Walid Belal
Sept 2020

Summary

This tutorial shows how to install hadoop and IDE (Intellij Idea) on your laptop.

Install Hadoop

You will be deploying a single Hadoop installation on your laptop. It is the easiest to configure and it uses your file system instead of HDFS. This type of deployment is suitable for test environments where you need to test code without getting to the complications of configuring Hadoop Cluster.

1-Download Hadoop from https://hadoop.apache.org/releases.html

Choose the binary release for version 3.1.3

2- Unpack the tar file into a folder on your laptop. For example /users/<your user>/Apps

Step 3, 4 and, 5 are only for windows clients
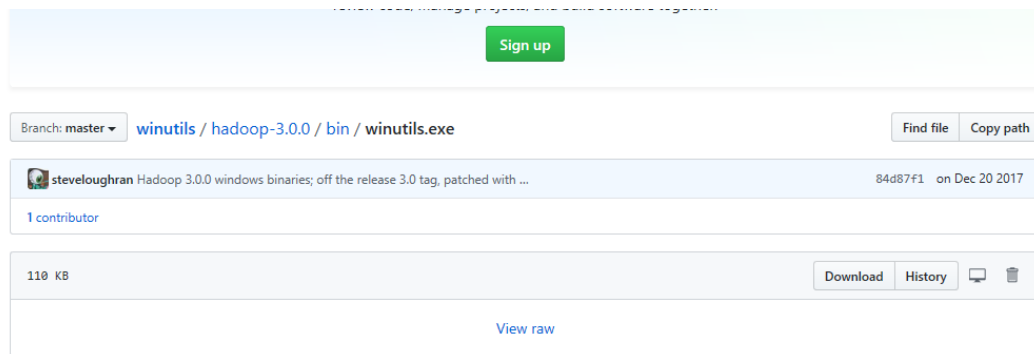
3- download winutil.exe from https://github.com/steveloughran/winutils

Choose Hadoop 3

Download winutil.exe to the bin folder of Hadoop. In the case of this document the folder is C:\Users\belalwa\Apps\hadoop-3.1.3\bin

Branch: master ▾   winutils / hadoop-3.0.0 / bin / winutils.exe   Find file   Copy path

steveloughran Hadoop 3.0.0 windows binaries; off the release 3.0 tag, patched with ...   84d87f1   on Dec 20 2017
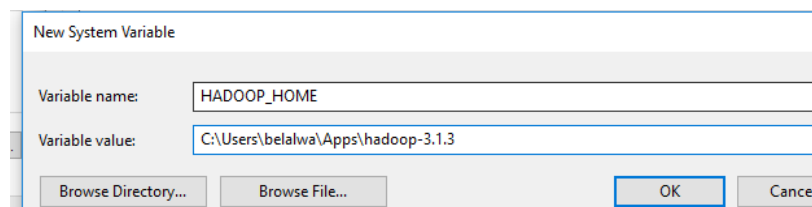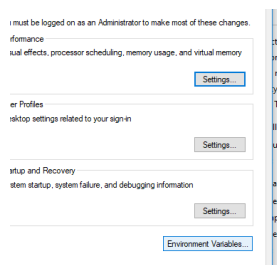
1 contributor

110 KB   Download   History   🖥 🗑

View raw
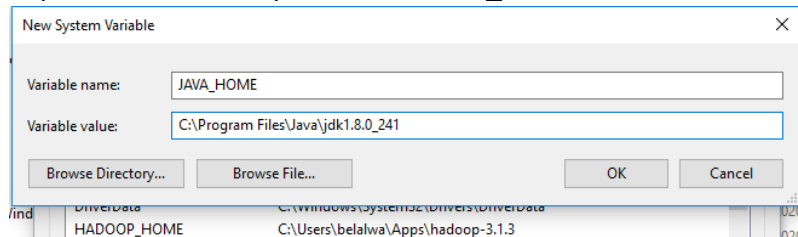
4- Add Hadoop_Home and JAVA_HOME Environment variables
Go to Advanced System Settings from control panel > system and sectury > System
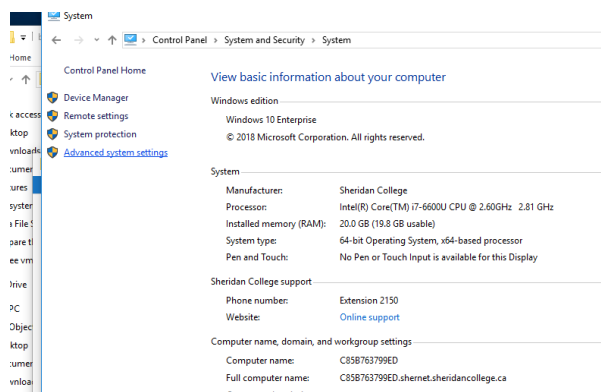
Choose environment variables and create a new system variable called HADOOP_HOME and the value is the path of the directory directly above bin folder

Repeat the same step to create JAVA_HOME variable

New System Variable   ×

Variable name: JAVA_HOME

Variable value: C:\Program Files\Java\jdk1.8.0_241

Browse Directory...   Browse File...   OK   Cancel

~ind   DriverData   C:\Windows\System32\Drivers\DriverData   020-
          HADOOP_HOME   C:\Users\belalwa\Apps\hadoop-3.1.3   020-

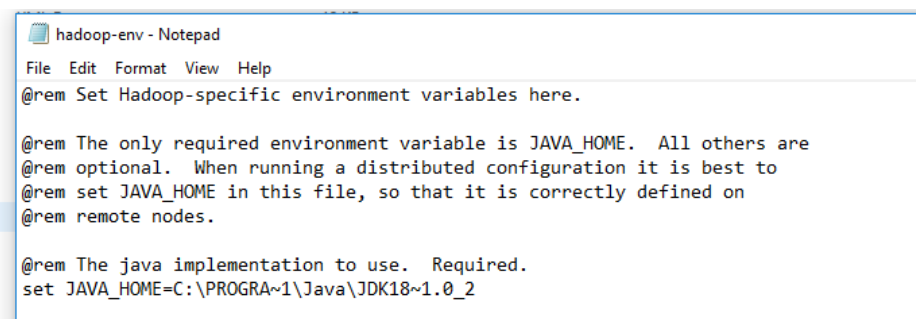The value should be the home folder of your jdk 8

5- First perform following task to create short name;-

Most probable you will need to use the short name because Hadoop will give an
error if there is a space in java home path. Below are the instructions on how to
get the short name
 cd to the java home folder then type " for %I in (.) do echo %~sI "
The returned string is the short name

```
C:\Users\belalwa>cd C:\Program Files\Java\jdk1.8.0_241

C:\Program Files\Java\jdk1.8.0_241>for %I in (.) do echo %~sI

C:\Program Files\Java\jdk1.8.0_241>echo C:\PROGRA~1\Java\JDK18~1.0_2
C:\PROGRA~1\Java\JDK18~1.0_2
```

Copy short name provided above and then Open CopyHadoop-env.cmd with a text editor and
change JAVA_HOME parameter in the file to the short name of java home. (Don't change
Hadop-env.ssh file, only command file needs to be changed)

hadoop-env - Notepad

File Edit Format View Help
@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA_HOME. All others are
@rem optional. When running a distributed configuration it is best to
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.

@rem The java implementation to use. Required.
set JAVA_HOME=C:\PROGRA~1\Java\JDK18~1.0_2

You will find hadoop-env.cmd file under $HADOOP_HOME\etc\hadoop

Install IntelliJ IDEA

1- Download IntelliJ IDEA Community Edition
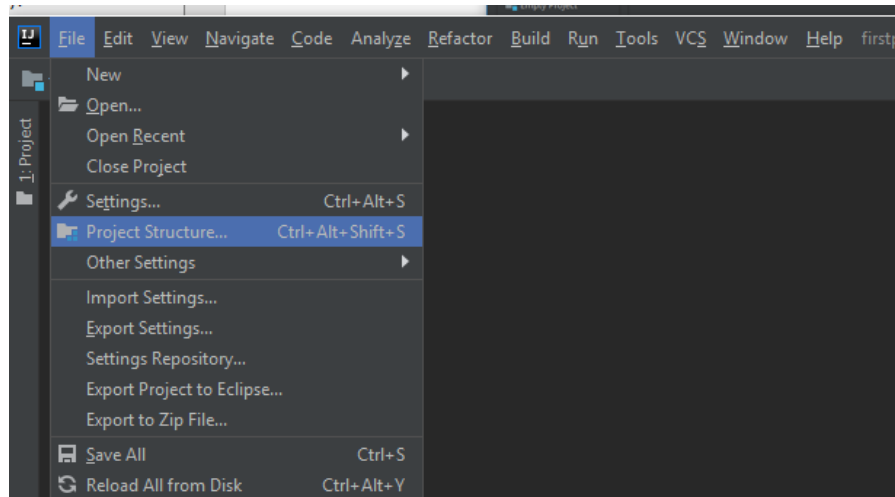https://www.jetbrains.com/idea/download/#section=windows

2- Install the software

Configure IntelliJ IDEA

1- Create a new Java project in IntelliJ. Name the project HadoopProj and store it in your class folder
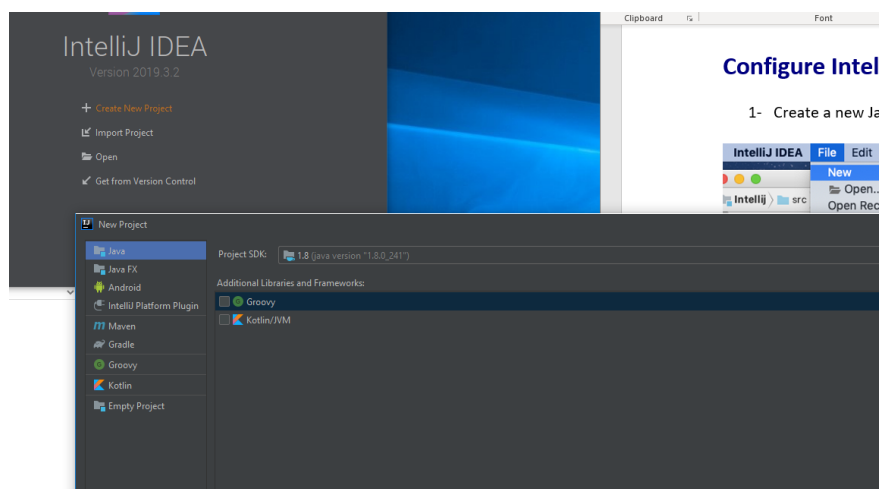
Add Hadoop module dependencies to the project

Choose Modules from the left and then Dependencies tab from the horizontal pane. Select whatever version of java sdk you have in your system. It can work with any jdk setup, need not be 1.8 version. Preferable it should be java 1.8 or above any version.
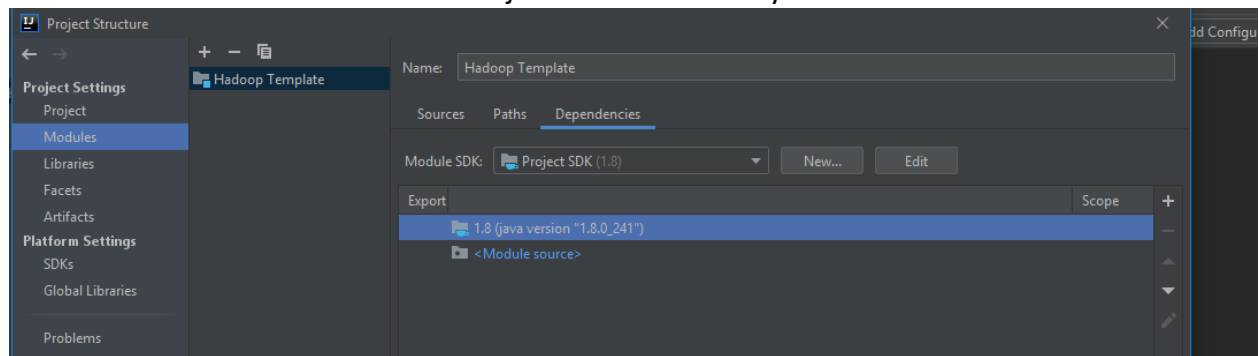
Click on the add button on the right of the window and Choose add JAR or Directory
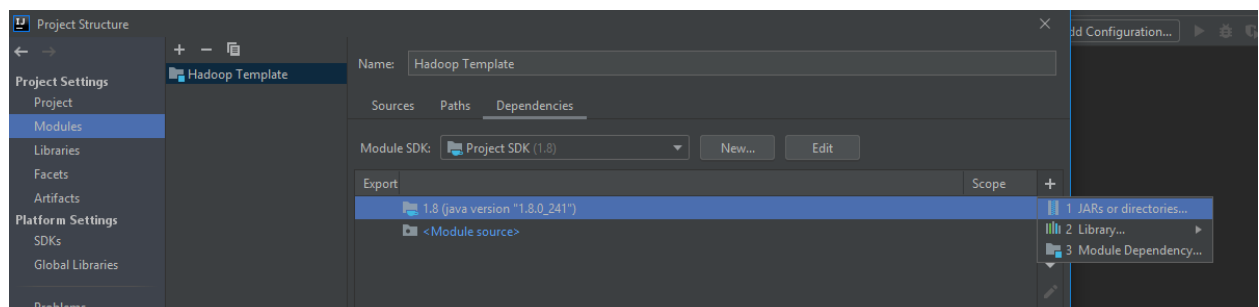
Choose all the folders under share/Hadoop folder and click open.
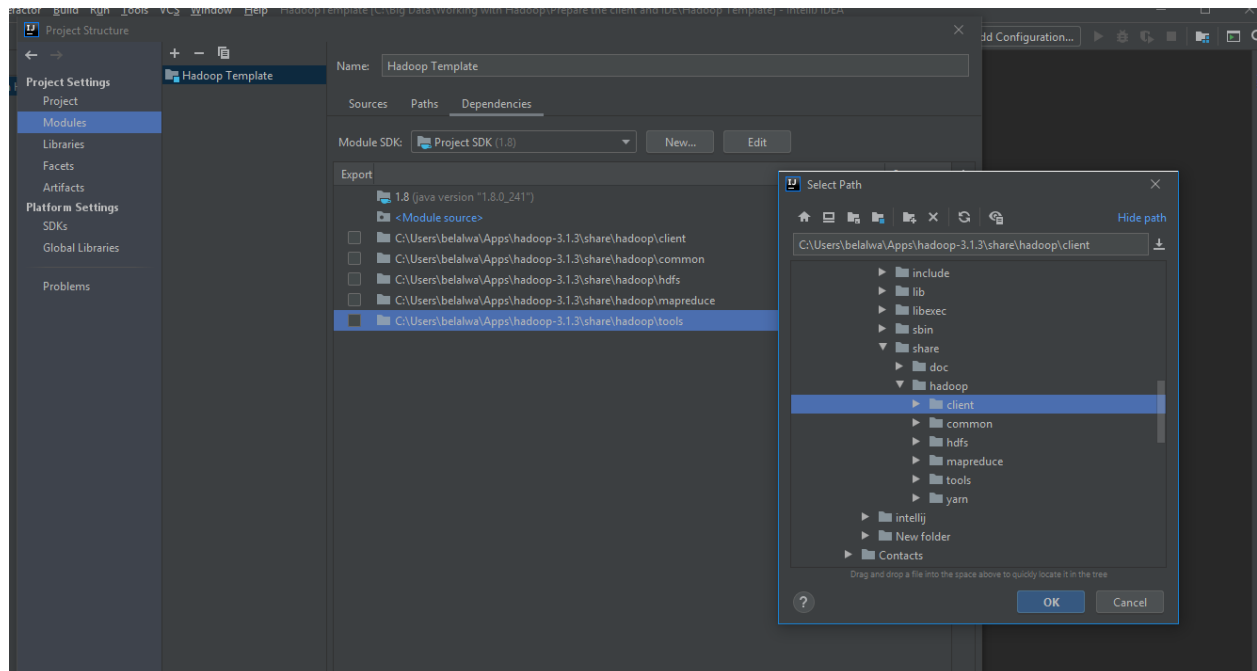
Click ok to the Project Structure Window

Save the project and close it

Now your project has all the hadoop libraries that you need to work with