

Labsheet-2

Map Reduce

Name : Anuvind M P

Roll : AM.EN.U4AIE22010

```
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

```
cd $HADOOP_HOME/sbin
```

```
./start-all.sh
```

Output :

```
anuvind@root:~/hadoop-3.4.0$ jps
16851 Jps
15734 NameNode
16136 SecondaryNameNode
16347 ResourceManager
15884 DataNode
16493 NodeManager
```

1. *hdfs dfs -mkdir /bigdata*

```
anuvind@root:~/hadoop-3.4.0/sbin$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - anuvind supergroup 0 2025-02-22 23:00 /bigdata
```

2. *hdfs dfs -touchz /bigdata/input*

```
anuvind@root:~/hadoop-3.4.0/sbin$ hdfs dfs -touchz /bigdata/input
anuvind@root:~/hadoop-3.4.0/sbin$ hdfs dfs -ls /bigdata
Found 1 items
-rw-r--r-- 1 anuvind supergroup 0 2025-02-22 23:21 /bigdata/input
```

3. *hdfs dfs -appendToFile -/bigdata /input*

aaaaa bbbb cc aaaa cc

cc bbbb zzzz ctrl+d ctrl+ d

```

anuvind@root:~/hadoop-3.4.0/sbin$ echo -e "aaaaa bbbb cc aaaaa cc\ncc bbbb zzzz" | hdfs dfs -appendToFile - /bigdata/input
anuvind@root:~/hadoop-3.4.0/sbin$ hdfs dfs -cat /bigdata/input
aaaaa bbbb cc aaaaa cc
cc bbbb zzzz

```

4. ***cd \$HADOOP_HOME***

5. ***ls***

6. ***cd share***

7. ***ls***

8. ***cd Hadoop***

9. ***ls***

10. ***cd mapreduce***

11. ***ls***

```

anuvind@root:~/hadoop-3.4.0/sbin$ cd $HADOOP_HOME
apreduce
ls
anuvind@root:~/hadoop-3.4.0$ ls
bin etc include lib libexec LICENSE-binary licenses-binary LICENSE.txt logs NOTICE-binary NOTICE.txt README.txt/sbin share
anuvind@root:~/hadoop-3.4.0$ cd share
anuvind@root:~/hadoop-3.4.0/share$ ls
doc hadoop
anuvind@root:~/hadoop-3.4.0/share$ cd hadoop
anuvind@root:~/hadoop-3.4.0/share/hadoop$ ls
client common hdfs mapreduce tools yarn
anuvind@root:~/hadoop-3.4.0/share/hadoop$ cd mapreduce
anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$ ls
hadoop-mapreduce-client-app-3.4.0.jar      hadoop-mapreduce-client-hs-plugins-3.4.0.jar      hadoop-mapreduce-client-shuffle-3.4.0.jar      sources
hadoop-mapreduce-client-common-3.4.0.jar   hadoop-mapreduce-client-jobclient-3.4.0.jar      hadoop-mapreduce-client-uploader-3.4.0.jar
hadoop-mapreduce-client-core-3.4.0.jar     hadoop-mapreduce-client-jobclient-3.4.0-tests.jar hadoop-mapreduce-examples-3.4.0.jar
hadoop-mapreduce-client-hs-3.4.0.jar       hadoop-mapreduce-client-native-task-3.4.0.jar     jdiff
anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$ |

```

13. ***hadoop jar hadoop-mapreduce-examples-3.3.0.jar wordcount /bigdata /out***

```

anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.4.0.jar wordcount /bigdata /out
2025-02-22 23:27:07,154 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-02-22 23:27:07,783 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/anuvind/.staging/job_1740246489276_0001
2025-02-22 23:27:08,199 INFO input.FileInputFormat: Total input files to process : 1
2025-02-22 23:27:09,120 INFO mapreduce.JobSubmitter: number of splits:1
2025-02-22 23:27:09,314 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1740246489276_0001
2025-02-22 23:27:09,314 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-02-22 23:27:09,574 INFO conf.Configuration: resource-types.xml not found
2025-02-22 23:27:09,574 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'
2025-02-22 23:27:10,709 INFO impl.YarnClientImpl: Submitted application application_1740246489276_0001
2025-02-22 23:27:10,762 INFO mapreduce.Job: The url to track the job: http://root:8088/proxy/application_1740246489276_0001/
2025-02-22 23:27:10,762 INFO mapreduce.Job: Running job: job_1740246489276_0001
2025-02-22 23:27:24,098 INFO mapreduce.Job: Job job_1740246489276_0001 running in uber mode : false
2025-02-22 23:27:24,100 INFO mapreduce.Job: map 0% reduce 0%
2025-02-22 23:27:35,218 INFO mapreduce.Job: map 100% reduce 0%
2025-02-22 23:27:39,271 INFO mapreduce.Job: map 100% reduce 100%
2025-02-22 23:27:40,305 INFO mapreduce.Job: Job job_1740246489276_0001 completed successfully
2025-02-22 23:27:40,442 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=617183
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=136
  HDFS: Number of bytes written=27
  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)=6587
  Total time spent by all reduces in occupied slots (ms)=4021
  Total time spent by all map tasks (ms)=6587
  Total time spent by all reduce tasks (ms)=4021
  Total vcore-milliseconds taken by all map tasks=6587
  Total vcore-milliseconds taken by all reduce tasks=4021
  Total megabyte-milliseconds taken by all map tasks=6745088
  Total megabyte-milliseconds taken by all reduce tasks=4117504
Map-Reduce Framework
  Map input records=2
  Map output records=8
  Map output bytes=68
  Map output materialized bytes=49
  Input split bytes=100
  Combine input records=8
  Combine output records=4
  Reduce input groups=4
  Reduce shuffle bytes=49

```

```

Reduce shuffle bytes=49
Reduce input records=4
Reduce output records=4
Spilled Records=8
Shuffled Maps=1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=3063
CPU time spent (ms)=22100
Physical memory (bytes) snapshot=608637440
Virtual memory (bytes) snapshot=5122162688
Total committed heap usage (bytes)=608676928
Peak Map Physical memory (bytes)=365277184
Peak Map Virtual memory (bytes)=2556772352
Peak Reduce Physical memory (bytes)=235360256
Peak Reduce Virtual memory (bytes)=2565390336

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=36
File Output Format Counters
  Bytes Written=27

```

14. `hdfs dfs -ls /out`

```

anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$ hdfs dfs -ls /out
Found 2 items
-rw-r--r--  1 anuvind supergroup          0 2025-02-22 23:27 /out/_SUCCESS
-rw-r--r--  1 anuvind supergroup        27 2025-02-22 23:27 /out/part-r-00000
anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$


```

15. `hdfs dfs -cat /out/part*`

```

anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$ hdfs dfs -cat /out/part*
aaaaa  2
bbbb   2
cc      3
zzzz   1
anuvind@root:~/hadoop-3.4.0/share/hadoop/mapreduce$

```



ResourceManager

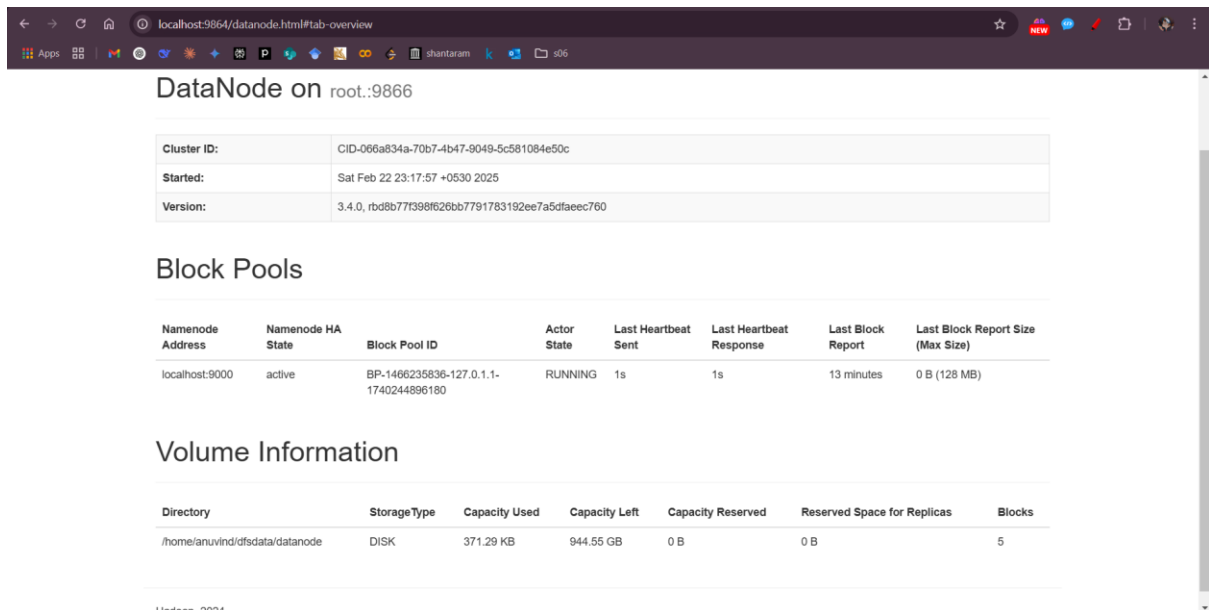
NodeManager

Node Information
List of Applications
List of Containers

















Tools

NodeManager information

Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total VCores allocated for Containers	8
Resource types	memory-mb (unit=Mi), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Sat Feb 22 23:30:09 IST 2025
NodeHealthReport	
NodeManager started on	Sat Feb 22 23:18:07 IST 2025
NodeManager Version:	3.4.0 from bd8b77f398f626bb7791783192ee7a5dfaeec760 by root source checksum 934da0c5743762b7851c0ff9f8ca2 on 2024-03-04T06:54Z
Hadoop Version:	3.4.0 from bd8b77f398f626bb7791783192ee7a5dfaeec760 by root source checksum f7fe694a3613358b38812ae9c31114e on 2024-03-04T06:35Z



Browse Directory

/										Go!							
Show	25	entries											Search:				
<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		drwxr-xr-x		anuvind		supergroup		0 B		Feb 22 23:21		0		0 B		bigdata	
<input type="checkbox"/>		drwxr-xr-x		anuvind		supergroup		0 B		Feb 22 23:27		0		0 B		out	
<input type="checkbox"/>		drwxr-xr-x		anuvind		supergroup		0 B		Feb 22 23:27		0		0 B		tmp	

Hadoop, 2024.

In operation

DataNode State

All

Show 25 entries

Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Block pool usage StdDev	Version
<div>✓/default-rack/root-.9866 (127.0.0.1:9866)</div>	http://root-.9864	1s	18m	371.29 KB	11.09 GB	<div>1006.85 GB</div> <div></div>	5	371.29 KB (0%)	0%	3.4.0

Showing 1 to 1 of 1 entries

Previous

1

Next