# Session 2:

## **YARN**

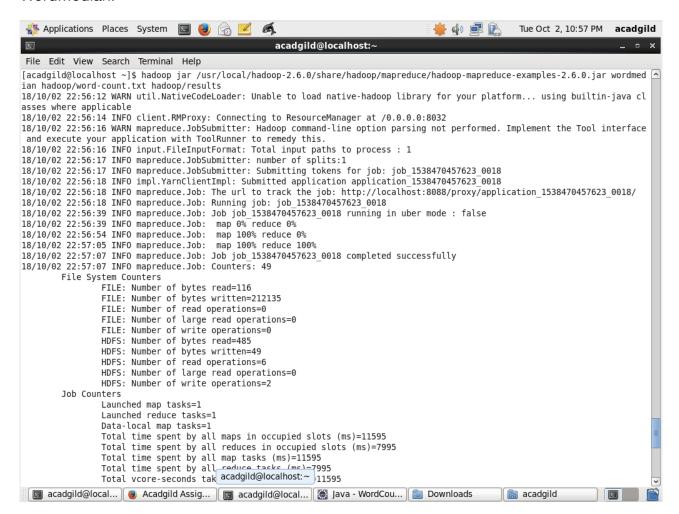
### **MR-INTRODUCTION**

# **Assignment 1**

#### Task 1:

Execute WordMedian, WordMean, WordStandardDeviation programs using hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM.

#### WordMedian:



```
💸 Applications Places System 国 📵 🍙 🗹 🙈
                                                                                                  👚 🐠 🚅 🖺
                                                                                                                      Tue Oct 2, 10:59 PM acadgild
                                                             acadgild@localhost:~
File Edit View Search Terminal Help
                   Total time spent by all reduce tasks (ms)=7995
Total vcore-seconds taken by all map tasks=11595
Total vcore-seconds taken by all reduce tasks=7995
Total megabyte-seconds taken by all map tasks=11873280
                   Total megabyte-seconds taken by all reduce tasks=8186880
         Map-Reduce Framework
                   Map input records=3
                   Map output records=61
Map output bytes=488
                   Map output materialized bytes=116
                   Input split bytes=122
                   Combine input records=61
                   Combine output records=11
                   Reduce input groups=11
                   Reduce shuffle bytes=116
Reduce input records=11
                   Reduce output records=11
I
                   Spilled Records=22
                   Shuffled Maps =1
                   Failed Shuffles=0
                   Merged Map outputs=1
                   GC time elapsed (ms)=279
                   CPU time spent (ms)=1790
                   Physical memory (bytes) snapshot=270798848
                   Virtual memory (bytes) snapshot=4113940480
                   Total committed heap usage (bytes)=137498624
         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=363
         File Output Format Counters
                   Bytes Written=49
The median is: 4
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -ls hadoop/results
18/10/02 23:03:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
             1 acadgild supergroup
                                                 0 2018-10-02 23:02 hadoop/results/_SUCCESS
-rw-r--r--
-rw-r--r-- 1 acadgild supergroup 49 2018-10-02 23:02 had
[acadgild@localhost ~]$ hadoop fs -cat hadoop/results/part-r-00000
                                               49 2018-10-02 23:02 hadoop/results/part-r-00000
18/10/02 23:03:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
2
        12
3
4
        16
        1
6
8
        5
9
        4
10
        3
12
]3 1
[acadgild@localhost ~]$ ■
```

#### WordMean:

```
👫 Applications Places System 国 📵 🙈 🗾 🍕
                                                                                                                👚 🚯 🚅 🦺 Tue Oct 2, 11:08 PM acadgild
                                                                     acadgild@localhost:~
 File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar /usr/local/hadoop-2.6.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar wordmea 🔄
n hadoop/word-count.txt hadoop/results
18/10/02 23:07:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/10/02 23:07:06 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/10/02 23:07:08 INFO input.FileInputFormat: Total input paths to process : 1
18/10/02 23:07:08 INFO mapreduce.JobSubmitter: number of splits:1
18/10/02 23:07:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538470457623_0020
18/10/02 23:07:09 INFO impl.YarnClientImpl: Submitted application application_1538470457623_0020
18/10/02 23:07:09 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538470457623_0020/
18/10/02 23:07:09 INFO mapreduce.Job: Running job: job_1538470457623_0020
18/10/02 23:07:24 INFO mapreduce.Job: Job job_1538470457623_0020 running in uber mode : false
18/10/02 23:07:24 INFO mapreduce.Job: map 0% reduce 0%
18/10/02 23:07:36 INFO mapreduce.Job: map 100% reduce 0%
18/10/02 23:07:48 INFO mapreduce.Job: map 100% reduce 100%
18/10/02 23:07:49 INFO mapreduce.Job: Job job_1538470457623_0020 completed successfully
18/10/02 23:07:49 INFO mapreduce.Job: Counters: 49
           File System Counters
                      FILE: Number of bytes read=39
                      FILE: Number of bytes written=212245
                      FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
                      HDFS: Number of bytes read=485
                      HDFS: Number of bytes written=20
                      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)=10440 Total time spent by all reduces in occupied slots (ms)=7921
                      Total time spent by all map tasks (ms)=10440
Total time spent by all reduce tasks (ms)=7921
                      Total vcore-seconds taken by all map tasks=10440
                      Total vcore-seconds taken by all reduce tasks=7921
                                                                                                                                                         27 Items in Trash
                      Total megabyte-seconds taken by all map tasks=10690560
```

```
acadgild@localhost:~
File Edit View Search Terminal Help
                 Total time spent by all reduce tasks (ms)=7921
Total vcore-seconds taken by all map tasks=10440
                 Total vcore-seconds taken by all reduce tasks=7921
                 Total megabyte-seconds taken by all map tasks=10690560
                 Total megabyte-seconds taken by all reduce tasks=8111104
        Map-Reduce Framework
                 Map input records=3
Map output records=122
                 Map output bytes=1769
                 Map output materialized bytes=39
                 Input split bytes=122
                 Combine input records=122
                 Combine output records=2
                 Reduce input groups=2
Reduce shuffle bytes=39
                 Reduce input records=2
                 Reduce output records=2
                 Spilled Records=4
                 Shuffled Maps =1
                 Failed Shuffles=0
                 Merged Map outputs=1
                 GC time elapsed (ms)=296
                 CPU time spent (ms)=2080
                 Physical memory (bytes) snapshot=271011840
                 Virtual memory (bytes) snapshot=4113776640
                 Total committed heap usage (bytes)=137498624
        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=363
        File Output Format Counters
                 .
Bytes Written=20
The mean is: 4.770491803278689
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -cat hadoop/results/part-r-00000
18/10/02 23:10:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 61
length 291
[acadgild@localhost ~]$ |
```

#### WordStandardDeviation:

```
🌞 🐠 🚅 🖺
💸 Applications Places System 国 📵 🍙 🗹 🍕
                                                                                                               Tue Oct 2, 11:15 PM acadgild
                                                        acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar /usr/local/hadoop-2.6.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar wordsta 🔄
ndarddeviation hadoop/word-count.txt hadoop/results
18/10/02 23:12:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/10/02 23:12:14 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032 18/10/02 23:12:16 INFO input.FileInputFormat: Total input paths to process : 1
18/10/02 23:12:16 INFO mapreduce.JobSubmitter: number of splits:1
18/10/02 23:12:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538470457623_0021
18/10/02 23:12:17 INFO impl.YarnClientImpl: Submitted application application_1538470457623_0021
18/10/02 23:12:18 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538470457623_0021/18/10/02 23:12:18 INFO mapreduce.Job: Running job: job_1538470457623_0021
18/10/02 23:12:30 INFO mapreduce.Job: Job job 1538470457623 0021 running in uber mode : false
18/10/02 23:12:30 INFO mapreduce.Job: map 0% reduce 0%
18/10/02 23:12:42 INFO mapreduce.Job:
                                           map 100% reduce 0%
18/10/02 23:12:53 INFO mapreduce.Job: map 100% reduce 100%
18/10/02 23:12:53 INFO mapreduce.Job: Job job_1538470457623_0021 completed successfully
18/10/02 23:12:53 INFO mapreduce.Job: Counters: 49
        File System Counters
                 FILE: Number of bytes read=56
                 FILE: Number of bytes written=212439
                 FILE: Number of read operations=0
                 FILE: Number of large read operations=0
FILE: Number of write operations=0
                  HDFS: Number of bytes read=485
                  HDFS: Number of bytes written=32
                 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)=9610
                  Total time spent by all reduces in occupied slots (ms)=7434
                  Total time spent by all map tasks (ms)=9610
                 Total time spent by all reduce tasks (ms)=7434
Total vcore-seconds taken by all map tasks=9610
                  Total vcore-seconds taken by all reduce tasks=7434
                                                                                                                  acadgild
                 Total megabyte-seconds taken by all map tasks=9840640
          Map-Reduce Framework
                   Map input records=3
                   Map output records=183
                   Map output bytes=2684
                   Map output materialized bytes=56
                   Input split bytes=122
                   Combine input records=183
Combine output records=3
                   Reduce input groups=3
                   Reduce shuffle bytes=56
                   Reduce input records=3
                   Reduce output records=3
Spilled Records=6
                   Shuffled Maps =1
                   Failed Shuffles=0
                   Merged Map outputs=1
                   GC time elapsed (ms)=209
                   CPU time spent (ms)=1830
                   Physical memory (bytes) snapshot=268406784
Virtual memory (bytes) snapshot=4113920000
                   Total committed heap usage (bytes)=137498624
          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=363
          File Output Format Counters
                   Bytes Written=32
 The standard deviation is: 2.8823569739611794
                                                                                                                                           I
 [acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -cat hadoop/results/part-r-00000
18/10/02 23:17:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl asses where applicable count 61
length 291
square 1895
[acadgild@localhost ~]$
```

#### Task 2:

#### Task 2.1:

Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

Please refer to <a href="https://github.com/itsmanish13/ACADGILD">https://github.com/itsmanish13/ACADGILD</a> BIG DATA ENGINEERING/tree/master/map-reduce-prog for code. Driver class is same in all utilities.

A filter mapper is written for this problem which will filter out invalid records (records having NA in 1st and 3rd column)

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar map-reduce-prog-0.1.jar filter hadoop/television.txt /home/acadgild/results
18/10/02 23:29:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/10/02 23:30:00 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/10/02 23:30:01 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/10/02 23:30:02 INFO input.FileInputFormat: Total input paths to process : 1
18/10/02 23:30:03 INFO mapreduce.JobSubmitter: number of splits:1
18/10/02 23:30:04 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538470457623_0024
18/10/02 23:30:04 INFO impl.YarnClientImpl: Submitted application application_1538470457623_0024
18/10/02 23:30:04 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538470457623_0024/
18/10/02 23:30:04 INFO mapreduce.Job: Running job: job_1538470457623_0024
18/10/02 23:30:19 INFO mapreduce.Job: Job job 1538470457623_0024 running in uber mode : false 18/10/02 23:30:19 INFO mapreduce.Job: map 0% reduce 0%
18/10/02 23:30:30 INFO mapreduce.Job:
                                          map 100% reduce 0%
18/10/02 23:30:31 INFO mapreduce.Job: Job job_1538470457623_0024 completed successfully
18/10/02 23:30:31 INFO mapreduce.Job: Counters: 30
        File System Counters
                 FILE: Number of bytes read=0
                 FILE: Number of bytes written=105796
                 FILE: Number of read operations=0
                 FILE: Number of large read operations=0
                 FILE: Number of write operations=0
                 HDFS: Number of bytes read=855
                 HDFS: Number of bytes written=646
                 HDFS: Number of read operations=5
                 HDFS: Number of large read operations=0
                 HDFS: Number of write operations=2
        Job Counters
                 Launched map tasks=1
                 Data-local map tasks=1
                 Total time spent by all maps in occupied slots (ms)=8529
                 Total time spent by all reduces in occupied slots (ms)=0
                 Total time spent by all map tasks (ms)=8529
                 Total vcore-seconds taken by all map tasks=8529
                                                                                                                                      I
                 Total megabyte-seconds taken by all map tasks=8733696
        Map-Reduce Framework
                 Map input records=18
                 Map output records=16
```

```
[acadgild@localhost ~]$ hadoop fs -ls /home/acadgild/results
18/10/02 23:30:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--
              1 acadgild supergroup
                                                0 2018-10-02 23:30 /home/acadgild/results/ SUCCESS
-rw-r--r--
              1 acadgild supergroup
                                              646 2018-10-02 23:30 /home/acadgild/results/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /home/acadgild/results/part-m-00000
18/10/02 23:31:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Jamsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Akai|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Onida|Lucid|18|Uttar Pradesh|232401|16200
Onida|Decent|14|Uttar Pradesh|232401|16200
Lava|Attention|20|Assam|454601|24200
Zen|Super|14|Maharashtra|619082|9200
Samsung|Optima|14|Madhya Pradesh|132401|14200
Samsung|Decent|16|Kerala|922401|12200
Lava|Attention|20|Assam|454601|24200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
Samsung|Super|14|Maharashtra|619082|9200
[acadgild@localhost ~]$ ■
```

#### **Task 2.2**

Write a Map Reduce program to calculate the total units sold for each Company.

Please refer to <a href="https://github.com/itsmanish13/ACADGILD\_BIG\_DATA\_ENGINEERING/tree/master/map-reduce-prog">https://github.com/itsmanish13/ACADGILD\_BIG\_DATA\_ENGINEERING/tree/master/map-reduce-prog</a> for code. Driver class is same in all utilities.

Both mapper and reduce is written for utility totalSalesByBrand for this problem.

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar map-reduce-prog-0.1.jar totalSalesByBrand hadoop/television.txt hadoop/results
18/10/02 23:35:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/10/02 23:35:14 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032 18/10/02 23:35:16 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/10/02 23:35:16 INFO input.FileInputFormat: Total input paths to process : 1
18/10/02 23:35:16 INFO mapreduce.JobSubmitter: number of splits:1
18/10/02 23:35:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538470457623_0025
18/10/02 23:35:17 INFO impl.YarnClientImpl: Submitted application application_1538470457623_0025
18/10/02 23:35:17 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538470457623_0025/
18/10/02 23:35:17 INFO mapreduce.Job: Running job: job_1538470457623_0025
18/10/02 23:35:30 INFO mapreduce.Job: Job job 1538470457623_0025 running in uber mode : false 18/10/02 23:35:30 INFO mapreduce.Job: map 0% reduce 0%
18/10/02 23:35:42 INFO mapreduce.Job:
                                               map 100% reduce 0%
18/10/02 23:35:54 INFO mapreduce.Job:
                                               map 100% reduce 100%
18/10/02 23:35:55 INFO mapreduce.Job: Job job_1538470457623_0025 completed successfully
18/10/02 23:35:56 INFO mapreduce.Job: Counters: 49
         File System Counters
                   FILE: Number of bytes read=225
                   FILE: Number of bytes written=213425
                   FILE: Number of read operations=0
                   FILE: Number of large read operations=0
                   FILE: Number of write operations=0
                   HDFS: Number of bytes read=855
                   HDFS: Number of bytes written=43
                   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)=10537
                   Total time spent by all reduces in occupied slots (ms)=7698
                   Total time spent by all map tasks (ms)=10537
                   Total time spent by all reduce tasks (ms)=7698
Total vcore-seconds taken by all map tasks=10537
                   Total vcore-seconds taken by all reduce tasks=7698
```

```
[acadgild@localhost ~]$ hadoop fs -cat hadoop/results/part-r-00000

18/10/02 23:36:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable

AKAI 1

LAVA 3

NA 1

ONIDA 4

SAMSUNG 7

ZEN 2

[acadgild@localhost ~]$ 

[acadgild@localhost ~]$
```

#### Task 2.3

Write a Map Reduce program to calculate the total units sold in each state for Onida company.

Please refer to <a href="https://github.com/itsmanish13/ACADGILD">https://github.com/itsmanish13/ACADGILD</a> BIG DATA ENGINEERING/tree/master/map-reduce-prog for code. Driver class is same in all utilities.

Both mapper and reduce is written for utility totalSalesByState for this problem.

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar map-reduce-prog-0.1.jar totalSalesByState hadoop/television.txt hadoop/results
18/10/02 23:40:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/10/02 23:40:05 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/10/02 23:40:06 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface
and execute your application with ToolRunner to remedy this.
18/10/02 23:40:06 INFO input.FileInputFormat: Total input paths to process: 1
18/10/02 23:40:06 INFO mapreduce.JobSubmitter: number of splits:1
18/10/02 23:40:07 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538470457623_0026
18/10/02 23:40:07 INFO impl.YarnClientImpl: Submitted application application_1538470457623_0026
18/10/02 23:40:07 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538470457623_0026/
18/10/02 23:40:07 INFO mapreduce.Job: Running job: job_1538470457623_0026
18/10/02 23:40:21 INFO mapreduce.Job: Job job_1538470457623_0026 running in uber mode : false
18/10/02 23:40:21 INFO mapreduce.Job: map 0% reduce 0%
18/10/02 23:40:32 INFO mapreduce.Job:
                                             map 100% reduce 0%
18/10/02 23:40:42 INFO mapreduce.Job: map 100% reduce 100%
18/10/02 23:40:44 INFO mapreduce.Job: Job job_1538470457623_0026 completed successfully
18/10/02 23:40:44 INFO mapreduce.Job: Counters: 49
         File System Counters
                  FILE: Number of bytes read=425
                  FILE: Number of bytes written=213825
                  FILE: Number of read operations=0
                  FILE: Number of large read operations=0
I
                  FILE: Number of write operations=0
                  HDFS: Number of bytes read=855
                  HDFS: Number of bytes written=165
                  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)=8972
                  Total time spent by all reduces in occupied slots (ms)=7569
                   Total time spent by all map tasks (ms)=8972
                  Total time spent by all reduce tasks (ms)=7569
                  Total vcore-seconds taken by all map tasks=8972
                  Total vcore-seconds taken by all reduce tasks=7569
```

```
[acadgild@localhost ~]$ hadoop fs -cat hadoop/results/part-r-00000

18/10/02 23:42:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable

AKAI-KERALA 1

LAVA-ASSAM 3

NA-UTTAR PRADESH 1

PNIDA-KERALA 1

UNIDA-UTTAR PRADESH 3

SAMSUNG-KERALA 1

SAMSUNG-KERALA 1

SAMSUNG-MADHYA PRADESH 3

SAMSUNG-MADHYA PRADESH 3

SAMSUNG-MAHARASHTRA 3

ZEN-MAHARASHTRA 2

[acadgild@localhost ~]$
```