

1NT19IS108
CHAITANYA P
C1 BATCH

Exercise-3: MAPREDUCE

Use the Hadoop framework to write a custom MapReduce program to perform word count operation on a custom data set.

Initially create a new project, package and class in eclipse to run a java code.

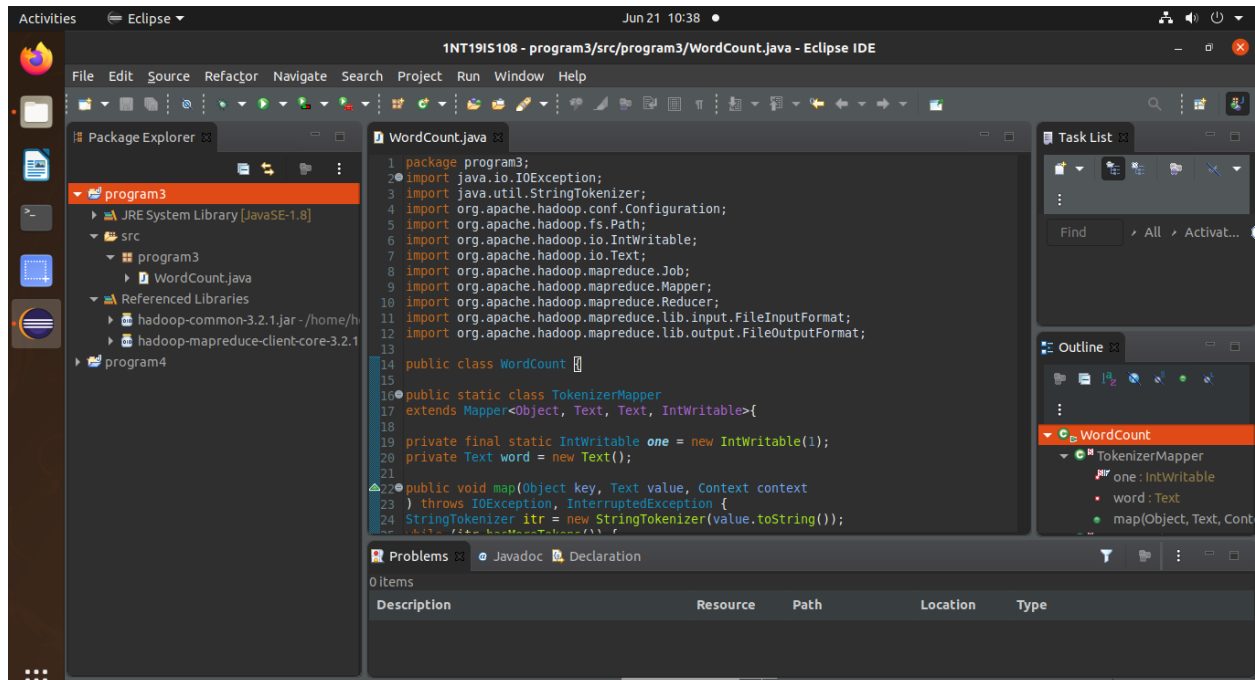
To install jar files:

Right click on project (Mapreduce)

Click on -> build a path -> add external archives -> Hadoop 3.2.1 -> share

In share 1. Click on common ->open hadoop-common-3.2.1.jar

2. Click on mapreduce -> open hadoop-mapreduce-client-core-3.2.1.jar



IN TERMINAL:

Run the commands:

```
cd $HADOOP_HOME
```

```
cd sbin
```

```
jps
```

```
start-all.sh
```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
4999 ResourceManager
7241 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
4811 SecondaryNameNode
5148 NodeManager
4428 NameNode
4572 DataNode
8511 Jps
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 4428. Stop it first.
Starting datanodes
localhost: datanode is running as process 4572. Stop it first.
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
admin1-HP-280-G4-MT-Business-PC: secondarynamenode is running as process 4811. Stop it first.
Starting resourcemanager
resourcemanager is running as process 4999. Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 5148. Stop it first.

```

hdfs dfs -mkdir -p ~/inputchait

hdfs dfs -appendToFile - ~/inputchait/1008.txt

Create a file and add content to it. ->(ctrl D two times)

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p ~/inputchait
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - ~/inputchait/1008.txt
hat is the hat of all of all the hats2022-06-21 10:30:18,771 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostT
rusted = false

```

hadoop jar /home/hadoop/Desktop/chaitanya0.jar ~/inputchait ~/outputchait

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/chaitanya0.jar ~/inputchait ~/outputchait
2022-06-21 10:32:35,638 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-21 10:32:35,869 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2022-06-21 10:32:35,899 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1655785365793_0002
2022-06-21 10:32:35,980 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-21 10:32:36,074 INFO input.FileInputFormat: Total input files to process : 1
2022-06-21 10:32:36,123 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-21 10:32:36,165 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-21 10:32:36,173 INFO mapreduce.JobSubmitter: number of splits:1
2022-06-21 10:32:36,256 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-21 10:32:36,273 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655785365793_0002
2022-06-21 10:32:36,273 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-21 10:32:36,385 INFO conf.Configuration: resource-types.xml not found
2022-06-21 10:32:36,385 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-21 10:32:36,428 INFO impl.YarnClientImpl: Submitted application application_1655785365793_0002
2022-06-21 10:32:36,454 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655785365793_0002/
2022-06-21 10:32:36,455 INFO mapreduce.Job: Running job: job_1655785365793_0002
2022-06-21 10:32:41,587 INFO mapreduce.Job: Job job_1655785365793_0002 running in uber mode : false
2022-06-21 10:32:41,589 INFO mapreduce.Job: map 0% reduce 0%
2022-06-21 10:32:44,647 INFO mapreduce.Job: map 100% reduce 0%
2022-06-21 10:32:48,686 INFO mapreduce.Job: map 100% reduce 100%
2022-06-21 10:32:49,710 INFO mapreduce.Job: Job job_1655785365793_0002 completed successfully
2022-06-21 10:32:49,766 INFO mapreduce.Job: Counters: 54
    File System Counters
      FILE: Number of bytes read=65
      FILE: Number of bytes written=451057
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=194
      HDFS: Number of bytes written=35
      HDFS: Number of read operations=8
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=2
      HDFS: Number of bytes read erasure-coded=0
    Job Counters
      Launched map tasks=1
      Launched reduce tasks=1
      Data-local map tasks=1
      Total time spent by all maps in occupied slots (ms)=1380

```

hdfs dfs -cat ~/outputchait/part*

```

    Bytes Written=35
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat ~/outputchait/part*
2022-06-21 10:33:14,555 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
all      2
bat      2
hats     1
is       1
of       2
the      2
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$

```