



Install All DI Components



This document supports Pentaho Business Analytics Suite 5.0 GA and Pentaho Data Integration 5.0 GA, documentation revision August 22, 2013, copyright © 2013 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

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

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.

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 340
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

Contents

Introduction to DI Wizard Installation.....	4
Wizard Installation Process Overview.....	5
Prepare Environment.....	6
Create User Account.....	6
Windows User Account.....	6
Linux User Account.....	6
Download Installation File.....	7
Disable Anti-Spyware and Anti-Virus Programs.....	7
Check Memory Settings for PostgreSQL.....	7
Install PDI.....	8
Install PDI with the Wizard.....	8
Verify Installation.....	9
Next Steps.....	11

Introduction to DI Wizard Installation

These instructions explain how to use the Wizard to install Pentaho Data Integration (PDI). You can use PDI 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.

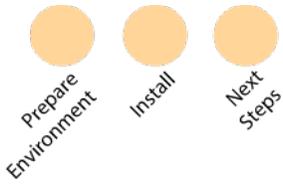
Tools

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.

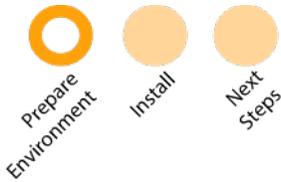
Wizard Installation Process Overview



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. During this step you perform these tasks.

- [Create User Account](#)
- [Download Installation Package](#)
- [Disable AntiSpyware or Antivirus Programs](#)

If you plan to install the software on a Mac, you also need to perform one additional step.

- [Adjust Amount of Memory Mac 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. Open a **Terminal** window on the computer that you want to install the software. If you plan to install the software on a remote computer, establish an OpenSSH session to the remote server.
2. In the **Terminal** window, log in as the **root** user by entering this command at the prompt.

```
su root
```

3. When prompted, type the password in the **Terminal** window.
4. In the **Terminal** window, create a new user account called **pentaho**, along with the `pentaho` home directory, by typing this line.

```
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.

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

```
sudo passwd pentaho
```

6. 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.

Download Installation File

Download the installation file from the Pentaho website. There are installation files for the Windows, Linux, and Mac operating systems. Download the correct installation file for your operating system. Make note of where you downloaded the file because you will need to find it again in a future step.

- **Windows:** `pdi-5.0.0-x64.exe`
- **Linux:** `pdi-5.0.0-x64.bin`
- **Mac:** `pdi-5.0.0-x64.app.tar.gz`

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, reenable them.

Check Memory Settings for PostgreSQL

If you plan to install the software on a Mac OS, you need to increase the amount of memory that it allocates for PostgreSQL. You can skip these instructions if you plan to install the software on Windows or Linux and go to the [Install PDI](#) step.

PostgreSQL is the name of the database that contains audit, schedule, and other data that you create when you use PDI. When you run the Installation Wizard, it automatically installs PostgreSQL for you. 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.

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, this is also an account that has administrator privileges, such as the [pentaho account that you created](#) during the previous step.
2. Do one of these things to launch the installation file that you [downloaded](#).
 - a) **If you are using Windows:** Open **Windows Explorer** and navigate to the installation file you just downloaded. Double-click the `pdi-5.0.0-x64.exe` file to launch it.
 - b) **If you are using Mac:** Double-click the `pdi-5.0.0-x64.app.tar.gz` installation app.
 - c) **If you are using 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-5.0.0-x64.bin
```

- d) **If you are using 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-5.0.0-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-5.0.0-x64.bin - mode xwindow
```

3. The Pentaho Wizard Splash Screen appears, then the **Welcome** window. Click **Next**.
4. Read the license agreement. To continue, accept the license agreement, then click **Next**.
5. When the **Installation Folder** window appears, do one of these things.
 - 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.



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.

6. 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: ' " & < > \ ` . Consider making note of the password you choose and putting it in a secure place. Click **Next** to continue.
7. The **Ready to Install** window appears. Click the **Next** button to install the software. 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.
8. 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.

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

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. By default, 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
../server
../../data-integration-server
../../../../data
../../../../licenses
../../../../pentaho-solutions
../../../../tomcat
../design-tools
../data-integration
../common
../documentation
../java
../licenses
../postgresql
../scripts

```

3. If you have trouble with the installer, check the [Troubleshooting](#) documentation.
4. Take note of these files and paths before you move to the next section of these instructions.

Top-level files:

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

- **ctlscript:** (**.sh** or **.bat**) Starts, stops, restarts, and shows the status of Pentaho services.
- **installation-summary.txt:** Contains the information from the summary screen at the end of the installation process.
- **pentaho.log:** The log file for the Pentaho Web application.
- **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/`

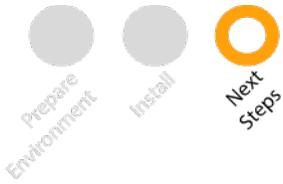
- **License Installer:** `/pentaho/license-installer/`

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

Note: The license installer can also be found in the data-integration directory. All of these license installers perform the same functions, but are available in separate places to account for instances where only specific parts of the Pentaho suite are installed to a particular server or workstation. It does not matter which one you use. You can also install licenses through the User Console if you have already installed the BA Server. See the [next step](#) for more information on licenses.

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 DI](#) 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 [DI Upgrade](#) instructions.