

```
[cloudera@quickstart ~]$ hdfs dfs
Usage: hadoop fs [generic options]
    [-appendToFile <localsrc> ... <dst>]
    [-cat [-ignoreCrc] <src> ...]
    [-checksum <src> ...]
    [-chgrp [-R] GROUP PATH...]
    [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
    [-chown [-R] [OWNER][:([GROUP])] PATH...]
    [-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
    [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-count [-q] [-h] <path> ...]
    [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
    [-createSnapshot <snapshotDir> [<snapshotName>]]
    [-deleteSnapshot <snapshotDir> <snapshotName>]
    [-df [-h] [<path> ...]]
    [-du [-s] [-h] <path> ...]
    [-expunge]
    [-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-getfacl [-R] <path>]
    [-getfattr [-R] {-n name | -d} [-e en] <path>]
    [-getmerge [-nl] <src> <localdst>]
    [-help [cmd ...]]
    [-ls [-d] [-h] [-R] [<path> ...]]
    [-mkdir [-p] <path> ...]
    [-moveFromLocal <localsrc> ... <dst>]
    [-moveToLocal <src> <localdst>]
    [-mv <src> ... <dst>]
    [-put [-f] [-p] [-l] <localsrc> ... <dst>]
    [-renameSnapshot <snapshotDir> <oldName> <newName>]
    [-rm [-f] [-r|-R] [-skipTrash] <src> ...]
    [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
    [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>}] [--set <acl_spec> <path>]]
    [-setfattr {-n name [-v value] | -x name} <path>]
    [-setrep [-R] [-w] <rep> <path> ...]
    [-stat [format] <path> ...]
    [-tail [-f] <file>]
    [-test -[defsz] <path>]
    [-text [-ignoreCrc] <src> ...]
    [-touchz <path> ...]
    [-usage [cmd ...]]
```

Generic options supported are

- conf <configuration file> specify an application configuration file
- D <property=value> use value for given property
- fs <local|namenode:port> specify a namenode
- jt <local|resourcemanager:port> specify a ResourceManager
- files <comma separated list of files> specify comma separated files to be copied to the map reduce cluster
- libjars <comma separated list of jars> specify comma separated jar files to include in the classpath.
- archives <comma separated list of archives> specify comma separated archives to be unarchived on the compute machines.

The general command line syntax is
bin/hadoop command [genericOptions] [commandOptions]




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

```
Usage: hadoop fs [generic options]
```

```
[-appendToFile <localsrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]
[-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
[-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-count [-q] [-h] <path> ...]
[-cp [-f] [-p] [-p[topax]] <src> ... <dst>]
[-createSnapshot <snapshotDir> [<snapshotName>]]
[-deleteSnapshot <snapshotDir> <snapshotName>]
[-df [-h] [<path> ...]]
[-du [-s] [-h] <path> ...]
[-expunge]
[-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-getfacl [-R] <path>]
[-getfattr [-R] {-n name | -d} [-e en] <path>]
[-getmerge [-nl] <src> <localdst>]
[-help [cmd ...]]
[-ls [-d] [-h] [-R] [<path> ...]]
[-mkdir [-p] <path> ...]
[-moveFromLocal <localsrc> ... <dst>]
[-moveToLocal <src> <localdst>]
[-mv <src> ... <dst>]
[-put [-f] [-p] [-l] <localsrc> ... <dst>]
[-renameSnapshot <snapshotDir> <oldName> <newName>]
[-rm [-f] [-r] [-R] [-skipTrash] <src> ...]
[-rmdir [--ignore-fail-on-non-empty] <dir> ...]
[-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>][[--set <acl_spec> <path>]]
[-setfattr {-n name [-v value] | -x name} <path>]
[-setrep [-R] [-w] <rep> <path> ...]
[-stat [format] <path> ...]
[-tail [-f] <file>]
[-test [-d] [-s] <path>]
[-text [-ignoreCrc] <src> ...]
[-touchz <path> ...]
[-usage [cmd ...]]
```

Generic options supported are

```
-conf <configuration file>      specify an application configuration file
-D <property=value>             use value for given property
-fs <local|namenode:port>       specify a namenode
-jt <local|resourcemanager:port> specify a ResourceManager
-files <comma separated list of files> specify comma separated files to be copied to the map reduce cluster
-libjars <comma separated list of jars> specify comma separated jar files to include in the classpath.
-archives <comma separated list of archives> specify comma separated archives to be unarchived on the compute machines.
```

The general command line syntax is

```
bin/hadoop command [genericOptions] [commandOptions]
```

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

```
Found 2 items
```

```
drwxr-xr-x  - cloudera cloudera          0 2022-09-27 04:50 cloudera2
```

```
-rw-r--r--  1 cloudera cloudera         41 2022-09-27 04:48 vf.txt
```

```
[cloudera@quickstart ~]$ cat wordct.txt
```

```
cat: wordct.txt: No such file or directory
```

```
[cloudera@quickstart ~]$ vi wordct.txt
```

```
[1]+  Stopped                  vim wordct.txt
```

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

```
ls: Unknown command
```

Did you mean -ls? This command begins with a dash.


```

[cloudera@quickstart ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - cloudera cloudera          0 2022-09-27 04:50 cloudera2
-rw-r--r-- 1 cloudera cloudera          41 2022-09-27 04:48 vf.txt
[cloudera@quickstart ~]$ hdfs dfs -mkdir inputword
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 3 items
drwxr-xr-x - cloudera cloudera          0 2022-09-27 04:50 cloudera2
drwxr-xr-x - cloudera cloudera          0 2022-11-08 02:42 inputword
-rw-r--r-- 1 cloudera cloudera          41 2022-09-27 04:48 vf.txt
[cloudera@quickstart ~]$ hdfs dfs -put wordct.txt inputword/
[cloudera@quickstart ~]$ cat inputword
cat: inputword: No such file or directory
[cloudera@quickstart ~]$ cat inputword.txt
cat: inputword.txt: No such file or directory
hadoop jar WCFFile.jar WCDriver inputword/wordct.txt WCOuput
22/11/08 03:27:31 INFO client.RMPProxy: Connecting to ResourceManager at quickstart.cloudera/127.0.0.1:8032
22/11/08 03:27:31 INFO client.RMPProxy: Connecting to ResourceManager at quickstart.cloudera/127.0.0.1:8032
22/11/08 03:27:32 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed.
Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/11/08 03:27:32 INFO mapred.FileInputFormat: Total input paths to process : 1
22/11/08 03:27:32 INFO mapreduce.JobSubmitter: number of splits:2
22/11/08 03:27:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667903581833_0005
22/11/08 03:27:33 INFO impl.YarnClientImpl: Submitted application application_1667903581833_0005
22/11/08 03:27:33 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/
application_1667903581833_0005/
22/11/08 03:27:33 INFO mapreduce.Job: Running job: job_1667903581833_0005
22/11/08 03:27:40 INFO mapreduce.Job: Job job_1667903581833_0005 running in uber mode : false
22/11/08 03:27:40 INFO mapreduce.Job: map 0% reduce 0%
22/11/08 03:27:46 INFO mapreduce.Job: map 100% reduce 0%
22/11/08 03:27:53 INFO mapreduce.Job: map 100% reduce 100%
22/11/08 03:27:53 INFO mapreduce.Job: Job job_1667903581833_0005 completed successfully
22/11/08 03:27:53 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=74
    FILE: Number of bytes written=342038
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=293
    HDFS: Number of bytes written=30
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=1007104
    Total time spent by all reduces in occupied slots (ms)=445056
    Total time spent by all map tasks (ms)=7868
    Total time spent by all reduce tasks (ms)=3477
    Total vcore-seconds taken by all map tasks=7868
    Total vcore-seconds taken by all reduce tasks=3477
    Total megabyte-seconds taken by all map tasks=1007104
    Total megabyte-seconds taken by all reduce tasks=445056
  Map-Reduce Framework
    Map input records=11
    Map output records=10
    Map output bytes=77
    Map output materialized bytes=85
    Input split bytes=236
    Combine input records=0
    Combine output records=0
    Reduce input groups=5
    Reduce shuffle bytes=85
    Reduce input records=10

```



```

Reduce output records=5
Spilled Records=20
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=375
CPU time spent (ms)=2390
Physical memory (bytes) snapshot=414773248
Virtual memory (bytes) snapshot=2252845056
Total committed heap usage (bytes)=142082048

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=57
File Output Format Counters
Bytes Written=30
0

```

```
[cloudera@quickstart ~]$ hadoop fs -cat WCOutput/part-00000
```

```

abc      3
df       4
fdsf     1
fgh      1
fsf      1

```

```
[cloudera@quickstart ~]$ ls
```

bn.txt	Downloads	kerberos	Public	WCFile.txt
cloudera-manager	eclipse	lib	Templates	WCFile.txt~
cm_api.py	enterprise-deployment.json	mapreduce	vf.txt	wordct.txt
Desktop	express-deployment.json	Music	Videos	wordct.txt~
Documents	first.txt	Pictures	WCFile.jar	workspace

```
[cloudera@quickstart ~]$ ls workspace/
```

```
training wordcount
```

```
[cloudera@quickstart ~]$
```

GURU99 ASSIGN

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir sales
```

```
[cloudera@quickstart ~]$ hdfs dfs -put SalesJan2009.csv sales/
```

```
put: 'SalesJan2009.csv': No such file or directory
```

```
[cloudera@quickstart ~]$ cd workspace
```

```
[cloudera@quickstart workspace]$ hdfs dfs -put SalesJan2009.csv sales/
```

```
[cloudera@quickstart workspace]$ hadoop jar mapreduce.jar SalesCountryDriver sales SCOutput
```

```
22/11/08 04:04:39 INFO client.RMProxy: Connecting to ResourceManager at quickstart.cloudera/127.0.0.1:8032
```

```
22/11/08 04:04:39 INFO client.RMProxy: Connecting to ResourceManager at quickstart.cloudera/127.0.0.1:8032
```

```
22/11/08 04:04:39 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed.
```

```
Implement the Tool interface and execute your application with ToolRunner to remedy this.
```

```
22/11/08 04:04:39 INFO mapred.FileInputFormat: Total input paths to process : 1
```

```
22/11/08 04:04:39 INFO mapreduce.JobSubmitter: number of splits:2
```

```
22/11/08 04:04:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667903581833_0006
```

```
22/11/08 04:04:40 INFO impl.YarnClientImpl: Submitted application application_1667903581833_0006
```

```
22/11/08 04:04:40 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1667903581833_0006/
```

```
22/11/08 04:04:40 INFO mapreduce.Job: Running job: job_1667903581833_0006
```

```
22/11/08 04:04:45 INFO mapreduce.Job: Job job_1667903581833_0006 running in uber mode : false
```

```
22/11/08 04:04:45 INFO mapreduce.Job: map 0% reduce 0%
```

```
22/11/08 04:04:51 INFO mapreduce.Job: map 100% reduce 0%
```

```
22/11/08 04:04:57 INFO mapreduce.Job: map 100% reduce 100%
```

```
22/11/08 04:04:57 INFO mapreduce.Job: Job job_1667903581833_0006 completed successfully
```

```
22/11/08 04:04:57 INFO mapreduce.Job: Counters: 49
```

```
File System Counters
```

```
FILE: Number of bytes read=1559
```


FILE: Number of bytes written=345339
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=127595
HDFS: Number of bytes written=661
HDFS: Number of read operations=9
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters

Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=958336
Total time spent by all reduces in occupied slots (ms)=423296
Total time spent by all map tasks (ms)=7487
Total time spent by all reduce tasks (ms)=3307
Total vcore-seconds taken by all map tasks=7487
Total vcore-seconds taken by all reduce tasks=3307
Total megabyte-seconds taken by all map tasks=958336
Total megabyte-seconds taken by all reduce tasks=423296

Map-Reduce Framework

Map input records=999
Map output records=999
Map output bytes=15743
Map output materialized bytes=1856
Input split bytes=240
Combine input records=0
Combine output records=0
Reduce input groups=58
Reduce shuffle bytes=1856
Reduce input records=999
Reduce output records=58
Spilled Records=1998
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=286
CPU time spent (ms)=2390
Physical memory (bytes) snapshot=412852224
Virtual memory (bytes) snapshot=2272555008
Total committed heap usage (bytes)=144179200

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=127355

File Output Format Counters

Bytes Written=661

[cloudera@quickstart workspace]\$ hadoop fs -cat SCOutput/part-00000

Argentina	1
Australia	38
Austria	7
Bahrain	1
Belgium	8
Bermuda	1
Brazil	5
Bulgaria	1
CO	1
Canada	76
Cayman Isls	1
China	1
Costa Rica	1

```
Country 1
Czech Republic 3
Denmark 15
Dominican Republic 1
Finland 2
France 27
Germany 25
Greece 1
Guatemala 1
Hong Kong 1
Hungary 3
Iceland 1
India 2
Ireland 49
Israel 1
Italy 15
Japan 2
Jersey 1
Kuwait 1
Latvia 1
Luxembourg 1
Malaysia 1
Malta 2
Mauritius 1
Moldova 1
Monaco 2
Netherlands 22
New Zealand 6
Norway 16
Philippines 2
Poland 2
Romania 1
Russia 1
South Africa 5
South Korea 1
Spain 12
Sweden 13
Switzerland 36
Thailand 2
The Bahamas 2
Turkey 6
Ukraine 1
United Arab Emirates 6
United Kingdom 100
United States 462
[cloudera@quickstart workspace]$
```