Get Acquainted with Pentaho Components

Help and Support Resources

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

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

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

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

Liability Limits and Warranty Disclaimer

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

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

Trademarks

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

Third-Party Open Source Software

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

Contact Us

Global Headquarters Pentaho Corporation
Citadel International, Suite 340
5950 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

Contents

Introducing Pentaho Business Analytics

The Pentaho Business Analytics (BA) Suite helps you analyze data quickly so you can answer complex business questions. With our integrated toolset, you can create business intelligence solutions, as well as dynamically updating data marts and warehouses. You can connect to any data source, from relational databases to big data sources, XML files, or even simple CSV files. No matter where your data is stored, we make it easy to view analytic results immediately. Pentaho's software components are designed for a wide variety of users ranging from Business Solution Consumers and Business Analysts to Data Scientists.

The Pentaho BA Suite consists of both BA and Data Integration (DI) components. *Business Analytics* (BA) components help you create compelling visualizations, reports, and dashboards, based on well-formed data models that you design. *Data Integration* (DI) components allow you to connect to and extract data from diverse data sources such as relational databases, NoSQL databases, and Hadoop. DI allows provides a visual interface analyzes thyou can use to transform your data to support business analytics.

Business Analytics

Pentaho Business Analytics (BA) comprises a suite of several business intelligence tools, which provide valuable insight into business trends and performance.

- The *User Console* allows you to create business analytics content, display and schedule reports, and administer the BA Server.
- The *BA Server* hosts Pentaho-created and user-created content. It houses the BA Repository which contains audit, scheduling information as well as content.
- *Report Designer* lets you create production-ready, print-quality reports.
- *Data Modeling* tools help you build and refine data models.

These tools are based on the open source *Pentaho BI Platform/Server* project.

Business Analytics Server

The BA Server hosts the User Console, and processes all reporting, analysis, and dashboard content. The BA Server hosts the centralized BA repository for secure sharing of all BA data solutions. It also houses scheduling and audit tables.

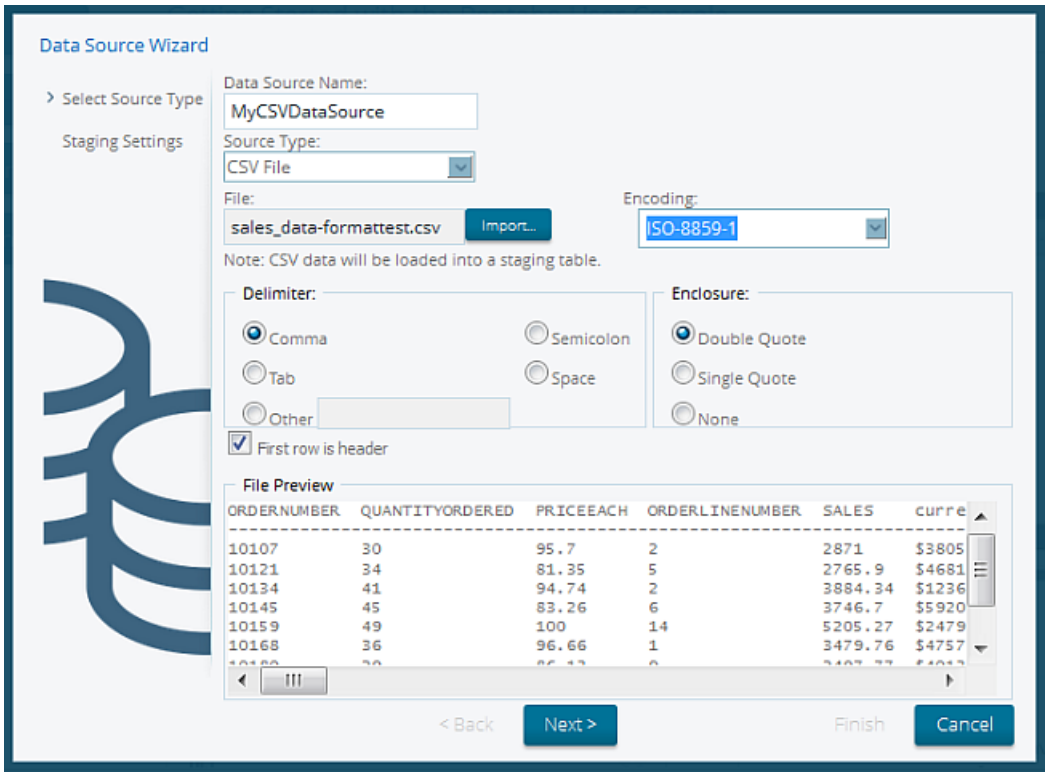
Manage the BA server through its web-based tool, the *User Console*.

User Console

The User Console's web-based tools make it easy for business users to *analyze data*. Create *interactive reports* and build integrated *dashboards* to identify and share business intelligence. Explore data from a wide variety of sources, regardless of your technical expertise. Share your creations in a public folder, or save your work to PDF, Excel, or CSV files.

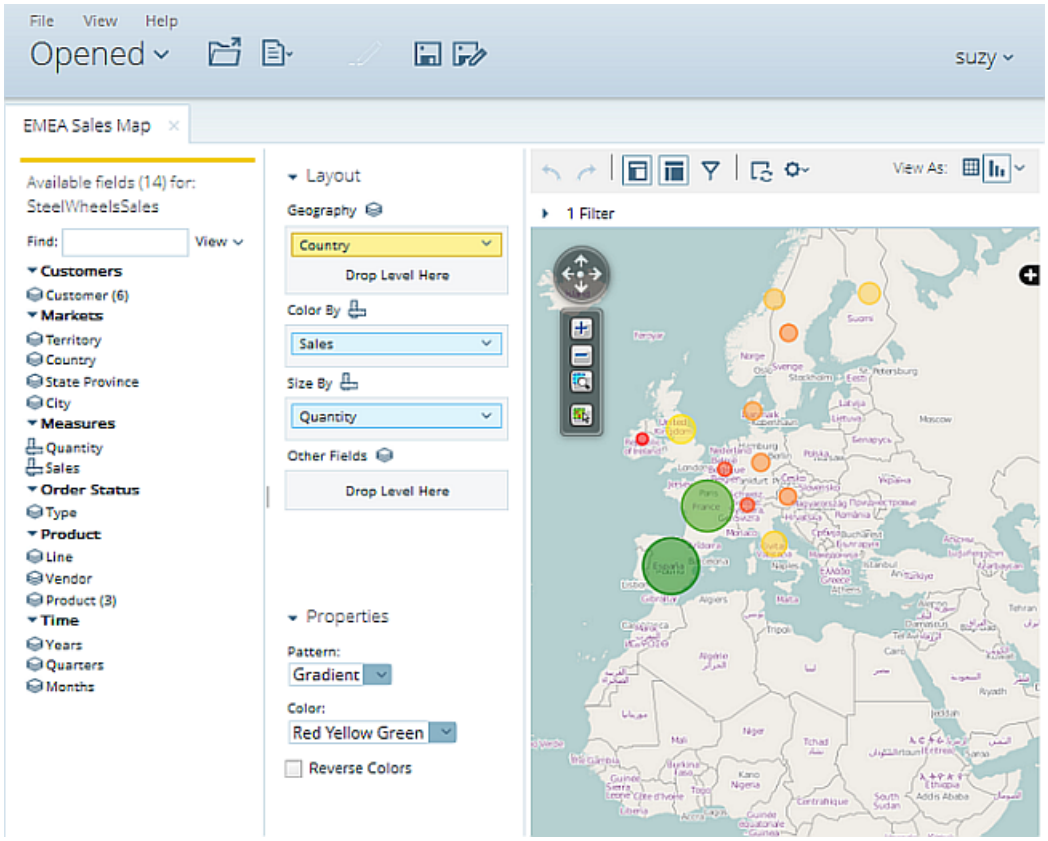
The User Console also allows system administrators to configure and manage the *BA Server*.

Data Source Wizard



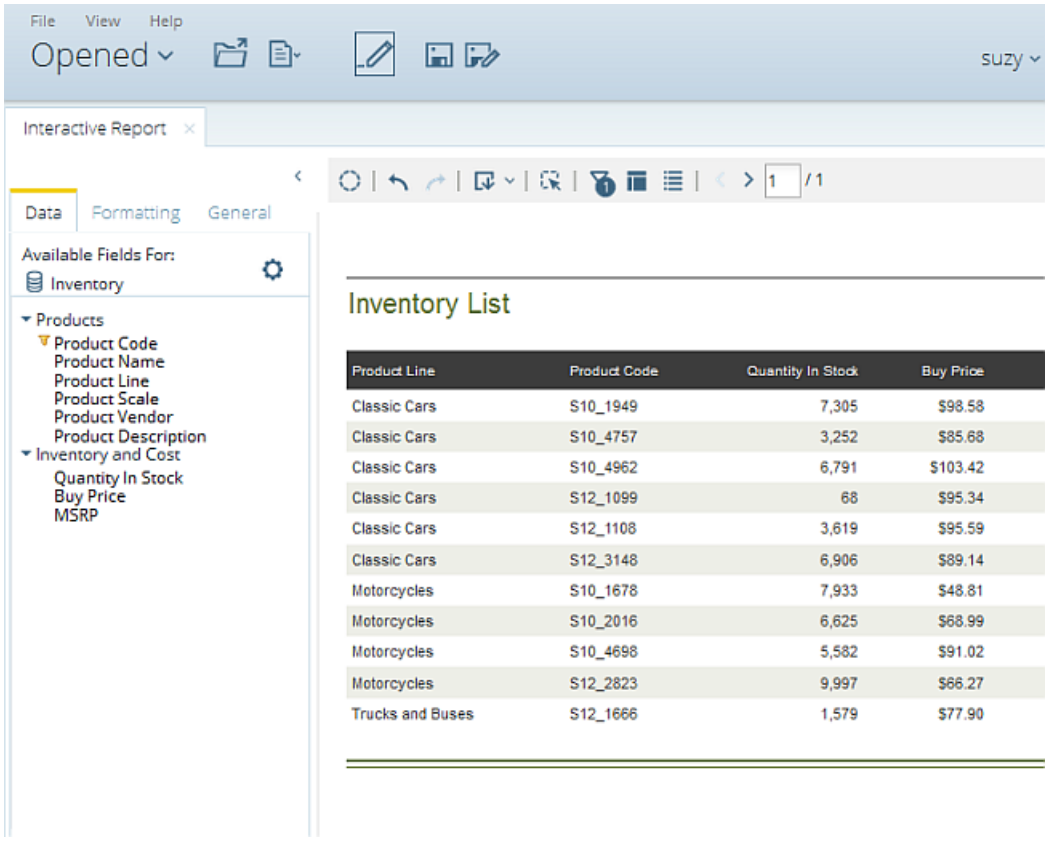
The Data Source Wizard guides you through defining relationships within your data. Use the wizard to get initial results from both simple, relational data models, and more robust, multidimensional data models. Each model type has specific advantages, and as your business needs evolve, so will the models you build.

Analyzer



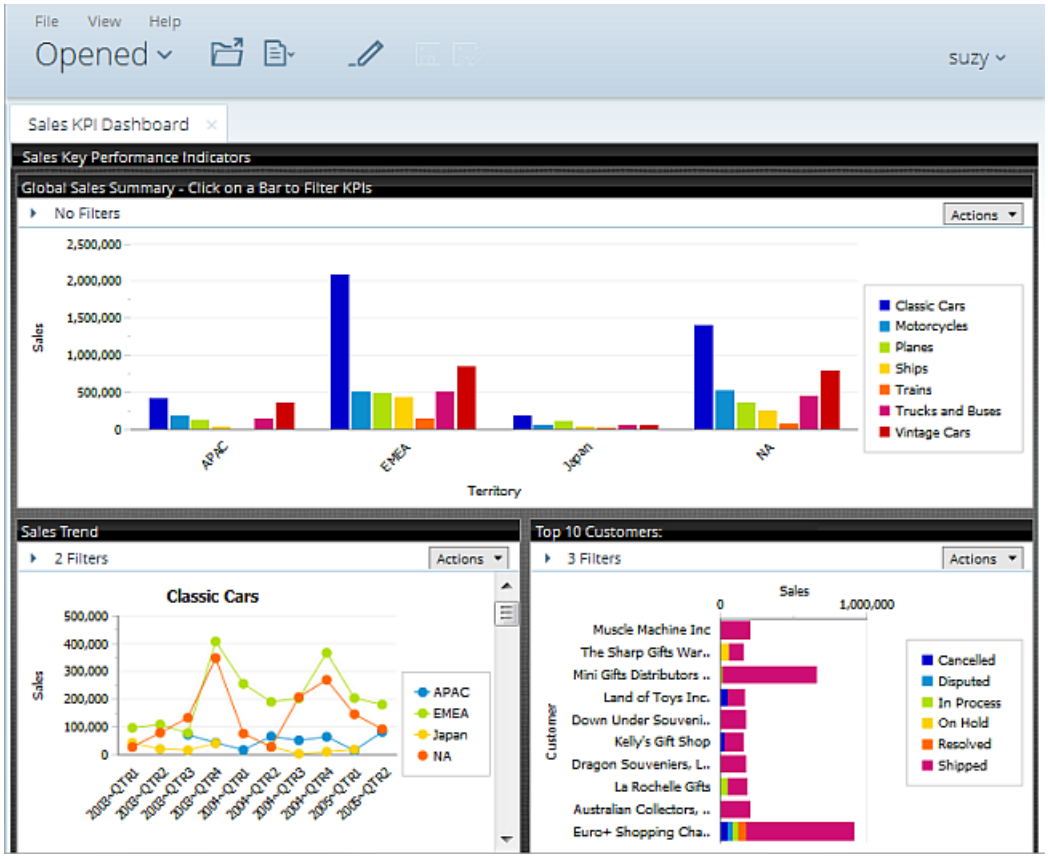
Analyzer helps you filter and visualize data to make informed business decisions. Simply drag items such as product lines and customer attributes onto a canvas, then create calculated fields and hierarchies to refine the results. Analyzer brings your data to life when you add sophisticated visualizations, such as maps, charts, and grids. These visualizations reveal complex patterns and outliers to make your analysis effortless.

Interactive Reports



Create template-based, on-demand reports in an intuitive, drag-and-drop environment with Interactive Reports (IR). Narrow your scope with filters and prompts to focus viewer attention. Generate reports with sorts, groupings, and summaries.

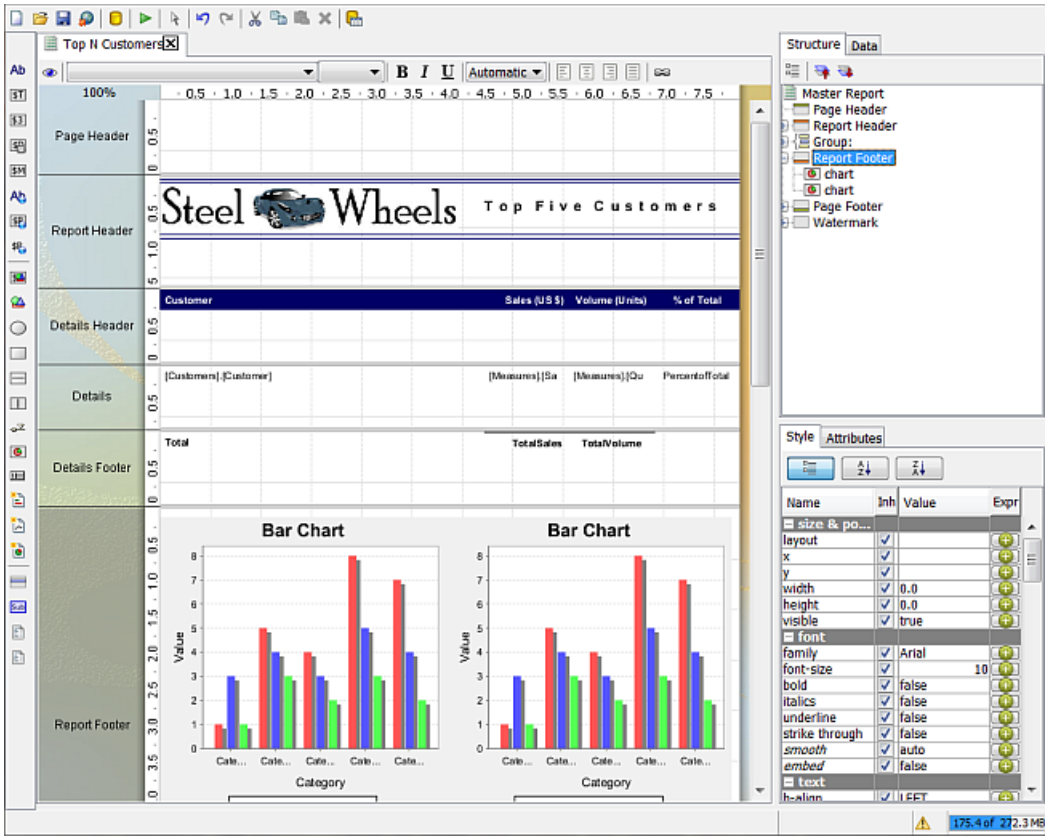
Dashboard Designer



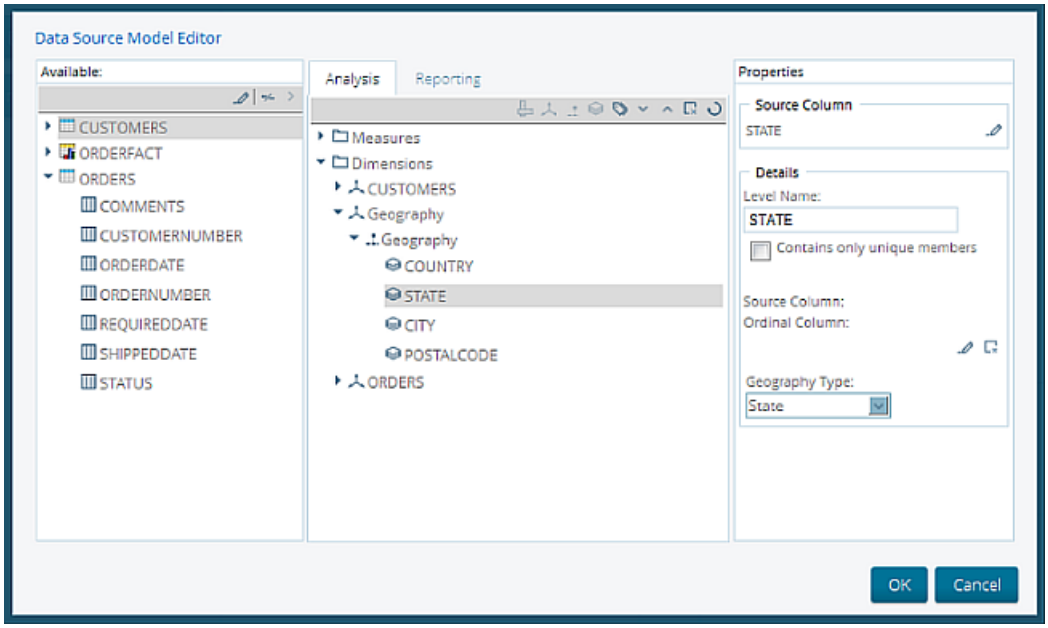
With this designer, transform collections of data into intuitive, visually attractive dashboards that help decision makers recognize critical knowledge at a glance. Embed a wide variety of content in the web-based display, including Interactive Reports, Analyzer visualizations, graphics, charts, web pages, even collaborative content. Mobile and desktop users can personalize dynamic dashboards to reflect their preferences.

Report Designer

Use Report Designer when you want to create highly detailed, print-quality reports based on data from almost any source, including RDBMS, OLAP, big data, and NoSQL data sources. This standalone tool, based on the open source *Pentaho Reporting Project*, lets you create reports, and generate CSV or XLS files, HTML, and PDFs.



Data Modeling Tools



While business users can create data models with the Data Source Wizard, data design professionals may need more powerful options. Use our standalone modeling tools to create very complex data models. Fine-tune your models by adding calculated fields, applying security, and localizing to meet your requirements. The only limiting factor is your imagination.

- Use the *Metadata Editor* to build and refine relational data models, optimized to provide immediate, up-to-date information.

- Use *Schema Workbench* to create and enhance multidimensional data models, with robust features for expanding analysis across time. These data models are based on models in the open source *Pentaho Mondrian Project*.
- Use *Aggregation Designer* to boost system performance for multidimensional data models.

Next Steps

REWRITE: The BA server and tool set are the core foundation of the Pentaho Business Analytics suite. Our Pentaho Data Integration (PDI) tool set also plays an important role by populating data marts with structured data.

Learn More

- Data design professionals *use PDI* side-by-side with business partners to build and enhance business intelligence solutions.

Data Integration

While business analytics deliver detailed information to decision makers throughout your organization, insightful analysis depends on the quality and structure of your data. By assembling an effective, integrated *data mart* or *warehouse*, you can make sure your analytics are grounded in clean, well-formed data. Pentaho's Data Integration (PDI) toolset, also known as the open source *Pentaho Kettle Project*, answers these challenges.

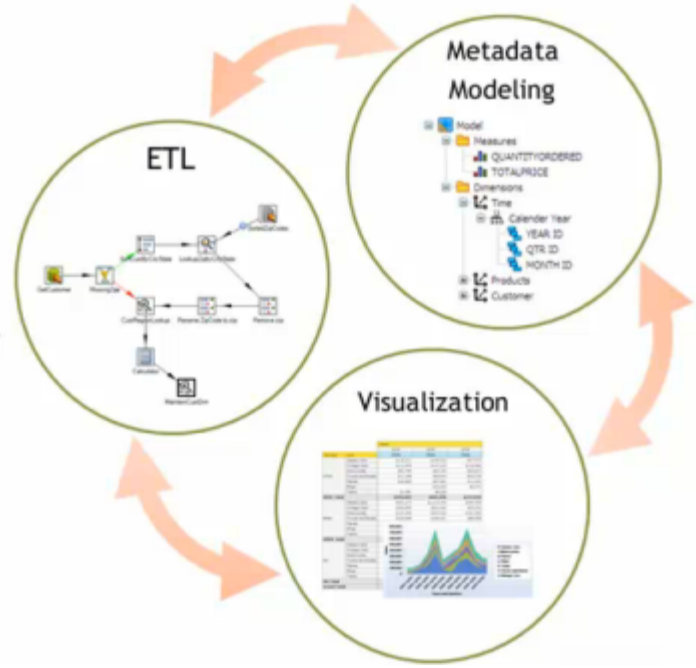
PDI helps you blend data from diverse sources, and create automated *Extract, Transform, Load* (ETL) processes to populate data marts.

- **Extract** data from operational stores, social media streams, flat files, and big data sources such as Hadoop or NoSQL.
- **Transform** your data into structures that fit reporting and analytical needs by integrating data from multiple sources, cleansing to make the data consistent, and supplementing with additional data.
- **Load** data into reporting databases, data marts, data warehouses, big data, or other systems so business users can select and analyze that data.

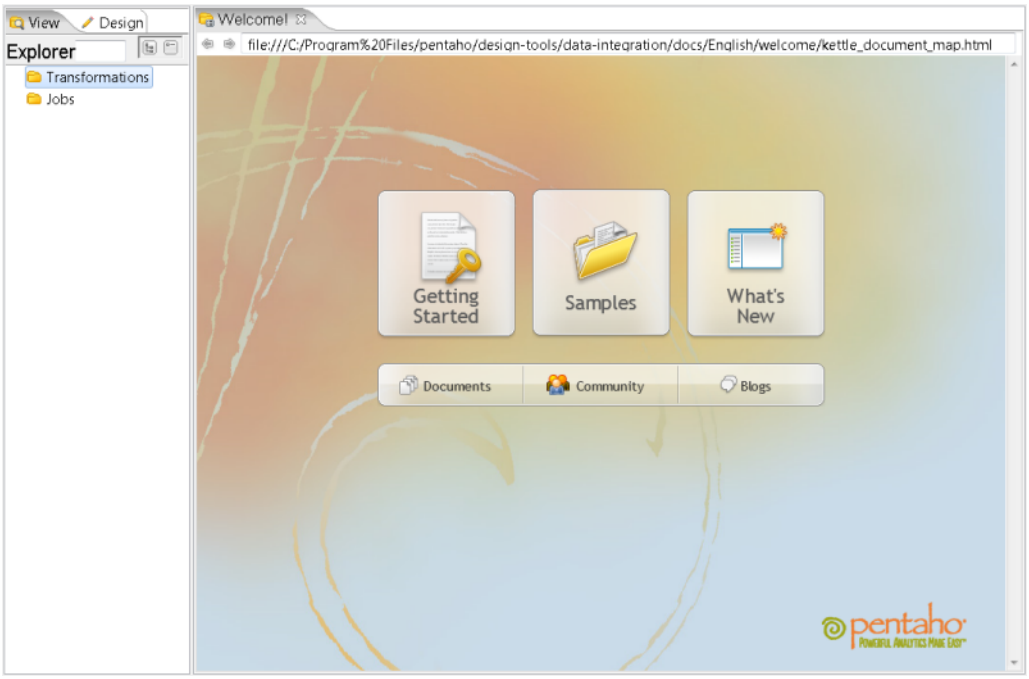
Beyond simply transforming and moving data, PDI provides intelligent orchestration for operationalizing ETL and business processes. Accomplish any type of *Data Integration* (DI), including structuring, modeling, and other ETL tasks.

If you are familiar with the structures in your data, and are comfortable with ETL and data modeling concepts, you can use PDI tools to create a production data mart within a single afternoon.

- The design tool, *Spoon*, provides a graphic environment that helps you build a data mart.
- Use *Instaview*, a template-based multidimensional modeling tool, to visualize your data immediately.
- The *Data Integration (DI) Server* is the core of the PDI system.
- Several *command line utilities* permit power users to work outside of Spoon.
- Process large data sets across clustered computers with *Pentaho MapReduce*.



Spoon



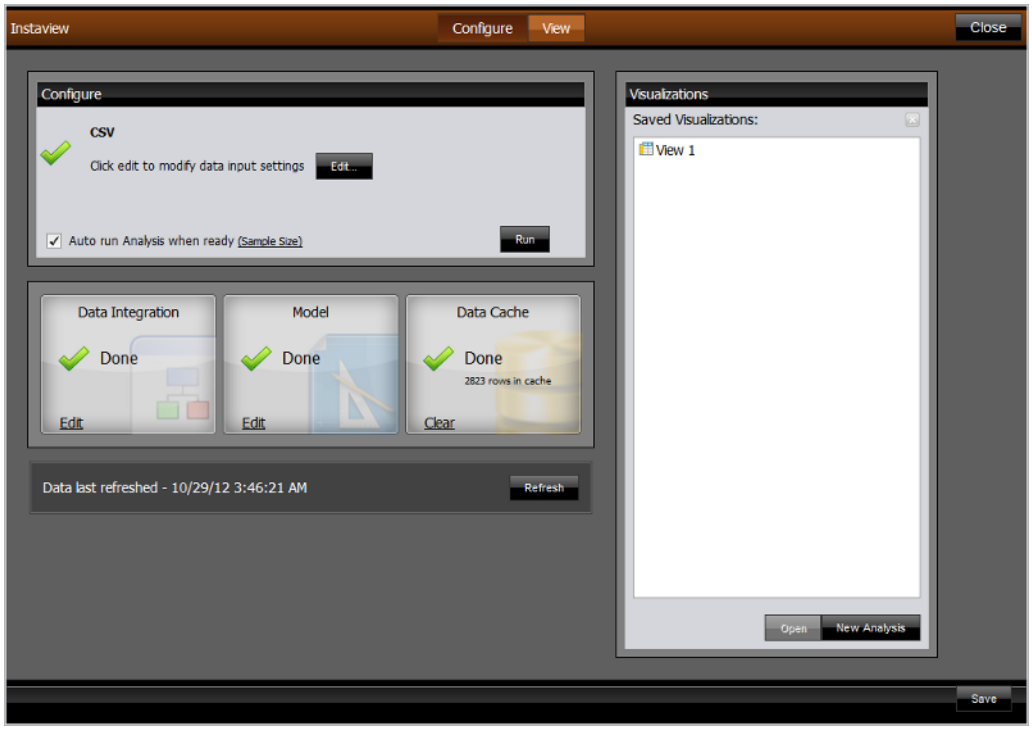
Spoon is a graphical design tool that lets you create, schedule, and run complex ETL and DI processes, without writing or generating code. Drag job entries onto the Spoon canvas, or choose from a rich library of more than 200 pre-built steps to create a series of data integration processing instructions. Refine input and output steps to perform specialized tasks, such as creating connections to data, and integrating with legacy systems.

When you run data integration processes, PDI executes instructions automatically, either on demand or from a schedule.

Agile BI

Agile BI is an accelerated development approach, which links Spoon's data integration to *Analyzer* and its visualizations. You can immediately see analyzed results as you change your data mart or data models, without leaving Spoon. This technique helps data design professionals and business users work together to rapidly resolve business analysis concerns.

Instaview



Sometimes you simply need a quick snapshot of information that is buried in data sets, databases, or other data sources. For example, you could be interested in knowing how often a particular key word appears in customer comments on your website.

Even if you do not understand DI concepts or know how to write code, Instaview lets you automatically extract, transform, and stage data so you can quickly populate a data mart and analyze data. Simply choose from a collection of pre-built templates that provide instant connectivity to the most popular data sources, such as MongoDB, Hive, and HBase. Then, let Instaview generate the data model for you. After it is created, manipulate model elements to organize data to meet your needs. Finally, use Analyzer to create visualizations, such as heat maps and bubble charts, to view results.

Data Integration Server

The DI Server runs centrally stored transformations and jobs. The DI Server also hosts the DI repository and processing engine, provides a service layer for security and authentication, and allows scheduling.

Manage the DI Server through its related tool, *Spoon*.

Command Line Utilities

Although Spoon lets you run PDI transformations and jobs within an intuitive, graphical environment, you can also use PDI by command line.

- Use Pan to execute PDI transformations, which represent a data stream through a set of independent tasks.
- Use Kitchen to orchestrate PDI jobs, which contain transformations and other job entries as part of a larger business process.
- Use Carte to set up dedicated, remote PDI servers, so you can coordinate jobs across a collection of *clustered* computers, and execute transformations within a cluster of Carte cluster nodes.

Pentaho MapReduce

Pull information you need from a Hadoop cluster by creating visual MapReduce applications. Even without programming experience, you can build jobs that use the embeddable PDI engine to extract and manipulate large data sets from distributed file systems.

Next Steps

The *BA and DI Installation Overview* provides all the information you need to get started installing Business Analytics, so you can achieve your business intelligence goals.

Learn More

- Have you completed the Pentaho test drive? Take a few minutes to *Try Pentaho Now*!