Pentaho Data Integration

For the latest information, please visit our web site at www.pentaho.com

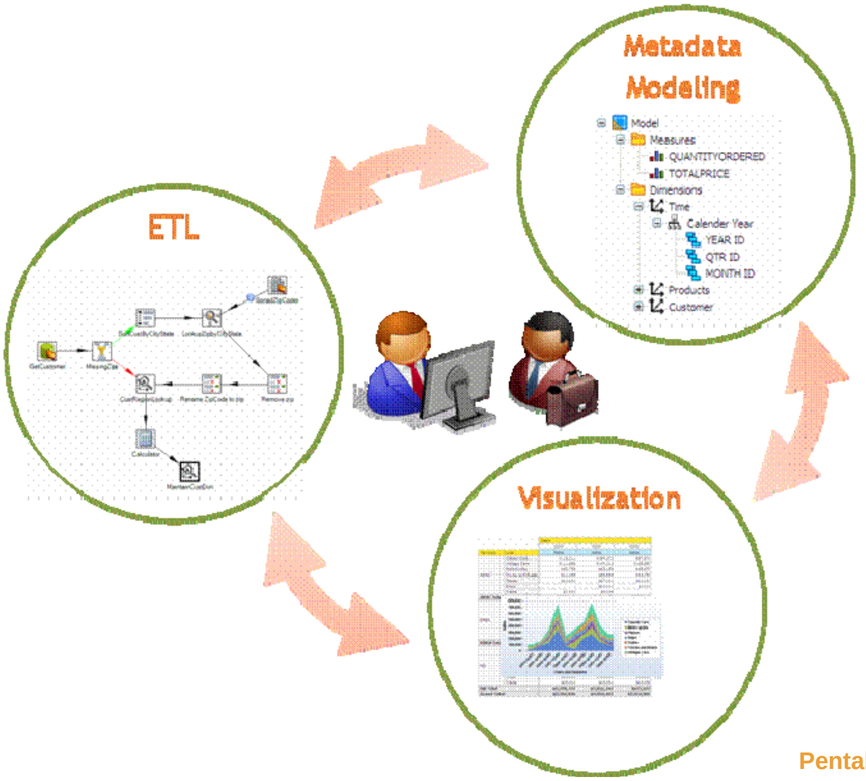
Pentaho Data Integration Overview

Data is everywhere including the ever increasing volumes of “Big Data” being generated in social media, online gaming and other rapidly growing markets. Providing a consistent, single version of the truth across all sources of information is a key challenge faced by IT organizations today. Likewise, the processes and methodologies to deliver data and analytics to users are cumbersome, expensive, and slow. Pentaho Data Integration delivers powerful Extraction, Transformation and Loading (ETL) capabilities using an innovative, metadata-driven approach and unifies the beginning to end BI project development process to increase productivity and dramatically accelerate the delivery of successful BI projects. And, Pentaho’s extensible, standards based architecture ensures that you will never be forced to adopt proprietary methodologies into your ETL solution.

Agile BI

Most data warehousing, data migration and analytics projects are expensive, have long deployment cycles and high risks of failure. Applying the concepts of agile programming and rapid application development to business intelligence applications, Pentaho has pioneered **Agile BI** which redefines the way BI projects are built and deployed. **Agile BI** unifies currently separate ETL, modeling and visualization processes into a single development environment for building business intelligence projects. Pentaho’s Agile BI solution:

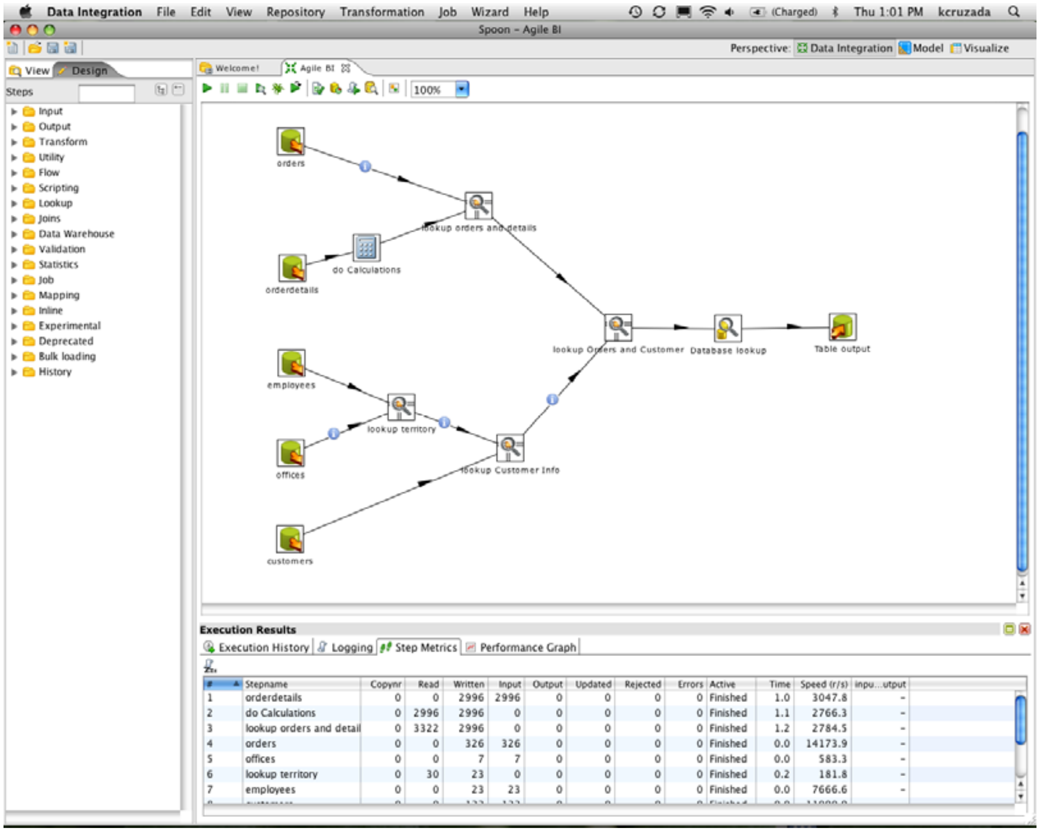
- Powers instantaneous, iterative BI application development
- Enables seamless collaboration between developers and end users
- Merges complex BI development into a single process
- Dramatically reduces time and difficulty of building and deploying BI apps
- Combines the strength of rapid BI application development with robust ETL for complex data integration needs



Ease of Use

Pentaho Data Integration's metadata-driven approach means you simply specify WHAT you want to do, but not HOW you want to do it. Now administrators can create complex transformations and jobs in a graphical, drag-and-drop environment without having to generate any custom code. Pentaho Data Integration is a full-feature ETL solution including:

- Rich transformation library with over 150 out-of-the-box mapping objects
- Comprehensive integration with Hadoop designed to lower the technical barriers to adopting Hadoop for Big Data projects
- Advanced data warehousing support for Slowly Changing and Junk Dimensions
- Enterprise-class performance and scalability
- Metadata injection exposes all properties of the step and enables injection of file names, the removal or renaming of fields and other metadata properties
- ERP connectors and data quality plug-ins also available
- Unified ETL, modeling and visualization development environment



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	inpu...utput
1	orderdetails	0	0	2996	2996	0	0	0	0	Finished	1.0	3047.8	-
2	do Calculations	0	2996	2996	0	0	0	0	0	Finished	1.1	2766.3	-
3	lookup orders and detail	0	3322	2996	0	0	0	0	0	Finished	1.2	2784.5	-
4	orders	0	0	326	326	0	0	0	0	Finished	0.0	14173.9	-
5	offices	0	0	7	7	0	0	0	0	Finished	0.0	583.3	-
6	lookup territory	0	30	23	0	0	0	0	0	Finished	0.2	181.8	-
7	employees	0	0	23	23	0	0	0	0	Finished	0.0	7666.6	-

Graphical, model-driven approach to ETL

Pentaho Data Integration Enterprise Edition

Pentaho Data Integration Enterprise Edition extends Pentaho's best-in-class open source business intelligence (BI) capabilities with additional software and services designed to help you and your organization:

- **Achieve BI success**
- **Save time, resources, and money**
- **Mitigate risk**

Achieve BI success

What makes the difference between success and failure in business intelligence or data warehousing projects? There is ample evidence from IT professionals, consultants, and industry analysts that success or failure with business intelligence is often driven far more by "people and process" issues rather than technology. Poor planning, lack of commitment, inadequate resources or skill sets, and inability to deliver initial results quickly can doom a BI project regardless of the selected software products and technology. While open source software is rapidly transforming the IT landscape and has provided new levels of flexibility and freedom for customers, open source software alone does not address the traditional pitfalls of BI projects. Pentaho Data Integration Enterprise Edition and Pentaho's **Agile BI** approach provides the product capabilities and value-added services to help you deliver a successful BI project for your organization, including consultative support and product expertise, software maintenance, management and monitoring tools, and more.

Save Time, Resources, and Money

Even large organizations have fewer IT resources than they would like, and they strive to get the most out of their investments in time, people, and technology. There are numerous public examples of Pentaho customers who have realized the Total Cost of Ownership (TCO) advantages of commercial open source BI from Pentaho, recognizing that investing in a relationship with Pentaho saves time, resources, and money not just in the long-term, but in the *short term* as they initiate BI projects. "Going it alone" with free BI software not only increases your risk of failure, it turns out to be more expensive. Pentaho Data Integration Enterprise Edition delivers critical benefits like stabilized software, enhanced deployment capabilities, direct access to product expertise, and committed response times to help you save time, resources, and money.

Mitigate Risk

Business intelligence risk comes in many shapes and forms. Risk of project failure, risk of late delivery, risk of going over budget, and legal risk as well. Beyond providing the software enhancements and services to reduce project risk, Pentaho provides a lower-cost model for enterprise-class business intelligence software that reduces budget risk by eliminating large, up-front software license fees. Pentaho Data Integration Enterprise Edition also includes legal protection to minimize your company's risk and exposure to potential legal issues related to intellectual property in open source software.

Promote Collaboration Between IT and End Users

With the introduction of Pentaho's **Agile BI** approach in unifying the ETL, modeling and visualization processes into a single environment, Pentaho Data Integration drives closer collaboration and cooperation between IT staff and business users. By being able to work closely together in real-time to accurately define and refine just what analytics are needed, Pentaho Data Integration results in faster delivery of higher quality business intelligence projects.

Pentaho Data Integration for Hadoop

Pentaho Data Integration Enterprise Edition 4.1 delivers comprehensive integration with Hadoop designed to lower the technical barriers to adopting Hadoop for Big Data projects. Using Pentaho Data Integration's easy-to-use, graphical design environment, ETL designers and data architects can now harness the power of Hadoop with zero programming to address common data integration use cases including:

- Moving data files into and out of the Hadoop Distributed File System (HDFS)
- Input/Output data to and from Hadoop using standard SQL statements
- Coordination and execution of Hadoop tasks as part of larger data integration and BIworkflows
- Graphical design of new MapReduce jobs taking advantage of vast library of pre-built mapping and data transformation steps

Pentaho Data Integration Enterprise Edition supports the latest releases of Apache Hadoop as well as popular commercial distributions such as Cloudera Distribution for Hadoop and Amazon Elastic MapReduce.

Pentaho Data Integration Enterprise Edition Features

Pentaho Data Integration Enterprise Edition allows you to deploy the best-in-class capabilities of Pentaho Data Integration in production with confidence, security, and far lower total cost of ownership than proprietary alternatives. Pentaho Data Integration Enterprise Edition provides additional capabilities including professional support, software maintenance, enhanced software functionality, certified software, product expertise, and the best software assurance program in the industry.

Software and Services	Community Edition	Enterprise Edition
Data Integration / ETL	Open Source	Certified
Business Intelligence Platform	Open Source	Certified
Community Forums Interaction	✓	✓
Agile BI Unified Development Environment	✓	✓
Community Web Documentation (wiki)	✓	✓
Professional Support		
• Telephone support (toll-free)		✓
• E-mail support		✓
• Service Level Agreement		✓
• Unlimited support cases		✓
Software Maintenance		
• Software maintenance	By in-house staff	✓ By Pentaho Engineers
• Patch releases		✓
• Fixes included in future releases		✓
Enhanced Functionality		
• Pentaho Data Integration Enterprise Console		✓
• Performance monitoring		✓
• Remote administration		✓
• Alerting		✓
• Enterprise Security		✓
• Integrate with 3rd party LDAP/MSAD		✓
• End user/role administration controlling users/roles actions		✓
• Content permissions security (owner, create, read, update, delete)		✓
• Enhanced repository security		✓
• Team repository versioning and locking		✓
• Enhanced content management		✓
• Remote job scheduling and management		✓
• Pentaho Data Integration for Hadoop		✓ (Add-on)
Certified Software		
• Stabilized software		✓
• Managed release cycle		✓
• Optimized builds		✓
Product Expertise		
• Professional documentation		✓
• Knowledge base		✓
• Consultative support		✓
• Remote assistance packages		✓ Optional Add-On

• Installation/configuration packages		✓
• Design and integration packages		✓
• Troubleshooting and optimization packages		✓
• Enterprise Edition online forum		✓
• Web based training		✓ Optional Add-On

For more information on the features and benefits of Pentaho's Enterprise Editions, please see the Pentaho BI Suite Enterprise Edition brochure.

Pentaho Data Integration Feature Details

Unified ETL, Modeling and Visualization Environment

The cornerstone of Pentaho's **Agile BI** approach, the new unified BI development environment collapses the traditional cumbersome and step intensive BI development process. Pentaho Data Integration enables teams of technical and business users to work interactively together in a single environment to seamlessly design, automatically model and then instantly visualize the results of that design in a single environment. This approach enables the deployment of iterative and fully vetted data integration and business intelligence applications in a fraction of the time of other solutions.

Pentaho Data Integration includes not only ETL functionality but also a desktop version of Pentaho's modeling, visualization and reporting capabilities. Together with the ETL functionality, these capabilities enable the closed-loop design capabilities at the heart of Agile BI. Once the BI application design is complete, the full functionality of Pentaho's Analyzer and Reporting solutions are used together with Pentaho Data Integration to deploy the applications to end users.

Modern, Standards-based Architecture

Pentaho Data Integration's open, standards-based architecture is a natural fit for any environment or BI solution. Major benefits of the architecture include:

- 100% Java with broad, cross-platform support
- Complete separation of user interface, data, and metadata
- Fully integrated with the Pentaho BI Suite providing advanced scheduling, security, reporting, and analysis

Enterprise-Class ETL

- Broad out-of-the-box data source support including packaged applications, over 30 open source and proprietary database platforms, flat files, Excel documents, Hadoop and more
- Extensible architecture makes custom plug-in and connector development a breeze
- Repository-based providing easy re-use of transformation components, multi-developer and team collaboration including versioning and locking, and structured management of models, connections, logs, and more
- Enterprise class performance and scalability with support for massively parallel processing (MPP) through clustered execution of transformations
- Fully integrated with the Pentaho BI Suite providing advanced scheduling, security, reporting, and analysis
- Integrated debugger to streamline troubleshooting of data integration processes

- Data Integration Enterprise Console allowing administrators to analyze job performance trends over time, to stop, pause, and restart live jobs, and set execution thresholds

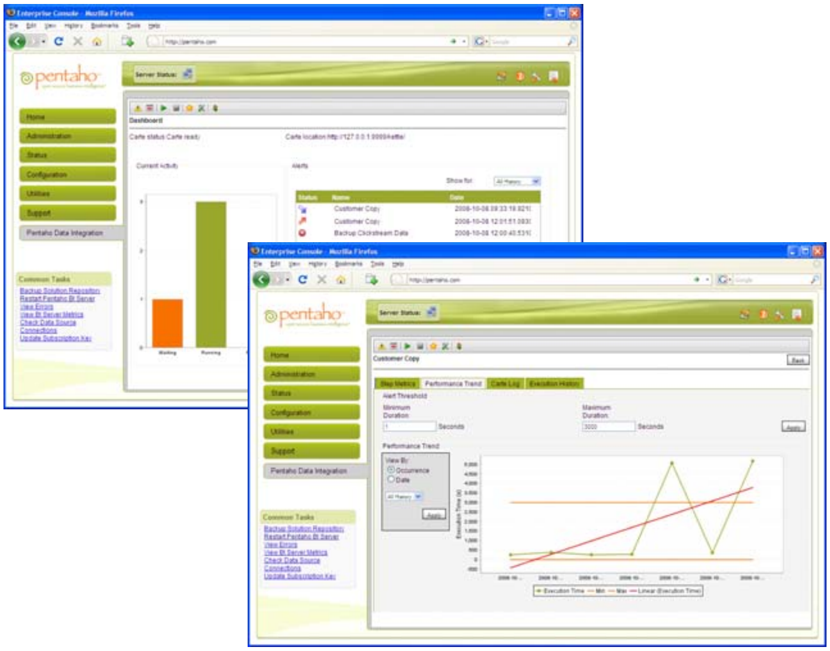
Common Use Cases

- Data warehouse population with built-in support for slowly changing dimensions, junk dimensions and much, much more.
- Export of database(s) to text-file(s) or other databases
- Import of data into databases, ranging from text-files to excel sheets
- Data migration between database applications
- Big Data analytics including support for hybrid data model of integrated Hadoop and traditional, structured data
- Exploration of data in existing databases (tables, views, etc.)
- Instantaneous prototyping of BI applications
- Information enrichment by looking up data in various information stores
- Data cleansing by applying complex conditions in data transformations
- Application integration

Pentaho Data Integration Enterprise Console

Pentaho Data Integration provides a scalable, parallel processing architecture to provide robust ETL performance even when working with very large data volumes. Many organizations deploy Pentaho Data Integration across multiple physical servers. The Pentaho Data Integration Enterprise Console provides a centralized, thin-client administration environment for managing and monitoring enterprise extract, transform, and load (ETL) deployments.

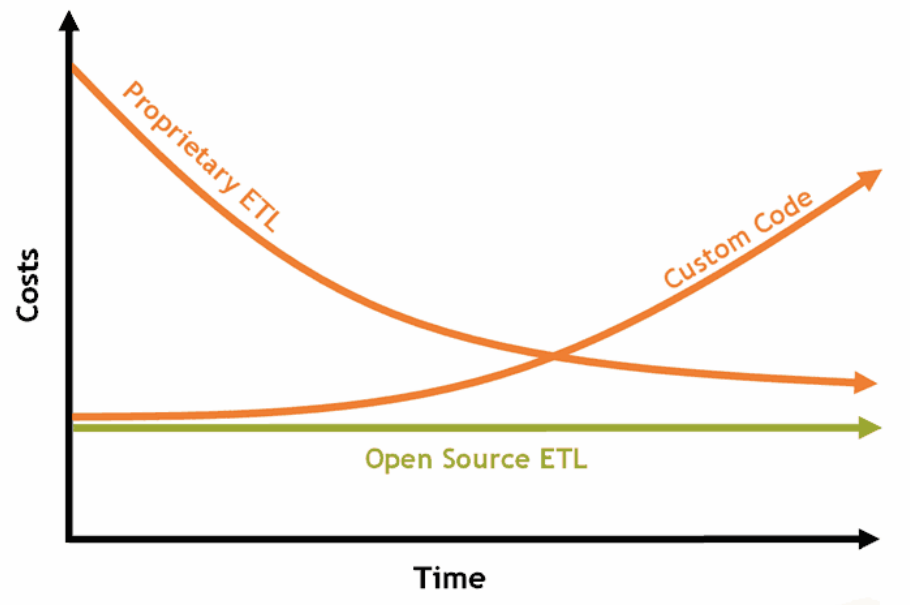
Pentaho Data Integration Enterprise Console allows administrators to monitor performance trends for ETL jobs to identify potential performance degradation issues before they impact batch data processing windows. Administrators can also set minimum and maximum thresholds for jobs, and even pause, stop, and restart live ETL jobs to direct maximum processing power to the most resource-intensive integration jobs.



Pentaho Data Integration Enterprise Console allows administrators to monitor ETL job performance trends and stop, pause, and restart live data integration jobs.

The End of 'Build vs. Buy'

One of the most difficult decisions in any data warehousing project is whether to populate your data warehouse manually using custom code or choose a proprietary ETL tool like Informatica or Oracle Warehouse Builder.



The 'build' solution is appealing in that there are no up front costs associated with software licensing and you can build the solution to your exact specifications. However, businesses today are in a constant state of change and the ongoing costs to maintain a custom solution often negate the initial savings. Proprietary ETL offerings will get your project off the ground faster and provide dramatic savings in maintenance costs over time, but often carry a six figure price tag just to get started. Pentaho Data Integration delivers the best of both worlds with no up front license costs and a significant reduction in TCO compared to custom built solutions. An annual subscription providing professional support, enhanced functionality, certified software, software maintenance and software assurance is also available at a fraction of the cost of proprietary offerings.

Customer Examples

swissport

"We needed fully functional reporting and data integration tools but wanted to see if an open source alternative would give us a similar experience to Oracle. After looking at what was out there, Pentaho had the complete tool set, and after further testing, our users noticed no difference in the features they need."

Swissport uses Pentaho BI Suite Enterprise Edition to automate the integration of their financial, flight and cargo data and deliver operational and strategic reports, metrics and analytics to top executives. A replacement for Oracle BI, Pentaho provided a full BI suite at a lower cost of ownership and helps Swissport answer operational questions such as "how much traffic to expect during holidays" and "is more de-icing staff required at particular airports.”.

"Pentaho provides the same or better functionality and more support versus the competition. They provide a lower total cost of ownership versus proprietary systems. We've been very happy with Pentaho Data Integration Enterprise Edition so far."

Cardiac Sciences uses Pentaho Data Integration Enterprise Edition to source data from their ERP system and provide timely and accurate information for marketing, sales and finance. Selected over Informatica and Microsoft SQL Server Integration Services, Pentaho Data Integration helps Cardiac Science overcome the heavy reliance on IT to generate reports and facilitates the nightly archival of their complex and growing production database.

“We selected Pentaho for its ease-of-use. Pentaho addressed many of our requirements -- from reporting and analysis to dashboards, OLAP and ETL, and offered our business users the Excel-based access that they wanted.”

MySQL uses Pentaho Data Integration Enterprise Edition to integrate data sources from across the organization including Cost Center Rollups stored in Microsoft Excel. This unified data source is used for reporting and analysis of operational expenses by department and cost center using the Pentaho BI Suite.

"With professional support and world-class ETL from Pentaho, we've been able to simplify our IT environment and lower our costs. We were also surprised at how much faster Pentaho Data Integration was than our prior solution."

ZipRealty (NASD: ZIPR) uses Pentaho Data Integration Enterprise Edition to integrate data from multiple sources. They replaced a home-grown ETL application with Pentaho Data Integration, allowing them to respond more quickly to business needs and infrastructure needs and dramatically reduce their maintenance costs.

"Pentaho provides a great solution for us, addressed our technical and business requirements, was quick to deploy, and provided far better value than other alternatives."

Unionfidi chose Pentaho Data Integration Enterprise Edition to build and maintain their data warehouse which supports unified dashboarding, reporting, and analysis to users across the enterprise.

"The simplicity of the interface actually allows Lifetime Entertainment Services to give direct access to business analysts, allowing them to understand and manage the business rules governing the integration of information. That wasn't previously possible with complex hand-coded integration jobs."

Lifetime Networks uses Pentaho Data Integration Enterprise Edition to integrate data from a variety of packaged and custom applications as well as market and viewership information from Nielsen Media Research to provide a centralized view of information to support business intelligence. They chose Pentaho Data Integration Enterprise Edition to replace a homegrown data integration application after a four-week evaluation that included traditional proprietary ETL tools and other open source alternatives.