



Choose Your Workflows



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

## Help and Support Resources

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

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

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

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

## Liability Limits and Warranty Disclaimer

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

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

## Trademarks

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

## Third-Party Open Source Software

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

## Contact Us

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

Sales Inquiries: [sales@pentaho.com](mailto:sales@pentaho.com)

# Contents

Prepare for the Evaluation.....	4
Choose Your Evaluation Track.....	4
Choose Evaluation Method.....	4
Use the Hosted Demo.....	5
Custom Pentaho Prototype.....	5
Trial Download.....	5
Next Steps.....	6
Explore and Learn Pentaho Business Analytics.....	7
Tour the User Console and Create Your First Reports.....	7
Explore and Learn Data Source Basics.....	7
Learn About Report Designer.....	8
Learn More About Pentaho Business Analytics.....	9
Explore and Learn Pentaho Data Integration.....	10
PDI Basics.....	10
Explore Big Data and Instaview.....	10
Learning About Kitchen, Pan, and Carte.....	11
Next Steps.....	11

# Prepare for the Evaluation

We provide two tracks, the Business Analytics track and the Data Integration track, that you can take as you evaluate our products. 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.

## Choose Your Evaluation Track

Choose the track for evaluation.

**Table 1: Pentaho Evaluation Decision Table**

Explore Considerations	Choose Options	
	<i>Business Analytics Evaluation</i>	<i>Data Integration Evaluation</i>
<b>Expertise</b>	<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>
<b>Data Set Description</b>	<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>
<b>Reporting Options</b>	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.
<b>Data Storage Types</b>	<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>
<b>Recommendation</b>	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

## Choose Evaluation Method

Now that you have chosen your installation path, choose how the evaluation software will be hosted. Review the decision table and choose the option that best meets your needs.

**Table 2: Pentaho Evaluation Decision Table**

Explore Considerations	Choose Options		
	Hosted Demo	Custom Prototype	Trial Download
<b>Track</b>	BA	BA or DI	BA or DI

Explore Considerations	Choose Options		
	Hosted Demo	Custom Prototype	Trial Download
<b>Summary</b>	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.	With the complimentary help of experienced Pentaho analysts and data integration specialists, plan and build a custom prototype that illustrates what Pentaho can do with your data.	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 install and administrate, through create your first data models and build reports, analysis, dashboards, and data integration ETL transformations.
<b>Data Source</b>	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
<b>Hardware/Software Requirements</b>	Web Browser	Varies, depending on your requirements.	One computer that meets the server requirements stated <a href="#">here</a> .
<b>Recommendation</b>	<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>All evaluators, particularly any Big Data or Data Integration evaluators.</p> <ul style="list-style-type: none"> <li>Recommended for evaluators who want to explore Business Analytics and Data Integration features 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>Technical support is available to help if you have questions.</li> </ul>

## Use the Hosted Demo

To run the hosted demo, visit the [Pentaho website](#).

## Custom Pentaho Prototype

To use the custom demo, simply contact us by clicking the Your Data, Our Servers link at [Pentaho](#). Once you have contacted us, a representative will guide you through the process.

## Trial Download

To complete this process, you access selected areas of the Pentaho web site and documentation links that provide you with the software, information, and guidance to complete the evaluation. At the end of this process you have a fully configured system that contains all of the Pentaho Business Analytics and Data Integration components.

1. Download the Pentaho Business Analytics Trial Version of the software from the [Pentaho website](#). Select the appropriate file for your operating system. The trial is only valid for 30 days, so you might not want to download until you are ready to fully configure the software and start your evaluation.
2. Use [Install All BA Components](#) and when it comes to selecting components, select them all.
3. The [User Console](#) automatically launches. Verify that your [evaluation licenses](#) are installed. If you need to, you can install them now.

4. Evaluate *Business Analytics*.
5. Evaluate *Data Integration*.

## Next Steps

---

If you went through the Hosted Demo and want to learn more about Pentaho, select the Custom Pentaho Prototype or Trial Download evaluation options to continue. See the Evaluation Installation Decision Table for more details. If you used the Trial Download, remember we are here to help you with your *Business Analytics* and *Data Integration* evaluation. After your evaluation contact your Pentaho sales representative to purchase Pentaho.

# Explore and Learn Pentaho Business Analytics

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.

This section is grouped into three parts that you can work through at your own pace. Each module contains several lessons that contain links to videos, tutorials, user guides, and reference content that are designed to introduce you to the basic capabilities of Pentaho Business Analytics.

- [Tour the User Console and Create Your First Reports](#)
- [Explore and Learn Data Source Basics](#)
- [Learn About Report Designer](#)
- [Learn More About Pentaho Business Analytics](#)

## Tour the User Console and Create Your First Reports

The **User Console** is a fun, easy to use, 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 Business Analytics (BA) Server, maintaining Pentaho licenses, setting up security, managing report scheduling, and tailoring system performance to meet your requirements.

**Table 3: Explore and Learn Pentaho Reporting Capabilities**

Lesson	Watch, Read, and Do This	Benchmarks
<b>Tour the User Console</b>	<ul style="list-style-type: none"> <li>• <a href="#">Pentaho User Console Overview</a></li> <li>• <a href="#">Tour the User Console</a></li> </ul>	<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>
<b>Create Your First Reports and Dashboards</b>	<ul style="list-style-type: none"> <li>• <a href="#">Creating an Interactive Report</a></li> <li>• <a href="#">Creating an Analysis Report</a></li> <li>• <a href="#">Creating a Dashboard</a></li> <li>• <a href="#">Learn About Analysis, Interactive Reports, and Dashboards</a></li> </ul>	<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>
<b>Schedule Your Report</b>	<ul style="list-style-type: none"> <li>• <a href="#">Schedule Reports</a></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 and analysis.

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 4: Explore and Learn Data Sources Basics

Lesson	Watch, Read, and Do This	Benchmarks
<b>Create Your First Data Source</b>	<ul style="list-style-type: none"> <li>• <a href="#">Creating a Data Source</a></li> <li>• <a href="#">Get Started with Data Models</a></li> <li>• <a href="#">Tour the Data Source Wizard</a></li> </ul>	<ul style="list-style-type: none"> <li>• Understand how the BA 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>
<b>Choose Data Source Types</b>	<ul style="list-style-type: none"> <li>• <a href="#">Choose a Data Source Type</a></li> </ul>	<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>
<b>Create Your First CSV Data Source</b>	<ul style="list-style-type: none"> <li>• <a href="#">Create CSV Data Source</a></li> </ul>	<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>
<b>Refine Your Data Source Model</b>	<ul style="list-style-type: none"> <li>• <a href="#">Edit Data Source Models</a></li> </ul>	(Optional) Edit your evaluation data source model using the Data Source Model Editor.

## Learn About Report Designer

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

Objectives are to get familiar with the Report Designer interface, and use it to create and publish a print-quality report - keep business goals/use cases in mind while doing so.

Table 5: Explore and Learn about Report Designer

Lesson	Watch, Read, or Do This	Benchmarks
<b>Explore the Report Designer Interface</b>	<ul style="list-style-type: none"> <li>• <a href="#">Explore Report Designer</a></li> </ul>	<ul style="list-style-type: none"> <li>• Tour the Report Designer interface before you begin building reports.</li> </ul>
<b>Report Designer Workflow Overview</b>	<ul style="list-style-type: none"> <li>• <a href="#">Report Designer Workflow</a></li> </ul>	<ul style="list-style-type: none"> <li>• Look over the work flow concepts for Report Designer.</li> </ul>
<b>Create Your First Report</b>	<ul style="list-style-type: none"> <li>• <a href="#">Create Your First Print-Quality Reports</a></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>
<b>Refine the Look of Your Report</b>	<ul style="list-style-type: none"> <li>• <a href="#">Design Print-Quality Reports</a></li> </ul>	<ul style="list-style-type: none"> <li>• Explore more advanced features of Report Designer, beginning with report elements.</li> </ul>
<b>Add a PDI Data Source</b>	<ul style="list-style-type: none"> <li>• <a href="#">Adding a PDI Data Source</a></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>



## Learn More About Pentaho Business Analytics

---

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

# Explore and Learn Pentaho Data Integration

This evaluation track acquaints you with some of the major features of Pentaho Data Integration (PDI). PDI is a robust extract, transform, and load (ETL) solution that you can use to integrate, discover, and manipulate 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.

As you explore, you will be introduced to the major components, watch videos, work through hands-on examples, and read about the different features. Go at your pace! Feel free to dig into the documentation or to contact us if you have questions.

The PDI evaluation track is divided into three sections.

- [PDI Basics](#) introduces you to PDI concepts and terminology, then guides you through the process of creating your first transformations and jobs.
- [Explore Big Data and Instaview](#) guides you through hands-on tutorials for how PDI interfaces with Hadoop, Cassandra, MongoDB, and MapR. You also learn how to use and configure Instaview, then how to edit transformations and metadata models.
- [Learn About Kitchen, Pan, and Carte](#) provides an overview of the command line tools that you can use to execute jobs and transformations in Spoon.

## PDI Basics

This section familiarizes you with PDI and learn about basic terminology and concepts. Then, you learn how to access and configure Spoon and take a spin through the interface. Once you are familiar with Spoon, you will be guided through the creation of your first transformation and jobs as well as how to use Agile BI to visualize your data. Review the instructions on how to use the Marketplace to install plugins.

Lesson	Watch, Read, or Do This	Benchmarks
What is Pentaho Data Integration?	<ul style="list-style-type: none"> <li>• Watch the <a href="#">Pentaho Data Integration</a> video.</li> <li>• Learn about <a href="#">Terminology and Basic Concepts</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Gain an overview of PDI.</li> <li>• Get acquainted with PDI terminology and concepts.</li> </ul>
Introducing Spoon	<ul style="list-style-type: none"> <li>• Set up your <a href="#">Spoon work environment</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Start and configure Spoon.</li> <li>• Tour the Spoon interfaces.</li> </ul>
Introduction to Transformation and Jobs	<ul style="list-style-type: none"> <li>• Create <a href="#">transformations</a>.</li> <li>• Build and schedule <a href="#">jobs</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Create transformations from flat files; filter records.</li> <li>• Load and retrieve data from relational databases.</li> <li>• Run transformations and schedule jobs.</li> </ul>
Introduction to Agile BI	<ul style="list-style-type: none"> <li>• <a href="#">Use Agile BI to create reports</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Learn how to use Agile BI to visualize your data and to enhance data quality.</li> </ul>
Adding Plugins to Spoon	<ul style="list-style-type: none"> <li>• <a href="#">Add Plugins to Spoon</a></li> </ul>	<ul style="list-style-type: none"> <li>• Learn how to use the marketplace to add plugins to Spoon.</li> </ul>

## Explore Big Data and Instaview

Quickly accelerate the time to Big Data value by using PDI to connect to a variety of data sources, including Hadoop, NoSQL, and Analytical databases. Learn how PDI interfaces with Big Data sources, work through tutorials, and peruse our steps and transformations. Also, check out our template of the week, view a demonstration of how to mine social data, and take a look at Instaview.

Topic	Watch, Read, or Do This	Benchmarks
What is Big Data?	<ul style="list-style-type: none"> <li>• <a href="#">Big Data Video</a>.</li> <li>• Read about <a href="#">Big Data Resources</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Gain an overview of Big Data and PDI.</li> <li>• Check out Cassandra, Splunk, MongoDB, and Hadoop Big Data resources.</li> </ul>
What is Instaview?	<ul style="list-style-type: none"> <li>• Learn about <a href="#">Instaview Video</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Gain an overview of instaview.</li> </ul>
Introduction to DI Big Data Steps and Transformations	<ul style="list-style-type: none"> <li>• Find out what <a href="#">Big Data steps are available</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Review available Big Data transformation steps and job entries.</li> </ul>
PDI, Hadoop, Cassandra, MongoDB, and MapR	<ul style="list-style-type: none"> <li>• Work through <a href="#">Big Data Tutorials</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Explore how to load, transform, extract, and report on data in a Hadoop and MapR clusters.</li> <li>• Learn how to write and read data to and from Cassandra, and how to create reports.</li> <li>• Learn how to create MongoDB reports and how to read and write data to and from MongoDB.</li> </ul>
Using and Configuring Instaview	<ul style="list-style-type: none"> <li>• Learn how to <a href="#">Use and Configure Instaview</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Connecting to data</li> <li>• Creating an analyzer report.</li> <li>• Correcting data quality and adding filters.</li> <li>• Opening, saving, and closing Instaview</li> <li>• Editing Data sources and managing visualizations.</li> </ul>
Beyond the Basics: Editing transformations and Metadata Models	<ul style="list-style-type: none"> <li>• Edit <a href="#">Transformations and Metadata Models</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Edit Data Transformations</li> <li>• Edit Metadata Models</li> </ul>

## Learning About Kitchen, Pan, and Carte

Learn how to use Kitchen and Pan, which are command lines tool for executing jobs and transformations modeled in Spoon. Get familiar with Carte, which is a web server that enables remote execution of jobs and transformations.

Topic	Watch, Read, or Do This	Benchmarks
What are Kitchen, Pan, and Carte?	<ul style="list-style-type: none"> <li>• <a href="#">About Kitchen, Pan, and Carte</a></li> </ul>	<ul style="list-style-type: none"> <li>• Gain an overview of instaview.</li> </ul>
Configure Kitchen, Pan, and Carte	<ul style="list-style-type: none"> <li>• <a href="#">Configure Kitchen, Pan, and Carte</a>.</li> </ul>	<ul style="list-style-type: none"> <li>• Configure Kitchen, Pan, and Carte</li> </ul>

## Next Steps

Now that you have completed an evaluation of PDI, dig a little deeper.

- Review the [PDI User Guide](#) to learn more about features that make PDI distinctive.
- Watch the [predictive analytics overview](#) to learn more about how PDI can be used to operationalize your data.
- Visit our [user communities](#).

- Contact Pentaho Pre-Sales to learn more about how Pentaho can be customized to meet your needs. The flexibility of PDI means that you can explore, process, transform, export, and visualize data in a myriad of ways.