

Big Data Management (BDMA & MIRI Masters)

IntelliJ IDEA Manual

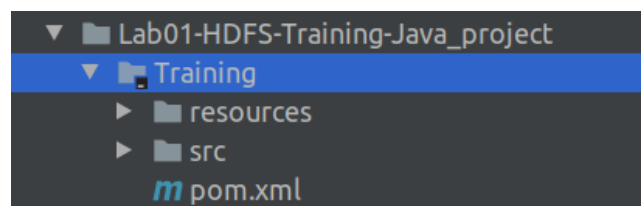
IntelliJ IDEA is an IDE to develop and run Java programs which is installed in the Windows 10 image at FIB's laboratories. This manual briefly describes its basic usage functionalities you will need to run code in the lab sessions.

Loading a project

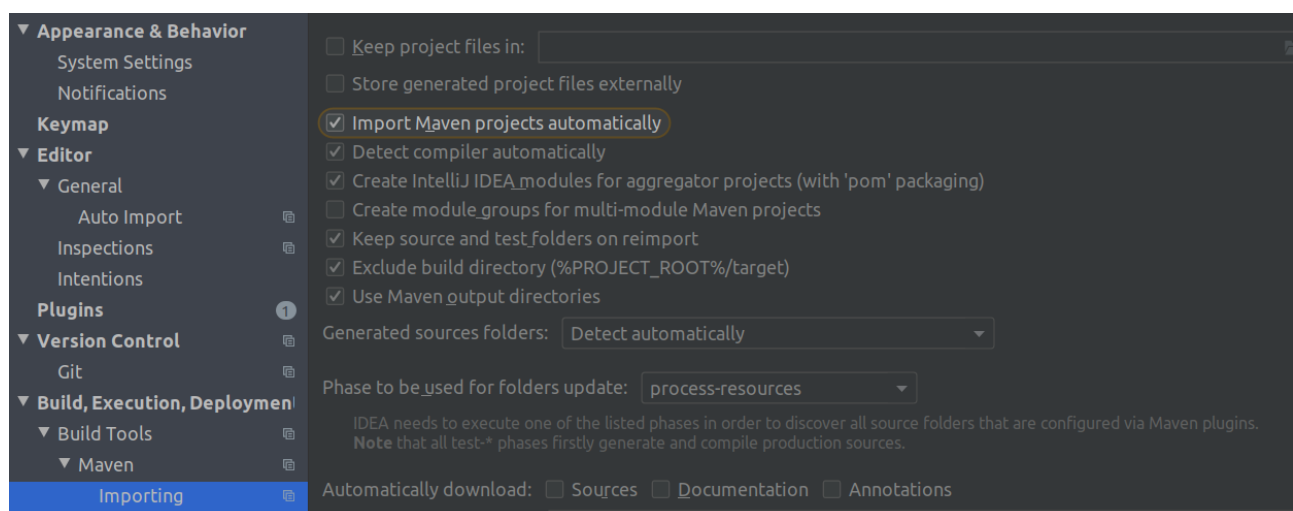
All Java projects, either training or session, have the same structure. After you unzip the file, the content is as follows:

- **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):



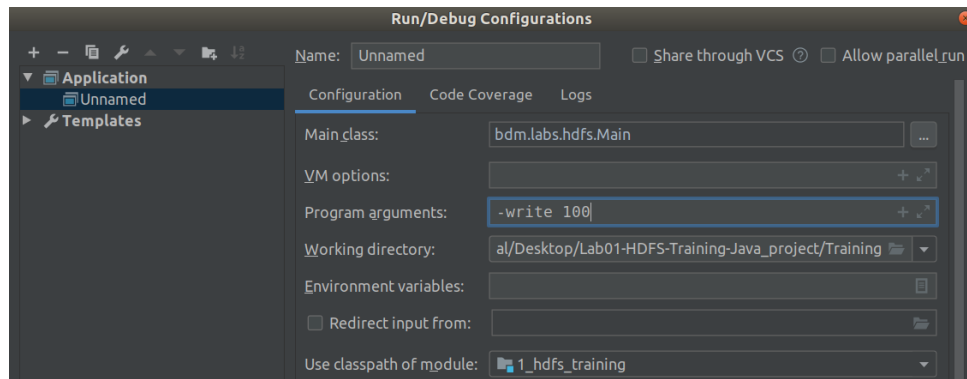
IntelliJ IDEA will load the project. Make sure the *Import Maven projects automatically* option is set in case dependencies are not downloaded automatically (you can find this menu in **File / Settings**).



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 and Spark

Running MapReduce and Spark 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 `%HADOOP_HOME%/bin` must be specified in the variable `PATH`.

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

