HITACHI
Inspire the Next

Best Practices for Installation & Upgrade

HITACHI
Inspire the Next

Change log (if you want to use it):

Date	Version	Author	Changes

Contents

This page intentionally left blank.

Overview

This document is intended to provide a series of best practices around the installation of the Pentaho suite of software.

Topics are arranged in a series of groups with individual best practices for the topic explained. It is not intended to demonstrate how to implement each best practice or provide templates based on the best practices defined within the document.

Our intended audience is Pentaho administrators, or anyone with a background in big data, database creation or administration, who is interested in installing Pentaho.

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, 8.0

The Components Reference in Pentaho Documentation has a complete list of supported software and hardware.

Best Practices for Installing Pentaho

We have collected this series of best practices to help you with fine-tuning your Pentaho installation.

Keep in mind that if you are running ETL overnight and Analyzer during the day, it is OK to run both on the same server. However, if you run ETL and Analyzer at the same time, it is probably better to have separate servers, one for ETL and one for reporting.

You can find details on these best practices in the following sections:

General Installation

We recommend using the archive installation method, instead of the graphical installation wizard.

The graphical installer is used for evaluation and demonstration purposes, and is not suitable for a large production installation.

Using the archive installation method for your production environment will ensure that you can properly configure the Pentaho software. You can still use the graphical installation wizard to install Pentaho design tools on individual workstations.

Memory Allocation

Our recommendations for allocating memory are to:

Make Sure to Use Enough Resources

- **Recommendation:** Allow the Java Virtual Machine (JVM) to allocate as much memory as possible, up to 24GB.
- **Rationale:** Make sure that the Java Runtime Environment (JRE) instance that runs the Pentaho servers has adequate resources available for optimal performance. Adjusting the memory limit is an easy configuration change, and depends on the client tool or web application server you are using.
- **Solution:** Increase the memory limit for JVM. Change the Java VM Memory Limits has instructions based on operating system.

Use Larger and Faster Disk Volume

- **Recommendation:** Update the `java.io.tmpdir` and `CATALINA_TMPDIR` variables to use a larger and faster disk volume.
- **Rationale:** By default, the Pentaho installation uses a path within the Pentaho install location. Depending on the install, this may not be the optimal choice. Typically, a larger and faster disk volume should be used for these operations.
- **Solution:** Make the following changes:
  1. In `start-pentaho.sh`:
     Add `–Djava.io.tmpdir = /path/to/disk`
  2. In `catalina.sh`:
     Add `CATALINA_TMPDIR=/path/to/disk`

Session-Related Timeouts

We recommend changing your session-related timeouts from the defaults:

- **Recommendation:** Change default timeouts to protect sensitive server data.
- **Rationale:** The default session timeout in `…\tomcat\webapps\pentaho\WEB-INF\web.xml` is 30 minutes.
- **Solution:** We recommend that you change the following timeouts:
  1. `web.xml`:
     Change the default timeout in the `…\tomcat\webapps\pentaho\WEB-INF\web.xml` file to 5 or 10 minutes.
  2. `server.xml`:
     Update the connection timeout in `…\tomcat\conf\server.xml` from `20000` seconds to an appropriate number depending on your needs.

Solution Database Repository for Enterprise Environments

For enterprise environments, we recommend using an enterprise-ready database solution with database administration (DBA) support to facilitate backups, as well as high availability.

PostgreSQL is supplied as the default repository database with Pentaho. However, this installer version of PostgreSQL version is intended only for demonstration purposes, and is not suited for production/pre-production/user acceptance testing (UAT)/enterprise environments.

Pentaho does support PostgreSQL and other database vendors as Pentaho Repository databases. It is simply the default installation of PostgreSQL that we do not recommend using for production environments.

Post-Install Cleanup

There are a couple of things that we recommend as part of your post-install cleanup:

Remove Sample Data

- **Recommendation:** Remove all remnants of sample data in your production environment.
- **Rationale:** In production environments, remove the `plugin-samples.zip` and `samples.zip` files from the default content before starting the server for the first time.
- **Solution:** Remove any references to samples in `pentaho.xml`, `web.xml`, `sessionStartupActions.xml` and any other configuration files. The samples are great for evaluation, proof-of-concept, and prototype environments, but not production. They can also slow down the performance of the application and log unnecessary errors.

Remove Unused Plugins

- **Recommendation:** Remove unused plugins from the solution.
- **Rationale:** Pentaho ships with a wide variety of plugins that may or may not be needed. Unused plugins consume additional memory and processing power that would be better utilized by the necessary plugins. Typical candidates include the big data, geo, and mobile plugins.
- **Solution:** Plugins that are not needed should be removed.

Web Applications

Make sure that you change the fully qualified URL for your web application from `localhost` to DNS or IP.

For web applications embedding Pentaho products, external references such as image tags refer to the Pentaho server and will not render properly when `localhost` is used. Therefore, you should modify the `web.xml` file for the application to include the proper fully-qualified URL for the application with DNS and port number.

Related Information

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

- Archive Installation of the Pentaho Server
- Change the Java VM Memory Limits
- Changing the Pentaho Server Fully Qualified URL
- Java Virtual Machine
- Pentaho Components Reference
- Remove Sample Data from the Pentaho Server
- Solution Database 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 use the archive installation method for your production environment?	YES_____ NO_____	
Did you allow the JVM to allocate as much memory as possible?	YES_____ NO_____	
Did you update the `java.io.tmpdir` and `CATALINA_TMPDIR` variables to use a larger and faster disk volume?	YES_____ NO_____	
Did you change your session-related timeouts to reflect your needs?	YES_____ NO_____	
Did you use an enterprise-ready database solution with DBA support?	YES_____ NO_____	
Did you remove all sample data from your production environment?	YES_____ NO_____	
Did you remove unused plugins?	YES_____ NO_____	
Did you change the fully qualified URL for your web application from `localhost` to DNS or IP?	YES_____ NO_____	