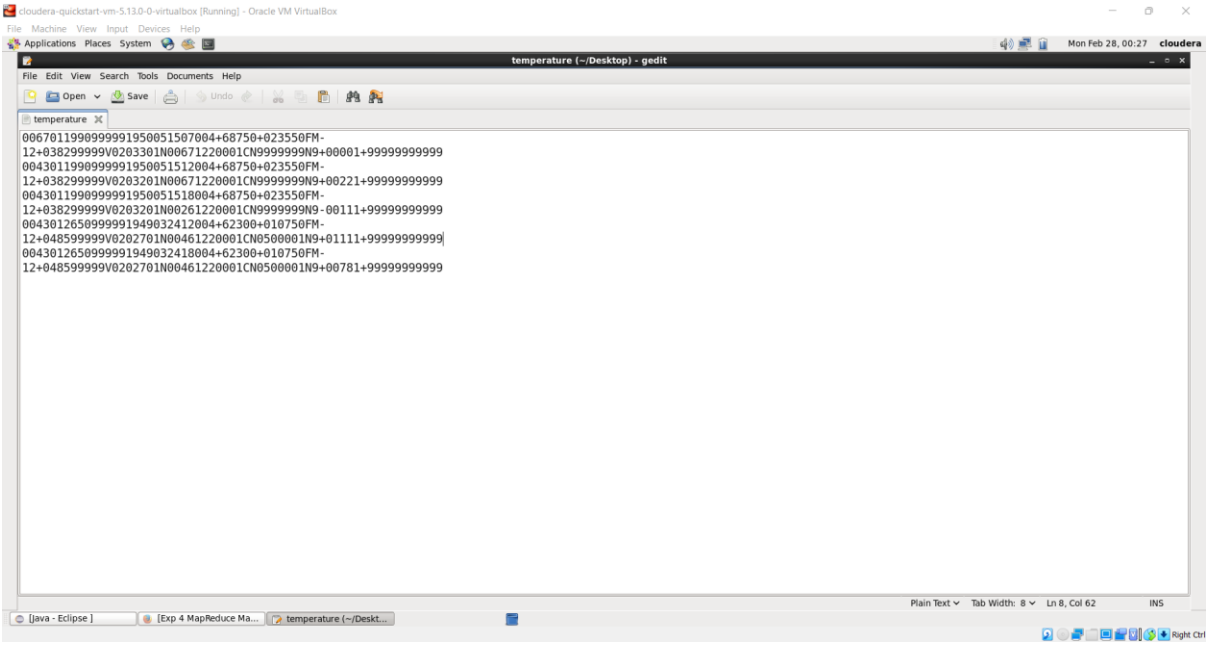
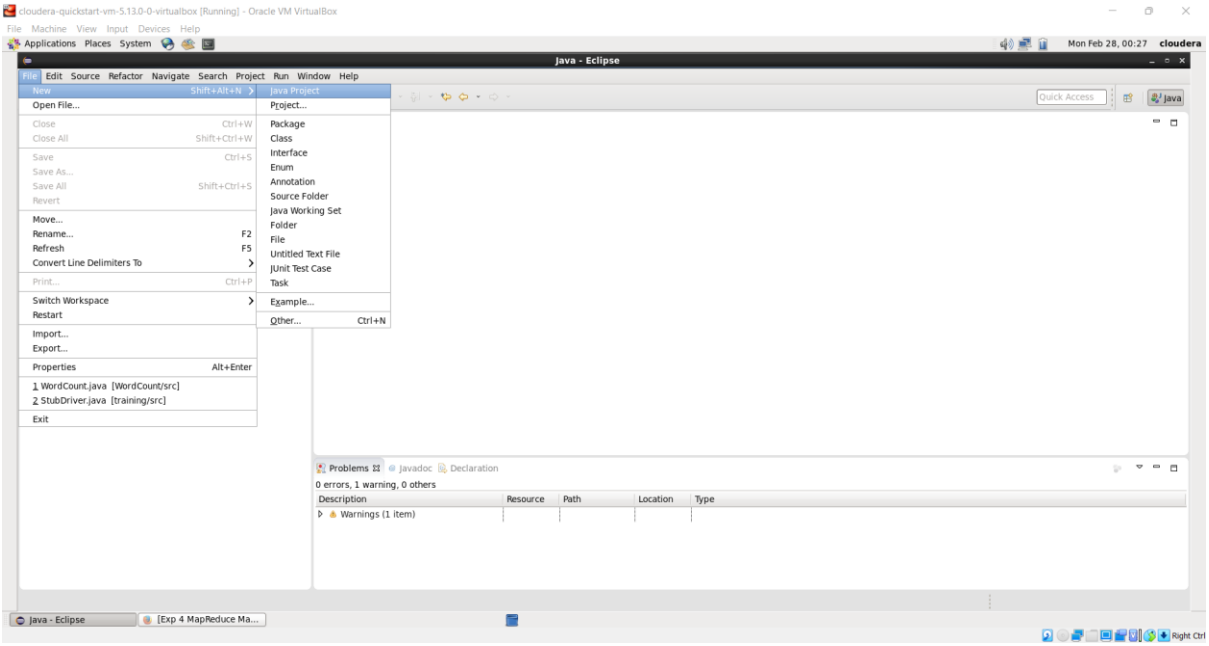
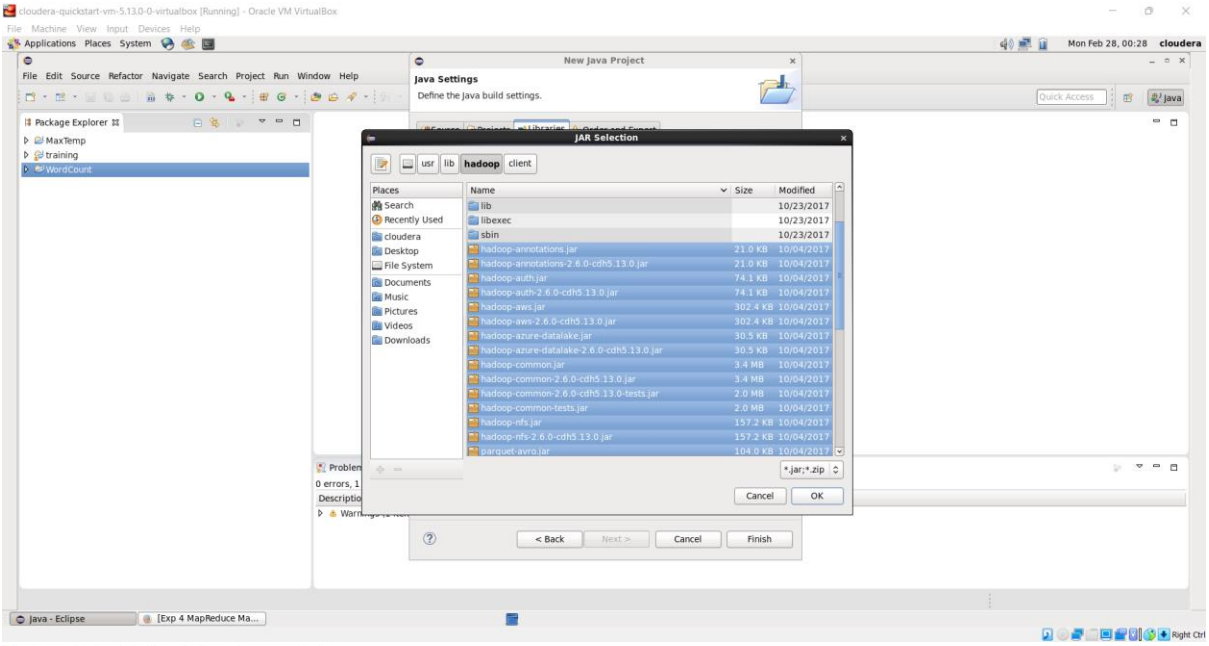
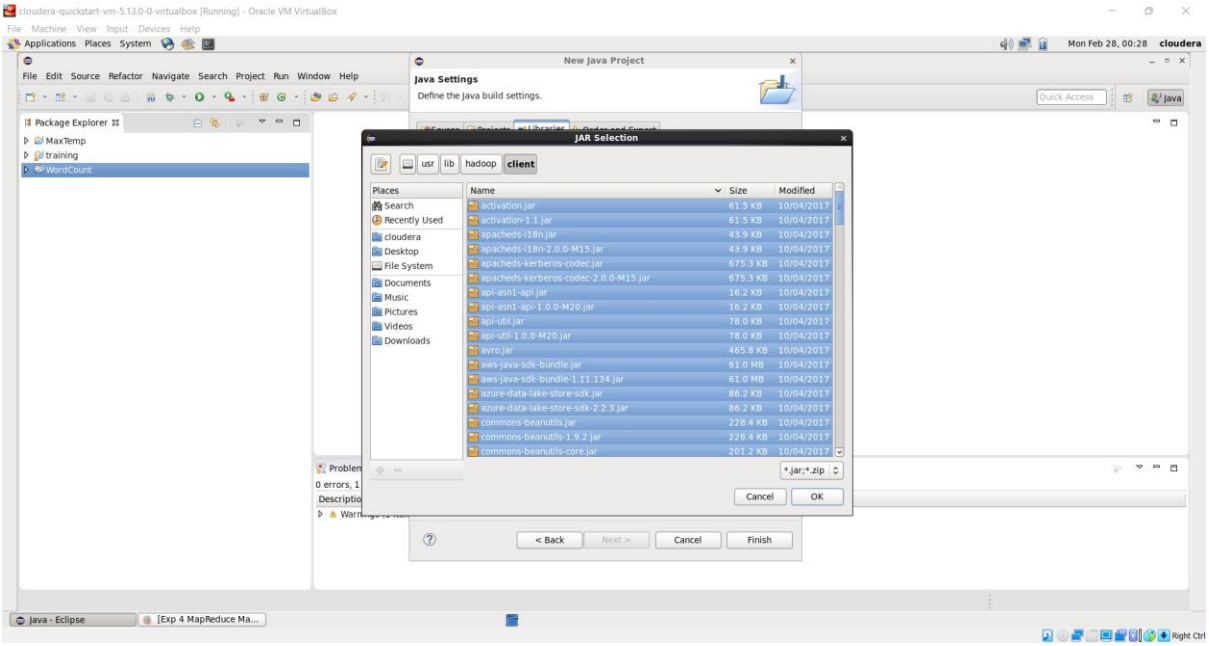
Practical 4
To determine maximum temperature using Hadoop MapReduce

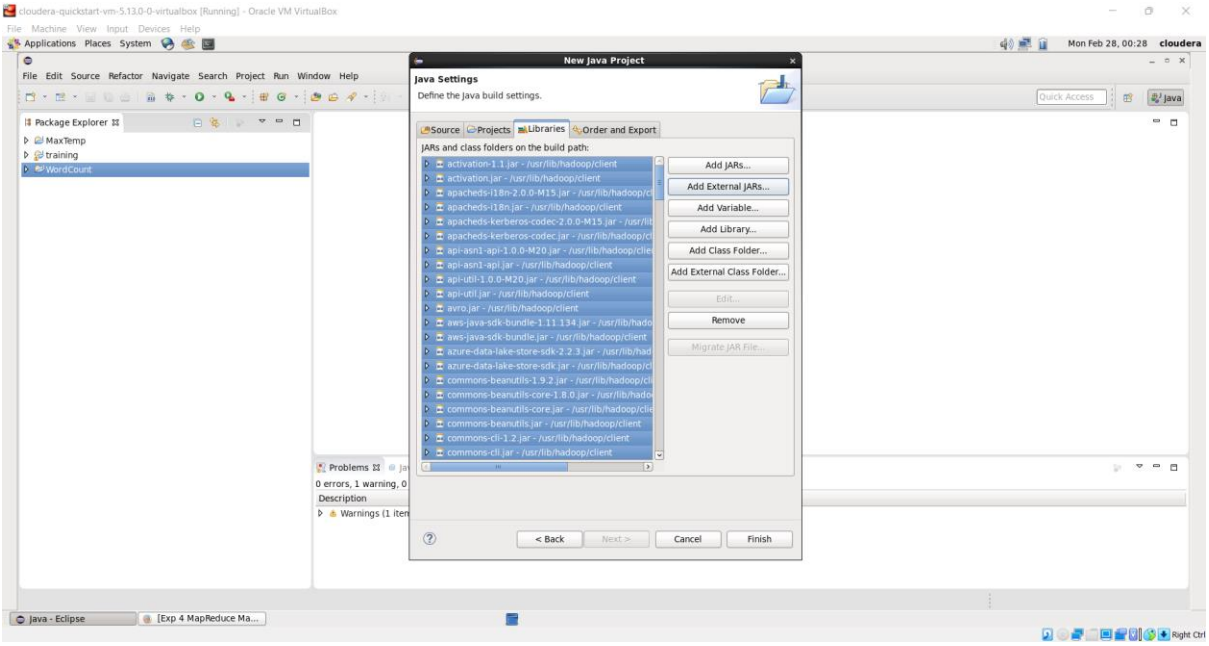
Dataset – temperature.txt

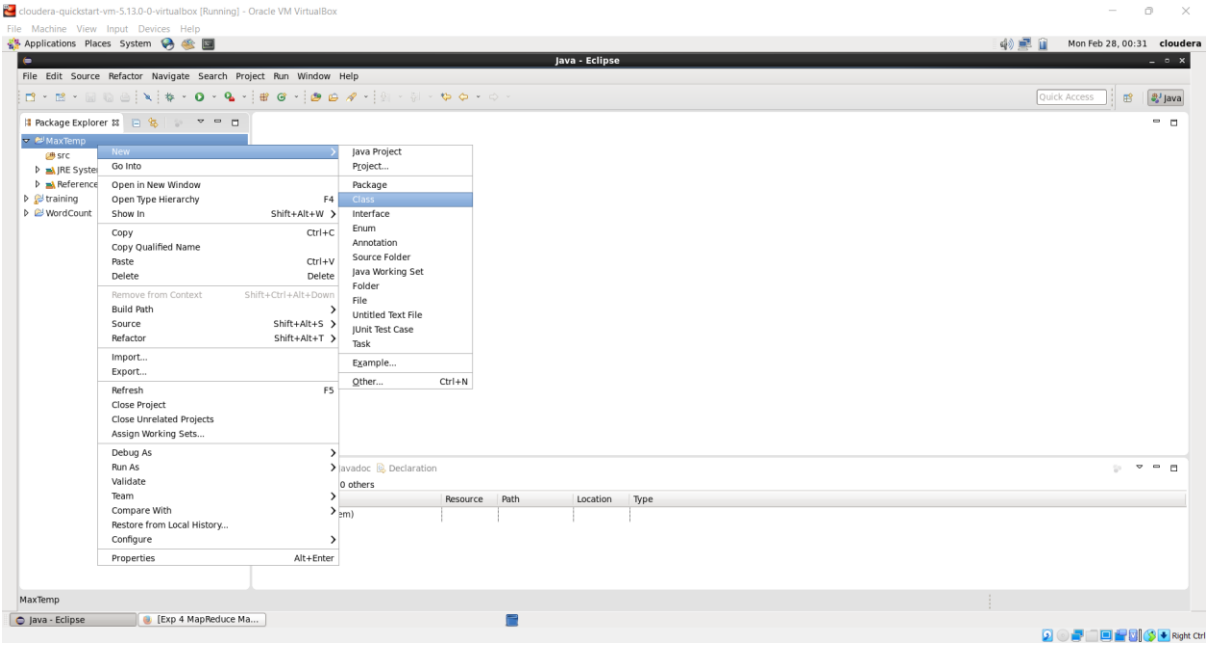


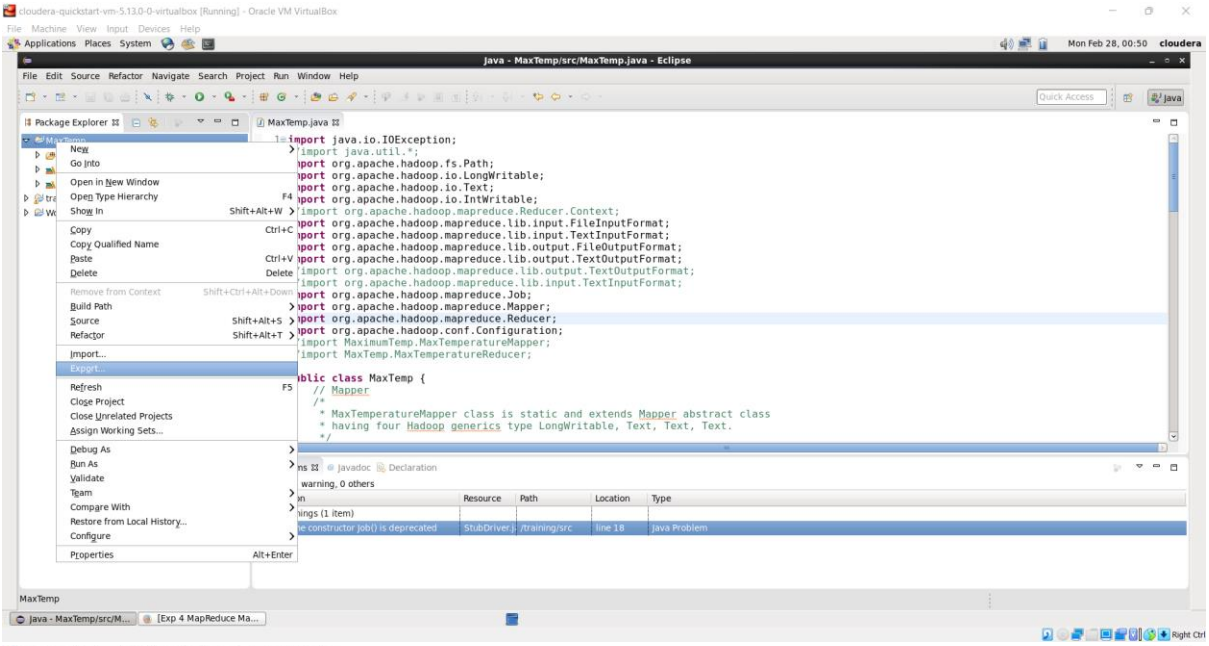


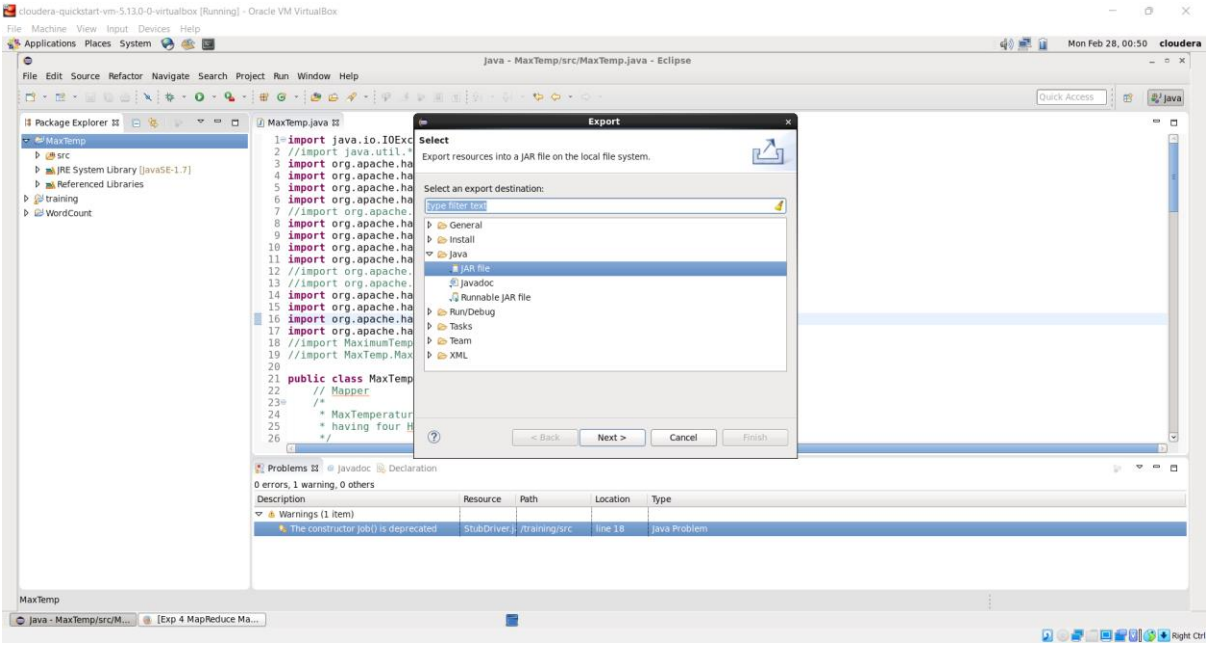

New Java Project
Java Settings
Define the Java build settings.
JAR Selection
usr
lib
hadoop
client
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name
Size
Modified
lib
libexec
sbin
hadoop-annotations.jar
hadoop-annotations-2.6.0-cdh5.13.0.jar
hadoop-auth.jar
hadoop-auth-2.6.0-cdh5.13.0.jar
hadoop-aws.jar
hadoop-aws-2.6.0-cdh5.13.0.jar
hadoop-azure-datalake.jar
hadoop-azure-datalake-2.6.0-cdh5.13.0.jar
hadoop-common.jar
hadoop-common-2.6.0-cdh5.13.0.jar
hadoop-common-2.6.0-cdh5.13.0-tests.jar
hadoop-common-tests.jar
hadoop-nfs.jar
hadoop-nfs-2.6.0-cdh5.13.0.jar
parquet-avro.jar
.jar;.zip
Cancel
OK
< Back
Next >
Cancel
Finish

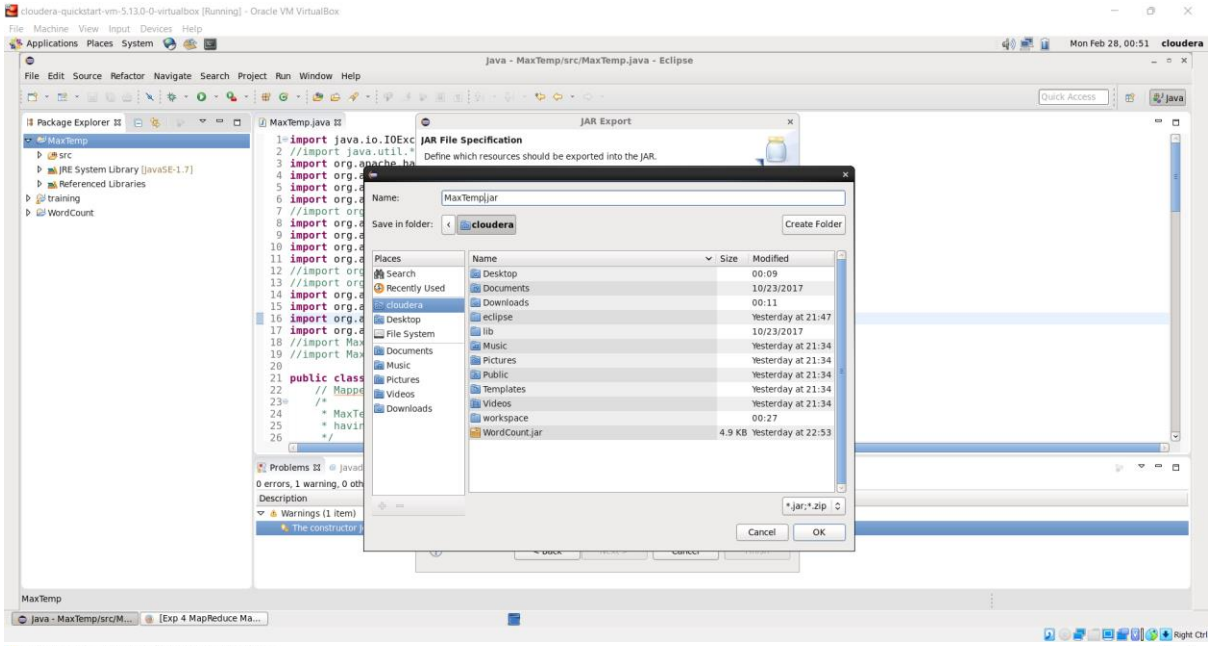

New Java Project
Java Settings
Define the Java build settings.
JAR Selection
usr
lib
hadoop
client
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name
Size
Modified
activation.jar
activation-1.1.jar
apacheds-i18n.jar
apacheds-i18n-2.0.0-M15.jar
apacheds-kerberos-codec.jar
apacheds-kerberos-codec-2.0.0-M15.jar
api-asn1-api.jar
api-asn1-api-1.0.0-M20.jar
api-util.jar
api-util-1.0.0-M20.jar
avro.jar
aws-java-sdk-bundle.jar
aws-java-sdk-bundle-1.11.134.jar
azure-data-lake-store-sdk.jar
azure-data-lake-store-sdk-2.2.3.jar
commons-beanutils.jar
commons-beanutils-1.9.2.jar
commons-beanutils-core.jar
.jar;.zip
Cancel
OK
< Back
Next >
Cancel
Finish

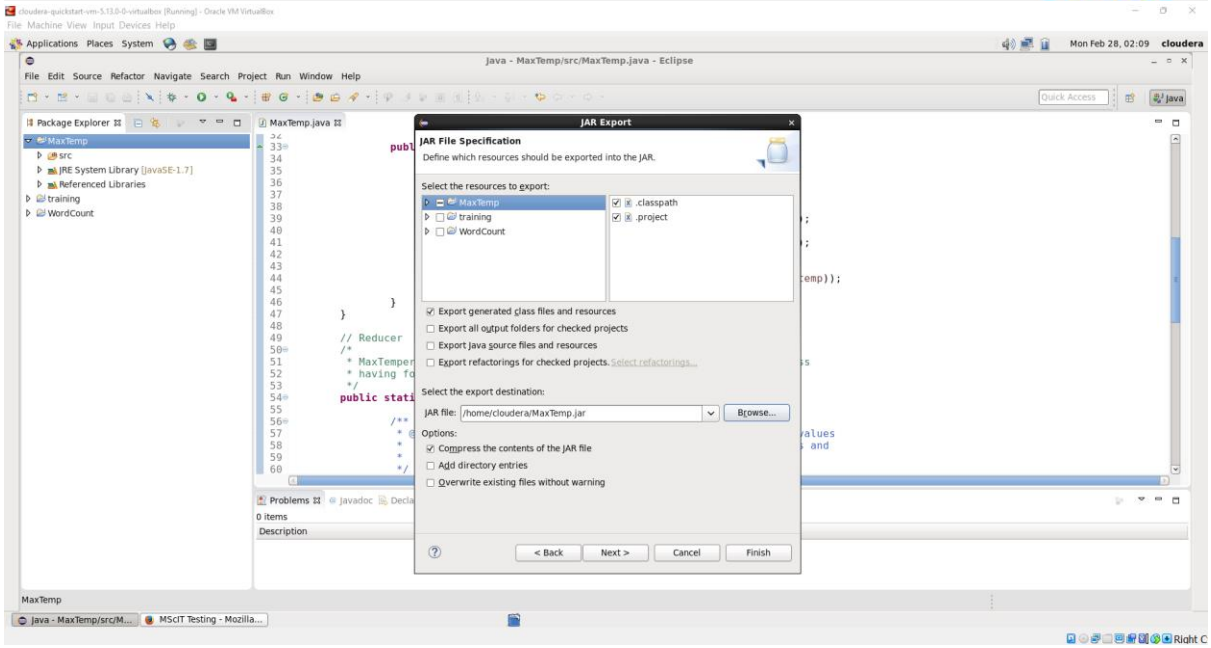

New Java Project
Java Settings
Define the Java build settings.
Source
Projects
Libraries
Order and Export
JARs and class folders on the build path:
Add JARs...
Add External JARs...
Add Variable...
Add Library...
Add Class Folder...
Add External Class Folder...
Edit...
Remove
Migrate JAR File...
< Back
Next >
Cancel
Finish

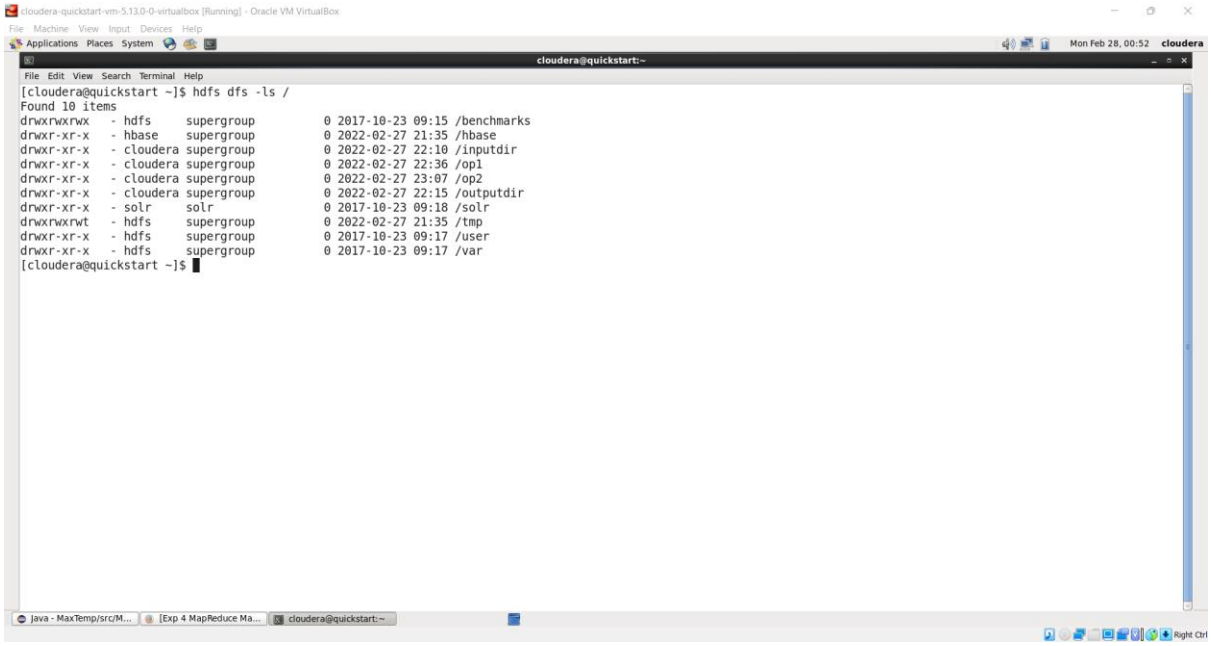

Java - Eclipse
New
Go Into
Open in New Window
Open Type Hierarchy
Show In
Copy
Copy Qualified Name
Paste
Delete
Build Path
Source
Refactor
Import...
Export...
Refresh
Close Project
Close Unrelated Projects
Assign Working Sets...
Debug As
Run As
Validate
Team
Compare With
Restore from Local History...
Configure
Properties
Java Project
Project...
Package
Class
Interface
Enum
Annotation
Source Folder
Java Working Set
Folder
File
Untitled Text File
JUnit Test Case
Task
Example...
Other...

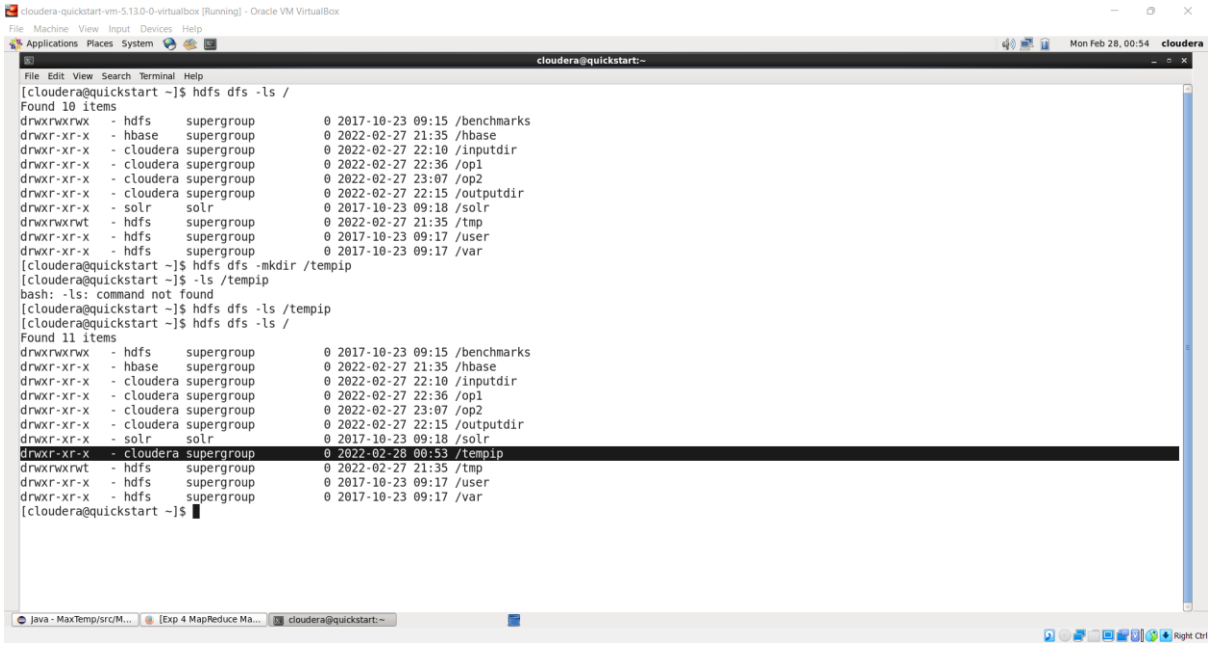

Java - MaxTemp/src/MaxTemp.java - Eclipse
Package Explorer
MaxTemp.java
New
Go Into
Open in New Window
Open Type Hierarchy
Show In
Copy
Copy Qualified Name
Paste
Delete
Build Path
Source
Refactor
Import...
Export...
Refresh
Close Project
Close Unrelated Projects
Assign Working Sets...
Debug As
Run As
Validate
Team
Compare With
Restore from Local History...
Configure
Properties

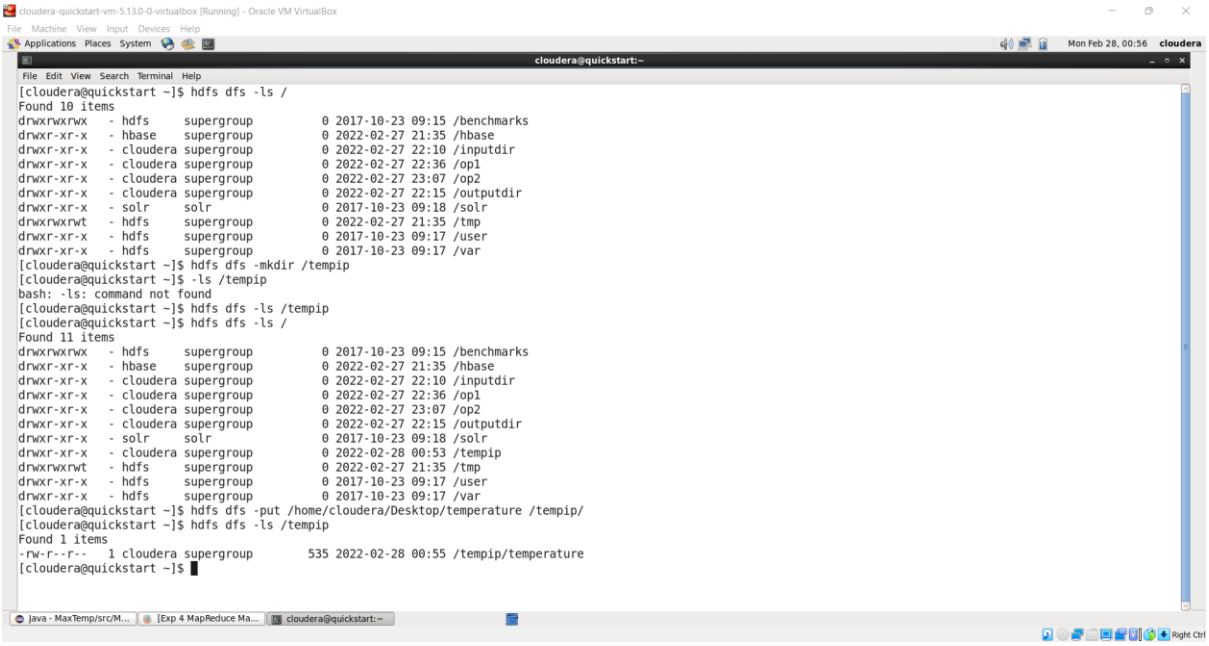

Java - MaxTemp/src/MaxTemp.java - Eclipse
Export
Select
Export resources into a JAR file on the local file system.
Select an export destination:
General
Install
Java
JAR file
Javadoc
Runnable JAR file
Run/Debug
Tasks
Team
XML
< Back
Next >
Cancel
Finish

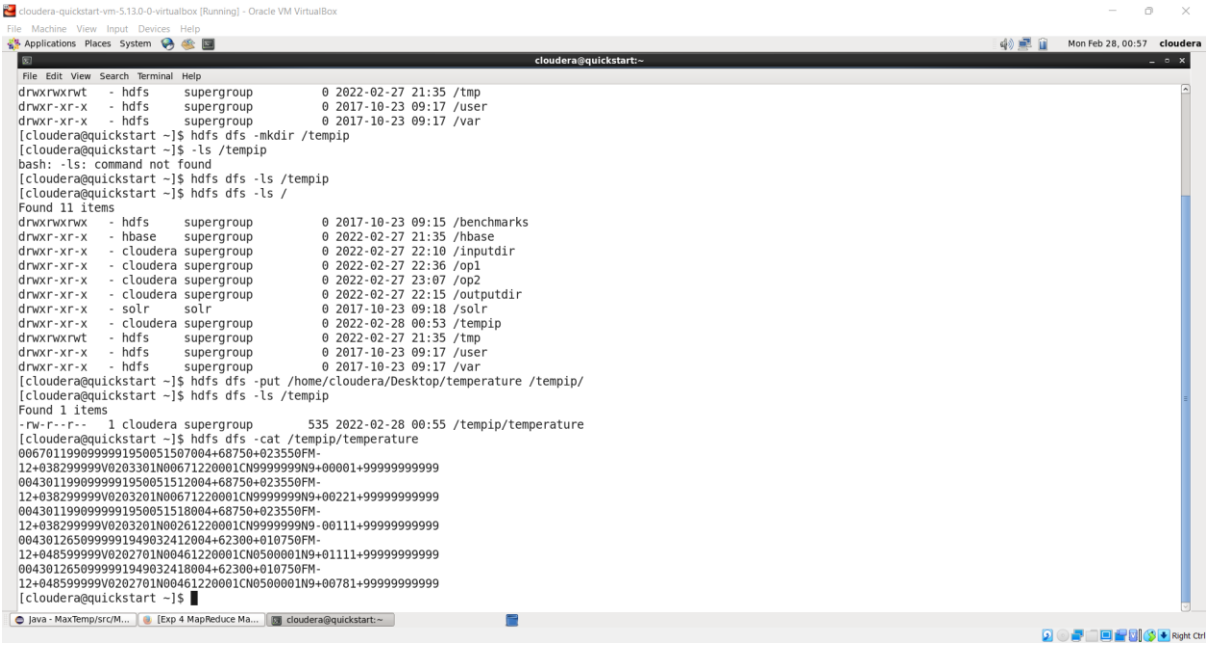

Java - MaxTemp/src/MaxTemp.java - Eclipse
JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Name: MaxTemp.jar
Save in folder: cloudera
Create Folder
Cancel
OK

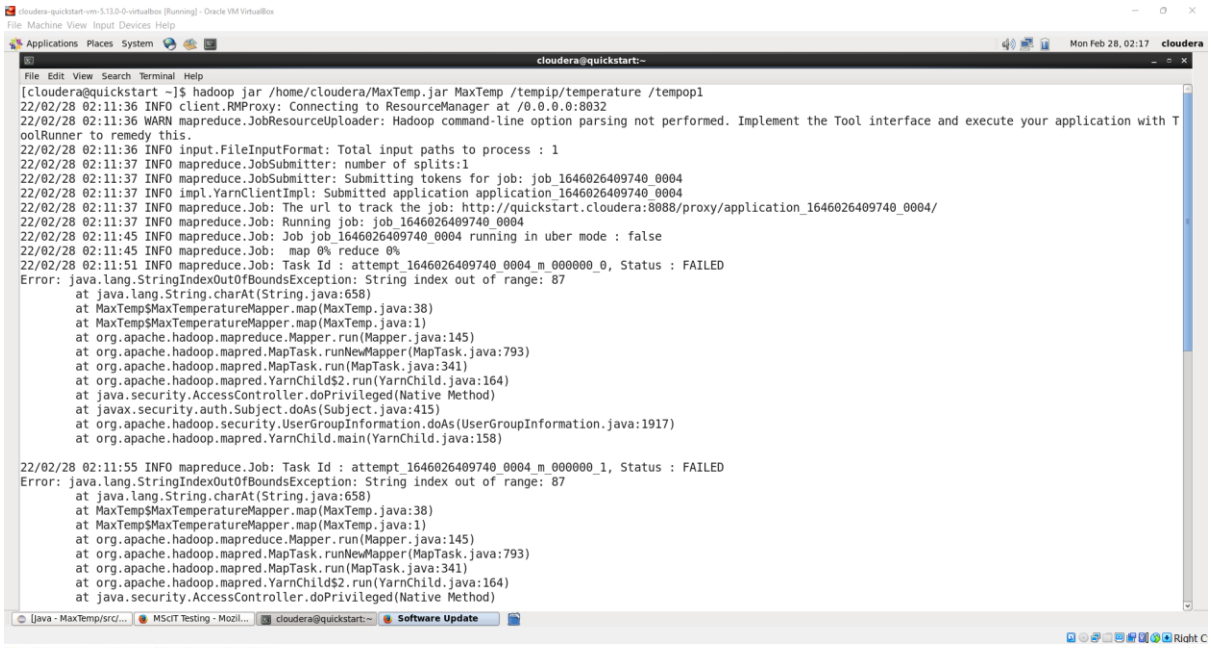

Java - MaxTemp/src/MaxTemp.java - Eclipse
JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Select the resources to export:
Export generated class files and resources
Export all output folders for checked projects
Export Java source files and resources
Export refactorings for checked projects.
Select the export destination:
JAR file: /home/cloudera/MaxTemp.jar
Browse...
Options:
Compress the contents of the JAR file
Add directory entries
Overwrite existing files without warning
< Back
Next >
Cancel
Finish

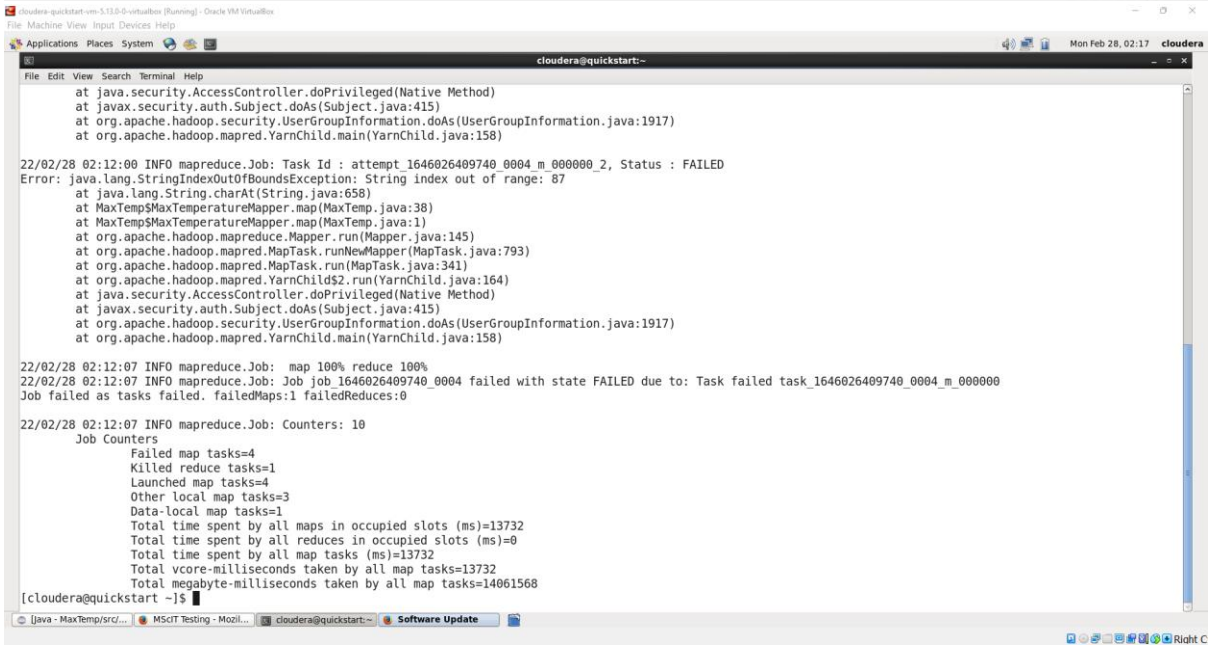

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Mon Feb 28, 00:52 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 10 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:36 /op1
drwxr-xr-x - cloudera supergroup 0 2022-02-27 23:07 /op2
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$
Java - MaxTemp/src/M...
[Exp 4 MapReduce Ma...
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Mon Feb 28, 00:54 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 10 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:36 /op1
drwxr-xr-x - cloudera supergroup 0 2022-02-27 23:07 /op2
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /tempip
[cloudera@quickstart ~]$ -ls /tempip
bash: -ls: command not found
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 11 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:36 /op1
drwxr-xr-x - cloudera supergroup 0 2022-02-27 23:07 /op2
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxr-xr-x - cloudera supergroup 0 2022-02-28 00:53 /tempip
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$
Java - MaxTemp/src/M...
[Exp 4 MapReduce Ma...
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Mon Feb 28, 00:56 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 10 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:36 /op1
drwxr-xr-x - cloudera supergroup 0 2022-02-27 23:07 /op2
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /tempip
[cloudera@quickstart ~]$ -ls /tempip
bash: -ls: command not found
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 11 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:36 /op1
drwxr-xr-x - cloudera supergroup 0 2022-02-27 23:07 /op2
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxr-xr-x - cloudera supergroup 0 2022-02-28 00:53 /tempip
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/temperature /tempip/
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
Found 1 items
-rw-r--r-- 1 cloudera supergroup 535 2022-02-28 00:55 /tempip/temperature
[cloudera@quickstart ~]$
Java - MaxTemp/src/M...
[Exp 4 MapReduce Ma...
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Mon Feb 28, 00:57 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /tempip
[cloudera@quickstart ~]$ -ls /tempip
bash: -ls: command not found
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 11 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 21:35 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:10 /inputdir
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:36 /op1
drwxr-xr-x - cloudera supergroup 0 2022-02-27 23:07 /op2
drwxr-xr-x - cloudera supergroup 0 2022-02-27 22:15 /outputdir
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxr-xr-x - cloudera supergroup 0 2022-02-28 00:53 /tempip
drwxrwxrwt - hdfs supergroup 0 2022-02-27 21:35 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/temperature /tempip/
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
Found 1 items
-rw-r--r-- 1 cloudera supergroup 535 2022-02-28 00:55 /tempip/temperature
[cloudera@quickstart ~]$ hdfs dfs -cat /tempip/temperature
0067011990999991950051507004+68750+023550FM-
12+038299999V0203301N00671220001CN9999999N9+00001+99999999999
0043011990999991950051512004+68750+023550FM-
12+038299999V0203201N00671220001CN9999999N9+00221+99999999999
0043011990999991950051518004+68750+023550FM-
12+038299999V0203201N00261220001CN9999999N9-00111+99999999999
0043012650999991949032412004+62300+010750FM-
12+048599999V0202701N00461220001CN0500001N9+01111+99999999999
0043012650999991949032418004+62300+010750FM-
12+048599999V0202701N00461220001CN0500001N9+00781+99999999999
[cloudera@quickstart ~]$
Java - MaxTemp/src/M...
[Exp 4 MapReduce Ma...
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Mon Feb 28, 02:17 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/MaxTemp.jar MaxTemp /tempip/temperature /tempop1
22/02/28 02:11:36 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/28 02:11:36 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
22/02/28 02:11:36 INFO input.FileInputFormat: Total input paths to process : 1
22/02/28 02:11:37 INFO mapreduce.JobSubmitter: number of splits:1
22/02/28 02:11:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646026409740_0004
22/02/28 02:11:37 INFO impl.YarnClientImpl: Submitted application application_1646026409740_0004
22/02/28 02:11:37 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1646026409740_0004/
22/02/28 02:11:37 INFO mapreduce.Job: Running job: job_1646026409740_0004
22/02/28 02:11:45 INFO mapreduce.Job: Job job_1646026409740_0004 running in uber mode : false
22/02/28 02:11:45 INFO mapreduce.Job: map 0% reduce 0%
22/02/28 02:11:51 INFO mapreduce.Job: Task Id : attempt_1646026409740_0004_m_000000_0, Status : FAILED
Error: java.lang.StringIndexOutOfBoundsException: String index out of range: 87
at java.lang.String.charAt(String.java:658)
at MaxTemp$MaxTemperatureMapper.map(MaxTemp.java:38)
at MaxTemp$MaxTemperatureMapper.map(MaxTemp.java:1)
at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:145)
at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:793)
at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:415)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1917)
at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)
22/02/28 02:11:55 INFO mapreduce.Job: Task Id : attempt_1646026409740_0004_m_000000_1, Status : FAILED
Error: java.lang.StringIndexOutOfBoundsException: String index out of range: 87
at java.lang.String.charAt(String.java:658)
at MaxTemp$MaxTemperatureMapper.map(MaxTemp.java:38)
at MaxTemp$MaxTemperatureMapper.map(MaxTemp.java:1)
at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:145)
at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:793)
at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
at java.security.AccessController.doPrivileged(Native Method)
[java - MaxTemp/src/...
MSCIT Testing - Mozil...
cloudera@quickstart:~
Software Update


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Mon Feb 28, 02:17 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:415)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1917)
at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)
22/02/28 02:12:00 INFO mapreduce.Job: Task Id : attempt_1646026409740_0004_m_000000_2, Status : FAILED
Error: java.lang.StringIndexOutOfBoundsException: String index out of range: 87
at java.lang.String.charAt(String.java:658)
at MaxTemp$MaxTemperatureMapper.map(MaxTemp.java:38)
at MaxTemp$MaxTemperatureMapper.map(MaxTemp.java:1)
at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:145)
at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:793)
at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:415)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1917)
at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)
22/02/28 02:12:07 INFO mapreduce.Job: map 100% reduce 100%
22/02/28 02:12:07 INFO mapreduce.Job: Job job_1646026409740_0004 failed with state FAILED due to: Task failed task_1646026409740_0004_m_000000
Job failed as tasks failed. failedMaps:1 failedReduces:0
22/02/28 02:12:07 INFO mapreduce.Job: Counters: 10
Job Counters
Failed map tasks=4
Killed reduce tasks=1
Launched map tasks=4
Other local map tasks=3
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=13732
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=13732
Total vcore-milliseconds taken by all map tasks=13732
Total megabyte-milliseconds taken by all map tasks=14061568
[cloudera@quickstart ~]$
[java - MaxTemp/src/...
MSCIT Testing - Mozil...
cloudera@quickstart:~
Software Update