

[Check for Java version]

```
Hadoop@nanobot-ubuntu:~$ java -version
```

```
openjdk version "11.0.2" 2019-01-15
OpenJDK Runtime Environment 18.9 (build 11.0.2+9)
OpenJDK 64-Bit Server VM 18.9 (build 11.0.2+9, mixed mode)
```

[Check for Hadoop version]

```
hadoop@nanobot-ubuntu:~$ hadoop version
```

```
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r
1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /home/hadoop/hadoop/share/hadoop/common/hadoop-
common-3.3.6.jar
```

[Localhost]

```
hadoop@nanobot-ubuntu:~$ ssh localhost
```

Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.5.0-28-generic x86\_64)

- \* Documentation: <https://help.ubuntu.com>
- \* Management: <https://landscape.canonical.com>
- \* Support: <https://ubuntu.com/pro>

Expanded Security Maintenance for Applications is enabled.

11 updates can be applied immediately.  
9 of these updates are standard security updates.  
To see these additional updates run: `apt list --upgradable`

Failed to connect to <https://changelogs.ubuntu.com/meta-release-lts>. Check your Internet connection or proxy settings

Last login: Tue Apr 30 23:29:05 2024 from 127.0.0.1

[Start]

```
hadoop@nanobot-ubuntu:~$ 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]  
Starting datanodes  
Starting secondary namenodes [nanobot-ubuntu]  
Starting resourcemanager  
Starting nodemanagers

[Check if all **6** nodes have been started]

hadoop@nanobot-ubuntu:~\$ jps

3376 SecondaryNameNode  
4114 Jps  
3140 DataNode  
3592 ResourceManager  
2937 NameNode  
3770 NodeManager

[Make (**user**)directory for programming]

hadoop@nanobot-ubuntu:~\$ hdfs dfs -mkdir /salman

[make input file (text) containing words for **wordcount**]

hadoop@nanobot-ubuntu:~\$ nano shera.txt

[put created file in another directory inside previous one , you just need to write this command  
(**input**)directory creation happens inside automatically]

hadoop@nanobot-ubuntu:~\$ hdfs dfs -put shera.txt /salman/arbaaz

[change directory to **hadoop**]

hadoop@nanobot-ubuntu:~\$ cd hadoop

[Look for **share**]

hadoop@nanobot-ubuntu:~/hadoop\$ ls

bin include libexec licenses-binary logs NOTICE-binary README.txt sbin  
etc lib LICENSE-binary LICENSE.txt mydata.txt NOTICE.txt rtest share

[change directory to **share**]

```
hadoop@nanobot-ubuntu:~/hadoop$ cd share
```

[Look for **hadoop**]

```
hadoop@nanobot-ubuntu:~/hadoop/share$ ls
```

```
doc  hadoop
```

[change directory to **hadoop**]

```
hadoop@nanobot-ubuntu:~/hadoop/share$ cd hadoop
```

[Look for **mapreduce**]

```
hadoop@nanobot-ubuntu:~/hadoop/share/hadoop$ ls
```

```
client  common  hdfs  mapreduce  tools  yarn
```

[change directory to **mapreduce**]

```
hadoop@nanobot-ubuntu:~/hadoop/share/hadoop$ cd mapreduce
```

[Look for **hadoop-mapreduce-examples-3.3.6.jar**]

```
hadoop@nanobot-ubuntu:~/hadoop/share/hadoop/mapreduce$ ls
```

```
hadoop-mapreduce-client-app-3.3.6.jar  hadoop-mapreduce-client-hs-plugins-3.3.6.jar  
hadoop-mapreduce-client-shuffle-3.3.6.jar  lib-examples  
hadoop-mapreduce-client-common-3.3.6.jar  hadoop-mapreduce-client-jobclient-3.3.6.jar  
hadoop-mapreduce-client-uploader-3.3.6.jar  sources  
hadoop-mapreduce-client-core-3.3.6.jar  hadoop-mapreduce-client-jobclient-3.3.6-tests.jar  
hadoop-mapreduce-examples-3.3.6.jar  
hadoop-mapreduce-client-hs-3.3.6.jar  hadoop-mapreduce-client-nativetask-3.3.6.jar  jdiff
```

[come outside of the directories]

```
hadoop@nanobot-ubuntu:~/hadoop/share/hadoop/mapreduce$ cd
```

[run command to the path of **hadoop-mapreduce-examples-3.3.6.jar** that is obtained by above steps followed by **/user/inputfolder /user/outputfolder** creation will happen inside]

```
hadoop@nanobot-ubuntu:~$ hadoop jar ~/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar wordcount /salman/arbaaz /salman/sohail
```

```
2024-04-30 23:39:24,121 INFO impl.MetricsConfig: Loaded properties from hadoop-
metrics2.properties
2024-04-30 23:39:24,191 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10
second(s).
2024-04-30 23:39:24,191 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-30 23:39:24,368 INFO input.FileInputFormat: Total input files to process : 1
2024-04-30 23:39:24,419 INFO mapreduce.JobSubmitter: number of splits:1
2024-04-30 23:39:24,498 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_local1325504371_0001
2024-04-30 23:39:24,498 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-04-30 23:39:24,578 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-04-30 23:39:24,578 INFO mapreduce.Job: Running job: job_local1325504371_0001
2024-04-30 23:39:24,579 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-04-30 23:39:24,583 INFO output.PathOutputCommitterFactory: No output committer factory
defined, defaulting to FileOutputCommitterFactory
2024-04-30 23:39:24,584 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 2
2024-04-30 23:39:24,584 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup
_temporary folders under output directory:false, ignore cleanup failures: false
2024-04-30 23:39:24,584 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-04-30 23:39:24,613 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-04-30 23:39:24,613 INFO mapred.LocalJobRunner: Starting task:
attempt_local1325504371_0001_m_000000_0
2024-04-30 23:39:24,627 INFO output.PathOutputCommitterFactory: No output committer factory
defined, defaulting to FileOutputCommitterFactory
2024-04-30 23:39:24,627 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 2
2024-04-30 23:39:24,627 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup
_temporary folders under output directory:false, ignore cleanup failures: false
2024-04-30 23:39:24,640 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-04-30 23:39:24,644 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/salman/
arbaaz:0+87
2024-04-30 23:39:24,722 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-04-30 23:39:24,722 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-04-30 23:39:24,722 INFO mapred.MapTask: soft limit at 83886080
2024-04-30 23:39:24,722 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-04-30 23:39:24,722 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-04-30 23:39:24,726 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
```

2024-04-30 23:39:24,958 INFO mapred.LocalJobRunner:  
2024-04-30 23:39:24,960 INFO mapred.MapTask: Starting flush of map output  
2024-04-30 23:39:24,960 INFO mapred.MapTask: Spilling map output  
2024-04-30 23:39:24,960 INFO mapred.MapTask: bufstart = 0; bufend = 167; bufvoid = 104857600  
2024-04-30 23:39:24,960 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600  
2024-04-30 23:39:24,968 INFO mapred.MapTask: Finished spill 0  
2024-04-30 23:39:24,975 INFO mapred.Task: Task:attempt\_local1325504371\_0001\_m\_000000\_0 is done. And is in the process of committing  
2024-04-30 23:39:24,978 INFO mapred.LocalJobRunner: map  
2024-04-30 23:39:24,978 INFO mapred.Task: Task 'attempt\_local1325504371\_0001\_m\_000000\_0' done.  
2024-04-30 23:39:24,982 INFO mapred.Task: Final Counters for attempt\_local1325504371\_0001\_m\_000000\_0: Counters: 24  
File System Counters  
FILE: Number of bytes read=281507  
FILE: Number of bytes written=925380  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=87  
HDFS: Number of bytes written=0  
HDFS: Number of read operations=5  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=1  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Map input records=1  
Map output records=20  
Map output bytes=167  
Map output materialized bytes=153  
Input split bytes=100  
Combine input records=20  
Combine output records=14  
Spilled Records=14  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=154  
Total committed heap usage (bytes)=639107072  
File Input Format Counters  
Bytes Read=87  
2024-04-30 23:39:24,982 INFO mapred.LocalJobRunner: Finishing task: attempt\_local1325504371\_0001\_m\_000000\_0  
2024-04-30 23:39:24,982 INFO mapred.LocalJobRunner: map task executor complete.  
2024-04-30 23:39:24,984 INFO mapred.LocalJobRunner: Waiting for reduce tasks  
2024-04-30 23:39:24,984 INFO mapred.LocalJobRunner: Starting task: attempt\_local1325504371\_0001\_r\_000000\_0  
2024-04-30 23:39:24,991 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory  
2024-04-30 23:39:24,991 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

2024-04-30 23:39:24,991 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup  
\_temporary folders under output directory:false, ignore cleanup failures: false  
2024-04-30 23:39:24,991 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]  
2024-04-30 23:39:24,994 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:  
org.apache.hadoop.mapreduce.task.reduce.Shuffle@69d916f3  
2024-04-30 23:39:24,995 WARN impl.MetricsSystemImpl: JobTracker metrics system already  
initialized!  
2024-04-30 23:39:25,006 INFO reduce.MergeManagerImpl: MergeManager:  
memoryLimit=2497812736, maxSingleShuffleLimit=624453184, mergeThreshold=1648556416,  
ioSortFactor=10, memToMemMergeOutputsThreshold=10  
2024-04-30 23:39:25,007 INFO reduce.EventFetcher: attempt\_local1325504371\_0001\_r\_000000\_0  
Thread started: EventFetcher for fetching Map Completion Events  
2024-04-30 23:39:25,026 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map  
attempt\_local1325504371\_0001\_m\_000000\_0 decomp: 149 len: 153 to MEMORY  
2024-04-30 23:39:25,029 INFO reduce.InMemoryMapOutput: Read 149 bytes from map-output for  
attempt\_local1325504371\_0001\_m\_000000\_0  
2024-04-30 23:39:25,030 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of  
size: 149, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory -> 149  
2024-04-30 23:39:25,031 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning  
2024-04-30 23:39:25,032 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2024-04-30 23:39:25,032 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory  
map-outputs and 0 on-disk map-outputs  
2024-04-30 23:39:25,036 INFO mapred.Merger: Merging 1 sorted segments  
2024-04-30 23:39:25,036 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left  
of total size: 145 bytes  
2024-04-30 23:39:25,037 INFO reduce.MergeManagerImpl: Merged 1 segments, 149 bytes to disk  
to satisfy reduce memory limit  
2024-04-30 23:39:25,037 INFO reduce.MergeManagerImpl: Merging 1 files, 153 bytes from disk  
2024-04-30 23:39:25,038 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from  
memory into reduce  
2024-04-30 23:39:25,038 INFO mapred.Merger: Merging 1 sorted segments  
2024-04-30 23:39:25,038 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left  
of total size: 145 bytes  
2024-04-30 23:39:25,039 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2024-04-30 23:39:25,059 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead,  
use mapreduce.job.skiprecords  
2024-04-30 23:39:25,094 INFO mapred.Task: Task:attempt\_local1325504371\_0001\_r\_000000\_0 is  
done. And is in the process of committing  
2024-04-30 23:39:25,096 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2024-04-30 23:39:25,096 INFO mapred.Task: Task attempt\_local1325504371\_0001\_r\_000000\_0 is  
allowed to commit now  
2024-04-30 23:39:25,114 INFO output.FileOutputCommitter: Saved output of task  
'attempt\_local1325504371\_0001\_r\_000000\_0' to hdfs://localhost:9000/salman/sohail  
2024-04-30 23:39:25,115 INFO mapred.LocalJobRunner: reduce > reduce  
2024-04-30 23:39:25,115 INFO mapred.Task: Task 'attempt\_local1325504371\_0001\_r\_000000\_0'  
done.  
2024-04-30 23:39:25,115 INFO mapred.Task: Final Counters for  
attempt\_local1325504371\_0001\_r\_000000\_0: Counters: 30  
File System Counters  
FILE: Number of bytes read=281845  
FILE: Number of bytes written=925533  
FILE: Number of read operations=0

FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=87  
HDFS: Number of bytes written=91  
HDFS: Number of read operations=10  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=3  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Combine input records=0  
Combine output records=0  
Reduce input groups=14  
Reduce shuffle bytes=153  
Reduce input records=14  
Reduce output records=14  
Spilled Records=14  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=0  
Total committed heap usage (bytes)=639107072  
Shuffle Errors  
BAD\_ID=0  
CONNECTION=0  
IO\_ERROR=0  
WRONG\_LENGTH=0  
WRONG\_MAP=0  
WRONG\_REDUCE=0  
File Output Format Counters  
Bytes Written=91  
2024-04-30 23:39:25,116 INFO mapred.LocalJobRunner: Finishing task:  
attempt\_local1325504371\_0001\_r\_000000\_0  
2024-04-30 23:39:25,116 INFO mapred.LocalJobRunner: reduce task executor complete.  
2024-04-30 23:39:25,582 INFO mapreduce.Job: Job job\_local1325504371\_0001 running in uber  
mode : false  
2024-04-30 23:39:25,583 INFO mapreduce.Job: map 100% reduce 100%  
2024-04-30 23:39:25,584 INFO mapreduce.Job: Job job\_local1325504371\_0001 completed  
successfully  
2024-04-30 23:39:25,589 INFO mapreduce.Job: Counters: 36  
File System Counters  
FILE: Number of bytes read=563352  
FILE: Number of bytes written=1850913  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=174  
HDFS: Number of bytes written=91  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=4  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework

Map input records=1  
Map output records=20  
Map output bytes=167  
Map output materialized bytes=153  
Input split bytes=100  
Combine input records=20  
Combine output records=14  
Reduce input groups=14  
Reduce shuffle bytes=153  
Reduce input records=14  
Reduce output records=14  
Spilled Records=28  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=154  
Total committed heap usage (bytes)=1278214144  
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=87  
File Output Format Counters  
Bytes Written=91

[Check if our (**user**) directory **salman** was formed or not]

hadoop@nanobot-ubuntu:~\$ hdfs dfs -ls /

Found 11 items

drwxr-xr-x	- hadoop supergroup	0 2024-04-29 23:35 /32251
drwxr-xr-x	- hadoop supergroup	0 2024-04-29 00:55 /dan
drwxr-xr-x	- hadoop supergroup	0 2024-04-28 16:33 /input
drwxr-xr-x	- hadoop supergroup	0 2024-04-24 12:00 /logs
drwxr-xr-x	- hadoop supergroup	0 2024-04-29 22:45 /lord
drwxr-xr-x	- hadoop supergroup	0 2024-04-30 23:39 /salman
drwxr-xr-x	- hadoop supergroup	0 2024-04-24 12:00 /test1
drwxr-xr-x	- hadoop supergroup	0 2024-04-25 12:09 /user
drwxr-xr-x	- hadoop supergroup	0 2024-04-25 11:08 /user1
drwxr-xr-x	- hadoop supergroup	0 2024-04-25 11:28 /user39
drwxr-xr-x	- hadoop supergroup	0 2024-04-25 22:22 /user40



[Write following command for output ]

```
hadoop@nanobot-ubuntu:~$ hdfs dfs -cat /salman/sohail/part-r-00000
```

```
, 1  
8436 1  
arbaaz 1  
as 1  
driver 2  
is 1  
sohail 1  
of 3  
one 1  
otp 2  
the 3  
this 1  
was 1  
what 1
```

**shera.txt**

arbaaz sohail what is the otp of this driver , as the otp of one of the driver was 8436