**LARGE CSV FOR 3 DIGIT SELECTION:**

**OUTPUT FOR BLOCK SIZE:**

Status: HEALTHY

Total size: 2183910056 B

Total dirs: 1

Total files: 1

Total blocks (validated): 33 (avg. block size 66179092 B)

Minimally replicated blocks: 33 (100.0 %)

Over-replicated blocks: 0 (0.0 %)

Under-replicated blocks: 0 (0.0 %)

Mis-replicated blocks: 0 (0.0 %)

Default replication factor: 1

Average block replication: 1.0

Corrupt blocks: 0

Missing replicas: 0 (0.0 %)

Number of data-nodes: 3

Number of racks: 1

**NUMBER OF BLOCKS ON DATANODE1 DATNODE2 AND DATANODE 3 RESP:**

hadoop fsck /user/nxh130330/input\_files4 -files -location -blocks -racks | grep /default-rack/10.176.92.131 | wc -l

492

hadoop fsck /user/nxh130330/input\_files4 -files -location -blocks -racks | grep /default-rack/10.176.92.132 | wc -l

390

hadoop fsck /user/nxh130330/input\_files4 -files -location -blocks -racks | grep /default-rack/10.176.92.133 | wc -l

459

**Job Completion Summary:**

Job number: job\_201411071216\_0046

14/11/07 15:11:39 INFO mapred.JobClient: Counters: 31

14/11/07 15:11:39 INFO mapred.JobClient: Job Counters

14/11/07 15:11:39 INFO mapred.JobClient: Launched reduce tasks=1

14/11/07 15:11:39 INFO mapred.JobClient: SLOTS\_MILLIS\_MAPS=782963

14/11/07 15:11:39 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0

14/11/07 15:11:39 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0

14/11/07 15:11:39 INFO mapred.JobClient: Rack-local map tasks=4

14/11/07 15:11:39 INFO mapred.JobClient: Launched map tasks=33

14/11/07 15:11:39 INFO mapred.JobClient: Data-local map tasks=29

14/11/07 15:11:39 INFO mapred.JobClient: SLOTS\_MILLIS\_REDUCES=14621

14/11/07 15:11:39 INFO mapred.JobClient: File Input Format Counters

14/11/07 15:11:39 INFO mapred.JobClient: Bytes Read=2184041128

14/11/07 15:11:39 INFO mapred.JobClient: File Output Format Counters

14/11/07 15:11:39 INFO mapred.JobClient: Bytes Written=11580108

14/11/07 15:11:39 INFO mapred.JobClient: FileSystemCounters

14/11/07 15:11:39 INFO mapred.JobClient: FILE\_BYTES\_READ=195884221

14/11/07 15:11:39 INFO mapred.JobClient: HDFS\_BYTES\_READ=2184044791

14/11/07 15:11:39 INFO mapred.JobClient: FILE\_BYTES\_WRITTEN=295899668

14/11/07 15:11:39 INFO mapred.JobClient: HDFS\_BYTES\_WRITTEN=11580108

14/11/07 15:11:39 INFO mapred.JobClient: Map-Reduce Framework

14/11/07 15:11:39 INFO mapred.JobClient: Map output materialized bytes=98072645

14/11/07 15:11:39 INFO mapred.JobClient: Map input records=15669889

14/11/07 15:11:39 INFO mapred.JobClient: Reduce shuffle bytes=98072645

14/11/07 15:11:39 INFO mapred.JobClient: Spilled Records=10564957

14/11/07 15:11:39 INFO mapred.JobClient: Map output bytes=418563958

14/11/07 15:11:39 INFO mapred.JobClient: Total committed heap usage (bytes)=11667832832

14/11/07 15:11:39 INFO mapred.JobClient: CPU time spent (ms)=207830

14/11/07 15:11:39 INFO mapred.JobClient: Map input bytes=2183910056

14/11/07 15:11:39 INFO mapred.JobClient: SPLIT\_RAW\_BYTES=3663

14/11/07 15:11:39 INFO mapred.JobClient: Combine input records=16641318

14/11/07 15:11:39 INFO mapred.JobClient: Reduce input records=3523540

14/11/07 15:11:39 INFO mapred.JobClient: Reduce input groups=479684

14/11/07 15:11:39 INFO mapred.JobClient: Combine output records=4721889

14/11/07 15:11:39 INFO mapred.JobClient: Physical memory (bytes) snapshot=11580551168

14/11/07 15:11:39 INFO mapred.JobClient: Reduce output records=479684

14/11/07 15:11:39 INFO mapred.JobClient: Virtual memory (bytes) snapshot=35663065088

14/11/07 15:11:39 INFO mapred.JobClient: Map output records=15442969

**JOBTRACKER SUMMARY:**

2014-11-07 15:10:58,841 INFO org.apache.hadoop.mapred.JobTracker: jobToken generated and stored with users keys in /home/hadoop/mapred/job\_201411071216\_0046/jobToken

2014-11-07 15:10:58,855 INFO org.apache.hadoop.mapred.JobInProgress: job\_201411071216\_0046: nMaps=33 nReduces=1 max=-1

2014-11-07 15:10:58,855 INFO org.apache.hadoop.mapred.JobTracker: Job job\_201411071216\_0046 added successfully for user 'nxh130330' to queue 'default'

2014-11-07 15:10:58,855 INFO org.apache.hadoop.mapred.AuditLogger: USER=nxh130330 IP=10.176.92.129 OPERATION=SUBMIT\_JOB TARGET=job\_201411071216\_0046 RESULT=SUCCESS

2014-11-07 15:10:58,856 INFO org.apache.hadoop.mapred.JobTracker: Initializing job\_201411071216\_0046

2014-11-07 15:10:58,857 INFO org.apache.hadoop.mapred.JobInProgress: Initializing job\_201411071216\_0046

2014-11-07 15:10:58,880 INFO org.apache.hadoop.mapred.JobInProgress: Input size for job job\_201411071216\_0046 = 2183910056. Number of splits = 33

2014-11-07 15:10:58,882 INFO org.apache.hadoop.mapred.JobInProgress: job\_201411071216\_0046 LOCALITY\_WAIT\_FACTOR=1.0

2014-11-07 15:10:58,882 INFO org.apache.hadoop.mapred.JobInProgress: Job job\_201411071216\_0046 initialized successfully with 33 map tasks and 1 reduce tasks.

2014-11-07 15:11:38,989 INFO org.apache.hadoop.mapred.JobInProgress: Job job\_201411071216\_0046 has completed successfully.

2014-11-07 15:11:38,989 INFO org.apache.hadoop.mapred.JobInProgress$JobSummary: jobId=job\_201411071216\_0046,submitTime=1415394658841,launchTime=1415394658882,firstMapTaskLaunchTime=1415394660196,firstReduceTaskLaunchTime=1415394682970,firstJobSetupTaskLaunchTime=1415394658989,firstJobCleanupTaskLaunchTime=1415394697780,finishTime=1415394698989,numMaps=33,numSlotsPerMap=1,numReduces=1,numSlotsPerReduce=1,user=nxh130330,queue=default,status=SUCCEEDED,mapSlotSeconds=782,reduceSlotsSeconds=14,clusterMapCapacity=96,clusterReduceCapacity=96,jobName=RegionCrimeCount

2014-11-07 15:11:38,996 INFO org.apache.hadoop.mapred.JobHistory: Moving file:/home/hadoop/logs/history/job\_201411071216\_0046\_1415394658841\_nxh130330\_RegionCrimeCount to file:/home/hadoop/logs/history/done/version-1/cshdfs\_1415384193928\_/2014/11/07/000000

2014-11-07 15:11:38,999 INFO org.apache.hadoop.mapred.JobHistory: Moving file:/home/hadoop/logs/history/job\_201411071216\_0046\_conf.xml to file:/home/hadoop/logs/history/done/version-1/cshdfs\_1415384193928\_/2014/11/07/000000

**NAMENODE SUMMARY:**

2014-11-07 15:10:58,606 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.jar. blk\_207677569474527804\_195892

2014-11-07 15:10:58,639 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.jar from client DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,639 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.jar is closed by DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,640 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem: Increasing replication for /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.jar. New replication is 10

2014-11-07 15:10:58,683 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem: Increasing replication for /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.split. New replication is 10

2014-11-07 15:10:58,688 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.split. blk\_-5601357457987241634\_195893

2014-11-07 15:10:58,738 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.split from client DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,738 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.split is closed by DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,751 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.splitmetainfo. blk\_-6344914551388653288\_195894

2014-11-07 15:10:58,754 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.splitmetainfo from client DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,754 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.splitmetainfo is closed by DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,821 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.xml. blk\_7754961465313187256\_195895

2014-11-07 15:10:58,825 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.xml from client DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,825 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /tmp/hadoop-hadoop/mapred/staging/nxh130330/.staging/job\_201411071216\_0046/job.xml is closed by DFSClient\_NONMAPREDUCE\_1609303504\_1

2014-11-07 15:10:58,832 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /home/hadoop/mapred/job\_201411071216\_0046/job-info. blk\_-8548791862448500662\_195896

2014-11-07 15:10:58,834 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /home/hadoop/mapred/job\_201411071216\_0046/job-info from client DFSClient\_NONMAPREDUCE\_-851221293\_1

2014-11-07 15:10:58,834 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /home/hadoop/mapred/job\_201411071216\_0046/job-info is closed by DFSClient\_NONMAPREDUCE\_-851221293\_1

2014-11-07 15:10:58,837 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /home/hadoop/mapred/job\_201411071216\_0046/jobToken. blk\_8163970350897054082\_195897

2014-11-07 15:10:58,840 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /home/hadoop/mapred/job\_201411071216\_0046/jobToken from client DFSClient\_NONMAPREDUCE\_-851221293\_1

2014-11-07 15:10:58,840 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /home/hadoop/mapred/job\_201411071216\_0046/jobToken is closed by DFSClient\_NONMAPREDUCE\_-851221293\_1

2014-11-07 15:10:58,871 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /user/nxh130330/output\_files5/\_logs/history/job\_201411071216\_0046\_conf.xml. blk\_6400584345065098909\_195899

2014-11-07 15:10:58,876 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /user/nxh130330/output\_files5/\_logs/history/job\_201411071216\_0046\_conf.xml from client DFSClient\_NONMAPREDUCE\_1184788864\_25

2014-11-07 15:10:58,876 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /user/nxh130330/output\_files5/\_logs/history/job\_201411071216\_0046\_conf.xml is closed by DFSClient\_NONMAPREDUCE\_1184788864\_25

2014-11-07 15:11:24,473 INFO org.apache.hadoop.hdfs.StateChange: BLOCK\* allocateBlock: /user/nxh130330/output\_files5/\_logs/history/job\_201411071216\_0046\_1415394658841\_nxh130330\_RegionCrimeCount. blk\_8985314796248516025\_195899

2014-11-07 15:11:38,994 INFO org.apache.hadoop.hdfs.StateChange: Removing lease on /user/nxh130330/output\_files5/\_logs/history/job\_201411071216\_0046\_1415394658841\_nxh130330\_RegionCrimeCount from client DFSClient\_NONMAPREDUCE\_1184788864\_25

2014-11-07 15:11:38,994 INFO org.apache.hadoop.hdfs.StateChange: DIR\* completeFile: /user/nxh130330/output\_files5/\_logs/history/job\_201411071216\_0046\_1415394658841\_nxh130330\_RegionCrimeCount is closed by DFSClient\_NONMAPREDUCE\_1184788864\_25

**MULTIPLE REDUCERS:**

**OUTPUT FOR BLOCK SIZE:**

.Status: HEALTHY

Total size: 2183910056 B

Total dirs: 1

Total files: 1341

Total blocks (validated): 1341 (avg. block size 1628568 B)

Minimally replicated blocks: 1341 (100.0 %)

Over-replicated blocks: 0 (0.0 %)

Under-replicated blocks: 0 (0.0 %)

Mis-replicated blocks: 0 (0.0 %)

Default replication factor: 1

Average block replication: 1.0

Corrupt blocks: 0

Missing replicas: 0 (0.0 %)

Number of data-nodes: 3

Number of racks: 1

**NUMBER OF BLOCKS ON DATANODE1 DATNODE2 AND DATANODE 3 RESP:**

hadoop fsck /user/nxh130330/input\_files16 -files -location -blocks -racks | grep /default-rack/10.176.92.131 | wc -l

465

hadoop fsck /user/nxh130330/input\_files16 -files -location -blocks -racks | grep /default-rack/10.176.92.132 | wc -l

438

hadoop fsck /user/nxh130330/input\_files16 -files -location -blocks -racks | grep /default-rack/10.176.92.133 | wc -l

438