Auto-tuning Hadoop Map Reduce

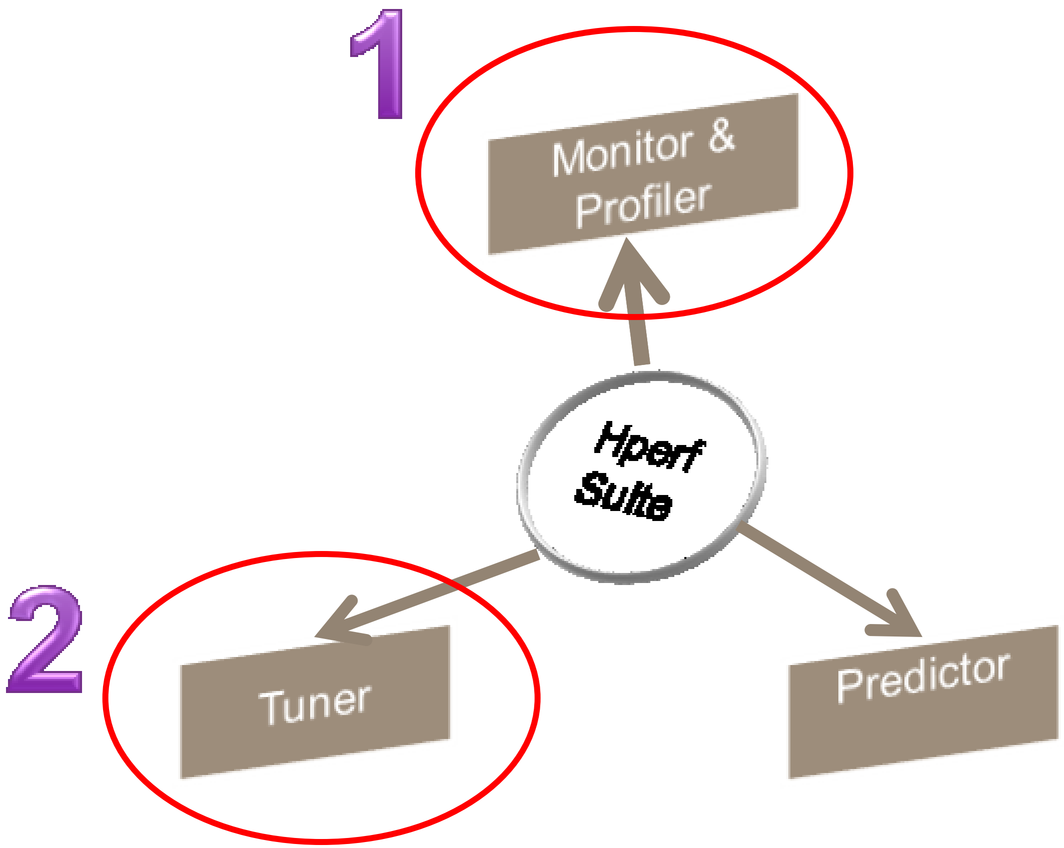
Manoj Nambiar,Ishtiyaque M. Shaikh

Performance Engineering Research Centre, TCS

Agenda

- Hperf Profiler
- Hperf Tuner
- Demo screenshots
- Case studies for Hperf Tuner

Hadoop based Tools and Experience



Hperf Profiler

How Hperf Profiler is different from other products?

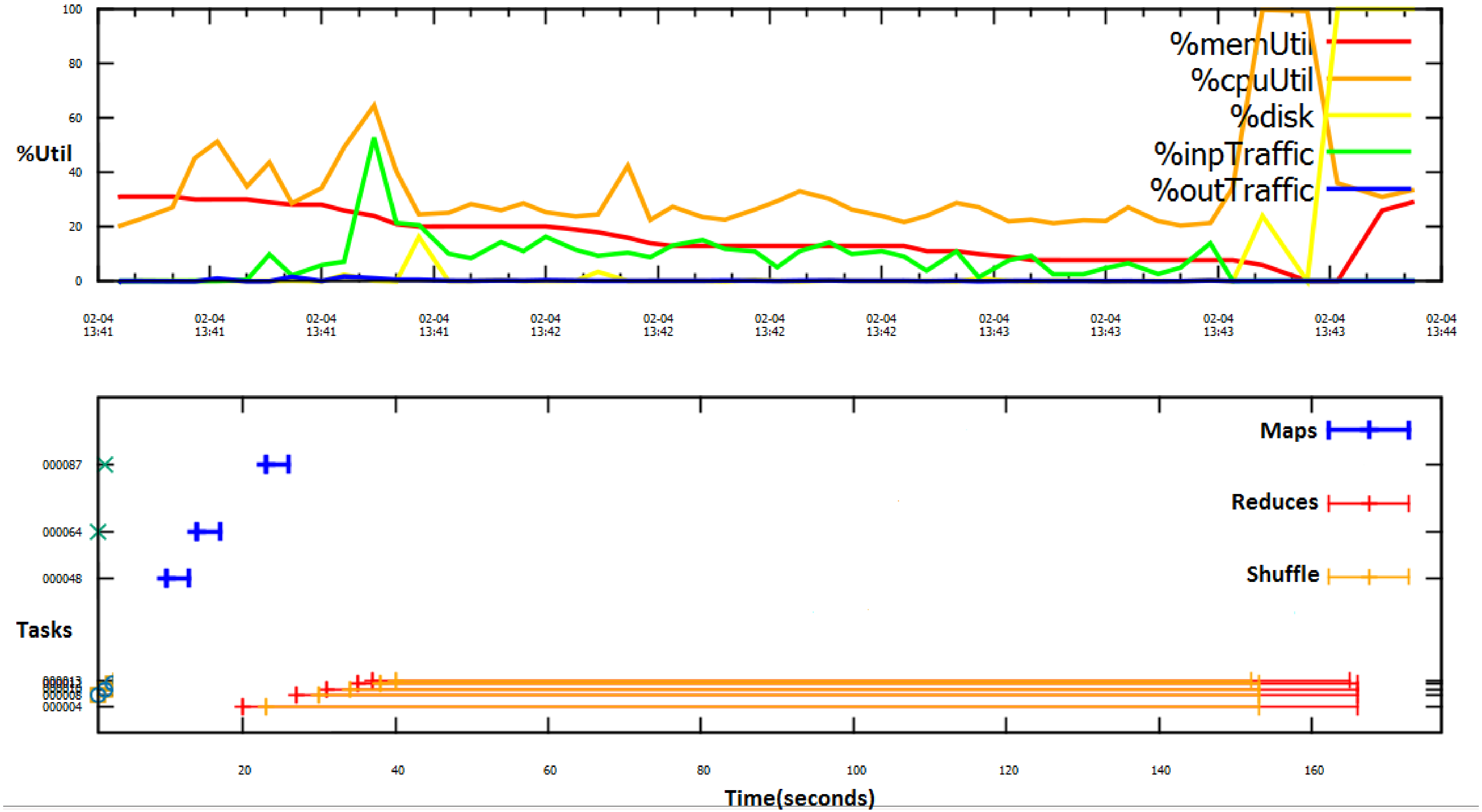
- Simple to install
- Better co-relation between map-reduce task, job and system counters
- In house – Free
- No Instrumentation
- Provides various views to represent system and MR job level details.

Hperf Profiler views:

- ❑ *Detailed MR Job view*
- ❑ *Consolidated System Utilization*
- ❑ *Detailed CPU view*
- ❑ *Task Level System Utilization*
- ❑ *Detailed Disk/Network View*

Profiler data for one of the nodes

Map/Reduce task with system utilization for n221

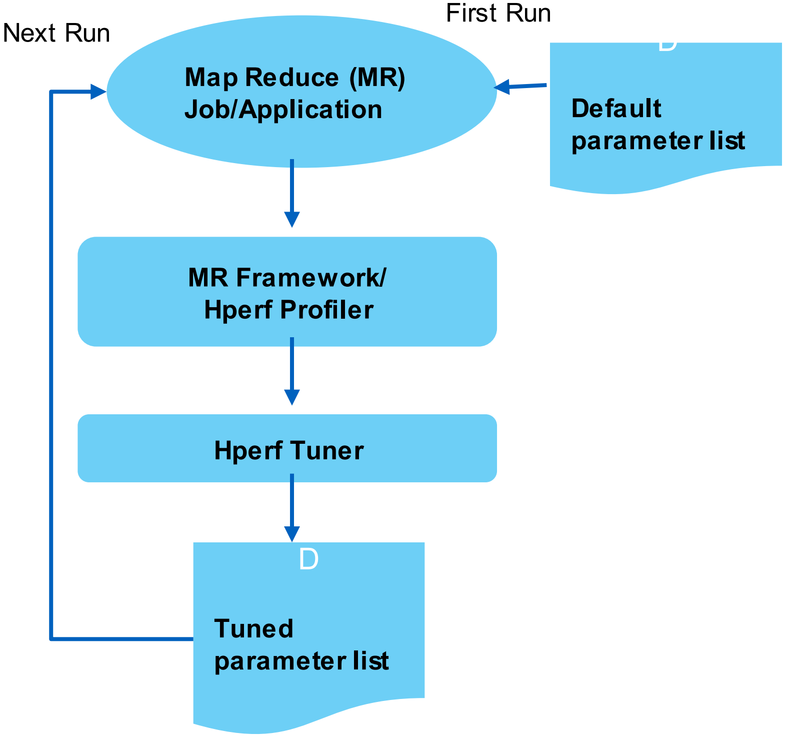


Easy analysis - Helps in finding optimization opportunities

Hperf Tuner

How Hperf Tuner is different from other products?

- No free open source MR tuner is available.
- Rule based tuning (Recommender) [2]
- Analytical optimization based tuning [1]
- Auto tuning capability.



Approach:

- Around 200 MR parameters
- OS and system level parameters
- Tune MR parameters to improve application performance

PERC Publication:

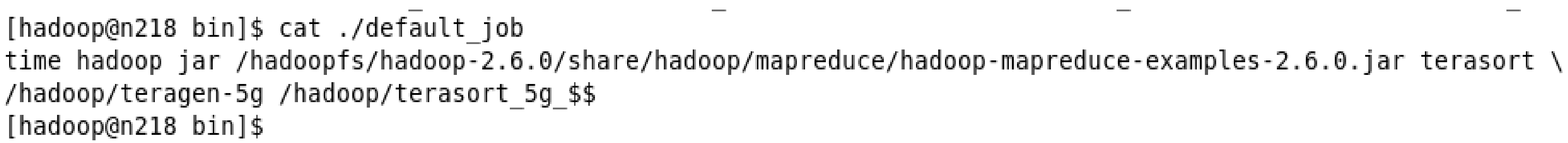
Scalable Resource Monitoring Tool for Hadoop 2, I Shaikh, Rekha Singhal, CMG INDIA, 2015.

References:

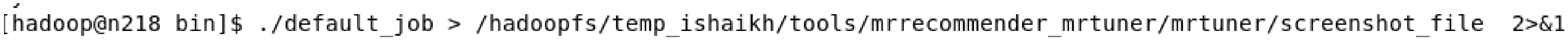
1. MRTuner: a toolkit to enable holistic optimization for mapreduce jobs, Proceedings of the VLDB Endowment , 7 Issue 13, August 2014 ,Pages 1319-1330 (Cost Based Optimization)
2. Hadoop Performance Tuning- A Pragmatic & Iterative Approach, Dominqu Heger, CMG USA, 2013. (Rule Based techniques)

Demo – Running the job

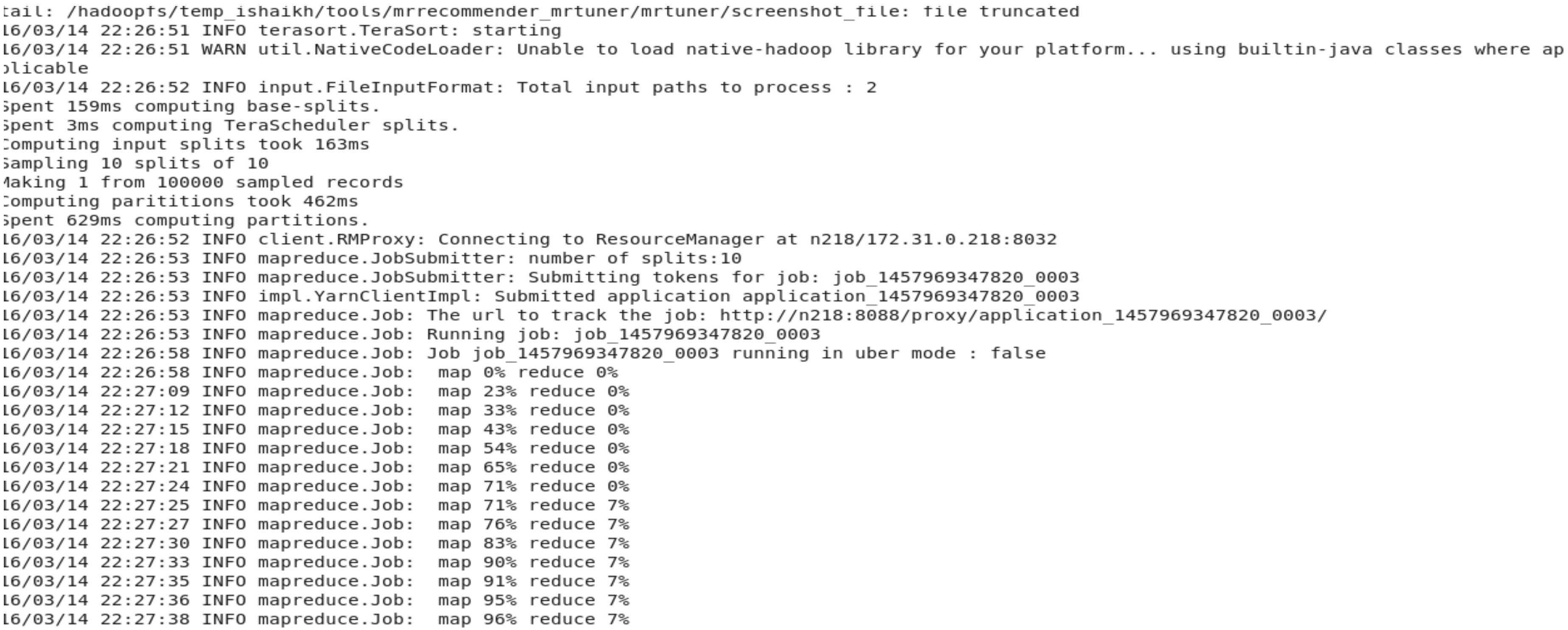
Running MR job with default settings



```
[hadoop@n218 bin]$ cat ./default_job
time hadoop jar /hadoopfs/hadoop-2.6.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar terasort \
/hadoop/teragen-5g /hadoop/terasort_5g_$$
[hadoop@n218 bin]$
```

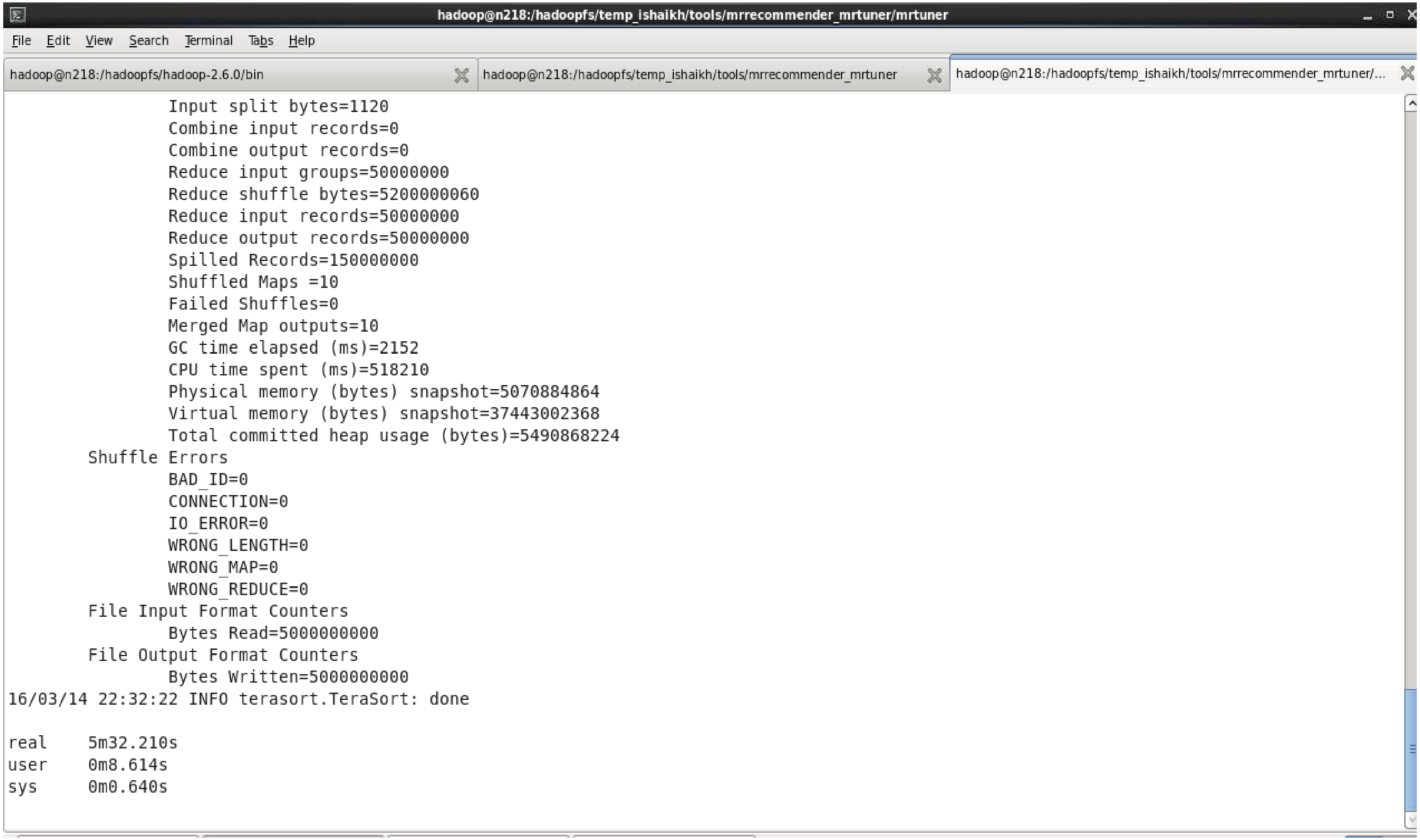


```
[hadoop@n218 bin]$ ./default_job > /hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner/mrtuner/screenshot_file  2>&1
```



```
tail: /hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner/mrtuner/screenshot_file: file truncated
16/03/14 22:26:51 INFO terasort.TeraSort: starting
16/03/14 22:26:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where ap
plicable
16/03/14 22:26:52 INFO input.FileInputFormat: Total input paths to process : 2
Spent 159ms computing base-splits.
Spent 3ms computing TeraScheduler splits.
Computing input splits took 163ms
Sampling 10 splits of 10
Making 1 from 100000 sampled records
Computing parititions took 462ms
Spent 629ms computing partitions.
16/03/14 22:26:52 INFO client.RMProxy: Connecting to ResourceManager at n218/172.31.0.218:8032
16/03/14 22:26:53 INFO mapreduce.JobSubmitter: number of splits:10
16/03/14 22:26:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1457969347820_0003
16/03/14 22:26:53 INFO impl.YarnClientImpl: Submitted application application_1457969347820_0003
16/03/14 22:26:53 INFO mapreduce.Job: The url to track the job: http://n218:8088/proxy/application_1457969347820_0003/
16/03/14 22:26:53 INFO mapreduce.Job: Running job: job_1457969347820_0003
16/03/14 22:26:58 INFO mapreduce.Job: Job job_1457969347820_0003 running in uber mode : false
16/03/14 22:26:58 INFO mapreduce.Job:  map 0% reduce 0%
16/03/14 22:27:09 INFO mapreduce.Job:  map 23% reduce 0%
16/03/14 22:27:12 INFO mapreduce.Job:  map 33% reduce 0%
16/03/14 22:27:15 INFO mapreduce.Job:  map 43% reduce 0%
16/03/14 22:27:18 INFO mapreduce.Job:  map 54% reduce 0%
16/03/14 22:27:21 INFO mapreduce.Job:  map 65% reduce 0%
16/03/14 22:27:24 INFO mapreduce.Job:  map 71% reduce 0%
16/03/14 22:27:25 INFO mapreduce.Job:  map 71% reduce 7%
16/03/14 22:27:27 INFO mapreduce.Job:  map 76% reduce 7%
16/03/14 22:27:30 INFO mapreduce.Job:  map 83% reduce 7%
16/03/14 22:27:33 INFO mapreduce.Job:  map 90% reduce 7%
16/03/14 22:27:35 INFO mapreduce.Job:  map 91% reduce 7%
16/03/14 22:27:36 INFO mapreduce.Job:  map 95% reduce 7%
16/03/14 22:27:38 INFO mapreduce.Job:  map 96% reduce 7%
```

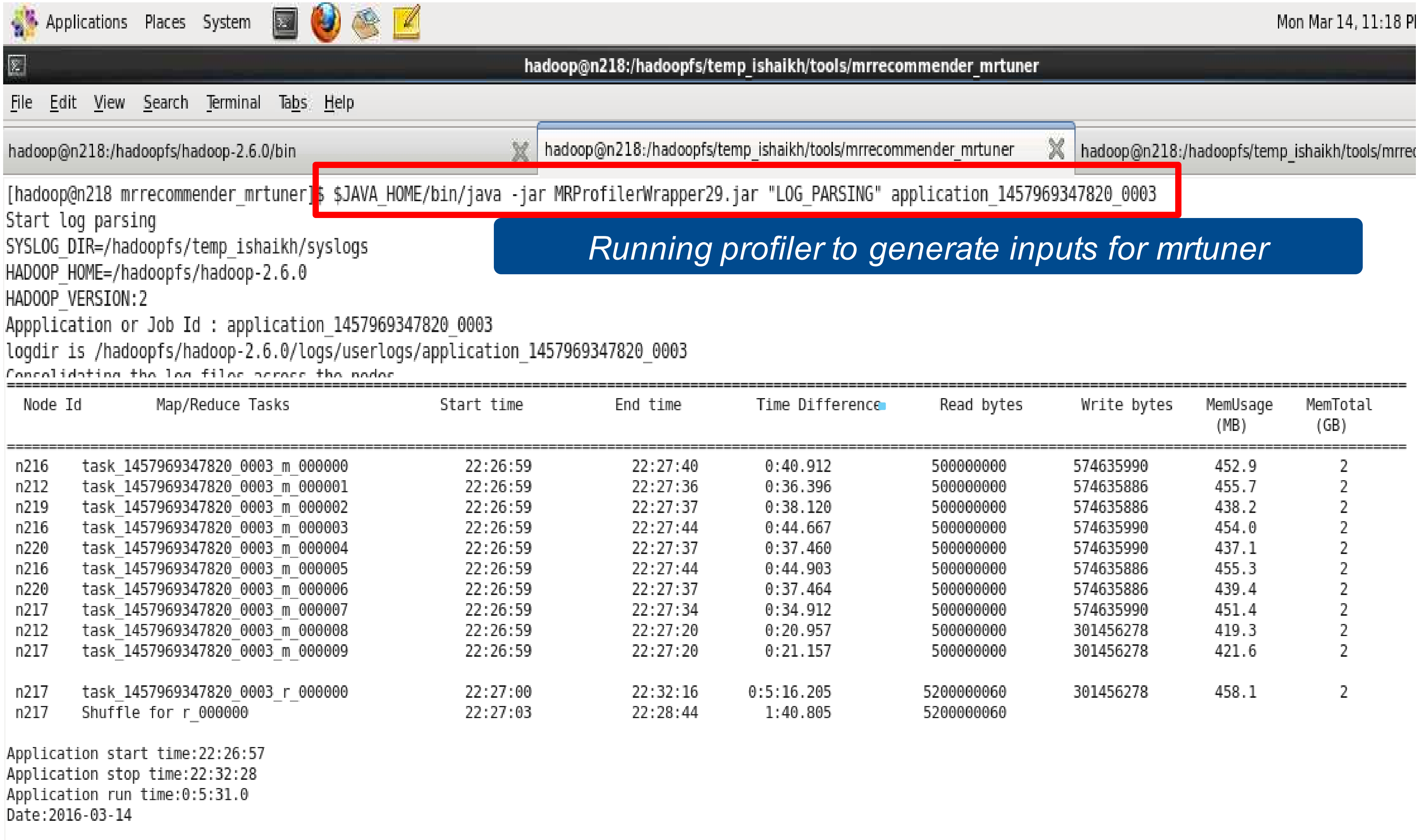
Demo – Job completes


hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner/mrtuner
File Edit View Search Terminal Tabs Help
hadoop@n218:/hadoopfs/hadoop-2.6.0/bin
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner/...

```
                Input split bytes=1120
                Combine input records=0
                Combine output records=0
                Reduce input groups=50000000
                Reduce shuffle bytes=5200000060
                Reduce input records=50000000
                Reduce output records=50000000
                Spilled Records=150000000
                Shuffled Maps =10
                Failed Shuffles=0
                Merged Map outputs=10
                GC time elapsed (ms)=2152
                CPU time spent (ms)=518210
                Physical memory (bytes) snapshot=5070884864
                Virtual memory (bytes) snapshot=37443002368
                Total committed heap usage (bytes)=5490868224
        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=5000000000
        File Output Format Counters
                Bytes Written=5000000000
16/03/14 22:32:22 INFO terasort.TeraSort: done

real    5m32.210s
user    0m8.614s
sys     0m0.640s
```

TATA CONSULTANCY SERVICES
Experience certainty.

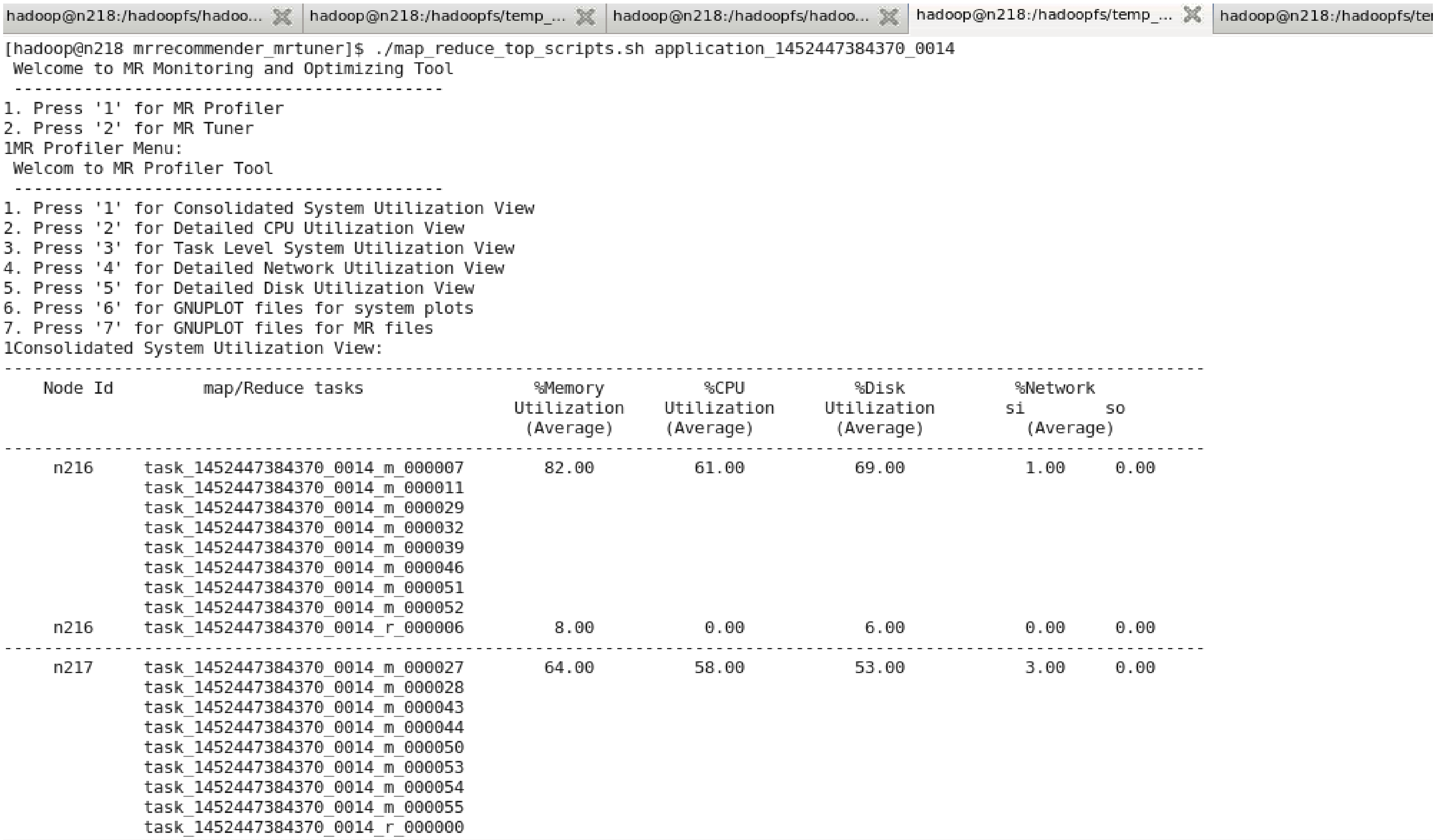
MR Profiler demo – Task Level View


Applications
Places
System
Mon Mar 14, 11:18 PI
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner
File
Edit
View
Search
Terminal
Tabs
Help
hadoop@n218:/hadoopfs/hadoop-2.6.0/bin
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrec

Running profiler to generate inputs for mrtuner

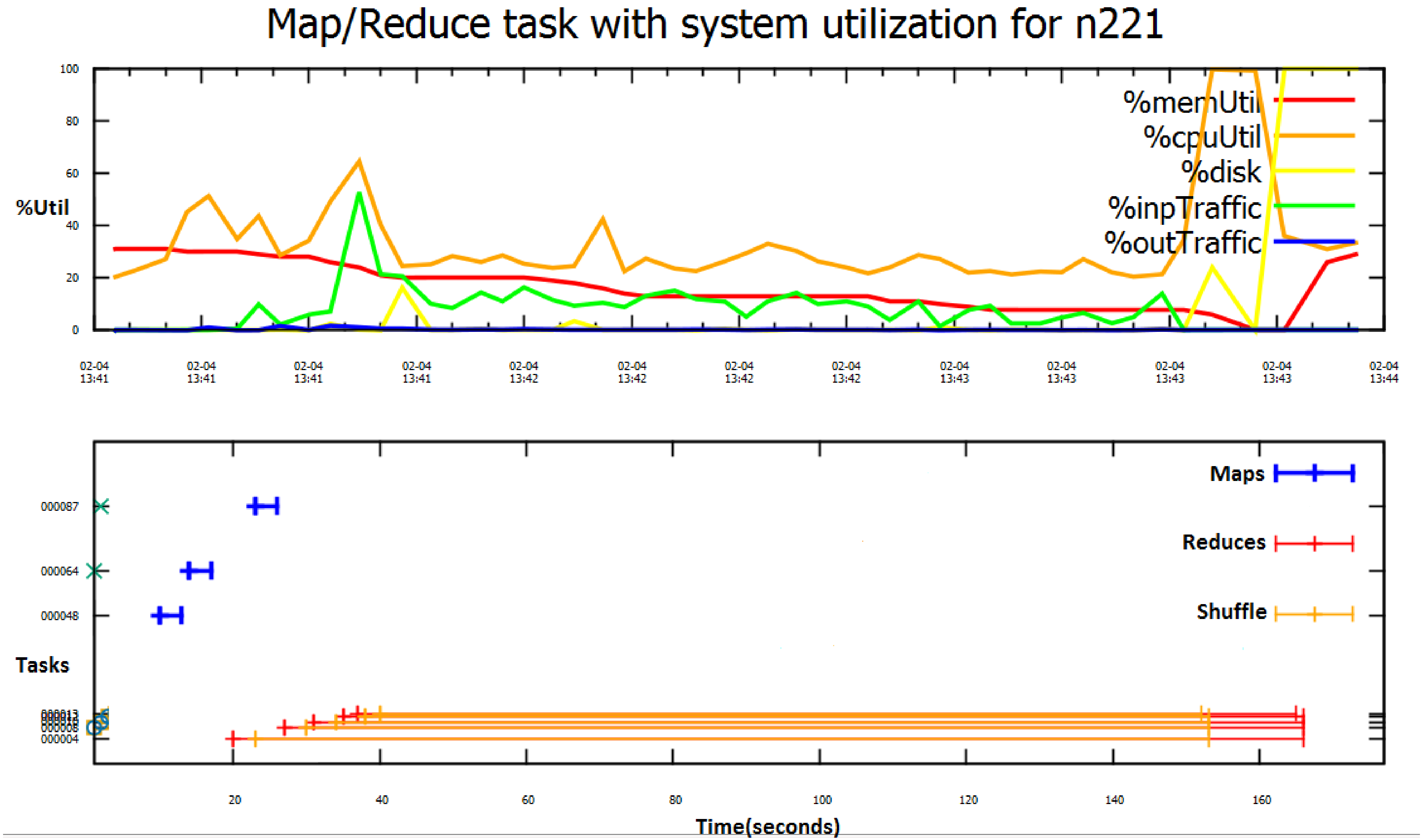
```
[hadoop@n218 mrrecommender_mrtuner]$ $JAVA_HOME/bin/java -jar MRProfilerWrapper29.jar "LOG_PARSING" application_1457969347820_0003
Start log parsing
SYSLOG_DIR=/hadoopfs/temp_ishaikh/syslogs
HADOOP_HOME=/hadoopfs/hadoop-2.6.0
HADOOP_VERSION:2
Appplication or Job Id : application_1457969347820_0003
logdir is /hadoopfs/hadoop-2.6.0/logs/userlogs/application_1457969347820_0003
Consolidating the log files across the nodes
====================================================================================================================================================
  Node Id        Map/Reduce Tasks                 Start time          End time          Time Difference        Read bytes        Write bytes     MemUsage    MemTotal
                                                                                                                                                   (MB)        (GB)
====================================================================================================================================================
 n216    task_1457969347820_0003_m_000000       22:26:59          22:27:40          0:40.912          500000000         574635990        452.9          2
 n212    task_1457969347820_0003_m_000001       22:26:59          22:27:36          0:36.396          500000000         574635886        455.7          2
 n219    task_1457969347820_0003_m_000002       22:26:59          22:27:37          0:38.120          500000000         574635886        438.2          2
 n216    task_1457969347820_0003_m_000003       22:26:59          22:27:44          0:44.667          500000000         574635990        454.0          2
 n220    task_1457969347820_0003_m_000004       22:26:59          22:27:37          0:37.460          500000000         574635990        437.1          2
 n216    task_1457969347820_0003_m_000005       22:26:59          22:27:44          0:44.903          500000000         574635886        455.3          2
 n220    task_1457969347820_0003_m_000006       22:26:59          22:27:37          0:37.464          500000000         574635886        439.4          2
 n217    task_1457969347820_0003_m_000007       22:26:59          22:27:34          0:34.912          500000000         574635990        451.4          2
 n212    task_1457969347820_0003_m_000008       22:26:59          22:27:20          0:20.957          500000000         301456278        419.3          2
 n217    task_1457969347820_0003_m_000009       22:26:59          22:27:20          0:21.157          500000000         301456278        421.6          2

 n217    task_1457969347820_0003_r_000000       22:27:00          22:32:16        0:5:16.205         5200000060         301456278        458.1          2
 n217    Shuffle for r_000000                   22:27:03          22:28:44          1:40.805         5200000060

Application start time:22:26:57
Application stop time:22:32:28
Application run time:0:5:31.0
Date:2016-03-14
```

MR Profiler Demo – Node Level View

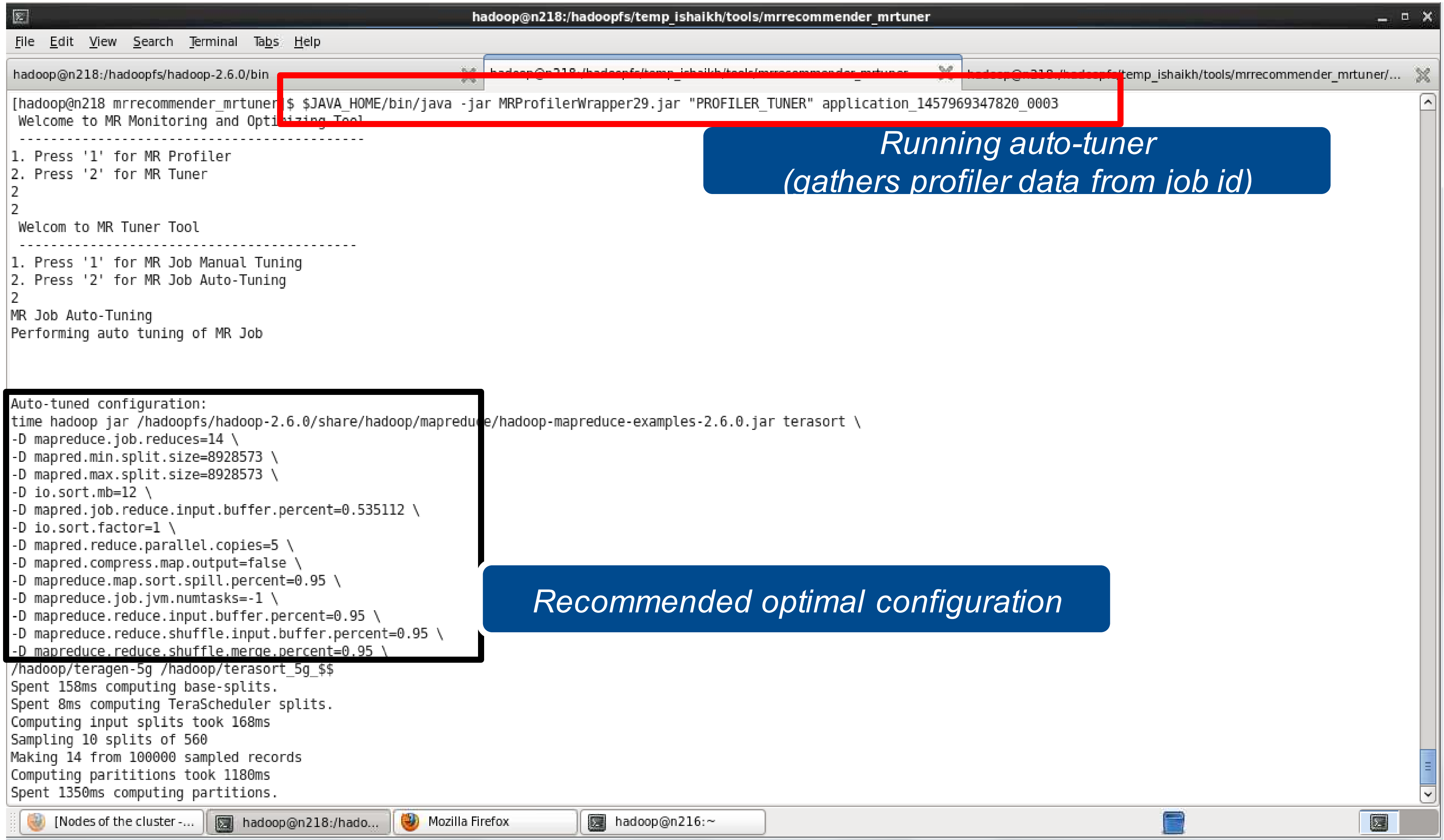

hadoop@n218:/hadoopfs/hadoo...
hadoop@n218:/hadoopfs/temp_...
hadoop@n218:/hadoopfs/hadoo...
hadoop@n218:/hadoopfs/temp_...
hadoop@n218:/hadoopfs/te

```
[hadoop@n218 mrrecommender_mrtuner]$ ./map_reduce_top_scripts.sh application_1452447384370_0014
 Welcome to MR Monitoring and Optimizing Tool
 -------------------------------------------
1. Press '1' for MR Profiler
2. Press '2' for MR Tuner
1MR Profiler Menu:
 Welcom to MR Profiler Tool
 -------------------------------------------
1. Press '1' for Consolidated System Utilization View
2. Press '2' for Detailed CPU Utilization View
3. Press '3' for Task Level System Utilization View
4. Press '4' for Detailed Network Utilization View
5. Press '5' for Detailed Disk Utilization View
6. Press '6' for GNUPLOT files for system plots
7. Press '7' for GNUPLOT files for MR files
1Consolidated System Utilization View:
------------------------------------------------------------------------------------------------------------------------
    Node Id         map/Reduce tasks                 %Memory          %CPU           %Disk           %Network
                                                   Utilization    Utilization     Utilization       si        so
                                                    (Average)     (Average)        (Average)          (Average)
------------------------------------------------------------------------------------------------------------------------
     n216     task_1452447384370_0014_m_000007        82.00          61.00           69.00           1.00     0.00
              task_1452447384370_0014_m_000011
              task_1452447384370_0014_m_000029
              task_1452447384370_0014_m_000032
              task_1452447384370_0014_m_000039
              task_1452447384370_0014_m_000046
              task_1452447384370_0014_m_000051
              task_1452447384370_0014_m_000052
     n216     task_1452447384370_0014_r_000006         8.00           0.00            6.00           0.00     0.00
------------------------------------------------------------------------------------------------------------------------
     n217     task_1452447384370_0014_m_000027        64.00          58.00           53.00           3.00     0.00
              task_1452447384370_0014_m_000028
              task_1452447384370_0014_m_000043
              task_1452447384370_0014_m_000044
              task_1452447384370_0014_m_000050
              task_1452447384370_0014_m_000053
              task_1452447384370_0014_m_000054
              task_1452447384370_0014_m_000055
              task_1452447384370_0014_r_000000
```

Graphical Plot from MR Profiler



MR Tuner demo



hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner
File Edit View Search Terminal Tabs Help
hadoop@n218:/hadoopfs/hadoop-2.6.0/bin
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner
hadoop@n218:/hadoopfs/temp_ishaikh/tools/mrrecommender_mrtuner/...
[Nodes of the cluster -...
hadoop@n218:/hado...
Mozilla Firefox
hadoop@n216:~

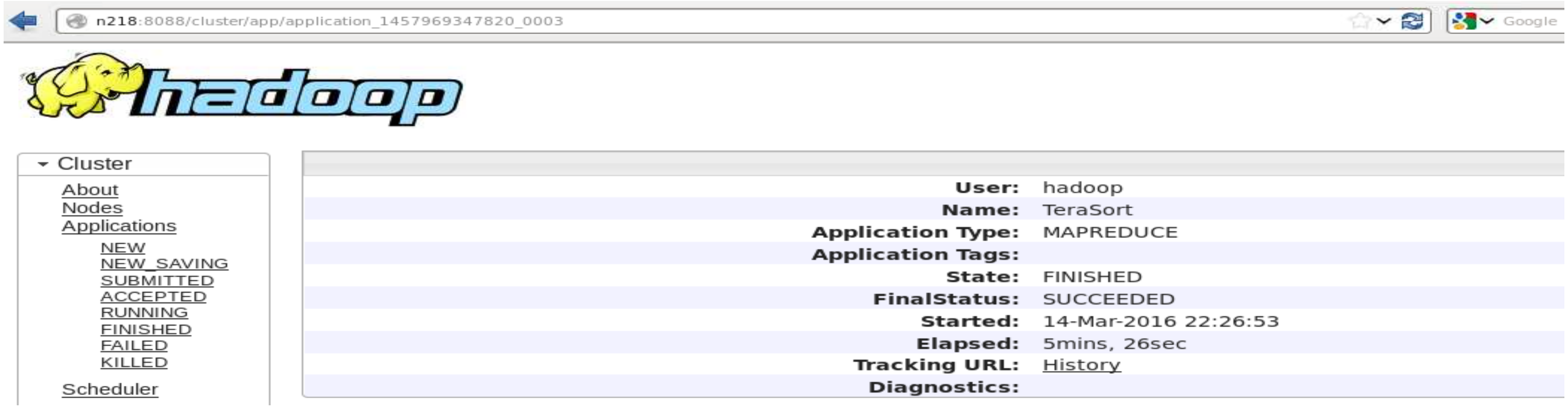
```
[hadoop@n218 mrrecommender_mrtuner]$ $JAVA_HOME/bin/java -jar MRProfilerWrapper29.jar "PROFILER_TUNER" application_1457969347820_0003
 Welcome to MR Monitoring and Optimizing Tool
 --------------------------------------------
1. Press '1' for MR Profiler
2. Press '2' for MR Tuner
2
2
 Welcom to MR Tuner Tool
 -------------------------------------------
1. Press '1' for MR Job Manual Tuning
2. Press '2' for MR Job Auto-Tuning
2
MR Job Auto-Tuning
Performing auto tuning of MR Job


Auto-tuned configuration:
time hadoop jar /hadoopfs/hadoop-2.6.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar terasort \
-D mapreduce.job.reduces=14 \
-D mapred.min.split.size=8928573 \
-D mapred.max.split.size=8928573 \
-D io.sort.mb=12 \
-D mapred.job.reduce.input.buffer.percent=0.535112 \
-D io.sort.factor=1 \
-D mapred.reduce.parallel.copies=5 \
-D mapred.compress.map.output=false \
-D mapreduce.map.sort.spill.percent=0.95 \
-D mapreduce.job.jvm.numtasks=-1 \
-D mapreduce.reduce.input.buffer.percent=0.95 \
-D mapreduce.reduce.shuffle.input.buffer.percent=0.95 \
-D mapreduce.reduce.shuffle.merge.percent=0.95 \
/hadoop/teragen-5g /hadoop/terasort_5g_$$
Spent 158ms computing base-splits.
Spent 8ms computing TeraScheduler splits.
Computing input splits took 168ms
Sampling 10 splits of 560
Making 14 from 100000 sampled records
Computing parititions took 1180ms
Spent 1350ms computing partitions.
```

Running auto-tuner
(gathers profiler data from job id)

Recommended optimal configuration

MR Tuner demo

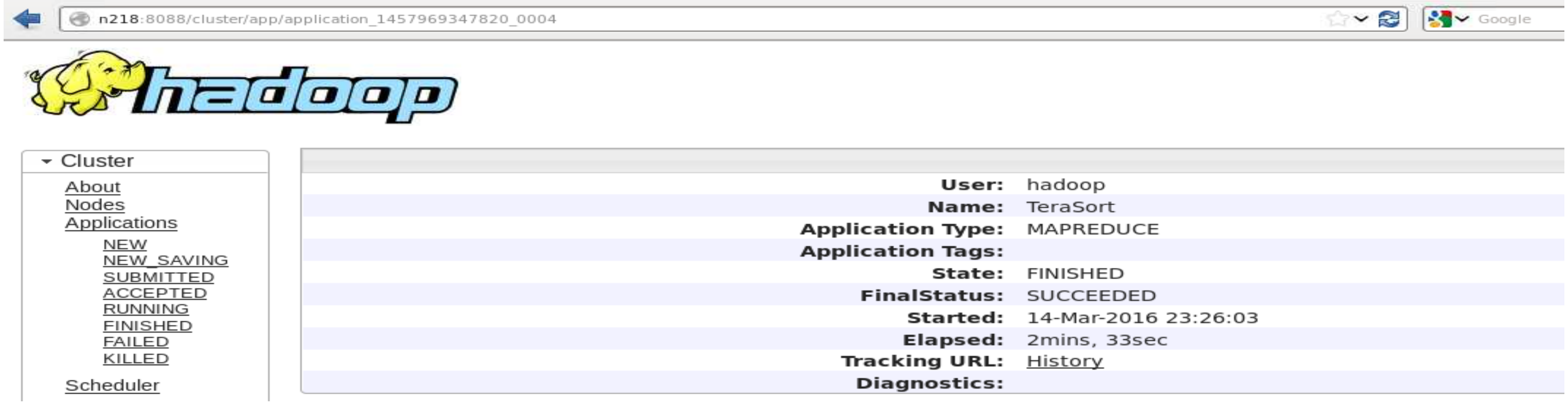
Default terasort



n218:8088/cluster/app/application_1457969347820_0003

User:	hadoop
Name:	TeraSort
Application Type:	MAPREDUCE
Application Tags:	
State:	FINISHED
FinalStatus:	SUCCEEDED
Started:	14-Mar-2016 22:26:53
Elapsed:	5mins, 26sec
Tracking URL:	History
Diagnostics:	

Tuned terasort



n218:8088/cluster/app/application_1457969347820_0004

User:	hadoop
Name:	TeraSort
Application Type:	MAPREDUCE
Application Tags:	
State:	FINISHED
FinalStatus:	SUCCEEDED
Started:	14-Mar-2016 23:26:03
Elapsed:	2mins, 33sec
Tracking URL:	History
Diagnostics:	

Case Study for Hive query Optimization

Parameters	Default	Tuned	Description
mapreduce.job.reduces	-1	4	The default number of reduce tasks per job. -1 indicate hive decide number of reduces.
mapreduce.task.io.sort.mb	100	1	The total amount of buffer memory to use while sorting files, in megabytes.
mapreduce.input.fileinputformat.split.minsize	268435456	1475104145	The minimum size chunk that map input should be split into.
mapreduce.task.io.sort.factor	10	1	The number of streams to merge at once while sorting files. This determines the number of open file handles.
Summary :			
No of maps	88	16	
No of reduces	24	4	
Input data size	20GB	20GB	
Query execution time	224.5sec	127sec	Gain = 43%
Configuration: 4 node each with 8 cores, 4 GB RAM.			
Query: select count(*),logeventid from hadoop_bpo_history_log_data_final group by logeventid;			

Case Studies for Hperf Tuner

Applications	Configuration	Data size	Performance gain/ job execution time
TCS Financial	Number of Nodes=8 cores=4, RAM=16GB	5GB	40%
	Number of Nodes=8 cores=4, RAM=16GB	40GB	36%
	Number of Nodes=8 Cores=56, RAM=132GB	7TB	13%+
Terasort	Number of nodes=8 Cores=4, RAM=16GB	5GB	31%
	Number of nodes=8 Cores=4, RAM=16GB	10GB	37%
Telecom Benchmark	Number of nodes=3 Cores=4, RAM=16GB	32GB	24%
Internal application Hive query	Number of nodes=4 Cores=8, RAM=4GB	22GB	47%

THANK YOU

QUESTIONS ??