

```
[cloudera@quickstart ~]$ mkdir bigdata
[cloudera@quickstart ~]$ hdfs dfs -put bigdata/
[cloudera@quickstart ~]$ hdfs dfs -ls
```

Found 22 items

```
drwx----- - cloudera cloudera      0 2022-01-12 01:17 .staging
drwxr-xr-x - cloudera cloudera      0 2022-01-12 01:17 CHARACTER_SETS
drwxr-xr-x - cloudera cloudera      0 2020-05-23 23:09 avro_json_write
drwxr-xr-x - cloudera cloudera      0 2022-01-20 03:15 bigdata
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:15 csv_dir
-rw-r--r-- 1 cloudera cloudera    39 2022-01-20 02:50 example.txt
drwxr-xr-x - cloudera cloudera      0 2022-01-20 02:32 hdfs
drwxr-xr-x - cloudera cloudera      0 2020-06-04 08:36 import_avro
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_avro_1
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:13 json_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:39 json_orc
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_orc_1
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:38 json_parquet
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_parquet_1
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:11 orc_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:14 parquet_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-23 23:11 parquet_json_write
drwxr-xr-x - cloudera cloudera      0 2020-05-22 13:40 part_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-22 14:01 part_dir2
-rw-r--r-- 1 cloudera cloudera    37 2022-01-09 22:32 sample.txt
-rw-r--r-- 1 cloudera cloudera    49 2022-01-11 03:08 sample_kpi.csv
drwxr-xr-x - cloudera cloudera      0 2020-06-04 09:04 zeyo_dir
```

```
[cloudera@quickstart ~]$ rm -r bigdata
```

```
rm: remove directory `bigdata'? y
```

```
[cloudera@quickstart ~]$ hdfs dfs -get /bigdata
```

```
get: `/bigdata': No such file or directory
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls
```

Found 22 items

```
drwx----- - cloudera cloudera      0 2022-01-12 01:17 .staging
drwxr-xr-x - cloudera cloudera      0 2022-01-12 01:17 CHARACTER_SETS
drwxr-xr-x - cloudera cloudera      0 2020-05-23 23:09 avro_json_write
drwxr-xr-x - cloudera cloudera      0 2022-01-20 03:15 bigdata
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:15 csv_dir
-rw-r--r-- 1 cloudera cloudera    39 2022-01-20 02:50 example.txt
drwxr-xr-x - cloudera cloudera      0 2022-01-20 02:32 hdfs
drwxr-xr-x - cloudera cloudera      0 2020-06-04 08:36 import_avro
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_avro_1
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:13 json_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:39 json_orc
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_orc_1
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:38 json_parquet
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_parquet_1
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:11 orc_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:14 parquet_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-23 23:11 parquet_json_write
drwxr-xr-x - cloudera cloudera      0 2020-05-22 13:40 part_dir
```

```

drwxr-xr-x   - cloudera cloudera          0 2020-05-22 14:01 part_dir2
-rw-r--r--   1 cloudera cloudera          37 2022-01-09 22:32 sample.txt
-rw-r--r--   1 cloudera cloudera          49 2022-01-11 03:08 sample_kpi.csv
drwxr-xr-x   - cloudera cloudera          0 2020-06-04 09:04 zeyo_dir
[cloudera@quickstart ~]$ vi practice.txt
[cloudera@quickstart ~]$ cat practice.txt
Hello Everyone Welcom to KPI PARTNERS
This is my first file in vmware
[cloudera@quickstart ~]$ hadoop fs -put practice.txt /user/cloudera
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera
Found 23 items
drwx-----   - cloudera cloudera          0 2022-01-12 01:17
/user/cloudera/.staging
drwxr-xr-x   - cloudera cloudera          0 2022-01-12 01:17
/user/cloudera/CHARACTER_SETS
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 23:09
/user/cloudera/avro_json_write
drwxr-xr-x   - cloudera cloudera          0 2022-01-20 03:15 /user/cloudera/bigdata
drwxr-xr-x   - cloudera cloudera          0 2020-05-22 22:15 /user/cloudera/csv_dir
-rw-r--r--   1 cloudera cloudera          39 2022-01-20 02:50
/user/cloudera/example.txt
drwxr-xr-x   - cloudera cloudera          0 2022-01-20 02:32 /user/cloudera/hdfs
drwxr-xr-x   - cloudera cloudera          0 2020-06-04 08:36
/user/cloudera/import_avro
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 22:56
/user/cloudera/json_avro_1
drwxr-xr-x   - cloudera cloudera          0 2020-05-22 22:13
/user/cloudera/json_dir
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 22:39
/user/cloudera/json_orc
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 22:56
/user/cloudera/json_orc_1
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 22:38
/user/cloudera/json_parquet
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 22:56
/user/cloudera/json_parquet_1
drwxr-xr-x   - cloudera cloudera          0 2020-05-22 22:11 /user/cloudera/orc_dir
drwxr-xr-x   - cloudera cloudera          0 2020-05-22 22:14
/user/cloudera/parquet_dir
drwxr-xr-x   - cloudera cloudera          0 2020-05-23 23:11
/user/cloudera/parquet_json_write
drwxr-xr-x   - cloudera cloudera          0 2020-05-22 13:40
/user/cloudera/part_dir
drwxr-xr-x   - cloudera cloudera          0 2020-05-22 14:01
/user/cloudera/part_dir2
-rw-r--r--   1 cloudera cloudera          70 2022-01-20 03:20
/user/cloudera/practice.txt
-rw-r--r--   1 cloudera cloudera          37 2022-01-09 22:32
/user/cloudera/sample.txt
-rw-r--r--   1 cloudera cloudera          49 2022-01-11 03:08

```

```

/user/cloudera/sample_kpi.csv
drwxr-xr-x - cloudera cloudera          0 2020-06-04 09:04
/user/cloudera/zeyo_dir
[cloudera@quickstart ~]$ hadoop jar
/usr/lib/hadoop-mapreduce/hadoop-map-reduce-examples.jar
Not a valid JAR: /usr/lib/hadoop-mapreduce/hadoop-map-reduce-examples.jar
[cloudera@quickstart ~]$ hadoop jar
/usr/lib/hadoop-mapreduce/hadoop-map-reduce-examples.jar
Not a valid JAR: /usr/lib/hadoop-mapreduce/hadoop-map-reduce-examples.jar
[cloudera@quickstart ~]$ hadoop jar
/usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcount
/user/cloudera/practice.txt /user/cloudera/output
22/01/20 03:27:57 INFO client.RMProxy: Connecting to ResourceManager at
quickstart.cloudera/127.0.0.1:8032
22/01/20 03:27:57 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.
java:967)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:7
05)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/01/20 03:27:57 INFO input.FileInputFormat: Total input paths to process : 1
22/01/20 03:27:57 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.
java:967)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:7
05)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/01/20 03:27:57 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.
java:967)
    at

```

org.apache.hadoop.hdfs.DFSOutputStream\$DataStreamer.endBlock(DFSOutputStream.java:705)

at

org.apache.hadoop.hdfs.DFSOutputStream\$DataStreamer.run(DFSOutputStream.java:894)

22/01/20 03:27:57 INFO mapreduce.JobSubmitter: number of splits:1

22/01/20 03:27:58 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1642677741747_0001

22/01/20 03:27:58 INFO impl.YarnClientImpl: Submitted application
application_1642677741747_0001

22/01/20 03:27:58 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1642677741747_0001/

22/01/20 03:27:58 INFO mapreduce.Job: Running job: job_1642677741747_0001

22/01/20 03:28:07 INFO mapreduce.Job: Job job_1642677741747_0001 running in uber
mode : false

22/01/20 03:28:07 INFO mapreduce.Job: map 0% reduce 0%

22/01/20 03:28:12 INFO mapreduce.Job: map 100% reduce 0%

22/01/20 03:28:18 INFO mapreduce.Job: map 100% reduce 100%

22/01/20 03:28:19 INFO mapreduce.Job: Job job_1642677741747_0001 completed
successfully

22/01/20 03:28:19 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=147

FILE: Number of bytes written=295189

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=193

HDFS: Number of bytes written=96

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)=1543680

Total time spent by all reduces in occupied slots (ms)=1772544

Total time spent by all map tasks (ms)=3015

Total time spent by all reduce tasks (ms)=3462

Total vcore-milliseconds taken by all map tasks=3015

Total vcore-milliseconds taken by all reduce tasks=3462

Total megabyte-milliseconds taken by all map tasks=1543680

Total megabyte-milliseconds taken by all reduce tasks=1772544

Map-Reduce Framework

Map input records=2

Map output records=13

Map output bytes=122

Map output materialized bytes=143

Input split bytes=123

Combine input records=13

```

Combine output records=13
Reduce input groups=13
Reduce shuffle bytes=143
Reduce input records=13
Reduce output records=13
Spilled Records=26
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=72
CPU time spent (ms)=720
Physical memory (bytes) snapshot=255373312
Virtual memory (bytes) snapshot=3890036736
Total committed heap usage (bytes)=101449728

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=70

File Output Format Counters
  Bytes Written=96

```

```

[cloudera@quickstart ~]$ hadoop dfs -ls /user/cloudera/output
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

```

```

Found 2 items
-rw-r--r--  1 cloudera cloudera          0 2022-01-20 03:28
/user/cloudera/output/_SUCCESS
-rw-r--r--  1 cloudera cloudera        96 2022-01-20 03:28
/user/cloudera/output/part-r-00000
[cloudera@quickstart ~]$ hadoop dfs -cat /user/cloudera/output/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

```

```

Everyone      1
Hello 1
KPI 1
PARTNERS      1
This 1
Welcom 1
file 1
first 1
in 1
is 1
my 1
to 1

```

```

vmware 1
[cloudera@quickstart ~]$ ^C
[cloudera@quickstart ~]$ dir
avro          cm_api.py  devices.json  eclipse
express-deployment.json  HiveDirectory  kpi_hadoop  parcels          Pictures
rakeshdata    sample.txt    Videos
CHARACTER_SETS.java  data1        Documents    emp.java
external_jars          input.txt    lib          parquet_write  practice.txt
rakeshdata1    sparkjars_exec  workspace
cloudera-manager  Desktop      Downloads    enterprise-deployment.json
external-unified  kerberos     Music        part_dir        Public
sample_kpi.csv  Templates    zeyo_tab.java
[cloudera@quickstart ~]$ mkdir hadoop
[cloudera@quickstart ~]$ dir
avro          cm_api.py  devices.json  eclipse
express-deployment.json  hadoop  kerberos     Music          part_dir        Public
sample_kpi.csv  Templates  zeyo_tab.java
CHARACTER_SETS.java  data1        Documents    emp.java
external_jars          HiveDirectory  kpi_hadoop  parcels          Pictures
rakeshdata    sample.txt    Videos
cloudera-manager  Desktop      Downloads    enterprise-deployment.json
external-unified  input.txt    lib          parquet_write  practice.txt
rakeshdata1    sparkjars_exec  workspace
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$ cd /home/cloudera/hadoop
[cloudera@quickstart hadoop]$ vi sample1.txt
[cloudera@quickstart hadoop]$ cat sample1.txt
Hello KPI Partners!
This is Tanuja
Nice to meet you guys.
[cloudera@quickstart hadoop]$ ls
sample1.txt
[cloudera@quickstart hadoop]$ ls -lst
total 4
4 -rw-rw-r-- 1 cloudera cloudera 58 Jan 20 03:50 sample1.txt
[cloudera@quickstart hadoop]$ hdfs dfs -ls
Found 24 items
drwx----- - cloudera cloudera      0 2022-01-20 03:28 .staging
drwxr-xr-x - cloudera cloudera      0 2022-01-12 01:17 CHARACTER_SETS
drwxr-xr-x - cloudera cloudera      0 2020-05-23 23:09 avro_json_write
drwxr-xr-x - cloudera cloudera      0 2022-01-20 03:15 bigdata
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:15 csv_dir
-rw-r--r-- 1 cloudera cloudera    39 2022-01-20 02:50 example.txt
drwxr-xr-x - cloudera cloudera      0 2022-01-20 02:32 hdfs
drwxr-xr-x - cloudera cloudera      0 2020-06-04 08:36 import_avro
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_avro_1
drwxr-xr-x - cloudera cloudera      0 2020-05-22 22:13 json_dir
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:39 json_orc
drwxr-xr-x - cloudera cloudera      0 2020-05-23 22:56 json_orc_1

```

```
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
drwxr-xr-x - cloudera cloudera
-rw-r--r-- 1 cloudera cloudera
-rw-r--r-- 1 cloudera cloudera
-rw-r--r-- 1 cloudera cloudera
drwxr-xr-x - cloudera cloudera
[cloudera@quickstart hadoop]$ ls
sample1.txt
[cloudera@quickstart hadoop]$
```

```
0 2020-05-23 22:38 json_parquet
0 2020-05-23 22:56 json_parquet_1
0 2020-05-22 22:11 orc_dir
0 2022-01-20 03:28 output
0 2020-05-22 22:14 parquet_dir
0 2020-05-23 23:11 parquet_json_write
0 2020-05-22 13:40 part_dir
0 2020-05-22 14:01 part_dir2
70 2022-01-20 03:20 practice.txt
37 2022-01-09 22:32 sample.txt
49 2022-01-11 03:08 sample_kpi.csv
0 2020-06-04 09:04 zeyo_dir
```