



FREE eBook

LEARNING pentaho

Free unaffiliated eBook created from
Stack Overflow contributors.

#pentaho

Table of Contents

About.....	1
Chapter 1: Getting started with pentaho.....	2
Remarks.....	2
Examples.....	2
Installation or Setup of Pentaho Data Integration.....	2
Installation steps:.....	2
Chapter 2: Pentaho Data Integration.....	3
Examples.....	3
Hello World in Pentaho Data Integration.....	3
Create a Transformation:.....	3
Configuring the CSV file input Step.....	3
Configuring the Modified JavaScript Value Step.....	4
Configuring the XML Output Step.....	5
RUN.....	5
Installation of Pentaho DI.....	5
Installation steps:.....	5
Credits.....	7

About

You can share this PDF with anyone you feel could benefit from it, downloaded the latest version from: [pentaho](#)

It is an unofficial and free pentaho ebook created for educational purposes. All the content is extracted from [Stack Overflow Documentation](#), which is written by many hardworking individuals at Stack Overflow. It is neither affiliated with Stack Overflow nor official pentaho.

The content is released under Creative Commons BY-SA, and the list of contributors to each chapter are provided in the credits section at the end of this book. Images may be copyright of their respective owners unless otherwise specified. All trademarks and registered trademarks are the property of their respective company owners.

Use the content presented in this book at your own risk; it is not guaranteed to be correct nor accurate, please send your feedback and corrections to info@zzzprojects.com

Chapter 1: Getting started with pentaho

Remarks

This section provides an overview of what pentaho is, and why a developer might want to use it.

It should also mention any large subjects within pentaho, and link out to the related topics. Since the Documentation for pentaho is new, you may need to create initial versions of those related topics.

Examples

Installation or Setup of Pentaho Data Integration

Pentaho Data Integration comes in two varieties:

- Community Edition (CE) - Free version for developers
- Enterprise Edition (EE) - Paid version for enterprise use

Installation steps:

1. You can download Pentaho Data Integration **Community Edition** from [Sourceforge.net](https://sourceforge.net).

For the **Enterprise Edition** : [Download a 30-Day Trial](#)

2. **Prerequisites** - PDI V7 requires the Oracle Java Runtime Environment (JRE) version 8.

3. **PDI does not require installation.** Simply unpack the zip file into a folder of your choice. On Unix-like operating systems, you may need to make the shell scripts executable by using the chmod command:

```
cd data-integration
chmod +x *.sh
```

4. **Running:** PDI comes with a graphical user interface called Spoon, command-line scripts (Kitchen, Pan) to execute transformations and jobs, and other utilities.

Read [Getting started with pentaho online](https://riptutorial.com/pentaho/topic/4177/getting-started-with-pentaho): <https://riptutorial.com/pentaho/topic/4177/getting-started-with-pentaho>

Chapter 2: Pentaho Data Integration

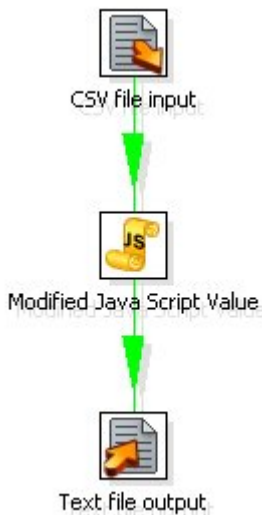
Examples

Hello World in Pentaho Data Integration

Lets create a simple transformation to convert a CSV into an XML file.

Our Transformation has to do the following:

- Read the CSV file
- Build the greetings message
- Save the greetings in the XML file



Create a Transformation:

Here's how to start the Transformation:

1. To the left of the workspace is the Steps Palette. Select the Input category.
2. Drag the CSV file onto the workspace on the right.
3. Select the Scripting category.
4. Drag the Modified JavaScript Value icon to the workspace.
5. Select the Output category.
6. Drag the XML Output icon to the workspace.

Now link the CSV file input with the Modified Java Script Value by creating a Hop:

1. Select the first Step.
2. Hold the Shift key and drag the icon onto the second Step.
3. Link the Modified Java Script Value with the XML Output via this same process.

Configuring the CSV file input Step

1. Double-click on the CSV file input Step. The configuration window for the step will appear. Here you'll indicate the file location, file format (e.g. delimiters, enclosure characters, etc.) and column metadata (e.g. column name, data type, etc)
2. Change the step name with one that is more representative of this Step's function. In this case, type in name list.
3. For the Filename field, click Browse and select the input file.
4. Click Get Fields to add the list of column names of the input file to the grid. By default, the Step assumes that the file has headers (the Header row present checkbox is checked).
5. The grid has now the names of the columns of your file. 6. Click Preview to ensure that the file will be read as expected. A window showing data from the file will appear. Click **OK** to finish defining the Step **CSV file input**.

Configuring the Modified JavaScript Value Step

1. Double-click on the Modified JavaScript Value Step.
2. The Step configuration window will appear. This is different from the previous Step config window in that it allows you to write JavaScript code. You will use it to build the message "Hello, " concatenated with each of the names.
3. Name this Step **Greetings**.
4. The main area of the configuration window is for coding. To the left, there is a tree with a set of available functions that you can use in the code. In particular, the last two branches have the input and output fields, ready to use in the code. In this example there are two fields: last_name and name. Write the following code:

```
var msg = 'Hello, ' + name + "!";
```

5. At the bottom you can type any variable created in the code. In this case, you have created a variable named msg. Since you need to send this message to the output file, you have to write the variable name in the grid.
6. Click OK to finish configuring the Modified Java Script Value step.
7. Select the Step you just configured. In order to check that the new field will leave this Step, you will now see the Input and Output Fields. Input Fields are the data columns that reach a Step. Output Fields are the data columns that leave a Step. There are Steps that simply transform the input data. In this case, the input and output fields are usually the same. There are Steps, however, that add fields to the Output - Calculator, for example. There are other Steps that filter or combine data causing that the Output has less fields that the Input - Group

by, for example.

8. Right-click the Step to bring up a context menu.
9. Select Show Input Fields. You'll see that the Input Fields are last_name and name, which come from the CSV file input Step.
10. Select Show Output Fields. You'll see that not only do you have the existing fields, but also the new msg field.

Configuring the XML Output Step

1. Double-click the XML Output Step. The configuration window for this kind of Step will appear. Here you're going to set the name and location of the output file, and establish which of the fields you want to include. You may include all or some of the fields that reach the Step.
2. Name the Step File: **Greetings**.
3. In the File box write:

```
${Internal.Transformation.Filename.Directory}/Hello.xml
```

4. Click Get Fields to fill the grid with the three input fields.
5. Save the Transformation again.

RUN

Click on the **RUN** button on the menu bar and Launch the transformation.

You may also create a Job which may be used to schedule multiple transformations and then run it.

Alternatively, Pan allows you to execute Transformations from a terminal window. The script is Pan.bat on Windows, or pan.sh on other platforms, and it's located in the installation folder.

```
Pan /file <Tutorial_folder_path>/Hello.ktr /norep
```

Installation of Pentaho DI

Pentaho Data Integration comes in two varieties:

- Community Edition (CE) - Free version for developers
- Enterprise Edition (EE) - Paid version for enterprise use

Installation steps:

1. You can download Pentaho Data Integration **Community Edition** from [Sourceforge.net](https://sourceforge.net).

For the **Enterprise Edition** : [Download a 30-Day Trial](#)

2. **Prerequisites** - PDI V7 requires the Oracle Java Runtime Environment (JRE) version 8.

3. **PDI does not require installation.** Simply unpack the zip file into a folder of your choice. On Unix-like operating systems, you may need to make the shell scripts executable by using the `chmod` command:

```
cd data-integration
chmod +x *.sh
```

4. **Running:** PDI comes with a graphical user interface called Spoon__,_ command-line scripts (Kitchen, Pan) to execute transformations and jobs, and other utilities.

Read Pentaho Data Integration online: <https://riptutorial.com/pentaho/topic/5523/pentaho-data-integration>

Credits

S. No	Chapters	Contributors
1	Getting started with pentaho	AlainD , Ani Menon , Community , jipipayo
2	Pentaho Data Integration	AlainD , Ani Menon