Explore What's New in DI

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

Trademarks

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

Enterprise Deployment Enhancements

Job Restartability

- Set checkpoints
- Restart jobs from last successful check point

Transactional Job Execution

- Provides the ability to 'roll back' job execution on failure

Security for Database Connections

- Full permissions: read, write, and delete
- Securely share connections with other users or groups

Data Movement Load Balancing

Load Balancing of data within transformations

Load balancing of data over multiple cluster nodes with using transformation clustering

Data Federation

Data Services: Query data via the Web, stream results using JSON, XML or Text Format from any transformation

Kettle JDBC driver allows you to query data using SQL/JDBC from any transformation

Use cases include:

- Rapidly deliver prepared (joined, cleaned, enriched, transformed) data mashups from multiple data sources
- Integrate prepared mashups with ESBs using Data Services
- Leverage prepared mashups with other JDBC compliant applications including Pentaho Reporting and Analytics

Marketplace

Share and/or download new plugins

Supports steps, job entries, database types, perspectives, and more

Open in Spoon under Help > Marketplace

Marketplace Highlight: Data Cleaner

Data profiling by Human Interface

- Analyzer tables and columns in preparation for ETL
- Profile step output in an ETL transformation to analyze prepared data
- Broad set of profiling analytics
  - Numeric: cardinality, min/max values, nulls, and more
  - String: min/max characters, case, whitespace, and more
  - Others: analytic options like pattern matching, de-duplication, character set distribution, data gap, date/time and more

Additional Data Integration Highlights

- WebSphere MQ/MQSeries Integration
- Inline help links on transformation steps and job entries
- Simplified looping - call jobs and transformations within a transformation and loop through the rows
- Detailed timing metrics for low level operations – helps to analyze bottle necks in more detail, e.g. database connect and query time vs. transformation time
- Extended monitoring of sub jobs and transformations in the Carte- and DI-Server (Expand remote job option)
- Pluggable and new data types, e.g. Internet Address, Time stamp
- Introduction of REST services (Carte and DI-Server)

- Several new Transformation Steps and Job Entries including: Splunk input/output, table compare, zip file, OpenERP input/output, Telnet and WMI Input

Expanded Partner Ecosystem Support

Enhanced Functionality

- New InstaView use case templates for Hadoop and Splunk
- Expanded NoSQL Integration

Expanded Ecosystem Support

- Hadoop high availability support
- New integrations: RedShift, Impala, Splunk
- New Hadoop Certs: Intel, Hortonworks, DataStax
- Support for latest versions of CDH, MapR, MongoDB, and Cassandra

Enhanced NoSQL Support

Leadership in Big Data Integration and Analytics

10Gen

- Added support for Aggregation Framework, Replica Sets, and Tag Sets
- Metadata discovery--samples documents to automatically determine available fields and data types
- Enhanced control for inserting/upserting data into MongoDB collections
- Enhanced Instaview Template for MongoDB analytics
- All enhancements available for Reporting and Data Integration

Datastax

- Added support for CQL-3 (continuing support for CQL-2)
- Enhanced capabilities for writing data
- Composite key support
- Ability to set Time To Live
- Query preview
- Support non-textural column names

Adaptive Big Data Layer

Transparent Access to and Integration of Big Data

- Insulates from changing versions, vendors, data stores
- Give customers broad flexibility of choice, rapid time to value, reduced risk
- Provides native integration into big data ecosystem
- Broadest, deepest Big Data Support