

## For NASA Data:

---

### Step 1:

To Count the number of rows in NASA data, Follow the commands in Achyutha\_NagavetiBhavaniSanthoshi\_README\_2

---

### Step 2:

To run LineCount jar file

```
anagave@hadoop-nn001:~$ hadoop jar lc.jar LineCount /user/anagave/NASA__data/2021/02/27/21/FlumeData.* /user/anagave/output_NASADATA
```

```
2021-03-12 19:03:37,465 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
2021-03-12 19:03:38,276 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting
to ResourceManager at hadoop-nn001.cs.okstate.edu/192.168.122.2:8032
2021-03-12 19:03:38,480 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting
to ResourceManager at hadoop-nn001.cs.okstate.edu/192.168.122.2:8032
2021-03-12 19:03:38,730 WARN mapreduce.JobResourceUploader: Hadoop command-
line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
2021-03-12 19:03:38,750 INFO mapreduce.JobResourceUploader: Disabling Erasure
Coding for path: /tmp/hadoop-yarn/staging/anagave/.staging/job_1615570959549_0062
2021-03-12 19:03:39,153 INFO mapred.FileInputFormat: Total input files to process : 82
2021-03-12 19:03:39,618 INFO mapreduce.JobSubmitter: number of splits:82
2021-03-12 19:03:39,866 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1615570959549_0062
2021-03-12 19:03:39,866 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-03-12 19:03:40,216 INFO conf.Configuration: resource-types.xml not found
2021-03-12 19:03:40,218 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
2021-03-12 19:03:40,410 INFO impl.YarnClientImpl: Submitted application
application_1615570959549_0062
2021-03-12 19:03:40,514 INFO mapreduce.Job: The url to track the job: http://hadoop-
nn001.cs.okstate.edu:8088/proxy/application_1615570959549_0062/
2021-03-12 19:03:40,517 INFO mapreduce.Job: Running job: job_1615570959549_0062
2021-03-12 19:03:45,644 INFO mapreduce.Job: Job job_1615570959549_0062 running
in uber mode : false
2021-03-12 19:03:45,647 INFO mapreduce.Job: map 0% reduce 0%
2021-03-12 19:03:50,768 INFO mapreduce.Job: map 20% reduce 0%
2021-03-12 19:03:52,792 INFO mapreduce.Job: map 22% reduce 0%
2021-03-12 19:03:54,816 INFO mapreduce.Job: map 39% reduce 0%
2021-03-12 19:03:58,859 INFO mapreduce.Job: map 55% reduce 0%
2021-03-12 19:04:02,909 INFO mapreduce.Job: map 68% reduce 0%
2021-03-12 19:04:05,938 INFO mapreduce.Job: map 73% reduce 0%
2021-03-12 19:04:06,949 INFO mapreduce.Job: map 82% reduce 0%
2021-03-12 19:04:07,959 INFO mapreduce.Job: map 82% reduce 11%
2021-03-12 19:04:08,969 INFO mapreduce.Job: map 84% reduce 11%
2021-03-12 19:04:09,980 INFO mapreduce.Job: map 88% reduce 16%
```

2021-03-12 19:04:10,991 INFO mapreduce.Job: map 95% reduce 16%  
2021-03-12 19:04:12,002 INFO mapreduce.Job: map 100% reduce 16%  
2021-03-12 19:04:13,012 INFO mapreduce.Job: map 100% reduce 100%  
2021-03-12 19:04:13,029 INFO mapreduce.Job: Job job\_1615570959549\_0062  
completed successfully  
2021-03-12 19:04:13,199 INFO mapreduce.Job: Counters: 55

#### File System Counters

FILE: Number of bytes read=1506  
FILE: Number of bytes written=23048516  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=20609894  
HDFS: Number of bytes written=17  
HDFS: Number of read operations=271  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=10  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Launched map tasks=82  
Launched reduce tasks=5  
Data-local map tasks=79  
Rack-local map tasks=3  
Total time spent by all maps in occupied slots (ms)=1127565  
Total time spent by all reduces in occupied slots (ms)=405600  
Total time spent by all map tasks (ms)=225513  
Total time spent by all reduce tasks (ms)=81120  
Total vcore-milliseconds taken by all map tasks=225513  
Total vcore-milliseconds taken by all reduce tasks=81120  
Total megabyte-milliseconds taken by all map tasks=1154626560  
Total megabyte-milliseconds taken by all reduce tasks=415334400

#### Map-Reduce Framework

Map input records=3296  
Map output records=3296  
Map output bytes=52736  
Map output materialized bytes=3936  
Input split bytes=12546  
Combine input records=3296  
Combine output records=82  
Reduce input groups=1  
Reduce shuffle bytes=3936  
Reduce input records=82  
Reduce output records=1  
Spilled Records=164  
Shuffled Maps =410  
Failed Shuffles=0  
Merged Map outputs=410  
GC time elapsed (ms)=1611  
CPU time spent (ms)=51780  
Physical memory (bytes) snapshot=30508232704  
Virtual memory (bytes) snapshot=555142189056  
Total committed heap usage (bytes)=68488790016

Peak Map Physical memory (bytes)=367734784  
Peak Map Virtual memory (bytes)=6388592640  
Peak Reduce Physical memory (bytes)=292577280  
Peak Reduce Virtual memory (bytes)=6401871872  
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=20597348  
File Output Format Counters  
Bytes Written=17

---

### Step 3:

To view the contents in output file

```
[anagave@hadoop-nn001:~$ hdfs dfs -cat /user/anagave/output_NASADATA/part*  
2021-03-12 19:07:25,252 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
Total Lines      3296
```

## For Mars Data:

---

### Step 1:

To Count the number of rows in Mars data, Follow the commands in Achyutha\_NagavetiBhavaniSanthoshi\_README\_2

---

### Step 2:

To run LineCount

```
anagave@hadoop-nn001:~$ hadoop jar lc.jar LineCount /user/anagave/Mars__data/2021/02/27/20/FlumeData.* /user/anagave/output_MarsDATA
```

```
2021-03-12 19:21:58,691 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-03-12 19:21:59,543 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at hadoop-nn001.cs.okstate.edu/192.168.122.2:8032
2021-03-12 19:21:59,775 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at hadoop-nn001.cs.okstate.edu/192.168.122.2:8032
2021-03-12 19:22:00,174 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-03-12 19:22:00,222 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/anagave/.staging/job_1615570959549_0071
2021-03-12 19:22:00,597 INFO mapred.FileInputFormat: Total input files to process : 49
2021-03-12 19:22:00,858 INFO mapreduce.JobSubmitter: number of splits:49
2021-03-12 19:22:01,011 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1615570959549_0071
2021-03-12 19:22:01,013 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-03-12 19:22:01,248 INFO conf.Configuration: resource-types.xml not found
2021-03-12 19:22:01,249 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-03-12 19:22:01,355 INFO impl.YarnClientImpl: Submitted application application_1615570959549_0071
2021-03-12 19:22:01,457 INFO mapreduce.Job: The url to track the job: http://hadoop-nn001.cs.okstate.edu:8088/proxy/application_1615570959549_0071/
2021-03-12 19:22:01,459 INFO mapreduce.Job: Running job: job_1615570959549_0071
2021-03-12 19:22:06,583 INFO mapreduce.Job: Job job_1615570959549_0071 running in uber mode : false
2021-03-12 19:22:06,586 INFO mapreduce.Job: map 0% reduce 0%
2021-03-12 19:22:11,722 INFO mapreduce.Job: map 33% reduce 0%
2021-03-12 19:22:14,756 INFO mapreduce.Job: map 35% reduce 0%
2021-03-12 19:22:15,769 INFO mapreduce.Job: map 65% reduce 0%
2021-03-12 19:22:18,807 INFO mapreduce.Job: map 71% reduce 0%
2021-03-12 19:22:19,819 INFO mapreduce.Job: map 88% reduce 0%
2021-03-12 19:22:21,848 INFO mapreduce.Job: map 92% reduce 0%
2021-03-12 19:22:22,860 INFO mapreduce.Job: map 96% reduce 0%
2021-03-12 19:22:23,872 INFO mapreduce.Job: map 100% reduce 60%
2021-03-12 19:22:24,882 INFO mapreduce.Job: map 100% reduce 100%
```

2021-03-12 19:22:24,899 INFO mapreduce.Job: Job job\_1615570959549\_0071 completed successfully

2021-03-12 19:22:25,070 INFO mapreduce.Job: Counters: 57

#### File System Counters

FILE: Number of bytes read=894  
FILE: Number of bytes written=14305526  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=9449983  
HDFS: Number of bytes written=17  
HDFS: Number of read operations=172  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=10  
HDFS: Number of bytes read erasure-coded=0

#### Job Counters

Killed map tasks=1  
Launched map tasks=49  
Launched reduce tasks=5  
Other local map tasks=1  
Data-local map tasks=46  
Rack-local map tasks=2  
Total time spent by all maps in occupied slots (ms)=695275  
Total time spent by all reduces in occupied slots (ms)=196560  
Total time spent by all map tasks (ms)=139055  
Total time spent by all reduce tasks (ms)=39312  
Total vcore-milliseconds taken by all map tasks=139055  
Total vcore-milliseconds taken by all reduce tasks=39312  
Total megabyte-milliseconds taken by all map tasks=711961600  
Total megabyte-milliseconds taken by all reduce tasks=201277440

#### Map-Reduce Framework

Map input records=1729  
Map output records=1729  
Map output bytes=27664  
Map output materialized bytes=2334  
Input split bytes=7501  
Combine input records=1729  
Combine output records=48  
Reduce input groups=1  
Reduce shuffle bytes=2334  
Reduce input records=48  
Reduce output records=1  
Spilled Records=96  
Shuffled Maps =245  
Failed Shuffles=0  
Merged Map outputs=245  
GC time elapsed (ms)=1049  
CPU time spent (ms)=32770  
Physical memory (bytes) snapshot=18673999872  
Virtual memory (bytes) snapshot=344546791424  
Total committed heap usage (bytes)=42549116928  
Peak Map Physical memory (bytes)=366071808

Peak Map Virtual memory (bytes)=6388482048  
Peak Reduce Physical memory (bytes)=266899456  
Peak Reduce Virtual memory (bytes)=6397992960  
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=9442482  
File Output Format Counters  
Bytes Written=17

---

### Step 3:

To view the contents in output file

```
[anagave@hadoop-nn001:~$ hdfs dfs -cat /user/anagave/output_MarsDATA/part* ]  
2021-03-12 19:25:21,967 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
Total Lines      1729
```