BANASTHALI VIDYAPITH

BIG DATA ANALYTICS PRACTICAL FILE

HDFS, WORD COUNT AND HBASE COMMANDS

NAME: Sneha Baser

CLASS: B. TECH VIIIth

Semester SEC: C

ROLL NO.: 1812883

HDFS COMMANDS

1. $hadoop version

```
cloudera@quickstart:~
[cloudera@quickstart ~]$ hadoop version
Hadoop 2.6.0-cdh5.13.0
Subversion http://github.com/cloudera/hadoop -r 42e8860b182e55321bd5f5605264da4adc8882be
Compiled by jenkins on 2017-10-04T18:08Z
Compiled with protoc 2.5.0
From source with checksum 5e84c185f8a22158e2b0e4b8f85311
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.13.0.jar
[cloudera@quickstart ~]$
```

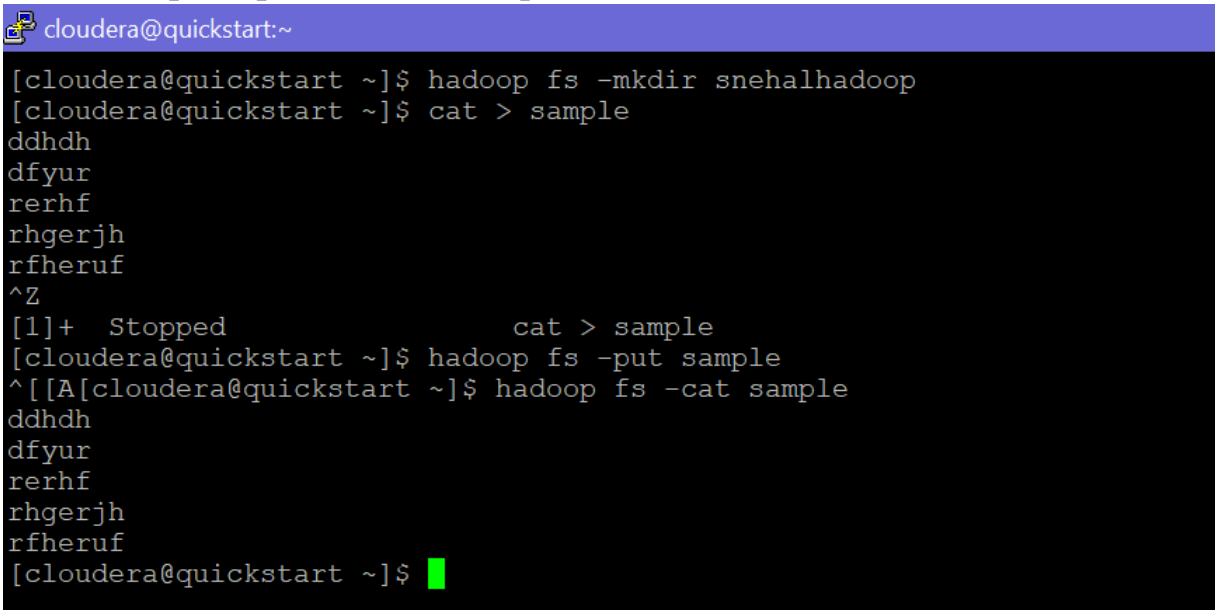
2. $hadoop fs -ls

```
[cloudera@quickstart ~]$ hadoop fs -ls
Found 514 items
-rw-r--r--   1 cloudera cloudera          0 2021-11-30 18:37 1946282_ananya
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 02:09 2046281_lav
drwxr-xr-x   - cloudera cloudera          0 2021-11-30 17:58 26_wordcount
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 00:56 AditiPal
-rw-r--r--   1 cloudera cloudera        105 2021-12-01 00:54 Annu
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:35 Annu1
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:42 Annu2
-rw-r--r--   1 cloudera cloudera        105 2021-12-01 01:05 Anu
-rw-r--r--   1 cloudera cloudera         28 2021-11-30 08:15 Asd.txt
drwxr-xr-x   - cloudera cloudera          0 2021-10-29 00:49 D1
drwxr-xr-x   - cloudera cloudera          0 2022-01-26 22:47 Dir1
-rw-r--r--   1 cloudera cloudera         33 2022-01-26 23:08 HDFS
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:26 K1
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:35 K2
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:44 Laxita2
-rw-r--r--   1 cloudera cloudera         42 2021-12-01 01:39 Ni.txt
-rw-r--r--   1 cloudera cloudera         65 2021-08-26 02:02 Nidhi
-rw-r--r--   1 cloudera cloudera         70 2021-12-18 20:18 Nidhi.txt
-rw-r--r--   1 cloudera cloudera         42 2021-12-01 01:48 Nidhi_001.txt
-rw-r--r--   1 cloudera cloudera         99 2021-12-01 01:44 Nidhi_01.txt
drwxr-xr-x   - cloudera cloudera          0 2021-12-18 20:26 Nidhi_19
-rw-r--r--   1 cloudera cloudera        100 2021-12-01 00:58 Nidhi_table_01
drwxr-xr-x   - cloudera cloudera          0 2021-09-29 05:18 Nidhi_wcount
drwxr-xr-x   - cloudera cloudera          0 2022-01-28 20:57 Noor
drwxr-xr-x   - cloudera cloudera          0 2021-12-03 02:29 Out_01
drwxr-xr-x   - cloudera cloudera          0 2021-12-03 02:39 Out_1
drwxr-xr-x   - cloudera cloudera          0 2021-12-11 10:18 Output12
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:35 Pallavi1
drwxr-xr-x   - cloudera cloudera          0 2021-11-17 00:59 Pallavi_PigD
-rw-r--r--   1 cloudera cloudera          0 2021-10-28 22:41 Pallavi_TableData
drwxr-xr-x   - cloudera cloudera          0 2021-10-28 23:20 Pallavi_Tiwari
drwxr-xr-x   - cloudera cloudera          0 2021-09-04 02:14 Pari
drwxr-xr-x   - cloudera cloudera          0 2021-12-03 04:27 PigData_sneha
drwxr-xr-x   - cloudera cloudera          0 2021-12-19 00:48 PigOutput
drwxr-xr-x   - cloudera cloudera          0 2021-11-20 01:16 Pig_Output
drwxr-xr-x   - cloudera cloudera          0 2021-11-26 06:24 Pig_Output_nishtha
drwxr-xr-x   - cloudera cloudera          0 2021-12-19 00:56 Pig_Outputt
drwxr-xr-x   - cloudera cloudera          0 2021-12-04 01:24 Pig_Outputy
drwxr-xr-x   - cloudera cloudera          0 2021-12-03 23:38 Pig_YOutput
drwxr-xr-x   - cloudera cloudera          0 2021-11-26 08:26 Pigs_Output
drwxr-xr-x   - cloudera cloudera          0 2021-12-01 01:40 Prashansha
drwxr-xr-x   - cloudera cloudera          0 2021-10-29 01:10 Q1
```

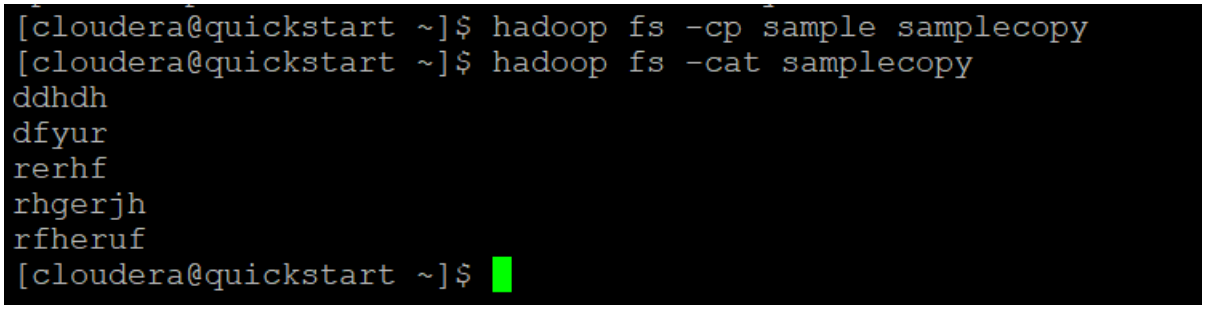
3. $hadoop fs -mkdir

```
cloudera@quickstart:~
[cloudera@quickstart ~]$ hadoop fs -mkdir snehalhadoop
[cloudera@quickstart ~]$
```

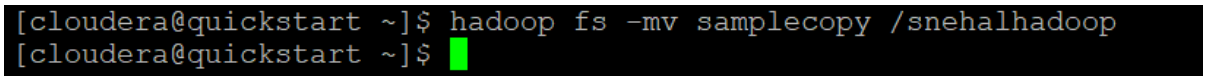
4. $hadoop fs -put and $hadoop fs -cat



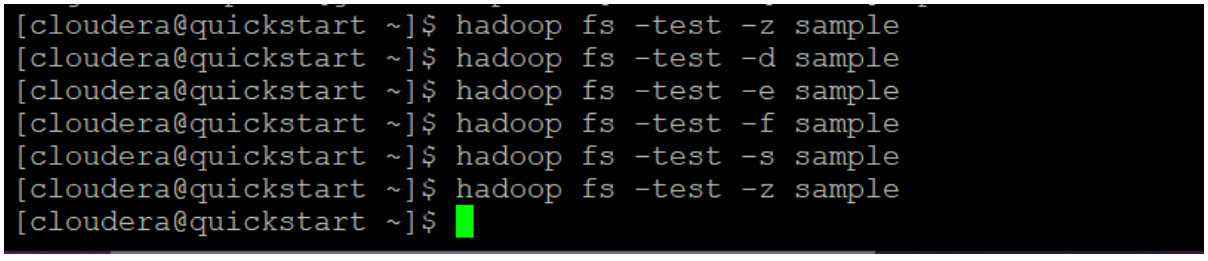
5. $hadoop fs -cp



6. $hadoop fs -mv



7. $hadoop fs -test



8. $hadoop fs -stat

```
[cloudera@quickstart ~]$ hadoop fs -stat sample
2022-02-20 18:44:03
[cloudera@quickstart ~]$ hadoop fs -stat %b  sample
34
[cloudera@quickstart ~]$ hadoop fs -stat %g sample
cloudera
[cloudera@quickstart ~]$ hadoop fs -stat %n sample
sample
[cloudera@quickstart ~]$ hadoop fs -stat %o sample
134217728
[cloudera@quickstart ~]$ hadoop fs -stat %r sample
1
[cloudera@quickstart ~]$ hadoop fs -stat %u sample
cloudera
[cloudera@quickstart ~]$ hadoop fs -stat %y sample
2022-02-20 18:44:03
[cloudera@quickstart ~]$
```

9. $hadoop get and copyToLocal

```
[cloudera@quickstart ~]$ cd shreyaJ/HDFS/
[cloudera@quickstart HDFS]$ cat > file2.txt
hello
new file
ok bye
^Z
[2]+  Stopped                 cat > file2.txt
[cloudera@quickstart HDFS]$ hadoop fs -copyFromLocal file2.txt shreyaJ/
[cloudera@quickstart HDFS]$ hadoop fs -cat shreyaJ/file2.txt
hello
new file
ok bye
[cloudera@quickstart HDFS]$ cd ..
[cloudera@quickstart shreyaJ]$ mkdir copyToLocal
[cloudera@quickstart shreyaJ]$ ls
aai  copyToLocal  firsti  grepfiles  hi  HDFS  n1  n2  newfile
[cloudera@quickstart shreyaJ]$ mv copyToLocal/ cplocal
[cloudera@quickstart shreyaJ]$ ls
aai  cplocal  firsti  grepfiles  hi  HDFS  n1  n2  newfile
[cloudera@quickstart shreyaJ]$ hadoop fs -get shreyaJ/file2.txt cplocal/
[cloudera@quickstart shreyaJ]$ hadoop fs -copyToLocal shreyaJ/file1.txt cplocal/
[cloudera@quickstart shreyaJ]$ cd cplocal/
[cloudera@quickstart cplocal]$ ls
file1.txt  file2.txt
[cloudera@quickstart cplocal]$ cat file1.txt
hi
HDFS commands lab1
bye
[cloudera@quickstart cplocal]$ cat file2.txt
hello
new file
ok bye
[cloudera@quickstart cplocal]$ cd ../..
[cloudera@quickstart ~]$ cd shreyaJ
[cloudera@quickstart shreyaJ]$ hadoop fs -mkdir shreyaJnew
[cloudera@quickstart shreyaJ]$ hadoop fs -cp shreyaJ/file1.txt shreyaJnew
[cloudera@quickstart shreyaJ]$ hadoop fs -cat shreyaJnew/file1.txt
hi
HDFS commands lab1
bye
```

10.$hadoop rm

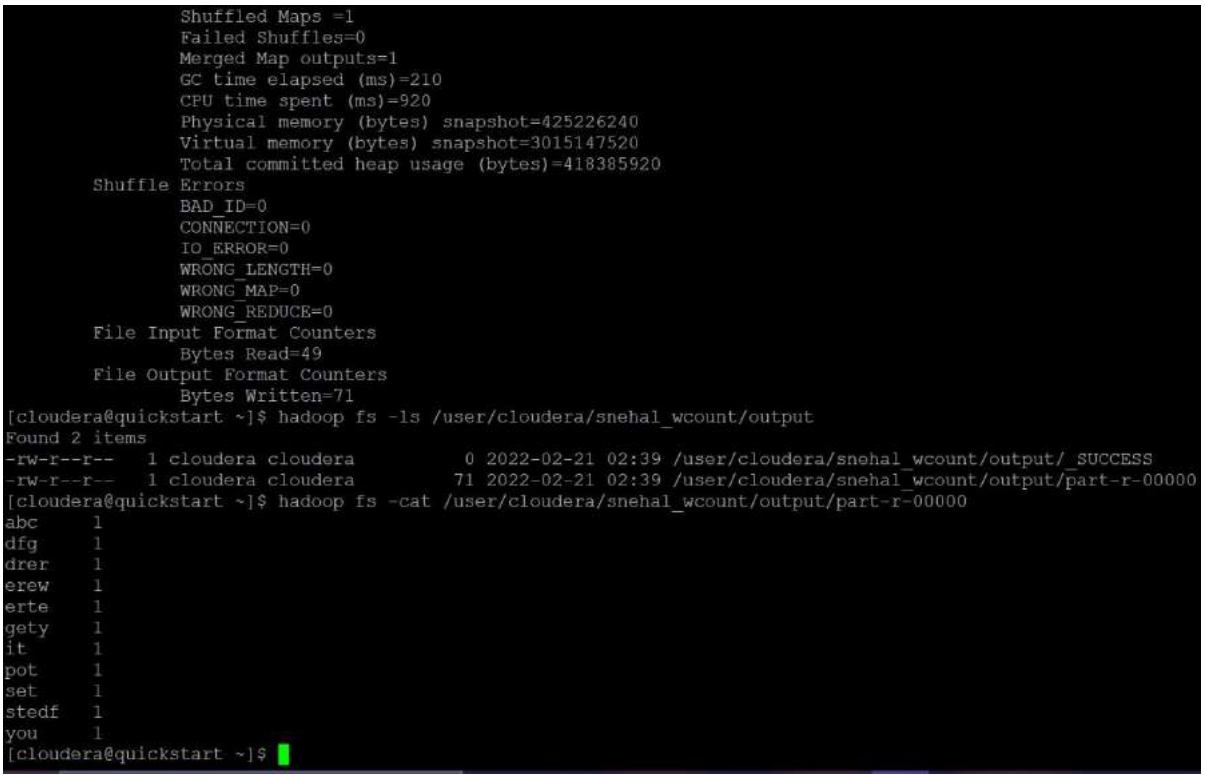
```
[cloudera@quickstart shreyaJ]$ hadoop fs -rm zzzz2
rm: `zzzz2': Is a directory
[cloudera@quickstart shreyaJ]$ hadoop fs -rm shreyaJnew/file1.txt
Deleted shreyaJnew/file1.txt
```

WORD COUNT

```
[cloudera@quickstart ~]$ hadoop fs -mkdir snehal_wcount
[cloudera@quickstart ~]$ cat > nfile
abc
dfg
drer
erew
erte
stedf
set
gety
you
it
bot
^Z
[1]+  Stopped                 cat > nfile
[cloudera@quickstart ~]$ hadoop fs -put n1

[cloudera@quickstart ~]$ hadoop fs -put nfile /user/cloudera/snehal_wcount/input
[cloudera@quickstart ~]$ hadoop jar wordcount.jar org.myorg.WordCount /user/cloudera/snehal_wcount/
utput
22/02/21 02:39:01 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/21 02:39:02 INFO input.FileInputFormat: Total input paths to process : 1
22/02/21 02:39:02 INFO mapreduce.JobSubmitter: number of splits:1
22/02/21 02:39:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1645434840893_0002
22/02/21 02:39:03 INFO impl.YarnClientImpl: Submitted application application_1645434840893_0002
22/02/21 02:39:03 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/pro
22/02/21 02:39:03 INFO mapreduce.Job: Running job: job_1645434840893_0002
22/02/21 02:39:12 INFO mapreduce.Job: Job job_1645434840893_0002 running in uber mode : false
22/02/21 02:39:12 INFO mapreduce.Job:  map 0% reduce 0%
22/02/21 02:39:22 INFO mapreduce.Job:  map 100% reduce 0%
22/02/21 02:39:31 INFO mapreduce.Job:  map 100% reduce 100%
22/02/21 02:39:31 INFO mapreduce.Job: Job job_1645434840893_0002 completed successfully
22/02/21 02:39:31 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=121
                FILE: Number of bytes written=287015
```

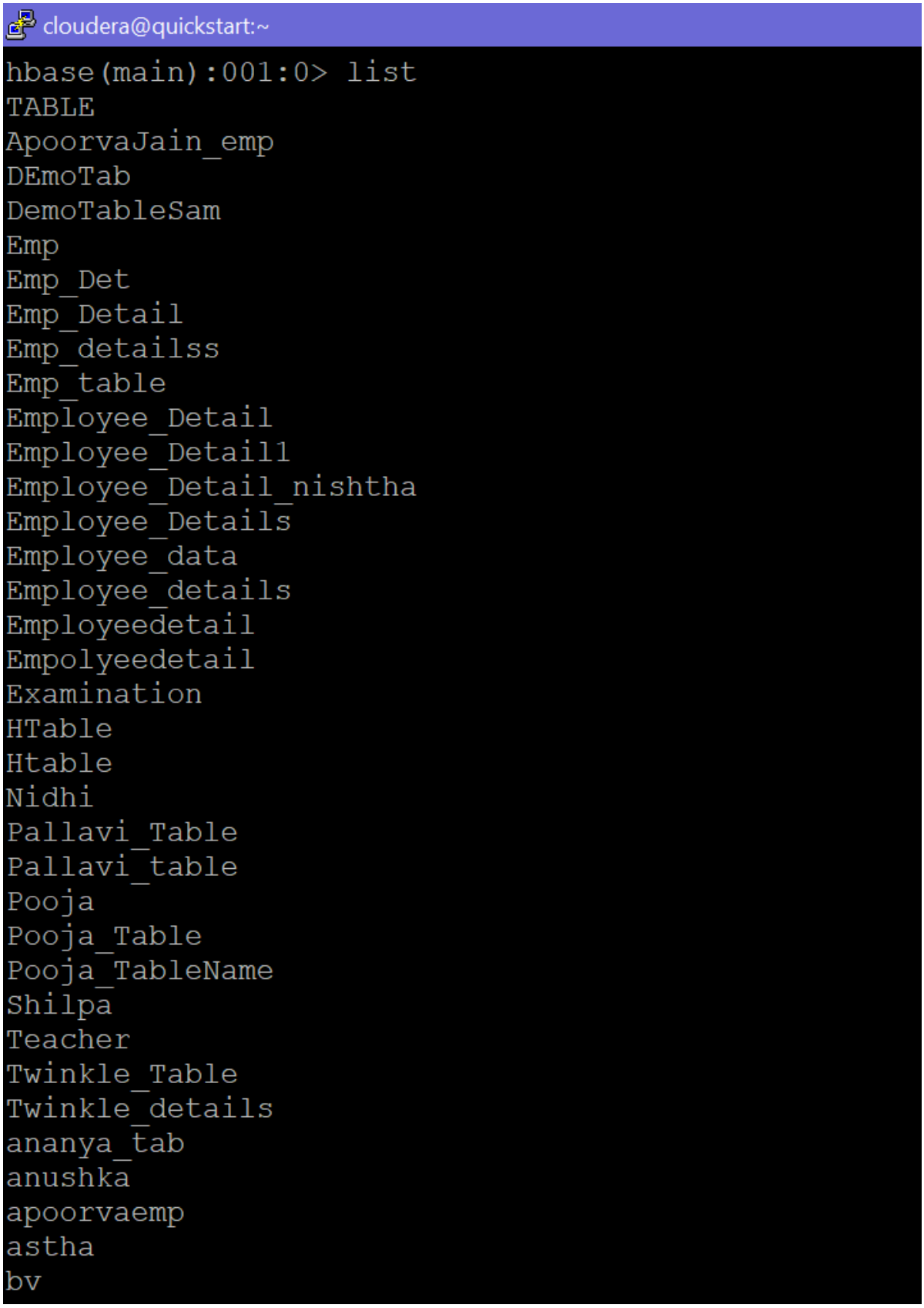
```
        File System Counters
                FILE: Number of bytes read=121
                FILE: Number of bytes written=287015
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=179
                HDFS: Number of bytes written=71
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        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)=5558
                Total time spent by all reduces in occupied slots (ms)=5404
                Total time spent by all map tasks (ms)=5558
                Total time spent by all reduce tasks (ms)=5404
                Total vcore-milliseconds taken by all map tasks=5558
                Total vcore-milliseconds taken by all reduce tasks=5404
                Total megabyte-milliseconds taken by all map tasks=5691392
                Total megabyte-milliseconds taken by all reduce tasks=5533696
        Map-Reduce Framework
                Map input records=11
                Map output records=11
                Map output bytes=93
                Map output materialized bytes=121
                Input split bytes=130
                Combine input records=0
                Combine output records=0
                Reduce input groups=11
                Reduce shuffle bytes=121
                Reduce input records=11
                Reduce output records=11
                Spilled Records=22
```


Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=210
CPU time spent (ms)=920
Physical memory (bytes) snapshot=425226240
Virtual memory (bytes) snapshot=3015147520
Total committed heap usage (bytes)=418385920
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=49
File Output Format Counters
Bytes Written=71
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/snehal_wcount/output
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2022-02-21 02:39 /user/cloudera/snehal_wcount/output/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 71 2022-02-21 02:39 /user/cloudera/snehal_wcount/output/part-r-00000
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/snehal_wcount/output/part-r-00000
abc 1
dfg 1
drer 1
erew 1
erte 1
gety 1
it 1
pot 1
set 1
stedf 1
you 1
[cloudera@quickstart ~]$

HBASE COMMANDS

1. HBASE shell

2. List


cloudera@quickstart:~
hbase(main):001:0> list
TABLE
ApoorvaJain_emp
DEmoTab
DemoTableSam
Emp
Emp_Det
Emp_Detail
Emp_detailss
Emp_table
Employee_Detail
Employee_Detail1
Employee_Detail_nishtha
Employee_Details
Employee_data
Employee_details
Employeedetail
Empolyeedetail
Examination
HTable
Htable
Nidhi
Pallavi_Table
Pallavi_table
Pooja
Pooja_Table
Pooja_TableName
Shilpa
Teacher
Twinkle_Table
Twinkle_details
ananya_tab
anushka
apoorvaemp
astha
bv

```
sup
sup_table
sup_table1
table_taru
table_tulika
tableau
tableau1
tabledivs
tabledivs2
vandana
vandana_data
vartika_table
varuni_hbase
vid
yogk
ys
ytable
102 row(s) in 0.3260 seconds

=> ["ApoorvaJain_emp", "DEmoTab", "DemoTableSam", "Emp", "Emp_Det", "Emp_Detail", "Emp_detailss", "Emp_table", "Employee_Detail", "Empl
oyee_Detail1", "Employee_Detail_nishtha", "Employee_Details", "Employee_data", "Employee_details", "Employeedetail", "Empolyeedetail",
"Examination", "HTable", "Htable", "Nidhi", "Pallavi_Table", "Pallavi_table", "Pooja", "Pooja_Table", "Pooja_TableName", "Shilpa", "Tea
cher", "Twinkle_Table", "Twinkle_details", "ananya_tab", "anushka", "apoorvaemp", "astha", "bv", "bvya", "demo_taniya", "demot", "demp"
, "demp1", "dummy", "ekta", "emp", "emp1", "emp21", "emp6", "emp654", "emp_Ekta", "emp_nish", "empdata", "employee", "employee_details1
", "employee_sneha", "employees", "empruby", "emptable", "empx", "gamingtable", "hs", "jk", "khush", "khushi_tab", "laxmi_table", "li",
 "my_table", "nalini_t", "nalini_t2", "nishtha_table", "perdetail", "product", "r_student", "ritika", "saniya", "sh_table", "shef", "sh
ikha", "shub_emp", "sim", "sss", "stu", "stud", "student", "student12", "student_table", "student_table1", "studentdetail", "sup", "sup
_table", "sup_table1", "table_taru", "table_tulika", "tableau", "tableau1", "tabledivs", "tabledivs2", "vandana", "vandana_data", "vart
ika_table", "varuni_hbase", "vid", "yogk", "ys", "ytable"]
hbase(main):002:0>
```

3. Create

```
hbase(main):003:0> create 'studetails', 'id','name'
0 row(s) in 1.2250 seconds

=> Hbase::Table - studetails
hbase(main):004:0>
```

4. Put

```
hbase(main):004:0> put 'studetails','r1','id','001'
0 row(s) in 0.1150 seconds

hbase(main):005:0> put 'studetails','r1','name','abc'
0 row(s) in 0.0080 seconds

hbase(main):006:0>
```

5. Get

```
hbase(main):007:0> get 'studetails','r1'
COLUMN                          CELL
 id:                            timestamp=1645437233271, value=001
 name:                          timestamp=1645437281161, value=abc
2 row(s) in 0.0090 seconds

hbase(main):008:0>
```

6. Scan

```
hbase(main):008:0> scan 'studetails'
ROW                                   COLUMN+CELL
 r1                                   column=id:, timestamp=1645437233271, value=001
 r1                                   column=name:, timestamp=1645437281161, value=abc
1 row(s) in 0.0620 seconds

hbase(main):009:0>
```

7. Describe

```
hbase(main):009:0> describe 'studetails'
Table studetails is ENABLED
studetails
COLUMN FAMILIES DESCRIPTION
{NAME => 'id', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1', CO
IN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false',
{NAME => 'name', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS => '1',
 MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false'
}
2 row(s) in 0.0400 seconds

hbase(main):010:0>
```

8. Disable

```
hbase(main):010:0> disable 'studetails'
0 row(s) in 2.2530 seconds

hbase(main):011:0> describe 'studetails'
Table studetails is DISABLED
studetails
COLUMN FAMILIES DESCRIPTION
{NAME => 'id', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTE
IN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS
{NAME => 'name', DATA_BLOCK_ENCODING => 'NONE', BLOOMFIL
 MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CEL
}
2 row(s) in 0.0110 seconds

hbase(main):012:0>
```

9. Updating a row cell

```
hbase(main):020:0> put 'studetails','r1','name','ansh'
0 row(s) in 0.0040 seconds

hbase(main):021:0> scan 'studetails'
ROW                                COLUMN+CELL
 r1                                column=id:, timestamp=1645437233271, value=001
 r1                                column=name:, timestamp=1645437818366, value=ansh
 r2                                column=id:, timestamp=1645437767457, value=002
 r2                                column=name:, timestamp=1645437783613, value=xyz
2 row(s) in 0.0090 seconds

hbase(main):022:0>
```

10. Deleting a specific cell

```
hbase(main):022:0> delete 'studetails','r2','name',1645437783613
0 row(s) in 0.0290 seconds

hbase(main):023:0> scan 'studetails'
ROW                                COLUMN+CELL
 r1                                column=id:, timestamp=1645437233271, value=001
 r1                                column=name:, timestamp=1645437818366, value=ansh
 r2                                column=id:, timestamp=1645437767457, value=002
2 row(s) in 0.0190 seconds

hbase(main):024:0>
```

11. Deleting all cells

```
hbase(main):025:0> deleteall 'studetails','r1'
0 row(s) in 0.0100 seconds

hbase(main):026:0> scan 'studetails'
ROW                                COLUMN+CELL
 r2                                column=id:, timestamp=1645437767457, value=002
1 row(s) in 0.0190 seconds

hbase(main):027:0>
```

12. Count

```
hbase(main):029:0> count 'studetails'
2 row(s) in 0.0180 seconds

=> 2
hbase(main):030:0> scan 'studetails'
ROW                                COLUMN+CELL
 r2                                column=id:, timestamp=1645437767457, value=002
 r3                                column=id:, timestamp=1645438187428, value=003
 r3                                column=name:, timestamp=1645438172488, value=eis
2 row(s) in 0.0180 seconds

hbase(main):031:0>
```

13. Drop

```
hbase(main):032:0> disable 'studetails'
0 row(s) in 2.2380 seconds

hbase(main):033:0> drop 'studetails'
0 row(s) in 1.2430 seconds

hbase(main):034:0> list
TABLE
ApoorvaJain_emp
DEmoTab
```

```
stud
student
student12
student_table
student_table1
studentdetail
sup
sup_table
sup_table1
table_taru
table_tulika
tableau
tableau1
tabledivs
tabledivs2
vandana
vandana_data
vartika_table
varuni_hbase
vid
yogk
ys
ytable
104 row(s) in 0.0410 seconds

=> ["ApoorvaJain_emp", "DEmoTab", "DemoTableSam", "Emp", "Emp_Det", "Emp_Detail", "Emp_detailss", "Emp_tab
oyee_Detail1", "Employee_Detail_nishtha", "Employee_Details", "Employee_data", "Employee_details", "Employ
"Examination", "HTable", "Htable", "Nidhi", "Pallavi_Table", "Pallavi_table", "Pooja", "Pooja_Table", "Poo
eyashitable", "Teacher", "Train", "Twinkle_Table", "Twinkle_details", "ananya_tab", "anushka", "apoorvaemp
mo_taniya", "demot", "demp", "demp1", "dummy", "ekta", "emp", "emp1", "emp21", "emp6", "emp654", "emp_Ekta
ployee", "employee_details1", "employee_sneha", "employees", "empruby", "emptable", "empx", "gamingtable",
_tab", "laxmi_table", "li", "my_table", "nalini_t", "nalini_t2", "nishtha_table", "perdetail", "product",
a", "sh_table", "shef", "shikha", "shub_emp", "sim", "sss", "stu", "stud", "student", "student12", "studen
studentdetail", "sup", "sup_table", "sup_table1", "table_taru", "table_tulika", "tableau", "tableau1", "ta
ana", "vandana_data", "vartika_table", "varuni_hbase", "vid", "yogk", "ys", "ytable"]
hbase(main):035:0>
```