Installing Pentaho Data Integration CE

Installing Pentaho Data Integration (PDI) client tool is a simple three step process. Make sure you have the following installed before installing PDI:

Prerequisites

- Install a 64 bit Java SE in the workstation or the laptop where you want to run Pentaho Data Integration (PDI). Any Java SE version between 11 to 18 should work.

- For Linux/Ubuntu installations, libwebkitgtk-1.0-0 also needs to be installed for PDI to function properly. Please check the Hitahi Vantara Pentaho Community page https://community.hitachivantara.com/discussion/libwebkitgtk-10-0-on-ubuntu-2204-lts for more information.

Installation Steps

- **1. Download the Zip**

Download the appropriate version of the compressed PDI artifacts for your operating system from Hitachi Vantara Pentaho Community page https://www.hitachivantara.com/en-us/products/lumada-dataops/data-integration-analytics/pentaho-community-edition.html

The artifact version and build are named as:

pdi-ce-<version>-<build>.zip

- **2. Extract the Zip**

Expand the PDI artifact zip in your workstation where you want to install it

OS	Unzip Instruction
Windows	Use Windows Explorer Unzip feature, available from Windows 8 and higher to unzip the *pdi-ce-<version>-<build>.zip*. Locate the file in the Windows explorer and then press and hold (or right-click), select Extract All.
Linux/Ubuntu/ Mac	Extract the downloaded file *pdi-ce-<version>-<build>.zip* using an Unzip tool to the directory location that you want Pentaho data integration installed. Typically, most users use /usr/local folder as it has the write permission for the user.

- **3. Start PDI**

OS	Unzip Instruction
Windows	Navigate to the folder where you have unzipped the PDI and locate the spoon.bat. Double click on this file to start the PDI.
Linux/Ubuntu/ Mac	Open a terminal window, navigate to the folder where you have extracted the PDI and then start it by typing the following in the terminal window. ***./spoon.sh***

Installing Pentaho Hadoop addon

Applicable for all OS (Windows/Linux/Ubuntu/Mac)

- Extract the downloaded file ***pdi-ce-<version>-<build>-hadoop-addon.zip*** using an unzip tool to a temporary folder.
    NOTE for Mac users:
    - Safari automatically unzips any downloaded file in the *Downloads* folder, without giving you the option to select where to unzip. Uncheck *Open "safe" files after downloading* in the Safari Preferences/General tab.
    - Open a command window in the *Downloads* folder and use unzip to extract the file to a temporary folder
      unzip pdi-ce-<version>-<build>-hadoop-addon.zip -d ./<temp-folder>
    - The upzip will create the "***data-integration***" folder inside the temporary folder and all Hadoop addon related files and folders will be inside it.

- Review the *HADOOP_ADDON_README* file in the unzipped "*data-integration*" folder for more information.

- Copy all following files and folders from the temporary Hadoop addon "data-integration" folder to the "data-integration" folder of the already installed Pentaho Data Integration CE application. Please make sure to copy contents while maintaining their folder hierarchy. For example:
    - All .jar files in ***<temp-folder>/data-integration/lib*** should be copied in Pentaho Data Integration ***data-integration/lib*** folder.
    - All folders in ***<temp-folder>/data-integration/plugins*** should be copied in Pentaho Data Integration ***data-integration/plugins*** folder.
    - All files such as uninstall-hadoop-addon.sh etc. should be copied from ***<temp-folder>/data-integration*** to Pentaho Data Integration ***data-integration folder***.

- Start Pentaho Data Integration to load the Hadoop addon.