

# Group 17: Miniproject 01 Report

Abhigyan Kishor(abk171), Saleh Almutairi(saa451), Quinn K Wolter(Qkw3)

VM IPs:

128.105.146.176(Public)

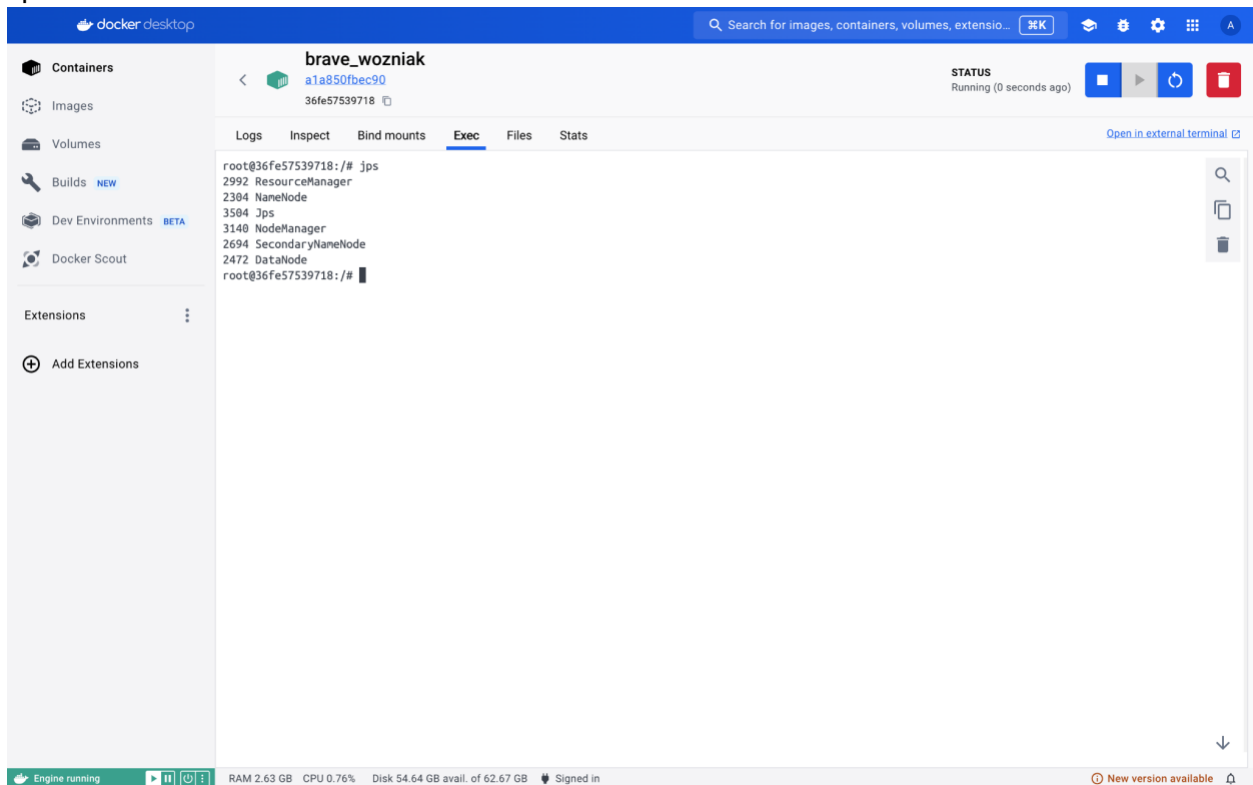
10.254.0.219(VM1),

10.254.3.84(VM2),

10.254.3.170(VM3)

Hadoop in Docker:

Jps



## Hdfs working

Containers

Images

Volumes

Builds NEW

Dev Environments BETA

Docker Scout

Extensions

Add Extensions

brave\_wozniak

a1a850bec90

36fe57539718

STATUS

Running (0 seconds ago)

Logs

Inspect

Bind mounts

Exec

Files

Stats

Open in external terminal

```
root@36fe57539718:/# jps
2992 ResourceManager
2304 NameNode
3140 NodeManager
2694 SecondaryNameNode
2472 DataNode
3532 Jps
root@36fe57539718:/# hdfs dfs -mkdir /input
2024-02-26 22:45:16,129 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applica
ble
root@36fe57539718:/# hdfs dfs -put /opt/hadoop/etc/hadoop/*.xml /input
2024-02-26 22:45:43,012 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applica
ble
2024-02-26 22:45:43,652 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:44,278 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:44,715 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:44,742 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:45,180 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:45,206 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:45,636 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:46,076 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-02-26 22:45:46,519 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
root@36fe57539718:/#
```

Engine running

RAM 2.69 GB CPU 0.25% Disk 54.64 GB avail. of 62.67 GB Signed in

New version available

Testing examples:

Hadoop jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples.3.2.1.jar  
wordcount /input /output

The screenshot shows the Docker Desktop application. On the left is a sidebar with navigation options: Containers, Images, Volumes, Builds (NEW), Dev Environments (BETA), Docker Scout, Extensions, and Add Extensions. The main area displays a terminal window for a container named 'brave\_wozniak' with ID 'a1a850fbec90' and IP '36fe57539718'. The terminal output shows Hadoop wordcount statistics:

```
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=237630
HDFS: Number of bytes written=9449
HDFS: Number of read operations=143
HDFS: Number of large read operations=0
HDFS: Number of write operations=12
HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework
  Map input records=745
  Map output records=3058
  Map output bytes=37534
  Map output materialized bytes=20421
  Input split bytes=969
  Combine input records=3058
  Combine output records=1219
  Reduce input groups=580
  Reduce shuffle bytes=20421
  Reduce input records=1219
  Reduce output records=580
  Spilled Records=2438
  Shuffled Maps =9
  Failed Shuffles=0
  Merged Map outputs=9
  GC time elapsed (ms)=30
  Total committed heap usage (bytes)=6111625216

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=27671
File Output Format Counters
  Bytes Written=9449

root@36fe57539718:/#
```

The bottom status bar indicates the Docker Engine is running, with system metrics: RAM 2.70 GB, CPU 1.02%, Disk 54.64 GB avail. of 62.67 GB, and a 'Signed in' status. A 'New version available' notification is also present.

Output:  
Hdfs dfs -cat /output/\*

Containers

Images

Volumes

Builds NEW

Dev Environments BETA

Docker Scout

Extensions

Add Extensions

brave\_wozniak

a1a850fbec90

36fe57539718

STATUS

Running (0 seconds ago)

Logs

Inspect

Bind mounts

Exec

Files

Stats

Open in external terminal

Timeline

timestamp

to

too

type="text/xml"

u:Xuser:Xuser

under

unique

updating

use

used

user

user.

user?

users

users,wheel"

uses

using

v2

valid.

value

value,

values

version="1.0"

version="1.0"?>

via

well

when

which

while

who

will

with

work

writing,

you

zero

root@36fe57539718:/#

RAM 2.70 GB

CPU 3.94%

Disk 54.64 GB avail. of 62.67 GB

Signed in

New version available

Hadoop in cluster:  
Namenode

```
ubuntu@vm-1:~$ jps
32307 Jps
30084 NameNode
30357 SecondaryNameNode
30588 ResourceManager
ubuntu@vm-1:~$
```

## Vm2

```
-- ssh -i ~/Downloads/group17 ubuntu@128.105.146.176 +
ubuntu@vm-2:~$ jps
5227 NodeManager
7086 Jps
4975 DataNode
ubuntu@vm-2:~$
```

## Vm3

```
-- ssh -i ~/Downloads/group17 ubuntu@128.105.146.176 +
ubuntu@vm-3:~$ jps
4432 Jps
4229 NodeManager
3978 DataNode
ubuntu@vm-3:~$
```

Q2.

hadoop jar ngram.jar ngram /q2input /output 3

```
-- ssh -i ~/Downloads/group17 ubuntu@128.105.146.176
-- ssh -i ~/Downloads/group17 ubuntu@128.105.146.176
+
Map-Reduce Framework
  Map input records=749
  Map output records=13952
  Map output bytes=111616
  Map output materialized bytes=32314
  Input split bytes=933
  Combine input records=13952
  Combine output records=3226
  Reduce input groups=1735
  Reduce shuffle bytes=32314
  Reduce input records=3226
  Reduce output records=1735
  Spilled Records=6452
  Shuffled Maps =9
  Failed Shuffles=0
  Merged Map outputs=9
  GC time elapsed (ms)=8366
  CPU time spent (ms)=9080
  Physical memory (bytes) snapshot=2023964672
  Virtual memory (bytes) snapshot=25375023104
  Total committed heap usage (bytes)=1269469184
  Peak Map Physical memory (bytes)=215252992
  Peak Map Virtual memory (bytes)=2536660992
  Peak Reduce Physical memory (bytes)=109428736
  Peak Reduce Virtual memory (bytes)=2545074176
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=26949
File Output Format Counters
  Bytes Written=10713
ubuntu@vm-1:~/miniproject1/q2$
```

hdfs dfs -cat /output/\*

```
-- ssh -i ~/Downloads/group17 ubuntu@128.105.146.176 -- ssh -i ~/Downloads/group17 ubuntu@128.105.146.176 +
y.c 4
y.d 2
y.h 1
y.i 2
y.j 2
y.m 3
y.n 2
y.p 3
y.q 3
y.r 17
y.w 1
y.z 1
y</ 5
y=2 1
y=4 1
y={ 2
yPr 2
ySc 2
yar 32
yin 1
yle 1
ynt 1
ype 1
ypi 1
ypt 5
yri 1
ys- 1
ysi 1
yst 1
yte 1
zat 1
zer 3
z kf 1
{na 2
{pr 2
|g| 1
ubuntu@vm-1:~/miniproject1/q2$
```



## Working with randomtextwriter

hadoop jar ~/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar

randomtextwriter /rtinput

```
2024-02-26 22:57:05,247 INFO mapreduce.Job: Job job_1708916616919_0017 running in uber mode : false
2024-02-26 22:57:05,251 INFO mapreduce.Job: map 0% reduce 0%
2024-02-26 22:57:15,470 INFO mapreduce.Job: map 100% reduce 0%
2024-02-26 22:57:16,527 INFO mapreduce.Job: Job job_1708916616919_0017 completed successfully
2024-02-26 22:57:16,728 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=225479
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=102
    HDFS: Number of bytes written=52569402
    HDFS: Number of read operations=6
    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
    Other local map tasks=1
    Total time spent by all maps in occupied slots (ms)=7800
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=7800
    Total vcore-milliseconds taken by all map tasks=7800
    Total megabyte-milliseconds taken by all map tasks=7987200
  Map-Reduce Framework
    Map input records=1
    Map output records=78780
    Input split bytes=102
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=218
    CPU time spent (ms)=2540
    Physical memory (bytes) snapshot=109314048
    Virtual memory (bytes) snapshot=2540879872
    Total committed heap usage (bytes)=32671392
    Peak Map Physical memory (bytes)=109314048
    Peak Map Virtual memory (bytes)=2540879872
  org.apache.hadoop.examples.RandomTextWriter$Counters
    BYTES WRITTEN=51628508
    RECORDS WRITTEN=78780
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=52569402
Job ended: Mon Feb 26 22:57:16 UTC 2024
The job took 25 seconds.
ubuntu@vm-1:~$
```

Running ngram.jar on rtinput

hadoop jar miniproject1/q2/ngram.jar ngram /rtinput /rtoutput 4

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=68317
Total time spent by all reduces in occupied slots (ms)=6972
Total time spent by all map tasks (ms)=68317
Total time spent by all reduce tasks (ms)=6972
Total vcore-milliseconds taken by all map tasks=68317
Total vcore-milliseconds taken by all reduce tasks=6972
Total megabyte-milliseconds taken by all map tasks=69956608
Total megabyte-milliseconds taken by all reduce tasks=7139328
Map-Reduce Framework
  Map input records=1291
  Map output records=28386171
  Map output bytes=256218939
  Map output materialized bytes=1191561
  Input split bytes=101
  Combine input records=28671714
  Combine output records=391444
  Reduce input groups=105901
  Reduce shuffle bytes=1191561
  Reduce input records=105901
  Reduce output records=105901
  Spilled Records=497345
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=998
  CPU time spent (ms)=63420
  Physical memory (bytes) snapshot=331976704
  Virtual memory (bytes) snapshot=5087260672
  Total committed heap usage (bytes)=170004480
  Peak Map Physical memory (bytes)=221532160
  Peak Map Virtual memory (bytes)=2541916160
  Peak Reduce Physical memory (bytes)=110444544
  Peak Reduce Virtual memory (bytes)=2545344512
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=52569402
File Output Format Counters
  Bytes Written=789969
ubuntu@vm-1:~$
ubuntu@vm-1:~$
```

Hdfs dfs -cat /rtoutput/\*

```
|int 3
|inv 1
|jap 1
|kar 1
|ken 1
|ker 1
|lau 1
|leu 1
|lim 1
|liq 1
|lit 2
|lop 2
|lou 3
|lyr 1
|mag 1
|mam 1
|man 1
|mar 2
|mas 1
|med 1
|men 1
|mes 1
|met 3
|min 2
|mon 2
|mut 1
|nau 1
|neu 1
|nig 1
|non 3
|obi 2
|oin 1
|ome 1
|ort 2
|out 2
|ove 3
|ovo 1
|pan 1
|par 1
|pel 2
|per 2
|phy 2
|pin 2
|pla 2
|plo 1
|pol 1
|pom 1
|por 1
|pos 1
```

Q3 - P1

~/hadoop/bin/hadoop jar ./SmiliesHitCounterJob.jar SmiliesHitCounterJob /q3input/ /p1/

```
Total time spent by all maps in occupied slots (ms)=776008
Total time spent by all reduces in occupied slots (ms)=90172
Total time spent by all map tasks (ms)=776008
Total time spent by all reduce tasks (ms)=90172
Total vcore-milliseconds taken by all map tasks=776008
Total vcore-milliseconds taken by all reduce tasks=90172
Total megabyte-milliseconds taken by all map tasks=794632192
Total megabyte-milliseconds taken by all reduce tasks=92336128
Map-Reduce Framework
  Map input records=9710351
  Map output records=330575
  Map output bytes=6611500
  Map output materialized bytes=298
  Input split bytes=1287
  Combine input records=330575
  Combine output records=10
  Reduce input groups=1
  Reduce shuffle bytes=298
  Reduce input records=10
  Reduce output records=1
  Spilled Records=20
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=16417
  CPU time spent (ms)=62940
  Physical memory (bytes) snapshot=2852020224
  Virtual memory (bytes) snapshot=35548774400
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=221028352
  Peak Map Virtual memory (bytes)=2544840704
  Peak Reduce Physical memory (bytes)=109932544
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=23
ubuntu@vm-1:~/miniproject1/q3/p1$
```

~/hadoop/bin/hdfs dfs -cat /p1/\*

```
ubuntu@vm-1:~/miniproject1/q3/p1$ ~/hadoop/bin/hdfs dfs -cat /p1/*
2024-02-26 03:54:51,343 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
/images/smilies 330575
```

**There were 330,575 hits to “/images/smilies”**

### Q3 - P2

~/hadoop/bin/hadoop jar ./IPHitCounterJob.jar IPHitCounterJob /q3input/ /p2/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=1959647
  Map output bytes=31354352
  Map output materialized bytes=168
  Input split bytes=1287
  Combine input records=1959647
  Combine output records=5
  Reduce input groups=1
  Reduce shuffle bytes=168
  Reduce input records=5
  Reduce output records=1
  Spilled Records=10
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=15312
  CPU time spent (ms)=59150
  Physical memory (bytes) snapshot=2887372800
  Virtual memory (bytes) snapshot=35530399744
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=215126016
  Peak Map Virtual memory (bytes)=2538905600
  Peak Reduce Physical memory (bytes)=109862912
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=20
ubuntu@vm-1:~/miniproject1/q3/p2$
```

~/hadoop/bin/hdfs dfs -cat /p2/\*

```
ubuntu@vm-1:~/miniproject1/q3/p2$ ~/hadoop/bin/hdfs dfs -cat /p2/*
2024-02-26 04:01:58,092 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
96.32.128.5 1959647
ubuntu@vm-1:~/miniproject1/q3/p2$
```

**There were 1,959,647 hits from the IP Address 96.32.128.5**

Q3 - P3

~/hadoop/bin/hadoop jar ./RequestMethodCounterJob.jar RequestMethodCounterJob /q3input/  
/p3/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=9710349
  Map output bytes=83850901
  Map output materialized bytes=845
  Input split bytes=1287
  Combine input records=9710349
  Combine output records=67
  Reduce input groups=16
  Reduce shuffle bytes=845
  Reduce input records=67
  Reduce output records=16
  Spilled Records=134
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=16035
  CPU time spent (ms)=70950
  Physical memory (bytes) snapshot=2896379904
  Virtual memory (bytes) snapshot=35548987392
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=216633344
  Peak Map Virtual memory (bytes)=2538762240
  Peak Reduce Physical memory (bytes)=108056576
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=146
ubuntu@vm-1:~/miniproject1/q3/p3$ .
```

~/hadoop/bin/hdfs dfs -cat /p3/\*

```
ubuntu@vm-1:~/miniproject1/q3/p3$ ~/hadoop/bin/hdfs dfs -cat /p3/*
2024-02-26 04:13:22,468 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
ASDE 1
DEBUG 5
DELETE 6
FLURP 6
GET 3542440
HEAD 5696259
INDEX 1
OPTIONS 26
POST 471552
PROPFIND 23
PUT 11
SEARCH 1
TRACE 10
TRACK 6
VIMF 1
get 1
ubuntu@vm-1:~/miniproject1/q3/p3$ .
```

**There were 5,696,259 HEAD requests, 3,542,441 GET requests, and 471,552 POST requests, along with a medley of other request types.**

Q3 - P4

~/hadoop/bin/hadoop jar ./MostHitPathJob.jar MostHitPathJob /q3input/ /p4/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=9710349
  Map output bytes=302733017
  Map output materialized bytes=322227377
  Input split bytes=1287
  Combine input records=0
  Combine output records=0
  Reduce input groups=5764203
  Reduce shuffle bytes=322227377
  Reduce input records=9710349
  Reduce output records=1
  Spilled Records=19420698
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=30213
  CPU time spent (ms)=121640
  Physical memory (bytes) snapshot=3683758080
  Virtual memory (bytes) snapshot=35521724416
  Total committed heap usage (bytes)=2617823232
  Peak Map Physical memory (bytes)=215457792
  Peak Map Virtual memory (bytes)=2536665088
  Peak Reduce Physical memory (bytes)=897118208
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=30
ubuntu@vm-1:~/miniproject1/q3/p4$
```

~/hadoop/bin/hdfs dfs -cat /p4/\*

```
ubuntu@vm-1:~/miniproject1/q3/p4$ ~/hadoop/bin/hdfs dfs -cat /p4/*
2024-02-26 04:24:23,519 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
/apache-log/access.log 630006
ubuntu@vm-1:~/miniproject1/q3/p4$
```

**The most hit path was “apache-log/access.log” with 63,006 hits.**

Q3 - P5

~/hadoop/bin/hadoop jar ./MostFrequentIPJob.jar MostFrequentIPJob /q3input/ /p5/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=9710351
  Map output bytes=172697742
  Map output materialized bytes=192118522
  Input split bytes=1287
  Combine input records=0
  Combine output records=0
  Reduce input groups=38197
  Reduce shuffle bytes=192118522
  Reduce input records=9710351
  Reduce output records=1
  Spilled Records=19420702
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=18488
  CPU time spent (ms)=87320
  Physical memory (bytes) snapshot=3063062528
  Virtual memory (bytes) snapshot=35521724416
  Total committed heap usage (bytes)=2172276736
  Peak Map Physical memory (bytes)=214335488
  Peak Map Virtual memory (bytes)=2536665088
  Peak Reduce Physical memory (bytes)=347066368
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=22
Job Complete!
```

~/hadoop/bin/hdfs dfs -cat /p5/\*

```
ubuntu@vm-1:~/miniproject1/q3/p5$ ~/hadoop/bin/hdfs dfs -cat /p5/*
2024-02-26 04:38:30,667 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
47.39.156.135 4305147
```

**The most prolific source of IP traffic was from 47.39.156.135 with 4,305,147 accesses.**



Q3 - P6

~/hadoop/bin/hadoop jar ./PostRequestCounterJob.jar PostRequestCounterJob /q3input/ /p6/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=471552
  Map output bytes=4243968
  Map output materialized bytes=155
  Input split bytes=1287
  Combine input records=471552
  Combine output records=7
  Reduce input groups=1
  Reduce shuffle bytes=155
  Reduce input records=7
  Reduce output records=1
  Spilled Records=14
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=15294
  CPU time spent (ms)=52860
  Physical memory (bytes) snapshot=2885169152
  Virtual memory (bytes) snapshot=35532210176
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=215494656
  Peak Map Virtual memory (bytes)=2538762240
  Peak Reduce Physical memory (bytes)=108924928
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=12
ubuntu@vm-1:~/miniproject1/q3/p6$
```

~/hadoop/bin/hdfs dfs -cat /p6/\*

```
ubuntu@vm-1:~/miniproject1/q3/p6$ ~/hadoop/bin/hdfs dfs -cat /p6/*
2024-02-26 04:55:24,044 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
POST 471552
```

**There were 471,552 POST requests.**

Q3 - P7

~/hadoop/bin/hadoop jar ./Status404CounterJob.jar Status404CounterJob /q3input/ /p7/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=7209667
  Map output bytes=57677336
  Map output materialized bytes=208
  Input split bytes=1287
  Combine input records=7209667
  Combine output records=13
  Reduce input groups=1
  Reduce shuffle bytes=208
  Reduce input records=13
  Reduce output records=1
  Spilled Records=26
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=14781
  CPU time spent (ms)=66090
  Physical memory (bytes) snapshot=2904866816
  Virtual memory (bytes) snapshot=35548987392
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=215994368
  Peak Map Virtual memory (bytes)=2538762240
  Peak Reduce Physical memory (bytes)=111607808
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=12
ubuntu@vm-1:~/miniproject1/q3/p7$
```

~/hadoop/bin/hdfs dfs -cat /p7/\*

```
ubuntu@vm-1:~/miniproject1/q3/p7$ ~/hadoop/bin/hdfs dfs -cat /p7/*
2024-02-26 05:01:12,213 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
404      7209667
```

**7,209,667 requests threw a 404 error.**

### Q3 - P8

~/hadoop/bin/hadoop jar ./DataRequestedOnDateJob.jar DataRequestedOnDateJob /q3input/ /p8/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=1135
  Map output bytes=18160
  Map output materialized bytes=96
  Input split bytes=1287
  Combine input records=1135
  Combine output records=1
  Reduce input groups=1
  Reduce shuffle bytes=96
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=15447
  CPU time spent (ms)=55170
  Physical memory (bytes) snapshot=2882641920
  Virtual memory (bytes) snapshot=35523821568
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=215080960
  Peak Map Virtual memory (bytes)=2538762240
  Peak Reduce Physical memory (bytes)=109633536
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=21
ubuntu@vm-1:~/miniproject1/q3/p8$
```

~/hadoop/bin/hdfs dfs -cat /p8/\*

```
ubuntu@vm-1:~/miniproject1/q3/p8$ ~/hadoop/bin/hdfs dfs -cat /p8/*
2024-02-26 05:07:51,349 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
19/Dec/2020    10337151
```

**10,337,151 bytes of data were requested on the 19th of December, 2020.**

Q3 - P9-1

~/hadoop/bin/hadoop jar TotalDataPerIPJob.jar TotalDataPerIPJob /q3input/ /p9-1/

```
total megabyte 11113seconds taken by Q1 Reduce
Map-Reduce Framework
  Map input records=9710351
  Map output records=9709284
  Map output bytes=211515723
  Map output materialized bytes=230934369
  Input split bytes=1287
  Combine input records=0
  Combine output records=0
  Reduce input groups=38156
  Reduce shuffle bytes=230934369
  Reduce input records=9709284
  Reduce output records=38156
  Spilled Records=19418568
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=18378
  CPU time spent (ms)=87590
  Physical memory (bytes) snapshot=3163213824
  Virtual memory (bytes) snapshot=35521724416
  Total committed heap usage (bytes)=2083418112
  Peak Map Physical memory (bytes)=215339008
  Peak Map Virtual memory (bytes)=2536665088
  Peak Reduce Physical memory (bytes)=380952576
  Peak Reduce Virtual memory (bytes)=2545078272
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=1754429284
File Output Format Counters
  Bytes Written=815741
buntu@vm-1:~/miniproject1/q3/p9-1$
```

~/hadoop/bin/hdfs dfs -cat /p9-1/\*

```
99.243.13.8 30078232
99.243.178.148 9931101
99.245.20.84 43377593
99.247.205.101 152333986
99.250.103.152 28401923
99.250.198.15 44196177
99.250.231.21 120105136
99.251.114.243 79630136
99.36.165.41 335986954
99.41.91.240 1966
99.42.33.227 21972450
99.44.48.177 19978692
99.46.24.122 135079985
99.47.88.187 173620180
99.6.132.105 24850025
99.6.134.237 783259914
99.6.135.72 391938008
99.67.58.183 2173550
99.7.17.72 20815092
99.8.58.234 1662328
99.9.78.146 27479648
99.90.186.205 6046464
99.96.70.125 3745317
c-73-169-232-206.hsd1.wa.comcast.net 25366
ip-095-208-160-029.um33.pools.vodafone-ip.de 19387
tor-exit-15.zbau.f3netze.de 223
buntu@vm-1:~/miniproject1/q3/p9-1$
```

**For this problem, we had to run two separate MapReduce operations. The first retrieved all IPs and the number of hits associated with them...**

~/hadoop/bin/hdfs dfs -cat /p9-1/\* > ~/p9-1.txt

~/hadoop/bin/hadoop fs -copyFromLocal ~/p9-1.txt /q3input-p9/

Q3 - P9-2

~/hadoop/bin/hadoop jar TopThreeIPsJob.jar TopThreeIPsJob /q3input-p9/ /p9-2/

```
10000 megabyte in 11111 seconds taken by 1111 Reduce
Map-Reduce Framework
  Map input records=38156
  Map output records=38156
  Map output bytes=845034
  Map output materialized bytes=921352
  Input split bytes=100
  Combine input records=0
  Combine output records=0
  Reduce input groups=38156
  Reduce shuffle bytes=921352
  Reduce input records=38156
  Reduce output records=3
  Spilled Records=76312
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=293
  CPU time spent (ms)=2930
  Physical memory (bytes) snapshot=321544192
  Virtual memory (bytes) snapshot=5082005504
  Total committed heap usage (bytes)=170004480
  Peak Map Physical memory (bytes)=212611072
  Peak Map Virtual memory (bytes)=2536660992
  Peak Reduce Physical memory (bytes)=108933120
  Peak Reduce Virtual memory (bytes)=2545344512
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=815741
File Output Format Counters
  Bytes Written=81
ubuntu@vm-1:~/miniproject1/q3/p9-2$
```

~/hadoop/bin/hdfs dfs -cat /p9-2/\*

```
ubuntu@vm-1:~/miniproject1/q3/p9-2$ ~/hadoop/bin/hdfs dfs -cat /p9-2/*
2024-02-26 05:24:16,693 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
14.139.110.137 55384128894
109.220.172.234 54013537755
51.210.183.78 36837061299
ubuntu@vm-1:~/miniproject1/q3/p9-2$
```

**And the next found the top three from the pile of data, giving us the impressive data flow numbers pictured above.**

Q3 - P10

~/hadoop/bin/hadoop jar Data200RequestedOnDateJob.jar Data200RequestedOnDateJob  
/q3input/ /p10/

```
Map-Reduce Framework
  Map input records=9710351
  Map output records=788
  Map output bytes=15760
  Map output materialized bytes=122
  Input split bytes=1287
  Combine input records=788
  Combine output records=2
  Reduce input groups=1
  Reduce shuffle bytes=122
  Reduce input records=2
  Reduce output records=1
  Spilled Records=4
  Shuffled Maps =13
  Failed Shuffles=0
  Merged Map outputs=13
  GC time elapsed (ms)=16883
  CPU time spent (ms)=52770
  Physical memory (bytes) snapshot=2825355264
  Virtual memory (bytes) snapshot=35523764224
  Total committed heap usage (bytes)=1819201536
  Peak Map Physical memory (bytes)=216055808
  Peak Map Virtual memory (bytes)=2538758144
  Peak Reduce Physical memory (bytes)=110145536
  Peak Reduce Virtual memory (bytes)=2545074176
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=1754429284
File Output Format Counters
  Bytes Written=23
ubuntu@vm-1:~/miniproject1/q3/p10$
```

~/hadoop/bin/hdfs dfs -cat /p10/\*

```
ubuntu@vm-1:~/miniproject1/q3/p10$ ~/hadoop/bin/hdfs dfs -cat /p10/*
2024-02-26 05:28:50,783 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
16/Jan/2022      2465568761
ubuntu@vm-1:~/miniproject1/q3/p10$
```

**Finally, we had 2465568761 bytes of data transferred in requests with a status of 200 "OK" on the 16th of January, 2022.**