

These components include:

### Pentaho User Console (PUC)

A design environment for accessing Analyzer, Interactive Reports, and Dashboard Designer. Pentaho User Console also offers system administration features for configuring your Pentaho Server.

### Analyzer

Helps you visualize data to make informed business decisions. You can create geographic, scatter chart, heat grid, and multi-chart visualizations. You can also filter data, add query parameters, configure drill-down links, apply conditional formatting, and to generate hyperlinks.

### Interactive Reports

A design interface used to create both simple and on-demand operational reports without relying on IT or report developers. You can quickly add elements to your report and format them any way you like.

### Dashboard Designer

Choose layout templates, themes, and content to create visually attractive dashboards that help decision makers gain critical knowledge at a glance. You can combine a wide variety of content including Interactive Reports, Analyzer visualizations, and collaborative content.

### CTools

A community-driven framework for creating dashboards by using web technologies such as JavaScript, CSS, and HTML. You can easily create dynamic dashboards for users to explore and understand large amounts of data using a variety of charts, tables, and other components.

## Data Integration and Analytics design tools

Use Pentaho design tools to develop and refine how your data values are reported, modeled, transformed, and stored.

These tools include:

#### Pentaho Data Integration (PDI)

Provides access to an Extraction, Transformation, and Loading (ETL) engine that captures the right data, cleanses the data, and stores data using a uniform format that is accessible and relevant to end users and IoT technologies.

#### Report Designer

Generate detailed, pixel-perfect reports using virtually any data source. It enables business intelligence professionals to create highly detailed, print-quality reports based on adequately prepared data.

#### Aggregation Designer

Provides a simple interface that allows you to create aggregate tables from levels within the dimensions you specify. Use this tool to improve the performance of your Pentaho Analysis (Mondrian) OLAP cubes.

#### Metadata Editor

Simplifies report building. Use the Metadata Editor to build Pentaho metadata domains and models. A Pentaho Metadata Model maps the physical structure of your database into a logical business model.

#### Schema Workbench

Edit and create multidimensional (Mondrian) models. You can create Mondrian models graphically or define them by hand-coding XML files.

## Evaluate Pentaho Data Integration and Analytics

Before setting up and using Pentaho Data Integration and Analytics in your production environment, you should evaluate it by performing the following steps:

1. [Install the evaluation version.](#)
2. [Learn about Analyzer, Interactive Reports, and Dashboard Designer.](#)
3. [Learn about Report Designer.](#)
4. [Learn about Pentaho Data Integration and the PDI client.](#)
5. [Learn about Pentaho, big data, and clusters.](#)
6. [Understand common workflows.](#)

## Data Integration and Analytics supported technologies

Reference material for supported components and JDBC drivers.

#### [Components Reference](#)

Reference matrix for supported components.

#### [JDBC Drivers Reference](#)

Reference matrix for supported JDBC drivers.

# Install the 30-day trial of Pentaho Data Integration and Analytics

The Pentaho Installation Wizard provides the easiest and quickest way to install, learn about, and evaluate the Pentaho Suite. With this 30-day trial, you can fully install and test Pentaho's Business Analytics (BA) and Data Integration (DI) components.

To get you started quickly after installation, the [Pentaho Data Integration \(PDI\) tutorial](#) explains how to build a transformation with PDI using sample data in a typical business scenario.

Note: As a best practice, use this installation method for evaluation purposes only. If you have any issues with your 30-day trial, contact the [Pentaho Trial Experts](#) for assistance.

With the Pentaho Installation Wizard you can choose one of two ways to install Pentaho components:

#### Default

Select the Keep it simple. Give me everything. option in the installation wizard.

#### Custom

Select the Let me decide for myself. option in the installation wizard.

## Process overview

Installation instructions are the same for the Windows, Linux, and Mac platforms.

1. [Download the software.](#)
2. [Start the installation wizard.](#)
3. Choose the [default installation](#) or [custom installation](#) method.
4. [Verify your installation.](#)
5. [Follow the steps in the tutorial to get started.](#)

Explore Considerations	
You Supply	A computer that meets Pentaho's <a href="#">operating system and hardware requirements</a> .
We Supply	<ul style="list-style-type: none"> <li>• Installation Package</li> <li>• Oracle JRE</li> <li>• Repository Platform (PostgreSQL)</li> <li>• Repository Platform JDBC Driver (PostgreSQL)</li> <li>• Web Application Server (Tomcat)</li> </ul>
Technologies Used	<ul style="list-style-type: none"> <li>• Tomcat web application server</li> <li>• PostgreSQL database to house the Pentaho Repository</li> </ul>
Expertise	<ul style="list-style-type: none"> <li>• Basic computer knowledge.</li> </ul>
Approximate Installation Time	30 minutes

## Download the software

To download the software, complete these steps.

1. Make sure that you are logged on to the computer on which you want to install the software.  
You should be logged on with an account that allows you to install software.
  - On Windows, this is typically an account that has administrator privileges.
  - On Linux, this is also an account that has administrator privileges, such as the pentaho account that you created during the previous step.
2. Navigate to the [Pentaho download](#) website.
3. Click Start a Free 30 Day Trial.  
The Submission Agreement form displays.
4. Fill in the requested information on the form and click Submit.  
The Download Pentaho EE Data Integration & Analytics dialog box displays.
5. Click Download in the Pentaho EE OnPrem section.
6. When prompted, choose a directory where you want to save the installer and wait for the installer to finish downloading.
7. If you are using Linux, make sure that you have the appropriate permission to execute the file.  
Open a Terminal window, navigate to the directory where you just downloaded the file, then enter



the following at the prompt.

```
chmod a+x ./pentaho-business-analytics-9.4.0-x64.bin
```

## Start the Pentaho Installation Wizard

To start the installation wizard, complete the following steps.

Note: Launch the installation wizard from a locally mounted hard drive only. Network-mounted drives (e.g. NFS or Netbeui) are not supported for performance reasons.

Do one of the following.

If you are using Windows

Open Windows Explorer and navigate to the installation file you just downloaded. Double-click the pentaho-business-analytics-9.4.0-x64.exe file to launch it.

If you are using Mac

Unpack the pentaho-business-analytics-9.4.0-x64.app.tar.gz archive, then double-click the pentaho-business-analytics-9.4.0-x64.app installation app.

If you are using Linux and your computer can display graphics

Open a Terminal window and navigate to the installation file you just downloaded. Enter the following command at the prompt, then press Enter.

```
./pentaho-business-analytics-9.4.0-x64.bin
```

If you are using Linux and your computer cannot display graphics

Open a Terminal window and navigate to the installation file you just downloaded. At the prompt enter one of the following commands, then press Enter.

If you have GTK libraries on your system and want to use the GTK Toolkit, enter this command :

```
./pentaho-business-analytics-9.4.0-x64.bin --mode text
```

If you do not have GTK libraries on your system or if you want to use OpenMotif, enter this command:

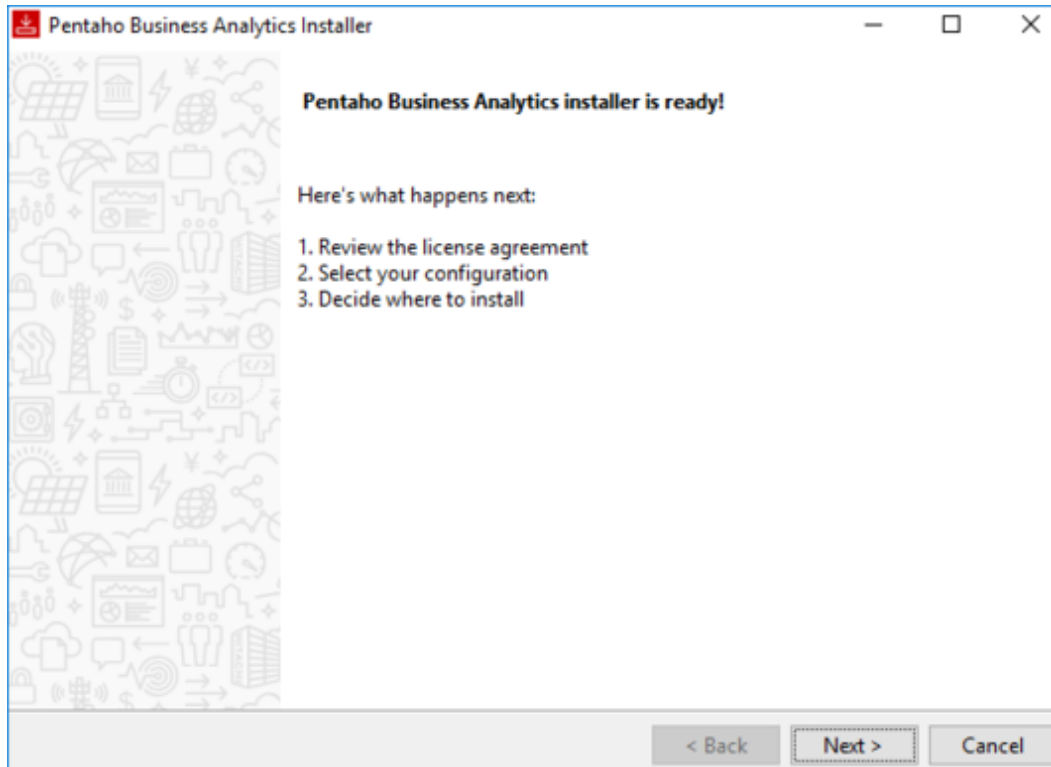
```
./pentaho-business-analytics-9.4.0-x64.bin --mode xwindow
```

## Default installation

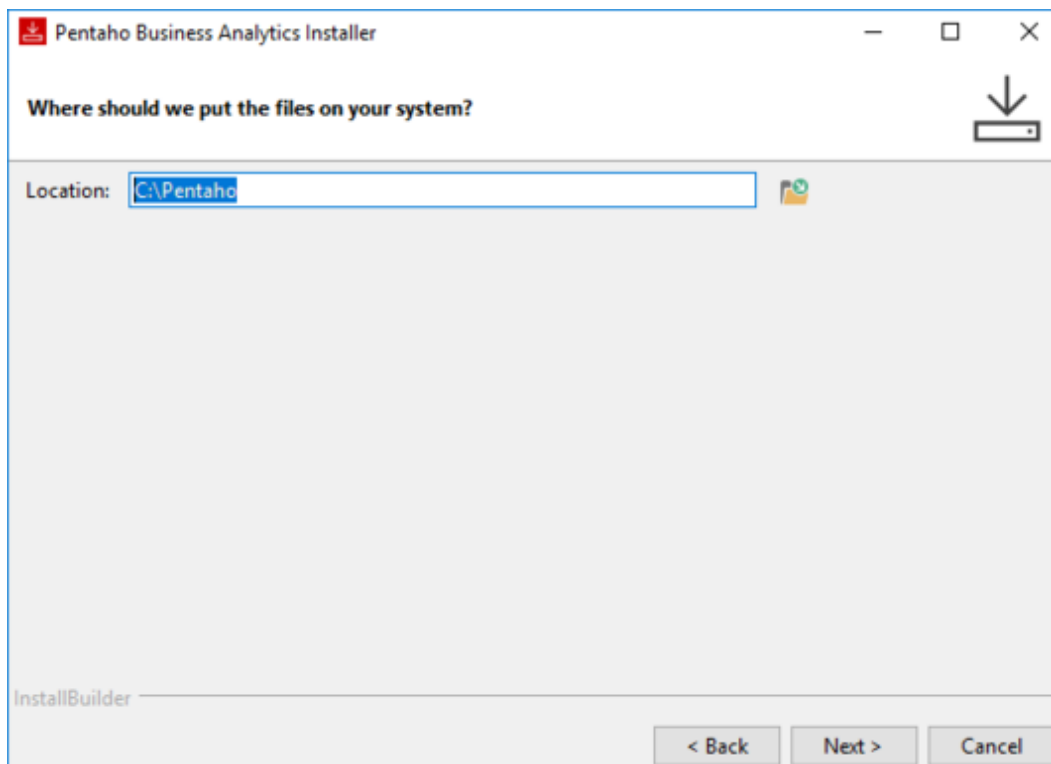
If you choose the Keep it simple. Give me everything. installation method, complete the following steps. If you want to install using a custom method, read the Custom Installation instructions.

Note: You cannot install into a directory that already exists. If you attempt to install into a directory that exists, an error message appears. If you are using Linux, the best practice is to not install the software in the /opt/ directory. The complicated permissions requirements and general isolation from the rest of the system might cause problems if the software is installed in this location.

1. After you start the installation wizard, the Pentaho Business Analytics Installation Wizard splash screen appears, then the Pentaho Business Analytics installation is ready! window. Click Next.



2. Read the license agreement. Click Accept, then click Next to continue.
3. Select the Keep it simple. Give me everything option in the Setup Type window, then click Next.
4. In the Installation folder window, accept the default directory or choose a different directory by entering the path in the Location text box where you want to install the software. Click Next.



5. If you selected to install the Pentaho Server, you are prompted to add a Postgres SQL password. Enter and confirm the password you want to assign to the postgres user who is assigned admin privileges for the PostgreSQL database.

Note:

Do not use these characters in the Password field because the installation wizard cannot process them:

' " & < > \

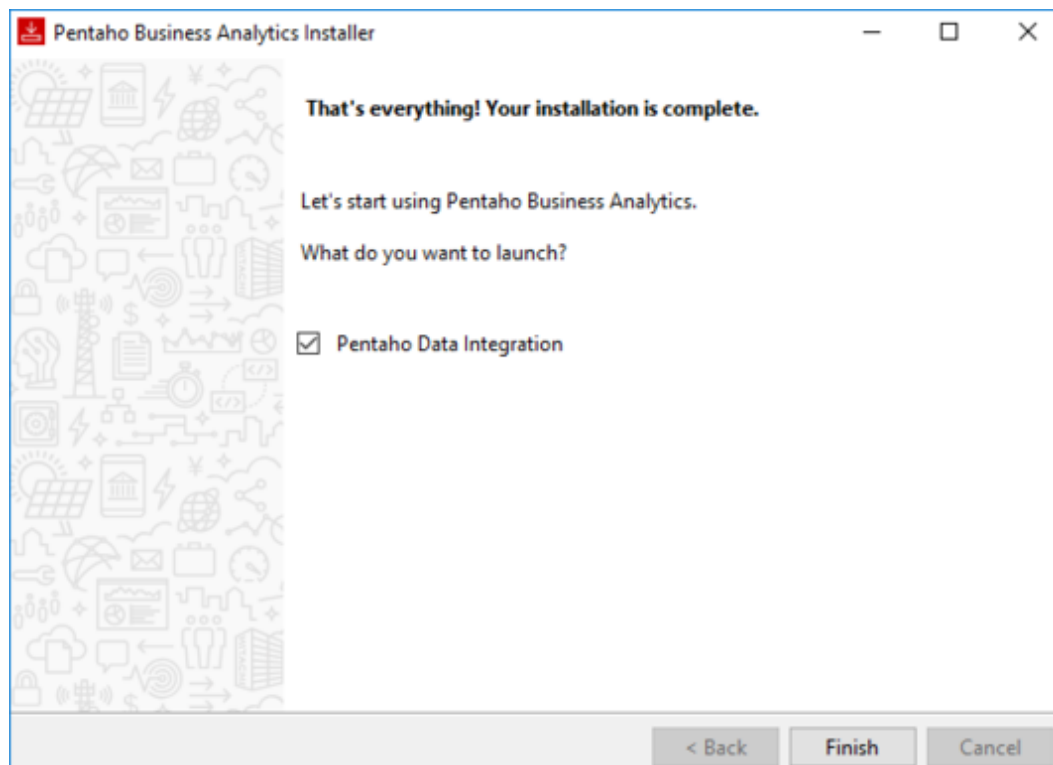
6. Click Next to continue.
7. If the Sample Database window appears, enter a port number in the Port field to use for the database which will house sample content, then click Next to continue.  
Note: The port request only appears if the Pentaho Server default port numbers (8080, 9092, 8443, 9443, and 5432) are not available on your system and if the additional ports that the installer has tried are not available.
8. When the Ready to Install window appears, click the Next button to install the software.

The Installation in Progress window appears.

The software takes approximately 30 minutes to install.

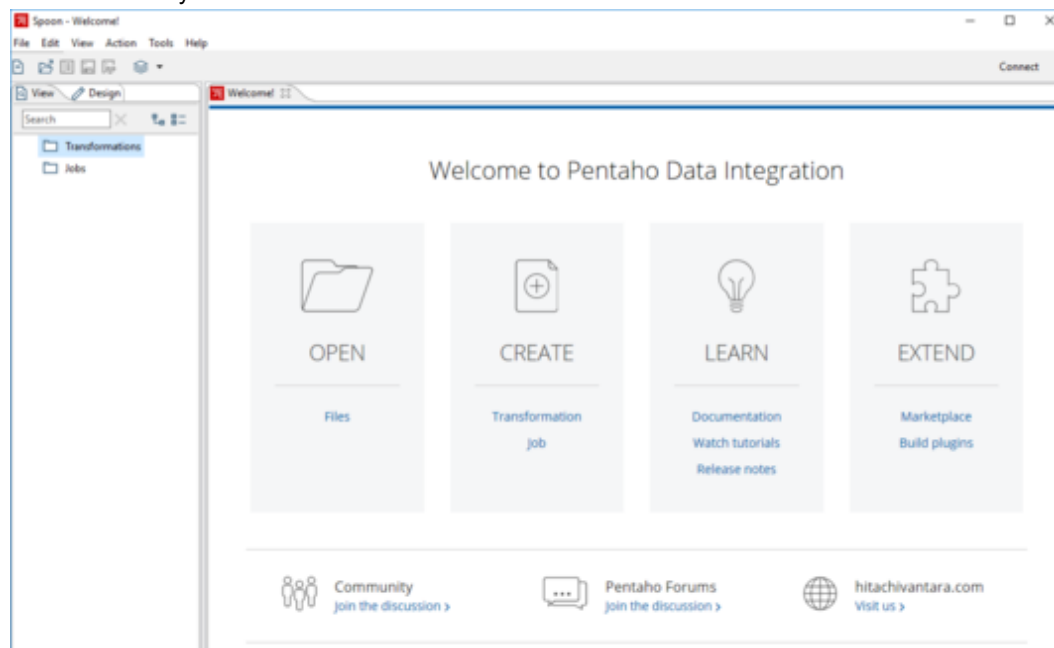
An error occurs if Pentaho Server port numbers (8080, 9092, 8443, 9443, and 5432) are not available on your system and if the additional ports that the installer has tried are not available. If the installer cannot find an available port, you are prompted to enter port numbers you want to use.

9. When installation is complete, the final window appears. Select the Pentaho Data Integration check boxes to launch Pentaho Data Integration (PDI).



10. Click Finish to end the installation process.

After PDI starts you will see the Welcome! window.



To get you started quickly after installation, the [Pentaho Data Integration \(PDI\) tutorial](#) explains how to build a transformation with PDI using sample data in a typical business scenario.

Note: As a best practice, use this installation method for evaluation purposes only. If you have any issues with your 30-day trial, contact the [Pentaho Trial Experts](#) for assistance.

If you have trouble with the installation wizard, see the Administer Pentaho Data Integration and Analytics document for details. After the installation is complete, the Pentaho Server runs with the default parameters. Terminal windows might remain open after installation. If you close them, the Pentaho Server stops. On Windows, the Pentaho Server is installed as a system service, meaning that it starts when you start your computer and stops when you shut down your computer. The User Console, Spoon, and other tools must be manually run. In the next step of these instructions there are links to documentation which explains how to start and stop servers manually.

## Custom installation

If you choose the Let me decide for myself. installation method, complete the following steps. If you want to install using the default method, read the Default Installation instructions.

Note: You cannot install into a directory that already exists. If you attempt to install into a directory that exists, an error message appears. If you are using Linux, the best practice is to not install the software in the /opt/ directory. The complicated permissions requirements and general isolation from the rest of the system might cause problems if the software is installed in this location.

1. After you start the installation wizard, the Pentaho Business Analytics Installation Wizard splash screen appears, then the Pentaho Business Analytics installation is ready! window. Click Next.
2. Read the license agreement. Click Accept, then click Next to continue.
3. Select the Let me decide for myself option in the Setup Type window, and then click Next.
4. In the Pentaho Applications window, select the Pentaho components that you want to install in the Install Set panel. When finished, click Next.
5. If you selected to install the Pentaho Server, then the Sample Content window appears. Sample content includes users, reports, dashboards, and analysis views which can be useful when evaluating the product. Select the Sure, give me sample content option to include this sample content with your installation, or select the No thanks, I'll create my own option to exclude this sample content. When finished, click Next.
6. In the Installation folder window, accept the default directory or choose a different directory by entering the path in the Location text box where you want to install the software. Click Next.
7. If you selected to install the Pentaho Server, you are prompted to add a Postgres SQL password. Enter and confirm the password you want to assign to the postgres user who is assigned admin privileges for the PostgreSQL database.

Note: Do not use these characters in the Password field because the installation wizard cannot process them:

' " & < > \

8. Click Next to continue.
9. If the Sample Database window appears, enter a port number in the Port field to use for the database which will house sample content, then click Next to continue.  
The port request only appears if the Pentaho Server default port numbers (8080, 9092, 8443, 9443, and 5432) are not available on your system and if the additional ports that the installer has tried are not available.
10. When the Ready to Install window appears, click the Next button to install the software.  
The Installation in Progress window appears.

The software takes approximately 30 minutes to install.

Note: An error occurs if Pentaho Server port numbers (8080, 9092, 8443, 9443, and 5432) are not available on your system and if the additional ports that the installer has tried are not available. If the installer cannot find an available port, you are prompted to enter port numbers you want to use.

11. When installation is complete, the final window appears. Select the Pentaho User Console and Pentaho Data Integration check boxes to launch the Pentaho User Console and Pentaho Data Integration (PDI).
12. Click Finish to end the process.

Note: If you have trouble with the installer, see the Administer Pentaho Data Integration and Analytics document for details. After the installation is complete, the Pentaho Server runs with the default parameters. Terminal windows might remain open after installation. If you close them, the Pentaho Server stops. On Windows, the Pentaho Server is installed as a system service, meaning that it starts when you start your computer and stops when you shut down your computer. The User Console, Spoon, and other tools must be manually run. In the next step of these instructions there are links to documentation that explains how to start and stop servers manually.

## Verify installation

To verify that the Pentaho products installed properly, review the software directory structure and the installation summary file.

1. Open a file explorer window such as Windows Explorer or a command line tool such as the Command Prompt or Terminal windows.
2. Navigate to the directory where you installed the software, then compare this directory structure with yours. If you installed the all components, the structure should look like this.

```
pentaho/
pentaho/design-tools/
pentaho/design-tools/aggregation-designer/
pentaho/design-tools/data-integration/
pentaho/design-tools/metadata-editor/
pentaho/design-tools/report-designer/
pentaho/design-tools/schema-workbench/
pentaho/documentation/
pentaho/java/
pentaho/jdbc-distribution/
pentaho/license-installer/
pentaho/licenses/pentaho/monetdb/
pentaho/postgresql/
pentaho/scripts/
pentaho/server/
pentaho/server/pentaho-server/
pentaho/server/pentaho-server/licenses/
pentaho/server/pentaho-server/logs/
pentaho/server/pentaho-server/pentaho-solutions/
```

pentaho/server/pentaho-server/third-party-tools/  
 pentaho/server/pentaho-server/tomcat/  
 pentaho/server/pentaho-server/wkhtmltoimage/  
 pentaho/server/hsqldb-sample-database

3. Open the installation-summary.txt file. Make sure that the design tools and plugins that you installed are listed.
4. Review the following locations of files.

Table. Top-level directories and files

File	Description
ctlscrip.sh	Starts, stops, restarts, and shows the status of Pentaho services. Available on Linux and Mac only.
installation-summary.txt	Contains the information from the summary screen at the end of the installation process.
uninstall	A script that removes Pentaho Business Analytics.

Table. Locations of specific programs

Tool/Plugin	Location
Pentaho Server	pentaho/server/pentaho-server/
Report Designer	pentaho/design-tools/report-designer/
Schema Workbench	pentaho/design-tools/schema-workbench/
Data Integration (Spoon)	pentaho/design-tools/data-integration/
Metadata Editor	pentaho/design-tools/metadata-editor/
Aggregation Designer	pentaho/design-tools/aggregation-designer/
Dashboard Designer	pentaho/server/pentaho-server/pentaho-solutions/system/dashboards/
Analyzer	pentaho/server/pentaho-server/pentaho-solutions/system/analyzer/
Interactive Reports	pentaho/server/pentaho-server/pentaho-solutions/system/pentaho-interactive-reporting/
License Installer	pentaho/license-installer/

Table. Location of logs

Log	Location
Pentaho Server Logs for BA configuration	pentaho/server/pentaho-server/logs/
Tomcat Logs for Pentaho Server for BA configuration	pentaho/server/pentaho-server/tomcat/logs/

For Mac OSX, you will have to copy the JDBC .jar into the location listed for Report Designer.

Table. Location of JDBC drivers

JDBC Driver	Location
Pentaho Server for BA configuration	pentaho/server/pentaho-server/tomcat/lib/
Report Designer	pentaho/design-tools/report-designer/lib/jdbc/
Schema Workbench	pentaho/design-tools/schema-workbench/drivers/
Aggregation Designer	pentaho/design-tools/aggregation-designer/drivers/
Metadata Editor	pentaho/design-tools/metadata-editor/libext/JDBC/
PDI client (Spoon)	pentaho/design-tools/data-integration/lib/

Table. Default port numbers

Port Number	Description
5432	PostgreSQL Server
8080	Pentaho Server Tomcat Web Server Startup Port
8012	Pentaho Server Shutdown Port
9001	HSQL Server Port
9092	Embedded H2 Database

Note: Your port numbers might differ if these ports are already used for other programs.

## Uninstalling the Pentaho Suite after evaluation

To uninstall the Pentaho Suite after the evaluation, complete the following steps:

1. Navigate to the pentaho directory, and run the uninstall file.
2. The Uninstall Wizard is launched.
3. When prompted, indicate whether you want to delete the data files.
4. When prompted at the end of the process, restart your computer.

Before installing a production version of Pentaho, you must first uninstall the evaluation version of Pentaho. See *Install Pentaho Data Integration and Analytics* for instructions on installing a production version.

## Getting started tutorial

To get you started quickly after installation, the [Pentaho Data Integration \(PDI\) tutorial](#) explains how to build a transformation with PDI.

## Tutorials

Review these tutorials to start using PDI, Pentaho reporting tools, and dashboards.

If you are new to PDI, start with [Getting Started with PDI](#). This tutorial shows you how to use Spoon, create transformations and jobs, and more.

Within this general tutorial, you can also view the following specific tutorials:

- [PDI Transformation Tutorial](#)
- [PDI Job Tutorial](#)
- [Getting Started with PDI and Hadoop](#)

The [Getting Started with Analyzer, Interactive Reports, and Dashboards](#) tutorial provides an overview of product features and related technologies. In addition, it contains recommendations on best practices, tutorials for getting started, and troubleshooting information for common situations.

Within this general tutorial, you can also view the following specific tutorials:

- [About Pentaho Business Analytics Tools](#)
- [Get Started with Pentaho Reporting Tools](#)
- [Quick Tour of the Pentaho User Console \(PUC\)](#)
- [Get Started with Interactive Reports](#)
- [Get Started with Analyzer Reports](#)
- [Get Started with Dashboard Designer](#)
- [Next Steps](#)

The [Getting Started with Report Designer](#) tutorial contains step-by-step instructions for creating reports in Report Designer.

Within this general tutorial, you can also view the following specific tutorials:

- [About Report Designer](#)

- [Create a Report with Report Designer](#)

# Getting Started with Analyzer, Interactive Reports, and Dashboard Designer

This guide provides an overview of product features and related technologies. In addition, it contains recommendations on best practices, tutorials for getting started, and troubleshooting information for common situations.

- [About Pentaho business analytics tools](#)

Pentaho helps you create reports and dashboards with our easy-to-use interface.

- [Quick tour of the Pentaho User Console \(PUC\)](#)

Take a quick tour of the different perspectives of the Pentaho User Console.

- [Get started with Analyzer Reports](#)

Analyzer is a visualization tool that helps you examine data contained in analysis data sources.

- [Get started with Pentaho Reporting tools](#)

Follow the directions in each tutorial to learn how to create your first Pentaho reports.

- [Get started with Interactive Reports](#)

Interactive Reports is a web-based interface which is used to create on-demand operational reports.

- [Get started with Dashboard Designer](#)

Examine different reports, have quick web access, or view charts and graphs from one screen.

- [Next steps](#)

After you have finished working through the tutorials, you are ready to learn more about Pentaho reporting.

## About Pentaho business analytics tools

The topics found in this section give you an overview of the reports and dashboards you create with the User Console, to help you become familiar with the look and feel of the console.

The Pentaho User Console is a web-based design environment where you can analyze data, create interactive reports, dashboard reports, and build integrated dashboards to share business intelligence solutions with others in your organization and on the internet. In addition to its design features, the User Console offers a wide variety of system administration features for configuring the Pentaho Server, maintaining the Pentaho licenses, setting up security, managing report scheduling, and tailoring system performance to meet your requirements.

## Prerequisites

Before you can work with the User Console, you must have the Pentaho software installed and the Pentaho Server configured. See *Install Pentaho Data Integration and Analytics* for installation and



configuration instructions.

## Expertise

The User Console does not require any special skills or knowledge to use its design environment. However, to use its system administration features you should know where your data is stored and how to access it, as well as details about your system configuration and security providers.

## Tools

Through the User Console you can access the Pentaho Repository on the server, as well as these Pentaho tools and features:

- Analyzer
- Interactive Reports
- Dashboard Designer
- Data Source Wizard
- Data Source Model Editor

## Login credentials

These tasks require that you [log on to the User Console](#) with an evaluator user name and password.

## Quick tour of the Pentaho User Console (PUC)

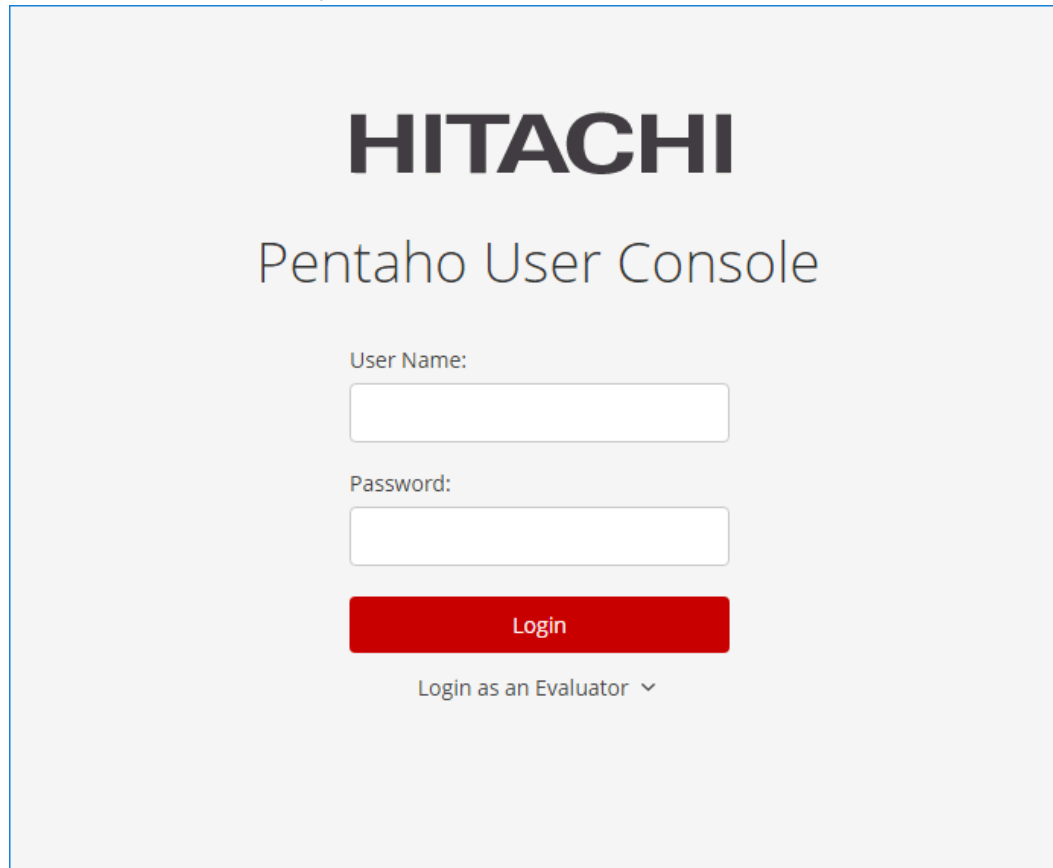
If you use file management tools or any web browser, you should feel right at home with the Pentaho User Console (PUC). To familiarize yourself with the different pages and controls of PUC, let us take you through a quick tour.

## Log in to PUC

Logging in to the PUC for the first time is easy with these steps:

1. Launch any web browser and enter the URL of the Pentaho server.

Your IT administrator can tell you the correct URL.



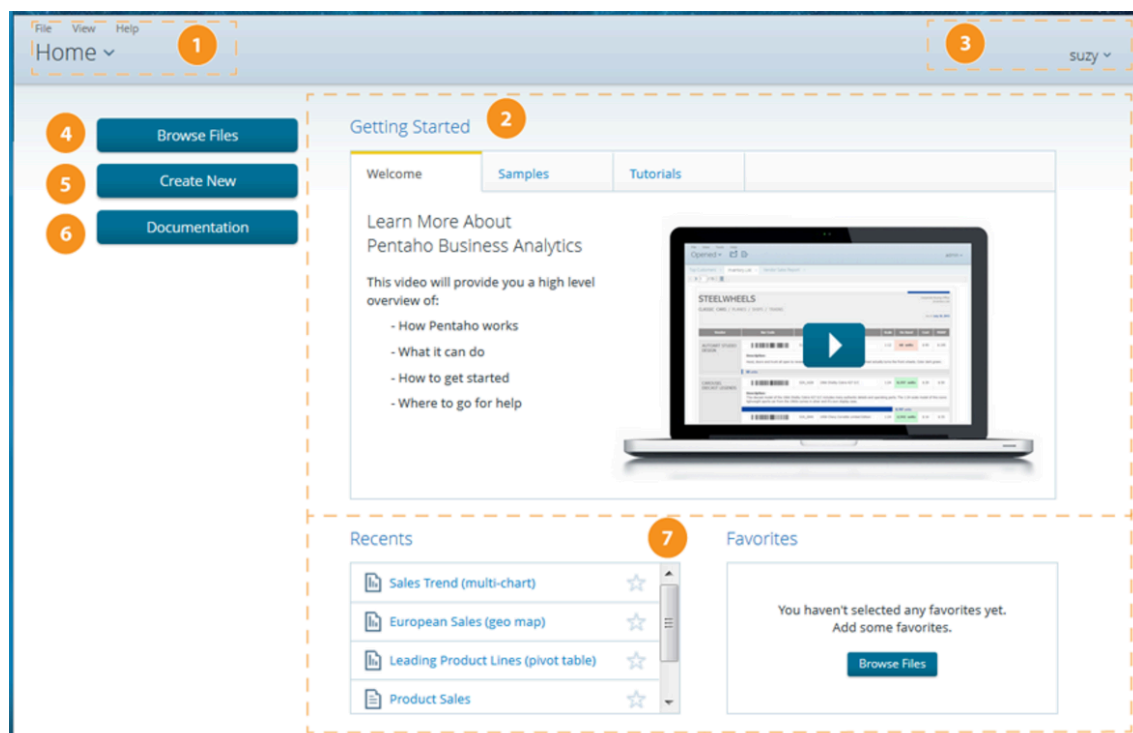
The image shows the Hitachi Pentaho User Console login interface. At the top, the 'HITACHI' logo is displayed in a large, bold, dark font, followed by 'Pentaho User Console' in a smaller, regular font. Below this, there are two input fields: 'User Name:' and 'Password:'. A red 'Login' button is positioned below the password field. At the bottom, there is a link that says 'Login as an Evaluator' with a downward arrow.

The page loads an introductory screen with a Login section.

2. Enter your user name and password and click Login, or use the Login as an Evaluator option.

## Home

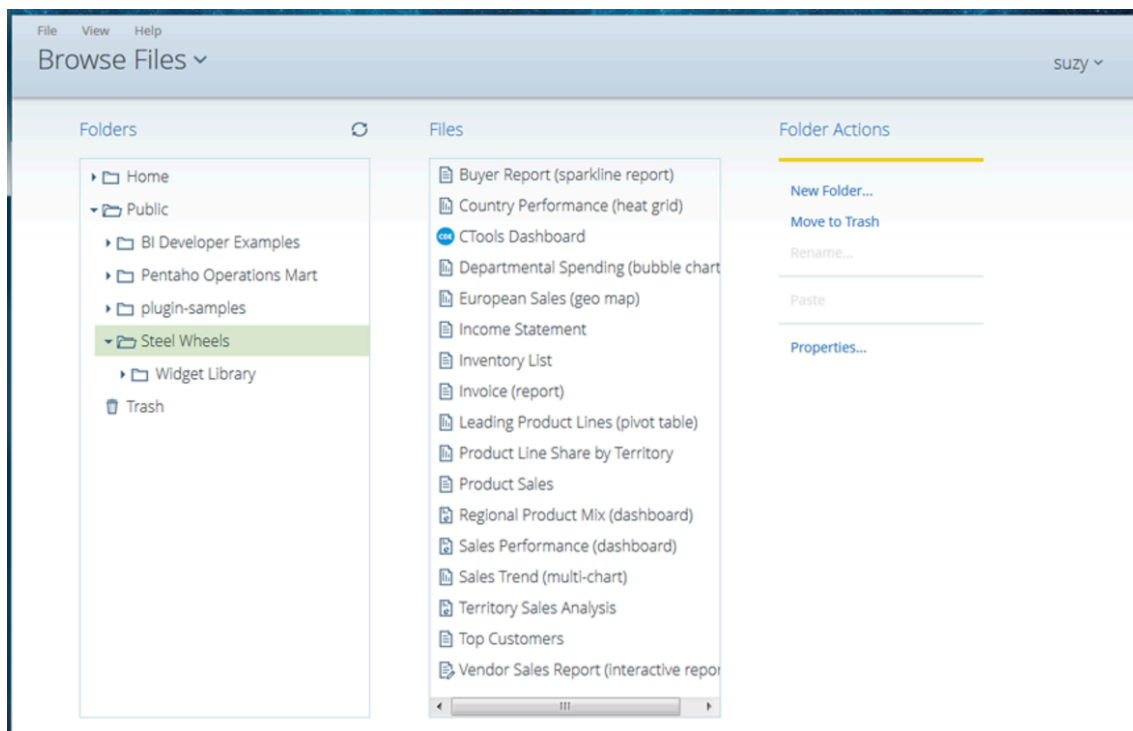
The first thing you see after you log in to the console is the Home page, which serves as the starting place for the User Console and all of the tasks that you do with it.



Item	Name	Function
1	Home view	The Home drop-down menu lets you flip easily from page to page, or return to your Home page.
2	Getting Started	<ul style="list-style-type: none"> <li>Welcome shows an introductory video about Pentaho products.</li> <li>Samples holds a variety of sample reports and dashboards that you can use to get familiar with the software.</li> <li>Tutorials contains a number of tutorial videos that give you a visual tour of the User Console, reports, and dashboards.</li> </ul>
3	Current User, and Log Out	Shows the name of the person currently logged on to the User Console. Clicking the arrow next to the name lets you log out of the User Console.
4	Browse Files	Brings you to the Browse Files page, where you can locate your files using the Browsing and Files panes, and manage them using the Actions pane. Any files that you open appear in a new window.
5	Create New	Gives you choices to create a new Analysis Report, Interactive Report, or Dashboard. You can also create a new Data Source, if you have permissions to work with data sources.
6	Documentation	Leads you to the <a href="#">Pentaho Documentation</a> , which opens in a new window or tab.
7	Recents, and Favorites	Shows a list of your most recently opened files. Clicking on the star next to a recently opened file adds it to your Favorites list.

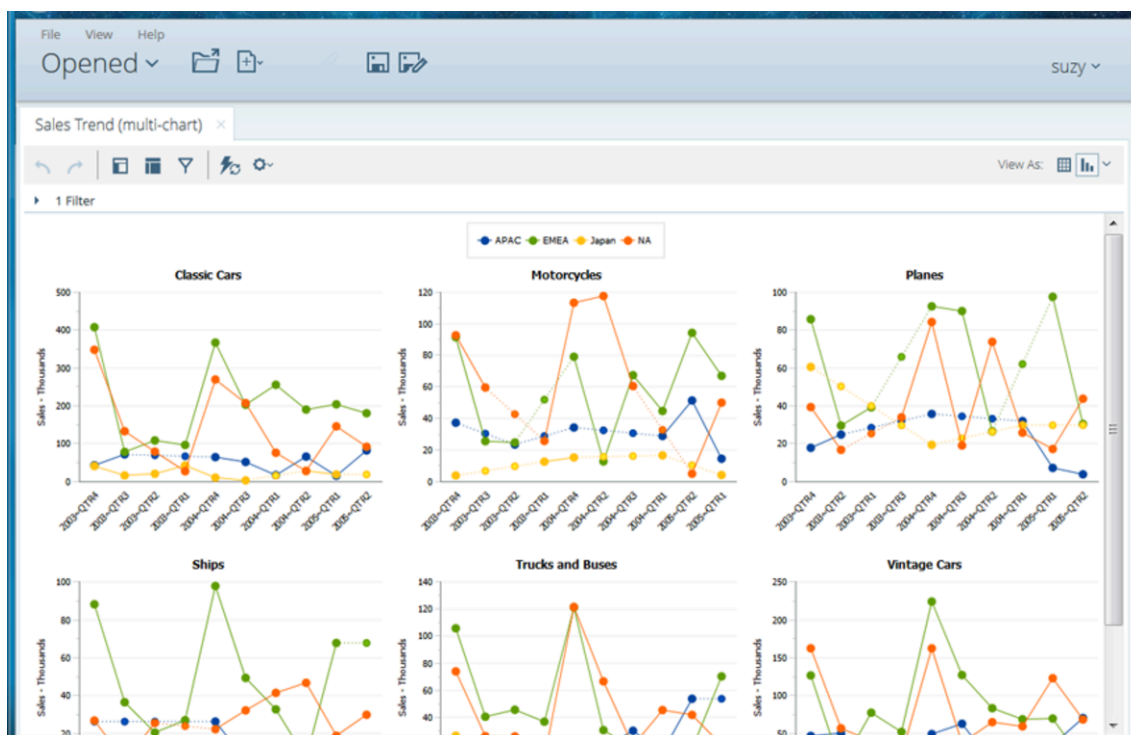
## Browse Files

The Browse Files page helps you keep your files organized and makes them easier for you to find and work with.



## Opened

The Opened page activates after you open a file from the Browse Files page and provides a simple space to work with your files.



### Use Pentaho tools

The Pentaho Business Analytics Suite download includes sample reports and a sample database called Steel Wheels. Steel Wheels is included so that you can quickly use the software and explore the Pentaho BA Suite.

The following table helps you decide what tools to use for your needs and provides links to the corresponding articles.

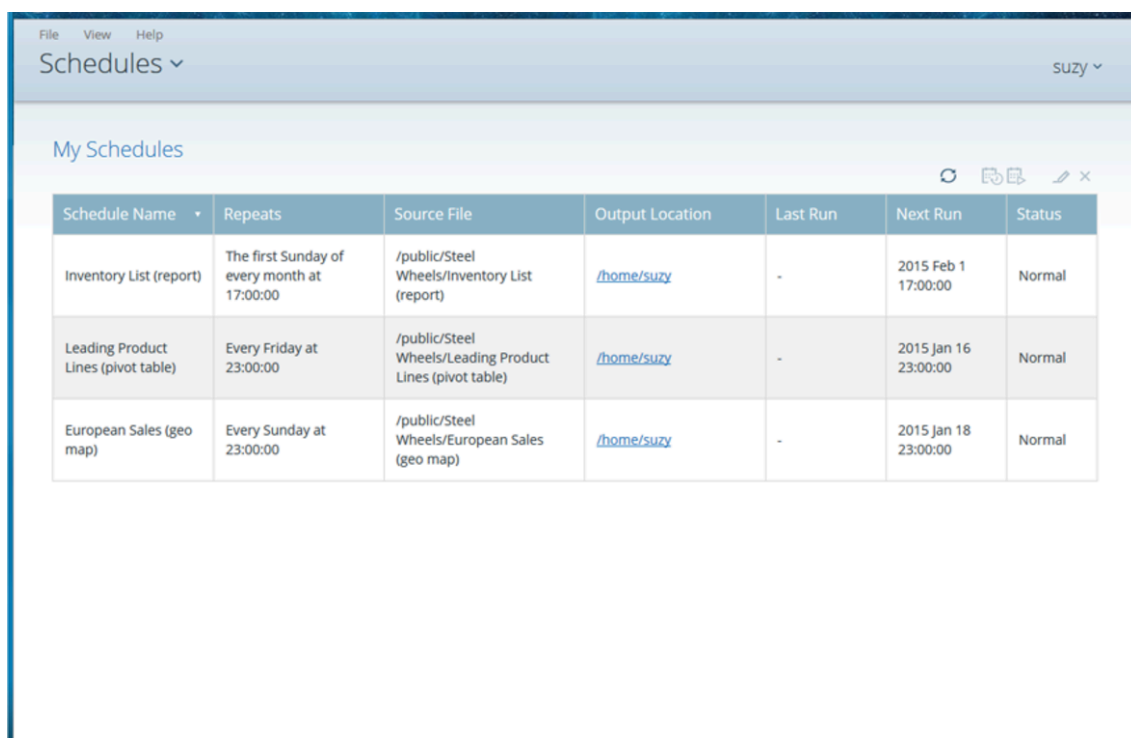
Explore Considerations	Choose Options		
	<a href="#">Interactive Reports</a>	<a href="#">Analyzer Reports</a>	<a href="#">Dashboard Designer</a>
Summary	Interactive Reports is a web-based design interface which is used to create both simple and on-demand operational reports without depending on IT or report developers.	Analyzer Reports is an intuitive analytical visualization tool that filters and drills down into business information contained in Pentaho analysis data sources.	Dashboard Designer allows users to create dashboards with little or no training. The dashboard is several different reports brought together inside one screen.
Expertise	Knowledge of basic computer functions, such as operating systems and web browsers.	Knowledge of basic computer functions, such as operating systems and web browsers.	Knowledge of basic computer functions, such as operating systems and web browsers.
Recommendation	Use Interactive Reports if you want to create a tabular report that answers an immediate business question, looks professional, can be printed quickly, and provides significant control over formatting elements	Use Analyzer Reports if you want to compile data quickly while visually exploring your data, perform advanced sorting and filtering of your data, and want to see chart	Use Dashboard Designer if you want to create an interface to view many different reports at once, have quick access to web pages that you visit often, or view dynamic charts and graphs within a

Explore Considerations	Choose Options		
	<a href="#">Interactive Reports</a>	<a href="#">Analyzer Reports</a>	<a href="#">Dashboard Designer</a>
	such as fonts, column width or sorting, background colors, and more.	visualizations that include detailed stop-lighting.	space while you create reports in another.

## Schedules

All of your active scheduled reports appear in the listing on the Schedules page, which you can get to by clicking the Home drop-down menu, then the Schedules link, in the upper-left corner of the PUC page.

My Schedules lists the reports scheduled to run, the recurrence pattern for the schedule, when it was last run, when it is set to run again, and the current state of the schedule. You can edit and maintain each of your schedules by using the controls above the schedules list, on the right end of the toolbar. See the *Pentaho Analytics* document for details.



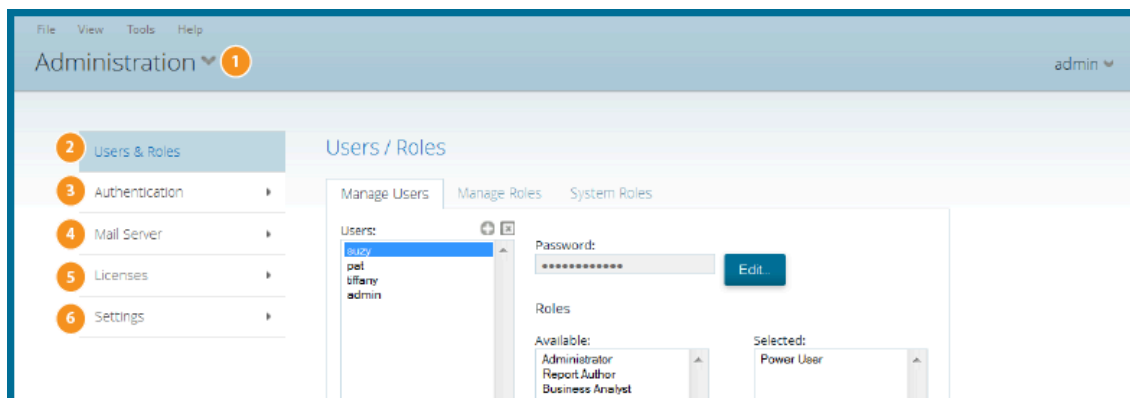
The screenshot shows the 'Schedules' page in the Pentaho user interface. At the top, there is a header bar with 'File', 'View', and 'Help' menus, and a user profile 'suzy'. Below the header, the page title 'Schedules' is displayed with a dropdown arrow. The main content area is titled 'My Schedules' and contains a table with the following data:

Schedule Name	Repeats	Source File	Output Location	Last Run	Next Run	Status
Inventory List (report)	The first Sunday of every month at 17:00:00	/public/Steel Wheels/Inventory List (report)	<a href="#">/home/suzy</a>	-	2015 Feb 1 17:00:00	Normal
Leading Product Lines (pivot table)	Every Friday at 23:00:00	/public/Steel Wheels/Leading Product Lines (pivot table)	<a href="#">/home/suzy</a>	-	2015 Jan 16 23:00:00	Normal
European Sales (geo map)	Every Sunday at 23:00:00	/public/Steel Wheels/European Sales (geo map)	<a href="#">/home/suzy</a>	-	2015 Jan 18 23:00:00	Normal

On the right side of the table, there is a toolbar with icons for refreshing, deleting, and other actions.

## Administration

In PUC, the Administration page allows logged-on users who have an assigned role that includes Administer Security permissions to perform system configuration and maintenance tasks. If you see Administration in the left drop-down menu on the Home page, you can click it to reveal menu items specific to administration of the Pentaho Server. If you do not have administration privileges, Administration does not appear within the console.



Item	Control Name	Function
1	Administration	In PUC, open the Administration page. The Administration page enables you to set up users, configure the mail server, change authentication settings on the Pentaho Server, and install software licenses for Pentaho.
2	Users & Roles	Manage the Pentaho users and roles for the Pentaho Server. See the <i>Pentaho Analytics</i> document for details.
3	Authentication	Set the security provider for the Pentaho Server to either the default Pentaho Security or LDAP/Active Directory. See the <i>Administer Pentaho Data Integration and Analytics</i> document for details.
4	Mail Server	Set up the outgoing email server and the account used to send reports through email.
5	Licenses	Manage Pentaho software licenses. See the <i>Administer Pentaho Data Integration and Analytics</i> document for details.
6	Settings	Manage settings for deleting older generated files, either manually or by creating a schedule for deletion.

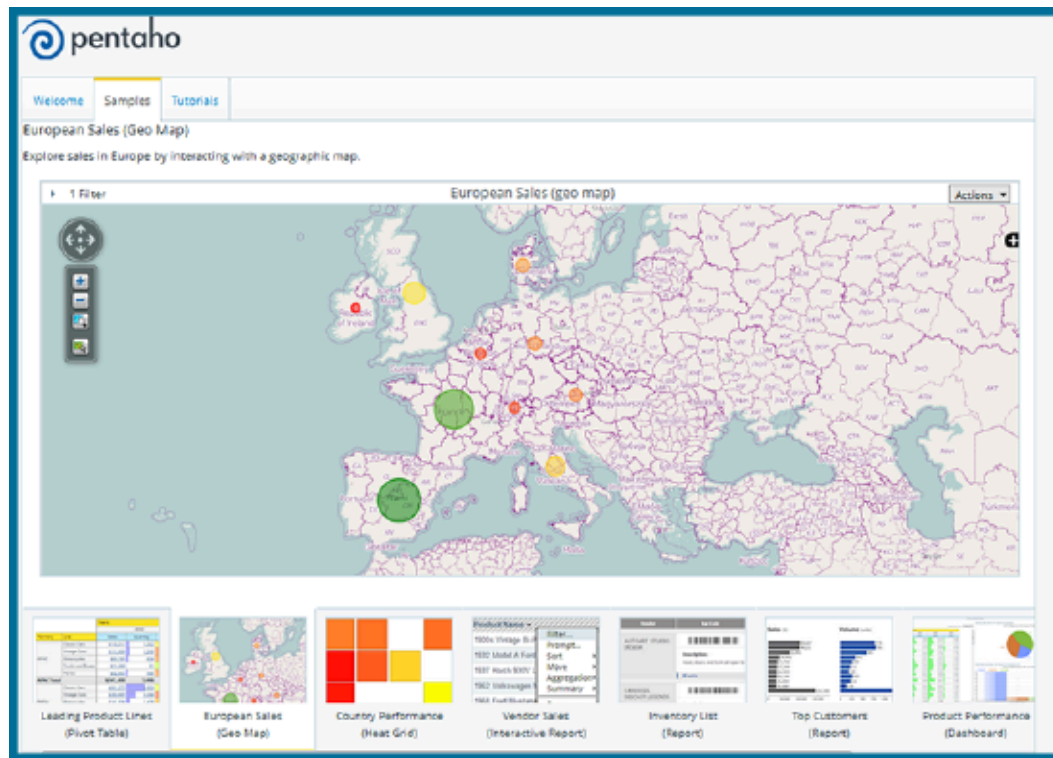
## Get started with Analyzer Reports

Analyzer Reports is an intuitive analytical visualization tool that filters and drills down into business information contained in Pentaho analysis data sources. Use Analyzer Reports if you want to compile data quickly in an interactive environment, perform advanced sorting and filtering of your data, and want to see chart visualizations that include conditional stop-lighting.

## View an Analyzer report sample

This section highlights some popular Analyzer capabilities that are available, using the sample report called European Sales located in the Getting Started widget.

1. In the Getting Started widget on the Home page, click the Samples tab.
2. In the scrolling panel, scroll down and click European Sales and then click Explore.

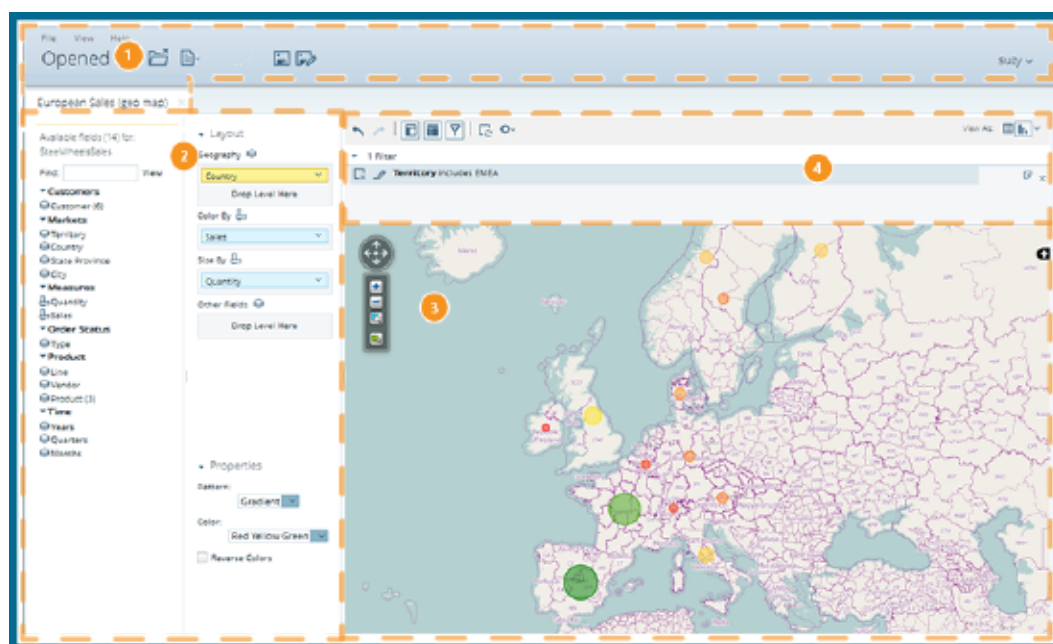


A new browser window will open. Click the Samples tab to see the report.

## Tour the Analyzer panels

You can view the editable version of European Sales in Analyzer by going to the Browse Files page in the User Console. Follow these steps:

1. From the User Console Home page, click the Browse Files button. In the Browsing pane, click to expand the Public folder, then click to expand the Steel Wheels folder.
2. In the center pane, double-click on European Sales.



The Opened page appears with the Analyzer report.

3. Click the Add More Fields and Rearrange Fields icons on the toolbar to expand the Available Fields and Layout panels.

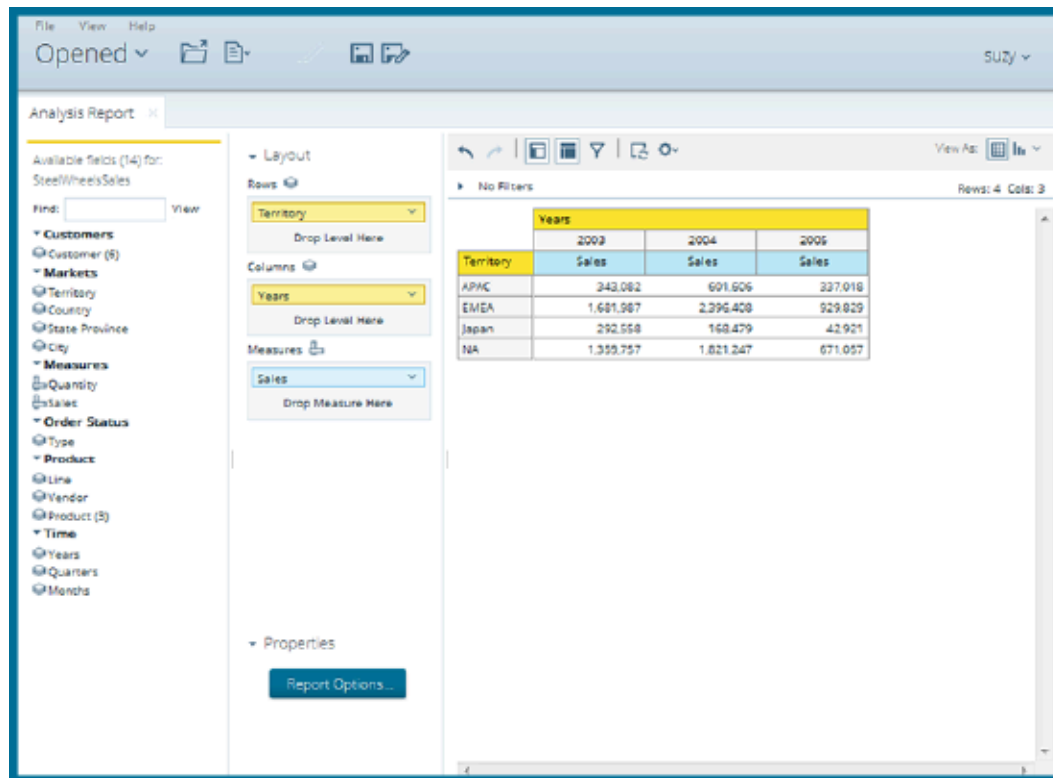
Item	Name	Function
1	Opened page	Displays quick access buttons across the top to create and save a new Analysis Report, Interactive Report, and Dashboard. Opened reports and files show as a series of tabs across the page.
2	Available Fields and Layout panels	<p>Use the Available Fields and Layout panels to drag levels and measures into a report.</p> <p>Your report displays changes in the report canvas as you drag items onto the Layout panel.</p> <p>Delete a level or measure from your report by dragging it from the Layout panel toward the lower right corner of the report canvas. As you drag it, a trash can icon appears in the lower right corner of the canvas.</p>
3	Report canvas	<p>Shows a dynamic view of your report as you work to build it. The look of your report changes constantly as you work with Available Fields and Layout panels to refine it.</p> <p>The report canvas shows different fields based on the chart type selected.</p>
4	Analyzer toolbar and filters	<p>Use the toolbar functions to undo or redo actions, hide lists of fields, add or hide filters, disable the auto-refresh function, adjust settings, and change the view of your report.</p> <p>Use the Filters panel to display a list of filters applied to the active report, or edit or delete filters.</p>

## Create your first Analyzer report

The instructions below guide you through the creation of your first Analyzer report, using the Steel Wheels sample data.

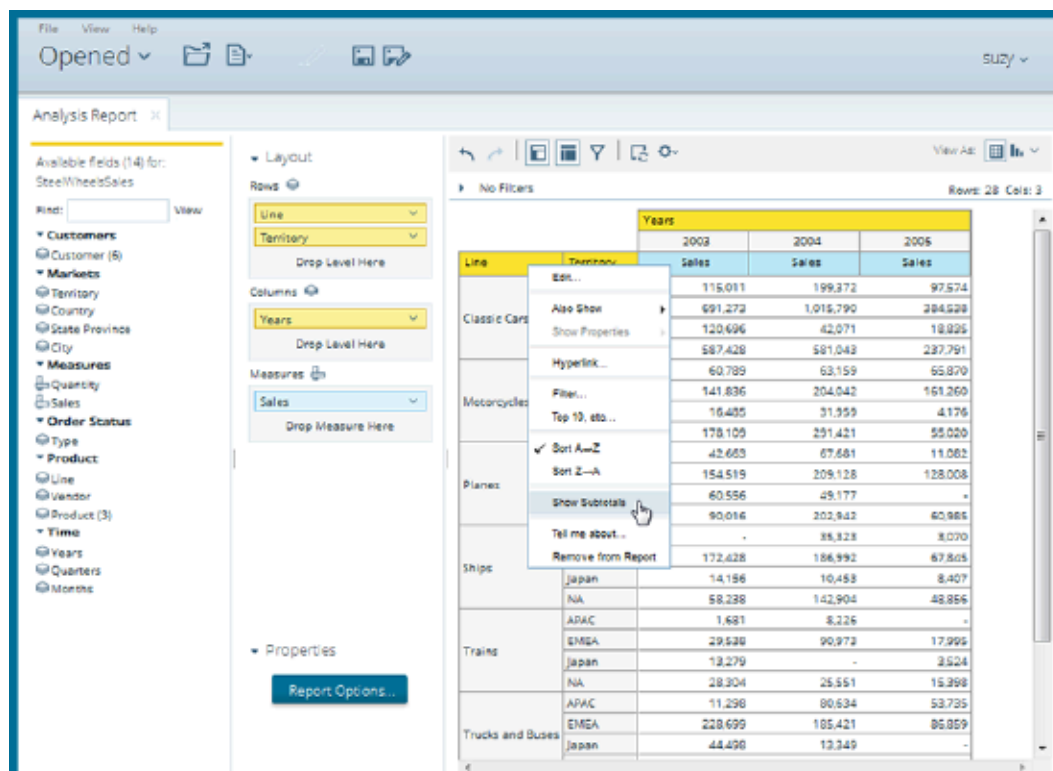
1. From the User Console Home page, click Create New, then choose Analysis Report.
2. Choose the SteelWheels:SteelWheelsSales data source from the Select Data Source dialog box. Click OK.  
A blank Analyzer report appears.
3. Click and drag the Territory element and drop it into the Rows field. Click and drag the Years element and drop it into the Columns field. Click and drag the Sales element into the Measures field.





A pivot table with the Territory, Years, and Sales data appears and populates with the information from the server.

- Click and drag the Line field and drop it above Territory in the Layout column on the canvas. Right-click the Line column header and select Show Subtotals from the menu that appears.



- Right-click the first Sales column and select Conditional Formatting > Data Bar - Green in the menu that appears.
- Right-click the same Sales column and select User Defined Measure > % of Rank, Running Sum. In the dialog box that appears, select the % of Sales radio button. Click Next.

New % of, Rank, Running Sum, etc.

Choose the measure field that you want to create:

☒ % of Sales

☐ Rank by Sales

☐ Running Sum of Sales

☐ % of Running Sum of Sales

Next Cancel

7. Select the Each Line Column/Row Subtotal (Subtotal is 100%) radio button. Click Done.

New % of, Rank, Running Sum, etc.

Name: % of Sales

Format: Percentage (%) Decimal Places: 2

Base Measure: Sales

% of Sales

☐ Column (Grand Total Column is 100%)

☐ Row (Grand Total Row is 100%)

☐ Grand Total (Table Grand Total is 100%)

☒ Each Line Column/Row Subtotal (Subtotal is 100%)

Back Done Cancel

8. Expand the filters canvas by clicking the Show Filters icon. Click and drag the Territory field from the Available Fields panel into the filter canvas.
- The Filter on Territory dialog box appears.
9. In the Filter on Territory dialog box, select APAC from the list and click the top, right-pointing arrow to move it to the box on the right.
10. Enable Parameter Name by clicking on the check box in the bottom left of the dialog box. Type region as your parameter name in the text box. Click OK.

### Filter on Territory

☒ Select from a list ([Includes](#) [Excludes](#))  
☐ Match a specific string ([Contains](#) [Doesn't Contain](#))

Choose values from list: Currently **Included** ▼ :

APAC  
EMEA  
Japan  
NA

>  
<  
>>  
<<

Showing all 4 values

☒ **region** Parameter Name

✓ APAC

1 value selected

The report updates and displays sales data for APAC exclusively.

- Click the line that separates each column to adjust it for better viewing. Move the line right or left as needed.

File View Help

Opened ▾

Summary ▾

Territory - Sales

Available fields (14) for StepViewSales

Find:  View

☒ Customers  
☒ Territory  
☒ Country  
☒ State/Province  
☒ City  
☒ Measures  
☒ Quantity  
☒ Sales  
☒ Order Status  
☒ Type  
☒ Product  
☒ Line  
☒ Vendor  
☒ Product (2)  
☒ Time  
☒ Years  
☒ Quarters  
☒ Months

Layout

Rows

Line  
Territory  
Drop Level Here

Columns

Years  
Drop Level Here

Measures

Sales  
% of Sales  
Drop Measure Here

Properties

Report Options

1 Filter

Territory includes APAC

Rows: 7 Cols: 6

Line	Territory	Years					
		2003		2004		2005	
		Sales	% of Sales	Sales	% of Sales	Sales	% of Sales
Classic Cars	APAC	115,011	100.00%	199,372	100.00%	90,574	
<b>Classic Cars Total</b>		<b>115,011</b>	<b>100.00%</b>	<b>199,372</b>	<b>100.00%</b>	<b>90,574</b>	
Motorcycles	APAC	60,789	100.00%	63,159	100.00%	65,870	
<b>Motorcycles Total</b>		<b>60,789</b>	<b>100.00%</b>	<b>63,159</b>	<b>100.00%</b>	<b>65,870</b>	
Planes	APAC	42,663	100.00%	67,681	100.00%	11,082	
<b>Planes Total</b>		<b>42,663</b>	<b>100.00%</b>	<b>67,681</b>	<b>100.00%</b>	<b>11,082</b>	
Ships	APAC	-		35,323	100.00%	3,070	
<b>Ships Total</b>		<b>-</b>		<b>35,323</b>	<b>100.00%</b>	<b>3,070</b>	
Trains	APAC	1,681	100.00%	8,226	100.00%	-	
<b>Trains Total</b>		<b>1,681</b>	<b>100.00%</b>	<b>8,226</b>	<b>100.00%</b>	<b>-</b>	
Trucks and Buses	APAC	11,298	100.00%	80,626	100.00%	53,735	
<b>Trucks and Buses Total</b>		<b>11,298</b>	<b>100.00%</b>	<b>80,626</b>	<b>100.00%</b>	<b>53,735</b>	
Vintage Cars	APAC	111,429	100.00%	147,212	100.00%	105,688	
<b>Vintage Cars Total</b>		<b>111,429</b>	<b>100.00%</b>	<b>147,212</b>	<b>100.00%</b>	<b>105,688</b>	

- Click the Save As icon in the toolbar. When the Save As dialog box appears, save your report as Territory - Sales in your user folder (be sure you can find this folder again later) and then click Save.

You have successfully created a simple Analyzer report from scratch. You may export the report to one of several different formats or close the report. See *Pentaho Analytics* for details on how to work with complex Analyzer reports.

## Get started with Interactive Reports

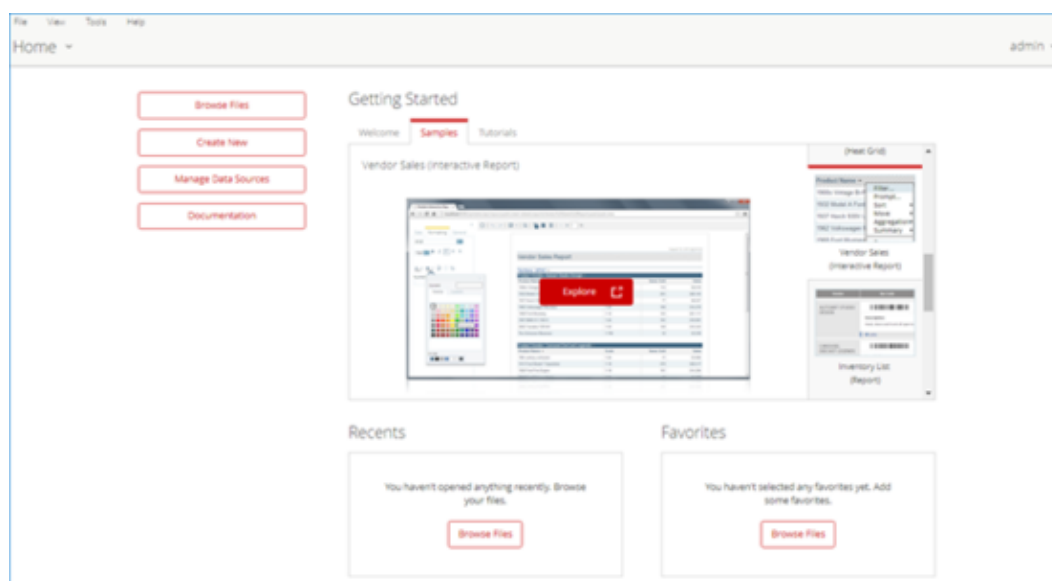
Interactive Reports is a web-based design interface which is used to create both simple and on-demand operational reports without depending on IT or report developers. Use Interactive Reports if you want to create a quick report that answers an immediate business question, looks professional, and provides significant control over formatting elements such as fonts, column width or sorting, background colors, and more.

If you want to start creating your first report right away, feel free to jump ahead to [Create your first Interactive report](#).

## View a Interactive report sample

This section highlights some popular Interactive Reports capabilities that are available, using the sample report called Vendor Sales Report, located in the Getting Started widget.

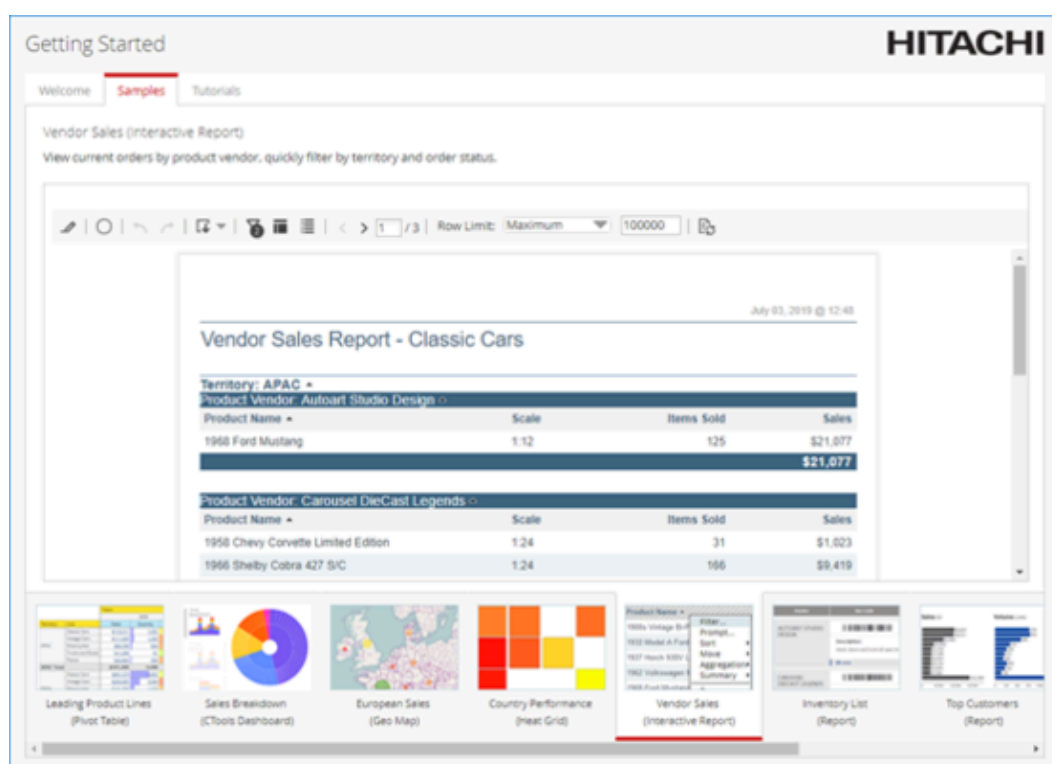
1. In the Getting Started widget on the Home page, click the Samples tab.



2. Click Vendor Sales from the scrolling panel on the right.

3. Click Explore in the Samples pane.

A new window opens showing the Vendor Sales sample report.

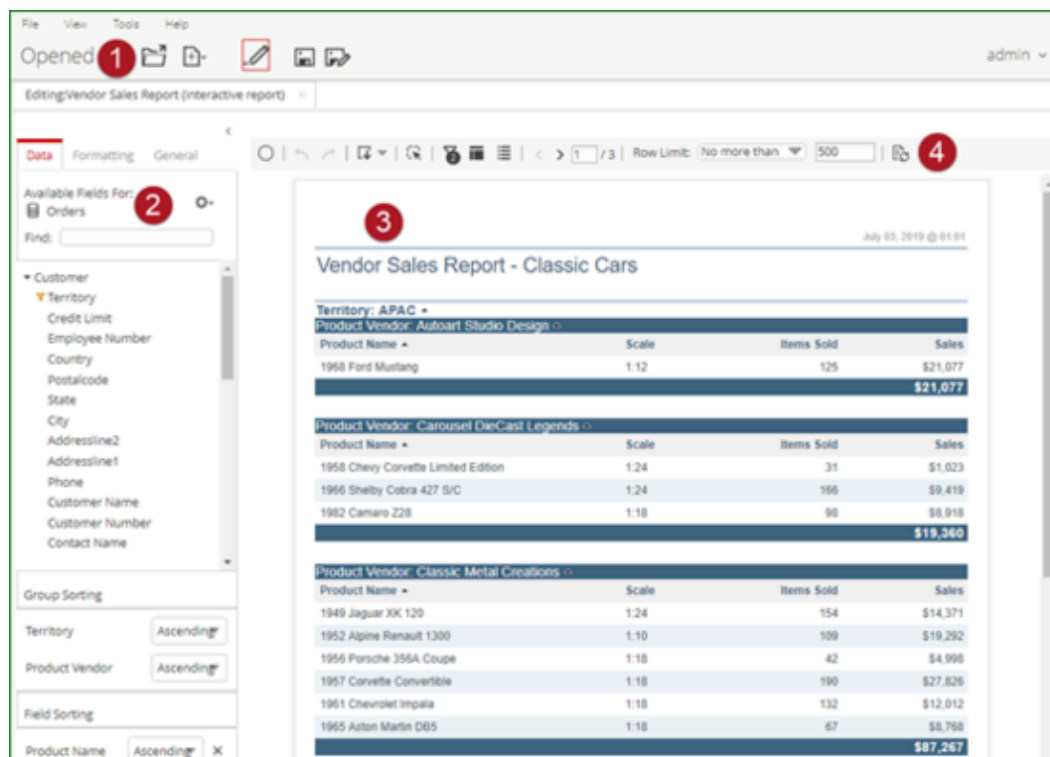


## Tour the Interactive panels

By going to the Browse Files page in the User Console, you can also view an editable version of the Vendor Sales report.

1. Switch to the Browse Files page in the User Console.
2. In the Folders pane, click to expand the Public folder, then click to highlight the Steel Wheels folder.
3. In the Files pane, click on Vendor Sales, and then click Edit in File Actions.

The Opened page appears with the interactive report and toolbars active.



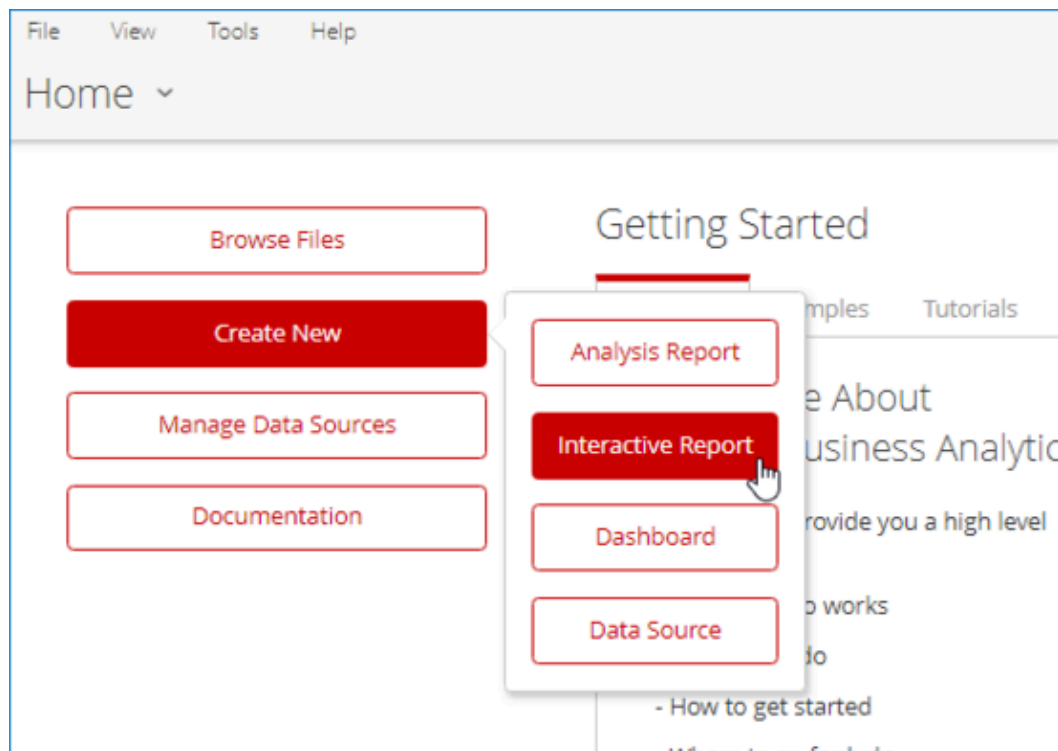
Item	Name	Function
1	Opened page	Displays quick access buttons across the top to create and save a new Analysis Report, Interactive Report, and Dashboard. Opened reports and files show as a series of tabs across the page.
2	Data, Formatting, and General panels	<p>Use the Data panel to drag information into a column or a row on the report. Your report display changes as you drag items onto the report canvas. Use Find to search for a specific field.</p> <p>Delete a level or measure from your report by dragging it from the Layout panel to the trash can, which appears in the lower right corner of the report canvas.</p> <p>The Formatting panel allows you to change the font size and type on the opened report.</p> <p>The General panel allows you to set preferences, select a paper size for printing, and select from a variety of templates for your report.</p>

Item	Name	Function
3	Report canvas	Shows a dynamic view of your report as you work to build it. The look of your report changes constantly as you work with Data, Formatting, and General panels to refine it.
4	Interactive toolbar and filters	Use the toolbar functions to undo or redo actions, hide lists of fields, add or hide filters, disable the auto-refresh function, adjust settings, change the view of your report, and limit the number of rows queried. Use the Filters panel to display a list of filters applied to the active report, or edit or delete filters.

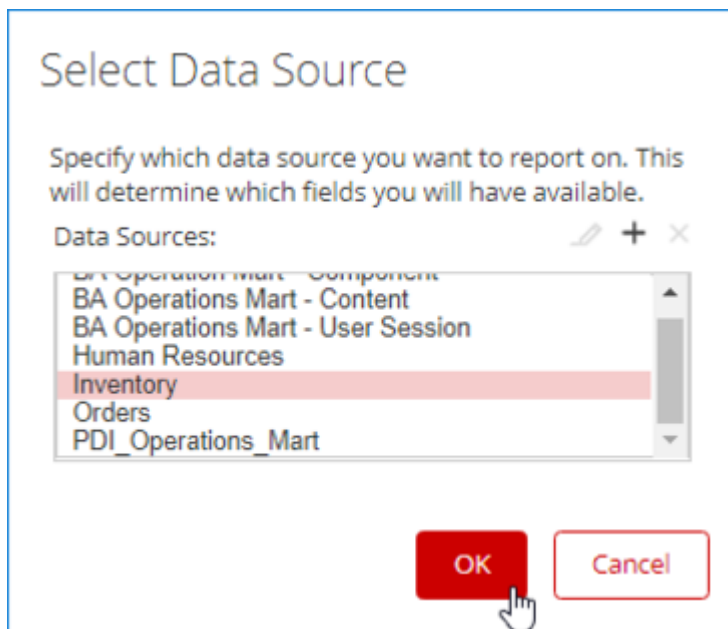
## Create your first Interactive report

The instructions below guide you through the creation of your first Interactive report using the Steel Wheels sample data.

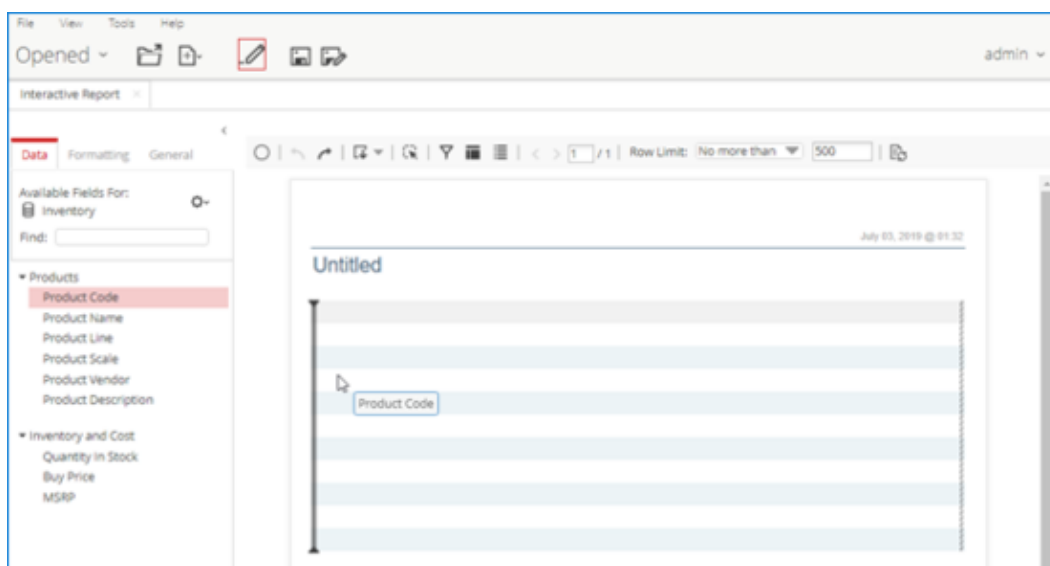
1. From the Home page, click Create New, then choose Interactive Report.



2. Choose the Inventory data source from the Select Data Source dialog box. Click OK.



3. Click Get Started on the dialog box that appears.  
A blank Interactive report canvas appears.
4. Click and drag the Product Code element onto the report canvas until a highlighted vertical line appears. Drop it onto the report canvas.



5. Continue dragging and dropping these fields onto the canvas: Product Name, Product Vendor, Quantity in Stock, MSRP, and Buy Price.

Interactive Report

Data Formatting General

Available Fields For: Inventory

Find:

▼ Products

- Product Code
- Product Name
- Product Line
- Product Scale
- Product Vendor
- Product Description

▼ Inventory and Cost

- Quantity in Stock
- Buy Price
- MSRP

Group Sorting

Field Sorting

July 08, 2019 @ 02:32

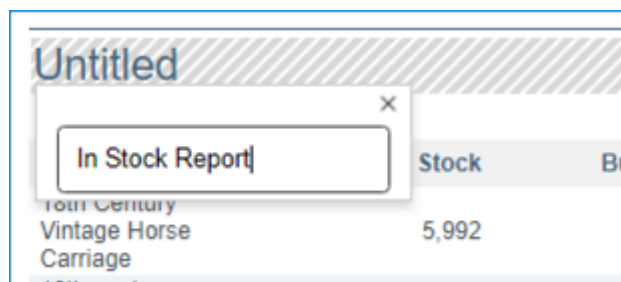
Untitled

Product Code	Product Name	Product Vendor	Quantity in Stock	MSRP	Buy Price
S10_1678	1969 Harley Davidson Ultimate Chopper	Min Lin Diecast	7,933	\$96.00	\$49.00
S10_1949	1952 Alpine Renault 1300	Classic Metal Creations	7,305	\$214.00	\$99.00
S10_2016	1996 Moto Guzzi 1100i	Highway 66 Mini Classics	6,625	\$119.00	\$69.00
S10_4698	2003 Harley-Davidson Eagle Drag Bike	Red Start Diecast	5,582	\$194.00	\$91.00
S10_4757	1972 Alfa Romeo GTA	Motor City Art Classics	3,252	\$136.00	\$86.00
S10_4962	1962 Lancia Delta 16V	Second Gear Diecast	6,791	\$148.00	\$103.00
S12_1099	1968 Ford Mustang	Autoart Studio Design	68	\$195.00	\$95.00
S12_1108	2001 Ferrari Enzo	Second Gear Diecast	3,619	\$208.00	\$96.00
S12_1666	1958 Setra Bus	Welly Diecast Productions	1,579	\$137.00	\$78.00
S12_2823	2002 Suzuki XREO	Unimax Art Galleries	9,997	\$151.00	\$66.00
S12_3148	1969 Convar Monza	Welly Diecast Productions	6,906	\$151.00	\$89.00
S12_3380	1968 Dodge Charger	Welly Diecast Productions	9,123	\$117.00	\$75.00
S12_3380	1968 Dodge Charger	Second Gear	9,123	\$117.00	\$75.00

The data from the chosen fields appears on the report canvas and populates with the information from the server.

Note: You can change the order of the columns by clicking the column heads and dragging the columns left or right until they are in the order you want. If you want to delete a column, just drag and drop the column title into the trash can.

- Rename your report by double-clicking on Untitled in the report canvas, and typing a name in the field that appears. In **Stock Report** is used in this example.



- After you have arranged your columns in order, apply a filter to the data by clicking on the Filter icon in the toolbar. After the Filter pane expands, drag the Product Code field onto the filter workspace.



The screenshot shows the Pentaho Data Integration and Analytics interface. The 'In Stock Report' table is displayed with the following data:

Product Code	Product Name	Product Vendor	Quantity in Stock	MSRP	Buy Price
S10_1678	1969 Harley Davidson Ultimate Chopper	Min Lin Diecast	7,933	\$96.00	\$49.00
S10_1949	1952 Alpine Renault 1300	Classic Metal Creations	7,305	\$214.00	\$99.00
S10_2016	1996 Moto Guzzi 1100i	Highway 66 Mini Classics	6,625	\$119.00	\$69.00
S10_4698	2003 Harley-Davidson Eagle Drag Bike	Red Start Diecast	5,582	\$194.00	\$91.00
S10_4757	1972 Alfa Romeo GTA	Motor City Art Classics	3,252	\$136.00	\$86.00
S10_4962	1962 Lancia Delta 16V	Second Gear Diecast	6,791	\$148.00	\$103.00
S12_1099	1968 Ford Mustang	Autoart Studio Design	68	\$195.00	\$95.00
S12_1108	2001 Ferrari Enzo	Second Gear Diecast	3,619	\$208.00	\$96.00
S12_1666	1958 Setra Bus	Welly Diecast Productions	1,579	\$137.00	\$78.00
S12_1673	1969 Buick Wildcat	Unimax Art	6,605	\$163.00	\$88.00

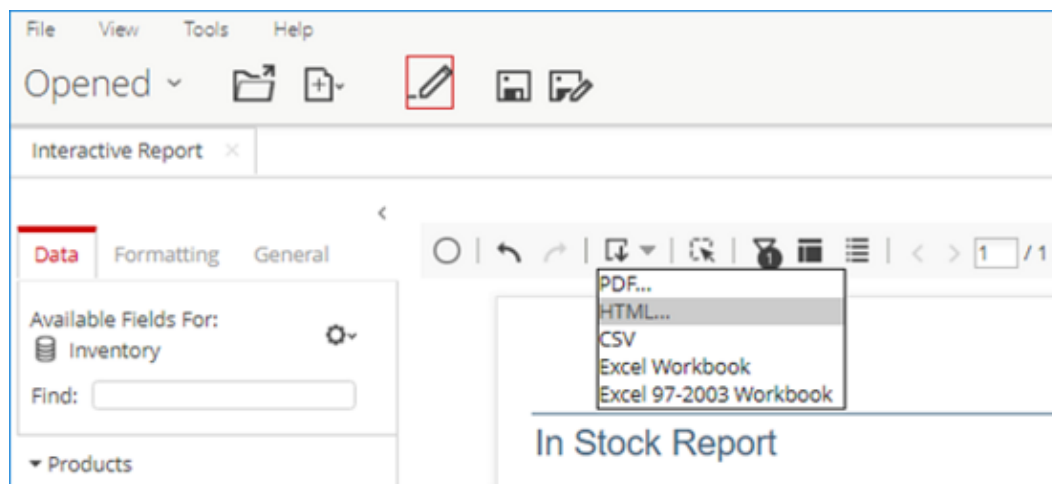
8. In the Filter on dialog box, click Select from a list.

The 'Filter on Product Code' dialog box is shown. The 'Select from a list (Includes, Excludes)' option is selected. The 'Parameter Name' field is empty. The 'Choose values from list' field contains the following values: S10\_1678, S10\_1949, S10\_2016, S10\_4698, S10\_4757, S10\_4962, and S12\_1099. The 'Currently:' dropdown is set to 'Included'. The 'Find' button is highlighted.

9. Choose items from the filter list by using one of these methods. Click on the arrows to move your selected filters on or off the filter list.

- o To choose more than one item from the filter list, hold down the Ctrl key and click the items that you want to filter by, then click the top arrow to move the items to the right panel.
- o Alternatively, you can choose a series from the list by holding down the Shift key and clicking on the first item you want to use, then clicking on the last item that you want to use.

- o Choose individual items by clicking to highlight them, and then clicking on the top arrow to move it to the right panel.
- 10. Click Ok, then click the Save As button in the toolbar.
  - a. When the Save As dialog box appears, save your report using the title that you used (in Step 6) for your report. "In Stock Report" is used in this example.
  - b. Choose your user folder as the location. Be sure to remember this folder location and report title, as the report will be used in a later tutorial. Click Save.
- 11. If you want to export the report, click the Export icon on the toolbar and chose a format from the dropdown list for the export.



The report exports in the selected format and a paper copy can be printed from the export.

Product Code	Product Name	Product Vendor	Quantity In Stock	MSRP	Buy Price
S10_1678	1969 Harley Davidson Ultimate Chopper	Min Lin Diecast	7,933	\$96.00	\$49.00
S10_1949	1952 Alpine Renault 1300	Classic Metal Creations	7,305	\$214.00	\$99.00
S10_2016	1996 Moto Guzzi 1100i	Highway 66 Mini Classics	6,625	\$119.00	\$69.00
S10_4698	2003 Harley-Davidson Eagle Drag Bike	Red Start Diecast	5,582	\$194.00	\$91.00
S10_4757	1972 Alfa Romeo GTA	Motor City Art Classics	3,252	\$136.00	\$86.00
S10_4962	1962 LanciaA Delta 16V	Second Gear Diecast	6,791	\$148.00	\$103.00
S12_1099	1968 Ford Mustang	Autoart Studio Design	68	\$195.00	\$95.00

You have successfully created a simple Interactive report from scratch. See *Pentaho Analytics* for details on how to work with more complex interactive reports.

## Get started with Pentaho Reporting tools

After you define the data sources for your Pentaho Server, you are ready to begin working with the Pentaho User Console to create your first reports. To create a report, just follow the directions in each tutorial. Each section uses the sample data sources that are included with the installation of the Pentaho Suite.

It is time to take a look at the User Console and learn how to create reports. Just follow the topics listed below as you work with the tutorials to create your first Pentaho Reports and Dashboards.

- [Quick tour of the Pentaho User Console \(PUC\)](#)
- [Get started with Interactive Reports](#)
- [Get started with Analyzer Reports](#)
- [Get started with Dashboard Designer](#)
- [Next steps](#)

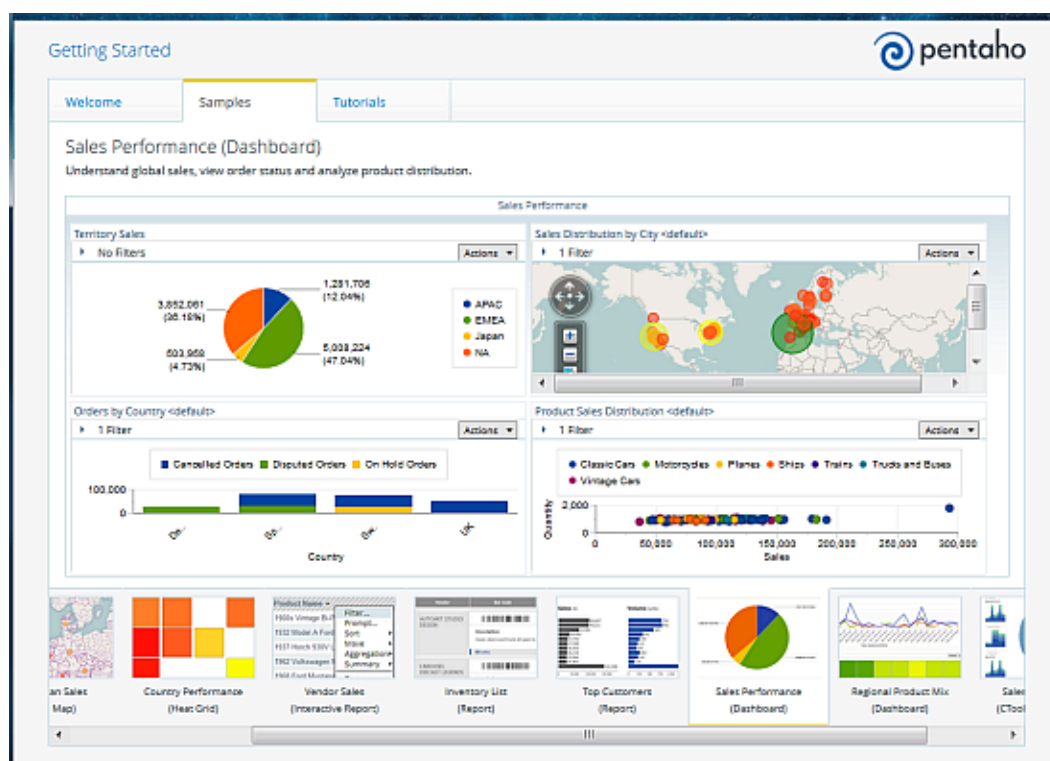
## Get started with Dashboard Designer

Dashboard Designer allows you to create dashboards with little or no training. The dashboard is several different reports brought together inside one screen. Use Dashboard Designer if you want to create an interface to view many different reports at once, have quick access to web pages that you visit often, or view dynamic charts and graphs within a space while you create reports in another.

## View a Dashboard sample

This section highlights some popular Dashboard Designer capabilities that are available, using the sample dashboard called Sales Performance (Dashboard), located in the Getting Started widget.

1. In the Getting Started widget on the Home page, click the Samples tab.
2. Scroll down to find the Sales Performance (Dashboard) sample. Click Explore to open a new browser screen, then click the Samples tab.
3. Scroll right in the horizontal list of items (at the bottom of the screen). Find Sales Performance (Dashboard) and click it to view the sample.



## Tour the Dashboard panels

You can view the editable version of the Sales Performance (Dashboard) in Dashboard Designer by clicking Browse Files on the User Console Home page.

Follow these steps:

1. In the Folders pane, click to expand the Public folder, then click to highlight the Steel Wheels folder.
2. In the center pane, double-click Sales Performance (Dashboard).

3. After the dashboard opens, click Edit in File Actions.

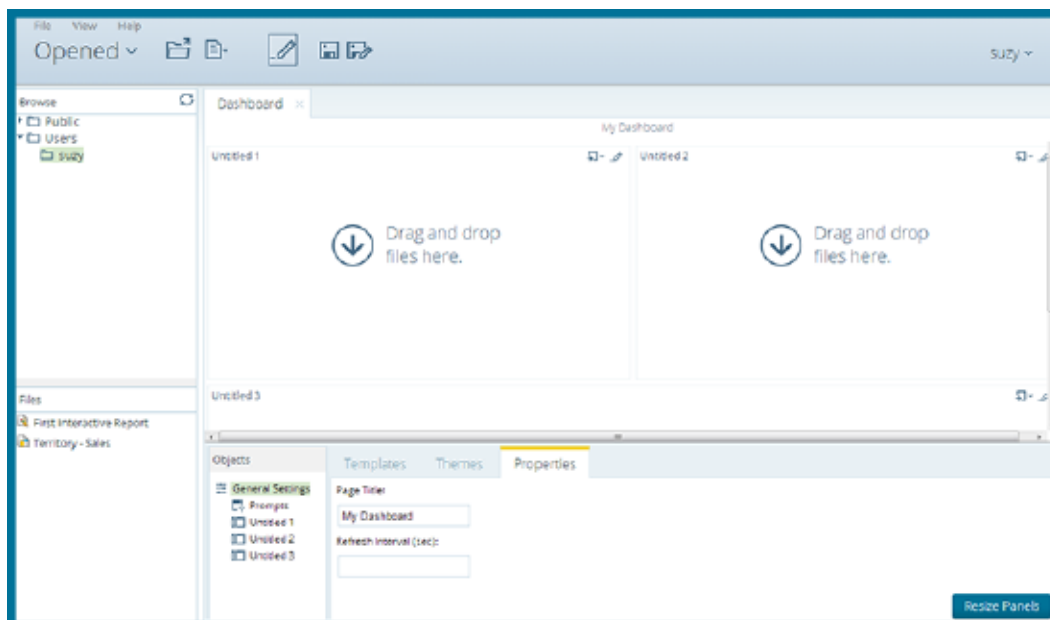


Item	Name	Function
1	Opened page	Displays quick access buttons across the top to create and save a new Analysis Report, Interactive Report, and Dashboard. Opened reports and files show as a series of tabs across the page.
2	Prompts panel	The prompts panel gives you a way to add filters to the individual parts of your dashboard.
3	Browse and Files panel	Locate your files using the Browse and Files panels, and add them to dashboards.
4	Dashboard canvas	Shows a dynamic view of your dashboard as you work to build it. The look of your dashboard refreshes as you add content from the Browse and Files panels, and work with the prompts or Objects panels.
5	Objects panel	Refine the look of your dashboard with the Objects panel by choosing a dashboard template or changing the titles for each object in the dashboard.

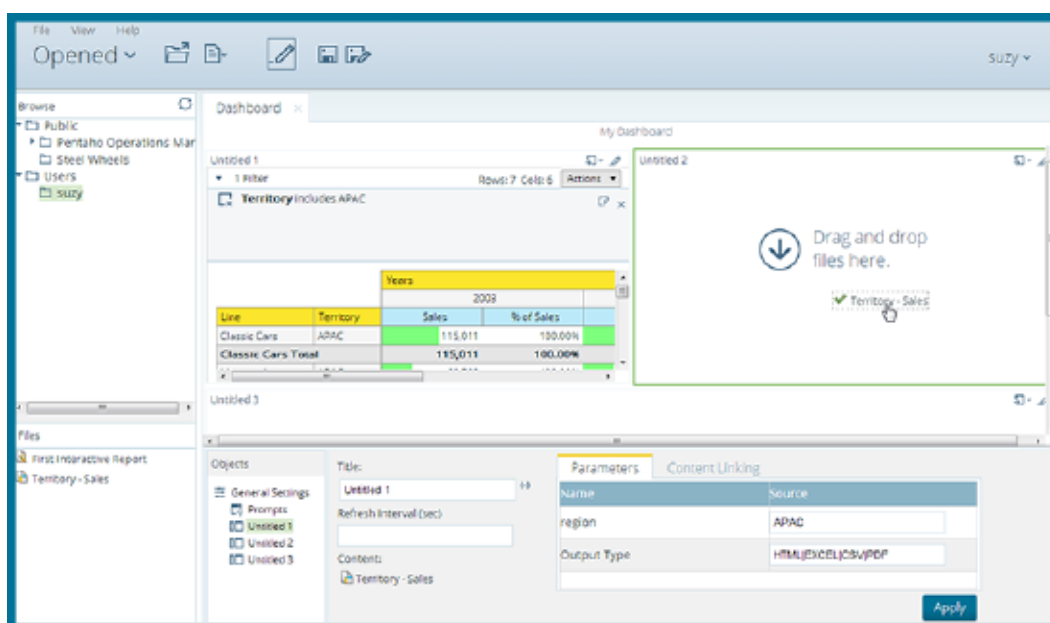
## Create your first dashboard

Follow the steps below to create your first dashboard:

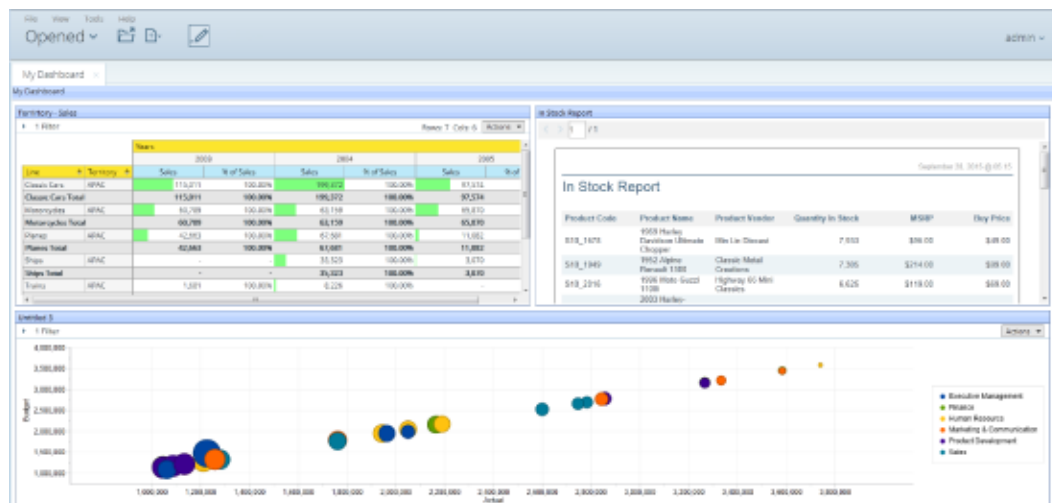
1. From the User Console Home page, click Create New, then choose Dashboard.
2. In the Edit pane, click the Templates tab and select the 2 over 1 layout to use for this exercise.
3. Click the Properties tab, and type My Dashboard in the Page Title text box.  
This is the title for your dashboard page.



4. Click the Themes tab and select the theme of your choice.  
The new theme will be applied to your dashboard immediately.
5. Locate the Analyzer report you created earlier by going to the Browse pane to the left of the dashboard canvas and clicking on the folder you used before. Click-and-drag the Territory - Sales file from the Files pane and onto the top-left dashboard panel.



6. Within the Edit pane, type Territory - Sales in the Title text box and click Apply.  
The dashboard panel is populated with the Territory - Sales report.
7. Locate your Interactive report through the Browse pane. Click-and-drag the In Stock Report file from the Files pane and onto the top-right dashboard panel.
8. Within the Edit pane, type In Stock Report in the Title text box and click Apply. Enter a title for this dashboard panel and click Apply.  
The right dashboard panel is populated by the In Stock report.
9. You can drag any report from the Steel Wheels folder in Public into the bottom dashboard panel.  
Enter a title for this dashboard panel and click Apply.
10. Click the Save as icon in the toolbar. When the Save As dialog box appears, save your dashboard as My Dashboard and click Save.
11. Close the dashboard (click the X to close it in the tabs at the top), then go to Browse Files and double-click on My Dashboard in the Files pane.



You have successfully created a simple dashboard from scratch. See *Pentaho Analytics* for details on how to work with complex dashboards.

## Next steps

After you have finished working through the walk-through tutorials, you are ready to learn more about Pentaho reporting with the following documents:

- *Pentaho Analytics*
- *Pentaho Report Designer*

# Getting started with Report Designer

This section contains some step-by-step instructions for creating and refining reports in Report Designer.

These topics are designed to quickly demonstrate basic Pentaho Report Designer features. See the Pentaho Report Designer document for more information about features and functions, from adding a data source to working with conditional formatting and formulas.

- Learn [About Pentaho Report Designer](#), a report creation tool that you can use alone, or as part of a Pentaho suite.
- The [Create a report with Report Designer](#) article shows you how to create a simple report with Report Designer:
- [Design your report](#)
- [Refine your report](#)
- [More about row banding, data formatting, and alignment](#)
- [Add a chart to your report](#)
- [Add parameters to your report](#)
- [Publish your report](#)

## About Pentaho Report Designer

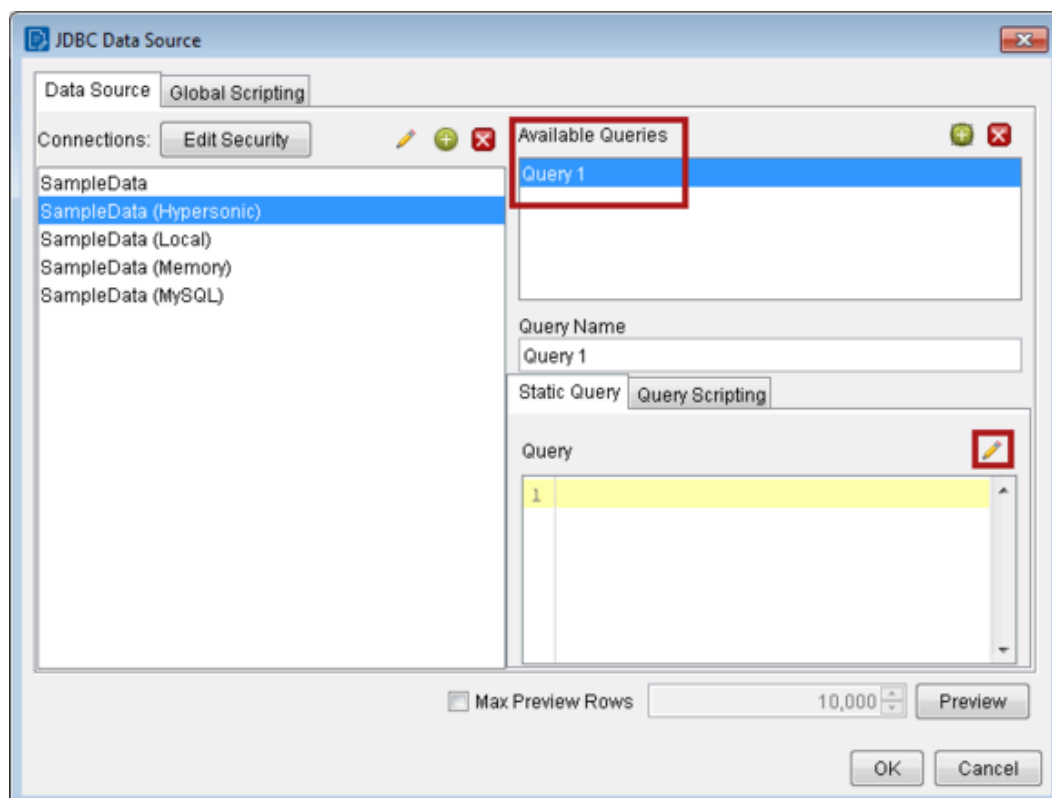
Pentaho Report Designer is a sophisticated report creation tool that you can use standalone, or as part of the larger Pentaho Business Analytics suite. It enables professionals to create highly detailed reports based on adequately prepared data from virtually any data source. The purpose of this document is to provide instructions for creating a simple Report Designer report.

This section contains walk-throughs for creating content in Report Designer. You must have Pentaho's sample database installed and available in order to follow the tutorials. Sample data is installed by default with Report Designer, though you or your system administrator may have removed it prior to production deployment. If it is been removed, you can simply reinstall Report Designer to get it back.


## Create a report with Report Designer

Perform the steps below to create a report using Report Designer:

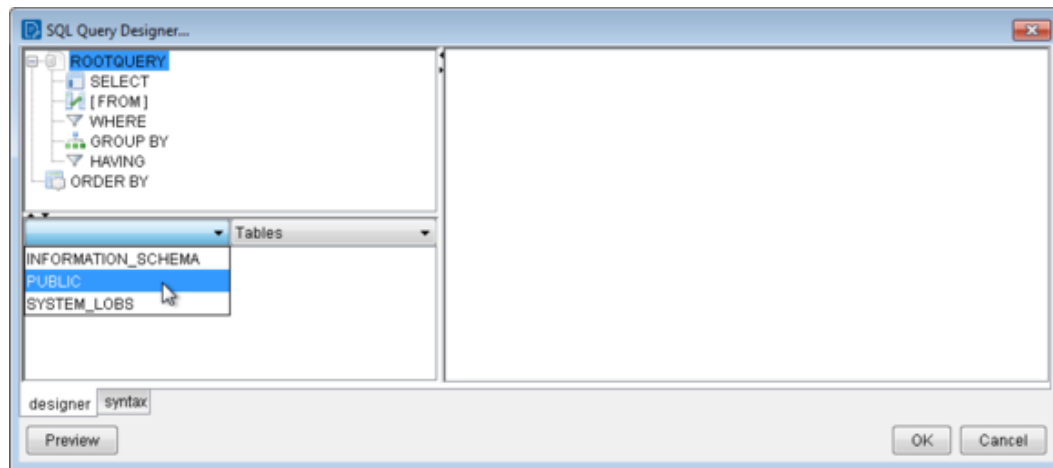
1. Start Report Designer. Go to Start > Programs > Pentaho Enterprise Edition > Design Tools > Report Designer.  
The Report Designer home page appears.  
Note: If you would like to change the size of the layout bands to give yourself more area to work in without changing the dimensions of the published report, you can click and drag the percentage number in the upper left corner of the workspace. The default is 100%, but if you click and drag it diagonally toward the upper left or lower right corners, the view will zoom in or out. If you want to reset the view to 100%, double-click the upper left corner where the percentage shows.
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 Data Source dialog box.  
The JDBC Data Source dialog box appears.
5. Under Connections, select SampleData (Hypersonic).
6. Next to Available Queries, click the Plus Sign to add queries.



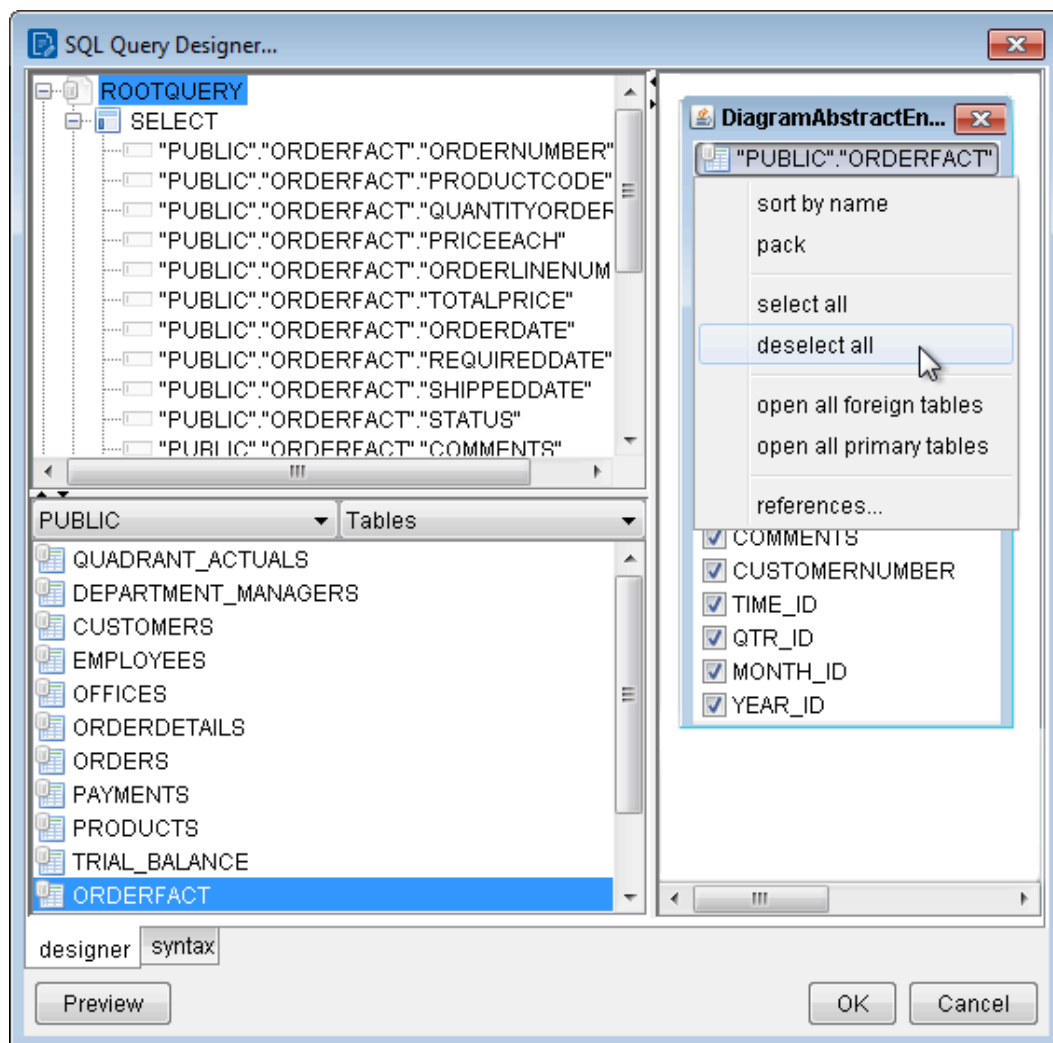
Query 1 appears under Available Queries and the edit (pencil) icon is now available, as shown below.

7. Click the editing icon (  ).  
The SQL Query Designer window opens. The SQL Query Designer provides you with a graphical environment which allows you to work with the data even if you do not understand SQL, the standard programming language for retrieving content from databases.
8. Select PUBLIC in the schema filter menu, and then double-click ORDERFACT so that the table appears in the workspace, as shown below.



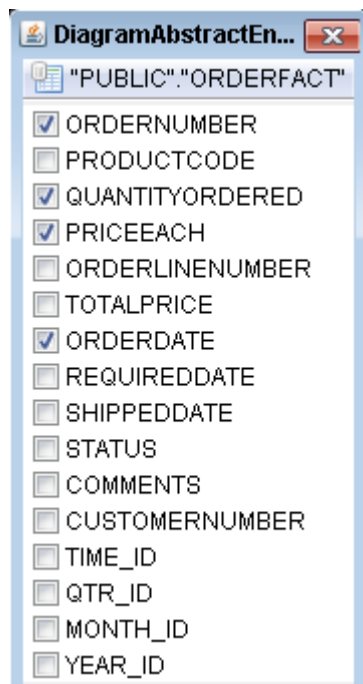


9. In the SQL Query Designer workspace, right-click **ORDERFACT** and choose **deselect all**, as shown below.

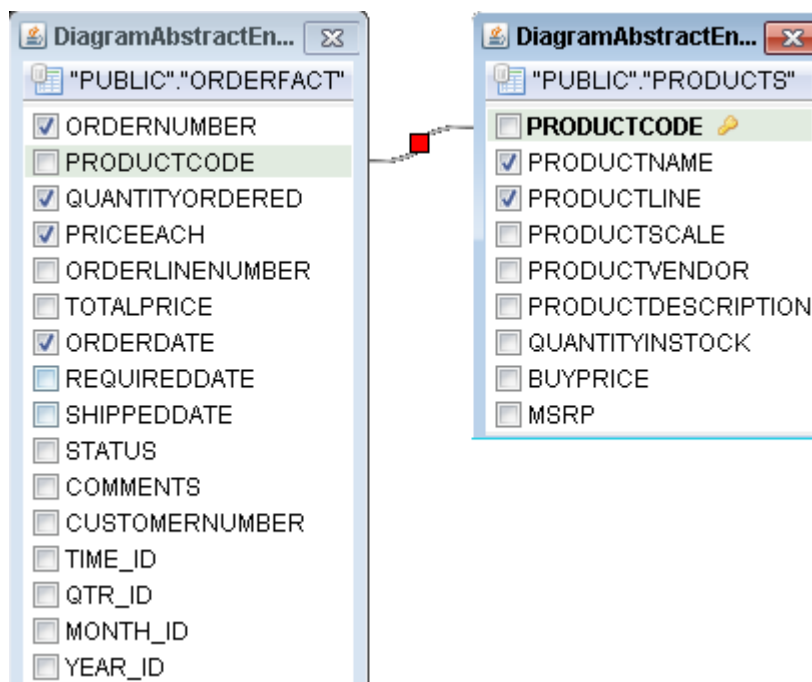


10. Now, select the **ORDERNUMBER**, **QUANTITYORDERED**, **PRICEEACH**, and **ORDERDATE** fields in the **ORDERFACT** table, as shown below.

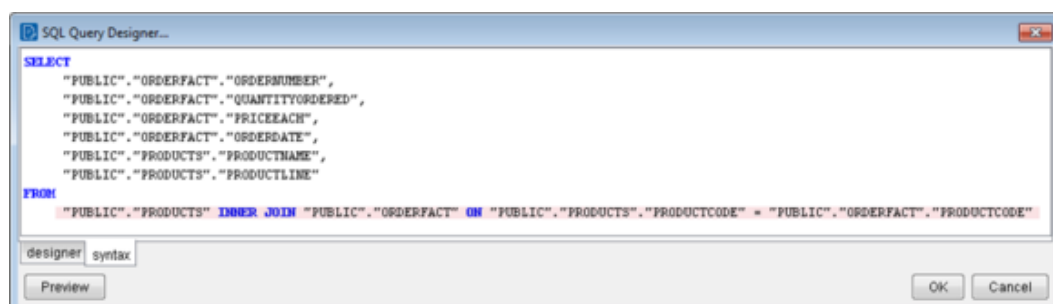




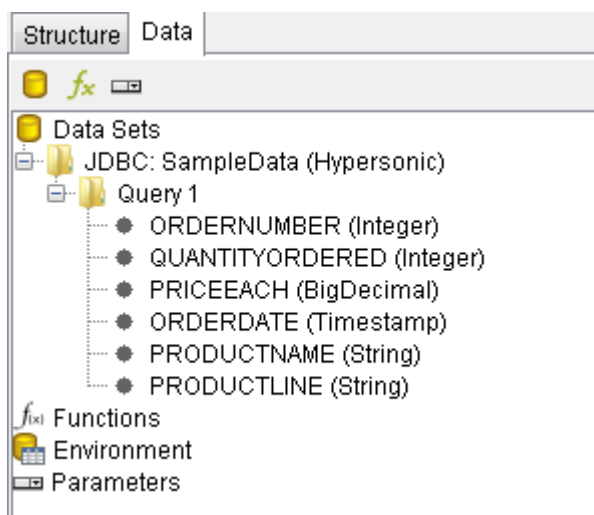
11. Double-click PRODUCTS so that the table also appears in the workspace.  
Notice the line joining the ORDERFACT and PRODUCTS tables together.
12. Clear all the PRODUCTS table fields, and then select just the PRODUCTNAME and PRODUCTLINE fields, as shown below.



13. For the purpose of this exercise, click the Syntax tab in the lower left portion of the SQL 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, as highlighted below:



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



The fields associated with your tables are now under Query 1, as shown below:

You are now ready to start [designing your report](#).

## Design 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 menu bar, 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, click and drag the ORDERNUMBER field into the Details band, ensuring 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 following example:



5. Click  (Preview) or select it from the View menu option to preview your report, as shown below:




10,107	Mon Feb 24 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	30	95.7
10,121	Wed May 07 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	34	81.35
10,134	Tue Jul 01 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	41	94.74
10,145	Mon Aug 25 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	45	83.26
10,159	Fri Oct 10 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	49	100
10,168	Tue Oct 28 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	36	96.66
10,180	Tue Nov 11 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	29	86.13
10,188	Tue Nov 18 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	48	100
10,201	Mon Dec 01 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	22	98.57
10,211	Thu Jan 15 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,223	Fri Feb 20 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	37	100
10,237	Mon Apr 05 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	23	100
10,251	Tue May 18 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	28	100
10,263	Mon Jun 28 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	34	100
10,275	Fri Jul 23 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	45	92.83
10,285	Fri Aug 27 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	36	100
10,299	Thu Sep 30 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	23	100
10,309	Fri Oct 15 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,318	Tue Nov 02 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	46	94.74
10,329	Mon Nov 15 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	42	100
10,341	Wed Nov 24 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,354	Sat Dec 04 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	42	86.13

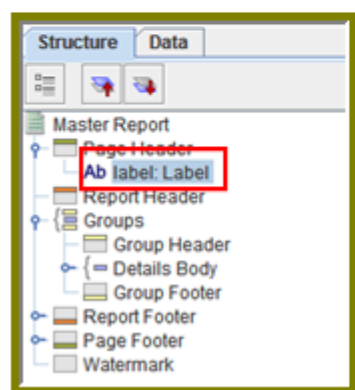
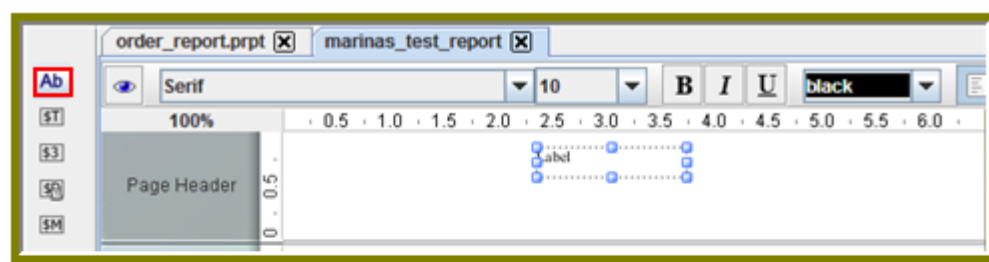
6. Click  (Edit) to return to the workspace view.

You have created your first report. But, wait... There's a problem. Without headers, report users will have a hard time understanding its content. You must continue refining your report, see [Refine your report](#).

## Refine your report

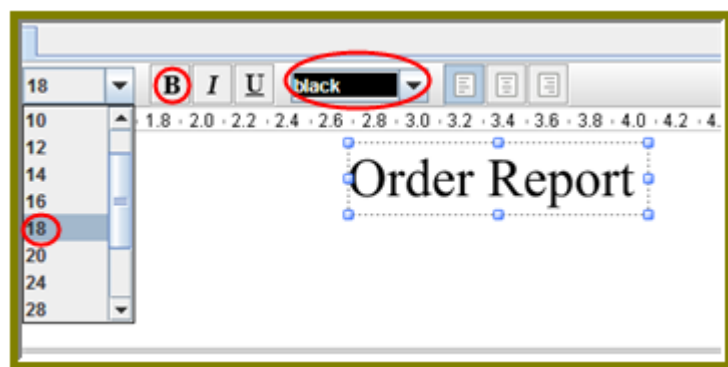
You have created a report in the [previous exercise](#) but now you need to make the report more descriptive so that users can understand the content in the report. Follow these instructions to refine your report.

1. Click and drag a label () from the tools palette in the left panel to the Page Header band. The Structure tab updates, as shown in the following figure:




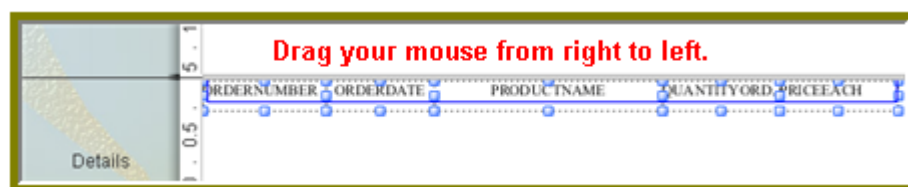
2. Click inside the Label item and type Order Report


3. Double-click inside the Order Report label to select the text, then select a larger font size (18 point) and apply boldface, as shown below.



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.

4. 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.
5. Now, create column headers by clicking the Details Header icon under the Structure tab in the right panel.  
The Style and Attributes tabs appear under the Structure tab.
6. Under common in the Attributes tab, change the Value of the hide-on-canvas option to False.  
The Details Header band appears in your workspace.
7. In the toolbar, click the Select Objects icon (  ).  
Notice that the icon changes to a cross hair as you move into the workspace.
8. 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, as shown in the following example:



9. 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.
10. Under Format in the Report Designer menu bar, select Morph > label.  
The column objects are changed to labels.
11. 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.
12. Click the Preview icon (  ) to display your report.  
The following image shows an example of how the report might appear. The report looks good but you may want to make it even easier to read by applying some banding.

Order Report				
Order No.	Order Date	Product Name	Quan.	Price Each
10,107	Mon Feb 24 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	30	95.7
10,121	Wed May 07 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	34	81.35
10,134	Tue Jul 01 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	41	94.74
10,145	Mon Aug 25 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	45	83.26
10,159	Fri Oct 10 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	49	100
10,168	Tue Oct 28 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	36	96.66
10,180	Tue Nov 11 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	29	86.13
10,188	Tue Nov 18 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	48	100
10,201	Mon Dec 01 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	22	98.57
10,211	Thu Jan 15 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,223	Fri Feb 20 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	37	100
10,237	Mon Apr 05 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	23	100
10,251	Tue May 18 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	28	100
10,263	Mon Jun 28 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	34	100
10,275	Fri Jul 23 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	45	92.83

13. In the toolbar, go to Format > Row Banding.

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

15. Click the Preview icon (  ) to display your report.

The following figure shows an example of how the report might appear:

Order Report				
Order No.	Order Date	Product Name	Quan.	Price Each
10,107	Mon Feb 24 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	30	95.7
10,121	Wed May 07 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	34	81.35
10,134	Tue Jul 01 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	41	94.74
10,145	Mon Aug 25 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	45	83.26
10,159	Fri Oct 10 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	49	100
10,168	Tue Oct 28 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	36	96.66
10,180	Tue Nov 11 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	29	86.13
10,188	Tue Nov 18 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	48	100
10,201	Mon Dec 01 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	22	98.57
10,211	Thu Jan 15 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,223	Fri Feb 20 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	37	100
10,237	Mon Apr 05 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	23	100
10,251	Tue May 18 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	28	100
10,263	Mon Jun 28 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	34	100
10,275	Fri Jul 23 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	45	92.83

16. In the menu bar, go to File > Save to save your report in the .../report-designer/samples folder.

Enter Orders into the File Name text box.

Note: See [More about row banding, data formatting, and alignment](#) for additional information about refining your report.

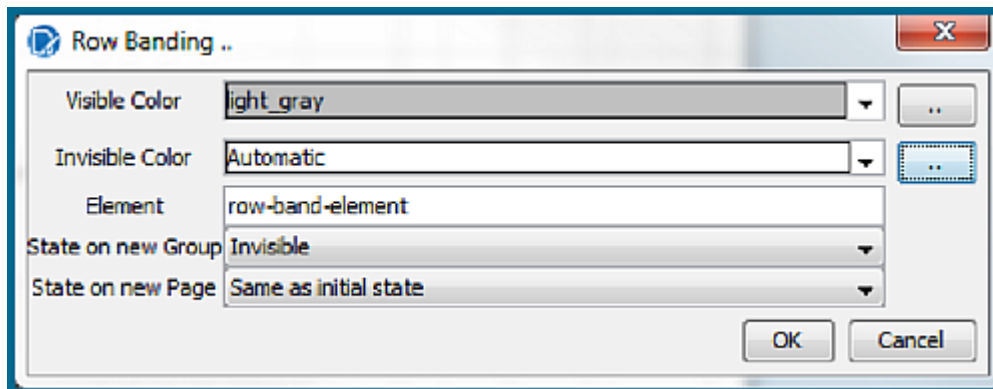
Now that your report is easier to read, you can consider [adding a chart](#).

## More about row banding, data formatting, and alignment

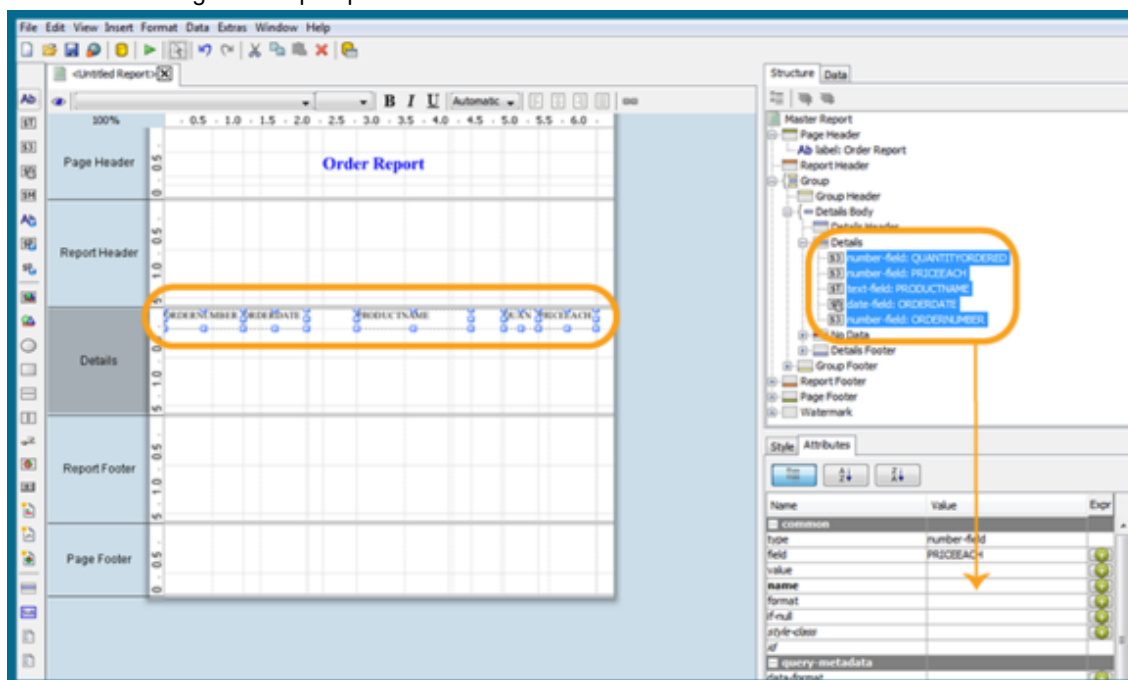
### Row banding

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.



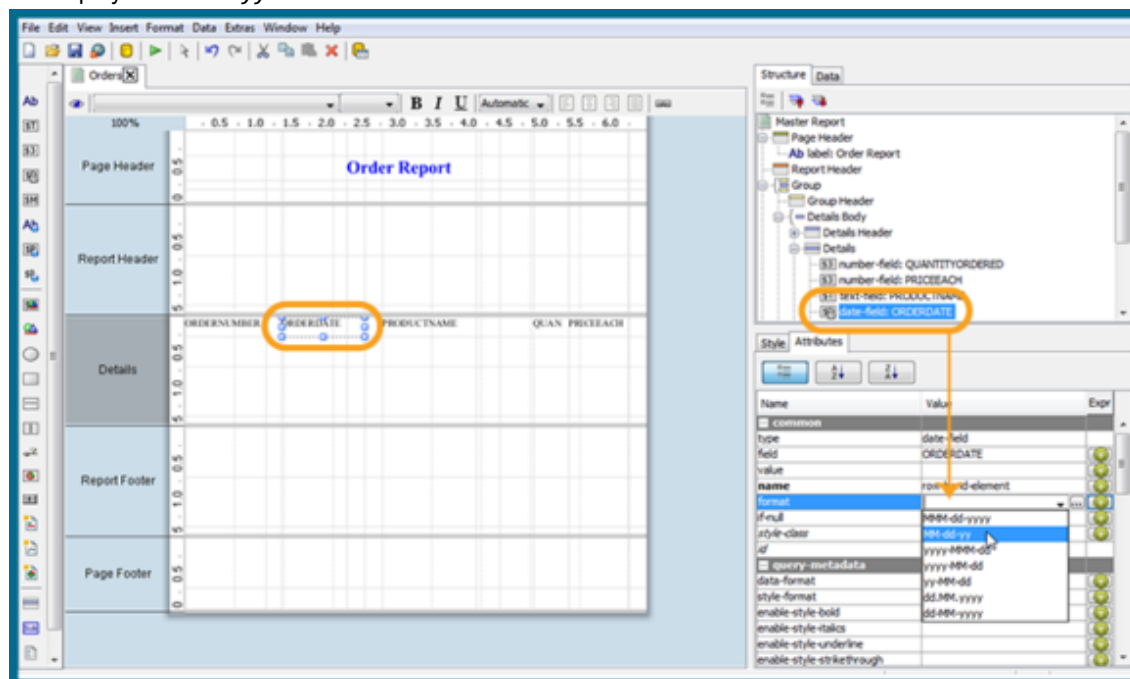
After you create your element, go back to the report and select the columns (fields) whose data will always be displayed with a row band. You must also type row-band-element in the name field under Attributes. In the example below, the data associated with each of the columns in the report will display 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:

Order No.	Order Date	Product Name	Quan.	Price Each
10,107	02-24-03	1969 Harley Davidson Ultimate Chooser	30	95.7
10,121	05-07-03	1969 Harley Davidson Ultimate Chooser	34	81.35
10,134	07-01-03	1969 Harley Davidson Ultimate Chooser	41	94.74
10,145	08-25-03	1969 Harley Davidson Ultimate Chooser	45	83.26
10,159	10-10-03	1969 Harley Davidson Ultimate Chooser	49	100
10,168	10-28-03	1969 Harley Davidson Ultimate Chooser	36	96.66
10,180	11-11-03	1969 Harley Davidson Ultimate Chooser	29	86.13
10,188	11-18-03	1969 Harley Davidson Ultimate Chooser	48	100
10,201	12-01-03	1969 Harley Davidson Ultimate Chooser	22	98.57
10,211	01-15-04	1969 Harley Davidson Ultimate Chooser	41	100
10,223	02-20-04	1969 Harley Davidson Ultimate Chooser	37	100
10,237	04-05-04	1969 Harley Davidson Ultimate Chooser	23	100
10,251	05-18-04	1969 Harley Davidson Ultimate Chooser	28	100
10,263	06-28-04	1969 Harley Davidson Ultimate Chooser	34	100
10,275	07-23-04	1969 Harley Davidson Ultimate Chooser	45	92.83
10,285	08-27-04	1969 Harley Davidson Ultimate Chooser	36	100
10,299	09-30-04	1969 Harley Davidson Ultimate Chooser	23	100
10,309	10-15-04	1969 Harley Davidson Ultimate Chooser	41	100
10,318	11-02-04	1969 Harley Davidson Ultimate Chooser	46	94.74
10,336	11-14-04	1969 Harley Davidson Ultimate Chooser	35	100

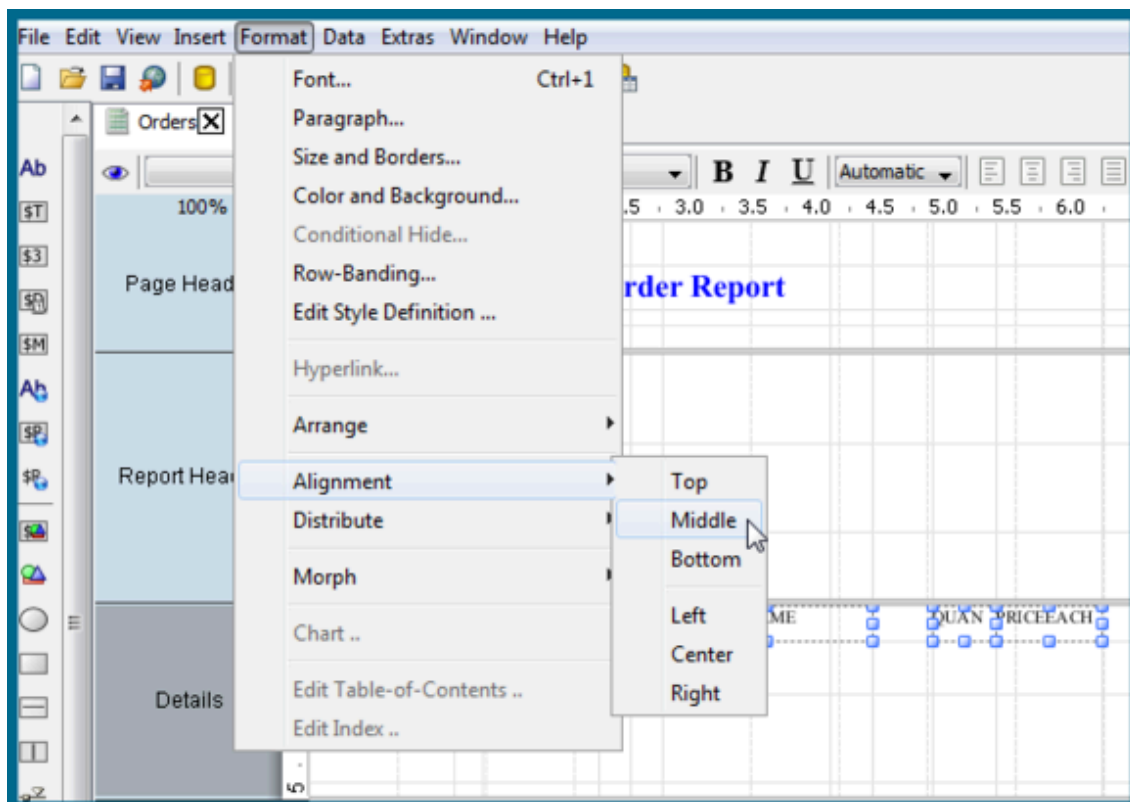
Note: You can type a value for your own format if you know the correct JavaScript string nomenclature.

## 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 the Select Objects icon (  ) 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.

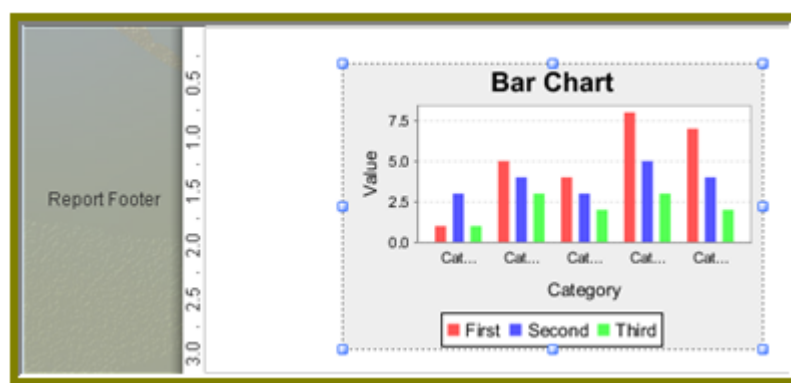




## Add a chart to your report

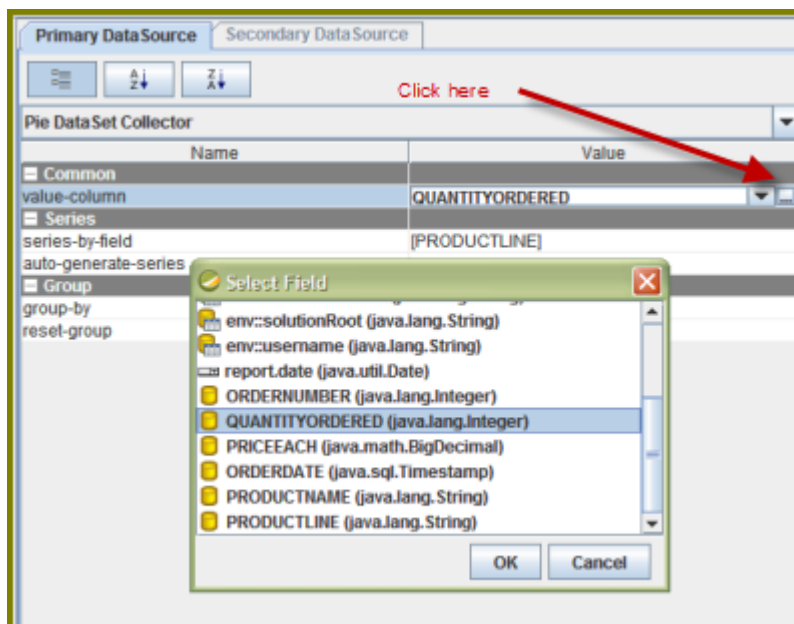
In the [previous exercise](#), you have seen a small subset of features associated with Report Designer. In this exercise, you will add a chart to your report.

1. Click File > Open and find the report you just saved. Click OK.
2. In the Palette, click and drag a Chart icon (📊) into the Report Footer band.
3. Use the resizing handles to center and stretch the chart. You can also adjust the width of the band, as shown in the following example:

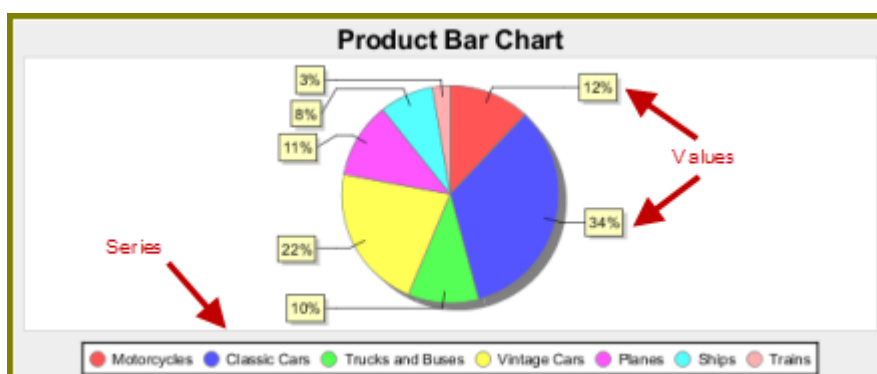


4. Double-click the sample chart.
5. The type of chart most appropriate for your report is a pie chart. In the top portion of the Edit Chart dialog box, select the pie chart icon.  
The pie chart properties that define its look and feel are listed in the left pane of the Edit Chart dialog box. Properties associated with the data in the chart are listed in the right pane.
6. Go to the Title properties. Next to the chart-title property, enter Product Pie Chart.
7. Under Common in the Primary DataSource tab, click the down arrow next to value-column then click the Ellipsis button to open the Select Field dialog box.  
The value-column specifies the actual values (measures) you want to chart. In this instance, you want to chart the quantity ordered, as shown in the following figure:



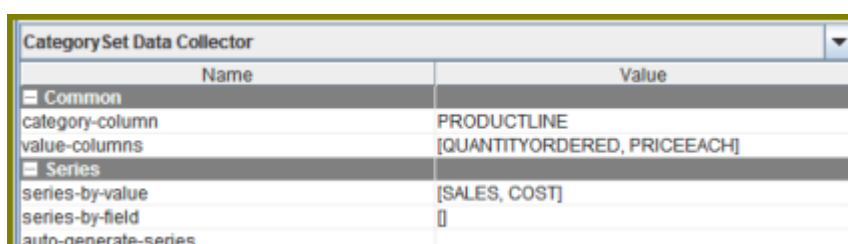


8. Select QUANTITYORDERED from the list and click OK.
9. Under Series, click the Ellipsis button next to series-by-field to open the Edit Array dialog box.  
Series-by-field specifies the field you are charting. In this instance you are charting by product line.  
Series are usually placed in the chart legend.
10. Click the Add icon (+).
11. Click in the blank field to expose the drop-down arrow.
12. Select PRODUCTLINE from the list and click OK.
13. Click OK to exit the Edit Chart dialog box.
14. Click the Preview icon (👁) to display your report.
15. When the report displays, go to end of the last page to see the following example:



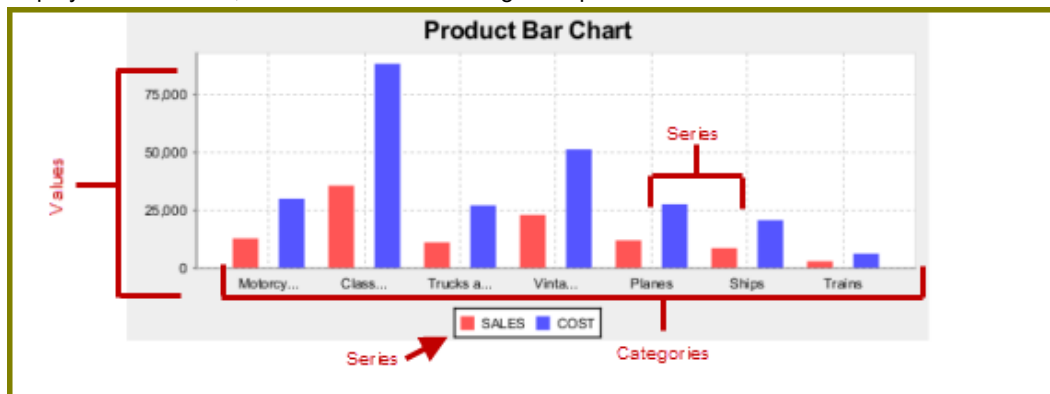
16. Save your report.

If you want to display a bar or line chart instead of a pie chart, use the chart settings, as shown in the following image:



You have to add the series-by-value setting manually in the Edit Array dialog box. Click the Add icon (+) and type SALES, and then click the Add icon again and type COST. Running the report




displays the bar chart, as shown in the following example:



You are now ready to [add a parameter to your report](#).

## Add parameters to your report


Previously, you added a [table](#) and a [chart](#) to your report. Now, you will make your report interactive by setting parameters. When you set parameters, users are prompted for a value or values when they run the report.

1. In the Report Designer, if it is not already open, click File > Open and select to open your Orders report.
2. In the menu bar, 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, enter enter\_prodline in the Name text field.
4. Enter Select Line in the Label text field.
5. Next to Display Type, select Drop Down so users can select a product line.
6. Click on JDBC (SampleData (Hypersonic)) under DataSources, and then click Edit icon (  ) to add a query that supplies the values (motorcycles, cars, ships, and so on) from which users of the report must choose.  
The JDBC Data Source dialog box appears.
7. Under Connections, select SampleData (Memory).
8. Next to Available Queries click Add icon (  ).  
A new query placeholder is added to the list (Query 2).
9. In the Query Name text field, enter prodlineList.
10. Enter your SQL query in the Query box. Either copy and paste the following SQL statements directly under Query in the Static Query tab:

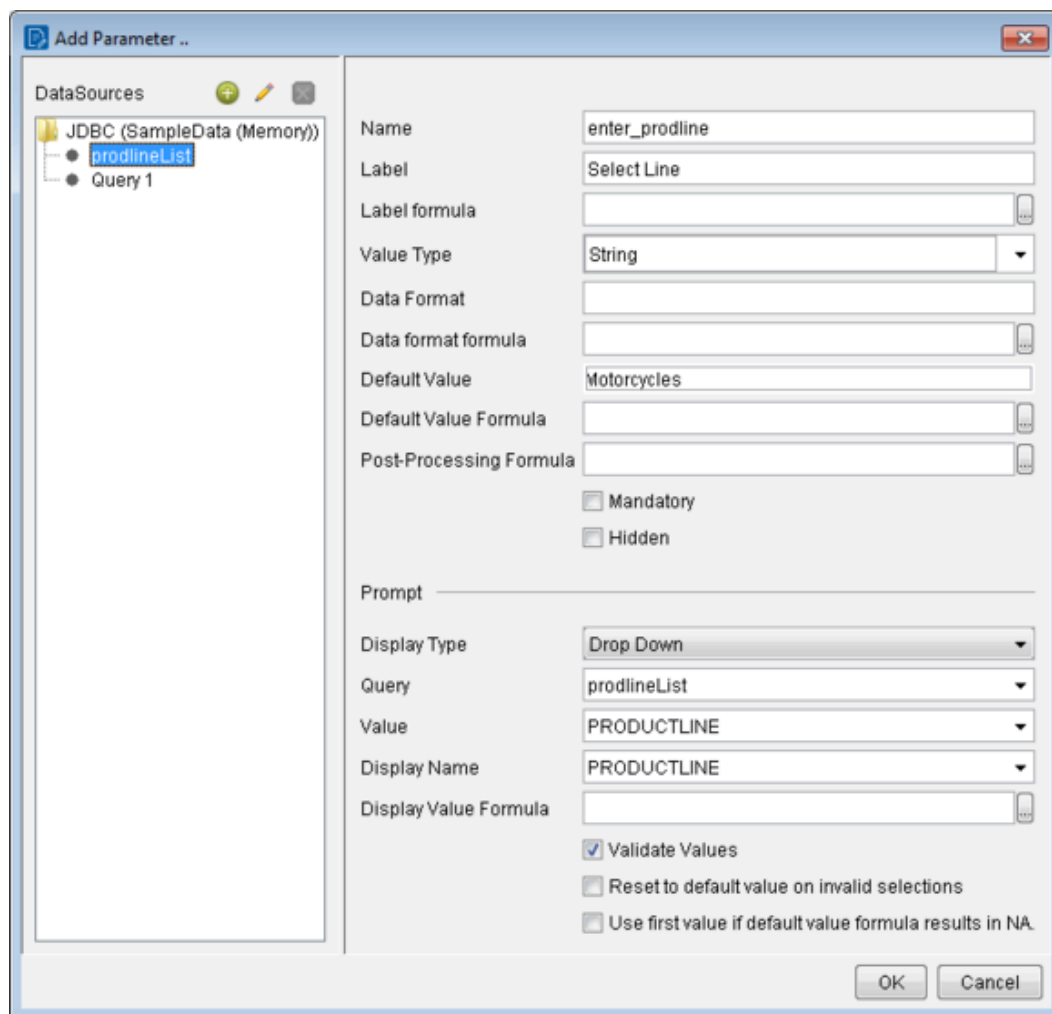
```
SELECT DISTINCT
    "PRODUCTS"."PRODUCTLINE"
FROM
    "PRODUCTS"
```

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

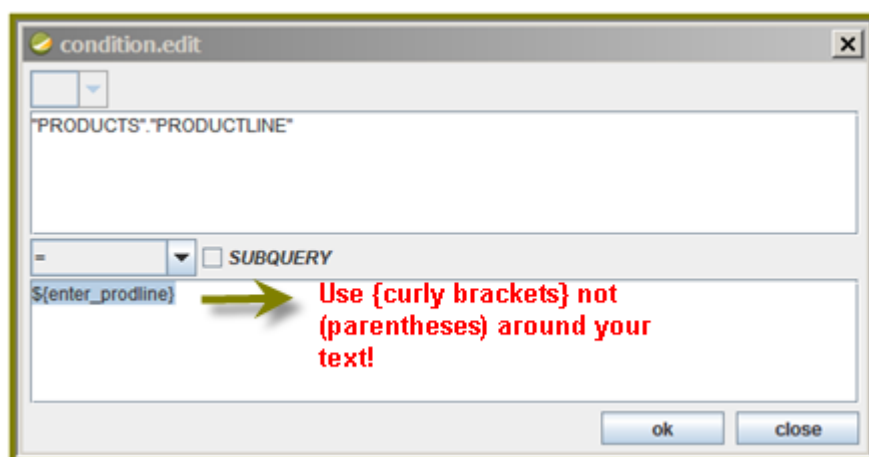
Or, use the SQL Query Designer to build your query as shown in the following steps:

- a. Click Edit icon (  ) to the upper right of the State Query tab.
  - b. In the schema filter menu of the SQL Query Designer, select PUBLIC.
  - c. Double-click the PRODUCTS table to select it.
  - d. In the right panel, click PRODUCTS and choose Deselect All.
  - e. Right-click SELECT in the upper left panel and choose Distinct.
  - f. In the right panel, select PRODUCTLINE.
  - g. Click OK to exit the SQL Query Designer and go to the next step.
11. Click OK to exit Data Source dialog box.
  12. In the Add Parameter dialog box under DataSources, select prodlineList.

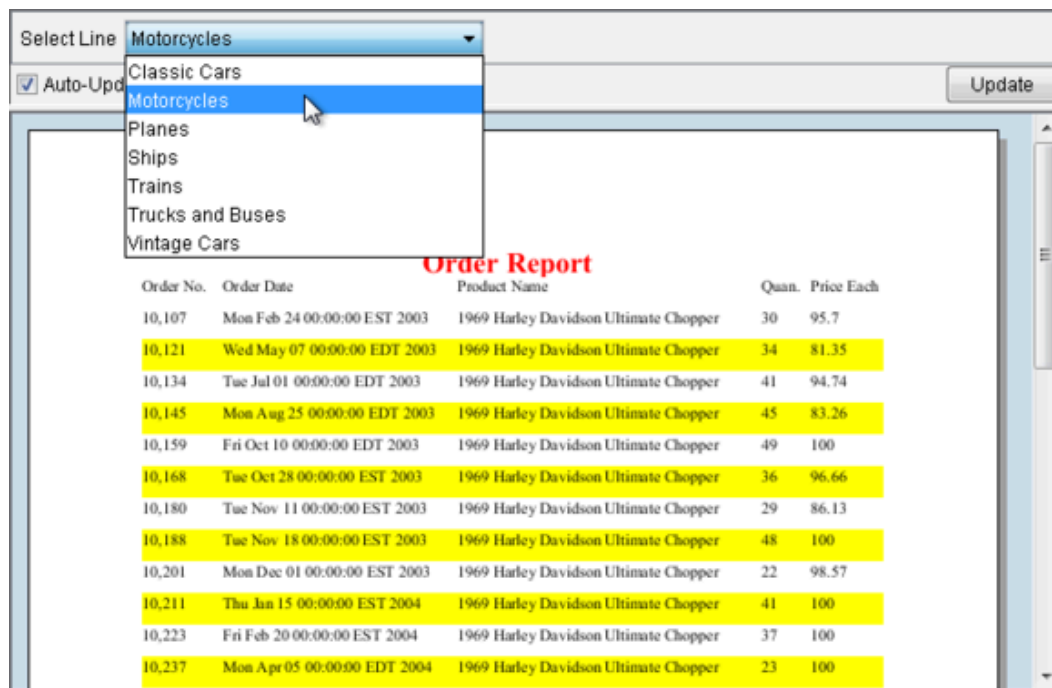
13. Next to Value Type, select String.
14. Optionally, type a default value (for example, Motorcycles) in the Default Value text box, as shown in the following example:



15. Click OK to exit the Add Parameter dialog box.
16. Now that you have created a product line parameter, you must map it back to your query (Query 1). Under Data, double-click Query 1, to open Query 1 in the JDBC Data Source dialog box.
17. Click Edit icon ( ) to the upper right of the State Query tab to access the SQL Query Designer, right-click PRODUCTLINE in the right panel, and then select add where condition. The condition.edit dialog box appears.
18. Type \${enter\_procline} into the edit area in the lower panel of the dialog box as shown in the following example, and then click OK:



19. Click OK to exit the SQL Query Designer.
20. Click OK to exit the Data Source dialog box.
21. Click the Preview button ( ) and notice the new product line menu, as shown below:




Order No.	Order Date	Product Name	Quan.	Price Each
10,107	Mon Feb 24 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	30	95.7
10,121	Wed May 07 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	34	81.35
10,134	Tue Jul 01 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	41	94.74
10,145	Mon Aug 25 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	45	83.26
10,159	Fri Oct 10 00:00:00 EDT 2003	1969 Harley Davidson Ultimate Chopper	49	100
10,168	Tue Oct 28 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	36	96.66
10,180	Tue Nov 11 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	29	86.13
10,188	Tue Nov 18 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	48	100
10,201	Mon Dec 01 00:00:00 EST 2003	1969 Harley Davidson Ultimate Chopper	22	98.57
10,211	Thu Jan 15 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	41	100
10,223	Fri Feb 20 00:00:00 EST 2004	1969 Harley Davidson Ultimate Chopper	37	100
10,237	Mon Apr 05 00:00:00 EDT 2004	1969 Harley Davidson Ultimate Chopper	23	100

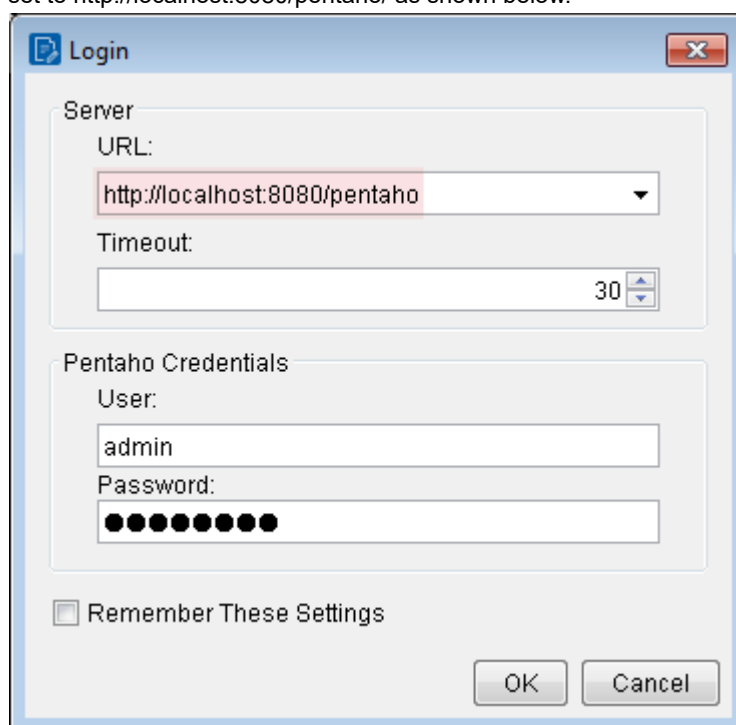
22. Save and close the report.

You are now ready to [publish your report](#).

## Publish your report

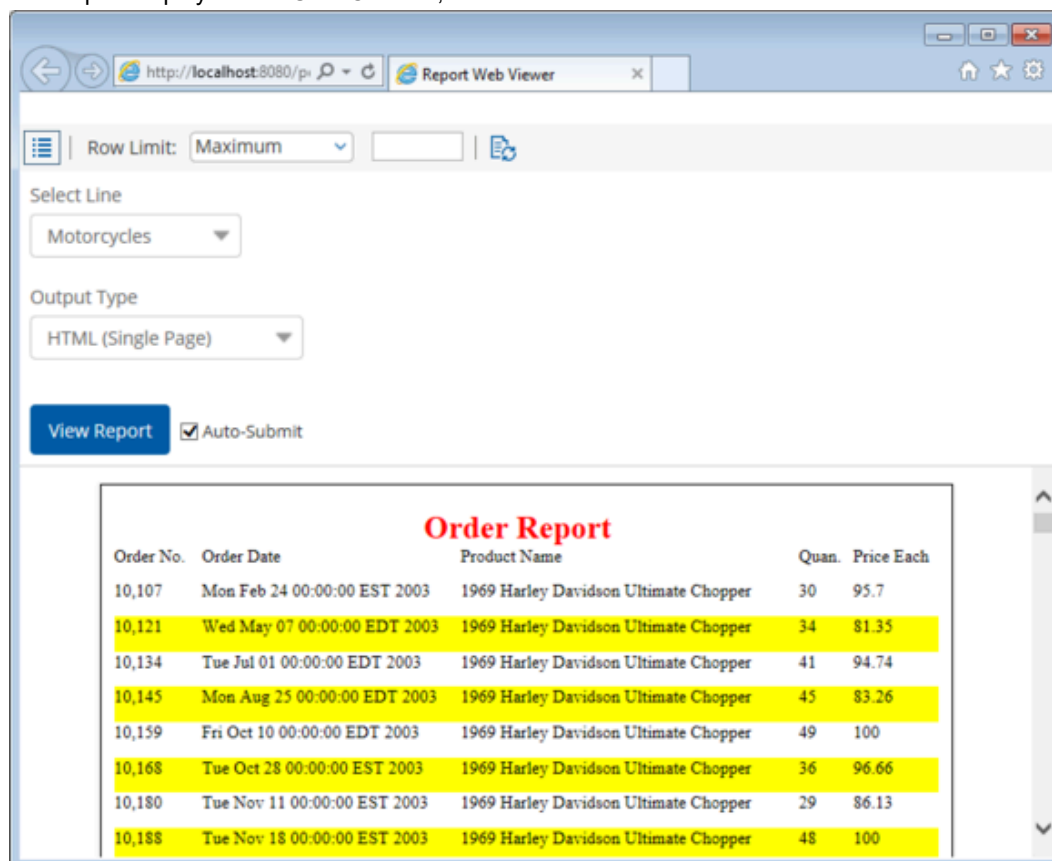
You have created and formatted [a simple report](#), [added a chart](#), [added a parameter](#), and now you are ready to share the report with your users.

1. In Report Designer, click File > Open to open the report you just created.
2. Click File > Publish. Alternatively, click the Publishes the report on a Pentaho server icon (  ).  
If you have not saved the report, a warning message reminds you to save it. The Login dialog box appears, pre-populated with credentials valid for the evaluation. Make sure that the server's URL is set to `http://localhost:8080/pentaho/` as shown below.



3. Click OK.  
The Publish to Server dialog box appears.
4. In the dialog box, type in a report title and description into the appropriate fields.
5. Under Location, save the report in the `...public/Steel Wheels` folder.

6. Select html as the Output Type and click OK.  
A success message appears.
7. Click Yes to go directly to the User Console to view the report you just published.  
If you want to access the report later, log on to the Pentaho Server by going to <http://localhost:8080/pentaho/> in a 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.
8. Log on as Admin.  
The default admin password is password.  
Your report displays in the User Console, as shown below:



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

You now have a report that users can view at any time.

## Getting Started with PDI

If you are new to Pentaho Data Integration, start here. These tutorials provide step-by-step instructions for creating and refining transformations and jobs using the PDI client (Spoon).

- [PDI Transformation Tutorial](#)
- [PDI Job Tutorial](#)
- [Getting Started with PDI and Hadoop](#)

The tutorials above are designed to quickly demonstrate basic PDI features. For more detailed information about PDI features and functions, see the following topics in the *Pentaho Data Integration* document:

- *Learn about the PDI Client*
- *Use Pentaho Repositories in PDI*
- *Schedule Perspective in the PDI Client*

### Pentaho Data Integration (PDI) tutorial

The following tutorial is intended for users who are new to the Pentaho suite or who are evaluating Pentaho as a data integration and business analysis solution. The tutorial consists of six basic steps, demonstrating how to build a data integration transformation and a job using the features and tools provided by Pentaho Data Integration (PDI).

The Data Integration perspective of PDI allows you to create two basic file types: transformations and jobs. Transformations describe the data flows for ETL such as reading from a source, transforming data and loading it into a target location. Jobs coordinate ETL activities such as defining the flow and dependencies for what order transformations should be run, or prepare for execution by checking conditions such as, "Is my source file available?" or "Does a table exist in my database?"

The aim of this tutorial is to walk you through the basic concepts and processes involved in building a transformation with PDI in a typical business scenario. In this scenario, you are loading a flat file (CSV) of sales data into a database to generate mailing lists. Several of the customer records are missing postal codes that must be resolved before loading into the database. In the preview feature of PDI, you will use a combination of steps to cleanse, format, standardize, and categorize the sample data. The six basic steps are:

[Step 1: Extract and load data](#)

[Step 2: Filter for missing codes](#)

[Step 3: Resolve missing data](#)

[Step 4: Clean the data](#)

[Step 5: Run the transformation](#)

[Step 6: Orchestrate with jobs](#)

## Prerequisites

To complete this tutorial, you need the following items:

- An installed version of the [Pentaho 30-day trial](#).

## Step 1: Extract and load data

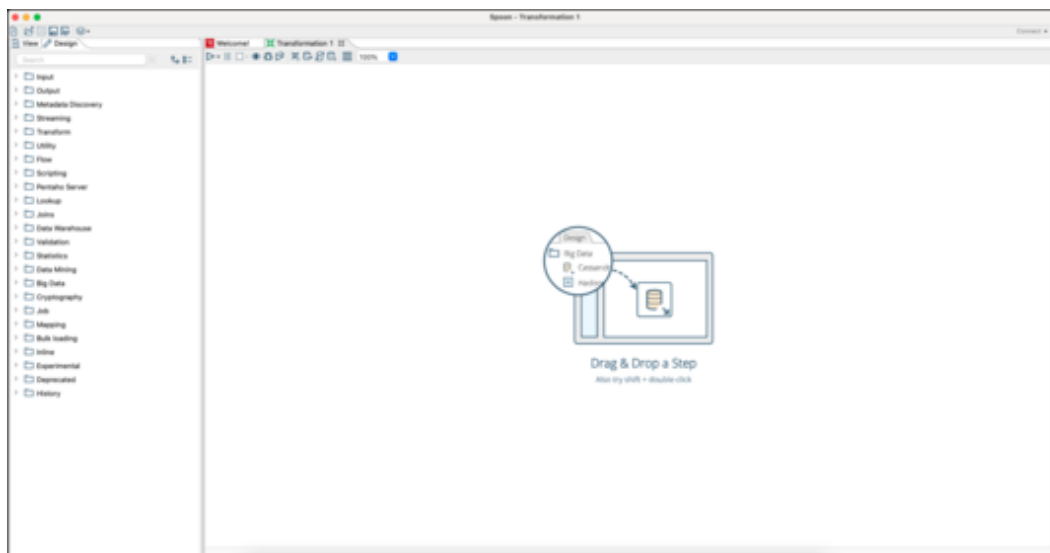
In Step 1, you will retrieve data from a CSV flat file and use the Text File Input step to connect to a repository, view the file schema, and retrieve the data contents.

### Create a new transformation

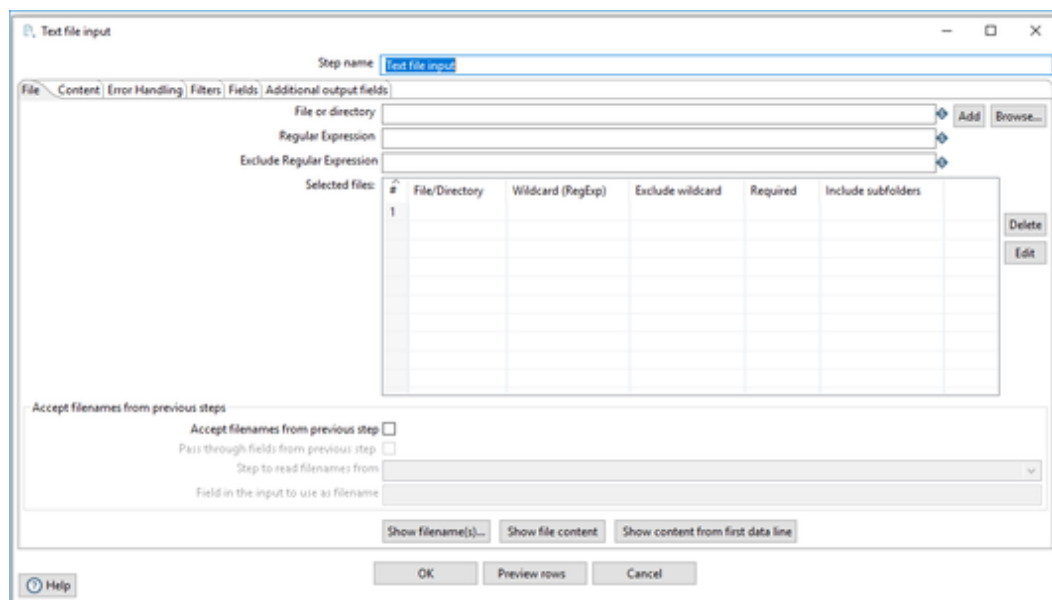
Follow these steps to create a new transformation.

If you want to insert a variable into a field that accepts variables, you can put your cursor in the fields and press CTRL+Spacebar to see a list of variables to insert. Fields that accept variables have a blue diamond.

1. Select File > New > Transformation in the upper-left corner of the PDI window.



2. Under the Design tab, expand the Input node, then select and drag a Text File Input step onto the canvas.
3. Double-click the Text File input step. In the Text file input window, you can set the properties of the step.



4. In the Step Name field, type Read Sales Data.  
The Text file input step is now renamed to Read Sales Data.
5. Click Browse to locate the sales\_data.csv source file in the ...\\design-tools\\data-integration\\samples\\transformations\\files folder. The Browse button appears in the upper-right side of the window near the File or Directory field.
6. Change File type to \*.csv. Select sales\_data.csv, then click OK.  
The path to the source file appears in the File or directory field.
7. Click Add.  
The path to the file appears under Selected Files.

### View the content in the sample file

Follow these steps to look at the contents of the sample file.

1. Click the Content tab, then set the Format field to Unix.
2. Click the File tab again and click the Show file content in the lower section of the window.
3. The Number of lines (0-all lines) window appears. Click OK to accept the default.
4. The Content of first file window displays the file. Examine the file to see how that input file is delimited, what enclosure character is used, and whether or not a header row is present.

In the sample, the input file is comma delimited, using the enclosure character of a quotation mark ("). It contains a single header row containing field names.

- Click the Close button to close the window.

### Edit and save the transformation

Follow these steps to provide information about the data's content.

- Click the Content tab. Use the fields under the Content tab to define how your data is formatted.
- Verify that the Separator is set to comma (,) and that the Enclosure is set to quotation mark ("). Select Header and enter 1 in the Number of header lines field.

- Click the Fields tab and click Get Fields to retrieve the input fields from your source file. When the Number of lines to sample window appears, enter 0 in the field, then click OK.
- If the Scan Result window displays, click Close to close the window.

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type
1	ORDERNUMBER	Integer	#		15	0	\$	-	-	-		none
2	QUANTITYORDERED	Integer	#		15	0	\$	-	-	-		none
3	PRICEEACH	Number	#,.		5	2	\$	-	-	-		none
4	ORDERLINENUMBER	Integer	#		15	0	\$	-	-	-		none
5	SALES	Number	#,.		7	2	\$	-	-	-		none
6	ORDERDATE	String			15		\$	-	-	-		none
7	STATUS	String			10		\$	-	-	-		none
8	QTR_ID	Integer	#		15	0	\$	-	-	-		none
9	MONTH_ID	Integer	#		15	0	\$	-	-	-		none
10	YEAR_ID	Integer	#		15	0	\$	-	-	-		none
11	PRODUCTLINE	String			12		\$	-	-	-		none
12	MSRP	Integer	#		15	0	\$	-	-	-		none
13	PRODUCTCODE	String			8		\$	-	-	-		none
14	CUSTOMERNAME	String			30		\$	-	-	-		none
15	PHONE	String			16		\$	-	-	-		none
16	ADDRESSLINE1	String			40		\$	-	-	-		none
17	ADDRESSLINE2	String			9		\$	-	-	-		none
18	CITY	String			14		\$	-	-	-		none
19	STATE	String			10		\$	-	-	-		none
20	POSTALCODE	String			8		\$	-	-	-		none
21	COUNTRY	String			13		\$	-	-	-		none
22	TERRITORY	String			5		\$	-	-	-		none
23	CONTACTLASTNAME	String			10		\$	-	-	-		none
24	CONTACTFIRSTNAME	String			9		\$	-	-	-		none

- To verify that the data is read correctly, click the Content tab, then click Preview Rows.
- In the Enter the number of rows you would like to preview window, click OK to accept the default. The Examine preview data window appears.



7. Review the data. Do you notice any missing, incomplete, or variations of the data?
  - o STATE & POSTALCODE both contain <null>
  - o COUNTRY contains both USA and United States.
8. Click OK to save the information that you entered in the step.
9. Enter a name for the transformation and provide additional properties using the Transformation Properties window. There are multiple ways to open the Transformation Properties window.
  - o Right-click on any empty space on the canvas and select Properties.
  - o Double-click on any empty space on the canvas to select Properties.
  - o Enter the CTRL-T keyboard combination.
10. In the Transformation Name field, enter Getting Started Transformation.  
Below the name, the filename is empty.
11. Click OK to close the Transformation Properties window.
12. To save the transformation, select File > Save.  
When saving your transformation for the first time, you are prompted for a file location and name of your choice. The file extension .ktr is the usual file extension for transformations.

### Load data into a relational database

Now you are ready to take all the records that are exiting the Filter rows step where the POSTALCODE was not null (the true condition), and load them into a database table. You will use the Table Output step and a hop from the Text File Input step to direct the data stream into a database table. This section of the tutorial uses a pre-existing database established during the Pentaho installation, which is started along with the server.

#### Create the Table Output step

Follow these instructions to create the Table Output step.

1. Under the Design tab, expand the contents of the Output node.
2. Click and drag a Table Output step into your transformation.
3. Create a hop between the Read Sales Data and Table Output steps. To create the hop:
  - a. Press the SHIFT key.
  - b. Click the Read Sales Data (Text File Input) step and drag the mouse to draw a line to the Table Output step.
  - c. Release the SHIFT key.
  - d. Click the Table Output step.
4. Double-click the Table Output step to open its Edit properties dialog box.
5. Rename your Table Output step to Write to Database.

#### Create a connection to the database

Follow these steps to create a connection to the database.

1. Click New next to the Connection field. You must create a connection to the database.  
The Database Connection window appears.
2. Provide the settings for connecting to the database.

Field	Setting
Connection Name	Sample Data
Connection Type	Hypersonic
Host Name	localhost
Database Name	sampledata
Port Number	9001
User Name	pentaho_admin

Field	Setting
Password	password (If password does not work, please check with your system administrator.)

- Click Test to verify your entries are correct. A success message appears. Click OK.

Note: If you get an error when testing your connection, ensure that you have provided the correct settings information as described in the table and that the sample database is running. See the *Install Pentaho Data Integration and Analytics* document for information about how to start the Pentaho Server.

- Click OK to exit the Database Connections window.

### Define the Data Definition Language (DDL)

DDLs are the SQL commands that define the different structures in a database such as CREATE TABLE. Fortunately, Pentaho can help you create the necessary DDL.

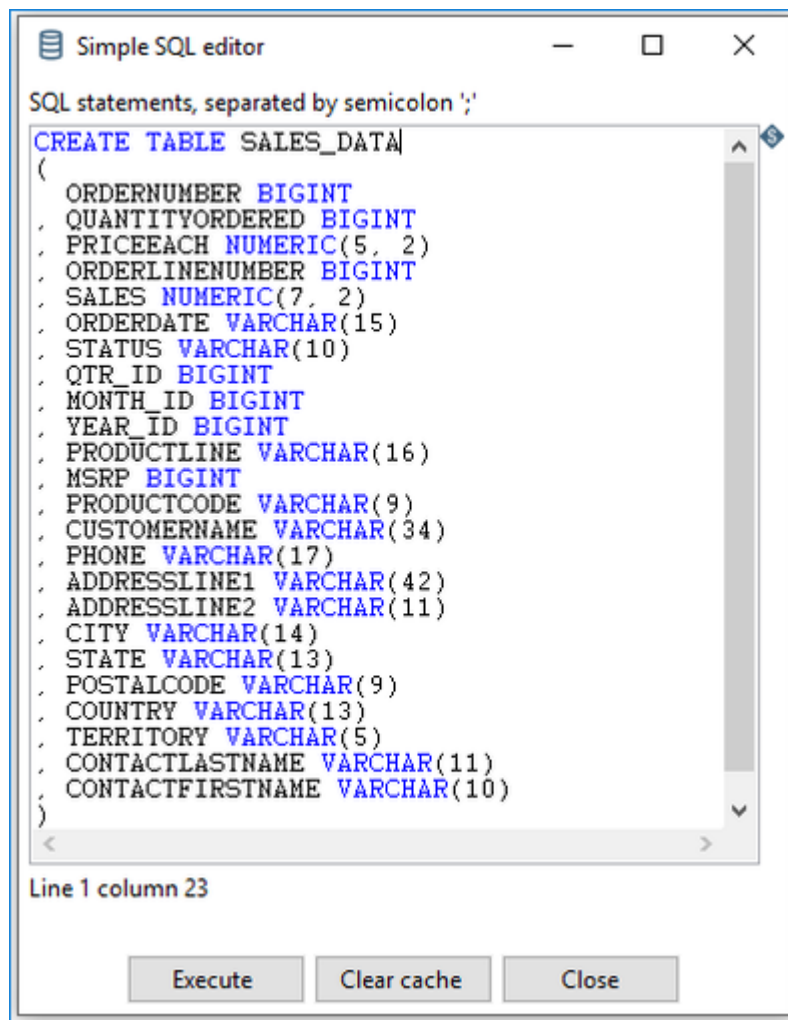
- Enter SALES\_DATA in the Target Table text field.
- This table does not exist in the target database, so Pentaho can generate the DDL to create the table and execute it. In this scenario, the DDL is based on the stream of data coming from the previous step, which is the Read Sales Data step.
- In the Table Output window, select the Truncate Table property.

The screenshot shows the 'Table output' dialog box with the following settings:

- Step name:** Write to Database
- Connection:** Sample Data
- Target schema:** (empty)
- Target table:** SALES\_DATA
- Commit size:** 1000
- Truncate table:** ☒
- Ignore insert errors:** ☐
- Specify database fields:** ☐
- Main options tab:**
  - Partition data over tables:** ☐
    - Partitioning field: (empty)
  - Partition data per month:** ☒
    - Partition data per day: ☐
  - Use batch update for inserts:** ☒
  - Is the name of the table defined in a field?** ☐
    - Field that contains name of table: (empty)
    - Store the tablename field: ☒
  - Return auto-generated key:** ☐
    - Name of auto-generated key field: (empty)

Buttons at the bottom: Help, OK, Cancel, SQL.

- Click the SQL button in the bottom of the Table output dialog box to generate the DDL for creating your target table.
- The Simple SQL editor window appears with the SQL statements needed to create the table.



6. Click Execute to execute the SQL statement.  
The Results of the SQL statements window appears.
7. Examine the results, then click OK to close the Results of the SQL statements window.
8. Click Close in the Simple SQL editor window
9. Click OK to close the Table output window.
10. Save your transformation.

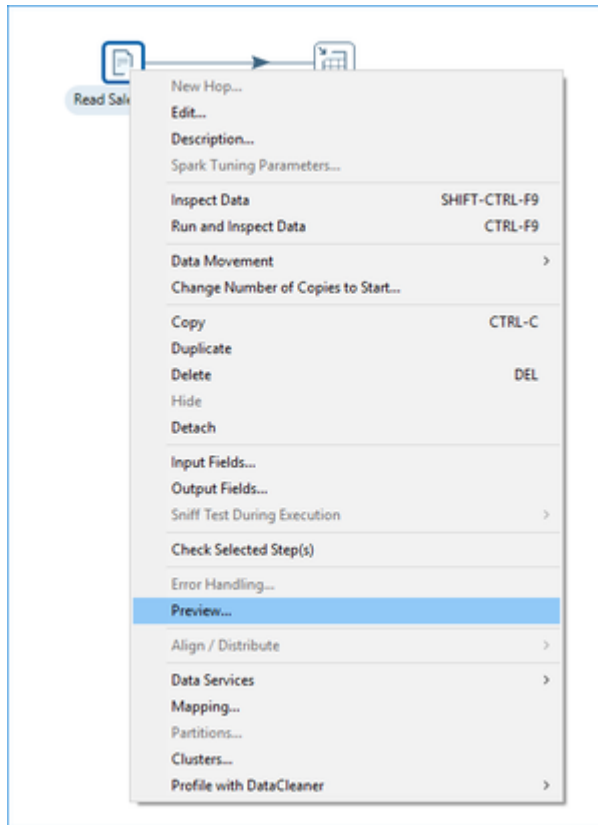
## Step 2: Filter for missing codes

After completing [Step 1: Extract and load data](#), you are ready to add a transformation component to your data pipeline. The source file contains several records that are missing postal codes. This section of the tutorial filters out those records that have missing postal codes, where the POSTALCODE is not null (the true condition), and ensures that only complete records are loaded into the database table.

### Preview the rows read by the input step

Follow these steps to preview the rows read by the input step.

1. Right-click on the Read Sales Data step and select Preview.



- Specify the number of rows to preview. Optionally, you can configure break-points which pause execution based on a defined condition, such as a field having a specific value or exceeding a threshold.
- Click the Quick Launch button. Preview the data and notice that several of the input rows are missing values for the POSTALCODE field.

Examine preview data

Rows of step: Read Sales Data (1000 rows)

ID	YEAR_ID	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME	PHONE	ADDRESSLINE1	ADDRESSLINE2	CITY	STATE	POSTALCODE	COUNTRY	TERRITORY
2	2003	Motorcycles	95	110,1678	Land of Toys Inc.	212337818	887 Long Airport Avenue		NYC	NY	10022	United States	NA
5	2003	Motorcycles	95	110,1678	Rams Collectables	26-43-1555	39 rue de l'Admiral		Paris		91100	France	EMEA
7	2003	Motorcycles	95	110,1678	Sgtm Souvenirs	+33 1 48 42 7055	27 rue du Colonel Pierre Aude		Paris		75008	France	EMEA
8	2003	Motorcycles	95	110,1678	ToysGreenToys.com	628037263	78834 Hillside Dr.		Pasadena	CA	90003	USA	NA
10	2003	Motorcycles	95	110,1678	Corporate Gift Ideas Co.	850033108	7754 Strong St.		San Francisco	CA		United States	NA
10	2003	Motorcycles	95	110,1678	Technica Stores Inc.	850039808	9408 Furth Circle		Burlingame	CA	94017	USA	NA
11	2003	Motorcycles	95	110,1678	Dandelion Designs Imports	20-18-1555	154, chemin de la Tourneil		Lille		59000	France	EMEA
11	2003	Motorcycles	95	110,1678	Hankku Gifts	+47 2287 3215	Drammens 121, Rte 744 Sentrum		Bergen		N 3004	Norway	EMEA
12	2003	Motorcycles	95	110,1678	Mini Wheels Co.	8500355787	3557 North Pendale Street		San Francisco	CA		United States	NA
1	2004	Motorcycles	95	110,1678	Auto Canal- Petit	(31 47 53 8333	25, rue Lorient		Paris		75016	France	EMEA
2	2004	Motorcycles	95	110,1678	Australian Collectors, Co.	61 9520 4555	638 St Kilda Road		Melbourne	Victoria	3004	Australia	APAC
4	2004	Motorcycles	95	110,1678	Vitachrome Inc.	212337818	2875 Kingston Rd.		NYC	NY	10022	USA	NA
5	2004	Motorcycles	95	110,1678	Taken Collectables Inc.	201939090	7478 Moss Rd.		Newark	NJ	08019	USA	NA
6	2004	Motorcycles	95	110,1678	Gift Depot Inc.	205552570	25503 South Bay Ln.		Bridgewater	CT	07562	United States	NA
7	2004	Motorcycles	95	110,1678	La Rochelle Gifts	40-67-8555	67, rue des Cinqante Oranges		Nantes		44000	France	EMEA
8	2004	Motorcycles	95	110,1678	Marlin's Replicas Co.	6175038555	38023 Spinnaker Dr.		Cambridge	MA	02147	USA	NA
9	2004	Motorcycles	95	110,1678	Toys of Finland, Co.	95-224 8555	Kesäkuja 45		Helsinki		00100	Finland	EMEA
10	2004	Motorcycles	95	110,1678	Beane Mini Imports	07-68 9555	Erling Skakke gate 78		Stavem		4110	Norway	EMEA
11	2004	Motorcycles	95	110,1678	Orcaat Classics Inc.	2155511555	7388 Frampton St.		Albany	NY	12207	USA	NA
11	2004	Motorcycles	95	110,1678	Land of Toys Inc.	212337818	887 Long Airport Avenue		NYC	NY	10022	United States	NA
11	2004	Motorcycles	95	110,1678	Selburg Collectables	6362-9555	Giesweg 14		Selburg		5020	Austria	EMEA
12	2004	Motorcycles	95	110,1678	Souvenirs And Things Co.	+61 2 8495 8335	Monter Money Building, 815 Pacific Hwy		Chatswood	NSW	2067	Australia	APAC
2	2005	Motorcycles	95	110,1678	La Rochelle Gifts	40-67-8555	67, rue des Cinqante Oranges		Nantes		44000	France	EMEA
3	2005	Motorcycles	95	110,1678	FunGifts.com	508552555	1703 First Street		New Bedford	MA	01905	USA	NA
4	2005	Motorcycles	95	110,1678	UK Collectables Ltd.	0770 335-2282	Berkley Gardens 12, Brewery		Liverpool		063 1ELT	UK	EMEA

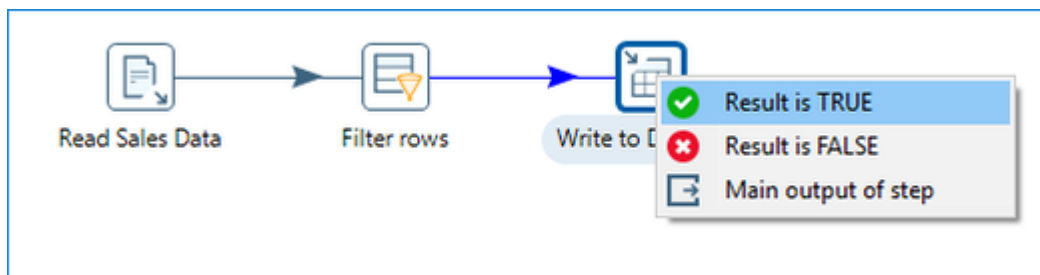
Close Stop Get more rows

- Click Stop on the preview window to end the preview.

### Separate the records with missing postal codes

Follow these instructions to use the Filter Rows transformation step to separate out those records missing postal codes. These records are resolved later in the tutorial.

- Add a Filter Rows step to your transformation. Under the Design tab, select Flow > Filter Rows.
- Insert your Filter Rows step between your Read Sales Data step and your Write to Database step.
  - Right-click and delete the hop between the Read Sales Data step and Write to Database steps.
  - Create a hop between the Read Sales Data step and the Filter Rows step. Create a hop by clicking on the step, hold the SHIFT key down and click-and-drag to draw a line to the next step.
  - Create a hop between the Filter Rows step and Write to Database step.
  - In the dialog box that appears, select Result is TRUE.



3. Double-click the Filter Rows step. The Filter Rows window appears.
4. In the Step Name field, enter Filter Missing Zips.
5. Click in The condition field to open the Fields window. The available conditions appear.
6. In the Fields window select POSTALCODE and click OK.
7. Click the comparison operator field, which is set to = by default. The Functions window appears.
8. Select IS NOT NULL from the list of functions, and then click OK to close the Functions window.

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

9. Click OK to exit the Filter Rows window.
- Note: You will return to this step later to configure the Send true data to step and Send false data to step settings after adding their target steps to your transformation.
10. Save your transformation.

## Step 3: Resolve missing data

After completing [Step 2: Filter for missing codes](#), you are ready to resolve the missing postal codes. In this section, you will learn how to use a second text file containing a list of cities, states, and postal codes, to look up the postal codes for those records in which the fields are missing, which is the false branch of your Filter rows step.

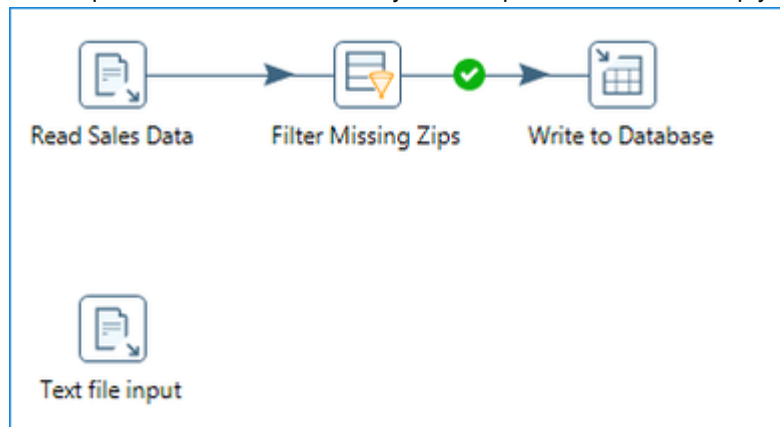
First, you will use a Text file input step to read from the source file. Then, you will use a Stream lookup step to bring the resolved postal codes into the stream. Lastly, you will use the Select values step to rename fields on the stream, remove unnecessary fields, and more.

### Retrieve data from your lookup file

Follow these steps to retrieve data from your lookup file.

1. Add a new Text File Input step to your transformation.

This step retrieves the records from your lookup file. Do not add a hop yet.



2. Open the Text File Input step window, then enter Read Postal Codes in the Step name property.
3. Click Browse to navigate to the Zipssortedbycitystate.csv source file located in the directory ...\\design-tools\\data-integration\\samples\\transformations\\files.
4. Change File type to \*.csv, select Zipsortedbycitystate.csv, and click OK.  
The path to the source file appears in the File or directory field.
5. Click Add.  
The path to the file appears under Selected files.

### View the contents of the sample file

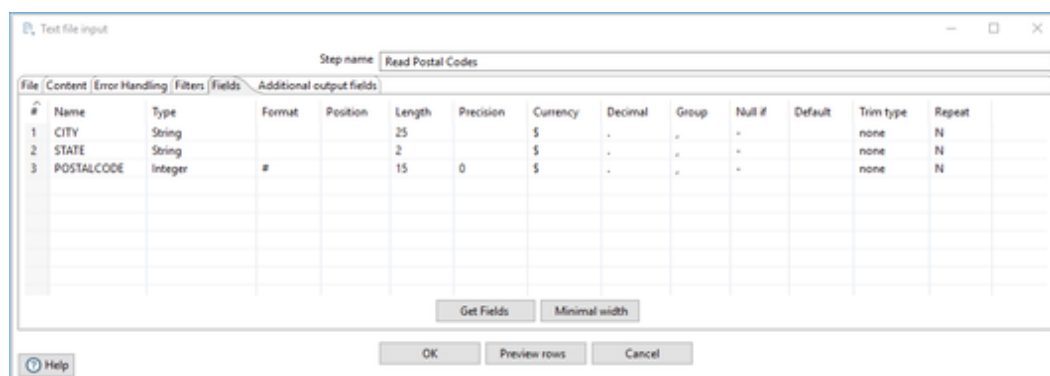
Follow these steps to view the contents of the sample file.

1. Click the Content tab, then set the Format field to Unix.
2. Click the File tab again and click the Show file content near the bottom of the window.
3. The Number of lines(0=all lines) window appears. Click the OK button to accept the default.
4. The Content of first file window displays the file. Examine the file to see how that input file is delimited, what enclosure character is used, and whether or not a header row is present. In the example, the input file is comma (,) delimited and the enclosure character is the quotation mark ("). A single header row contains field names.
5. Click Close to close the window.

### Edit and save the transformation

Follow these steps to edit and save your transformation.

1. In the Content tab, change the Separator character to a comma (,) and confirm that the Enclosure setting is a quotation mark ("). Verify that the Header option is selected.
2. Under the Fields tab, click Get Fields to retrieve the data from your CSV file.
3. The Number of lines to sample window appears. Enter 0 in the field, then click OK.



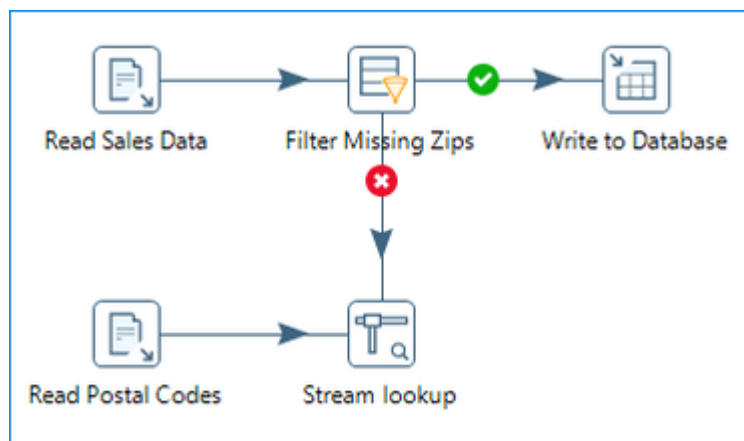
4. If the Scan Result window displays, click Close to close it.
5. Click Preview rows to verify that your entries are correct.
  - a. When prompted to enter the preview size, click OK.
  - b. Review the information in the window, then click Close.
6. Click OK to exit the Text File input window.

7. Save the transformation.

## Resolve missing zip code information

Follow these steps to resolve the missing postal code information.

1. Add a Stream Lookup step to your transformation by clicking the Design tab, expanding the Lookup folder, then selecting Stream Lookup.
2. Draw a hop from the Filter Missing Zips to the Stream lookup step. In the dialog box that appears, select Result is FALSE.
3. Create a hop from the Read Postal Codes step to the Stream lookup step.



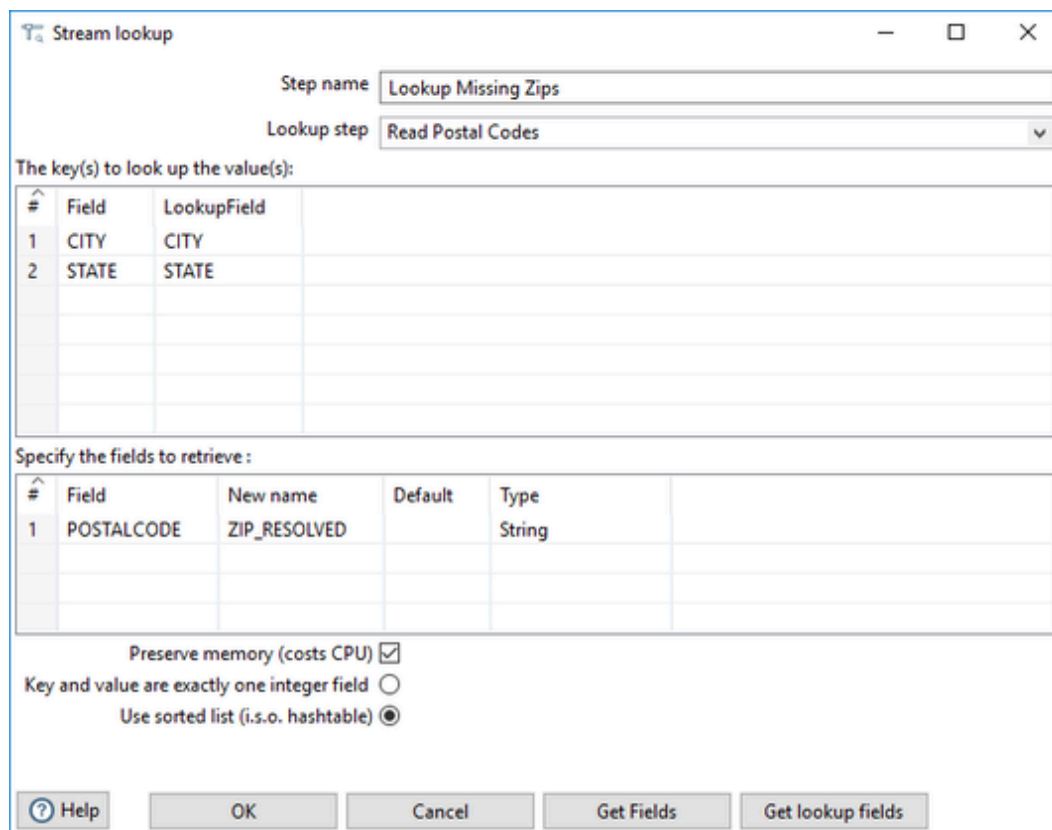
4. Double-click the Stream lookup step to open the Stream Value Lookup window.
5. Rename Stream Lookup to Lookup Missing Zips.
6. From the Lookup step drop-down box, select Read Postal Codes as the lookup step. Perform the following:
  - a. In the key(s) to look up the value(s) table, define the CITY and STATE fields .
  - b. In row #1, open the drop-down menu in the Field column and select CITY.
  - c. Click in the LookupField column and select CITY.
  - d. In row #2, open the drop-down menu in the Field column and select STATE.
  - e. Click in the LookupField column and select STATE.

#	Field	LookupField
1	CITY	CITY
2	STATE	STATE

#	Field	New name	Default	Type
1				

7. Click Get Lookup Fields to pull the three fields from the Read Postal Code step.
8. POSTALCODE is the only field you want to retrieve. To delete the CITY and STATE lines, right-click in the line and select Delete Selected Lines.

9. In the New Name field, change the name POSTALCODE to ZIP\_RESOLVED and verify that Type is set to String.
10. Select Use sorted list (i.s.o. hashtable).



The Stream lookup dialog box is shown with the following configuration:

- Step name:** Lookup Missing Zips
- Lookup step:** Read Postal Codes
- The key(s) to look up the value(s):**

#	Field	LookupField
1	CITY	CITY
2	STATE	STATE
- Specify the fields to retrieve:**

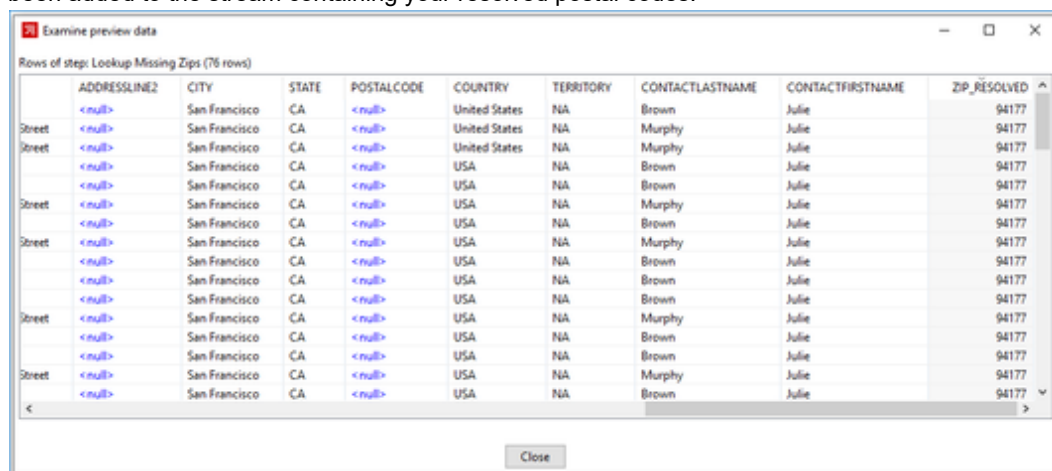
#	Field	New name	Default	Type
1	POSTALCODE	ZIP_RESOLVED		String
- Options:**
  - Preserve memory (costs CPU) ☒
  - Key and value are exactly one integer field ☐
  - Use sorted list (i.s.o. hashtable) ☒
- Buttons:** Help, OK, Cancel, Get Fields, Get lookup fields

11. Click OK to close the Stream Value Lookup edit properties dialog box.
12. Save your transformation.

### Preview your transformation

Follow these steps to preview your transformation.

1. To preview the data, select and right-click the Lookup Missing Zips step. From the menu that appears, select Preview.
2. In the Transformation debug dialog window, click Quick Launch to preview the data flowing through this step.
3. In the Examine preview data window that appears, note that the new field, ZIP\_RESOLVED, has been added to the stream containing your resolved postal codes.



The Examine preview data window shows the following data for the 'Lookup Missing Zips' step (76 rows):

	ADDRESSLINE2	CITY	STATE	POSTALCODE	COUNTRY	TERRITORY	CONTACTLASTNAME	CONTACTFIRSTNAME	ZIP_RESOLVED
	<null>	San Francisco	CA	<null>	United States	NA	Brown	Julie	94177
Street	<null>	San Francisco	CA	<null>	United States	NA	Murphy	Julie	94177
Street	<null>	San Francisco	CA	<null>	United States	NA	Murphy	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
Street	<null>	San Francisco	CA	<null>	USA	NA	Murphy	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
Street	<null>	San Francisco	CA	<null>	USA	NA	Murphy	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
Street	<null>	San Francisco	CA	<null>	USA	NA	Murphy	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
Street	<null>	San Francisco	CA	<null>	USA	NA	Murphy	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177
Street	<null>	San Francisco	CA	<null>	USA	NA	Murphy	Julie	94177
	<null>	San Francisco	CA	<null>	USA	NA	Brown	Julie	94177

4. Click Close to close the window.
5. If the Select the preview step window appears, click Close.

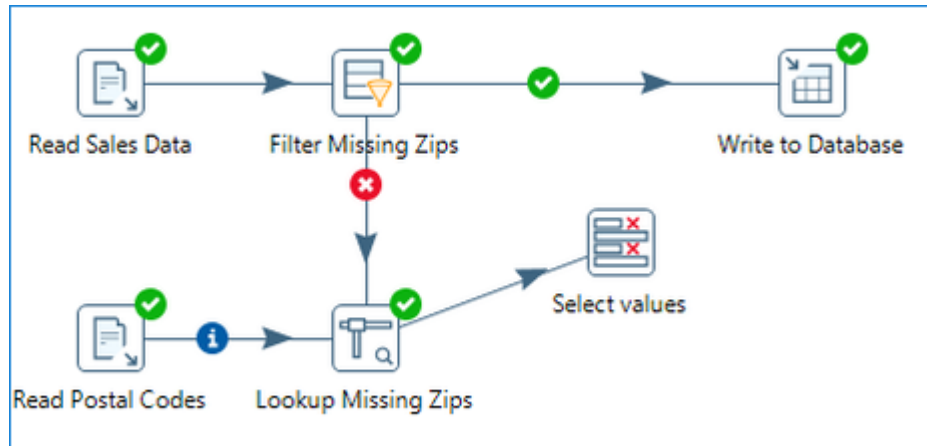
The execution results near the bottom of the PDI window display updated metrics in the Step Metrics tab.



## Apply formatting to your transformation

Follow these steps to clean up the field layout on your lookup stream so that it matches the format and layout of the other stream going to the Write to Database step.

1. Add a Select Values step to your transformation by expanding the Transform folder and clicking Select Values.
2. Create a hop from the Lookup Missing Zips to the Select Values step.



3. Double-click the Select Values step to open its properties dialog box.
4. Rename the Select Values step to Prepare Field Layout.
5. Click Get fields to select to retrieve all fields and begin modifying the stream layout.
6. In the Fields list, find the # column and click the number for the ZIP\_RESOLVED field.  
Use CTRL UP (MacOS, COMMAND UP ) to move ZIP\_RESOLVED just below the POSTALCODE field, which is the one that still contains null values.

Select values

Step name: Prepare Field Layout

Select & Alter Remove Meta-data

Fields:

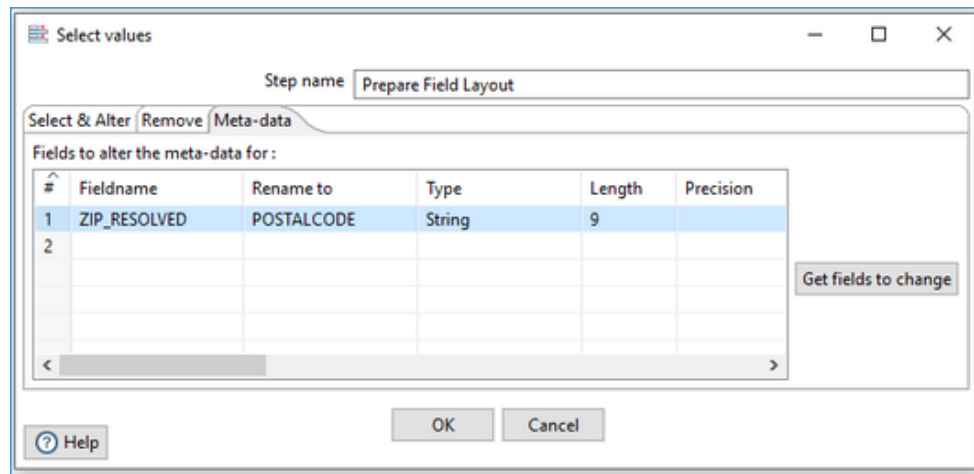
#	Fieldname	Rename to	Length	Precision
1	ORDERNUMBER			
2	QUANTITYORDERED			
3	PRICEEACH			
4	ORDERLINENUMBER			
5	SALES			
6	ORDERDATE			
7	STATUS			
8	QTR_ID			
9	MONTH_ID			
10	YEAR_ID			
11	PRODUCTLINE			
12	MSRP			
13	PRODUCTCODE			
14	CUSTOMERNAME			
15	PHONE			
16	ADDRESSLINE1			
17	ADDRESSLINE2			
18	CITY			
19	STATE			
20	POSTALCODE			
21	ZIP_RESOLVED			
22	COUNTRY			
23	TERRITORY			
24	CONTACTLASTNAME			
25	CONTACTFIRSTNAME			

Get fields to select Edit Mapping

Include unspecified fields, ordered by ☐

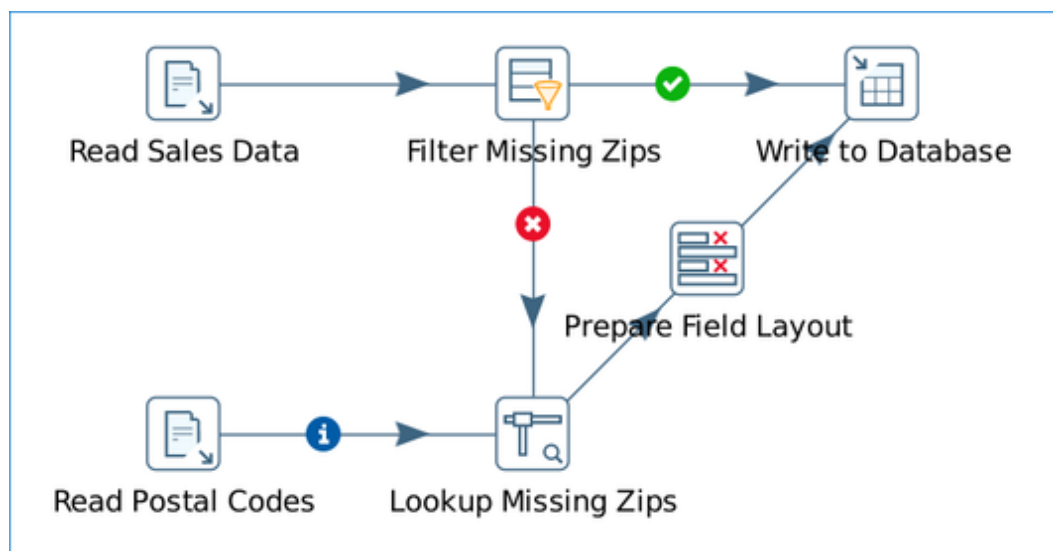
Help OK Cancel

7. Select the old POSTALCODE field in the list (line 20), right-click in the line, and select Delete Selected Lines
8. The original POSTALCODE field was formatted as a 9-character string. You must modify your new field to match the form. Click the Meta-Data tab.
9. In the first row of the Fields to alter table the meta-data for section, click in the Fieldname column and select ZIP\_RESOLVED. Perform the following steps:
  - a. Enter POSTALCODE in the Rename to column.
  - b. Select String in the Type column and enter 9 in the Length column.



c. Click OK to exit the edit properties dialog box.

10. Draw a hop from the Prepare Field Layout (Select values) step to the Write to Database (Table output) step.
11. When prompted, select the Main output of the step option.
12. Save your transformation.



## Step 4: Clean the data

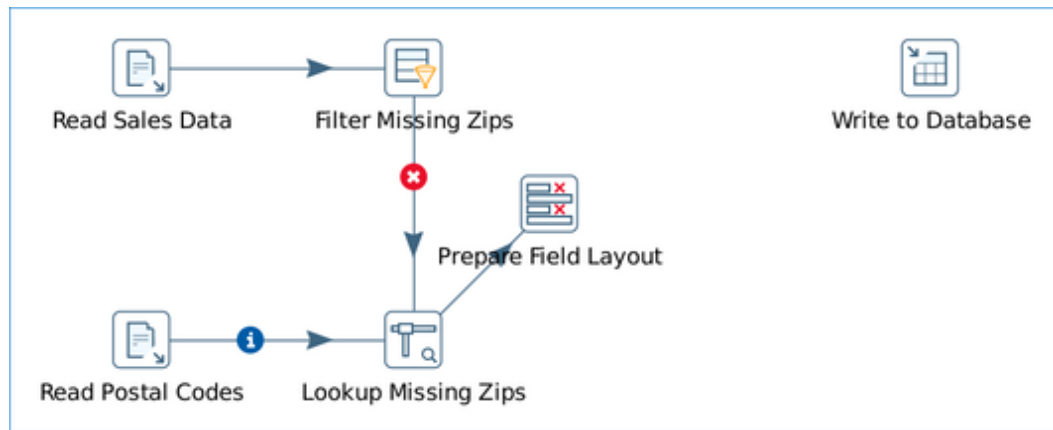
After completing [Step 3: Resolve missing data](#), you can further cleanse and categorize the data into buckets before loading it into a relational database. In this section, you will cleanse the COUNTRY field data by mapping United States to USA using the Value mapper step. Cleaning the data ensures there is only one version of USA.

In addition, you will learn how to use buckets for categorizing the SALES data into small, medium, and large categories using the Number range step. You will learn how to insert these cleaning and categorizing functions into your transformation just prior to the Write to Database step on the canvas.

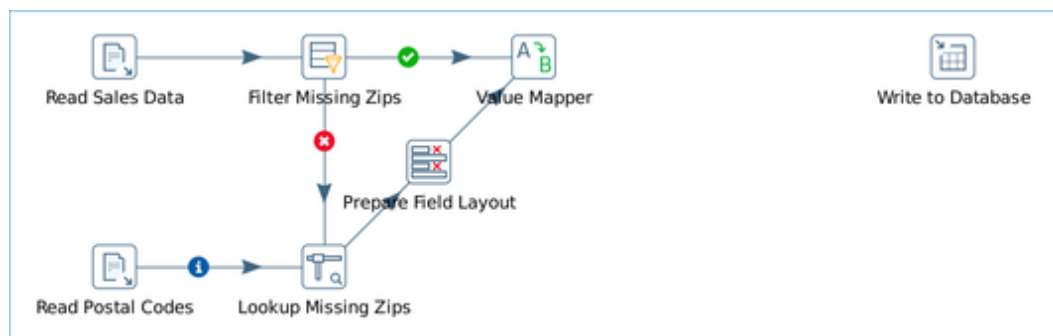
### Add a Value mapper step to the transformation

Follow these steps to add the Value mapper step to the transformation.

1. Delete both hops connected to the Write to Database step. For each hop, right-click and select Delete.
2. Create some extra space on the canvas. Drag the Write to Database step toward the right side of your canvas.



3. Add the Value mapper step to your transformation by expanding the Transform folder and choosing Value mapper.
4. Create a hop between the Filter Missing Zips and Value mapper steps. In the dialog box that appears, select Result is TRUE.
5. Create a hop between the Prepare Field Layout and Value mapper steps. When prompted, select the Main output of the step option.



### Set the properties in the Value Mapper step

Follow these steps to set the properties in the Value mapper step.

1. Double-click the Value mapper step to open its properties dialog box.
2. Click in the Fieldname to use field and select COUNTRY.
3. In the Field Values table, define the United States and USA field values.
  - a. In row #1, click the field in the Source value column and enter United States
  - b. Then, click the field in the Target value column and enter USA

Value mapper

Step name: Value mapper

Fieldname to use: COUNTRY

Target field name (empty=overwrite):

Default upon non-matching:

Field values:

#	Source value	Target value
1	United States	USA

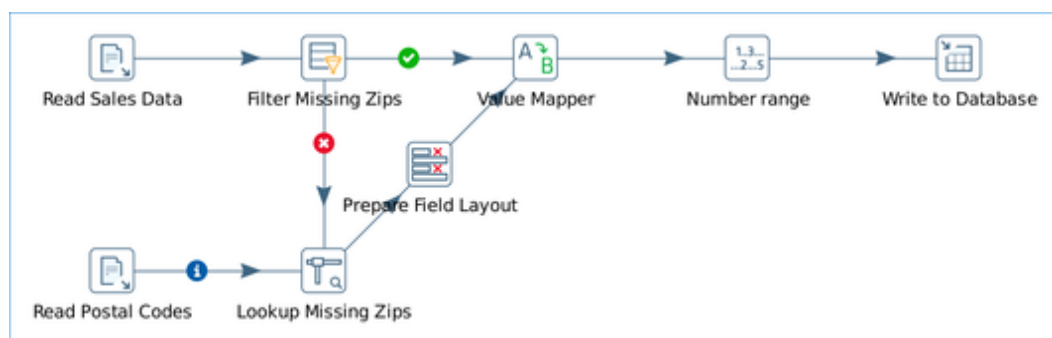
Help OK Cancel

4. Click OK.
5. Save your transformation.

### Apply ranges

Follow these steps to apply ranges to your transformation.

1. Add a Number range step to your transformation by expanding the Transform folder and selecting Number range.
2. Create a hop between the Value mapper and Number range steps.
3. Create a hop between the Number range and Write to Database (which was built using Table output) steps. When prompted, select the Main output of the step option.



4. Double-click the Number range step to open its properties dialog box.
5. Click in the Input field and select SALES from the list.
6. In the Output field enter DEALSIZE.
7. In the Ranges (min  $\leq$  x < max) table, define the Lower Bound and Upper Bound field ranges along with the bucket Value.
  - a. In row #1, click the field in the Upper Bound column and enter 3000.0. Then, click the field in the Value column and enter Small.
  - b. In row #2, click the field in the Lower Bound column and enter 3000.0. Then, click the field in the Upper Bound column and enter 7000.0. Click the field in the Value column and enter Medium.
  - c. In row #3, click the field in the Lower Bound column and enter 7000.0. Then, click the field in the Value column and enter Large.

**Number range**

Step name: Number range

Input field: SALES

Output field: DEALSIZE

Default value(if no range matches): unknown

Ranges (min  $\leq$  x < max):

#	Lower Bound	Upper Bound	Value
1		3000.0	Small
2	3000.0	7000.0	Medium
3	7000.0		Large

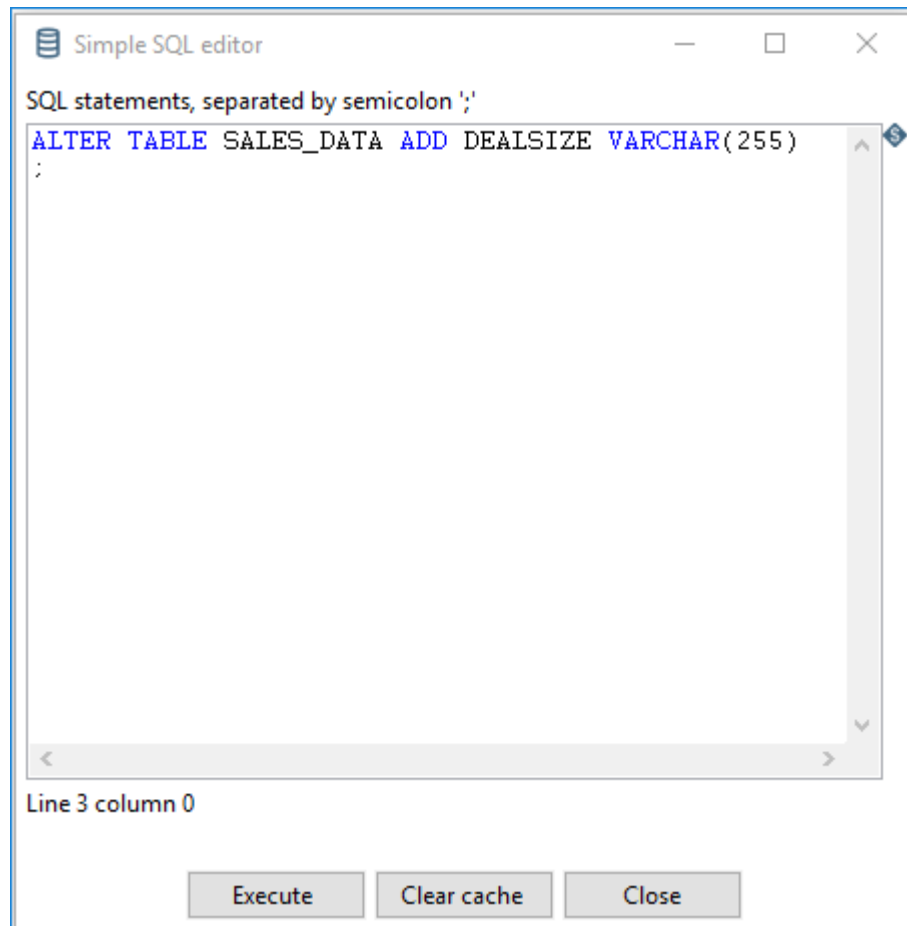
Help OK Cancel

8. Click OK.

### Execute the SQL statement

Your database table does not yet contain the field DEALSIZE. Perform these steps to execute the SQL statement.

1. Double-click the Write to Database step to open its properties dialog box.
2. Click the SQL button at the bottom of the window to generate the new DDL for editing your original target table. Note that the Write to Database step was built using Table output.
  - a. The Simple SQL editor window appears with the SQL statements needed to alter the table.



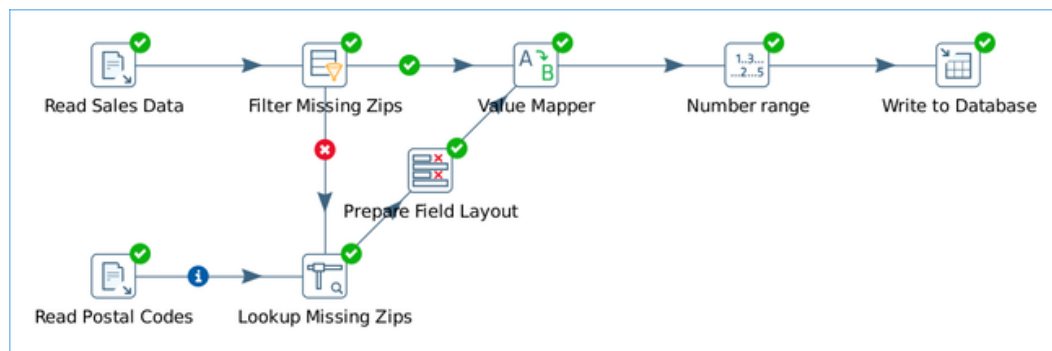
- b. Click Execute to execute the SQL statement.
  - c. The Results of the SQL statements window appears. Examine the results, then click OK to close the window.
  - d. Click Close in the Simple SQL editor window to close it.
  - e. Click OK to close the Write to Database window. Note that the Write to Database step was built using Table output
3. Save your transformation.

## Step 5: Run the transformation

Pentaho Data Integration provides a number of deployment options. The *Running a Transformation* section in the *Pentaho Data Integration* document explains these and other options available for execution. In this section of the tutorial, you create a transformation using the Local run option.

1. In the PDI client window, select Action > Run.  
The Run Options window appears.
2. Keep the default Pentaho local option for this exercise.  
It uses the native Pentaho engine and runs the transformation on your local machine. See the *Pentaho Data Integration* document if you are interested in setting up configurations that use another engine.
3. Click Run.

The transformation executes.



After the transformation runs, the Execution Results panel opens below the canvas.

### Viewing the execution results

Use the tabs in the Execution Results section of the window to view how the transformation executed, pinpoint errors, and monitor performance.

#### Step Metrics

Provides statistics for each step in your transformation including how many records were read, written, or caused an error, as well as processing speed (rows per second) and more. This tab also indicates whether an error occurred in a transformation step.

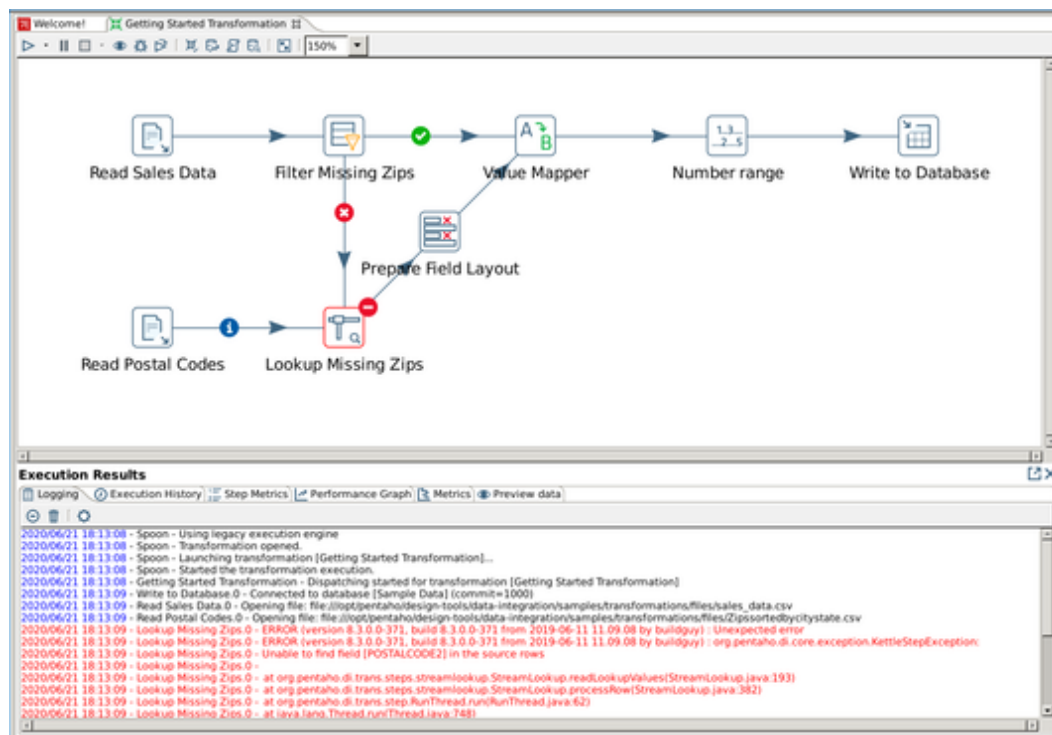
This tutorial introduces no intentional transformation errors, so the transformation should run correctly. If a mistake does occur, you can view the steps that caused the transformation to fail highlighted in red. In the example below, the Lookup Missing Zips step caused an error.

**Execution Results**

Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	Input/output
1 Read Sales Data	0	0	993	995	0	1	0	0	Stopped	0.1s	13.267	-
2 Read Postal Codes	0	0	9905	9906	0	1	0	0	Stopped	0.1s	137.583	-
3 Filter Missing Zips	0	1	0	0	0	0	0	0	Stopped	0.1s	10	-
4 Lookup Missing Zips	0	1	0	0	0	0	0	1	Stopped	0.1s	16	-
5 Prepare Field Layout	0	0	0	0	0	0	0	0	Stopped	0.1s	0	-
6 Value Mapper	0	0	0	0	0	0	0	0	Stopped	0.1s	0	-
7 Number range	0	0	0	0	0	0	0	0	Stopped	0.1s	0	-
8 Write to Database	0	0	0	0	0	0	0	0	Stopped	0.1s	0	-

#### Logging

Displays the logging details for the most recent execution of the transformation. It also allows you to drill deeper to determine where errors occur. Error lines are highlighted in red. In the example below, the Lookup Missing Zips step caused an error because it attempted to look up values on a field called POSTALCODE2 which did not exist in the lookup stream.



### Execution History

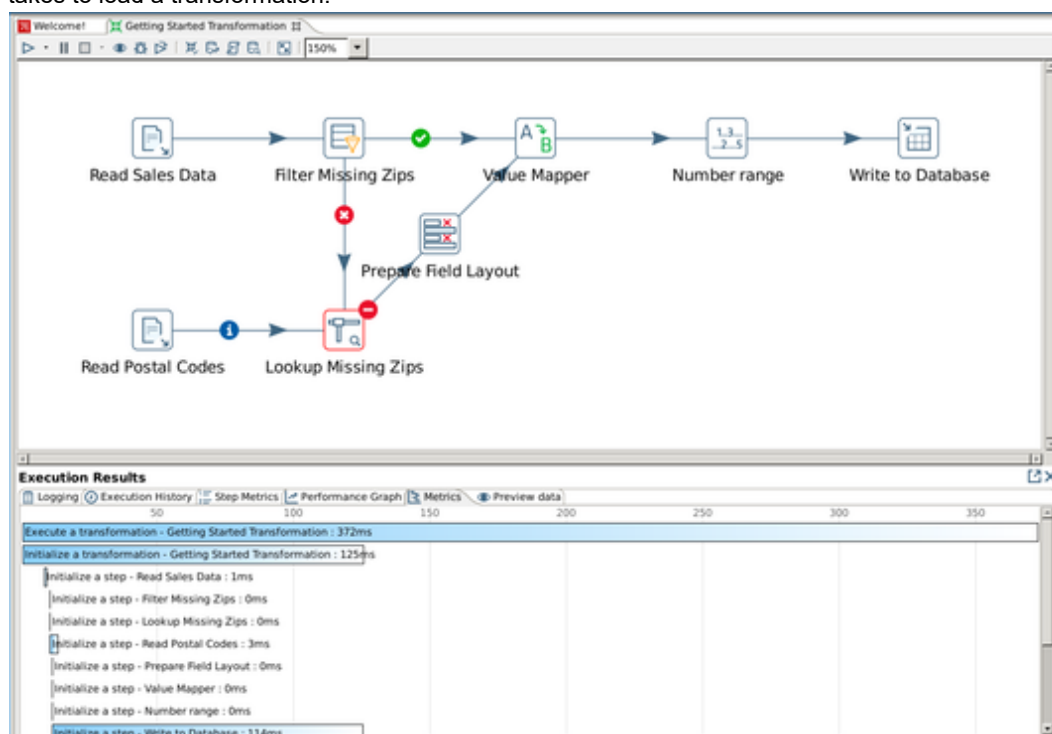
Provides access to the step metrics and log information from previous executions of the transformation. This feature works only if you have configured your transformation to log to a database through the Logging tab of the Transformation Settings dialog box.

### Performance Graph

Analyzes the performance of steps based on a variety of metrics including how many records were read, written, or caused an error, as well as processing speed (rows per second) and more. Like the execution history, this feature requires you to configure your transformation to log to a database through the Logging tab found in the Transformation Settings dialog box.

### Metrics tab

Displays a Gantt chart after the transformation or job runs. This information includes how long it takes to connect to a database, how much time is spent executing a SQL query, or how long it takes to load a transformation.





## Preview Data

Displays a preview of the data.

## Step 6: Orchestrate with jobs

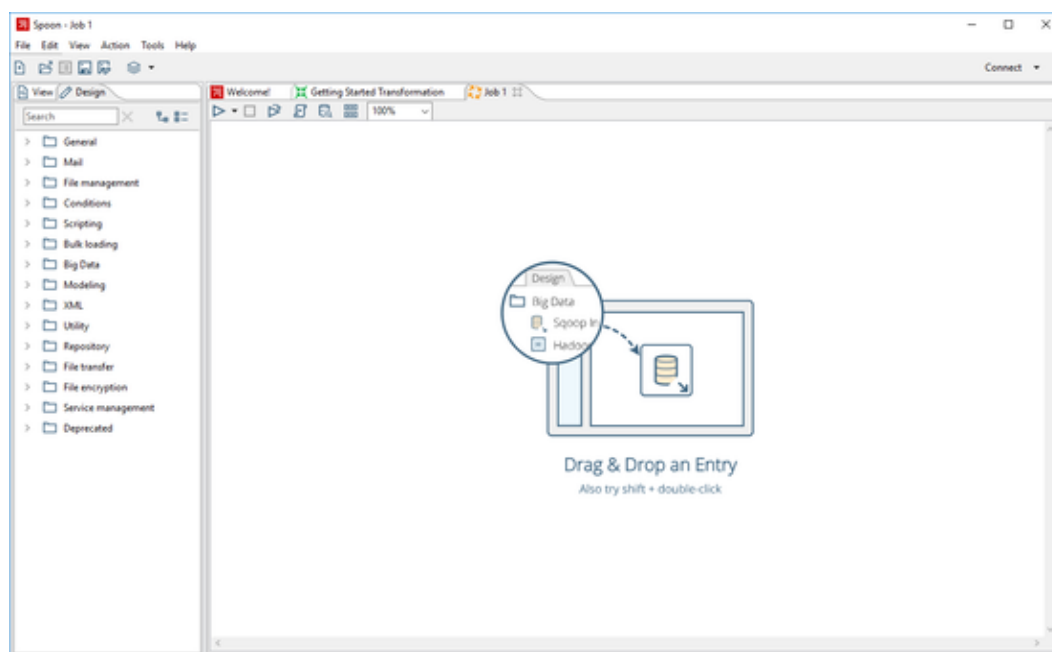
Jobs are used to coordinate ETL activities such as:

- Defining the flow and dependencies that control the linear order for the transformations to run.
- Preparing for execution by checking conditions such as, "Is my source file available?" or "Does a table exist?"
- Performing bulk load database operations.
- Assisting file management, such as posting or retrieving files using FTP, copying files, and deleting files.
- Sending success or failure notifications through email.

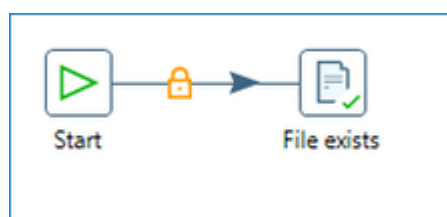
For this part of the tutorial, imagine that an external system is responsible for placing your `sales_data.csv` input in its source location every Saturday night at 9 p.m. You want to create a job that will verify that the file has arrived and then run the transformation to load the records into the database. In a subsequent exercise, you will schedule the job to run every Sunday morning at 9 a.m.

The following steps assume that you have built a Getting Started transformation as described in [Step 1: Extract and load data](#) of the tutorial.

1. Go to File > New > Job.



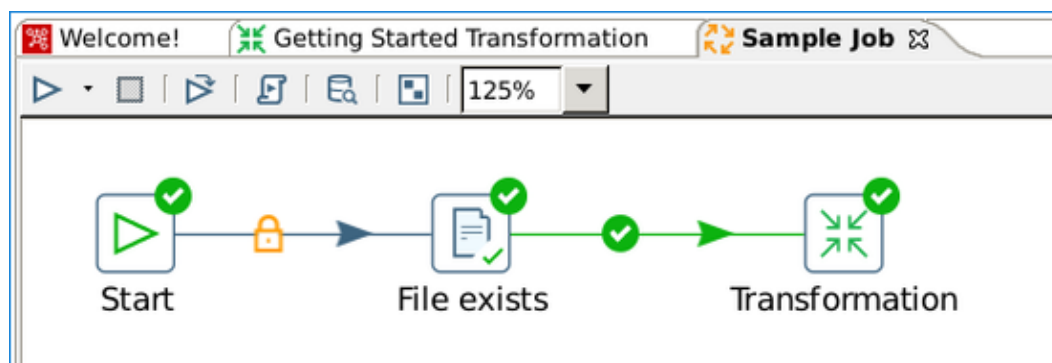
2. Expand the General folder and drag a Start job entry onto the canvas.  
The Start job entry defines where the execution will begin.  
Note: Jobs run in a sequential order of steps and transformations can run in a parallel order of steps.
3. Expand the Conditions folder and add a File Exists job entry.
4. Draw a hop from the Start job entry to the File Exists job entry.



5. Double-click the File Exists job entry to open its properties dialog box. Click Browse and set the filter near the bottom of the window to All Files. Select the `sales_data.csv` from the following

directory: ...\\design-tools\\data-integration\\samples\\transformations\\files.

6. Click OK to exit the Open File window.
7. Click OK to exit the Check if a file exists window.
8. Expand the General folder and add a Transformation job entry.
9. Draw a hop between the File Exists and the Transformation job entries.
10. Double-click the Transformation job entry to open its properties dialog box.
11. Click Browse to open the Select repository object window. Browse to and select the Getting Started transformation.
12. Click OK to close the Transformation window.
13. Save your job as Sample Job.
14. Click Run icon in the toolbar. When the Run Options window appears, select Local environment type and click Run. The Execution Results panel should open showing you the job metrics and log information for the job execution.



## PDI job tutorial

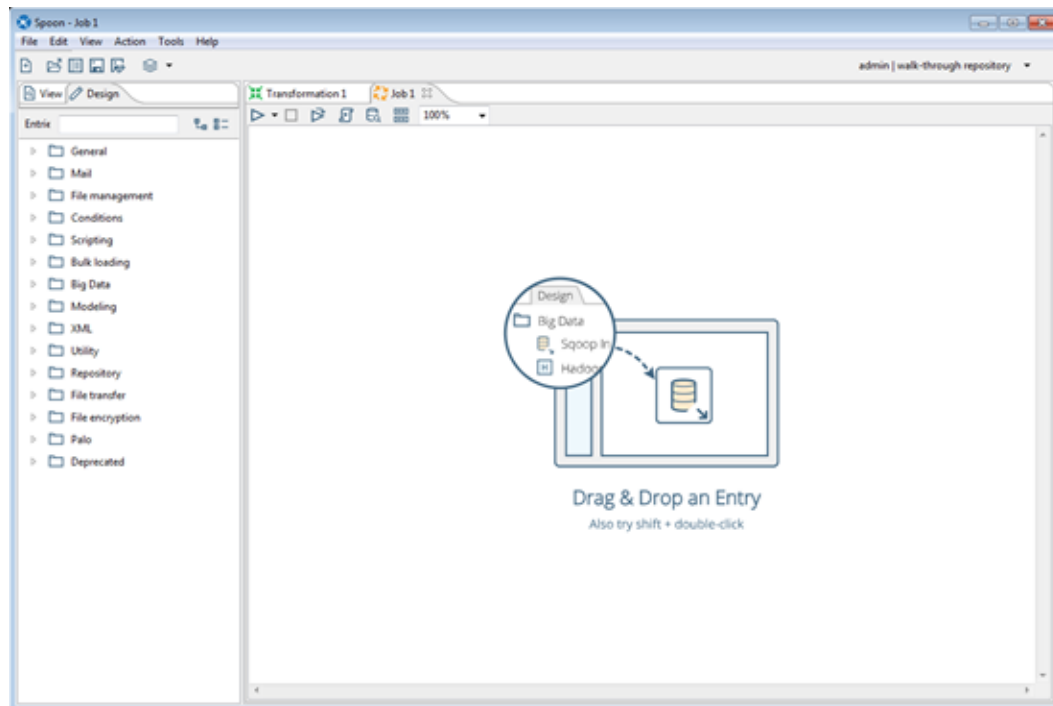
Jobs are used to coordinate ETL activities such as:

- Defining the flow and dependencies for what order transformations should be run.
- Preparing for execution by checking conditions such as, "Is my source file available?" or "Does a table exist?"
- Performing bulk load database operations.
- File management such as posting or retrieving files using FTP, copying files and deleting files.
- Sending success or failure notifications through email.

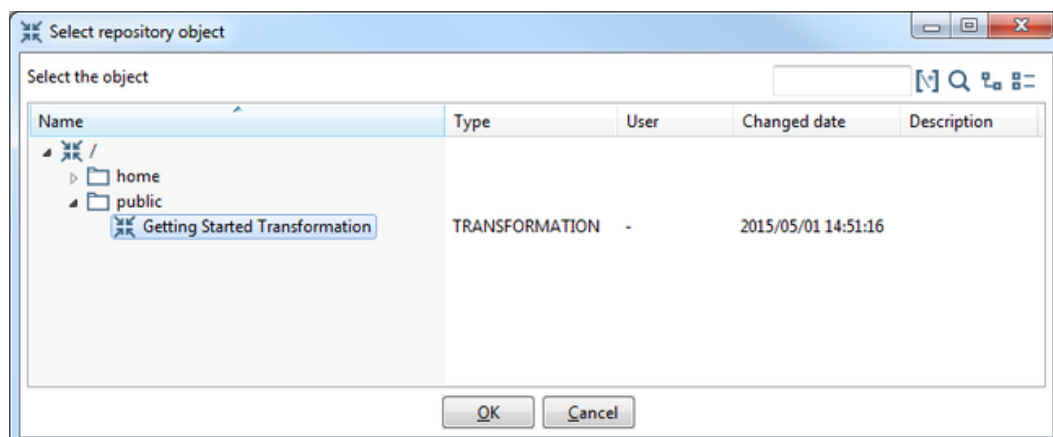
For this exercise, imagine that an external system is responsible for placing your sales\_data.csv input in its source location every Saturday night at 9 p.m. You want to create a job that will check to see that the file has arrived and run your transformation to load the records into the database. In a subsequent exercise, you will schedule the job to be run every Sunday morning at 9 a.m.

To complete this exercise, you must have completed the exercises in the [PDI Transformation Tutorial](#).

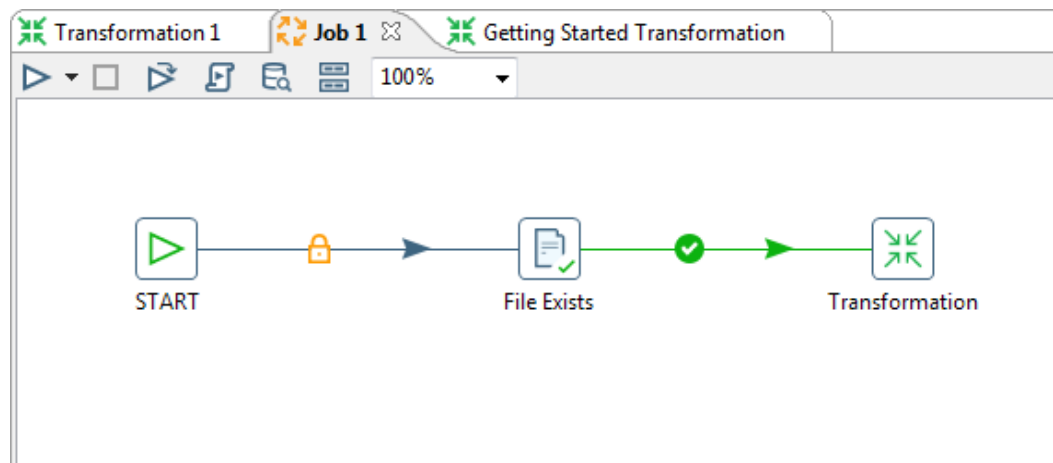
1. Go to File > New > Job.



2. Expand the General folder and drag a Start job entry onto the graphical workspace.  
The Start job entry defines where the execution will begin.
3. Expand the Conditions folder and add a File Exists job entry.
4. Draw a hop from the Start job entry to the File Exists job entry.
5. Double-click the File Exists job entry to open its Edit Properties dialog box. Click Browse and set the filter near the bottom of the window to All Files. Select the sales\_data.csv from the following location: ...\\design-tools\\data-integration\\samples\\transformations\\files.
6. Click OK to exit from the Open File window.
7. Click OK to exit from the Check if a file exists window.
8. In Spoon, expand the General folder and add a Transformation job entry.
9. Draw a hop between the File Exists and the Transformation job entries.
10. Double-click the Transformation job entry to open its edit Properties dialog box.
11. Click Browse to open the Select repository object window. Browse to and select the transformation you created in the [PDI Transformation Tutorial](#).
12. Expand the repository tree to find your sample transformation. Select it and click OK.



13. Save your job as Sample Job.
14. Click Run icon in the toolbar. When the Run Options window appears, choose Local environment type and click Run. The Execution Results panel should open showing you the job metrics and log information for the job execution.



## Getting started with PDI and Hadoop

Pentaho provides a complete big data analytics solution that supports the entire big data analytics process. From big data aggregation, preparation, and integration, to interactive visualization, analysis, and prediction, Pentaho allows you to harvest the meaningful patterns buried in big data stores. Analyzing your big data sets gives you the ability to identify new revenue sources, develop loyal and profitable customer relationships, and run your organization more efficiently and cost effectively.

# Pentaho, big data, and Hadoop

The term big data applies to very large, complex, or dynamic datasets that need to be stored and managed over a long time. To derive benefits from big data, you need the ability to access, process, and analyze data as it is being created. However, the size and structure of big data makes it very inefficient to maintain and process it using traditional relational databases.

Big data solutions re-engineer the components of traditional databases, such as data storage, retrieval, query, processing, and massively scales them.

## Big data overview

Pentaho increases speed-of-thought analysis against even the largest of big data stores by focusing on the features that deliver performance.

- **Instant access:** Pentaho provides visual tools to make it easy to define the sets of data that are important to you for interactive analysis. These data sets and associated analytics can be easily shared with others, and as new business questions arise, new views of data can be defined for interactive analysis.
- **High performance platform:** Pentaho is built on a modern, lightweight, high performance platform. This platform fully leverages 64-bit, multi-core processors and large memory spaces to efficiently leverage the power of contemporary hardware.
- **Extreme-scale, in-memory caching:** Pentaho is unique in leveraging external data grid technologies, such as Infinispan and Memcached to load vast amounts of data into memory so that it is instantly available for speed-of-thought analysis.
- **Federated data integration:** Data can be extracted from multiple sources, including big data and traditional data stores, integrated together and then flowed directly into reports, without needing an enterprise data warehouse or data mart.

## About Hadoop

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single

servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

A Hadoop platform consists of a Hadoop kernel, a [MapReduce](#) model, a distributed file system, and often a number of related projects—such as [Apache Hive](#), [Apache HBase](#), and others.

A Hadoop Distributed File System, commonly referred to as HDFS, is a Java-based, distributed, scalable, and portable file system for the Hadoop framework.

## Big data resources

The following resources may help in understanding big data architecture and components:

- [Pentaho Big Data Analytics Center](#)
- [Apache Hadoop project](#) -- A project that contains libraries that allows for the distributed processing of large data sets across clusters of computers using simple programming models. There are several modules, including the [Hadoop Distributed File System \(HDFS\)](#), which is a distributed file system that provides high-throughput access to application data and [Hadoop MapReduce](#), which is a key algorithm to distribute work around a cluster.
- [Avro](#)—A data serialization system
- [Cassandra](#)—A scalable multi-master database with no single points of failure
- [HBase](#)—A scalable, distributed database that supports structured data storage for large tables
- [Hive](#)—A data warehouse infrastructure that provides data summarization and on-demand querying
- [Pig](#)—A high-level, data-flow language and execution framework for parallel computation
- [ZooKeeper](#)—A high-performance coordination service for distributed applications
- [MongoDB](#)—A NoSQL open source document-oriented database system developed and supported by 10gen
- [Splunk](#) - A data collection, visualization and indexing engine for operational intelligence that is developed by Splunk, Inc.
- [CouchDB](#)—A NoSQL open source document-oriented database system developed and supported by Apache
- [Sqoop](#)—Software for transferring data between relational databases and Hadoop
- [Oozie](#)—A workflow scheduler system to manage Hadoop jobs

# About Pentaho workflows

The route that you take depends on different levels of expertise, what your unique business needs are, the data that you are using, and business interests when it comes to analyzing, reporting, and working with your data. You might find that you travel one track, and then the other, during your evaluation of Pentaho.

All of the products are integrated to work smoothly together, regardless of which track you ultimately choose. We provide specific details within the workflow discussions, however, here are the high-level use cases for each track.

- **Business Analytics (BA) Track:** Great for analysis and reporting. Meant primarily for business users and does not require special skills to successfully use the components involved. This track enables anyone to build Pentaho solutions without using programming or having deep understanding of data structures.
- **Data Integration (DI) Track:** Meant for data design professionals and requires a working knowledge of data structures and modeling, as well as extract, transform, and load (ETL) processes. With this track, you can directly manipulate data from multiple sources, making it scalable and efficient for enterprise-wide analysis and reporting.

Each track has three workflows: one for Evaluation, one for Development, and one for Production.

- **Evaluate and Learn:** If you used the trial download on the Pentaho website and want to get a hands-on feel for the components that are best for your implementation, follow the Evaluation Workflow.
- **Develop Pentaho Solutions:** After you have figured out which components are best for you and how to use them, the Develop Workflow is the process you use to build, change, and test Pentaho solutions until they meet your production requirements.
- **Go Live for Production:** When your solution is working just right, the Go Live Workflow shows how to move your solution from development to production.

## Prepare for the evaluation

This table guides you through the differences between the Business Analytics and Data Integration tracks, and helps you decide which track to follow for evaluation. You may choose to follow one track, and then the other, while you are exploring the software.

Table. Pentaho Track Decision Table

Explore Considerations	Choose Options	
	<a href="#">Business Analytics Evaluation</a>	<a href="#">Data Integration Evaluation</a>
Expertise	<ul style="list-style-type: none"> <li>• No special skills required</li> <li>• Knowledge of business requirements and what reports and analysis should show</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge of business requirements</li> <li>• Understanding of data structures and modeling</li> <li>• Knowledge of extract, transform, and load (ETL) processes</li> </ul>
Data Set Description	<ul style="list-style-type: none"> <li>• Single source of data</li> <li>• Data from multiple sources that have been transformed and joined into a single data mart or warehouse</li> <li>• Small data sets</li> </ul>	<ul style="list-style-type: none"> <li>• Multiple sources of data</li> <li>• Data you want to transform and join in one or more data marts or warehouses</li> <li>• Large to enormously vast data sets</li> </ul>
Reporting Options	Offers a wide variety of visualization and reporting options.	Offers more limited but focused reporting options that help you visualize and analyze data. BA tools can be used to generate reports based on DI-processed data.
Data Storage Types	<ul style="list-style-type: none"> <li>• Relational databases</li> <li>• CSV data sources</li> <li>• SQL queries</li> </ul>	<ul style="list-style-type: none"> <li>• Relational databases</li> <li>• NoSQL or Hadoop databases</li> <li>• Big data of any types</li> <li>• Data from a web service</li> </ul>
Recommendation	Best used by business analysts, managers, report designers, individual business units within an organization or enterprise	Best used by data scientists, data modelers, data integration and ETL developers, individual business units within an organization or enterprise, and enterprise-wide implementations

Now that you have an idea of which track you want to follow for evaluation, choose an evaluation method. This decision table explains the different options for evaluation so you can pick the option that works best for you.

Table. Pentaho Evaluation Decision Table

Explore Considerations	Choose Options		
	<a href="#">Hosted Demo</a>	<a href="#">Custom Prototype</a>	<a href="#">Trial Download</a>
Track	Business Analytics	Business Analytics or Data Integration	Business Analytics or Data Integration

Explore Considerations	Choose Options		
	<a href="#">Hosted Demo</a>	<a href="#">Custom Prototype</a>	<a href="#">Trial Download</a>
Summary	A cloud-based, hands-on, interactive exploration of Business Analytics reports, analysis, visualizations, and dashboards. Here you can see how easy and fun it can be to use Pentaho.	Work with Pentaho analysts and data integration specialists to plan and build a complimentary custom prototype that illustrates what Pentaho can do with your data. A representative will guide you through the entire process.	Using our trial software, tutorials, and documentation, install and configure your own work environment. Then, build a prototype to get a complete Pentaho experience from installation and administration, through creating your first data models and build reports, analysis, dashboards, and data integration ETL transformations.
Data Source	Pentaho sample data in CSV format	Your sample data, including a range of typical data characteristics in CSV format.	Your sample data, including a range of typical data characteristics in the format that you commonly use
Hardware/Software Requirements	Web browser	Varies, depending on your requirements.	One computer that meets the server requirements stated in the <a href="#">Components Reference</a> .
Recommendation	<p>Any evaluator who wants an overview of Business Analytics features.</p> <ul style="list-style-type: none"> <li>Recommended for business analysts and report designers.</li> </ul> <p>We recommend that you try out the <a href="#">Custom Prototype</a> or <a href="#">Trial Download</a> after you do the hosted demo.</p>	<p>All evaluators, particularly any big data or Data Integration evaluators.</p> <ul style="list-style-type: none"> <li>Recommended for evaluators who want to <a href="#">explore Business Analytics and Data Integration features</a> using a subset of their own data.</li> <li>Limited to first-time customers only.</li> </ul>	<p>Any evaluator who wants to independently work with Business Analytics, Data Integration tools, and big data.</p> <ul style="list-style-type: none"> <li>Recommended for evaluators who want to explore Business Analytics and Data Integration features using their own data.</li> <li><a href="#">Technical support</a> is available to help if you have questions.</li> </ul>

## Pentaho Data Integration workflows

Pentaho Data Integration is a robust extract, transform, and load (ETL) tool that you can use to integrate, manipulate, and visualize your data. You can use PDI to import, transform, and export data from multiple data sources, including flat files, relational databases, Hadoop, NoSQL databases, analytic databases, social media streams, and operational stores. You can also use PDI to clean and enrich the data, move data between databases, and to visualize your data.

- [Evaluate and Learn Pentaho Data Integration](#)
- [Develop Your DI Solution](#)
- [Go Live for Production - DI](#)

## Evaluate and learn Pentaho Data Integration (PDI)

As you explore Pentaho Data Integration (PDI), you will be introduced to the major components, watch videos, work through hands-on examples, and read about the different features. Review the documentation and contact Pentaho [sales support](#) if you have questions.

### PDI basics

This section familiarizes you with PDI and introduces you to basic terminology and concepts. Then, you learn how to start and configure Spoon and take a spin through the interface.

- Get a basic understanding of what PDI does.
- View a video that explains how PDI fits into the [Business Analytics Platform](#).
- Read about Pentaho Data Integration architecture in the *Pentaho Data Integration* document.

### Get acquainted with the PDI client

Spoon is the PDI design tool. In this section you will set up Spoon, take a tour of the Spoon interface, and learn about the different Spoon perspectives.

- Check out the [hardware and software requirements](#) for PDI.
- [Download Trial version](#) of the Pentaho Suite and install the software. (The platform includes PDI.)
- Learn how to install PDI only. See the *Install Pentaho Data Integration and Analytics* document for details.
- Configure the Pentaho Server. See the *Install Pentaho Data Integration and Analytics* document for details.
- Start the Pentaho Server. See the *Install Pentaho Data Integration and Analytics* document for details.
- Access the PDI client. See the *Pentaho Data Integration* document for details.
- Tour the PDI client perspective interfaces. See the *Pentaho Data Integration* document for details.
- Read about terminology and basic concepts in the *Pentaho Data Integration* document.

### Build transformations and jobs

Now that your environment is set up and you are familiar with the PDI client, you are ready to build transformations and jobs. Trying the following task may be helpful.

- Create a connection to the Pentaho Repository.
- Work through the exercise on [Creating a Transformation](#) that involves a flat file. Click through the links that are on the bottom of the page to complete the exercise.
- Create a job to execute the transformation.
- Schedule a job to execute the transformation at a later time.
- Review [commonly-used steps and job entries](#).

### Explore Big Data and Streamlined Data Refinery

In this section, you will learn how to use transformation steps to connect to a variety of big data data sources, including Hadoop, NoSQL, and analytical databases such as MongoDB. You can then try working through the detailed, step-by-step tutorials, and peruse the out-of-the-box steps that Spoon provides. Learn how to work with Streamlined Data Refinery. Then, you will have an opportunity to move beyond the basics and learn how to edit transformations and metadata models.

- Watch one of our [Big Data Videos](#).
- Learn how to work with Streamlined Data Refinery. See *Pentaho Data Integration* for details.
- Learn how to auto model using the Build Model. See *Pentaho Data Integration* for details. job entry and how this feature intersects with Analyzer.
- Find out what big data steps are available out-of-the-box, See *Install Pentaho Data Integration and Analytics* for details.



- Find out which Hadoop Distributions are available and how to configure them. See *Install Pentaho Data Integration and Analytics* for details.  
Note: You should already have a cluster set up to perform this task.
- Edit transformations and metadata models. See *Pentaho Data Integration* for details.
- Watch a video about how to use PDI to [blend Big Data](#).

### About Kitchen, Pan, and Carte

Kitchen, Pan, and Carte are command line tools for executing transformations and jobs modeled in the PDI client.

- Use Pan and Kitchen command line tools to work with transformations and jobs
- Use Carte clusters to:
  - Run transformations and jobs on a Carte cluster.
  - Scheduled jobs to run on a remote Carte server.
  - Start or stop Carte from the command line interface or a URL.
  - Run transformations and jobs from the repository on the Carte server

See the *Pentaho Data Integration* document for details on Kitchen, Pan, and Carte.

### Learn More

Now that you have completed an initial evaluation of PDI, dig a little deeper. Find out how to:

- Use newer steps and entries, like Spark Submit. See the *Install Pentaho Data Integration and Analytics* document for details.
- Read about how to turn a transformation into a data service. See the *Pentaho Data Integration* document for details.
- Use the ETL Metadata Injection step. See the *Pentaho Data Integration* document for details.
- Check out our *What's New* document.
- Create other Data Integration solutions. See the *Pentaho Data Integration* document for details.
- Administer PDI. See the *Administer Pentaho Data Integration and Analytics* document for details.
- Integrate with different security protocols, like Pentaho security, LDAP, MSAD, and Kerberos. See the *Administer Pentaho Data Integration and Analytics* document for details.
- Check out our developer center section in the *Administer Pentaho Data Integration and Analytics* document.

## Develop your PDI solution

This workflow helps you to set up and configure the DI development and test environments, then build, test, and tune your Pentaho DI Solution prototype. This process is similar to the Trial Download Evaluation experience, except that you will be completely configuring the Pentaho Server for data integration and working with your own ETL developers.

If you need extra help, Pentaho [professional services](#) is available. The end result of this is to learn DI implementation best practices and deploy your DI solution to a production server. Most development and testing for DI occurs in Spoon.

Before you begin developing your DI solution, we recommend that you attend Pentaho [training classes](#) to learn how to install and configure the Pentaho Server, as well as how to develop data models.

This section is grouped into parts that will guide you during the development of your DI solution. These parts are iterative and you might bounce between them during development. For example as you tune a job, you might find that although you have built a solution that produces the right results, it takes a long time to run. So, you might need to rebuild and test a transformation to improve efficiency, and then retest it.

### Design DI solution

Design helps you think critically about the problem you want to solve and possible solutions. Consider these questions as you gather your requirements and design the solution.

**Output**

What does the overall solution look like? What questions are posing and how do you want the answers formatted?

**Data Sources**

What type(s) of data sources are you querying? Where are they located? How much data do you need to process? Are you using big data? Are you using relational or non-relational data sources? Will you have a target data source? If so, where are they located?

**Content/Processing**

What data quality issues do you have? How is the input data mapped to the output data? Where do you want to process the content, in PDI or in the data source? What hardware will you include in your development environment? Will you need one or more quality assurance test environments or production environments?

Also, consider templates or standards, naming conventions, and other requirements of your end users if you have them. Consider how you will back up your data as well.

**Set up development environment**

Setting up the environment includes installing and configuring PDI on development computers, configuring clustering if needed, and connecting to data sources. If you have one or more quality assurance environments, you will need to set those up also.

Table. Table 1. PDI Set Up Checklist

Task	Do This	Objective
Verify System Requirements	<ul style="list-style-type: none"> <li>Consult <a href="#">Components Reference</a>.</li> <li>Consult <a href="#">JDBC Drivers Reference</a>.</li> </ul>	<ul style="list-style-type: none"> <li>Acquire one or more servers that meet the requirements.</li> <li>Obtain the correct drivers for your system.</li> </ul>
Obtain Software and Install PDI	<ul style="list-style-type: none"> <li>See <i>Install Pentaho Data Integration and Analytics</i>.</li> </ul>	<ul style="list-style-type: none"> <li>Get the software from your Sales Support representative.</li> <li>Install the software.</li> <li>Start the Pentaho Server and Spoon.</li> </ul>
Install licenses for the Pentaho Server	<ul style="list-style-type: none"> <li>See <i>Administer Pentaho Data Integration and Analytics</i>.</li> </ul>	<ul style="list-style-type: none"> <li>Add all relevant Pentaho licenses.</li> </ul>
Connect to the Pentaho Repository	<ul style="list-style-type: none"> <li>See <i>Pentaho Data Integration</i>.</li> </ul>	<ul style="list-style-type: none"> <li>Connect to the Pentaho Repository.</li> </ul>
Apply Advanced Security (if needed)	<ul style="list-style-type: none"> <li>See <i>Administer Pentaho Data Integration and Analytics</i>.</li> </ul>	<ul style="list-style-type: none"> <li>Determine whether you need to apply DI Advanced Security.</li> </ul>

**Build and test solution**

During this step, you develop transformations, jobs, and models, then test what you have developed. You will tune the transformations, jobs, and models for optimal performance.

Development occurs in the PDI client design tool. The PDI client's streamlined design tightly couples the build and test activities so that you can easily perform them iteratively. The PDI client has perspectives help you perform ETL and visualize data. The PDI client also provides a scheduling perspective that can be used to automate testing. Testing encompasses verifying the quality of transformations and jobs, reviewing visualizations, and debugging issues. One common method of testing is to include steps in a

transformation or job that calculate hash totals, checksums, record counts, and so forth to determine whether data is being properly processed. You can also visualize your data in analyzer and report designer and review the results as you develop. This can not only help you find errors and issues with processing, but can help you get a jump on user acceptance testing if you show these reports to your customers or business analysts to get early feedback.

One basic question, is how to determine the numbers of transformations and jobs needed, as well as the order in which they should be executed. A good rule of thumb is to create one transformation for each combination of source system and target tables. You can often identify combinations in your mapping documents. Once you've identified the number of transformations that you need, you can use the same process to determine that number of jobs that you need. When considering the order of execution for transformations and jobs, consider how referential integrity is enforced. Run target table transformations that have no dependencies first, then run transformations that are depend on those tables next, and so forth.

Table. Table 2. Build and Test Checklist - PDI client

Task	Do This	Objective
Understand the Basics	<ul style="list-style-type: none"> <li>Read the overview of the PDI client process in the <i>Pentaho Data Integration</i> document.</li> </ul>	<ul style="list-style-type: none"> <li>Review information about the process and perspectives.</li> </ul>
Review most often used steps and entries	<ul style="list-style-type: none"> <li>Review <a href="#">commonly-used steps and entries</a>.</li> </ul>	<ul style="list-style-type: none"> <li>Review available transformations and determine how you can use them for your solution.</li> <li>Review job step references to identify which steps can be used in your solution.</li> </ul>
Create and Run Transformations	<ul style="list-style-type: none"> <li>Create and run a transformation. See the <i>Pentaho Data Integration</i> document for details.</li> </ul>	<ul style="list-style-type: none"> <li>Identify the transformations needed for your job and implement them.</li> <li>Save transformation.</li> <li>Run transformations locally.</li> </ul>
Create and Run a Job	<ul style="list-style-type: none"> <li>Create and run a job. See the <i>Pentaho Data Integration</i> document for details.</li> </ul>	<ul style="list-style-type: none"> <li>Create a job.</li> <li>Arrange transformations in a job so that they execute logically.</li> <li>Run a job.</li> </ul>

#### Tune solution

Fine tune transformations and jobs to optimize performance. This involves using various tools such as the DI Operation and Audit Mart to determine where bottlenecks or other performance issues occur, and addressing them.

Table. Table 3. Tune Checklist

Task	Do This	Objective
Review the Performance Tuning Checklist and Make Changes to Transformations and Jobs	<ul style="list-style-type: none"> <li>Review tuning tips. See the <i>Administer Pentaho Data Integration and Analytics</i> document for tuning tips.</li> </ul>	<ul style="list-style-type: none"> <li>Get familiar with things that you can do to optimize performance.</li> </ul>

Task	Do This	Objective
		<ul style="list-style-type: none"> <li>• Apply tuning tips as needed.</li> </ul>
Consider other performance tuning options	<ul style="list-style-type: none"> <li>• Read about transactional databases. See the <i>Pentaho Data Integration</i> document for details on transactional databases.</li> <li>• Read about using logs. See the <i>Administer Pentaho Data Integration and Analytics</i> document for details on logging.</li> </ul>	<ul style="list-style-type: none"> <li>• Learn how to apply transactional databases.</li> <li>• Learn how to use logs to tune transformations and jobs.</li> </ul>

### Next steps

These resources will be helpful to you as you prepare to Go Live for Production:

- Prepare to [Go Live for Production](#).
- [Hitachi Vantara Lumada and Pentaho Support Portal](#): check with Support for service packs.

## Go Live for production - DI

Go Live is the process by which you migrate a prototype to production. This process is divided into four parts:

- Setting up the production environment
- Deploying the solution
- Tuning the solution
- Scheduling the runs

### Set up production environment

Setting up the environment includes installing the software on production computers, configuring clustering, and connecting to data sources. To set up the environment, install and configure the Pentaho Server, Spoon, and any plugins required. Then set up data sources and clusters.

Table. Set up Production Environment Checklist

Task	Do This	Objective
Verify system requirements	<ul style="list-style-type: none"> <li>• Consult the <a href="#">Components Reference</a>.</li> <li>• Consult the <a href="#">JDBC Drivers Reference</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Acquire one or more servers that meet the requirements.</li> <li>• Obtain the correct drivers for your system.</li> </ul>
Obtain software and install the Pentaho Server	<ul style="list-style-type: none"> <li>• Download and install the latest service pack.</li> <li>• Start the Pentaho Server. See <i>Install Pentaho Data Integration and Analytics</i> for details.</li> <li>• Start the PDI client. See <i>Pentaho Data Integration</i> for details.</li> <li>• Install the licenses (if necessary). See <i>Administer Pentaho Data Integration and Analytics</i> for details.</li> </ul>	<ul style="list-style-type: none"> <li>• Get the software from your Sales Support representative.</li> <li>• Install the software - we recommend using the installation wizard.</li> </ul>

Task	Do This	Objective
Change the Server Fully Qualified URL	<ul style="list-style-type: none"> <li>Change the ports and URLs. See <i>Administer Pentaho Data Integration and Analytics</i> for details.</li> </ul>	<ul style="list-style-type: none"> <li>Change the server's URL so that you do not have a conflict.</li> </ul>
Connect to the Pentaho Repository	<ul style="list-style-type: none"> <li>Create a connection to the Pentaho Repository. See <i>Pentaho Data Integration</i> for details.</li> </ul>	<ul style="list-style-type: none"> <li>Connect to the Pentaho Repository.</li> </ul>
Set up clusters	<ul style="list-style-type: none"> <li>Optional: Set up clusters. See <i>Pentaho Data Integration</i> for details.</li> </ul>	<ul style="list-style-type: none"> <li>Become familiar with clustering.</li> <li>Set up clusters, if they are needed in your environment.</li> </ul>
Copy configuration files	Copy shared.xml, repositories.xml, kettle.properties, and JAR files from the development environment to the production environment.	<ul style="list-style-type: none"> <li>System is set up and ready for production.</li> </ul>
Logging and monitoring your server	<ul style="list-style-type: none"> <li>Review logging and monitoring operations. See <i>Pentaho Data Integration</i> for details.</li> <li>Enable logging. See <i>Administer Pentaho Data Integration and Analytics</i> for details.</li> <li>Monitor PDI and SNMP traps. See <i>Administer Pentaho Data Integration and Analytics</i> for details.</li> </ul>	<ul style="list-style-type: none"> <li>Learn about the different ways to log and monitor Pentaho Server operations: <ul style="list-style-type: none"> <li>Log through Spoon and Carte</li> <li>Use SNMP traps with PDI</li> </ul> </li> </ul>

### Deploy solution

Export solutions from the Pentaho Repository that is in the development or test environments, to the Pentaho Repository that is in the production environment.

Table. Move Solution Checklist

Task	Do This	Objective
Export and Import Pentaho Repository	<ul style="list-style-type: none"> <li>See <i>Export and Import Pentaho Repository Content</i> in the <i>Administer Pentaho Data Integration and Analytics</i> document.</li> </ul>	<ul style="list-style-type: none"> <li>Export Pentaho Repository content from test environment</li> <li>Import Pentaho Repository content to production environment</li> </ul>

### Tune solution

Fine tune transformations and jobs to optimize performance. This involves using various tools such as the DI Operations and Audit Marts to determine where bottlenecks or other performance issues occur, and attempting to address them.

Table. Tune Checklist

Task	Do This	Objective
------	---------	-----------

Task	Do This	Objective
Review the Performance Tuning Checklist and Make Changes to Transformations and Jobs	<ul style="list-style-type: none"> <li>Consult the tuning tips. See the <i>Administer Pentaho Data Integration and Analytics</i> document for tuning tips.</li> </ul>	<ul style="list-style-type: none"> <li>Get familiar with things that you can do to optimize performance.</li> <li>Apply tuning tips as needed.</li> </ul>
Consider other performance tuning options	<ul style="list-style-type: none"> <li>Learn about transactional databases. See the <i>Pentaho Data Integration</i> document for details on transactional databases.</li> <li>Learn about using logs. See the <i>Administer Pentaho Data Integration and Analytics</i> document for details on logging.</li> </ul>	<ul style="list-style-type: none"> <li>Learn how to apply transactional databases.</li> <li>Learn how to use logs to tune transformations and jobs.</li> </ul>

### Schedule runs

Use the PDI client, Pan, or Kitchen to schedule executions of transformations and jobs.

Table. Schedule Runs Checklist

Task	Do This	Objective
Schedule Transformations and Jobs From Spoon	<ul style="list-style-type: none"> <li>Schedule transformations and jobs. See the <i>Pentaho Data Integration</i> document for details.</li> </ul>	<ul style="list-style-type: none"> <li>Schedule transformations and jobs</li> </ul>
Command Line Scripting Through Pan and Kitchen	<ul style="list-style-type: none"> <li>Learn about Pan's options. See the <i>Pentaho Data Integration</i> document for details.</li> <li>Learn about Kitchen's options. See the <i>Pentaho Data Integration</i> document for details.</li> </ul>	<ul style="list-style-type: none"> <li>Use Pan and Kitchen to schedule transformations and jobs.</li> </ul>

### Next steps

These resources will be helpful to you after your production server is live.

- Fine-tune Pentaho systems: Provides guidance on how to maintain and fine-tune your Pentaho Server. See the *Administer Pentaho Data Integration and Analytics* document for details.
- Pentaho [Training and Education](#)
- [Support Portal](#): Check with support for service packs.

## Commonly-used PDI steps and entries

Although there are over 330 transformation steps and job entries, some steps and entries are used more often than others. If you are creating a transformation and job, but do not know where to begin, this list might be helpful to you.

### Top ten transformation steps

PDI Transformation steps are documented in Documentation.

- Text File Input
- Table Input
- Excel Input
- Text File Output
- Table Output
- Excel Output
- Select Values
- Filter Rows
- Group By
- Stream Lookup

### Other commonly-used transformation steps

PDI Transformation steps are documented in Documentation

- INPUT: Generate Rows, Data Grid, Get Data from XML, CSV File Input, Fixed File Input
- OUTPUT: XML Output
- TRANSFORM: Split Fields, Calculator, Add Constants, Add Sequence, Replacing Strings, Split Fields, Sort Rows, String Operations, Strings Cut
- SCRIPTING: User Defined Java Class, Modified Java Script Value, User Defined Java Expression
- FLOW: Abort, Append Streams, Block this step until steps finish, Blocking Step, Detect Empty Stream, Dummy, ETL Metadata Injection, Filter Rows, Identify Last Row in a Stream, Java Filter, Job Executor, Prioritize Streams, Single Threader, Switch/Case, Transformation Executor
- LOOKUP
- JOINS: Join Rows, Merge Join
- JOB: Get Variables, Set Variables

### Commonly-used job entries

PDI Job entries are documented in Documentation.

- GENERAL: Start, Job, Transformation, Success
- UTILITY: Abort
- MAIL: Mail
- FILE MANAGEMENT: Add filenames to result, Compare folders, Convert file between Windows and Unix, Copy Files, Create a folder, Create file, Delete file, Delete filenames from result, Delete files, Delete folders, File Compare, HTTP, Move Files, Process result filenames, Unzip file, Wait for file, Write to file, Zip file
- UTILITIES: Write to log

## Pentaho Business Analytics workflow

Pentaho Business Analytics is a combined business analytics and data integration platform that allows business users, data scientists, and IT administrators to easily access, explore, and visualize their data. Pentaho empowers business users to make information-driven decisions that positively impact their organization's performance, data scientists to use a full-spectrum of tools to create robust data models, and IT to rapidly deliver a secure, scalable, flexible, and easy to manage business analytics platform for the broadest set of users.

- [Evaluate and learn Pentaho Business Analytics](#)
- [Develop your BA environment](#)
- [Go live for production - BA](#)

## Evaluate and learn Pentaho Business Analytics

As you explore Pentaho Business Analytics, you will be introduced to the major components, watch videos, work through hands-on examples, and learn about the different features.

Go at your own pace. Feel free to dig into the documentation or to contact Pentaho [sales support](#) if you have questions.

This section is divided into four parts that will help you get familiar with Business Analytics:

- Tour the User Console and create your first reports
- Explore and learn data source basics
- Learn about Report Designer
- Discover more about Pentaho Business Analytics

### Tour the User Console and create your first reports

The User Console is a web-based design environment where you can analyze data, create interactive reports, dashboard reports, and build integrated dashboards to share business intelligence solutions with others in your organization and on the internet. In addition to its design features, the User Console offers a wide variety of system administration features for configuring the Pentaho Server, maintaining Pentaho licenses, setting up security, managing report scheduling, and tailoring system performance to meet your requirements.

If you have installed the trial download on your laptop or desktop machine, you are ready to get started exploring. If you have the software installed on a server, and want to use your machine to point to it, see [Develop your BA environment](#) for details.

Table. Table 1. Explore and Learn Pentaho Reporting Capabilities

Lesson	Do This	Notes
Tour the User Console	<a href="#">Quick tour of the Pentaho User Console (PUC)</a>	<ul style="list-style-type: none"> <li>• Understand the features of the User Console</li> <li>• View the sample reports on the Samples tab of the Getting Started section</li> </ul>
Create Your First Reports and Dashboards	<a href="#">Learn about Analyzer, Interactive Reports, and Dashboards</a>	<ul style="list-style-type: none"> <li>• Created and saved an Interactive Report</li> <li>• Created and saved an Analysis Report</li> <li>• Created and saved a custom Dashboard</li> </ul>
Schedule Your Report	<ul style="list-style-type: none"> <li>• Learn about scheduling reports. See the <i>Pentaho Analytics</i> document for details.</li> </ul>	<ul style="list-style-type: none"> <li>• Scheduled a report to run and email automatically.</li> <li>• Received your report through email after the schedule runs.</li> </ul>

### Explore and learn data source basics

If you have already worked with the Steel Wheels sample data, and want to learn how to create your own data sources and data models with Pentaho, use the Data Source Wizard. The Data Source Wizard helps you define a data source that contains the data you want to use and guides you through the creation of your evaluation data model for use in creating reports.

After you define a data source, you can make it available to other evaluators so they can create reports and analysis by simply picking the data source from the data source list. Any number of reports can be created using a single data source.

Table. Table 2. Explore and Learn Data Sources Basics

Lesson	Do This	Notes
--------	---------	-------



Lesson	Do This	Notes
Create Your First Data Source	<ul style="list-style-type: none"> <li>Create a Data Source</li> <li>Tour the Data Source Wizard</li> </ul> <p>See <i>Pentaho Analytics</i> for instructions.</p>	<ul style="list-style-type: none"> <li>Understand how the Pentaho Server and Data Source Wizard work together to create usable data sources and data models.</li> <li>Explore the Data Source Wizard interface.</li> <li>Learn the basics of creating a data source using the Data Source Wizard.</li> </ul>
Choose Data Source Types	<ul style="list-style-type: none"> <li>Choose a data source type</li> </ul> <p>See <i>Pentaho Analytics</i> for instructions.</p>	<ul style="list-style-type: none"> <li>Learn about the different data source types supported by the Data Source Wizard.</li> </ul> <p>We recommend using a CSV data source for evaluation.</p>
Create Your First CSV Data Source	<ul style="list-style-type: none"> <li>Create a CSV data source</li> </ul> <p>See <i>Pentaho Analytics</i> for instructions.</p>	<ul style="list-style-type: none"> <li>Import a CSV data file using the Data Source Wizard.</li> <li>Create the CSV data source.</li> </ul> <p>We recommend creating a report using this new CSV data source, then refining the data model with the . Data Source Model Editor as needed</p>
Refine Your Data Source Model	<ul style="list-style-type: none"> <li>Edit multidimensional data source models.</li> </ul> <p>See <i>Pentaho Analytics</i> for instructions.</p>	(Optional) Edit your evaluation data source model using the Data Source Model Editor.
Inline Model Editing	<ul style="list-style-type: none"> <li>Read <i>Working with Analyzer measures</i> in the <i>Pentaho Analytics</i> document.</li> </ul>	<ul style="list-style-type: none"> <li>Learn how to edit your data models while working in Analyzer.</li> </ul>
Learn about Streamlined Data Refinery	<ul style="list-style-type: none"> <li>Learn how to work with Streamlined Data Refinery</li> </ul> <p>See <i>Pentaho Data Integration</i> for instructions.</p>	<ul style="list-style-type: none"> <li>Learn how Streamlined Data Refinery works.</li> </ul>

### Learn about Report Designer

Pentaho Report Designer is a report creation tool that you can use by itself, or as part of the Pentaho Suite. It allows professionals to create print-quality reports based on data from virtually any type of data source.

These resources in the *Pentaho Report Designer* document will help you get familiar with the Report Designer interface, and guide you through the creation and publishing of a print-quality report.

Table. Table 3. Explore and Learn about Report Designer

Lesson	Section in document	Notes
--------	---------------------	-------

Lesson	Section in document	Notes
Explore the Report Designer Interface	<ul style="list-style-type: none"> <li>• <i>Explore Report Designer</i></li> </ul>	<ul style="list-style-type: none"> <li>• Tour the Report Designer interface before you begin building reports.</li> </ul>
Report Designer Workflow Overview	<ul style="list-style-type: none"> <li>• <i>Learn about Report Designer workflow</i></li> </ul>	<ul style="list-style-type: none"> <li>• Look over the work flow concepts for Report Designer.</li> </ul>
Create Your First Report	<ul style="list-style-type: none"> <li>• <i>Create your first print-quality report</i></li> </ul>	<ul style="list-style-type: none"> <li>• Create a report.</li> <li>• Add a chart and parameters to your report.</li> <li>• View and then publish your report.</li> </ul>
Refine the Look of Your Report	<ul style="list-style-type: none"> <li>• <i>Design print-quality reports</i></li> </ul>	<ul style="list-style-type: none"> <li>• Explore more advanced features of Report Designer, beginning with report elements.</li> </ul>
Add a PDI Data Source	<ul style="list-style-type: none"> <li>• <i>Add a PDI data source</i></li> </ul>	<ul style="list-style-type: none"> <li>• Add a PDI data source and use it to create a report in Report Designer.</li> </ul>

### Discover more about Pentaho Business Analytics

- The Pentaho Analyzer, Interactive Reports, and Dashboard Designer plugins provide in-depth details about creating eye-catching business intelligence deliverables for your user community. See the *Pentaho Analytics* document for details.
- If you are a system administrator, check out the *Install Pentaho Data Integration and Analytics* document. Both have details on configuring and administering your Pentaho Server using the User Console, as well as a section on the variety of things you can do to maintain your server manually.

### Next steps

- Contact [Pentaho](#) to learn more about how Business Analytics can be tailored to meet your business needs.
- Learn how to [Develop your BA environment](#).

## Develop your BA environment

The Develop Your BA Environment workflow outlines the processes that you use to set up and configure your Pentaho Server for BA development, and to build, change, and test Pentaho solutions until they meet your unique business requirements. This process is similar to the Trial Download Evaluation experience, except that you will be completely configuring the Pentaho Server and working with your own report designers and data scientists, as well as Pentaho [professional services](#). The end result of this section is to learn BA implementation best practices and deploy your BA solution to a production server.

Before you begin developing your BA environment, we recommend that you attend Pentaho [training classes](#) to learn how to install and configure the Pentaho Server, as well as how to develop data models and applications.

### Set up your Pentaho Server

This table helps you verify that the hardware and software requirements are met, install the software, and configure the Pentaho Server and BA Design tools.

Table. Table 1. Pentaho Server Set-Up Checklist

Checklist	Steps	Notes
-----------	-------	-------

Checklist	Steps	Notes
Verify System Requirements	<ul style="list-style-type: none"> <li>Consult the <a href="#">Components Reference</a></li> <li>Consult the <a href="#">JDBC drivers reference</a></li> </ul>	<p>In preparation for this task, make sure that you:</p> <ul style="list-style-type: none"> <li>Acquire one or more servers that meet the requirements.</li> <li>Obtain the correct drivers for your system.</li> </ul>
Obtain Software and Install the Pentaho Server	<ul style="list-style-type: none"> <li>Get the software</li> <li><a href="#">Install all Pentaho components</a></li> <li><a href="#">Access the User Console</a></li> </ul>	<p>After you have the hardware, you will need to:</p> <ul style="list-style-type: none"> <li>Get the software from your Sales Support representative.</li> <li>Install the software - we recommend using the installation wizard.</li> <li>Access the User Console, toured the Administration page, and changed the default administrator password.</li> </ul>
Change the Pentaho Server Fully Qualified URL	<ul style="list-style-type: none"> <li>Change the Pentaho Server fully qualified URL. See the <i>Administer Pentaho Data Integration and Analytics</i> document for details.</li> </ul>	<p>If you are going to have many machines pointing to one server, make sure you change the fully qualified URL:</p> <ul style="list-style-type: none"> <li>Change the URL and verify that you can connect to it.</li> </ul>
Configure the Pentaho Server	<ul style="list-style-type: none"> <li>Manage the licenses. See the <i>Administer Pentaho Data Integration and Analytics</i> document for details.</li> <li>Specify data connections for the Pentaho Server. See the <i>Install Pentaho Data Integration and Analytics</i> document for details.</li> <li>Set up e-mails for scheduled reports. See the Pentaho Analytics document for details.</li> <li>Manage schedules. See the Pentaho Analytics document for details.</li> </ul>	<p>After installing the software and logging on to the User Console for the first time, you will need to:</p> <ul style="list-style-type: none"> <li>Add all relevant Pentaho licenses.</li> <li>Set up your data connections.</li> <li>Configure an email server through the Administration page.</li> <li>Learn how to manage schedules through the Schedules page. These will be set up during the production phase.</li> </ul>
Configure BA Design Tools	<ul style="list-style-type: none"> <li>Configure the design tools and utilities. See the <i>Install Pentaho Data Integration and Analytics</i> document for details.</li> <li>Specify data connections for Business Analytics tools. See the <i>Install Pentaho Data Integration and Analytics</i> document for details.</li> </ul>	<p>If you plan to use BA design Tools, make sure you do these tasks. Otherwise, skip this step:</p> <ul style="list-style-type: none"> <li>Configure your BA design tools.</li> <li>Specify the data connections required for each tool.</li> </ul> <p>We recommend that you NOT configure these tools on your production server. Rather, configure them on your development server, test them, then migrate them over to production.</p>

## Import data sources and data models

Follow the steps in the Data Sources and Data Models Checklist to create data sources and models that are ideal for agile development of your BA solution.

Table. Table 2. Data Sources and Data Models Checklist

Task	Steps	Notes
Choose Data Source Types	<ul style="list-style-type: none"> <li>Choose a Data Source type. See <i>Pentaho Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Review the information about data sources.</li> <li>Examine the differences between the different data models - relational data models vs. multidimensional data models.</li> </ul>
Create Data Sources and Models	<ul style="list-style-type: none"> <li>Tour the Data Source Wizard. See <i>Pentaho Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Learn how the Pentaho Server and Data Source Wizard work together to create usable data sources and data models.</li> </ul>
Create Database Table Data Sources	<ul style="list-style-type: none"> <li>Create a database table source. See <i>Pentaho Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Create Data Sources and preliminary data models using the Data Source Wizard.</li> </ul>
Learn about Mondrian Schemas	<ul style="list-style-type: none"> <li>Create and modify Mondrian schemas. See <i>Pentaho Schema Workbench</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Create Mondrian schemas and add a Mondrian data source</li> <li>Adapt the schema to work with Analyzer</li> <li>Modify the schema in Schema Workbench</li> </ul>

## Create reports and further refine data models

We recommend that you work with data scientists and business analysts at this stage of the process to achieve the best results. As you get ready to move to production, use data sources created using either the Pentaho Schema Workbench or Pentaho Metadata Editor. See those documents for details.

Table. Table 3. Create Reports and Refine Data Models Checklist

Task	Steps	Notes
Create Analysis Reports, Interactive Reports, and Dashboards	<ul style="list-style-type: none"> <li>See the <i>Pentaho Analytics</i> document for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Create Interactive and Analysis reports, as well as a Dashboard.</li> <li>Verify that the reports show the data that you want.</li> <li>If they do not, work with data scientists to get the results you need.</li> </ul>
Create a Report with Report Designer	<ul style="list-style-type: none"> <li>See the <i>Pentaho Report Designer</i> document for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>(Optional) Create a report in Report Designer.</li> </ul>

Task	Steps	Notes
Refine Your Data Source Model	<ul style="list-style-type: none"> <li>Edit multidimensional data source models. See the <i>Pentaho Analytics</i> document for instructions.</li> <li>Use Schema Workbench to refine multidimensional data models. See the <i>Pentaho Schema Workbench</i> document for instructions.</li> <li>Use Metadata Editor to refine relational data models. See the <i>Pentaho Metadata Editor</i> document for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Refine data models to get the data that you want to see in your reports.</li> <li>Create the reports again to verify the results.</li> <li>Repeat this process as needed.</li> </ul>

### Test environment quality

If you have decided to perform quality assurance testing, upload your content to the Pentaho Repository, then download it to the QA server. See *Administer Pentaho Data Integration and Analytics* for instructions.

Some organizations have an additional layer of User Acceptance Testing that is performed after QA testing is complete.

### Next steps

- Investigate Security: Decide which security solution works best for your business environment. See *Administer Pentaho Data Integration and Analytics* for details.
- Learn about Scheduling: Decide which schedules to set up on your production server. See *Pentaho Analytics* for details.
- Think about what content you want to move to your production server. See *Administer Pentaho Data Integration and Analytics* for details.
- [Support Portal](#): check with Support for service packs.
- Prepare to [Go live for production](#).

## Go live for production - BA

This section explains the process for moving Pentaho content and server settings among multiple server instances. Typically, this process includes either two or three separate servers with identical configurations: one for BA content development, one for testing and QA if desired, and one for production. We recommend that you work with a member of Pentaho [professional services](#) to help with the preparation work for deployment to production.

### Prepare for going live

This section is divided into two parts; the first part serves as a checklist for setting up a Pentaho Server, and the second consists of prerequisites that you must perform before going live with your production server. If you already have your production server ready to go, skip ahead to the tasks listed in the *Prerequisites for Go Live - Checklist for Production Server* table below.

Table. Pentaho Server Set Up Checklist

Task	Do This	Notes
Verify System Requirements	<ul style="list-style-type: none"> <li>Consult the <a href="#">Components Reference</a>.</li> <li>Consult the <a href="#">JDBC drivers reference</a>.</li> </ul>	<ul style="list-style-type: none"> <li>Acquired one or more servers that meet the requirements.</li> <li>Obtained the correct drivers for your system.</li> </ul>

Task	Do This	Notes
Obtain Software and Install the Pentaho Server	<ul style="list-style-type: none"> <li>• Install Pentaho Suite. See <i>Install Pentaho Data Integration and Analytics</i> for instructions.</li> <li>• Download and install the latest service pack. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> <li>• Access the User Console. See <i>Pentaho Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>• Installed the software.</li> <li>• Installed the latest service pack.</li> <li>• Accessed the User Console, toured the Administration page, and changed the default administrator password.</li> </ul> <p>If you need to, see the next segment to change the fully qualified URL for the Pentaho Server.</p>
Change the Server Fully Qualified URL	<ul style="list-style-type: none"> <li>• Change the Pentaho Server's fully qualified URL if necessary. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> </ul>	<p>If you are going to have many machines pointing to one server, make sure you change the fully qualified URL and verify that you can connect to it.</p>
Configure the Server	<ul style="list-style-type: none"> <li>• Manage the licenses. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> <li>• Specify the data connections for the Pentaho Server. See <i>Install Pentaho Data Integration and Analytics</i> for instructions.</li> <li>• Set up emails for scheduled reports. See <i>Pentaho Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>• Added all relevant Pentaho licenses.</li> <li>• Set up your data connections.</li> <li>• Configured an email server through the Administration page.</li> </ul>

This checklist guides you through the tasks you need to do to ensure a smooth transition from a development server to a live production server.

Table. Prerequisites for Go Live - Checklist for Production Server

Task	Do This	Notes
Compare configuration files	<ul style="list-style-type: none"> <li>• Compare configuration files.</li> <li>• Verify and increase memory settings. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>• Differences between configuration files are identified.</li> <li>• Unified properties file is committed to version control.</li> <li>• Memory settings have been increased as applicable.</li> </ul>
Verify Data Sources	<ul style="list-style-type: none"> <li>• Specify data connections for the Pentaho Server. See <i>Install Pentaho Data Integration and Analytics</i> for instructions.</li> <li>• Define JNDI connections for the Pentaho Server. See <i>Install Pentaho Data Integration and Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>• Make sure data sources are able to be promoted.</li> <li>• Established JNDI sources as replacements if needed.</li> </ul>

Task	Do This	Notes
Define Security	<ul style="list-style-type: none"> <li>Define security for the Pentaho Server. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> <li>Manage users and roles. See <i>Pentaho Analytics</i> for instructions.</li> <li>Implement advanced security. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Security is implemented.</li> <li>Users, roles, and permissions are defined.</li> </ul>
Upload Content	<ul style="list-style-type: none"> <li>Upload and download from the Pentaho Repository. See <i>Administer Pentaho Data Integration and Analytics</i> for instructions.</li> </ul>	<ul style="list-style-type: none"> <li>Uploaded files and folders.</li> </ul>

### Compare configuration files

The most important server configuration settings are stored in the `/server/pentaho-server/pentaho-solutions/system/` directory. There are also a few core settings inside of the Pentaho WAR archive deployed to your application server, though they should not change at all after your initial server setup is complete.

#### CAUTION:

Do not change the names of any content files, data sources, solution directories, or other file names during the promotion process. Names should be set during solution development and strictly maintained throughout content promotion. Modifying file names can result in complications with BA content that cannot be immediately detected, which will negatively impact your QA process.

To be absolutely certain that you have all of the server configuration files selected, you should compare the following directories in their entirety:

- `/pentaho-solutions/system/`
- `/WEB-INF/` inside your deployed `pentaho.war`
- `/META-INF/` inside your deployed `pentaho.war`
- For JBoss deployments, the `PentahoHibernate-ds.xml` and `quartz-ds.xml` files in the `/server/default/deploy/` directory

Note: There are binaries inside of the plugin directories for Analyzer, Dashboard Designer, Interactive Reports, and Community Dashboard Framework. While it may be useful to note any binary differences, which would indicate possible version differences among Pentaho Server or plugin deployments, in general you should be most concerned with XML configuration files and properties files. However, if you have done any plugin customization work, you will need to promote those changes as well.

### Move content to production server

This checklist is a compilation of best practices for you to use to promote your Pentaho Server settings, data sources, and content.

Before you dive into going from a development server to your live production server, make sure that you have completed all of the tasks in the [Prerequisites](#) checklist.

Task	Do This	Notes
Download Content	<ul style="list-style-type: none"> <li>Upload and Download from the Pentaho Repository</li> </ul>	<ul style="list-style-type: none"> <li>All desired content is moved to the production server.</li> <li>See <i>Administer Pentaho Data Integration and Analytics</i> for details.</li> </ul>

Task	Do This	Notes
Set Up Schedules and Blockout Times	<ul style="list-style-type: none"> <li>• Manage Schedules</li> <li>• Prevent Scheduling by Setting Blockout Times</li> </ul>	<ul style="list-style-type: none"> <li>• Set up schedules for production server.</li> <li>• Set up blockout times for server maintenance tasks.</li> <li>• See <i>Pentaho Analytics</i> for details.</li> </ul>

### Next steps

These resources will be helpful to you after your Production server is live.

- See *Administer Pentaho Data Integration and Analytics* for guidance on how to maintain and fine-tune your Pentaho Server.
- Pentaho [Training and Education](#)
- [Support Portal](#) - check with support for service packs.

## Components Reference

Pentaho aims to accommodate diverse computing environments. This list provides details about the environment components and versions we support. Where applicable, versions are listed as *certified* or *supported*:

### Certified

The version has been tested and validated for compatibility with Pentaho.

### Supported

Support is available for listed non-certified versions.

If you have questions about your particular computing environment, contact [Pentaho Support](#).

## Server

The Pentaho Server is hardware-independent and runs on server-class computers that comply with these specifications for minimum hardware and required operating systems:

Hardware—64 bit	Operating System—64 bit	
	Certified	Supported
Processor Intel EM64T or AMD64 Dual-Core RAM 8 GB with 4 GB dedicated to Pentaho servers Disk Space 20 GB free after installation	<ul style="list-style-type: none"> <li>• Microsoft Windows 2022 Server</li> <li>• CentOS Stream 8</li> <li>• Red Hat Enterprise 8</li> <li>• Ubuntu Server 20.04 LTS</li> </ul>	<ul style="list-style-type: none"> <li>• Microsoft Windows 2019 Server &amp; 2022 Server</li> <li>• CentOS Linux 7, Linux 8, &amp; Stream 8</li> <li>• Red Hat Enterprise 7.x &amp; 8.x</li> <li>• Ubuntu Server 18.x LTS &amp; 20.x LTS</li> </ul>

## Container deployment

Supported technology for deploying Pentaho in containers.

Technology	Certified	Supported
Docker	20.10.10	20.x



## Workstation

These Pentaho design tools are hardware-independent and run on client-class computers that comply with these specifications for minimum hardware and required operation systems.

- Pentaho Aggregation Designer
- Pentaho Data Integration
- Pentaho Metadata Editor
- Pentaho Report Designer
- Pentaho Schema Workbench

Hardware—64 bit	Operating System—64 bit	
	Certified	Supported
<b>Processors</b> <ul style="list-style-type: none"> <li>• Apple Macintosh Dual-Core</li> <li>• Intel EM64T or AMD64 Dual-Core</li> </ul> <b>RAM</b> 2 GB RAM for most of the design tools, PDI requires 2 GB dedicated	<ul style="list-style-type: none"> <li>• Ubuntu Desktop 20.04</li> <li>• Microsoft Windows 11</li> <li>• macOS 12 (Monterey)</li> </ul>	<ul style="list-style-type: none"> <li>• Ubuntu Desktop 18.x &amp; 20.x</li> <li>• Microsoft Windows 10 &amp; 11</li> <li>• macOS 11 (Big Sur) &amp; 12 (Monterey)</li> </ul>
<b>Disk Space</b> 2 GB free after installation		
<b>Minimum Screen Size</b> 1280 x 960		

## Embedded software

When embedding Pentaho software into other applications, the computing environment should comply with these specifications for minimum hardware and required operation systems.

- Embedded Pentaho Reporting
- Embedded Pentaho Analysis
- Embedded Pentaho Data Integration

Hardware—64 bit	Operating System—64 bit	
	Certified	Supported
<b>Processors</b> Intel EM64T or AMD64 Dual-Core	<ul style="list-style-type: none"> <li>• Microsoft Windows 2022 Server</li> <li>• CentOS 8</li> <li>• Red Hat Enterprise 8</li> <li>• Ubuntu Server 20.04 LTS</li> </ul>	<ul style="list-style-type: none"> <li>• Microsoft Windows 2019 Server &amp; 2022 Server</li> <li>• CentOS 7 &amp; 8</li> <li>• Red Hat Enterprise 7.x &amp; 8.x</li> <li>• Ubuntu Server 18.x LTS &amp; 20.x LTS</li> </ul>
<b>RAM</b> 8 GB with 4 GB dedicated to Pentaho servers		
<b>Disk Space</b> 20 GB free after installation		

## Application servers

Servers to which you deploy Pentaho software must run one of these application servers:

- JBoss EAP 7.3
- Tomcat 9.0.xx (shipped with 9.0.50) with Oracle Java 11.x (shipped with 11.0.13)

## Solution database repositories

Pentaho software stores processing artifacts in these database repositories:

Certified	Supported
PostgreSQL 14	PostgreSQL 12.x & 13.x *
MySQL 8.026	MySQL 5.7
Oracle 21c	Oracle 19c & 21c (including patched versions)
MS SQL Server 2019	Microsoft SQL Server 2017 & 2019 (including patched versions)

\* The default installed solution database.

## Hadoop Vendors

Pentaho software supports data sources from these Hadoop Vendors.

Vendor	Driver Version
Cloudera Data Platform (CDP) Private Cloud	7.1
Google Dataproc	1.4
EMR	5.21
Microsoft Azure HDInsight	4.0

## Data Sources: General

Pentaho software supports the following data sources. Check this list if you are evaluating Pentaho or checking for general compatibility with a specific vendor.

Data Source	Certified	Supported
Salesforce	54	54.x
Amazon Redshift	1.2.34.1058	1.2.34.x
Snowflake	3.13.16	3.13.x
Hitachi Content Platform	8.0.0.9	8.0.0.x

## Pentaho Tools

This table summarizes which data sources are compatible with the main Pentaho tools.

Pentaho Software	Data Source
Pentaho Reporting	<ul style="list-style-type: none"> <li>JDBC 3/4*</li> <li>ODBC</li> <li>OLAP4J</li> <li>XML</li> <li>Pentaho Analysis</li> <li>Pentaho Data Integration</li> <li>Pentaho Metadata</li> <li>Scriptable</li> <li>Snowflake</li> </ul>
Pentaho Server, Action Sequences	<ul style="list-style-type: none"> <li>Relational (JDBC)</li> <li>Hibernate</li> <li>Javascript</li> </ul>

Pentaho Software	Data Source
	<ul style="list-style-type: none"> <li>• Metadata (MQL)</li> <li>• Mondrian (MDX)</li> <li>• XML (XQuery)</li> <li>• Security User/Role List Provider</li> <li>• Snowflake</li> <li>• Data Integration Steps (PDI)</li> <li>• Other Action Sequences</li> <li>• Web Services</li> <li>• XMLA</li> </ul>
Pentaho Data Integration	<ul style="list-style-type: none"> <li>• JDBC 3/4*</li> <li>• OLAP4J</li> <li>• Salesforce</li> <li>• Snowflake</li> <li>• XML</li> <li>• CSV</li> <li>• Microsoft Excel</li> </ul>
<p>* Use a JDBC 3.x or 4.x compliant driver that is compatible with SQL-92 standards when communicating with relational data sources. For a list of drivers to use with relational JDBC databases, see the <a href="#">JDBC drivers reference</a>.</p>	

## Big Data Sources: General

Pentaho software supports the following Big Data sources. Check this list if you are evaluating Pentaho or checking for general compatibility with a specific vendor.

Data Source	Supported Version
Amazon EMR (via Hive)	5.21, 5.24, 5.32
Cloudera (via Hive or Impala)	6.1, 6.2, 6.3
Cloudera Data Platform (via Hive or Impala)	7.1.x
Datastax	4.6, 4.8
Google BigQuery	1.2.2.1004
Google Dataproc	1.4, 2.2
Greenplum	4.2, 4.3
Hortonworks (via Hive or Spark SQL)	3.0, 3.1
Microsoft Azure HDInsight	4.0
MongoDB	4.0.2
Netezza	7.1, 7.2
SAP HANA	SPS
Teradata	14.10, 15.0
Vertica	9.3.0.0

## Big Data Sources: Details

This table shows the Big Data sources that are compatible with specific Pentaho tools.

Data Source	Versions	Analyzer	PIR/PDD	Pentaho Reporting	DSW	PDIServer/Client	PRD	PSW	PME
Amazon EMR <sup>a</sup>	5.21, 5.24, 5.36	Yes	Yes	No	No	Yes	Yes	No	Yes
Cloudera Data Platform (CDP) Private Cloud	7.1.x (for job execution)	No	No	No	No	Yes	Yes	No	Yes
	<a href="#">via Impala<sup>b</sup></a> (as data source)	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
	via Hive3 <sup>c</sup> (as data source)	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Datastax	4.6, 4.8	No	No	No	No	Yes	No	No	No
<a href="#">Google BigQuery</a>	1.2.25 <sup>d</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Google Dataproc <sup>e</sup> (for job execution)	1.4, 2.2 <sup>f</sup>	No	No	No	No	Yes	Yes	No	No
	via Hive2 and Google BigQuery (as data source)	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Greenplum	4.2, 4.3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Microsoft Azure HDInsight	4.0	Yes	Yes	No	No	Yes	No	No	Yes
MongoDB	4.4	No	No	Yes	No	Yes	Yes	No	No
Netezza	7.1, 7.2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SAP HANA	SPS	No	No	No	No	Yes	No	No	No
Teradata	16.20	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vertica	10 & 11	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

**Notes:** A generic Apache Hadoop driver is included in the Pentaho distribution for version 9.4: Other supported drivers can be downloaded from the [Hitachi Vantara Lumada and Pentaho Support Portal](#).

<sup>a</sup> Use the EMR 5.21 driver for your EMR 5.24 or EMR 5.36 cluster. The EMR 5.21 driver is certified to work for EMR 5.24.

<sup>b</sup> You must have the current version of the Pentaho release to use the CDP 7.1.4 driver. The CDP 7.1.4 driver requires the [Impala JDBC Connector 2.6.4 Cloudera driver](#).

<sup>c</sup> Hive3 as a data source for CDP also supports Hive LLAP, and Hive3 on Tez.

<sup>d</sup> The Simba driver required for Google BigQuery is the JDBC 4.2-compatible version. See <https://cloud.google.com/bigquery/partners/simba-drivers/>.

<sup>e</sup> HBase is not supported with Google Dataproc.

Data Source	Versions	Analyzer	PIR/PDD	Pentaho Reporting	DSW	PDIServer/Client	PRD	PSW	PME
<sup>f</sup> Use the Google Dataproc 1.8 driver for your Google Dataproc 2.2 cluster. The Google Dataproc 1.8 driver is certified to work for Google Dataproc 2.2.									

## SQL dialect-specific

Pentaho software generates dialect-specific SQL when communicating with these data sources. Certified indicates the SQL dialect has been tested for compatibility with Pentaho.

Pentaho Software	Data Source
Pentaho Analyzer	Certified <ul style="list-style-type: none"> <li>• Amazon Redshift</li> <li>• Azure SQL</li> <li>• Impala</li> <li>• MySQL</li> <li>• Microsoft SQL Server</li> <li>• Oracle</li> <li>• PostgreSQL</li> <li>• Snowflake</li> </ul> Supported <ul style="list-style-type: none"> <li>• Access</li> <li>• Derby</li> <li>• Firebird</li> <li>• Greenplum</li> <li>• Hsqldb</li> <li>• IBM DB2</li> <li>• IBM MQ 9.2</li> <li>• Informix</li> <li>• Ingres</li> <li>• Interbase</li> <li>• Neoview</li> <li>• Netezza</li> <li>• SqlStream</li> <li>• Sybase</li> <li>• Teradata</li> <li>• Vectorwise</li> <li>• Vertica</li> <li>• Other SQL-89 compliant*</li> </ul>
Pentaho Metadata	Certified <ul style="list-style-type: none"> <li>• Azure SQL</li> <li>• Hive 2</li> <li>• Impala</li> <li>• MySQL</li> <li>• PostgreSQL</li> </ul> Supported <ul style="list-style-type: none"> <li>• Amazon Redshift</li> <li>• ASSQL</li> <li>• Firebird</li> <li>• H2</li> </ul>

Pentaho Software	Data Source
	<ul style="list-style-type: none"> <li>• Hypersonic</li> <li>• IBM DB2</li> <li>• IBM MQ 9.2</li> <li>• Ingres</li> <li>• Interbase</li> <li>• MS Access</li> <li>• MS SQL Server (JTDS Driver)</li> <li>• MS SQL Server (Microsoft Driver)</li> <li>• Netezza</li> <li>• Snowflake</li> <li>• Sybase</li> <li>• Vertica</li> <li>• Other SQL-92 compliant*</li> </ul>
Pentaho Data Integration	<p>Certified</p> <ul style="list-style-type: none"> <li>• Amazon Redshift</li> <li>• Azure SQL</li> <li>• Hive</li> <li>• Hive 2</li> <li>• Impala</li> <li>• MS SQL Server (JTDS Driver)</li> <li>• MS SQL Server (Microsoft Driver)</li> <li>• MySQL</li> <li>• Oracle</li> <li>• PostgreSQL</li> <li>• Snowflake</li> <li>• Vertica</li> </ul> <p>Supported</p> <ul style="list-style-type: none"> <li>• Apache Derby</li> <li>• AS/400</li> <li>• InfiniDB</li> <li>• Exasol 4</li> <li>• Firebird SQL</li> <li>• Greenplum</li> <li>• H2</li> <li>• Hypersonic</li> <li>• IBM DB2</li> <li>• IBM MQ 9.2</li> <li>• Informix</li> <li>• Ingres</li> <li>• Ingres VectorWise</li> <li>• MaxDB (SAP DB)</li> <li>• Neoview</li> <li>• Netezza</li> <li>• Oracle RDB</li> <li>• SAP HANA</li> <li>• SQLite</li> <li>• Teradata</li> <li>• UniVerse database</li> <li>• Other SQL-92 compliant*</li> </ul>
<p>* If your data source is not in this list and is compatible with SQL-92, Pentaho software uses a generic SQL dialect.</p>	

## Third-party libraries

Pentaho software is compatible with the following third-party web framework, file system, engine, and utility libraries:

- AngularJS 1.7.8
- HTTPClient 4.5.9
- Apache VFS 2.3
- Apache Axis2 1.7.9
- Apache Log4j 2.17.1

## Security

Pentaho software integrates with these third-party security authentication systems:

- Active Directory
- CAS 5.x and CAS 6.5
- Integrated Microsoft Windows Authentication
- LDAP
- RDBMS

## Java virtual machine

Pentaho software requirements for Java Runtime Environment (JRE).

Pentaho Software	Certified	Supported
All Pentaho software	<ul style="list-style-type: none"> <li>• Oracle Java 11.0.13</li> <li>• Oracle OpenJDK 11.0.13</li> </ul>	<ul style="list-style-type: none"> <li>• Oracle Java 8.x &amp; 11.x</li> <li>• Oracle OpenJDK 8.x &amp; 11.x</li> <li>• AdoptOpenJDK</li> <li>• Zulu from Azul Systems</li> </ul>

Note:

- Some Hadoop clusters using Java 8 may not be fully compatible when running Pentaho with Java 11.
- The PDI client requires Java 11.x to run on Windows 11.

## Web browsers

Pentaho supports major versions of web browsers that are publicly available six weeks prior to the finalization of a Pentaho release.

Certified Browsers	Supported Browsers
<ul style="list-style-type: none"> <li>• Apple Safari 15.6.1 (On macOS only)</li> <li>• Google Chrome 106.0.5249.119</li> <li>• Microsoft Edge 106.0.1370.42</li> <li>• Mozilla Firefox 105</li> </ul>	<ul style="list-style-type: none"> <li>• Apple Safari 15 (On macOS only) and later</li> <li>• Google Chrome 106 and later</li> <li>• Microsoft Edge 106 and later</li> <li>• Mozilla Firefox 105 and later</li> </ul>
*Linux requires libwebkitgtk-1.0. See <i>Install Pentaho Data Integration and Analytics</i> for more information.	

## Support statement for Analyzer on Impala

These are the minimum requirements for Analyzer to work with Impala:

- Pentaho 7.1 or later
- Cloudera CDH5.x, CDH 6.1, Impala 1.3.x or later
- Recommend using Parquet compressed file format for tables in Impala

- Recommendations for the Hive and Simba drivers. The driver to use depends on the following scenarios:

Scenario	Recommended Driver
Pentaho 8.3 or later with the CDH 5.14 shim	<a href="#">Impala JDBC Connector 2.5.43 Cloudera</a> driver.
Pentaho 8.3 or later with the CDH 6.1 driver	<a href="#">Impala JDBC Connector 2.6.4. Cloudera</a> driver.
Pentaho 9.0 or later with the CDH 6.1 driver	<a href="#">Impala JDBC Connector 2.6.4. Cloudera</a> driver.
Pentaho 9.1 or later with the CDP 7.1.4 driver	<a href="#">Impala JDBC Connector 2.6.4. Cloudera</a> driver.

- Make sure that the JDBC driver is dropped into the Pentaho Server and Schema Workbench directories. See the *Install Pentaho Data Integration and Analytics* document for details.
- Turn off connection pooling in Pentaho Server.
- Set global order by limit in Cloudera manager.
- In Mondrian schemas, divide dimension tables with high cardinality into several levels

Note: As with any data source, the performance of Pentaho Analyzer on Impala will be dependent upon the data shape, Impala's configuration, and the types of queries. See the best practice, "Pentaho Analyzer with Impala as a Data Source" located at: <https://support.pentaho.com/hc/en-us/articles/208652846> or download the [PDF](#).

There are some compiled Mondrian automated test suite results for Analyzer on Impala with OEM Simba, as well as the community Apache Hive driver:

- [Analyzer on Impala with OEM Simba](#)
- [Analyzer on Impala with community Apache Hive driver](#)

## Google BigQuery

You can use Google BigQuery as a data source with the Pentaho User Console or with the PDI client.

Before you begin, you must have a Google account and must create service account credentials in the form of a key file in JSON format to connect to Google BigQuery. To create service account credentials, see the [Google Cloud Storage Authentication documentation](#).

Additionally, you must set permissions for your BigQuery and Google Cloud accounts. To configure your service account authentication, see the [Google Service Account documentation](#).

Perform the following steps to create a JDBC connection to a Google BigQuery data source from the User Console or PDI client.

- Stop the Pentaho Server.
- Download the ZIP file containing the Simba version 1.2.2.1004 JDBC 4.2 driver for Google BigQuery from the [Google Cloud Simba Drivers site](#).
- Navigate to the server/pentaho-server/tomcat/webapps/pentaho/WEB-INF/lib directory for the User Console or the design-tools/data-integration/lib directory for the PDI client and delete any files associated with previous versions of Google BigQuery.  
Visually verify each file to ensure the older version is deleted.
- Extract the following files to the server/pentaho-server/tomcat/webapps/pentaho/WEB-INF/lib folder for the User Console or the design-tools/data-integration/lib directory for the PDI client.
  - animal-sniffer-annotations-1.14.jar
  - api-common-1.7.0.jar
  - avro-1.9.0.jar
  - checker-compact-qual-2.5.2.jar
  - error\_prone\_annotations-2.1.3.jar
  - gax-1.42.0.jar
  - gax-grpc-1.42.0.jar
  - google-api-client-1.28.0.jar
  - google-api-services-bigquery-v2-rev426-1.25.0.jar
  - google-auth-library-credentials-0.15.0.jar
  - google-auth-library-oauth2-http-0.13.0.jar



- o GoogleBigQueryJDBC42.jar
- o google-cloud-bigquerystorage-0.85.0-alpha.jar
- o google-cloud-core-1.67.0.jar
- o google-cloud-core-grpc-1.67.0.jar
- o google-http-client-1.29.0.jar
- o google-http-client-apache-2.0.0.jar
- o google-http-client-jackson2-1.28.0.jar
- o google-oauth-client-1.28.0.jar
- o grpc-alts-1.18.0.jar
- o grpc-auth-1.18.0.jar
- o grpc-context-1.18.0.jar
- o grpc-core-1.18.0.jar
- o grpc-google-cloud-bigquerystorage-v1beta1-0.50.0.jar
- o grpc-grpclb-1.18.0.jar
- o grpc-netty-shaded-1.18.0.jar
- o grpc-protobuf-1.18.0.jar
- o grpc-protobuf-lite-1.18.0.jar
- o grpc-stub-1.18.0.jar
- o gson-2.7.jar
- o j2objc-annotations-1.1.jar
- o javax.annotation-api-1.3.2.jar
- o jsr305-3.0.2.jar
- o opencensus-api-0.18.0.jar
- o opencensus-contrib-grpc-metrics-0.18.0.jar
- o opencensus-contrib-http-util-0.18.0.jar
- o protobuf-java-3.7.0.jar
- o protobuf-java-util-3.7.0.jar
- o proto-google-cloud-bigquerystorage-v1beta1-0.50.0.jar
- o proto-google-common-protos-1.15.0.jar
- o proto-google-iam-v1-0.12.0.jar
- o threetenbp-1.3.3.jar

Note: The Google BigQuery connection name does not display in the User Console Database Connection dialog box until you copy these files.

5. Restart the Pentaho Server.
6. Log on to the User Console or the PDI client, then open the Database Connection dialog box. See the *Install Pentaho Data Integration and Analytics* document for more information on the Database Connection dialog box.
7. In the Database Connection dialog box, select General, then select Google BigQuery as the Database Type.
8. In the Settings area, enter the information for your Google BigQuery account.
  - o The Host Name is the URL to Google's BigQuery web services API. For example, <https://www.googleapis.com/bigquery/v2>
  - o The Project ID in the PDI client and the Database name in the User Console are identical.
  - o The Port Number is 443.
9. Click Options, then add the following parameters and values.

Parameter	Value
OAuthType	0 (Zero)
OAuthServiceAcctEmail	Specify your service account email address.
OAuthPvtKeyPath	Specify the path to your private key credential file.
Timeout	Specify the amount of time, in seconds, before the server closes the connection. The recommended value is 120 seconds.

10. Click Test to verify that you can connect to your data.

## Compatibility issues running Pentaho on Java 11 with your Hadoop cluster

Pentaho running on Java 11 is compatible with Java 11 on your Hadoop clusters. If you only have Java 8 on your Hadoop clusters, some components may not be compatible when running Pentaho on Java 11. The following sections indicate which components are compatible with Pentaho running on Java 11 when using Java 8 on your Hadoop clusters.

## Cloudera 7.1 secured cluster

When connecting to a Cloudera 7.1 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 7.1 driver, the following component compatibilities apply:

### Compatible components

HDFS, Avro, Parquet, ORC, HBase, Hive, HadoopExecutor\*, Oozie, Pig, Pentaho MapReduce

### Non-compatible components

Sqoop

\* Compatible as long as the source JAR is compiled with Java 8.

## Hortonworks 3.1 secured cluster

When connecting to a Hortonworks 3.1 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Hortonworks 3.0 driver, the following component compatibilities apply:

### Compatible components

HDFS, Avro, Parquet, ORC, HBase, Hive, HadoopExecutor\*

### Non-compatible components

Sqoop, Oozie, Pig

\* Compatible as long as the source JAR is compiled with Java 8.

Note: The Pentaho MapReduce component was not tested.

## Cloudera 6.3 secured cluster

When connecting to a Cloudera 6.3 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 7.1 driver, the following component compatibilities apply:

### Compatible components

HDFS, Avro, Parquet, ORC, Oozie, Pig, Pentaho MapReduce

### Non-compatible components

HBase, Hive, HadoopExecutor, Sqoop

When connecting to a Cloudera 6.3 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 6.1 driver, the following component compatibilities apply:

### Compatible components

HDFS, Avro, Parquet, ORC, HBase, Hive, HadoopExecutor, Oozie, Pentaho MapReduce

### Non-compatible components

Sqoop, Pig

## Cloudera 6.2 secured cluster

When connecting to a Cloudera 6.2 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 7.1 driver, the following component compatibilities apply:

### Compatible components

HDFS, Avro, Parquet, ORC, HBase, HadoopExecutor, Oozie, Pig, Pentaho MapReduce

### Non-compatible components

Hive, Sqoop

When connecting to a Cloudera 6.2 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 6.1 driver, the following component compatibilities apply:

Compatible components

HDFS, Avro, Parquet, ORC, HBase, Hive, HadoopExecutor, Oozie, Pentaho MapReduce

Non-compatible components

Sqoop, Pig

## Cloudera 6.1 secured cluster

When connecting to a Cloudera 6.1 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 7.1 driver, the following component compatibilities apply:

Compatible components

HDFS, Avro, Parquet, ORC, HBase, HadoopExecutor\*, Oozie, Pig, Pentaho MapReduce

Non-compatible components

Hive, Sqoop

\* Compatible as long as the source JAR is compiled with Java 8.

When connecting to a Cloudera 6.1 secured cluster running in Java 8 from Pentaho running Java 11 while using the Pentaho Cloudera 6.1 driver, the following component compatibilities apply:

Compatible components

HDFS, Avro, Parquet, ORC, HBase, Hive, HadoopExecutor\*, Oozie

Non-compatible components

Sqoop, Pig\*\*, Pentaho MapReduce\*\*

\* Compatible as long as the source JAR is compiled with Java 8.

\*\* Missing method errors occur with Pig and Pentaho MapReduce.

## JDBC drivers reference

Attention: This reference article will eventually be removed from the product. If you have questions regarding JDBC drivers, contact your vendor or the [Support Portal](#).

This reference is subject to change. You can report any issues with this page to the [Support Portal](#).

Database	Vendor	URL
<a href="#">Amazon Redshift</a>	Amazon	<a href="http://docs.aws.amazon.com/redshift/latest/mgmt/configure-jdbc-connection.html">http://docs.aws.amazon.com/redshift/latest/mgmt/configure-jdbc-connection.html</a>
<a href="#">Apache Derby</a>	Apache	<a href="http://db.apache.org/derby/derby_downloads.html">http://db.apache.org/derby/derby_downloads.html</a>
<a href="#">Caché</a>	InterSystems	<a href="http://www.cachemonitor.de/intersystems-documentation/cache-jdbc-driver">http://www.cachemonitor.de/intersystems-documentation/cache-jdbc-driver</a>
<a href="#">CUBRID</a>	CUBRID	<a href="http://www.cubrid.org/?mid=downloadsitem=jdbc_driver">http://www.cubrid.org/?mid=downloadsitem=jdbc_driver</a>
<a href="#">Daffodil DB</a>	Daffodil Software	<a href="http://sourceforge.net/projects/daffodildb/">http://sourceforge.net/projects/daffodildb/</a>
<a href="#">DB2 AS/400</a>	IBM	<a href="https://www-03.ibm.com/systems/power/software/i/toolbox/downloads.html">https://www-03.ibm.com/systems/power/software/i/toolbox/downloads.html</a>
<a href="#">DB2 Universal Database</a>	IBM	<a href="http://www-01.ibm.com/support/docview.wss?uid=swg21363866">http://www-01.ibm.com/support/docview.wss?uid=swg21363866</a>

Database	Vendor	URL
<a href="#">Firebird</a>	Firebird Foundation	<a href="http://www.firebirdsql.org/en/jdbc-driver/">http://www.firebirdsql.org/en/jdbc-driver/</a>
<a href="#">FrontBase</a>	FrontBase	<a href="http://www.frontbase.com/cgi-bin/WebObjects/FBWebSite">http://www.frontbase.com/cgi-bin/WebObjects/FBWebSite</a>
<a href="#">Google BigQuery</a>	Google BigQuery	<a href="https://cloud.google.com/bigquery">https://cloud.google.com/bigquery</a>
<a href="#">Greenplum</a>	EMC2	<a href="http://jdbc.postgresql.org/download.html">http://jdbc.postgresql.org/download.html</a>
<a href="#">H2 Database</a>	H2	<a href="http://www.h2database.com">http://www.h2database.com</a>
<a href="#">Hive</a>	Apache	<a href="http://hive.apache.org/">http://hive.apache.org/</a>
<a href="#">Hive2</a>	Apache	<a href="http://hive.apache.org/">http://hive.apache.org/</a>
<a href="#">HSQLDB</a>	HyperSQL	<a href="http://sourceforge.net/projects/hsqldb/">http://sourceforge.net/projects/hsqldb/</a>
<a href="#">Impala</a>	Cloudera	<a href="https://www.cloudera.com/documentation.html">https://www.cloudera.com/documentation.html</a>
<a href="#">Informix</a>	IBM	<a href="https://www-01.ibm.com/software/data/informix/">https://www-01.ibm.com/software/data/informix/</a>
<a href="#">Ingres</a>	Action	<a href="http://esd.action.com/product/drivers/JDBC/java">http://esd.action.com/product/drivers/JDBC/java</a>
<a href="#">InterBase</a>	Embarcadero	<a href="http://edn.embarcadero.com">http://edn.embarcadero.com</a>
<a href="#">jTDS Free MS SQL Sybase</a>	jTDS	<a href="http://jtds.sourceforge.net/">http://jtds.sourceforge.net/</a>
<a href="#">MariaDB</a>	MariaDB	<a href="https://downloads.mariadb.org/connector-java/">https://downloads.mariadb.org/connector-java/</a>
<a href="#">MaxDB</a>	SAP	<a href="http://maxdb.sap.com">http://maxdb.sap.com</a>
<a href="#">Mckoi SQL Database</a>	Mckoi SQL Database	<a href="http://www.mckoi.com/originalmckoisql/index.html">http://www.mckoi.com/originalmckoisql/index.html</a>
<a href="#">Mimer</a>	Mimer Information Technology	<a href="http://www.mimer.com">http://www.mimer.com</a>
<a href="#">MySQL</a>	Oracle	<a href="https://dev.mysql.com/downloads/connector/j/">https://dev.mysql.com/downloads/connector/j/</a>
<a href="#">Neoview</a>	HP	Contact your local HP representative for information on this product.
<a href="#">Netezza</a>	IBM	<a href="http://www.netezza.com">http://www.netezza.com</a>
<a href="#">OpenBase SQL</a>	OpenBase International	<a href="http://www.openbase.com/index.php/products/downloads">http://www.openbase.com/index.php/products/downloads</a>
<a href="#">Oracle</a>	Oracle	<a href="http://www.oracle.com/technetwork/database/features/jdbc/index.html">http://www.oracle.com/technetwork/database/features/jdbc/index.html</a>
<a href="#">Pervasive</a>	Pervasive	<a href="http://www.pervasivedb.com/download/Pages/PDBDownloads.aspx">http://www.pervasivedb.com/download/Pages/PDBDownloads.aspx</a>
<a href="#">PostgreSQL</a>	PostgreSQL Global Development Group	<a href="http://jdbc.postgresql.org/">http://jdbc.postgresql.org/</a>
<a href="#">SAP ASE (formerly Sybase ASE)</a>	SAP	<a href="https://support.sap.com/swdc">https://support.sap.com/swdc</a>
<a href="#">SAP DB</a>	SAP MaxDB	<a href="https://support.sap.com/software.html">https://support.sap.com/software.html</a>
<a href="#">SAP HANA</a>	SAP	<a href="http://help.sap.com/hana">http://help.sap.com/hana</a>
<a href="#">SAP SQL Anywhere</a>	SAP	<a href="https://support.sap.com/software.html">https://support.sap.com/software.html</a>
<a href="#">SmallSQL</a>	SmallSQL	<a href="http://www.smallsql.de/download.html">http://www.smallsql.de/download.html</a>

Database	Vendor	URL
<a href="#">Snowflake</a>	Snowflake	<a href="https://repo1.maven.org/maven2/net/snowflake/snowflake-jdbc">https://repo1.maven.org/maven2/net/snowflake/snowflake-jdbc</a>
<a href="#">SQLite</a>	Xerial	<a href="http://www.xerial.org/trac/Xerial/wiki/SQLiteJDBC">http://www.xerial.org/trac/Xerial/wiki/SQLiteJDBC</a>
<a href="#">SQL Server</a>	Microsoft	<a href="http://msdn.microsoft.com/en-us/sqlserver/aa937724.aspx">http://msdn.microsoft.com/en-us/sqlserver/aa937724.aspx</a>
<a href="#">Teradata</a>	Teradata	<a href="http://downloads.teradata.com/download/connectivity/jdbc-driver">http://downloads.teradata.com/download/connectivity/jdbc-driver</a>
<a href="#">Vertica</a>	HP	<a href="http://www.vertica.com">http://www.vertica.com</a>

## Amazon Redshift

Vendor Name	Details	
Recommended Native Driver		
Amazon Web Services	Company URL: <a href="http://aws.amazon.com/redshift/">http://aws.amazon.com/redshift/</a>	
	Driver URL: <a href="https://s3.amazonaws.com/redshift-downloads/drivers/RedshiftJDBC4.jar">https://s3.amazonaws.com/redshift-downloads/drivers/RedshiftJDBC4.jar</a>	
	JDBC URL Syntax by Type:  jdbc:redshift://<endpoint>:<port>/<database>?tcpKeepAlive=true	Default Port:  5439
	JDBC Class:  com.amazon.redshift.jdbc4.Driver	JDBC JAR File Name:  RedshiftJDBC4.jar
	Comments:  Download and use only the RedshiftJDBC4.jar.	

## Apache Derby

Vendor Name	Details
Recommended Native Driver	
IBM	Company URL: <a href="http://www.ibm.com">http://www.ibm.com</a>
	Driver URL: <a href="http://db.apache.org/derby/derby_downloads.html">http://db.apache.org/derby/derby_downloads.html</a>

Vendor Name	Details	
	JDBC URL Syntax by Type:  Server jdbc:derby://<server>[:<port>]/<databaseName>[;<URL attribute>=>value>]  Embedded jdbc:derby:<databaseName>[;create=true]	Default Port:  1527
	JDBC Class:  org.apache.derby.jdbc.ClientDriver  org.apache.derby.jdbc.EmbeddedDriver	JDBC JAR File Name:  derby.jar
	Shipped with Pentaho products:  Pentaho Data Integration	
	Comments:  Open source database.	

## Caché

Vendor Name	Details	
Recommended Native Driver		
InterSystems	Company URL:  <a href="http://www.cachemonitor.de">http://www.cachemonitor.de</a>	
	Driver URL:  <a href="http://www.cachemonitor.de/intersystems-documentation/cache-jdbc-driver">http://www.cachemonitor.de/intersystems-documentation/cache-jdbc-driver</a>	
	JDBC URL Syntax by Type:  Server  jdbc:Cache://<server>[:<port>]/<namespace>	Default Port:  1972
	JDBC Class:  com.intersys.jdbc.CacheDriver	JDBC JAR File Name:  cachedb.jar

## CUBRID

Vendor Name	Details
Recommended Native Driver	

Vendor Name	Details	
CUBRID	Company URL: <a href="http://www.cubrid.org">http://www.cubrid.org</a>	
	Driver URL: <a href="http://www.cubrid.org/?mid=downloads&amp;item=jdbc_driver">http://www.cubrid.org/?mid=downloads&amp;item=jdbc_driver</a>	
	JDBC URL Syntax by Type:  Server jdbc:cubrid:<server>:<port>: <databaseName>:<username>: <password> :[?<URL attribute>=<value> [&<URL attribute> =<value>] ... ]	Default Port:  33000
	JDBC Class:  cubrid.jdbc.driver.CUBRIDDriver	JDBC JAR File Name:  N/A
	Comments:  Open source database highly optimized for Web applications.	

## Daffodil DB

Vendor Name	Details	
Recommended Native Driver		
Daffodil Software	Company URL: <a href="http://db.daffodilsw.com">http://db.daffodilsw.com</a>	
	Driver URL: <a href="http://sourceforge.net/projects/daffodildb/">http://sourceforge.net/projects/daffodildb/</a>	
	JDBC URL Syntax by Type:  Server jdbc:daffodilDB://<server>[:<port>]/<data baseName>  Embedded jdbc:daffodilDB_embedded:<databaseName>	Default Port:  3456  N/A
	JDBC Class:  in.co.daffodil.db.rmi.RmiDaffodilDBDriver  in.co.daffodil.db.jdbc.DaffodilDBDriver	JDBC JAR File Name:  DaffodilDB_client.jar  DaffodilDB_Embedded.jar,

Vendor Name	Details	
		DaffodilDB_Common.jar
	Comments: Open source database.	

## DB2 AS/400

Vendor Name	Details	
Recommended Native Driver		
IBM	Company URL: <a href="http://www.ibm.com">http://www.ibm.com</a>	
	Driver URL: <a href="http://www-01.ibm.com/support/docview.wss?uid=swg21363866">http://www-01.ibm.com/support/docview.wss?uid=swg21363866</a> <a href="http://www-03.ibm.com/systems/power/software/i/toolbox/downloads.html">http://www-03.ibm.com/systems/power/software/i/toolbox/downloads.html</a>	
	JDBC URL Syntax by Type:  Server  jdbc:as400://<server>naming=sql;errors=full	Default Port:  N/A
	JDBC Class:  com.ibm.as400.access.AS400JDBCDriver	JDBC JAR File Name:  jt400.jar

## DB2 Universal Database

Vendor Name	Details	
Recommended Native Driver		
IBM	Company URL: <a href="http://www.ibm.com">http://www.ibm.com</a>	
	Driver URL: <a href="http://www-01.ibm.com/support/docview.wss?uid=swg21363866">http://www-01.ibm.com/support/docview.wss?uid=swg21363866</a>	
	JDBC URL Syntax by Type:  Server  jdbc:db2://<server>[:<port>]/<databaseName>[:<URL attribute>=<value>;<URL attribute>=<value>]	Default Port:  50000



Vendor Name	Details	
	JDBC Class: <code>com.ibm.db2.jcc.DB2Driver</code>	JDBC JAR File Name: <code>db2jcc4.jar</code>

## Firebird

Vendor Name	Details	
Recommended Native Driver		
Firebird Foundation	Company URL: <a href="http://www.firebirdsql.org">http://www.firebirdsql.org</a>	
	Driver URL: <a href="http://www.firebirdsql.org/en/jdbc-driver/">http://www.firebirdsql.org/en/jdbc-driver/</a>	
	JDBC URL Syntax by Type:	Default Port:
	Server <code>jdbc:firebirdsql:&lt;server&gt;[/&lt;port&gt;]:/&lt;database-file&gt;</code>  (JDBC Type 4, official format)	3050  3050  3050
	Server <code>jdbc:firebirdsql://&lt;server&gt;[:&lt;port&gt;]/&lt;database-file&gt;</code>  (JDBC Type 4, compatibility format)	3050  N/A
Server <code>jdbc:firebirdsql:native//&lt;server&gt;[/&lt;port&gt;]:/&lt;database-file&gt;</code>  (JDBC Type 2, compatibility format)		
Server <code>jdbc:firebirdsql:native://&lt;server&gt;[:&lt;port&gt;]/&lt;database-file&gt;</code>  JDBC Type 2, compatibility format. Requires libraries)		
Embedded <code>jdbc:firebirdsql:embedded:/&lt;local-database-file&gt;</code>  (JDBC Type 2, compatibility format. Requires libraries)		
JDBC Class: <code>org.firebirdsql.jdbc.FBDriver</code>	JDBC JAR File Name:  <code>jaybird-full-xxx.jar</code>	

Vendor Name	Details
	<p>Shipped with Pentaho products:</p> <p>Pentaho Data Integration</p>

## FrontBase

Vendor Name	Details	
Recommended Native Driver		
FrontBase	Company URL: <a href="http://www.frontbase.com">http://www.frontbase.com</a>	
	Driver URL: <a href="http://www.frontbase.com/cgi-bin/WebObjects/FBWebSite">http://www.frontbase.com/cgi-bin/WebObjects/FBWebSite</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:FrontBase://&lt;host&gt;[:&lt;port&gt;]/&lt;databaseName&gt;</code>	Default Port:  N/A
	JDBC Class:  <code>com.frontbase.jdbc.FBDriver</code>	JDBC JAR File Name:  <code>frontbasejdbc.jar</code>

## Google BigQuery

Vendor Name	Details
Recommended Native Driver	
Google BigQuery	For installation instructions, see <a href="#">Google BigQuery</a> .
	Company URL:
	<a href="https://cloud.google.com/bigquery">https://cloud.google.com/bigquery</a>
	Driver URL:
	<a href="https://cloud.google.com/bigquery/partners/simba-drivers/">https://cloud.google.com/bigquery/partners/simba-drivers/</a>
	Host URL:
	<a href="https://bigquery.googleapis.com/discovery/v1/apis/bigquery/v2/rest">https://bigquery.googleapis.com/discovery/v1/apis/bigquery/v2/rest</a>
	Default Port:

Vendor Name	Details
	443
	Simba JDBC Driver URL: <a href="https://cloud.google.com/bigquery/providers/simba-drivers">https://cloud.google.com/bigquery/providers/simba-drivers</a>

## Greenplum

Vendor Name	Details	
Recommended Native Driver		
Greenplum	Company URL: <a href="http://greenplum.org/">http://greenplum.org/</a>	
	Driver URL: <a href="http://jdbc.postgresql.org/download.html">http://jdbc.postgresql.org/download.html</a>	
	JDBC URL Syntax by Type:  Server  jdbc:postgresql://<server>[:<port>]/<databaseName>	Default Port:  5342
	JDBC Class:  org.postgresql.Driver	JDBC JAR File Name:  postgresql-8.x-xxx.jdbc4.jar
	Comments:  Greenplum uses the Postgresql JDBC driver.	

## H2 Database

Vendor Name	Details	
Recommended Native Driver		
H2	Company URL: <a href="http://www.h2database.com">http://www.h2database.com</a>	
	Driver URL: <a href="http://www.h2database.com">http://www.h2database.com</a>	
	JDBC URL Syntax by Type:	Default Port:

Vendor Name	Details	
	Server jdbc:h2:tcp://server[:port]/file-path Embedded jdbc:h2:file-name	9092  N/A
	JDBC Class:  jdbc:h2:tcp://server[:port]/file-path  org.h2.Driver	JDBC JAR File Name:  h2-x.x.xxx.jar
	Shipped with Pentaho products: <ul style="list-style-type: none"> <li>• Pentaho Server</li> <li>• Pentaho Data Integration</li> <li>• Pentaho Metadata Editor</li> <li>• Pentaho Report Designer</li> </ul>	
	Comments:  Open source Java SQL database.	

## Hive

Vendor Name	Details	
Recommended Native Driver		
Apache	Company URL:  <a href="http://hive.apache.org/">http://hive.apache.org/</a>	
	Driver URL:  <a href="https://www.cloudera.com/downloads/connectors/hive/jdbc/2-6-21.html">https://www.cloudera.com/downloads/connectors/hive/jdbc/2-6-21.html</a>  or  <a href="https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc">https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc</a>	
	JDBC URL Syntax by Type:  Server  jdbc:hive://<server>[:<port>]/default	Default Port:  10000
	JDBC Class:  org.apache.hadoop.hive.jdbc.HiveDriver	JDBC JAR File Name:  pentaho-hadoop-shims-common-fragment-Vx-x.x.x.jar
	Shipped with Pentaho products:	

Vendor Name	Details
	<ul style="list-style-type: none"> <li>• Pentaho Server</li> <li>• Pentaho Data Integration</li> <li>• Pentaho Metadata Editor</li> <li>• Pentaho Report Designer</li> </ul>
	<p>Comments:</p> <p>The pentaho-hadoop-shims-common-fragment-Vx-x.x.x.jar library includes a proxy driver. The actual Hive JDBC implementation for the specific distribution and version of Hadoop is located in the Pentaho driver for that distro. See the <i>Install Pentaho Data Integration and Analytics</i> document for more information.</p> <p>Hive does not support the full SQL capabilities. It uses a subset and is more accurately referred to as HiveQL.</p>

## Hive2

Vendor Name	Details	
Recommended Native Driver		
Apache	Company URL: <a href="http://hive.apache.org/">http://hive.apache.org/</a>	
	Driver URL: <a href="https://www.cloudera.com/downloads/connectors/hive/jdbc/2-6-21.html">https://www.cloudera.com/downloads/connectors/hive/jdbc/2-6-21.html</a>  or <a href="https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc">https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc</a>	
	JDBC URL Syntax by Type:  Server  jdbc:hive2://<server>[:<port>]/<db>	Default Port:  10000
	JDBC Class:  org.apache.hive.jdbc.HiveDriver	JDBC JAR File Name:  pentaho-hadoop-shims-common-fragment-Vx-x.x.x.jar
	Shipped with Pentaho products: <ul style="list-style-type: none"><li>• Pentaho Server</li><li>• Pentaho Data Integration</li><li>• Pentaho Metadata Editor</li><li>• Pentaho Report Designer</li></ul>	

Vendor Name	Details
	<p>Comments:</p> <p>The pentaho-hadoop-shims-common-fragment-Vx-x.x.x.jar library includes a proxy driver. The actual Hive JDBC implementation for the specific distribution and version of Hadoop is located in the Pentaho driver for that distro. See the <i>Install Pentaho Data Integration and Analytics</i> document for more information.</p> <p>Hive2 does not support the full SQL capabilities. It uses a subset and is more accurately referred to as HiveQL.</p>

## HSQldb

Vendor Name	Details	
Recommended Native Driver		
HyperSQL	Company URL: <a href="http://www.hsqldb.org">http://www.hsqldb.org</a>	
	Driver URL: <a href="http://sourceforge.net/projects/hsqldb/">http://sourceforge.net/projects/hsqldb/</a>	
	JDBC URL Syntax by Type:  Server jdbc:hsqldb:hsqldb://<server>[:<port>]/<databaseName>  Embedded Memory jdbc:hsqldb:mem:<databaseName>  Embedded File jdbc:hsqldb:file:<database-file>	Default Port:  9001  N/A  N/A
	JDBC Class:  org.hsqldb.jdbcDriver	JDBC JAR File Name:  hsqldb.jar
	Shipped with Pentaho products: <ul style="list-style-type: none"><li>• Pentaho User Console</li><li>• Pentaho Server</li><li>• Pentaho Aggregation Designer</li><li>• Pentaho Data Integration</li><li>• Pentaho Metadata Editor</li><li>• Pentaho Report Designer</li></ul>	

## Impala

Vendor Name	Details	
Recommended Native Driver		
Cloudera	Company URL: <a href="http://cloudera.com">http://cloudera.com</a>	
	Driver URL:  Built in	
	JDBC URL Syntax by Type:  jdbc:hive2://<server>[<port>]//;auth=noSasl	Default Port:  21050
	JDBC Class:  org.apache.hadoop.hive.jdbc.HiveDriver	JDBC JAR File Name:  pentaho-hadoop-shims-common-fragment-Vx-x.x.x.jar
	Shipped with Pentaho products: <ul style="list-style-type: none"><li>• Pentaho Server</li><li>• Pentaho Data Integration</li><li>• Pentaho Metadata Editor</li><li>• Pentaho Report Designer</li></ul>	
	Comments:  The pentaho-hadoop-shims-common-fragment-Vx-x.x.x.jar library includes a proxy driver. The actual Hive JDBC implementation for the specific distribution and version of Hadoop is located in the Pentaho driver for that distro. See the <i>Install Pentaho Data Integration and Analytics</i> document for more information.  Hive does not support the full SQL capabilities. It uses a subset and is more accurately referred to as HiveQL.	

## Informix

Vendor Name	Details
Recommended Native Driver	
IBM	Company URL: <a href="http://www.ibm.com">http://www.ibm.com</a>
	Driver URL: <a href="http://www14.software.ibm.com/webapp/download/search.jsp?go=y&amp;rs=ifxjdbc">http://www14.software.ibm.com/webapp/download/search.jsp?go=y&amp;rs=ifxjdbc</a>

Vendor Name	Details	
	JDBC URL Syntax by Type:  Server jdbc:informix-sqli://<server>[:<port>]/<databaseName>: informixserver= <dbservername>	Default Port:  1533
	JDBC Class:  com.informix.jdbc.IfxDriver	JDBC JAR File Name:  ifxjdbc.jar

## Ingres

Vendor Name	Details	
Recommended Native Driver		
Action	Company URL: <a href="http://www.action.com/">http://www.action.com/</a>	
	Driver URL: <a href="http://esd.action.com/product/drivers/JDBC/java">http://esd.action.com/product/drivers/JDBC/java</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:ingres://&lt;server&gt;[:&lt;port&gt;]/&lt;databaseName&gt;</code>	Default Port:  21071
	JDBC Class:  <code>com.ingres.jdbc.IngresDriver</code>	JDBC JAR File Name:  <code>ijjdbc.jar</code>
	Comments:  Open source relational database management system.	

## InterBase

Vendor Name	Details
Recommended Native Driver	
Embarcadero	Company URL:  <a href="http://edn.embarcadero.com">http://edn.embarcadero.com</a>



Vendor Name	Details	
	Driver URL: N/A	
	JDBC URL Syntax by Type:  Server <code>jdbc:interbase://&lt;server&gt;/&lt;full_db_path&gt;</code>	Default Port:  N/A
	JDBC Class  <code>interbase.interclient.Driver</code>	JDBC JAR File Name:  <code>interclient.jar</code>

## jTDS Free MS SQL Sybase

Vendor Name	Details	
Recommended Native Driver		
jTDS	Company URL: <a href="http://jtds.sourceforge.net/">http://jtds.sourceforge.net/</a>	
	Driver URL:  N/A	
	JDBC URL Syntax by Type:  SQL Server jdbc:jtds:<server_type>//<server>[:<port>][/<database>][;<property>=<value>[;...]]]  Sybase jdbc:jtds:<server_type>://<server>[:<port>][/<database>e]	Default Port:  1433  7100
	JDBC Class:  interbase.interclient.Driver	JDBC JAR File Name:  jtds-x.x.x.jar
	Shipped with Pentaho products: <ul style="list-style-type: none"><li>• Pentaho User Console</li><li>• Pentaho Server</li><li>• Pentaho Aggregation Designer</li><li>• Pentaho Data Integration</li><li>• Pentaho Metadata Editor</li><li>• Pentaho Report Designer</li></ul>	

## MariaDB

Vendor Name	Details	
Recommended Native Driver		
MariaDB	Company URL: <a href="http://mariadb.org">http://mariadb.org</a>	
	Driver URL: <a href="https://downloads.mariadb.org/connector-java/">https://downloads.mariadb.org/connector-java/</a>	
	JDBC URL Syntax by Type:  Server <pre>jdbc:mariadb://&lt;hostname&gt;[,&lt;failoverhost&gt;][:&lt;port&gt;] / &lt;dbname&gt;[?&lt;URL attribute&gt;=&lt;value&gt;[&amp;&lt;URL attribute&gt;=&lt;v alue&gt;] ... ]</pre>	Default Port:  3306
	JDBC Class:  org.mariadb.jdbc.Driver	JDBC JAR File Name:  mariadb-java-client-2.1.2.jar

## MaxDB

Vendor Name	Details	
Recommended Native Driver		
SAP	Company URL: <a href="http://www.sap.com">http://www.sap.com</a>	
	Driver URL: <a href="http://maxdb.sap.com">http://maxdb.sap.com</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:sapdb://&lt;server&gt;[:&lt;port&gt;]/&lt;databaseName&gt;</code>	Default Port:  7210
	JDBC Class: <code>com.sap.dbtech.jdbc.DriverSapDB</code>	JDBC JAR File Name: <code>sapdbc.jar</code>
	Comments:	

Vendor Name	Details
	Database management system developed and supported by SAP AG.

## Mckoi SQL Database

Vendor Name	Details
Recommended Native Driver	
Mckoi	Company URL: <a href="http://www.mckoi.com">http://www.mckoi.com</a>
	Driver URL: <a href="http://www.mckoi.com/originalmckoisql/index.html">http://www.mckoi.com/originalmckoisql/index.html</a>
	JDBC URL Syntax by Type: Server <code>jdbc:mckoi://&lt;server&gt;[:&lt;port&gt;][/&lt;schema&gt;]/</code>
	Default Port: 9157
	JDBC Class: <code>com.mckoi.JDBCdriver</code>
	DBC JAR File Name: <code>mckoidb.jar</code>
	Comments: Open source SQL database written in Java.

## Mimer

Vendor Name	Details
Recommended Native Driver	
Mimer Information Technology	Company URL: <a href="http://www.mimer.com">http://www.mimer.com</a>
	Driver URL: N/A
	JDBC URL Syntax by Type: Server <code>jdbc:mimer:&lt;protocol&gt;://&lt;server&gt;[:&lt;port&gt;]/&lt;database&gt;</code>
	Default Port: 1360

Vendor Name	Details	
	JDBC Class: <code>com.mimer.jdbc.Driver</code>	JDBC JAR File Name: <code>mimer.jar</code>

## MySQL

Vendor Name	Details	
Recommended Native Driver		
Oracle	Company URL: <a href="http://www.mysql.com">http://www.mysql.com</a>	
	Driver URL: <a href="https://dev.mysql.com/downloads/connector/j/">https://dev.mysql.com/downloads/connector/j/</a>	
	JDBC URL Syntax by Type:  Server  jdbc:mysql://<hostname>[,<failoverhost>][:<port>] /<dbname>[?<URL attribute>=<value>[&<URL attribute>=<value>] ... ]	Default Port:  3306
	JDBC Class:  com.mysql.jdbc.Driver (official class name)  org.gjt.mm.mysql.Driver (older class name)	JDBC JAR File Name:  mysql-connector-java-5.x.xx-bin.jar
	Comments: Version 5.7 is supported, .8.026 is certified.	

## Neoview

Vendor Name	Details	
Recommended Native Driver		
HP	Company URL: <a href="http://www.hp.com">http://www.hp.com</a>	
	Driver URL: <a href="http://storiedigitali.live/hp-neoview-jdbc-63.html">http://storiedigitali.live/hp-neoview-jdbc-63.html</a>	
	JDBC URL Syntax by Type:	Default Port:

Vendor Name	Details	
	Server jdbc:hpt4jdbc://<system>[:<port>]/[:][<URL attribute>=<value>[;<URL attribute>=<value> ...]	18650
	JDBC Class: com.hp.t4jdbc.HPT4Driver	JDBC JAR File Name: N/A

## Netezza

Vendor Name	Details	
Recommended Native Driver		
IBM	Company URL: <a href="http://www.netezza.com">http://www.netezza.com</a>	
	Driver URL: N/A	
	JDBC URL Syntax by Type: <code>jdbc:netezza://&lt;server&gt;[:&lt;port&gt;]/&lt;database&gt;</code>	Default Port: 5480
	JDBC Class: <code>org.netezza.Driver</code>	JDBC JAR File Name: N/A

## OpenBase SQL

Vendor Name	Details	
Recommended Native Driver		
OpenBase International	Company URL: <a href="http://www.openbase.com">http://www.openbase.com</a>	
	Driver URL: <a href="http://www.openbase.com/index.php/products/downloads">http://www.openbase.com/index.php/products/downloads</a>	
	JDBC URL Syntax by Type:  Server  jdbc:openbase://<server>/<databaseName> >	Default Port:  N/A

Vendor Name	Details	
	JDBC Class: com.openbase.jdbc.ObDriver	JDBC JAR File Name: OpenBaseJDBC.jar

## Oracle

Vendor Name	Details	
Recommended Native Driver		
Oracle	Company URL: <a href="http://www.oracle.com">http://www.oracle.com</a>	
	Driver URL: <a href="http://www.oracle.com/technetwork/database/features/jdbc/index.html">http://www.oracle.com/technetwork/database/features/jdbc/index.html</a>	
	JDBC URL Syntax by Type:  Thin Server jdbc:oracle:thin:@<server>[:<port>]: <sid> OCI Server jdbc:oracle:oci:@<server>[:<port>]: <sid>	Default Port:  1521
	JDBC Class:  oracle.jdbc.driver.OracleDriver  oracle.jdbc.OracleDriver	JDBC JAR File Name:  ojdbcx.jar  orai18n.jar
	Comments:  The OCI server requires OCI libraries.	

## Pervasive

Vendor Name	Details	
Recommended Native Driver		
Pervasive	Company URL: <a href="http://www.pervasivedb.com/Pages/default.aspx">http://www.pervasivedb.com/Pages/default.aspx</a>	
	Driver URL: <a href="http://www.pervasivedb.com/download/Pages/PDBDownloads.aspx">http://www.pervasivedb.com/download/Pages/PDBDownloads.aspx</a>	
	JDBC URL Syntax by Type:	Default Port:

Vendor Name	Details	
	Server jdbc:pervasive://<server>[:<port>]/<datasource>	1583
	JDBC Class: com.pervasive.jdbc.v2.Driver	JDBC JAR File Name: N/A
	Shipped with Pentaho products: <ul style="list-style-type: none"> <li>• Pentaho Data Integration</li> <li>• Pentaho Report Designer</li> </ul>	
	Comments: The data source is the ODBC DSN.	

## PostgreSQL

Vendor Name	Details	
Recommended Native Driver		
PostgreSQL Global Development Group	Company URL: <a href="http://www.postgresql.org/">http://www.postgresql.org/</a>	
	Driver URL: <a href="http://jdbc.postgresql.org/">http://jdbc.postgresql.org/</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:postgresql://&lt;server&gt;[:&lt;port&gt;]/&lt;databaseName&gt;</code>	Default Port:  5342
	JDBC Class:  org.postgresql.Driver	JDBC JAR File Name:  postgresql-9.x-xxx.jdbc4.2.jar
	Shipped with Pentaho products: <ul style="list-style-type: none"><li>• Pentaho Data Integration</li><li>• Pentaho Report Designer</li></ul>	

## SAP ASE (formerly Sybase ASE)

Vendor Name	Details	
Recommended Native Driver		
SAP	Company URL: <a href="https://www.sap.com/products/sybase-ase.html">https://www.sap.com/products/sybase-ase.html</a>	
	Driver URL:  The jConnect JDBC driver can only be installed from the SAP Adaptive Server Enterprise Installer. See the SAP website for more information.	
	JDBC URL Syntax by Type:  Server  jdbc:sybase:Tds:<server>[:<port>]/<databaseName>	Default Port:  5000
	JDBC Class:  com.sybase.jdbc4.jdbc.SybDriver	JDBC JAR File Name:  N/A
	Comments:  The open source jTDS driver works with SAP ASE (formerly Sybase) as well. Note that although you can use jTDS open source JDBC driver, we recommend that you use the SAP-supplied JDBC driver instead. Connections might not work reliably if you use the jTDS JDBC driver.	

## SAP DB

Vendor Name	Details	
Recommended Native Driver		
SAP DB	Company URL:  N/A	
	Driver URL:  <a href="http://www.sapdb.org/sap_db_jdbc.htm">http://www.sapdb.org/sap_db_jdbc.htm</a>	
	JDBC URL Syntax by Type:  Server  jdbc:sapdb://<server>/<database_name>	Default Port:  N/A
	JDBC Class:  com.sap.dbtech.jdbc.DriverSapDB	JDBC JAR File Name:  sapdbc-x.x.x.jar



Vendor Name	Details
	Shipped with Pentaho products: Pentaho Data Integration
	Comments: FREE Enterprise Open Source Database.

## SAP HANA

Vendor Name	Details	
Recommended Native Driver		
SAP HANA	Company URL: <a href="http://hana.sap.com">http://hana.sap.com</a>	
	Driver URL:  For SAP customers, the driver is part of your client tools. Contact your SAP representative for more information.	
	JDBC URL Syntax by Type:  Server <code>jdbc:sap://&lt;server&gt;:&lt;port&gt;/?autocommit=false</code>	Default Port:  30015
	JDBC Class:  <code>com.sap.db.jdbc.Driver</code>	JDBC JAR File Name:  <code>ngdbc.jar</code>
	Comments:  Note that the default port number is '30015' where '00' is the instance of the machine you are connecting to. For example, you can connect to the same machine using '30015', '30115', or '31015'.	

## SAP SQL Anywhere

Vendor Name	Details
Recommended Native Driver	
SAP	Company URL: <a href="https://www.sap.com/products/sql-anywhere.html">https://www.sap.com/products/sql-anywhere.html</a>
	Driver URL:

Vendor Name	Details	
	The jConnect JDBC driver can only be installed from the SAP Adaptive Server Enterprise Installer. Visit the SAP website for more information.	
	JDBC URL Syntax by Type:  Server <code>jdbc:sybase:Tds:&lt;server&gt;[:&lt;port&gt;]/&lt;databaseName&gt;</code>	Default Port:  2638
	JDBC Class:  <code>com.sybase.jdbc4.jdbc.SybDriver</code>	JDBC JAR File Name:  N/A
	Comments:  This open source jTDS driver works with SAP ASE (formerly Sybase) as well. Note that although you can use jTDS open source JDBC driver, we recommend that you use the SAP-supplied JDBC driver instead. Connections might not work reliably if you use the jTDS JDBC driver.	

## SmallSQL

Vendor Name	Details	
Recommended Native Driver		
SmallSQL	Company URL: <a href="http://www.smallsql.de/">http://www.smallsql.de/</a>	
	Driver URL: <a href="http://www.smallsql.de/download.html">http://www.smallsql.de/download.html</a>	
	JDBC URL Syntax by Type:  Embedded <code>jdbc:smallsql:databaseName[?URL attribute=value[URL attribute=value]...]</code>	Default Port:  N/A
	JDBC Class:  <code>smallsql.database.SSDriver</code>	JDBC JAR File Name:  <code>smallsql.jar</code>
	Comments  Java desktop SQL database engine.	

## Snowflake

Vendor Name	Details	
Recommended Native Driver		
Snowflake	Company URL: <a href="http://snowflake.com">http://snowflake.com</a>	
	Driver URL: <a href="https://repo1.maven.org/maven2/net/snowflake/snowflake-jdbc">https://repo1.maven.org/maven2/net/snowflake/snowflake-jdbc</a>	
	JDBC URL Syntax by Type:  Embedded <code>jdbc:snowflake://&lt;account_name&gt;.snowflakecomputing.com/?&lt;connection_params&gt;</code>	Default Port:  443
	JDBC Class:  <code>net.snowflake.client.jdbc.SnowflakeDriver</code>	JDBC JAR File Name:  <code>snowflake-jdbc-3.6.28.jar</code>
	Comments: Version 3.6.28 is the minimum version supported. If timeout errors occur, see <i>Pentaho Data Integration</i> to troubleshoot.	

## SQLite

Vendor Name	Details	
Recommended Native Driver		
Xerial	Company URL:  N/A	
	Driver URL:  <a href="http://www.xerial.org/trac/Xerial/wiki/SQLiteJDBC">http://www.xerial.org/trac/Xerial/wiki/SQLiteJDBC</a>	
	JDBC URL Syntax by Type:  Server  jdbc:sqlite:<filename.db>	Default Port:  N/A
	JDBC Class:  org.sqlite.JDBC	JDBC JAR File Name:  sqlite-jdbc-x.x.x.jar
	Shipped with Pentaho products:	

Vendor Name	Details
	Pentaho Data Integration

## SQL Server

Vendor Name	Details	
Recommended Native Driver		
Microsoft	Company URL: <a href="http://www.microsoft.com">http://www.microsoft.com</a>	
	Driver URL: <a href="http://msdn.microsoft.com/en-us/sqlserver/aa937724.aspx">http://msdn.microsoft.com/en-us/sqlserver/aa937724.aspx</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:sqlserver://&lt;server&gt;[:&lt;port&gt;];DatabaseName=&lt;databaseName&gt;</code>	Default Port:  1433
	JDBC Class:  <code>com.microsoft.sqlserver.jdbc. SQLServerDriver</code>	JDBC JAR File Name:  <code>sqljdbc4.jar</code>
	Comments:  The open source jtds driver also works with MSSQL.	

## Teradata

Vendor Name	Details	
Please see the <a href="#">Teradata JDBC Driver Reference</a> for information about required drivers.		
Teradata	Company URL: <a href="http://www.teradata.com">http://www.teradata.com</a>	
	Driver URL: <a href="http://downloads.teradata.com/download/connectivity/jdbc-driver">http://downloads.teradata.com/download/connectivity/jdbc-driver</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:teradata://&lt;dbshost&gt;[/&lt;URL attribute&gt;[:&lt;URL attribute&gt;]...]</code>	Default Port:  N/A

Vendor Name	Details	
	JDBC Class: <code>com.teradata.jdbc.TeraDriver</code>	JDBC JAR File Name: <code>terajdbc4.jar</code> <code>tdgssconfig.jar</code>

## Vertica

Vendor Name	Details	
Recommended Native Driver		
HP	Company URL: <a href="http://www.vertica.com">http://www.vertica.com</a>	
	Driver URL: <a href="http://www.vertica.com/support/myvertica/">http://www.vertica.com/support/myvertica/</a>	
	JDBC URL Syntax by Type:  Server <code>jdbc:vertica://&lt;server&gt;[:&lt;port&gt;]/&lt;databaseName&gt;</code>	Default Port:  5433
	JDBC Class:  <code>com.vertica.Driver</code>	JDBC JAR File Name:  N/A

## Install drivers with the JDBC distribution tool

To connect to a database, including the Pentaho Repository database, you will need to download and install a JDBC driver to the appropriate places for Pentaho components as well as on the the web application server that contains the Pentaho Server.

Perform the following steps to download a JDBC driver and install the JDBC driver using the JDBC Distribution Tool.

Note: Due to licensing restrictions, Pentaho cannot redistribute some third-party database drivers. You must download the file yourself and install it yourself.

1. Download a [JDBC driver](#) JAR from your database vendor or a third-party driver developer.
2. Copy the JDBC driver JAR you just downloaded to the `pentaho/jdbc-distribution` directory.
3. Open a cmd prompt or shell tool, navigate to the `pentaho/jdbc-distribution` directory and enter one of the following:
  - o Windows:

```
distribute-files.bat <name of JDBC driver JAR>
```

- o Linux:

```
./distribute-files.sh
```

4. If you have run this utility as part of the installation process, you are done. Proceed to the next step of the installation instructions.
5. If you have run this utility so that you can connect to a new repository, restart the Pentaho Server and design tools, then try to connect to the new repository. If you cannot connect, verify that the drivers are installed as shown in the table below. Restart your Pentaho Server and client tools.

List of Products and Corresponding Locations for JDBC Drivers	
Server or Design Tool	Directory
Pentaho Server	pentaho/server/pentaho-server/tomcat/lib
Pentaho Data Integration (Spoon)	pentaho/design-tools/data-integration/lib
Pentaho Report Designer (PRD)	pentaho/design-tools/report-designer/lib/jdbc
Pentaho Aggregation Designer (PAD)	pentaho/design-tools/aggregation-designer/drivers
Pentaho Schema Workbench (PSW)	pentaho/design-tools/schema-workbench/drivers
Pentaho Metadata Editor (PME)	pentaho/design-tools/metadata-editor/libext/JDBC