



# Install All DI Components

## Copyright Page

---

This document supports Pentaho Business Analytics Suite 5.4 GA and Pentaho Data Integration 5.4 GA, documentation revision June 9th, 2015, copyright © 2015 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](mailto: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](mailto:sales@pentaho.com)

# Introduction to DI Wizard Installation

---

The Wizard is the easiest and quickest way to install Pentaho Data Integration (DI). You can use DI to create and schedule data integration processes, model data, and perform other processing related to traditional and big data sources. The Wizard can be used to install the DI Server, Spoon design tool, and the Kitchen, Pan, and Carte utilities. The Wizard also installs plugins such as Instaview and Big Data.

## Prerequisites

If you are not sure which installation method to use, read [Select DI Installation Option](#) before you begin.

## Expertise

These instructions are written for anyone who wants to install PDI.

## Tools

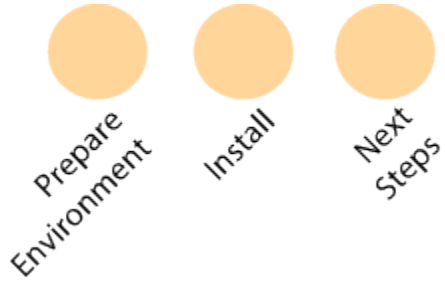
You must supply a server or workstation that meets the hardware and operating system requirements indicated in [support technologies](#).

## Login Credentials

All of the tasks in these instructions require that you have the computer accesses or privileges you need to install software on your server or workstation.

## Wizard Installation Process Overview

---

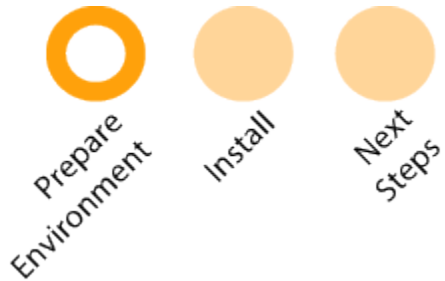


The Wizard installation process consists of these steps.

- The [Prepare Environment](#) step explains how to prepare your computer for installation of the software.
- The [Install](#) step explains how to use the wizard to install the software.
- [Next Steps](#) indicates what to do after you install the software.

## Prepare Environment

---



This step explains how to prepare your computer for the installation of the software. If you plan to install the software on a Mac, you also need to adjust the amount of memory the Mac OS allocates for PostgreSQL.

## Create User Account

### Windows User Account

If you plan to install PDI on a computer that runs Windows, you do not need to create a special user account. Just make sure that you have the appropriate privileges needed to install software. Often, this means that you need administrative-level privileges for the account that you use to install the software. Some Windows systems have expanded user access and permissions restrictions that change the way the software is installed and accessed by system users. If this happens, the installation wizard will require that you execute it with elevated permissions (**Run Elevated** status) so that the DI Server is accessible to all users.

### Linux User Account

If you plan to install PDI on a computer that runs Linux, create a user account named **pentaho** on the computer.

1. Establish a connection to the remote server.
2. In the **Terminal** window, create a new user account called **pentaho**, along with the `pentaho` home directory, by typing this command.

```
sudo useradd -s /bin/bash -m pentaho
```

**Note:** `/bin/bash` indicates that the user account should be created using the Bash shell. In many Linux distributions, the default new user shell is `/bin/sh` or some equivalent, such as Dash, that might not use the `~/ .bashrc` configuration file by default. If you don't have or want to use Bash, adjust the instructions accordingly.

3. In the **Terminal** window, assign a password for the **pentaho** user by typing this line.

```
sudo passwd pentaho
```

4. Verify that you can log in using the newly-created **pentaho** user account.

a. In the **Terminal** window, log in by typing this command.

```
su pentaho
```

b. Type the password for the **pentaho** user account if you are prompted.

c. Use the **Terminal** window to navigate to the `pentaho` directory to verify that it has been created. By default, it is in the `/home` directory.

d. Close the **Terminal** window.

## Disable Anti-Spyware and Anti-Virus Programs

Disable any anti-spyware and anti-virus software before installing PDI. Some types of software firewalls and anti-virus scanners may also block the DI Server or prevent it from operating normally. After the server is installed and you have started it, enable the anti-spyware and anti-virus programs again.

## Advanced Topics

Complete the instructions in this section only if you have a headless node or if you plan to install on a Mac OS.

### Prepare a Headless Linux or Solaris Server

There are two headless server scenarios that require special procedures on Linux and Solaris systems. One is for a system that has no video card; the other is for a system that has a video card, but does not have an X server installed. In some situations -- particularly if your server doesn't have a video card -- you will have to perform both procedures in order to properly generate reports with the BA Server.

#### Systems without video cards

The `java.awt.headless` option enables systems without video output and/or human input hardware to execute operations that require them. To set this application server option when the BA Server starts, you will need to modify the startup scripts for either the BA Server, or your Java application server. You do not need to do this now, but you will near the end of these instruction when you perform the [Start BA Server](#) step. For now, add the following item to the list of CATALINA\_OPTS parameters: `-Djava.awt.headless=true`.

The entire line should look something like this:

```
export CATALINA_OPTS="-Djava.awt.headless=true -Xms4096m -Xmx6144m -
XX:MaxPermSize=256m -Dsun.rmi.dgc.client.gcInterval=3600000 -Dsun.rmi.dgc.server.
gcInterval=3600000"
```

If you intend to create a BA Server service control script, you must add this parameter to that script's CATALINA\_OPTS line.

Note: If you do not have an X server installed, you must also follow the below instructions.

#### Systems without X11

To generate charts, the Pentaho Reporting engine requires functionality found in X11. If you are unwilling or unable to install an X server, you can install the **xvfb** package instead. xvfb provides X11 framebuffer emulation, which performs all graphical operations in memory instead of sending them to the screen.

Use your operating system's package manager to properly install xvfb.

## **Adjust Amount of Memory Mac OS Allocates for PostgreSQL**

If you plan to install the software on a Mac OS, and you choose to use PostgreSQL, you need to increase the amount of memory that the Mac OS allocates for PostgreSQL. You can skip these instructions if you plan to install the software on Windows or Linux.

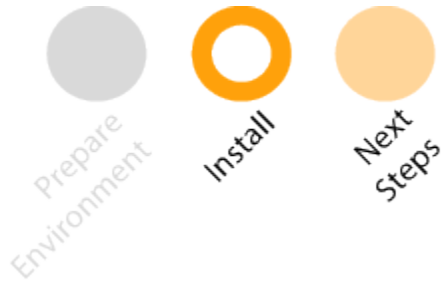
PostgreSQL is the name of the default database that contains audit, schedule and other data that you create.

PostgreSQL starts successfully only if your computer has allocated enough memory. Go to <http://www.postgresql.org/docs/devel/static/kernel-resources.html> and follow the instructions there on how to adjust the memory settings on your computer.



## Install PDI

---



After you have [prepared your environment](#), use the Wizard to install Pentaho Data Integration. You only need to perform two tasks.

- Install PDI with the Wizard - Run the installation wizard to install PDI.
- Verify Installation - Review the directory structure to ensure that PDI was installed correctly.

If you need to uninstall the software, information on how to do that appears below the installation tasks.

### Install PDI with the Wizard

Use the wizard to install PDI. The wizard installs PDI software, the Tomcat web application server, and PostgreSQL, which is the default database that houses the DI Repository. The DI Repository holds audit, scheduling, and solution content data.

Installation instructions are the same for the Windows, Linux, and Mac platforms.

To install the software, complete these steps.

1. Make sure that you are logged into the computer on which you want to install PDI. You should be logged on with an account that allows you to install software. On Windows, this is typically an account that has administrator privileges. On Linux, [use the pentaho account that you created during the Prepare Environment step](#).
2. Download the correct installation wizard's file for your operating system.
  - **Windows:** `pdi-5.3.0-x64.exe`
  - **Linux:** `pdi-5.3.0-x64.bin`
  - **Mac:** `pdi-5.3.0-x64.app.tar.gz`
3. If you are using Linux, make sure that you have the appropriate permission to execute the file. Open a **Terminal** window, navigate to the directory where you just downloaded the file, then enter the following at the prompt.

```
chmod a+x ./pdi-5.3.0-x64.bin
```

4. Do one of these things to launch the installation wizard.
  - **If you are on Windows:** Open **Windows Explorer** and navigate to the installation file you just downloaded. Double-click the `pdi-5.3.0-x64.exe` file to launch it.

- **If you are on Mac:** Unpack the `pdi-5.3.0-x64.app.tar.gz` archive, then double click the `pdi-5.0.0-x64.app` file to start installation.
- **If you are on Linux and your computer *can* display graphics:** Open a **Terminal** window and navigate to the installation file you just downloaded. Enter this command at the prompt, then press **Enter**.

```
./pdi-5.3.0-x64.bin
```

- **If you are on Linux and your computer *cannot* display graphics:** Open a **Terminal** window and navigate to the installation file you just downloaded. At the prompt enter *one* of these commands, then press **Enter**. Enter this command if you have **GTK libraries** on your system and want to use the **GTK Toolkit**:

```
./pdi-5.3.0-x64.bin --mode text
```

Enter this command if you do not have **GTK libraries** on your system or if you want to use **OpenMotif** instead:

```
./pdi-5.3.0-x64.bin --mode xwindow
```

5. The Pentaho Wizard splash screen appears, then the **Welcome** window. Click **Next**.
6. Read the license agreement. To continue, select **I accept the license agreement**, then click **Next**.
7. When the **Installation Folder** window appears, do one of these things.

#### NOTE:

You cannot install into a directory that already exists. If you try, you will get an error message. If you are using Linux, we do not recommend that you install the software in the `/opt/` directory. The complicated permissions requirements and general isolation from the rest of the system can cause unintended issues to arise if the software is installed there.

- a. Accept the default directory and click **Next** to continue.
- b. Enter a different directory by entering the path in the text box or click **Browse** to navigate to the place where you want PDI to be installed. Click **Next** to continue.
- c. When the **PostgreSQL postgres user password** window appears, enter the password you want to assign to the PostgreSQL database's admin user. Do not use these characters in the password because the Wizard cannot process them:

```
' " & < > /
```

- d. Click **Next** to continue. The **Ready to Install** window appears. Click the **Next** button to install the software.
- e. The **Generating Evaluation Licenses** window appears as evaluation licenses are installed. If they are not able to install, a window appears that shows the reasons the evaluation licenses were not installed.
- f. The **Installation in Progress** window appears. Make sure the software installs completely before moving on to the next step.

#### NOTE:

An error occurs if the DI Server port numbers (**9092**, **9080**, **9443**, and **5432**) are not available on your system and if the additional ports that the wizard has tried are not available. If the wizard cannot find an available port, you are prompted to enter port numbers you want to use.

8. The **Installation Complete** window appears. Click the **Open Pentaho Data Integration and Begin in Instaview** options to view **Spoon**. Click **Finish** to end the process.
9. **Spoon** and **Instaview** are launched.

**NOTE:**

If you are using Linux, the following error can occur when Spoon or

Instaview launches: *java.lang.UnsatisfiedLinkError: Could not load SWT library. Reasons: no swt-mozilla-gtk in java.library.path*. If this happens, you need to install **libwebkitgtk-1.0** on your system. You can use the command `sudo apt-get install libwebkitgtk-1.0-0` to install the library.

10. If you have trouble with the installer, check the [Troubleshooting](#) documentation.

**NOTE:**

Once the installation wizard is complete, the DI Server are running with the default parameters. Terminal windows might remain open after installation. If you close them, the DI Servers stop. On Windows, the DI Server are installed as system services, meaning that they start when you start your computer, and they stop when you shut down your computer. Spoon and other tools must be manually run. In the next step of these instructions there are links to documentation that explains how to start and stop servers manually.

## Verify Installation

To verify that the installation occurred properly, review the software directory structure.

1. Open a file explorer window like **Windows Explorer** or a command line tool such as the **Command Prompt** or **Terminal** windows.
2. Navigate to the directory where you installed PDI, then compare this directory structure with yours.

```
pentaho/  
pentaho/design-tools/  
pentaho/design-tools/data-integration/  
pentaho/documentation/  
pentaho/java/  
pentaho/jdbc-distribution  
pentaho/license-installer/  
pentaho/licenses/  
pentaho/monetdb/  
pentaho/postgresql/  
pentaho/scripts/  
pentaho/server/  
pentaho/server/data-integration-server/  
pentaho/server/data-integration-server/logs/  
pentaho/server/data-integration-server/pentaho-solutions/  
pentaho/server/data-integration-server/scripts/  
pentaho/server/data-integration-server/tomcat/
```

3. Open the `installation-summary.txt` file. Make sure that the design tools and plugins that you installed are listed.
4. Review the following locations of files and directories before going to the next step.

Top-level files:

These files are found in the top-level `/pentaho` directory:

- **ctlscrip.sh:** Starts, stops, restarts, and shows the status of Pentaho services. Available on Linux only.
- **installation-summary.txt:** Contains the information from the summary screen at the end of the installation process.
- **uninstall:** A script that removes what has been installed.

Locations of specific programs

- **DI Server:** `/pentaho/server/data-integration-server/`
- **Spoon:** `/pentaho/design-tools/data-integration/`
- **License Installer:** `/pentaho/license-installer/`

Location of logs:

- **DI Server Logs:** `pentaho/server/data-integration-server/logs/`
- **Tomcat Logs for DI Server:** `pentaho/server/data-integration-server/tomcat/logs/`

Location of JDBC drivers:

- **DI Server:** `pentaho/server/data-integration-server/tomcat/lib`
- **Spoon:** `pentaho/design-tools/data-integration/lib/`

Default Port Numbers (your port numbers might differ if these ports are already used for other programs):

- **5434:** PostgreSQL Server
- **9080:** DI Server Port
- **9001:** HSQL Server Port
- **9092:** Embedded H2 Database
- **50000 or 50006:** Monet Database Port

## Uninstall Pentaho

If you need to uninstall Pentaho, complete these steps.

1. Navigate to the `pentaho` directory, and run the `uninstall` file.
2. The Uninstall Wizard is launched.
3. When prompted, indicate whether you want to delete the data files.
4. When prompted at the end of the process, restart your computer.

## Next Steps

---



Congratulations! You have just installed PDI. Now, you are ready to do these things.

- See [Set up DI Server and Tools](#) to learn how to start the DI Server and Spoon, and to see the instructions on how to install licenses.
- Work through the [Get Started with PDI](#) tutorial.

**Note:** If you have installed the DI Server and you want to migrate content from the previous version of the servers to this one, make sure that your [license keys have been installed](#), then review the [Upgrade BA and DI](#) instructions.