Upgrading From PDI 4.1.0 to 4.1.1

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

Introduction

This guide shows current Pentaho Data Integration Enterprise Edition customers how to upgrade from PDI 4.1.0 to 4.1.1. If you are upgrading from a version of PDI older than 4.1.0, refer instead to the upgrade document or documents in the Pentaho Knowledge Base that apply to your specific situation.

Note: This guide is only for Enterprise Edition upgrades. You cannot upgrade a PDI 4.1.0 Community Edition installation to a PDI 4.1.1 Enterprise Edition instance using this process.

What's new in 4.1.1?

Pentaho Data Integration 4.1.1 primarily represents a collection of performance improvements in the enterprise repository. Here is a full list of improvements between 4.1.0 and 4.1.1: *http://jira.pentaho.com/secure/ IssueNavigator.jspa?reset=true&jqlQuery=project+%3D+PDI+AND+fixVersion+%3D+%224.1.1%22+AND+status+in+ %28Resolved%2C+Closed%2C+%22Ready+For+Test%22%29+ORDER+BY+priority+DESC*. Among them, these are the highlights:

- **Improved import/export performance:** operations involving loading objects from the enterprise repository no longer execute a large number of server calls.
  - **Impact:** Interaction with the enterprise repository should be dramatically improved.
- **Client-side shared objects caching:** objects used by database connections, slave servers, cluster schemas, and partition schemas are now refreshed from the enterprise repository when they are first requested (opening a transformation or job will fetch them). Once they have been loaded, you can clear them with the new **Clear Shared Objects Cache** function in the **Repository** section of the **Tools** menu.
  - **Impact:** BI developers should be aware of this new cache scheme so that they can use it appropriately.
- **EE Repository no longer restructures the Repository Explorer view for non-admin users:** PDI users can now see all directories in the repository that they have read permissions for. This alleviates the need to share all documents from a /public folder, although that practice is still recommended.
  - **Impact:** BI developers and administrators may need to adjust permissions settings to more appropriately meet their needs.
- **Locks on transformations and jobs are now retrieved when opening a dialogue:** while this improves the responsiveness of the Repository Explorer, it also prevents users from updating lock status by changing directories. Locks that change while the Repository Explorer is open will still take effect, but the status icon will not be updated until the Repository Explorer window has been closed and reopened.
  - **Impact:** BI developers should be aware of this change in content lock status display, and adjust their habits accordingly.
- **New command-line parameters for limiting log output:** Pan and Kitchen now have two parameters that limit in-memory log output: **maxloglines** and **maxlogtimeout**. The first is the number of lines stored in memory, the second is the maximum number of minutes each log line should be kept.
  - **Impact:** BI developers who use Pan and Kitchen should be aware of these new parameters and modify their scripts and jobs accordingly.

Upgrade Best Practices

All production software upgrades, including Pentaho Data Integration, should be performed during off-peak hours and with enough time to restore from a backup before off-peak ends if something should go wrong. Ideally you would perform the upgrade on a test machine that mirrors the production environment, take notes along the way, and perform the same procedure on the production server when you know how long the entire process will take and are sure that there will be no unexpected problems.

This guide contains instructions for performing the safest and fastest upgrade to PDI 4.1.1.

Always back up your production data, and test the backup before proceeding with an upgrade.

Obtaining the Archive Packages

Consult the Welcome Kit email that was sent to you after completing the sales process. This email contains user credentials for the Enterprise Edition FTP site, where you can download individual archive packages for the DI Server and Data Integration client tools. Here are the packages you need for each platform and distribution:

- **DI Server for Windows:** `pdi-ee-server-4.1.3-GA.zip`
- **DI Server for Linux/Solaris/OS X:** `pdi-ee-server-4.1.3-GA.tar.gz`
- **Data Integration client tool Windows package:** `pdi-ee-client-4.1.3-GA.zip`
- **Data Integration client tool Linux/Solaris/OS X package:** `pdi-ee-client-4.1.3-GA.tar.gz`
- **Data Integration for Hadoop:** `phd-ee-4.1.3-GA.zip`

If you download the **pdi-ee-server** package, you must also download the Pentaho Enterprise Console package:

- **Pentaho Enterprise Console for Linux/Solaris/OS X:** `pec-3.8.0.1-GA.tar.gz`
- **Pentaho Enterprise Console for Windows:** `pec-3.8.0.1-GA.zip`

Creating Backups

You should back up your content files or repository and your Kettle settings in case something goes wrong with the upgrade. Refer to the sections below that apply to your situation.

Backing Up Content Files

If you do not use a database or enterprise repository for storing PDI content, then you are saving individual KJB and KTR files on a local or network drive. Hopefully you have created a sensible directory structure and naming convention for them. If not, this may be a good time to organize them properly.

Once you have all of your content in one directory, simply create a Zip or tar archive of it and copy the archive to a safe location outside of your local machine, such as a network drive or removable media.

This should be part of your normal production backup routine outside of this upgrade process.

Backing Up a Database Repository

Backing up your PDI database repository is as simple as using the **Export complete repository to XML** functionality in Spoon's **Repository Explorer** dialogue, which is accessible from the **File** menu. Then copy the resulting file to a safe location outside of the machine you are upgrading.

This should be part of your normal production backup routine outside of this upgrade process.

How to Back Up the Enterprise Repository

Follow the instructions below to create a backup of your PDI enterprise repository.

> **Note:** If you've made any changes to the Pentaho Enterprise Console or DI Server Web application configuration, such as changing the port number or base URL, you will have to modify this procedure to include the entire `/pentaho/server/` directory.

1. Stop the DI Server.

   `/pentaho/server/data-integration-server/stop-pentaho.sh`

2. Create a backup archive or package of the `/pentaho/server/data-integration-server/pentaho-solutions/` directory.

   `tar -cf pdi_backup.tar /pentaho/server/data-integration-server/pentaho-solutions/`

3. Copy the backup archive to removable media or an online backup server.

4. Start the DI Server.

   `/pentaho/server/data-integration-server/start-pentaho.sh`

Your DI Server's stored content, settings, schedules, and user/role information is now backed up.

To restore from this backup, simply unpack it to the same location, overwriting all files that already exist there.

Backing Up the .kettle Directory

The **.kettle** directory stores all of your client tool configuration settings and preferences. It is located in `~/.kettle` on Linux, Solaris, and OS X; and `C:\Documents and Settings\username\.kettle` on Windows, where *username* refers to the user account that the PDI client tools are installed to.

Create a Zip or tar archive of this directory and copy the archive to a safe location before upgrading.

Upgrading a Data Integration Server

Ensure that the DI Server and Pentaho Enterprise Console are stopped before continuing. You must have a PDI 4.1.0 DI Server installed in order to follow this procedure; if you do not use the DI Server, this upgrade task is unnecessary.

> **Note:** For a smoother post-upgrade test experience, you should perform this procedure before upgrading your PDI workstations.

Follow the instructions below to upgrade your Data Integration Server to version 4.1.1.

1. Remove the `/pentaho/server/license-installer/` directory if it exists.

   A new license installer will be unpacked with the new PDI archive. The new license utility will work with the version of PDI you're upgrading from, so there is no need to back it up. There may not be an existing license-installer directory in your PDI 4.1.0 installation; this is provided with the archive package, but not the graphical installer.

2. Rename the `/data-integration-server/` directory to **data-integration-server-old**.

   > **Note:** If you are coming from a BI Server upgrade, you already have a **server_old** directory. If this is the case, use `/server_old/data-integration-server/` in place of `/data-integration-server-old/`.

   ```
   mv /home/pentaho/pentaho/server/data-integration-server/ /home/pentaho/pentaho/server/
   data-integration-server-old/
   ```

3. Unpack the **pdi-ee-server-4.1.1-GA** package to the parent of the directory you just renamed.

   ```
   tar zxvf ~/downloads/pdi-ee-server-4.1.1-GA.tar.gz -C /home/pentaho/pentaho/server/
   ```

4. Copy all of the **applicationContext** files from the `/data-integration-server-old/pentaho-solutions/system/` directory to the new one, overwriting the equivalent files that are already there.

   ```
   cp applicationContext-* ~/pentaho/server/data-integration-server/pentaho-solutions/
   system/
   ```

5. Copy the **pentaho-spring-beans.xml** file from the `/data-integration-server-old/pentaho-solutions/system/` directory to the new one, overwriting the equivalent file that is already there.

   ```
   cp pentaho-spring-beans.xml ~/pentaho/server/data-integration-server/pentaho-
   solutions/system/
   ```

6. Transfer the information about the **admin role** from the following two old files to the new ones: **/pentaho-solutions/system/pentaho.xml** and **/pentaho-solutions/system/repository.spring.xml**

   ```
   <acl-voter>
   <!-- What role must someone be in to be an ADMIN of Pentaho -->
       <admin-role>Admin</admin-role>
   </acl-voter>

   <!-- The name of the authority which is granted to all admin users in single-tenancy
    mode. -->
   <bean id="singleTenantAdminAuthorityName" class="java.lang.String">
       <constructor-arg value="Admin" />
   </bean>
   ```

7. Copy the entire **quartz** directory from `/data-integration-server-old/pentaho-solutions/` to the new one.

   ```
   cp -r ./quartz ~/pentaho/server/data-integration-server/pentaho-solutions/
   ```

8. Copy the entire **hibernate** directory from `/data-integration-server-old/pentaho-solutions/system/hibernate/` to the new one.

   ```
   cp -r ./hibernate ~/pentaho/server/data-integration-server/pentaho-solutions/system/
   ```

9. Copy the entire **repository** directory from `/data-integration-server-old/pentaho-solutions/system/jackrabbit/` to the new one.

   ```
   cp -r ./jackrabbit/repository/ ~/pentaho/server/data-integration-server/pentaho-
   solutions/system/jackrabbit/
   ```

10. Copy the **context.xml** file from `/data-integration-server-old/tomcat/webapps/pentaho-di/META-INF/` to the new one.

   ```
   cp ./context.xml ~/pentaho/server/data-integration-server/tomcat/webapps/pentaho-di/
   META-INF/
   ```

11. Copy the **log4j.xml** file from `/data-integration-server-old/tomcat/webapps/pentaho-di/WEB-INF/classes/` to the new one.

```
cp ./log4j.xml ~/pentaho/server/data-integration-server/tomcat/webapps/pentaho-di/WEB-INF/classes/
```

12. Copy the **scripts** directory from `/data-integration-server-old/` directory to the new data-integration-server directory.

👉 **Note:** The scripts directory will only exist if you installed PDI from a graphical installation utility. If you installed via archive packages, it won't be there. If you do not see a scripts directory, then skip this step.

13. Optional: Copy the entire **jre** directory from `/data-integration-server-old/` to the new one.

```
cp -r jre ~/pentaho/server/data-integration-server/
```

This step is optional. If you already have a supported JRE or JDK installed on your system, you can skip copying this directory and simply ensure that you have a JAVA_HOME or PENTAHO_JAVA_HOME system variable that points to your Java instance.

14. If you have not already done so, merge any custom changes you have made to DI Server configuration files from the old ones to the new ones.

Your DI Server is now upgraded to version 4.1.1. Continue on to the next subsection to upgrade the Pentaho Enterprise Console.

Special Task for 64-bit Windows Environments

This procedure is **only** for 64-bit Windows deployments. It is **not** for 32-bit Windows or Linux systems.

Due to Apache Tomcat packaging changes, some 64-bit Windows binaries are not installed with the PDI archive. In order for the DI Server to continue working as expected, you must download some files from the Tomcat Web site. Follow the instructions below.

1. Download the Apache Tomcat 64-bit Windows archive package: *http://apache.osuosl.org/tomcat/tomcat-6/v6.0.29/bin/apache-tomcat-6.0.29-windows-x64.zip*.
2. Unpack the archive to a temporary location.
3. Copy the following files from the unpacked Tomcat directory to `/pentaho/server/data-integration-server/tomcat/bin/`, overwriting the existing files there:

   - Service.bat
   - Tomcat6.exe
   - Tomcat6w.exe
   - tcnative-1.dll

4. Delete the unpacked Tomcat archive directory.

Your DI Server upgrade on 64-bit Windows is complete. Continue on to the section on upgrading Enterprise Console.

Upgrading the Pentaho Enterprise Console

The upgraded DI Server will not work properly without upgrading the Pentaho Enterprise Console. To upgrade, follow the below process.

1. Rename the `/pentaho/server/enterprise-console/` directory to **enterprise-console-old**.

```
mv enterprise-console enterprise-console-old
```

2. Unpack the `pec-ee-3.8.0-GA` zip or tar.gz file to `/pentaho/server/`.

```
tar zxvf ~/downloads/pec-ee-3.8.0-GA.tar.gz -C /home/pentaho/pentaho/server/
```

3. Copy the following files from your old `/pentaho/server/enterprise-console-old/resource/config/` directory into the new one:

   - console.xml
   - console.properties
   - login.properties

- login.conf
- log4j.xml

```
cp /home/pentaho/pentaho/server/enterprise-console-old/resource/config/console.* /
home/pentaho/pentaho/server/enterprise-console/resource/config/ && cp /pentaho/server/
enterprise-console-old/resource/config/log* /home/pentaho/pentaho/server/enterprise-
console/resource/config/
```

The Pentaho Enterprise Console has been upgraded to the latest version.

Upgrading a Data Integration Workstation

Ensure that the Data Integration client tools (Spoon, or Pan or Kitchen via scripts or schedules) are not running on the machine before proceeding.

Note: If you use an enterprise repository to store your jobs and transformations, you should perform the DI Server upgrade process before this in order to properly test DI Server connectivity from your upgraded workstations.

Follow the directions below to upgrade your PDI workstations to version 4.1.1. Adjust the installation paths to match your scenario.

1. Remove your existing `/data-integration/` directory.

   ```
   rm -rf /home/rwilco/pentaho/design-tools/data-integration/
   ```

2. Unpack the **pdi-ee-client-4.1.1-GA** package to the same location that you just deleted.

   ```
   cd /home/rwilco/pentaho/design-tools/ && tar zxvf ~/downloads/pdi-ee-client-4.1.1-
   GA.tar.gz
   ```

3. Start Data Integration and ensure that you can still connect to the DI Server (if you have an enterprise repository connection) and access all of your old content.

This workstation is now upgraded to PDI 4.1.1-GA. Repeat this procedure for other workstations that you have support entitlements for.

Testing and Cleanup

You should now have a complete PDI 4.1.1 environment, from the DI Server to individual client workstations. Before you go back into production, you should perform the following tests:

- Open old jobs and transformations and ensure that they execute properly.
- If you are using an enterprise repository, ensure that each PDI workstation can connect to it.
- Create a new job, transformation, and/or Agile BI analysis schema and save it as you normally would.
- Schedule a job or transformation and ensure that the schedule executes properly.
- Ensure that existing schedules are still valid.
- If you are using an enterprise repository, share a job or transformation between PDI users and verify that both can access it.
- Physically restart the server and ensure that the DI Server and Enterprise Console are automatically started as services, if you have them configured as such.

Once you're certain that your PDI environment is ready for production, you can remove any installation artifacts, such as ZIP or tar.gz archives and installers. You can also remove your **data-integration-server-old** directory.