Hadoop Exercises on macOS

Exercise 0: Start Hadoop on macOS

Start Hadoop services

start-dfs.sh
start-yarn.sh

Check running Hadoop processes

jps
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
tar -xzf ~/Downloads/hadoop-3.3.6.tar.gz
cd hadoop-3.3.6/share/hadoop/mapreduce
pwd
EXAMPLES_JAR=<absolute_path>hadoop-mapreduce-examples-3.3.6.jar

---

Exercise 1: Explore Hadoop File System (HDFS)

hdfs dfs -ls /
hdfs dfs -mkdir /user
hdfs dfs -mkdir /user/yourname
hdfs dfs -ls /user

Exercise 2: Upload Files to HDFS

echo "Hello Hadoop" > hello.txt
hdfs dfs -put hello.txt /user/yourname/
hdfs dfs -ls /user/yourname
hdfs dfs -cat /user/yourname/hello.txt

Exercise 3: Download & Delete Files from HDFS

hdfs dfs -get /user/yourname/hello.txt downloaded_hello.txt
hdfs dfs -rm /user/yourname/hello.txt

Exercise 4: HDFS Directory Operations

hdfs dfs -mkdir -p /user/yourname/data/input
hdfs dfs -rm -r /user/yourname/data

Exercise 5: Understand HDFS Replication

hdfs dfs -stat %r /user/yourname/hello.txt
hdfs dfs -setrep -w 3 /user/yourname/hello.txt

Exercise 6: Run Your First MapReduce Job (WordCount)

nano input.txt

Input content example:

hadoop is fun
hadoop is powerful
hadoop is scalable
hdfs dfs -mkdir /user/yourname/wordcount
hdfs dfs -put input.txt /user/yourname/wordcount/input
hadoop jar $EXAMPLES_JAR \
wordcount \
/user/hadoop/wordcount/input \
/user/hadoop/wordcount/output
hdfs dfs -cat /user/yourname/wordcount/output/part-r-00000

Expected output:

hadoop 3
is 3
fun 1
powerful 1
scalable 1

Exercise 7: Understand MapReduce Output Structure

hdfs dfs -ls /user/yourname/wordcount/output

Exercise 8: Modify Input & Re-run Job

```
echo "hadoop is distributed" >> input.txt
hdfs dfs -put -f input.txt /user/yourname/wordcount/input
hdfs dfs -rm -r /user/yourname/wordcount/output
hadoop jar ... (same command as before)
```

Exercise 9: View Hadoop Web UI

Open in browser:

NameNode UI: http://localhost:9870

YARN UI: http://localhost:8088

Exercise 10: Mini Project – Log Analysis

nano access.log

Example log content:

```
200 GET /index.html
404 GET /missing.html
200 POST /submit
500 GET /error
200 GET /home
hdfs dfs -mkdir /user/yourname/logs
hdfs dfs -put access.log /user/yourname/logs/input
hadoop jar \
/usr/local/Cellar/hadoop/*/libexec/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar \
wordcount \
/user/yourname/logs/input \
/user/yourname/logs/output
hdfs dfs -cat /user/yourname/logs/output/part-r-00000
```