



Explore What's New in DI



This document supports Pentaho Business Analytics Suite 5.0 GA and Pentaho Data Integration 5.0 GA, documentation revision September 6, 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.

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

New Features for Pentaho Data Integration..... 4

New Features for Pentaho Data Integration

Improved System Performance Monitoring

Administrators can use the Operations Mart on both the BA and DI Servers to monitor system performance.

- Improved usability of the Analyzer dimensions
- Improved usability with the data mart loading
- Minor bug fixes to the PDI-specific portions of the Pentaho Operations Mart

Deliver Data From Multiple Data Sources

The Data Services and Kettle JDBC driver enable you to deliver data from multiple data sources, while enriching, cleansing, and transforming the data. We use Data Services for seamless integration into Enterprise Service Buses (ESBs). Integrate into any JDBC compliant application, such as a reporting system or in-house Java application.

- **Data Services:** Query Data via the Web (DI Server and Carte) and return in JSON, XML or text format from any transformation

Kettle JDBC driver: Query data via JDBC that is returned from any transformation

Data Movement Load Balancing

PDI provides load balancing of data within transformations and over multiple cluster nodes when using transformation clustering.

Revert Changes in Job Database Transactions

Database connections can be used with all jobs. This enables commits and rollbacks on a job level. Prior to this release, this was only possible with transformations.

Improved Restart Experience with Checkpoints

- If a job fails, you can resume it after the most recent checkpoint.
- Checkpoints ensure that the status is serialized (variables, arguments, files, and rows from the result set) and that the job can be continued after a safe state that is defined during design.

Easily Add New Plugins

The PDI Marketplace makes it possible to share and download new plugins. These plugins include steps, job entries, database types, perspectives, and more.

Data Profiling Enhancements

- Data Profiling Perspective includes DataCleaner: Analyze Tables and Columns in preparation for ETL
- View common statistics such as cardinality, min/max values, and nulls.
- Profile output for steps within an ETL transformation to analyze prepared data
- See Kettle Data Profiling with DataCleaner, which you download with the Market Place.

Further Highlights

- Integration with WebSphere MQ / MQSeries
- Help Button and direct link to the documentation from any step and job entry
- Continues Preview Mode - helps to see the results of changes in a step (field names, types etc.) immediately
- Detailed timing metrics for low level operations – helps to analyze bottle necks in more detail, e.g. database connect & query time vs. transformation time
- Simplified looping - call jobs within a transformation and loop through the rows
- Calling a job with kitchen and specifying the -logfile option includes summary information within the .log file (Start Time, Stop Time, Processing Time)