

# Pentaho Backup and Recovery

# HITACHI

Inspire the Next

Change log (if you want to use it):

Date	Version	Author	Changes

# Contents

- Overview..... 1
- Back Up the Pentaho or BA Server ..... 2
  - Incremental Backup for the Pentaho or BA Server..... 3
  - Full Backup for the Pentaho or BA Server..... 3
- Back Up the DI Server (Pentaho 6.x Only)..... 4
  - Incremental Backup for the DI Server ..... 5
  - Full Backup for the DI Server..... 5
- Restore Pentaho Servers ..... 6
  - Command Line Notes ..... 6
    - Command Line Arguments for Pentaho ..... 6
    - Command Line Arguments for Postgres Database ..... 6
  - Import to the Repository from the DI Server..... 7
    - Export from Command Line..... 7
- Related Information ..... 8
- Finalization Checklist..... 9

This page intentionally left blank.

## Overview

The purpose of this document is to introduce best practices for backing up Pentaho servers. This document only covers a general approach for the user, due to the variety of environments. The following topics are covered in this document:

- Back up the Pentaho or BA server
- Back up the DI server (Pentaho 6.x)
- Restore Pentaho servers
- Command Line Tools

The intention of this document is to speak about topics generally; however, these are the specific versions covered here:

Software	Version(s)
Pentaho	6.x, 7.x

The [Components Reference](#) in Pentaho Documentation has a complete list of supported software and hardware.

# Back Up the Pentaho or BA Server

This section explains the recommended backup procedure for a single Pentaho or BA server. The server is mainly composed of three configuration and deployment containers, with some additional file settings. If you encounter several servers, apply the procedure to each one:

Table 1: Backup Procedure for Single Pentaho or BA Server

Web Application Files Using:	Files
<b>Tomcat</b>	tomcat/webapps/pentaho tomcat/webapps/pentaho-style tomcat/webapps/sw-style
<b>JBOSS</b>	jboss-xxx/standalone/deployments/pentaho.war jboss-??*/standalone/deployments/pentaho-style.war jboss-??*/standalone/deployments/sw-style.war

The Pentaho installation contains system configuration, plugins, files, and folders that can be modified during Pentaho use. The default folder's name is `pentaho-solutions`.

1. There are three Pentaho database repositories that store operational data and system data, although you only need to back up two of them (Hibernate and Quartz):
  - a. Jackrabbit
  - b. Hibernate
  - c. Quartz
2. Pentaho's Jackrabbit database requires synchronization with local files stored in the `pentaho-solutions/system/jackrabbit/repository` folder.
3. There are other tools, software, and scripts that also require a backup, in addition to the server operational files:
  - a. `License-installer` tool and `.installedLicences`
  - b. Pentaho Start and Stop scripts
4. Pentaho stores some information by default in the `USER_HOME`. This can be changed upon installation by [setting environment variables](#). Use the `.pentaho.kettle.pentaho-meta` default directory for the `USER_HOME` hidden folders.
5. Stop the Pentaho server before doing a backup. This will guarantee that the information on the database is in sync with the local files stored in the `pentaho-solutions` repositories folder.
6. If you are unable to stop the Pentaho server, make sure to back up the database repositories before using anything in the `pentaho-solutions` folder. Pentaho does not provide any backup scripts or software for this operation.

## Incremental Backup for the Pentaho or BA Server

Database files and low-level files should be incrementally backed up before you back up the BA server.



*We recommend daily incremental backups for Pentaho solutions and Pentaho database repositories, and weekly full backups of the Pentaho web application, Pentaho solutions, and database repositories.*

Here are the steps for an incremental backup:

1. Stop the server.
2. Perform an incremental backup for the Pentaho database repositories.



*You only need to back up the Hibernate and Quartz database repositories.*

3. Perform an incremental backup for the `pentaho-solutions` folder. If the system is installed as a cluster, back up each folder.
4. Start the server.

## Full Backup for the Pentaho or BA Server

Full backups require the backup of the Pentaho database repositories, Pentaho solutions folder, and Pentaho web applications folder.

Keep in mind that the server stores temporal information in `pentaho-solutions/system/tmp` that needs to be backed up. The [Manual Cleanup of the Temporary Folder](#) documentation has tips for this process.

Here are the steps for a full backup:

1. Stop the server.
2. Back up the Pentaho database repositories.



*You only need to back up the Hibernate and Quartz database repositories.*

3. Use the database utilities based on the repository database selected.
4. Back up the Pentaho solutions folder.
5. Back up the Pentaho web applications folder.
6. Include **OTHER PENTAHO FOLDERS**.
7. Start the server.



*If a system is installed as a cluster, back up each `pentaho-solutions` folder, and each `pentaho-applications` folder.*

Those default locations are the `pentaho-solutions` directory and `/tomcat/webapps/pentaho` directory, but they could be customized with customer installation.

## Back Up the DI Server (Pentaho 6.x Only)

This section explains the recommended backup procedure for a single DI server.



*You can still set up separate DI-only servers with Pentaho 7.x, but the repositories are named Jackrabbit, Hibernate, and Quartz.*

The DI server is mainly composed of three configuration containers and three deployment containers. If you encounter several servers, apply the following procedure to each one:

*Table 2: Backup Procedure for Single DI Server*

Web Application Files Using:	Procedure
<b>Tomcat</b>	tomcat/webapps/pentaho-di tomcat/webapps/pentaho-style
<b>JBOSS</b>	jboss-xxx/standalone/deployments/pentaho-di.war jboss-??*/standalone/deployments/pentaho-style.war

The Pentaho installation contains system configuration, plugins, files, and folders that can be modified during Pentaho use. The default folder's name is `pentaho-solutions`.

There are three Pentaho database repositories that store operational data and system data, although you only need to back up two of them (Hibernate and Quartz):

- `di_jackrabbit`
- `di_hibernate`
- `di_quartz`



*Pentaho's `di_jackrabbit` database requires synchronization with local files stored in the `pentaho-solutions/system/jackrabbit` repository folder.*

There are other tools, software, and scripts that also require a backup, in addition to server operational files:

- License-installer tool and `.installedLicences`
- Pentaho Start and Stop scripts

Pentaho stores some information by default in the `USER_HOME`. This can be changed upon installation by setting environment variables. Back up the `.pentaho`, `.kettle`, and `.pentaho-meta` folders which are hidden, by default, inside the `USER_HOME` directory.



*We recommend stopping the DI server before doing a backup. This will guarantee that the information on the database is in sync with the local files stored in the `pentaho-solutions` repositories folder.*

Make sure to back up the database repositories before using anything in `pentaho-solutions` folder, if you are **unable to stop** the server. Pentaho does not provide any backup scripts or software to do the backup.

## Incremental Backup for the DI Server

Database files and low-level files should be incrementally backed up, to back up the DI server.



*We recommend daily incremental backups for Pentaho solutions and Pentaho database repositories, and weekly full backups of the Pentaho web application, Pentaho solutions, and database repositories.*



*It is important to make sure all external clients using the repository are stopped before stopping the DI server. Make sure all CARTE servers and PDI Client tools are disconnected before stopping the server.*

Here are the steps to stop the server:

1. Stop the DI server.
2. Perform an incremental backup for the Pentaho database repositories.



*You only need to back up the Hibernate and Quartz database repositories.*

3. Perform an incremental backup for the `pentaho-solutions` folder. If a system is installed as a cluster, backup each `pentaho-solutions` folder.
4. Start the DI server.

## Full Backup for the DI Server

Full backups require the backup of Pentaho database repositories, Pentaho solutions folder, and Pentaho web applications folder.



*Make sure all external clients using the repository are stopped before stopping the DI server. Make sure all CARTE servers and PDI Client tools are disconnected before stopping the server.*

Here are the steps for making a backup:

1. Stop the DI server.
2. Back up the Pentaho database repositories.



*You only need to back up the Hibernate and Quartz database repositories.*

3. Use the database utilities based on the repository database selected.
4. Back up the Pentaho solutions folder.
5. Back up the Pentaho web applications folder.
6. Include **OTHER PENTAHO FOLDERS**.
7. Start the DI server.



*If a system is installed as a cluster, back up each `pentaho-solutions` folder, and each `pentaho-applications` folder.*

Those default locations are the `pentaho-solutions` directory and `/tomcat/webapps/Pentaho` directory, but they could be customized with customer installation.

## Restore Pentaho Servers

This section explains the recommendations for restoring a backup for the Pentaho/BA and DI servers. Restoring a server from a backup is as simple as recovering the Pentaho solutions, Pentaho web applications, and Pentaho databases.



*It is important that the **restore** locations and IP/DNS names address remain the same, as they are fixed in the configuration files of the platform.*

If any of this information has changed, the configuration files can be changed to point to any new location such as IP, users, or passwords. Refer to the server manual install procedure to review all the configuration files locations before starting up the server.

## Command Line Notes

This section explains some other useful commands that can be used for backup and restore.

### *Command Line Arguments for Pentaho*

Use the following command lines to back up or restore the file's repositories from a Pentaho/BA server. Pentaho documentation has a more in-depth reference for [command line arguments](#).



*For Linux: files use `import-export.sh` instead of `import-export.bat`.*

To back up:

---

```
import-export.bat --backup --url=http://localhost:8080/pentaho --
username=admin --password=password --file-path=c:/home/Downloads/backup.zip
--logfile=c:/temp/logfile.log
```

---

To restore:

---

```
import-export.bat --restore --url=http://localhost:8080/pentaho --
username=admin --password=password --file-path=c:/home/Downloads/backup.zip
--overwrite=true --logfile=c:/temp/logfile.log
```

---

### *Command Line Arguments for Postgres Database*

Refer to the correct version of [PostgreSQL documentation](#) for backup and restore command line arguments for PostgreSQL.

## Import to the Repository from the DI Server

Follow the instructions below to import the repository. You must be already logged into the repository in the PDI client before you can perform this task.

1. In the PDI client, go to **Tools --> Repository --> Import Repository**

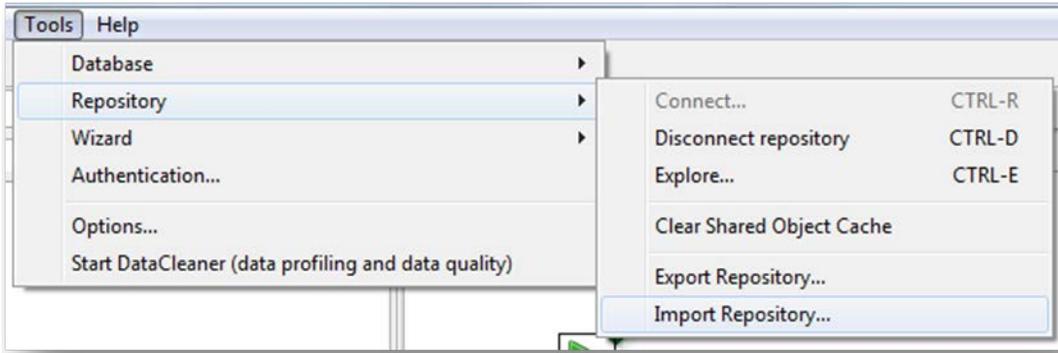


Figure 1: Import Repository

2. Locate the export (XML) file that contains the solution repository contents.
3. Select **Open**. The **Directory Selection** dialog box will appear.
4. Select the directory in which you want to import the repository.
5. Select **Ok**.
6. Enter a comment, if applicable.
7. Wait for the import process to complete.
8. Select **Close**.

## Export from Command Line

---

```
kitchen.bat /file:C:\Pentaho_samples\repository\repository_export.kjb
"/param:rep_name=PDI2000" "/param:rep_user=joe"
"/param:rep_password=password" "/param:rep_folder=/public/dev"
"/param:target_filename=C:\Pentaho_samples\repository\export\dev.xml"
```

---

## Related Information

Here are some links to information that you may find helpful while using this best practices document:

- [Import and Export PDI Content](#)
- [Set Kettle Variables](#)
- [Manual Cleanup of the Temporary Folder](#)
- [Command Line Arguments](#)
- [PostgreSQL Documentation](#)
- [Back Up and Restore Pentaho Repositories](#)

# Finalization Checklist

This checklist is designed to be added to any implemented project that uses this collection of best practices, to verify that all items have been considered and reviews have been performed.

Name of the Project: \_\_\_\_\_

Date of the Review: \_\_\_\_\_

Name of the Reviewer: \_\_\_\_\_

Item	Response	Comments
Did you perform incremental backups before backing up both the Pentaho and BA servers?	YES _____ NO _____	
Did you stop the servers before starting the incremental backups?	YES _____ NO _____	
Did you back up the Pentaho database repositories, Pentaho solutions folder, and Pentaho web applications folder before running a full backup for both the Pentaho and BA servers?	YES _____ NO _____	
Did you stop the server before backing up the database repositories?	YES _____ NO _____	
Did you perform the same steps above for backing up the DI server?	YES _____ NO _____	