PATC: Introduction to Big Data Analytics
(Big Data Management - Hands on)
Intellij IDEA Manual

This manual briefly describes the basic usage functionalities of IntelliJ IDEA, which you will need in order to run the MapReduce code in this hands-on session. First, download IntelliJ IDEA Community edition `https://www.jetbrains.com/idea/download/`, and install it following the instructions specific to your operating system.
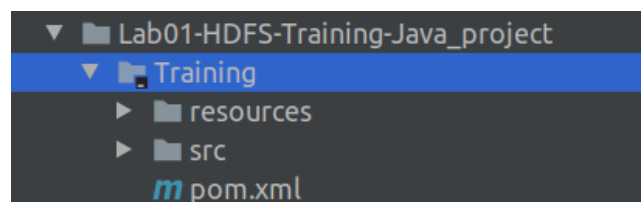
Important note: Although we provide IntelliJ instructions here, the provided project works as well with other IDEs that support Maven functionalities (e.g., Eclipse).

Loading a project

After you unzip the Java project zip we provided you with has the content is as follows:

- `resources` – a directory containing files you will need during the session (e.g., external schemas, datasets, etc.).
- `src` – a directory containing the source code
- `pom.xml` – a configuration file for Apache Maven, which contains the dependencies of the project. Maven will automatically download all external libraries that the project requires.

To load a project, click `Open` and select the root folder of the project as follows (note the root folder has an special icon):
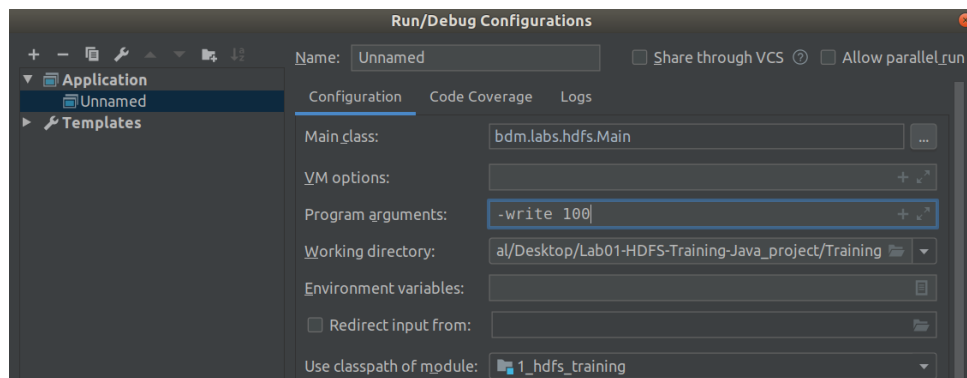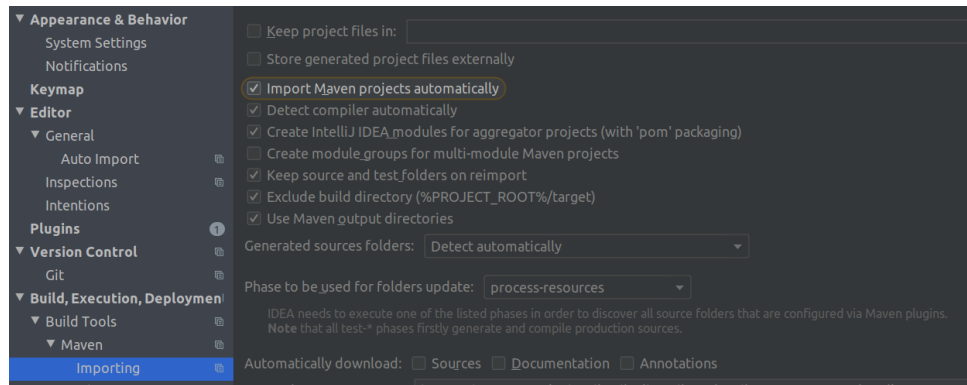

Intellij IDEA will load the project. Make sure the *Import Maven projects automatically* option is set in case dependencies are not downloaded automatically (you can find this menu in `File` / `Settings`).

Running programs

In order to run the different exercises you will need to create *Run Configurations*, which will drive the execution of the `Main` method. To create one, on the top-right corner, click *Add Configuration*. Click the `+` button, and select *Application*.

In the main class you need to select the one corresponding to the project at-hand. In *Program arguments* you can specify specific parameters required by the exercises.

Specific configuration for MapReduce

Running MapReduce jobs require special configuration. You should apply the following changes to the provided Java project:

1. In *pom.xml*, uncomment the `toolsjar` property and specify the correct path for Java's JDK in your system. Also, uncomment the `jdk.tools` dependency at the end of the file.

2. In *MainLocal.java*, specify in the variable `HADOOP_COMMON_PATH` the correct path to the `winutils` directory in the provided project. Also, uncomment the first line of the `main` method.

3. The directory `winutils/bin` must be specified in the variable `PATH`.

   From IntelliJ IDEA, you can modify the `PATH` variable from a run configuration as follows: