



What's New In PDI 4.4



This document supports Pentaho Business Analytics Suite 4.8 GA and Pentaho Data Integration 4.4 GA, documentation revision October 31, 2012.

This document is copyright © 2012 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 have questions that are not covered in this guide, or if you would like to report errors in the documentation, 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 on the topics covered in this guide, visit <http://www.pentaho.com/training>.

Limits of Liability and Disclaimer of Warranty

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.

Company Information

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>

E-mail: communityconnection@pentaho.com

Sales Inquiries: sales@pentaho.com

Documentation Suggestions: documentation@pentaho.com

Sign-up for our newsletter: <http://community.pentaho.com/newsletter/>

Contents

Introducing Instaview.....4

What's New in 4.4?.....5

Introducing Instaview

Pentaho Instaview is the fastest way to start using Pentaho Data Integration to analyze and visualize data.

Instaview uses templates to manage the complexities of data access and preparation. You can focus on selecting and filtering the data you want to explore, rather than spending time creating source connections and identifying measure and dimension fields.

Once the data has been selected, Instaview automatically generates transform and metadata models, executes them, and launches Pentaho Analyzer. This allows you to explore your data in the Analyzer desktop user interface.

With Instaview, you no longer have to set up a staging database, understand ETL transforms, learn about metadata models, or publish to a server before visualizing data. As your data requirements become more advanced, you have the ability to create your own templates and use the full power of Pentaho Data Integration (PDI).

See the *Getting Started with Pentaho Data Integration Instaview Guide* to understand and learn more about Pentaho Instaview.

What's New in 4.4?

There are several major changes between Pentaho Data Integration 4.3 and 4.4.

PDI Operations Mart

- **The PDI Operations Mart enables administrators to collect and query PDI log data into one centralized data mart for easy reporting and analysis.**
 - The operations mart has predefined samples for Pentaho Analyzer, Interactive Reporting, and Dashboards. You can create individualized reports to meet your specific needs.
- Sample inquiries include
- How many jobs or transformations have been successful compared to how many failed in a given period?
 - How many jobs or transformations are currently running?
 - What are the longest running jobs or transformations in a given period?
 - What is the highest failure rate of job or transformations in a given period?
 - How many rows have been processed in a particular time period? This enables you to see a trend of rows or time in time series for selected transformations.
 - The operations mart provides setup procedures for MySQL, Oracle, and PostgreSQL databases.
 - Install instructions for the PDI Operations Mart are available in the Pentaho InfoCenter.

Concat Fields Step

- **The Concat Fields step is used to join multiple fields into one target field.**
 - The fields can be delimited by a separator and the enclosure logic is completely compatible with the **Text File Output** step.
 - This step is very useful for joining fields as key/value pairs for the Hadoop **MapReduce Output** step.

SAS Input Step

- **The SAS Input step reads files in sas7bdat format created by SAS software.**
 - This step allows PDI developers to import files in sas7bdat format.

EDI to XML Step

- **The EDI to XML step converts EDI message text, which conforms to the ISO 9735 standard, to generic XML.**
 - The XML text is more accessible and enables selective data extraction using XPath and the **Get Data From XML** step.

New Pentaho Data Integration Software Development Kit

- ***Extending and Embedding Pentaho Data Integration* enables developers to utilize PDI beyond the out-of-the box functionality.**
 - This guide explains the mechanics of extending PDI plugins. It also explains embedding PDI functionality directly into Java applications. Pentaho provides sample code for all plugin types and embedding scenarios.