Version 5.2

Explore What's New in DI

Copyright Page

This document supports Pentaho Business Analytics Suite 5.2 GA and Pentaho Data Integration 5.2 GA, documentation revision October 7, 2014, copyright © 2014 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

To view the most up-to-date help content, visit https://help.pentaho.com.

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

The trademarks, logos, and service marks ("Marks") displayed on this website are the property of Pentaho Corporation or third party owners of such Marks. You are not permitted to use, copy, or imitate the Mark, in whole or in part, without the prior written consent of Pentaho Corporation or such third party. Trademarks of Pentaho Corporation include, but are not limited, to "Pentaho", its products, services and the Pentaho logo.

Trademarked names may appear throughout this website. Rather than list the names and entities that own the trademarks or inserting a trademark symbol with each mention of the trademarked name, Pentaho Corporation 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 460

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

New Features in Pentaho Data Integration 5.2

Pentaho Data Integration 5.2 delivers many exciting and powerful features that help you quickly and securely access, blend, transform, and explore data.

R Script Executor Step Improvements

The **R Script Executor**, **Weka Forecasting**, and **Weka Scoring** steps form the core of the Data Science Pack and transforms PDI into a powerful, predictive analytics tool. The R Script Executor step allows you to incorporate R scripts in your transformation so that you can include R-based statistical programming in your data flow. In PDI Version 5.2 you can now "plug and play" R scripts, without extra customization. Now you can pass incoming field metadata to the output field metadata, use a more intuitive user interface to run scripts by rows or by batches, and test scripts.

Related Content:

- R Script Executor Step
- Weka Forecasting and Weka Scoring Steps

New DI Server Administration Features

Porting content from one environment to another and performing general DI Repository maintenance is easier with the introduction of the new Purge Utility. The Purge Utility permanently purges the repository of versions of shared objects, such as database connection information, jobs, and transformations. You can also turn DI Repository versioning and comment capturing capabilities on and off.

Related Content:

- Set DI Version Control and Comment Tracking Options
- Purge Content and Shared Objects from the DI Repository

Kerberos Security Support for CDH 5.1 and HDP 2.1

If you are already using Kerberos to authenticate access to a Cloudera Distributed Hadoop 5.1 or Hortonworks Data Platform 2.1 cluster, with a little extra configuration, you can also use Kerberos to authenticate Pentaho DI users who need to access those clusters.

Related Content:

- Implement Kerberos Authentication

New Marketplace Plugins

Pentaho Marketplace continues to grow with many more of your contributions. Pentaho Marketplace is a home for community-developed plugins and a place where you can contribute, learn, benefit from, and connect to others. New contributions include:

- **LookupTimeDimensionStep**: Looks up and creates an entry on a data warehouse dimension time table and returns the ID.
- **Probabilistic Row Distributions**: Contains a collection of Row Distribution plugins for PDI that use probabilistic methods for determining the distribution of rows.
- **PDI Groovy Console**: Adds a Groovy console to the Help menu that has helper methods and classes that interact with the PDI environment.
- **Gremlin Script Step**: Provides a Gremlin script step for graph pipeline processing.

Related Content:

- [PDI Marketplace](#)
- [Pentaho Marketplace Community Site](#)

Improved Upgrade Experience

Upgrading PDI is easier because it is no longer a manual process. You can now upgrade from 5.1.x to 5.2 using the same upgrade utility used for patch releases.

Related Content:

- [Upgrade BA and DI](#)

New Documentation

There is now only one upgrade instead of two.

Related Content:

- [Upgrade BA and DI](#)

Minor Functionality Changes

To learn more about minor functionality changes that might impact your upgrade experience, see the [PDI 5.1 to 5.2 Functionality Changes](#) article.

PDI Version 5.2 Minor Functionality Changes

The following table describes minor functionality changes for PDI version 5.2. Provisions to preserve backward compatibility and related Jira cases are also listed.

Description	Related Jira Cases
The functionality of an export was changed for linked repository objects with the option "Specify by reference". In this case, the exported objects get replaced by "Specify by name" and the resolved path. This ensures a correct import to a different repository. To be compatible with the behavior to previous versions, set variable KETTLE_EXPORT_COMPATIBILITY_KEEP_REFERENCE to Y.	• PDI-9797
Special characters in the DI-Repository are no longer replaced by "_". To be compatible with the behavior to previous versions, set variable KETTLE_COMPATIBILITY_PUR_OLD_NAMING_MODE to Y.	• PDI-6394