



What's New in Pentaho Data Integration Enterprise Edition 4.0

Copyright © 2010 Pentaho Corporation. Redistribution permitted. All trademarks are the property of their respective owners.

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

Last Modified on March 29, 2010

Contents

Purpose of This Document	3
Pentaho Data Integration Enterprise Edition 4.0	3
Data Modeling and Visualization Perspectives	3
Pentaho Agile BI	3
Integrated Design Environment for Business Intelligence	4
Data Integration Perspective	4
Visualization Perspective	4
Modeling Perspective	5
Enterprise Edition Data Integration Server	5
Enterprise Repository and Content Management	5
Versioning	5
Locking	6
Enterprise Security	6
Integrated Scheduling	6
Usability Improvements	7
Enhanced Logging	7
New Transformation Steps	7
New Job Entries	8

Purpose of This Document

This document is intended for users who have a working familiarity with the capabilities of Pentaho Data Integration. This document is focused on introducing new capabilities delivered in Pentaho Data Integration 4.0. It is not intended to be a complete review of Pentaho Data Integration's functional capabilities.

Pentaho Data Integration Enterprise Edition 4.0

Pentaho Data Integration (PDI) Enterprise Edition 4.0 brings significant advances to the world's most popular data integration solution. This release includes new features designed to support collaborative team development, simplify ETL troubleshooting, and dramatically reduce the overall time it takes to deploy successful business intelligence applications.

Data Modeling and Visualization Perspectives

Pentaho Data Integration 4.0 introduces the concept of "perspectives" to Spoon, the graphical design interface. Perspectives enable entirely new functional capabilities to be integrated seamlessly, via plug-ins, into Spoon's design environment. With version 4.0, new Modeling and Visualization perspectives deliver the world's first unified ETL, modeling and data visualization development environment as part of Pentaho's Agile BI initiative.

Pentaho Agile BI

Historically, business intelligence (BI) projects have been fraught with out-dated development methodologies and software-related hurdles. Expensive per-user license fees, complex architectures and disparate toolsets with little-to-no integration have separated those responsible for implementing BI from those who will actually use it. For these reasons, BI projects typically take too long to deliver, run way over budget, and often do not meet end user needs. The end result is often failure.

Pentaho Agile BI, with the latest release of Pentaho Data Integration 4.0, redefines the way people build and deploy BI applications. Agile BI and PDI 4.0 present a radically different, fully-integrated development environment that collapses all of the steps from data integration and data modeling through end user visualization, so developers and business users can work side-by-side to build and deploy BI applications in minutes or hours rather than weeks or months. Combined with Pentaho's commercial open source model, Pentaho Agile BI breaks down the technical, financial and software accessibility barriers that have plagued BI projects for decades.



Pentaho Agile BI provides the fastest way to business intelligence success.

Integrated Design Environment for Business Intelligence

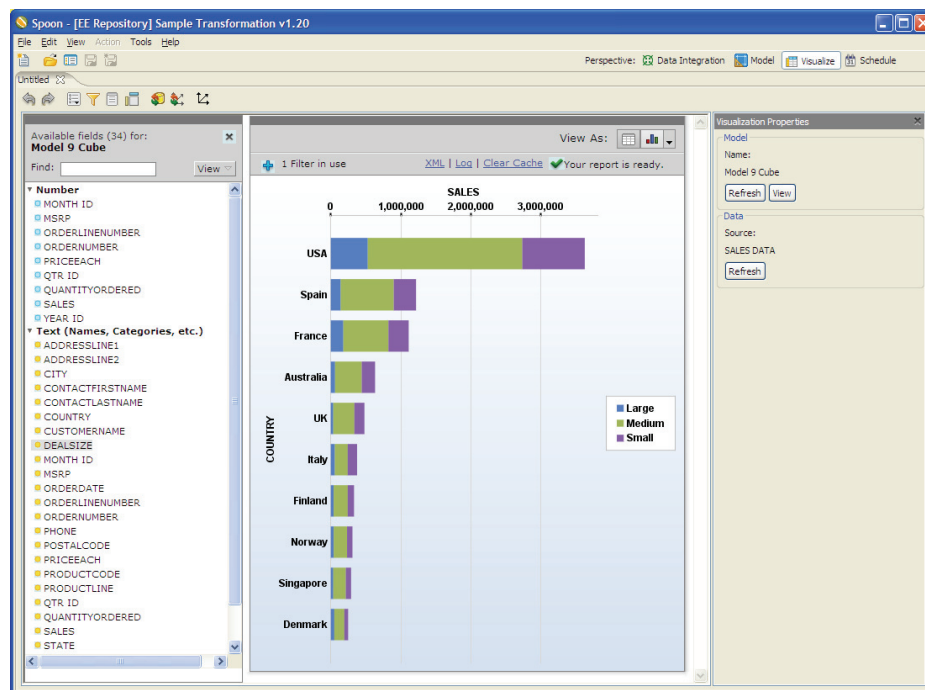
In support of Agile BI, Pentaho Data Integration's Spoon now provides an integrated design environment for performing all tasks related to building a BI solution including ETL, reporting and OLAP metadata modeling and end user visualization. In a few simple clicks, business users can instantly interact with data, building reports with zero knowledge of SQL or MDX, and work hand in hand with solution architects to refine the solution.

Data Integration Perspective

With the introduction of "perspectives" in Spoon, all of the powerful ETL capabilities now exist in a "Data Integration" perspective on the toolbar. This is the default perspective in Pentaho Data Integration and the starting point for graphically building data integration transformations and jobs to support a business intelligence project.

Visualization Perspective

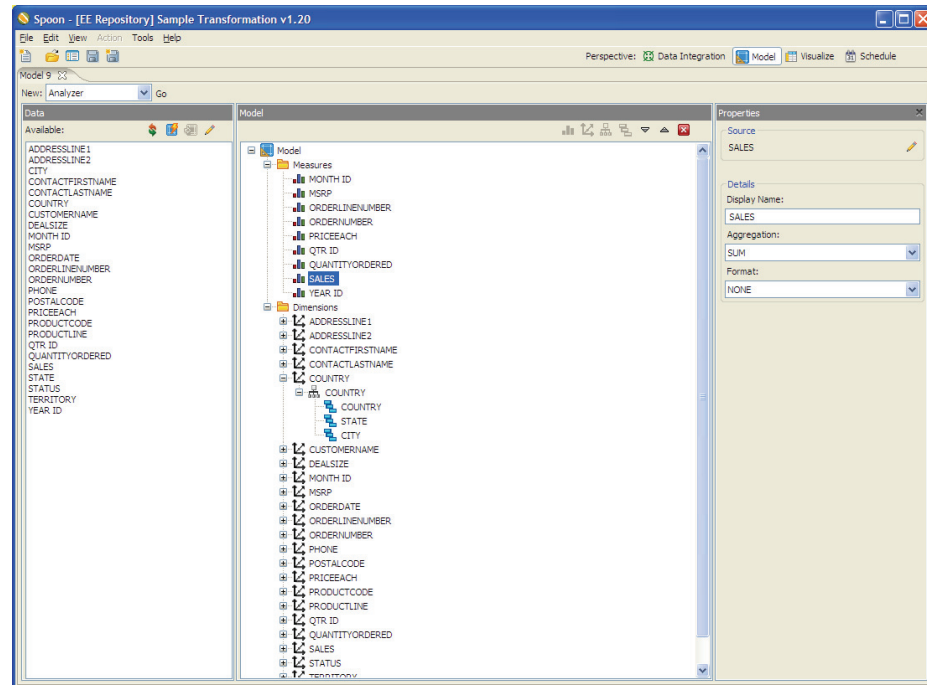
Once a data integration transformation has been created, users can now right-click on the target output step and instantly visualize the results in the form of ad hoc analytical views or production style reports. Under the covers, Pentaho Data Integration automatically generates the OLAP model and the reporting metadata that allows business users to begin visualizing and interacting with the data. This interaction takes place within the new Visualization perspective in Spoon. If a data quality problem is spotted or the business user requests a change in how the data is categorized or presented, users can quickly toggle back to the data integration or modeling perspectives to make updates in real-time. This level of collaboration and iteration is core to agile BI application development and helps ensure projects stay on track with the original business goals.



Integrated data visualization lets ETL developers and business users collaborate in real time.

Modeling Perspective

A new Modeling perspective in Pentaho Data Integration 4.0 lets users easily create or modify the underlying OLAP and reporting metadata models that support Pentaho's data visualization capabilities. Data modeling is a critical component to making information consumable by non-technical users, yet it has historically been one of the toughest and most time-consuming challenges to BI application development. By automating the initial metadata creation and integrating modeling in a unified data integration and visualization environment, Pentaho Data Integration 4.0 dramatically reduces the time it takes to deploy a scalable and user-friendly BI application.



Pentaho Data Integration 4.0 automatically creates the metadata models for reporting and ad hoc analytics.

Enterprise Edition Data Integration Server

Pentaho Data Integration Enterprise Edition 4.0 provides a new dedicated data integration server designed to simplify production deployments with enhanced security, content management capabilities, integrated scheduling and ETL job and transformation execution.

Enterprise Repository and Content Management

The new data integration server provides the ability to centrally store and manage ETL jobs and transformations via an enterprise repository. The enterprise repository includes critical capabilities for team environments, such as version control, content locking and integration with existing security mechanisms such as LDAP or vendor solutions like Microsoft Active Directory (MSAD).

Versioning

Built in content versioning allows ETL developers to keep full revision history of transformations and jobs through all stages of the development lifecycle. At any point, users have the ability to compare changes

made between different versions of a transformation or job and quickly restore previous versions if any rollbacks are necessary.

Locking

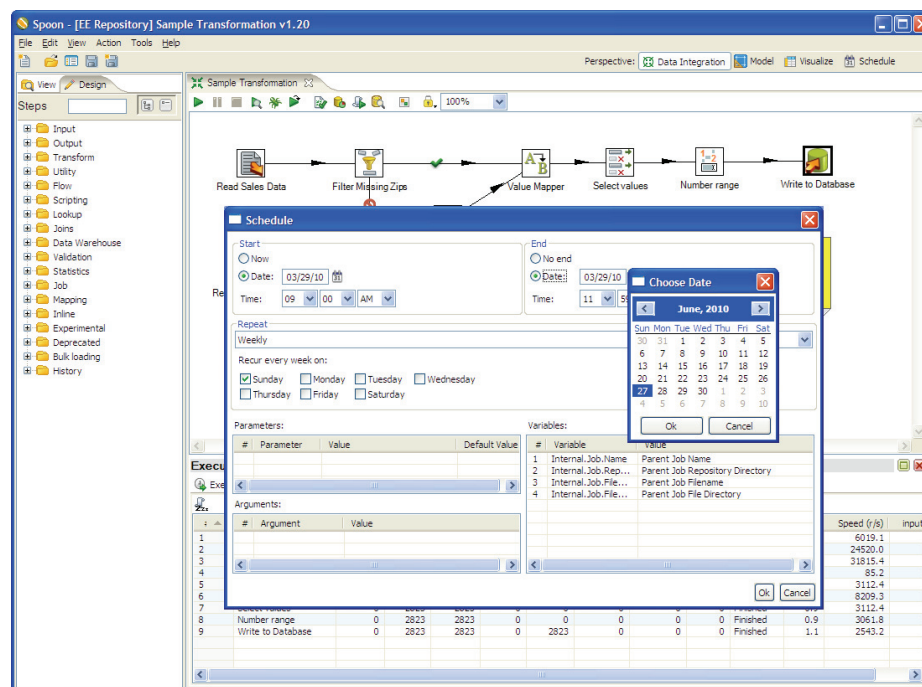
A critical feature for working in team environments, Pentaho Data Integration Enterprise Edition 4.0 provides content locking capabilities for ETL developers to lock a job or transformation from editing by other users. For collaborative projects where multiple developers are working on a common project, users can choose to lock a transformation or job while changes are made and tested.

Enterprise Security

New enterprise security lets organizations easily manage users, roles and permissions to control access to specific pieces of content in the repository. Pentaho security integration is provided by default, and organizations can also choose to leverage existing security providers such as LDAP or Microsoft Active Directory.

Integrated Scheduling

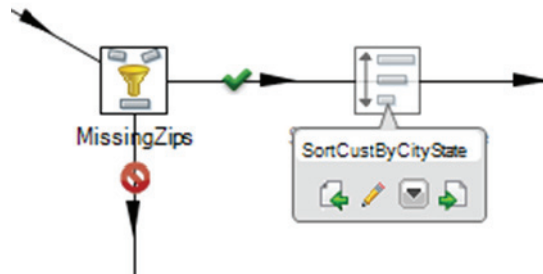
The data integration server in Pentaho Data Integration Enterprise Edition 4.0 provides services to schedule transformations and jobs directly from within the Spoon design environment. Much like scheduling a meeting with Microsoft Outlook, an intuitive interface provides the ability to define the start and stop dates, as well as the recurrence pattern for when the schedule will repeat (daily, weekly, monthly, etc.). With the addition of a Scheduler perspective in Spoon, users can monitor and coordinate scheduled activities.



Integrated scheduling makes it easy to coordinate the execution of jobs and transformations.

Usability Improvements

Many enhancements to the general design environment of Pentaho Data Integration have been incorporated in version 4.0. General improvements include hover-over menus and contextually aware hops that simplify the connection of steps as well as graphical indicators on hops representing the flow of information between steps. Likewise, new activity indicators on jobs help highlight current activity during execution.










Usability improvements in Pentaho Data Integration 4.0 simplify the creation of jobs and transformations.









Enhanced Logging

The logging architecture in Pentaho Data Integration 4.0 has been enhanced to provide robust control of transformation logs and job logs during execution. Enhanced highlighting on error and warning conditions simplify troubleshooting while the ability to drill from job logs into underlying transformation logs lets users navigate directly to the root cause. Additional enhancements include the ability to “sniff” input, output and error records during execution and the overall performance of the logging architecture has been significantly improved.

New Transformation Steps



In addition to the features listed above, Pentaho Data Integration 4.0 adds the following new steps for data integration transformations:

Icon	Step Name	Description
	SAP Input	Reads data from an SAP/R3 application server.
	Salesforce Input	Reads data from Salesforce.com.
	Salesforce Delete	Deletes record in Salesforce.com module.
	Salesforce Insert	Inserts record(s) in Salesforce.com module.
	Salesforce Update	Updates record(s) in Salesforce.com module.
	Salesforce Upsert	Insert or update records in Salesforce.com module.
	Data Grid	Allows for the creation of static rows of data for reference or testing purposes.

	Memory Group By	For smaller groups this allows intermediate statistical results to be kept in memory for faster performance.
	Teradata Fastload Bulk Loader	Bulk loads data into Teradata.
	OLAP Input	Reads data from an OLAP server using olap4j over XML/A (Mondrian, Palo, SSAS, SAP B/W).
	Add value field changing sequence	Creates a sequence that gets reset when the values in a set of fields changes (group sequence).
	User Defined Java Class	Creates custom plugin on the fly in a step.
	Send Message to Syslog	Sends a message to a Syslog server.
	Java Filter	Filter based on a User Defined Java Expression.
	Farrago streaming bulk loader	Loads data into LucidDB by using remote rows UDX.

New Job Entries

Pentaho Data Integration 4.0 adds the following new job entries:

Icon	Step Name	Description
	Check DB Connections	Checks for valid database connection before proceeding.
	Send Information using Syslog	Sends a message using a Syslog server.