[itv017244@g01 ~]$ ls data/trendytech/mapreduce\_jars

ls: cannot access data/trendytech/mapreduce\_jars: No such file or directory

[itv017244@g01 ~]$ ls data/

results

[itv017244@g01 ~]$ ls /data

addresses challenges git imdb nyse\_all retail\_db TetrasoftBigDataHackathon yelp-dataset-json

airlines\_all citibike gw03 linkedin nyse\_compressed retail\_db\_json training\_history

avro-tools-1.8.2.jar crime h1b names nyse\_symbols sports trendytech

cards gharchive hr\_db nyse relations tdata yelp-dataset

[itv017244@g01 ~]$ ls /data/trendytech

bigLog.txt customer-orders.csv google-ads-data.csv mapreduce\_jars search\_data.txt

boringwords.txt friends-data.csv kv1.txt samplefile.txt students.csv

[itv017244@g01 ~]$ ls /data/trendytech/mapreduce\_jars

mapreduce\_prog\_0\_reducer.jar mapreduce\_prog\_2\_reducer.jar mapreduce\_prog\_combiner.jar mapreduce\_prog\_cpartitioner.jar mapreduce\_prog.jar

[itv017244@g01 ~]$ hadoop fs -mkdir /user/itv017244/data/input

mkdir: `hdfs://m01.itversity.com:9000/user/itv017244/data': No such file or directory

[itv017244@g01 ~]$ hadoop fs -mkdir /user/itv017244/data/

[itv017244@g01 ~]$ hadoop fs -mkdir /user/itv017244/data/input/

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data

Found 1 items

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 08:24 /user/itv017244/data/input

[itv017244@g01 ~]$ vi inputfile.txt

[itv017244@g01 ~]$ hadoop fs -put inputfile.txt /user

put: Permission denied: user=itv017244, access=WRITE, inode="/user":hdfs:supergroup:drwxr-xr-x

[itv017244@g01 ~]$ hadoop fs -put inputfile.txt /user/itv017244/data/input/

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/input/

Found 1 items

-rw-r--r-- 3 itv017244 supergroup 18 2025-02-18 08:29 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ hadoop jar /data/trendytech/mapreduce\_jars/mapreduce\_prog.jar /user/itv017244/data/output

Exception in thread "main" java.lang.ArrayIndexOutOfBoundsException: 1

at Main.run(Main.java:30)

at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)

at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:90)

at Main.main(Main.java:39)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:498)

at org.apache.hadoop.util.RunJar.run(RunJar.java:323)

at org.apache.hadoop.util.RunJar.main(RunJar.java:236)

[itv017244@g01 ~]$ hadoop jar /data/trendytech/mapreduce\_jars/mapreduce\_prog.jar /user/itv017244/data/input/inputfile.txt /user/itv017244/data/output

2025-02-18 08:35:29,989 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at m02.itversity.com/172.16.1.104:8032

2025-02-18 08:35:30,209 INFO client.AHSProxy: Connecting to Application History server at m01.itversity.com/172.16.1.103:10200

2025-02-18 08:35:30,604 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/itv017244/.staging/job\_1737290322575\_19922

2025-02-18 08:35:35,916 INFO input.FileInputFormat: Total input files to process : 1

2025-02-18 08:35:36,107 INFO mapreduce.JobSubmitter: number of splits:1

2025-02-18 08:35:36,658 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1737290322575\_19922

2025-02-18 08:35:36,658 INFO mapreduce.JobSubmitter: Executing with tokens: []

2025-02-18 08:35:36,974 INFO conf.Configuration: resource-types.xml not found

2025-02-18 08:35:36,974 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2025-02-18 08:35:37,013 INFO impl.YarnClientImpl: Submitted application application\_1737290322575\_19922

2025-02-18 08:35:37,033 INFO mapreduce.Job: The url to track the job: http://m02.itversity.com:19088/proxy/application\_1737290322575\_19922/

2025-02-18 08:35:37,034 INFO mapreduce.Job: Running job: job\_1737290322575\_19922

2025-02-18 08:35:42,111 INFO mapreduce.Job: Job job\_1737290322575\_19922 running in uber mode : false

2025-02-18 08:35:42,176 INFO mapreduce.Job: map 0% reduce 0%

2025-02-18 08:35:46,251 INFO mapreduce.Job: map 100% reduce 0%

2025-02-18 08:35:50,288 INFO mapreduce.Job: map 100% reduce 100%

2025-02-18 08:35:51,315 INFO mapreduce.Job: Job job\_1737290322575\_19922 completed successfully

2025-02-18 08:35:51,406 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=48

FILE: Number of bytes written=532409

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=152

HDFS: Number of bytes written=26

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)=2244

Total time spent by all reduces in occupied slots (ms)=1314

Total time spent by all map tasks (ms)=2244

Total time spent by all reduce tasks (ms)=1314

Total vcore-milliseconds taken by all map tasks=2244

Total vcore-milliseconds taken by all reduce tasks=1314

Total megabyte-milliseconds taken by all map tasks=2297856

Total megabyte-milliseconds taken by all reduce tasks=1345536

Map-Reduce Framework

Map input records=1

Map output records=4

Map output bytes=34

Map output materialized bytes=48

Input split bytes=134

Combine input records=0

Combine output records=0

Reduce input groups=4

Reduce shuffle bytes=48

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)=42

CPU time spent (ms)=690

Physical memory (bytes) snapshot=582426624

Virtual memory (bytes) snapshot=5235810304

Total committed heap usage (bytes)=1248329728

Peak Map Physical memory (bytes)=292560896

Peak Map Virtual memory (bytes)=2613563392

Peak Reduce Physical memory (bytes)=289865728

Peak Reduce Virtual memory (bytes)=2622246912

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=18

File Output Format Counters

Bytes Written=26

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/input

Found 1 items

-rw-r--r-- 3 itv017244 supergroup 18 2025-02-18 08:29 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ cat /user/itv017244/data/input/inputfile.txt

cat: /user/itv017244/data/input/inputfile.txt: No such file or directory

[itv017244@g01 ~]$ hadoop fs -chmod 771 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/input/

Found 1 items

-rwxrwx--x 3 itv017244 supergroup 18 2025-02-18 08:29 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ hadoop fs -cat /user/itv017244/data/input/inputfile.txt

I am sneha Samita

[itv017244@g01 ~]$ hadoop fs -cat >> /user/itv017244/data/input/inputfile.txt

bash: /user/itv017244/data/input/inputfile.txt: No such file or directory

[itv017244@g01 ~]$ hadoop fs -cat /user/itv017244/data/input/inputfile.txt

I am sneha Samita

[itv017244@g01 ~]$ hadoop fs -cat > /user/itv017244/data/input/inputfile.txt

bash: /user/itv017244/data/input/inputfile.txt: No such file or directory

[itv017244@g01 ~]$ ls -lrt

total 8

drwxr-xr-x 3 itv017244 students 4096 Feb 18 07:40 data

-rw-r--r-- 1 itv017244 students 18 Feb 18 08:25 inputfile.txt

[itv017244@g01 ~]$ vi inputfile1.txt

[itv017244@g01 ~]$ ls -lrt

total 12

drwxr-xr-x 3 itv017244 students 4096 Feb 18 07:40 data

-rw-r--r-- 1 itv017244 students 18 Feb 18 08:25 inputfile.txt

-rw-r--r-- 1 itv017244 students 147 Feb 18 11:06 inputfile1.txt

[itv017244@g01 ~]$ cd data

[itv017244@g01 data]$ ls -lrt

total 4

drwxr-xr-x 2 itv017244 students 4096 Feb 18 07:41 results

[itv017244@g01 data]$ mkdir input

[itv017244@g01 data]$ ls -lrt

total 8

drwxr-xr-x 2 itv017244 students 4096 Feb 18 07:41 results

drwxr-xr-x 2 itv017244 students 4096 Feb 18 11:10 input

[itv017244@g01 data]$ cd input

[itv017244@g01 input]$ cd ..

[itv017244@g01 data]$ cd ..

[itv017244@g01 ~]$ cat inputfile1.txt

I am Sneha Samita

I like to learn new technologies

Currently i am learning BigData.

BigData is very interesting.

Learning new technologies is fun.

[itv017244@g01 ~]$ cd data

[itv017244@g01 data]$ cd input

[itv017244@g01 input]$ vi inputfile.txt

[itv017244@g01 input]$ cd ..

[itv017244@g01 data]$ cd ..

[itv017244@g01 ~]$ hadoop fs -ls /data

ls: `/data': No such file or directory

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data

Found 2 items

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 08:29 /user/itv017244/data/input

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 08:35 /user/itv017244/data/output

[itv017244@g01 ~]$ hadoop fs -cd /user/itv017244/data

-cd: Unknown command

Usage: hadoop fs [generic options]

[-appendToFile <localsrc> ... <dst>]

[-cat [-ignoreCrc] <src> ...]

[-checksum [-v] <src> ...]

[-chgrp [-R] GROUP PATH...]

[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]

[-chown [-R] [OWNER][:[GROUP]] PATH...]

[-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc> ... <dst>]

[-copyToLocal [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]

[-count [-q] [-h] [-v] [-t [<storage type>]] [-u] [-x] [-e] <path> ...]

[-cp [-f] [-p | -p[topax]] [-d] <src> ... <dst>]

[-createSnapshot <snapshotDir> [<snapshotName>]]

[-deleteSnapshot <snapshotDir> <snapshotName>]

[-df [-h] [<path> ...]]

[-du [-s] [-h] [-v] [-x] <path> ...]

[-expunge [-immediate] [-fs <path>]]

[-find <path> ... <expression> ...]

[-get [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]

[-getfacl [-R] <path>]

[-getfattr [-R] {-n name | -d} [-e en] <path>]

[-getmerge [-nl] [-skip-empty-file] <src> <localdst>]

[-head <file>]

[-help [cmd ...]]

[-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]

[-mkdir [-p] <path> ...]

[-moveFromLocal <localsrc> ... <dst>]

[-moveToLocal <src> <localdst>]

[-mv <src> ... <dst>]

[-put [-f] [-p] [-l] [-d] <localsrc> ... <dst>]

[-renameSnapshot <snapshotDir> <oldName> <newName>]

[-rm [-f] [-r|-R] [-skipTrash] [-safely] <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] [-s <sleep interval>] <file>]

[-test -[defswrz] <path>]

[-text [-ignoreCrc] <src> ...]

[-touch [-a] [-m] [-t TIMESTAMP ] [-c] <path> ...]

[-touchz <path> ...]

[-truncate [-w] <length> <path> ...]

[-usage [cmd ...]]

Generic options supported are:

-conf <configuration file> specify an application configuration file

-D <property=value> define a value for a given property

-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.

-jt <local|resourcemanager:port> specify a ResourceManager

-files <file1,...> specify a comma-separated list of files to be copied to the map reduce cluster

-libjars <jar1,...> specify a comma-separated list of jar files to be included in the classpath

-archives <archive1,...> specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:

command [genericOptions] [commandOptions]

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data

Found 2 items

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 08:29 /user/itv017244/data/input

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 08:35 /user/itv017244/data/output

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/input

Found 1 items

-rwxrwx--x 3 itv017244 supergroup 18 2025-02-18 08:29 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ hadoop fs rm /user/itv017244/data/input/inputfile.txt

rm: Unknown command

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

Usage: hadoop fs [generic options]

[-appendToFile <localsrc> ... <dst>]

[-cat [-ignoreCrc] <src> ...]

[-checksum [-v] <src> ...]

[-chgrp [-R] GROUP PATH...]

[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]

[-chown [-R] [OWNER][:[GROUP]] PATH...]

[-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] <localsrc> ... <dst>]

[-copyToLocal [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]

[-count [-q] [-h] [-v] [-t [<storage type>]] [-u] [-x] [-e] <path> ...]

[-cp [-f] [-p | -p[topax]] [-d] <src> ... <dst>]

[-createSnapshot <snapshotDir> [<snapshotName>]]

[-deleteSnapshot <snapshotDir> <snapshotName>]

[-df [-h] [<path> ...]]

[-du [-s] [-h] [-v] [-x] <path> ...]

[-expunge [-immediate] [-fs <path>]]

[-find <path> ... <expression> ...]

[-get [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]

[-getfacl [-R] <path>]

[-getfattr [-R] {-n name | -d} [-e en] <path>]

[-getmerge [-nl] [-skip-empty-file] <src> <localdst>]

[-head <file>]

[-help [cmd ...]]

[-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]

[-mkdir [-p] <path> ...]

[-moveFromLocal <localsrc> ... <dst>]

[-moveToLocal <src> <localdst>]

[-mv <src> ... <dst>]

[-put [-f] [-p] [-l] [-d] <localsrc> ... <dst>]

[-renameSnapshot <snapshotDir> <oldName> <newName>]

[-rm [-f] [-r|-R] [-skipTrash] [-safely] <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] [-s <sleep interval>] <file>]

[-test -[defswrz] <path>]

[-text [-ignoreCrc] <src> ...]

[-touch [-a] [-m] [-t TIMESTAMP ] [-c] <path> ...]

[-touchz <path> ...]

[-truncate [-w] <length> <path> ...]

[-usage [cmd ...]]

Generic options supported are:

-conf <configuration file> specify an application configuration file

-D <property=value> define a value for a given property

-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.

-jt <local|resourcemanager:port> specify a ResourceManager

-files <file1,...> specify a comma-separated list of files to be copied to the map reduce cluster

-libjars <jar1,...> specify a comma-separated list of jar files to be included in the classpath

-archives <archive1,...> specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:

command [genericOptions] [commandOptions]

[itv017244@g01 ~]$ hadoop fs -rm /user/itv017244/data/input/inputfile.txt

2025-02-18 11:14:14,049 INFO fs.TrashPolicyDefault: Moved: 'hdfs://m01.itversity.com:9000/user/itv017244/data/input/inputfile.txt' to trash at: hdfs://m01.itversity.com:9000/user/itv017244/.Trash/Current/user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ hadoop fs -put /home/itv017244/data/input/inputfile.txt /user/itv017244/data/input

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/input

Found 1 items

-rw-r--r-- 3 itv017244 supergroup 147 2025-02-18 11:15 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 ~]$ hadoop fs -cat /user/itv017244/data/input/inputfile.txt

I am Sneha Samita

I like to learn new technologies

Currently i am learning BigData.

BigData is very interesting.

Learning new technologies is fun.

[itv017244@g01 ~]$ ls-lrt

bash: ls-lrt: command not found...

[itv017244@g01 ~]$ ls -lrt

total 12

-rw-r--r-- 1 itv017244 students 18 Feb 18 08:25 inputfile.txt

-rw-r--r-- 1 itv017244 students 147 Feb 18 11:06 inputfile1.txt

drwxr-xr-x 4 itv017244 students 4096 Feb 18 11:10 data

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/

Found 2 items

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 11:15 /user/itv017244/data/input

drwxr-xr-x - itv017244 supergroup 0 2025-02-18 08:35 /user/itv017244/data/output

[itv017244@g01 ~]$ hadoop fs -rm -R /user/itv017244/data/output

2025-02-18 11:23:55,921 INFO fs.TrashPolicyDefault: Moved: 'hdfs://m01.itversity.com:9000/user/itv017244/data/output' to trash at: hdfs://m01.itversity.com:9000/user/itv017244/.Trash/Current/user/itv017244/data/output

[itv017244@g01 ~]$ hadoop jar /data/trendytech/mapreduce\_jars/mapreduce\_prog.jar /user/itv017244/data/input/inputfile.txt /user/itv017244/data/output

2025-02-18 11:25:44,511 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at m02.itversity.com/172.16.1.104:8032

2025-02-18 11:25:44,730 INFO client.AHSProxy: Connecting to Application History server at m01.itversity.com/172.16.1.103:10200

2025-02-18 11:25:45,042 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/itv017244/.staging/job\_1737290322575\_20021

2025-02-18 11:25:50,277 INFO input.FileInputFormat: Total input files to process : 1

2025-02-18 11:25:50,451 INFO mapreduce.JobSubmitter: number of splits:1

2025-02-18 11:25:50,587 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1737290322575\_20021

2025-02-18 11:25:50,587 INFO mapreduce.JobSubmitter: Executing with tokens: []

2025-02-18 11:25:50,808 INFO conf.Configuration: resource-types.xml not found

2025-02-18 11:25:50,808 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2025-02-18 11:25:50,862 INFO impl.YarnClientImpl: Submitted application application\_1737290322575\_20021

2025-02-18 11:25:50,880 INFO mapreduce.Job: The url to track the job: http://m02.itversity.com:19088/proxy/application\_1737290322575\_20021/

2025-02-18 11:25:50,881 INFO mapreduce.Job: Running job: job\_1737290322575\_20021

2025-02-18 11:25:55,954 INFO mapreduce.Job: Job job\_1737290322575\_20021 running in uber mode : false

2025-02-18 11:25:55,956 INFO mapreduce.Job: map 0% reduce 0%

2025-02-18 11:26:00,021 INFO mapreduce.Job: map 100% reduce 0%

2025-02-18 11:26:04,052 INFO mapreduce.Job: map 100% reduce 100%

2025-02-18 11:26:04,057 INFO mapreduce.Job: Job job\_1737290322575\_20021 completed successfully

2025-02-18 11:26:04,113 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=297

FILE: Number of bytes written=532907

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=281

HDFS: Number of bytes written=160

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)=1258

Total time spent by all reduces in occupied slots (ms)=1303

Total time spent by all map tasks (ms)=1258

Total time spent by all reduce tasks (ms)=1303

Total vcore-milliseconds taken by all map tasks=1258

Total vcore-milliseconds taken by all reduce tasks=1303

Total megabyte-milliseconds taken by all map tasks=1288192

Total megabyte-milliseconds taken by all reduce tasks=1334272

Map-Reduce Framework

Map input records=5

Map output records=24

Map output bytes=243

Map output materialized bytes=297

Input split bytes=134

Combine input records=0

Combine output records=0

Reduce input groups=19

Reduce shuffle bytes=297

Reduce input records=24

Reduce output records=19

Spilled Records=48

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=42

CPU time spent (ms)=650

Physical memory (bytes) snapshot=579887104

Virtual memory (bytes) snapshot=5236056064

Total committed heap usage (bytes)=1251999744

Peak Map Physical memory (bytes)=291606528

Peak Map Virtual memory (bytes)=2612699136

Peak Reduce Physical memory (bytes)=288280576

Peak Reduce Virtual memory (bytes)=2623356928

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=147

File Output Format Counters

Bytes Written=160

[itv017244@g01 ~]$ hadoop fs -ls /data/output

ls: `/data/output': No such file or directory

[itv017244@g01 ~]$ hadoop fs -ls /user/itv017244/data/output

Found 2 items

-rw-r--r-- 3 itv017244 supergroup 0 2025-02-18 11:26 /user/itv017244/data/output/\_SUCCESS

-rw-r--r-- 3 itv017244 supergroup 160 2025-02-18 11:26 /user/itv017244/data/output/part-r-00000

[itv017244@g01 ~]$ hadoop fs -cat /user/itv017244/data/output/part\*

BigData 1

BigData. 1

Currently 1

I 2

Learning 1

Samita 1

Sneha 1

am 2

fun. 1

i 1

interesting. 1

is 2

learn 1

learning 1

like 1

new 2

technologies 2

to 1

very 1

[itv017244@g01 ~]$ ls /data

addresses challenges git imdb nyse\_all retail\_db TetrasoftBigDataHackathon yelp-dataset-json

airlines\_all citibike gw03 linkedin nyse\_compressed retail\_db\_json training\_history

avro-tools-1.8.2.jar crime h1b names nyse\_symbols sports trendytech

cards gharchive hr\_db nyse relations tdata yelp-dataset

[itv017244@g01 ~]$ cd /data

[itv017244@g01 data]$ cd /trendytech

bash: cd: /trendytech: No such file or directory

[itv017244@g01 data]$ cd trendytech

[itv017244@g01 trendytech]$ ls -lrth

total 406M

-rw-r--r-- 1 root root 5.7K Oct 27 2020 kv1.txt

-rw-r--r-- 1 root root 18K Oct 28 2020 search\_data.txt

-rw-r--r-- 1 root root 134K Oct 29 2020 customer-orders.csv

-rw-r--r-- 1 root root 94K Oct 29 2020 google-ads-data.csv

-rw-r--r-- 1 root root 78K Oct 29 2020 boringwords.txt

-rw-r--r-- 1 root root 349M Oct 29 2020 bigLog.txt

-rw-r--r-- 1 root root 159 Oct 30 2020 samplefile.txt

-rw-r--r-- 1 root root 8.1K Oct 30 2020 friends-data.csv

-rwxr-xr-x 1 root root 58M Oct 30 2020 students.csv

drwxr-xr-x 2 itv005857 students 4.0K Apr 23 2023 mapreduce\_jars

[itv017244@g01 trendytech]$ head bigLog.txt

ERROR: Thu Jun 04 10:37:51 BST 2015

WARN: Sun Nov 06 10:37:51 GMT 2016

WARN: Mon Aug 29 10:37:51 BST 2016

ERROR: Thu Dec 10 10:37:51 GMT 2015

ERROR: Fri Dec 26 10:37:51 GMT 2014

ERROR: Thu Feb 02 10:37:51 GMT 2017

WARN: Fri Oct 17 10:37:51 BST 2014

ERROR: Wed Jul 01 10:37:51 BST 2015

WARN: Thu Jul 27 10:37:51 BST 2017

WARN: Thu Oct 19 10:37:51 BST 2017

[itv017244@g01 trendytech]$ pwd

/data/trendytech

[itv017244@g01 trendytech]$ cd ..

[itv017244@g01 data]$ cd ..

[itv017244@g01 /]$ hadoop fs -put /data/trendytech/bigLog.txt /user/itv017244/data/input

[itv017244@g01 /]$ hadoop fs -ls /user/itv017244/data/input

Found 2 items

-rw-r--r-- 3 itv017244 supergroup 365001114 2025-02-18 11:38 /user/itv017244/data/input/bigLog.txt

-rw-r--r-- 3 itv017244 supergroup 147 2025-02-18 11:15 /user/itv017244/data/input/inputfile.txt

[itv017244@g01 /]$ hadoop fs -rm -R /user/itv017244/data/output

2025-02-18 11:41:15,430 INFO fs.TrashPolicyDefault: Moved: 'hdfs://m01.itversity.com:9000/user/itv017244/data/output' to trash at: hdfs://m01.itversity.com:9000/user/itv017244/.Trash/Current/user/itv017244/data/output1739896875419

[itv017244@g01 /]$ hadoop jar /data/trendytech/mapreduce\_jars/mapreduce\_prog.jar /user/itv017244/data/input/bigLog.txt /user/itv017244/data/output

2025-02-18 11:41:40,744 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at m02.itversity.com/172.16.1.104:8032

2025-02-18 11:41:41,143 INFO client.AHSProxy: Connecting to Application History server at m01.itversity.com/172.16.1.103:10200

2025-02-18 11:41:41,545 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/itv017244/.staging/job\_1737290322575\_20033

2025-02-18 11:41:47,057 INFO input.FileInputFormat: Total input files to process : 1

2025-02-18 11:41:47,284 INFO mapreduce.JobSubmitter: number of splits:3

2025-02-18 11:41:47,773 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1737290322575\_20033

2025-02-18 11:41:47,773 INFO mapreduce.JobSubmitter: Executing with tokens: []

2025-02-18 11:41:47,978 INFO conf.Configuration: resource-types.xml not found

2025-02-18 11:41:47,978 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2025-02-18 11:41:48,032 INFO impl.YarnClientImpl: Submitted application application\_1737290322575\_20033

2025-02-18 11:41:48,051 INFO mapreduce.Job: The url to track the job: http://m02.itversity.com:19088/proxy/application\_1737290322575\_20033/

2025-02-18 11:41:48,051 INFO mapreduce.Job: Running job: job\_1737290322575\_20033

2025-02-18 11:41:55,150 INFO mapreduce.Job: Job job\_1737290322575\_20033 running in uber mode : false

2025-02-18 11:41:55,153 INFO mapreduce.Job: map 0% reduce 0%

2025-02-18 11:42:12,319 INFO mapreduce.Job: map 16% reduce 0%

2025-02-18 11:42:18,352 INFO mapreduce.Job: map 23% reduce 0%

2025-02-18 11:42:24,388 INFO mapreduce.Job: map 32% reduce 0%

2025-02-18 11:42:25,399 INFO mapreduce.Job: map 33% reduce 0%

2025-02-18 11:42:39,467 INFO mapreduce.Job: map 49% reduce 0%

2025-02-18 11:42:45,495 INFO mapreduce.Job: map 56% reduce 0%

2025-02-18 11:42:51,520 INFO mapreduce.Job: map 65% reduce 0%

2025-02-18 11:42:53,532 INFO mapreduce.Job: map 67% reduce 0%

2025-02-18 11:43:06,586 INFO mapreduce.Job: map 88% reduce 0%

2025-02-18 11:43:12,611 INFO mapreduce.Job: map 98% reduce 0%

2025-02-18 11:43:13,617 INFO mapreduce.Job: map 100% reduce 22%

2025-02-18 11:43:19,645 INFO mapreduce.Job: map 100% reduce 78%

2025-02-18 11:43:25,675 INFO mapreduce.Job: map 100% reduce 88%

2025-02-18 11:43:31,700 INFO mapreduce.Job: map 100% reduce 97%

2025-02-18 11:43:33,710 INFO mapreduce.Job: map 100% reduce 100%

2025-02-18 11:43:34,722 INFO mapreduce.Job: Job job\_1737290322575\_20033 completed successfully

2025-02-18 11:43:34,865 INFO mapreduce.Job: Counters: 55

File System Counters

FILE: Number of bytes read=1550002378

FILE: Number of bytes written=2326068179

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=365009699

HDFS: Number of bytes written=661

HDFS: Number of read operations=14

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed map tasks=1

Launched map tasks=4

Launched reduce tasks=1

Data-local map tasks=4

Total time spent by all maps in occupied slots (ms)=82483

Total time spent by all reduces in occupied slots (ms)=33208

Total time spent by all map tasks (ms)=82483

Total time spent by all reduce tasks (ms)=33208

Total vcore-milliseconds taken by all map tasks=82483

Total vcore-milliseconds taken by all reduce tasks=33208

Total megabyte-milliseconds taken by all map tasks=84462592

Total megabyte-milliseconds taken by all reduce tasks=34004992

Map-Reduce Framework

Map input records=10000000

Map output records=70000000

Map output bytes=635001114

Map output materialized bytes=775001132

Input split bytes=393

Combine input records=0

Combine output records=0

Reduce input groups=60

Reduce shuffle bytes=775001132

Reduce input records=70000000

Reduce output records=60

Spilled Records=210000000

Shuffled Maps =3

Failed Shuffles=0

Merged Map outputs=3

GC time elapsed (ms)=401

CPU time spent (ms)=102790

Physical memory (bytes) snapshot=2199371776

Virtual memory (bytes) snapshot=10474299392

Total committed heap usage (bytes)=3189768192

Peak Map Physical memory (bytes)=624832512

Peak Map Virtual memory (bytes)=2617180160

Peak Reduce Physical memory (bytes)=330403840

Peak Reduce Virtual memory (bytes)=2623967232

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=365009306

File Output Format Counters

Bytes Written=661

[itv017244@g01 /]$