

# Cloud Computing Lab DA-3

Arvind Babu  
22MIC0066

## HADOOP on Ubuntu 24.04:

### 1. Install OpenJDK

```
arvind@arvind-GF65-Thin-9SD:~$ sudo apt install openjdk-8-jdk -y
[sudo] password for arvind:
Sorry, try again.
[sudo] password for arvind:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
openjdk-8-jdk is already the newest version (8u422-b05-1~24.04).
The following packages were automatically installed and are no longer required:
  libb2-1 libliblept5 libqalculate-data libqalculate22t64 libqt6concurrent6t64
  libqt6core6t64 libqt6dbus6t64 libqt6gui6t64 libqt6network6t64
  libqt6opengl6t64 libqt6qml6 libqt6qmlmodels6 libqt6qmlworkerscript6
  libqt6quick6 libqt6quickshapes6 libqt6shadertools6 libqt6sql6-sqlite
  libqt6sql6t64 libqt6statemachine6 libqt6statemachineqml6 libqt6svg6
```

```
arvind@arvind-GF65-Thin-9SD:~$ java -version
openjdk version "21.0.4" 2024-07-16
OpenJDK Runtime Environment (build 21.0.4+7-Ubuntu-1ubuntu224.04)
OpenJDK 64-Bit Server VM (build 21.0.4+7-Ubuntu-1ubuntu224.04, mixed mode, sharing)
```

### 2. Move to new user and setup OpenSSH

```

arvind@arvind-GF65-Thin-9SD:~$ sudo apt install openssh-server openssh-client -y
[sudo] password for arvind:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
openssh-server is already the newest version (1:9.6p1-3ubuntu13.5).
openssh-client is already the newest version (1:9.6p1-3ubuntu13.5).
openssh-client set to manually installed.
The following packages were automatically installed and are no longer required:
  libb2-1 libblept5 libqalculate-data libqalculate22t64 libqt6concurrent6t64
  libqt6core6t64 libqt6dbus6t64 libqt6gui6t64 libqt6network6t64
  libqt6opengl6t64 libqt6qml6 libqt6qmlmodels6 libqt6qmlworkerscript6
  libqt6quick6 libqt6quickshapes6 libqt6shadertools6 libqt6sql6-sqlite
  libqt6sql6t64 libqt6statemachine6 libqt6statemachineqml6 libqt6svg6

```

```

arvind@arvind-GF65-Thin-9SD:~$ su - hdoop
Password:
hdoop@arvind-GF65-Thin-9SD:~$ ssh localhost
Welcome to Ubuntu 24.04 LTS (GNU/Linux 6.8.0-45-generic x86_64)
Already the newest version (1:9.6p1-3ubuntu13.5).
manually installed.
* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/pro
Expanded Security Maintenance for Applications is not enabled.

155 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

Last login: Thu Oct  3 18:29:29 2024 from 127.0.0.1

```

### 3. Install Hadoop from Apache.org using wget

```

hadoop@arvind-GF65-Thin-9SD:~$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
2024-10-25 16:43:40 https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 965537117 (921M) [application/x-gzip]
Saving to: 'hadoop-3.4.0.tar.gz'

hadoop-3.4.0.tar.gz  21%[==>                ] 200.60M   714KB/s   eta 17m 51s^
hadoop-3.4.0.tar.gz  27%[====>             ] 256.74M   710KB/s   eta 16m 20s

```

```

d.
HTTP request sent, awaiting response... 200 OK
Length: 965537117 (921M) [application/x-gzip]
Saving to: 'hadoop-3.4.0.tar.gz'
4.0 directory.

hadoop-3.4.0.tar.gz  21%[==>                ] 200.60M   714KB/s   eta 17m 51s^
hadoop-3.4.0.tar.gz 100%[=====] 920.81M   684KB/s   in 22m 49s

2024-10-25 17:06:29 (689 KB/s) - 'hadoop-3.4.0.tar.gz' saved [965537117/965537117]

a large cluster of networked servers
hadoop@arvind-GF65-Thin-9SD:~$ tar xzf hadoop-3.4.0.tar.gz
hadoop@arvind-GF65-Thin-9SD:~$

```

#### 4. Configure variables

```
GNU nano 7.2 .bashrc *
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi
#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.4.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^/ Go To Line
```

```
GNU nano 7.2 /home/hdoop/hadoop-3.4.0/etc/hadoop/hadoop-env.sh *
# JAVA_HOME=/usr/java/testing/hdfs -ls
# and add the full path to the OpenJDK
# Therefore, the vast majority (BUT NOT ALL!) of these defaults
# are configured for substitution and not append. If append
# is preferable, modify this file accordingly.

###
# Generic settings for HADOOP
###

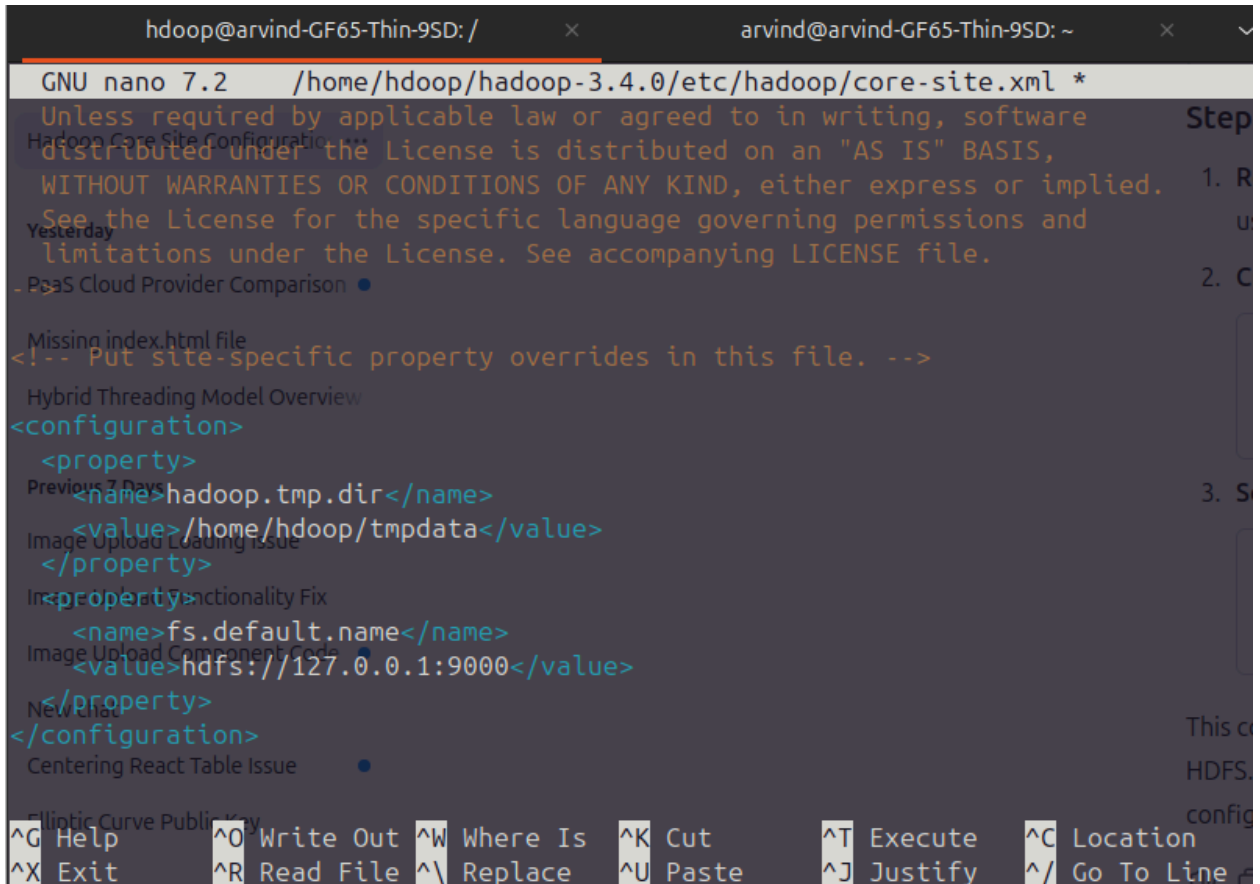
# Technically, the only required environment variable is JAVA_HOME.
# All others are optional. However, the defaults are probably not
# preferred. Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

# The language environment in which Hadoop runs. Use the English
# expected.
```

## 5. Edit core-site.xml

```
hadoop@arvind-GF65-Thin-9SD:/$ mkdir -p /home/hadoop/tmpdata
hadoop@arvind-GF65-Thin-9SD:/$ chmod 755 /home/hadoop/tmpdata
hadoop@arvind-GF65-Thin-9SD:/$ nano $HADOOP_HOME/etc/hadoop/core-site.xml
hadoop@arvind-GF65-Thin-9SD:/$
```



The screenshot shows a terminal window with two tabs. The active tab is titled 'hadoop@arvind-GF65-Thin-9SD: /' and displays the nano 7.2 editor. The editor is editing the file '/home/hadoop/hadoop-3.4.0/etc/hadoop/core-site.xml'. The content of the file is as follows:

```
GNU nano 7.2 /home/hadoop/hadoop-3.4.0/etc/hadoop/core-site.xml *
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/home/hadoop/tmpdata</value>
  </property>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://127.0.0.1:9000</value>
  </property>
</configuration>
```

The bottom of the screen shows the nano editor's command shortcuts:

^G Help	^O Write Out	^W Where Is	^K Cut	^T Execute	^C Location
^X Exit	^R Read File	^_ Replace	^U Paste	^J Justify	^_ Go To Line

## 6. Edit hdfs-site.xml

```
hadoop@arvind-GF65-Thin-9SD: /
arvind@arvind-GF65-Thin-9SD: ~
GNU nano 7.2 /home/hadoop/hadoop-3.4.0/etc/hadoop/hdfs-site.xml *
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>dfs.data.dir</name>
    <value>/home/hadoop/dfsdata/namenode</value>
  </property>
  <property>
    <name>dfs.data.dir</name>
    <value>/home/hadoop/dfsdata/datanode</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line
```

```
hadoop@arvind-GF65-Thin-9SD:/$ mkdir -p /home/hadoop/dfsdata/namenode
hadoop@arvind-GF65-Thin-9SD:/$ mkdir -p /home/hadoop/dfsdata/datanode
hadoop@arvind-GF65-Thin-9SD:/$ chmod -R 755 /home/hadoop/dfsdata
hadoop@arvind-GF65-Thin-9SD:/$
```

## 7. Edit mapred-site and yarn-site

```
hdoop@arvind-GF65-Thin-9SD: /  
arvind@arvind-GF65-Thin-9SD: ~  
GNU nano 7.2 /home/hdoop/hadoop-3.4.0/etc/hadoop/mapred-site.xml  
You may obtain a copy of the License at  
http://www.apache.org/licenses/LICENSE-2.0  
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License. See accompanying LICENSE file.  
-->  
!-- Put site-specific property overrides in this file. -->  
configuration>  
property>  
  <name>mapreduce.framework.name</name>  
  <value>yarn</value>  
/property>  
/configuration>  
G Help ^O Write Out ^W Where Is ^K Cut ^T Execute Put ^C Location  
X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^_ Go To Line
```



```
hdoop@arvind-GF65-Thin-9SD: / x arvind@arvind-GF65-Thin-9SD: ~ x
GNU nano 7.2 /home/hdoop/hadoop-3.4.0/etc/hadoop/yarn-site.xml *
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file
->
configuration>
property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
/property>
property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
/property>
property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
/property>
property>
  <name>yarn.acl.enable</name>
  <value>0</value>
/property>
property>
  <name>yarn.nodemanager.env-whitelist</name>
  <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH</value>
/property>
This site uses cookies. Some of them are
essential, while others help us improve
G Help ^O Write Out ^W Where Is ^K Cut ^T Execute 7.2 ^C Location
X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^_ Go To Line
```

## 8. Format hdfs namenode

```
hdoop@arvind-GF65-Thin-9SD: /$ hdfs namenode -format
WARNING: /home/hdoop/hadoop-3.4.0/logs does not exist. Creating.
2024-10-25 17:28:17,743 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = arvind-GF65-Thin-9SD/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.4.0
*****/
hdoop@phoenixnap: ~$ hdfs
WARNING: /home/hdoop/had
2024-09-09 13:08:42,739
/*****
STARTUP_MSG: Starting Na
```

## 9.Start the nodes



```
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [arvind-GF65-Thin-9SD]
arvind-GF65-Thin-9SD: Warning: Permanently added 'arvind-gf65-thin-9sd' (ED25519)
to the list of known hosts.
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$
```

```
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$
```

```
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ jps
9939 NameNode
10339 SecondaryNameNode
12901 Jps
12517 NodeManager
10091 DataNode
12332 ResourceManager
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$
```

## 10. Access from browser



The screenshot shows the Hadoop web interface at localhost:8080/cluster. The page title is 'All Applications'. On the left, there is a sidebar with a 'Cluster' menu and a 'Tools' menu. The main content area displays several metrics tables:

- Cluster Metrics:** A table with columns: Apps Submitted (0), Apps Pending (0), Apps Running (0), Apps Completed (0), Containers Running (0), Used Resources (<memory:0 B, vCores:0>), Total Resources (<memory:8 GB, vCores:8>), and Re (0).
- Cluster Nodes Metrics:** A table with columns: Active Nodes (1), Decommissioning Nodes (0), Decommissioned Nodes (0), Lost Nodes (0), and Unhealthy Nodes (0).
- Scheduler Metrics:** A table with columns: Scheduler Type (Capacity Scheduler), Scheduling Resource Type (memory-mb (unit-M), vcores), Minimum Allocation (<memory:1024, vCores:1>), Maximum Allocation (<memory:8192, vCores:4>), Maximum Cluster Application Priority (0), and Scheduler Busy % (0).

Below these metrics is a table titled 'Show 20 entries' with columns: ID, User, Name, Application Type, Application Tags, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU Vcores, Allocated Memory MB, Allocated GPUs, and Reser CPU VCor. The table is empty, and a message 'No data available in table' is displayed.

**Hadoop One Node Cluster is now working.**

**Q2. Test Hadoop using a program for it's APIs**

1. Install input from [https://github.com/MuhammadBilalYar/Hadoop-On-Window/blob/master/input\\_file.txt](https://github.com/MuhammadBilalYar/Hadoop-On-Window/blob/master/input_file.txt)
2. Start all daemons from hadoop

```

$ cd $HADOOP_HOME
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0$ cd sbin
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/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 9939. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 10091. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes [arvind-GF65-Thin-9SD]
arvind-GF65-Thin-9SD: secondarynamenode is running as process 10339. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
Starting resourcemanager
resourcemanager is running as process 12332. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 12517. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$

```

### 3. Create input directory and copy the file.

```

/tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ hadoop fs -mkdir /input_dir
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ hadoop fs -put ~/input_file.txt /input_dir
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ hadoop fs -ls /input_dir
Found 1 items
-rw-r--r-- 1 hadoop supergroup 1860 2024-10-25 18:30 /input_dir/input_file.txt
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$

```

### 4. Verify file contents

```
hadoop@arvind-GF65-Thin-9SD: /hadoop-3.4.0/sbin$ hadoop fs -cat /input_dir/input_
file.txt
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
```

5. Verify the File in HDFS to confirm that input\_file.txt is in HDFS. Use the following command:

View the File Contents: Display the contents of input\_file.txt in HDFS using cat.

```
hadoop fs -cat /input_dir/input_file.txt
```

Expected Result

After running these commands, you should see input\_file.txt listed in HDFS using hadoop fs -cat should display the file's contents in the terminal.

Let me know if any issues come up while following these steps!

## 5. Install MapReduceClient.jar

```

hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ wget https://github.com/Muhammad
BilalYar/Hadoop-On-Windows/raw/master/MapReduceClient.jar
--2024-10-25 18:33:16-- https://github.com/MuhammadBilalYar/Hadoop-On-Windows/ra
w/master/MapReduceClient.jar
Resolving github.com (github.com)... 20.207.73.82
Connecting to github.com (github.com)|20.207.73.82|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://raw.githubusercontent.com/MuhammadBilalYar/Hadoop-On-Windows/ma
ster/MapReduceClient.jar [following]
--2024-10-25 18:33:17-- https://raw.githubusercontent.com/MuhammadBilalYar/Hado
op-On-Windows/master/MapReduceClient.jar
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 2606:50c0:800
2::154, 2606:50c0:8000::154, 2606:50c0:8001::154
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|2606:50c0:80
02::154|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 270227 (264K) [application/octet-stream]
Saving to: 'MapReduceClient.jar'

MapReduceClient.jar 100%[=====] 263.89K  369KB/s  in 0.7s

2024-10-25 18:33:19 (369 KB/s) - 'MapReduceClient.jar' saved [270227/270227]

hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$

```

## 6. Run the MapReduce operation on the input file and store in output

```
hadoop@arvind-GF65-Thin-9SD:~/hadoop-3.4.0/sbin$ hadoop jar ./MapReduceClient.jar wordcount /input_dir /output_dir
2024-10-25 18:35:11,403 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2024-10-25 18:35:11,701 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1729860404954_0001
2024-10-25 18:35:11,919 INFO input.FileInputFormat: Total input files to process : 1
2024-10-25 18:35:11,960 INFO mapreduce.JobSubmitter: number of splits:1
2024-10-25 18:35:12,052 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1729860404954_0001
2024-10-25 18:35:12,052 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-10-25 18:35:12,176 INFO conf.Configuration: resource-types.xml not found
2024-10-25 18:35:12,176 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'
2024-10-25 18:35:12,562 INFO impl.YarnClientImpl: Submitted application application_1729860404954_0001
2024-10-25 18:35:12,585 INFO mapreduce.Job: The url to track the job: http://arvind-GF65-Thin-9SD:8088/proxy/application_1729860404954_0001/
2024-10-25 18:35:12,586 INFO mapreduce.Job: Running job: job_1729860404954_0001
2024-10-25 18:35:17,673 INFO mapreduce.Job: Job job_1729860404954_0001 running in uber mode : false
2024-10-25 18:35:17,676 INFO mapreduce.Job: map 0% reduce 0%
2024-10-25 18:35:21,778 INFO mapreduce.Job: map 100% reduce 0%
^[[1;5H2024-10-25 18:35:25,799 INFO mapreduce.Job: map 100% reduce 100%
2024-10-25 18:35:25,821 INFO mapreduce.Job: Job job_1729860404954_0001 completed successfully
2024-10-25 18:35:25,926 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=195
  FILE: Number of bytes written=617537
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
Result
After these steps, you can view the output of the wordcount job by running the following command:
```



```
←
Document tabs
  Tab 1
    Cloud Computing Lab 3
      Input split bytes=111
      Combine input records=390
      Combine output records=21
      Reduce input groups=21
      Reduce shuffle bytes=195
      Reduce input records=21
      Reduce output records=21
      Spilled Records=42
      Shuffled Maps =1
      Failed Shuffles=0
      Merged Map outputs=1
      GC time elapsed (ms)=68
      CPU time spent (ms)=930
      Physical memory (bytes) snapshot=551743488
      Virtual memory (bytes) snapshot=5141639168
      Total committed heap usage (bytes)=585629696
      Peak Map Physical memory (bytes)=333586432
      Peak Map Virtual memory (bytes)=2569056256
      Peak Reduce Physical memory (bytes)=218157056
      Peak Reduce Virtual memory (bytes)=2572582912
      Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
      File Input Format Counters
        Bytes Read=1860
      File Output Format Counters
        Bytes Written=120
hadoop@arvind-GF65-Thin-9SD: ~/hadoop-3.4.0/sbin$
```

## 7. Read the output

CONNECTION=0  
IO\_ERROR=0  
WRONG\_LENGTH=0  
WRONG\_MAP=0  
WRONG\_REDUCE=0

File Input Format Counters  
Bytes Read=1860  
File Output Format Counters  
Bytes Written=120


hadoop@arvind-GF65-Thin-9SD: ~/hadoop-3.4.0/sbin\$ hadoop fs -cat /output\_dir/\*

• /input\_dir  
• /output\_dir  
created by

Step 5: Verify the C  
1. Display Output  
output\_dir :  
bash  
hadoop fs -

This should sho

Result  
After these steps, yo  
successful execution



Cluster

About  
Nodes  
Node Labels  
Applications  
NEW  
NEW SAVING  
SUBMITTED  
ACCEPTED  
RUNNING  
FINISHED  
FAILED  
KILLED  
Scheduler

Tools

Cluster Metrics

Apps Submitted		Apps Pending		Apps Running		Apps Completed		Containers Running		Used Resources		Total Resources	
1	0	0	0	1	0	0	0	0	0	<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>	0	8

Cluster Nodes Metrics

Active Nodes		Decommissioning Nodes		Decommissioned Nodes		Lost Nodes		Unhealthy Nodes	
1	0	0	0	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type		Scheduling Resource Type		Minimum Allocation		Maximum Allocation		Maximum Cluster Application Priority		Scheduler	
Capacity Scheduler		memory-mb (unit-Mi), vcores		<memory:1024, vCores:1>		<memory:8192, vCores:4>		0		0	

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB
application_1729860404954_0001	hadoop	word count	MAPREDUCE		root.default	0	Fri Oct 25 18:35:12 +0550 2024	Fri Oct 25 18:35:12 +0550 2024	Fri Oct 25 18:35:24 +0550 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A

Showing 1 to 1 of 1 entries

# OpenStack

## 1. Install openstack-dashboard

```
arvind@arvind-GF65-Thin-9SD:~$ sudo apt-get install openstack-dashboard
[sudo] password for arvind:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following packages were automatically installed and are no longer required:
libb2-1 libblept5 libqalculate-data libqalculate22t64 libqt6concurrent6t64
libqt6core6t64 libqt6dbus6t64 libqt6gui6t64 libqt6network6t64 libqt6opengl6t64
libqt6qml6 libqt6qmlmodels6 libqt6qmlworkerscript6 libqt6quick6 libqt6quickshapes6
libqt6shadertools6 libqt6sql6-sqlite libqt6sql6t64 libqt6statemachine6
libqt6statemachineqml6 libqt6svg6 libqt6waylandclient6 libqt6waylandcompositor6
libqt6waylandeglclienthwintegration6 libqt6waylandeglcompositorhwintegration6
libqt6widgets6t64 libqt6wlsshellintegration6 libtesseract5 libts0t64
qml6-module-qt5compat-graphicaleffects qml6-module-qtqml-statemachine
qml6-module-qtqml-workerscript qml6-module-qtquick-shapes qml6-module-qtquick-window6
qt6-gtk-platformtheme qt6-gpa-plugins qt6-translations-l10n qt6-wayland
tesseract-ocr-eng tesseract-ocr-osd
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
apache2 apache2-bin apache2-data apache2-utils docutils-common
libapache2-mod-wsgi-py3 libapr1t64 libaprutil1-dbd-sqlite3 libaprutil1-ldap
```

## 2. Edit localsettings.py

Set controller as host

```
#EMAIL_HOST = 'smtp.my-company.com'
#EMAIL_PORT = 25
#EMAIL_HOST_USER = 'djangomail'
#EMAIL_HOST_PASSWORD = 'top-secret!'

OPENSTACK_HOST = "controller"
OPENSTACK_KEYSTONE_URL = "http://%s/identity/v3" % OPENSTACK_HOST

# The timezone of the server. This should correspond with the timezone
# of your entire OpenStack installation, and hopefully be in UTC.
TIME_ZONE = "UTC"

# Change this patch to the appropriate list of tuples containing
# a key, label and static directory containing two files:
# _variables.scss and _styles.scss
#AVAILABLE_THEMES = [
#    ('default', 'Default', 'themes/default'),
#    ('material', 'Material', 'themes/material'),
#    ('example', 'Example', 'themes/example'),
#]
```

◦ Allow all hosts to acc

ALLOWED\_HOSTS

◦ Configure the memca

SESSION\_ENGIN

```
CACHES = {
    'default': {
        'BACKEND': 'djang
        'LOCATION': 'LOC
    }
}
```

📌 Note

Comment out at

```
# We recommend you use memcached for development; otherwise after every reload
# of the django development server, you will have to login again. To use
# memcached set CACHES to something like below.
# For more information, see
# https://docs.djangoproject.com/en/1.11/topics/http/sessions/.
SESSION_ENGINE = 'django.contrib.sessions.backends.cache'
CACHES = {
    'default': {
        'BACKEND': 'django.core.cache.backends.memcached.PyMemcacheCache',
        'LOCATION': 'controller:11211',
    },
}

# If you use ``tox -e runserver`` for developments, then configure
# SESSION_ENGINE to django.contrib.sessions.backends.signed_cookies
# as shown below:
#SESSION_ENGINE = 'django.contrib.sessions.backends.signed_cookies'

# Send email to the console by default
```

= apt-get install openstack

openstack-dashboar

◦ Configure the dashboard to use

OPENSTACK\_HOST = "cont

◦ Allow all hosts to access the das

ALLOWED\_HOSTS = ['\*'],

◦ Configure the memcached sessi

SESSION\_ENGINE = 'djang

```
CACHES = {
    'default': {
        'BACKEND': 'd
        'LOCATION': 'c
```

```

OPENSTACK_HOST = "controller"
OPENSTACK_KEYSTONE_URL = "http://%s/identity/v3" % OPENSTACK_HOST
OPENSTACK_KEYSTONE_MULTIDOMAIN_SUPPORT = True
OPENSTACK_API_VERSIONS = {
    "identity": 3,
    "image": 2,
    "volume": 2,
}

OPENSTACK_KEYSTONE_DEFAULT_DOMAIN = "default"
OPENSTACK_NEUTRON_NETWORK = {
    ...
    'enable_router': False,
    'enable_quotas': False,
    'enable_distributed_router': False,
    'enable_ha_router': False,
    'enable_lb': False,
    'enable_firewall': False,
    'enable_vpn': False,
    'enable_fip_topology_check': False,
}
# The timezone of the server. This should correspond with the timezone
# of your entire OpenStack installation, and hopefully be in UTC.
TIME_ZONE = "IST"

```

```

arvind@arvind-GF65-Thin-9SD:~$ service apache2 reload
==== AUTHENTICATING FOR org.freedesktop.systemd1.manage-units ====
Authentication is required to reload 'apache2.service'.
Authenticating as: arvind
Password:
==== AUTHENTICATION COMPLETE ====

```

**Set memcached config to accept any IP:**

```

# Run the daemon as root. The start-memcached will default to running
# -u command is present in this config file
-u memcache

# Specify which IP address to listen on. The default is to listen on all
# This parameter is one of the only security measures that memcached has
# it's listening on a firewalled interface.
-l 0.0.0.0

# Limit the number of simultaneous incoming connections. The daemon
# -c 1024

# Lock down all paged memory. Consult with the README and homepage for
# -k

# Return error when memory is exhausted (rather than removing items)
# -M

# Maximize core file limit

```

**Add current ip address as controller to /etc/hosts**

```

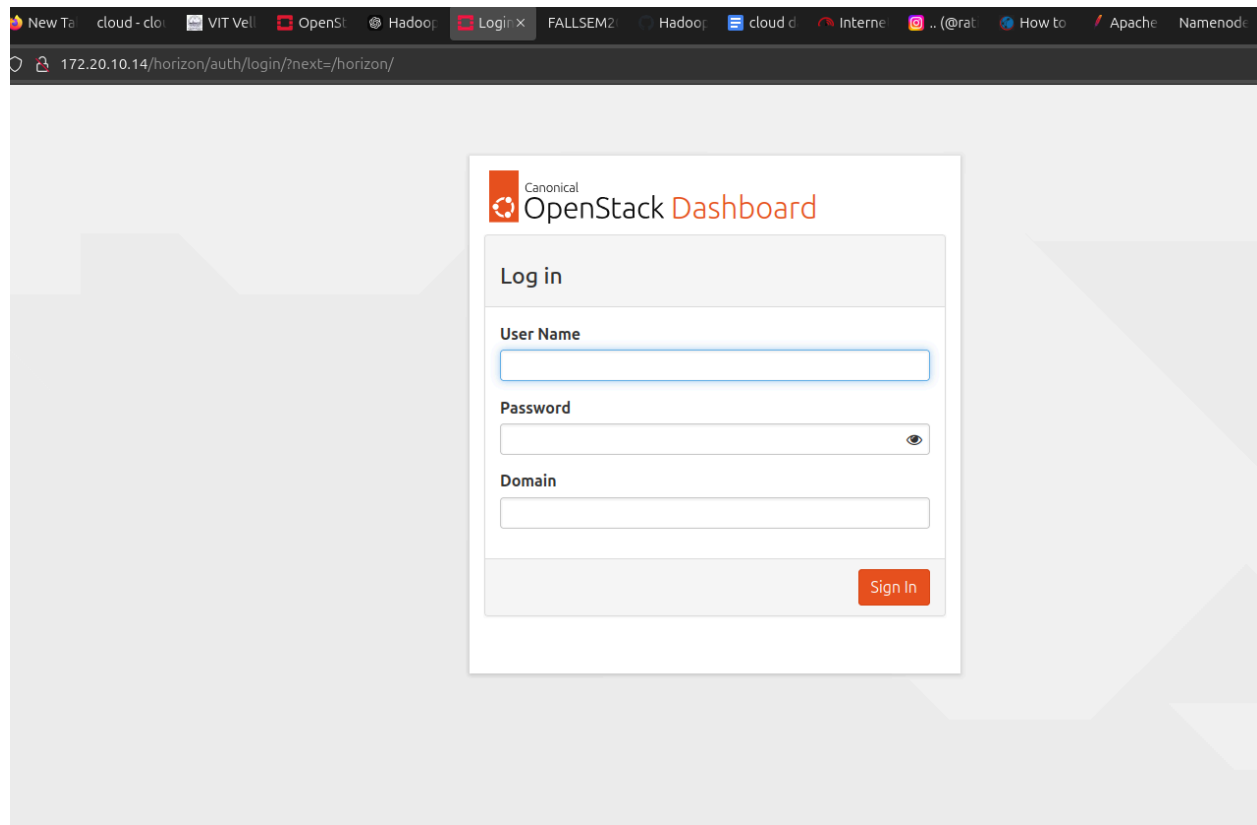
arvind@arvind-GF65-Thin-9SD: ~
GNU nano 7.2 /etc/hosts *
127.0.0.1 localhost
127.0.1.1 arvind-GF65-Thin-9SD
172.20.10.14 controller

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters

Yesterday
PaaS Cloud Provider Comparison
Missing index.html file
Hybrid Threading Model Overview
Previous 7 Days

```

## Access the dashboard at controller/horizon



The screenshot shows a web browser window with the address bar displaying `172.20.10.14/horizon/auth/login?next=/horizon/`. The browser's tab bar includes several open tabs: "New Ta...", "cloud - clo...", "VIT Vel...", "OpenSt...", "Hadoo...", "Login x", "FALLSEM2...", "Hadoo...", "cloud d...", "Interne...", "@rat...", "How to...", "Apache", and "Namenode...". The main content area displays the "OpenStack Dashboard" login interface. At the top left of the form is the Canonical OpenStack logo. Below the logo is a "Log in" heading. The form contains three input fields: "User Name" (with a blue border), "Password" (with a toggle eye icon), and "Domain". A red "Sign In" button is located at the bottom right of the form.

Canonical OpenStack Dashboard

Log in

User Name

Password

Domain

Sign In