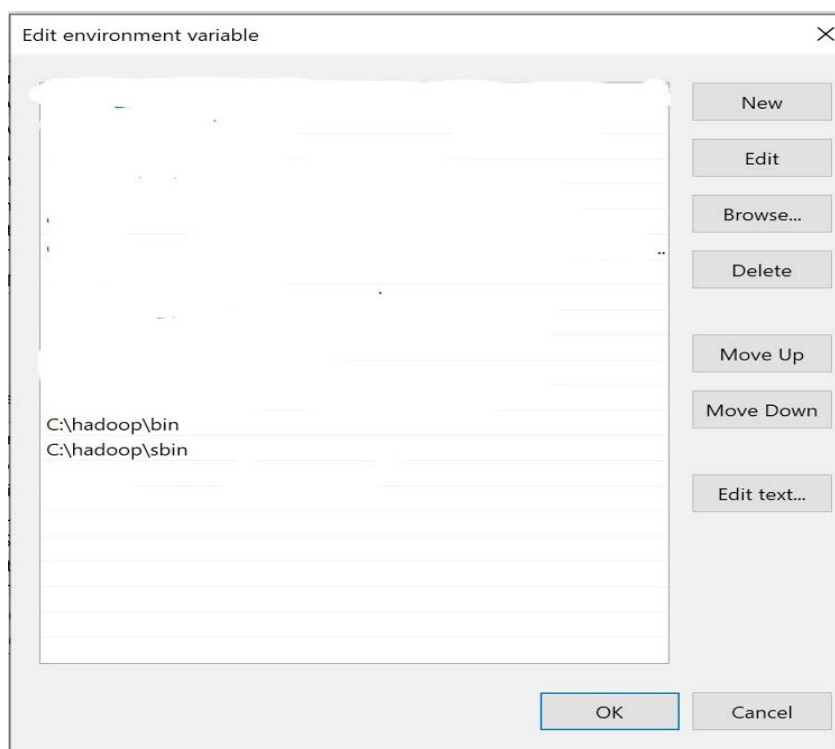


Running Map Reduce Program in Eclipse – Windows

1. Download and setup Apache Hadoop. Download tar.gz file for Hadoop ver 3.2.1 from link: <https://hadoop.apache.org/release/3.2.1.html>
2. Extract tar file at an appropriate location on your system. Say **C:\hadoop**
3. Add the location of the Hadoop directory to your system environment variables as HADOOP_HOME. Eg. if you have extracted Hadoop at **C:/hadoop** then add **HADOOP_HOME=C:/hadoop** as a system environment variable.



4. Add Hadoop bin to your system path variable. For Eg if Hadoop is present at **C:/hadoop**, add **C:/hadoop/bin** to the system path.



5. Download winutil binary files from link below Put them all in **hadoop/bin** folder if they are already not present in it.
<https://github.com/cdarlint/winutils/tree/master/hadoop-3.2.1/bin>

10. Provide Project Details: Inputs (1) Choose Create a simple project (skip archetype selection) (2) specify project location and click <next>

New Maven Project

Select project name and location

☒ Create a simple project (skip archetype selection)

☐ Use default Workspace location

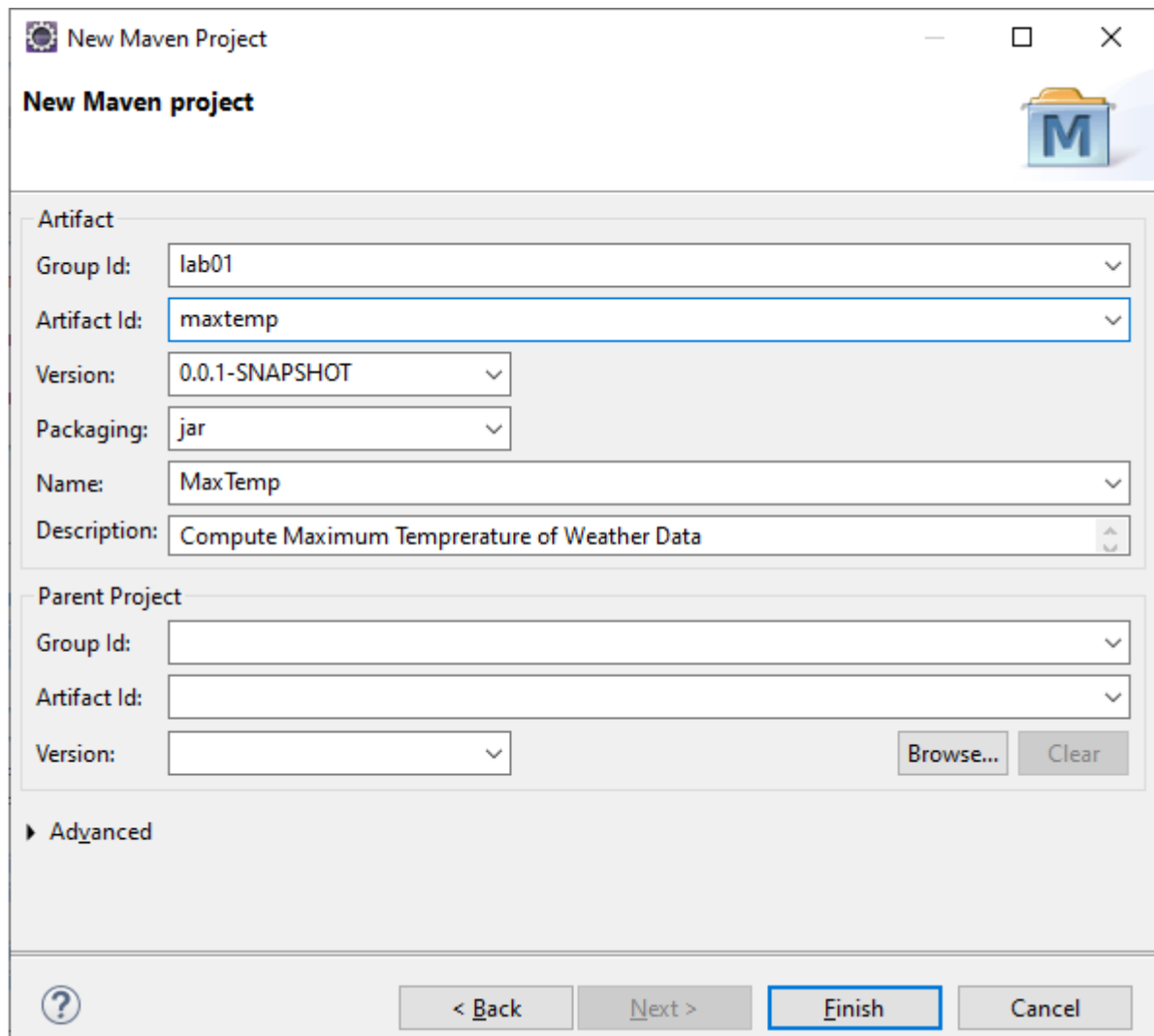
Location: Browse...

☐ Add project(s) to working set

Working set: More...

► Advanced

11. Provide Maven Project Details: Inputs (1) Group Id (2) Artifact Id (3) Name (4) Description. Leave other inputs blank or default value and click [Finish]



New Maven Project

New Maven project

Artifact

Group Id:

Artifact Id:

Version:

Packaging:

Name:

Description:

Parent Project

Group Id:

Artifact Id:

Version:

► **Advanced**

12. Add Hadoop dependencies to the project. Open pom.xml, it should like following

```
1 <project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.i
2
3   <modelVersion>4.0.0</modelVersion>
4   <groupId>lab01</groupId>
5   <artifactId>maxtemp</artifactId>
6   <version>0.0.1-SNAPSHOT</version>
7   <name>MaxTemp</name>
8   <description>Compute Maximum Temperature of Weather Data</description>
9
10  </project>
```

Add following dependencies in pom.xml (immediately above </project> tag.

```
<dependencies>
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-client</artifactId>
  <version>3.3.0</version>
</dependency>
<dependency>
  <groupId>org.slf4j</groupId>
  <artifactId>slf4j-api</artifactId>
  <version>1.7.30</version>
</dependency>
</dependencies>
```

Save the [pom.xml](#) file. It should start some downloads.

13. Create package “[lab01.maxtemp](#)”

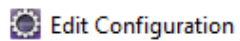
14. Add given three source files in this package

15. Change the name of the package at the top of all the three files from

“[pmj.mapr.maxtemp](#)” to “[lab01.maxtemp](#)”

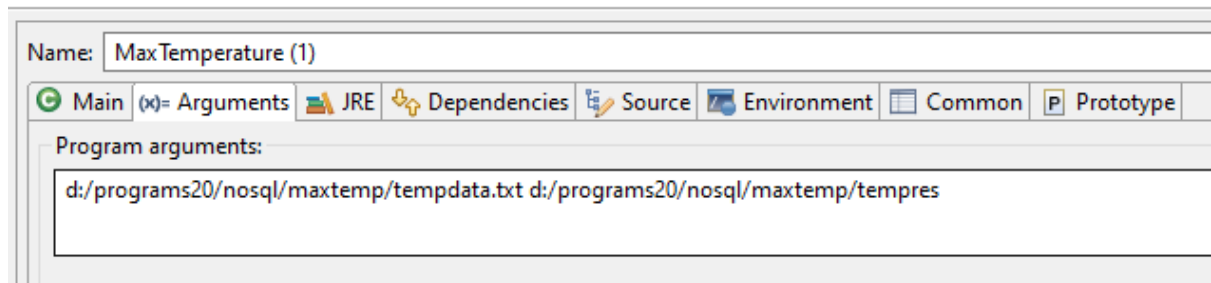
16. Create a directory a input data directory and put given data file in this.

17. Specify program parameters: Right click on “MaxTemperature.java”. Select properties. “Select Run/Debug settings”. Select this file. Add parameters.



Edit launch configuration properties

Run a Java application



18. Run: Right click on “MaxTemperature.java”. Choose “Run As” → “Java Application”