

2016

Hadoop Performance Analysis Using Mahout Recommender System

CSC#555 MINING BIG DATA
VAIBHAV SHARMA

Project Aim

1. To evaluate the performance of various Hadoop configurations using Mahout ALS Recommender Application
2. Develop a Recommender application using Map-Reduce

Data

To test the performance of various Hadoop Configurations, I started with the MovieLens data available at the following link: <http://grouplens.org/datasets/movielens/>

The movie ratings file was of 600 mb and had about 22 million records. As I needed files of different sized to have reasonable comparison, I wrote a python script which will generate random data very similar MovieLens data of different sizes. The data includes User Id, Movie Id and Ratings provided by the user for specific movie. This is a comma separated file. As the data was generated by the script, I did not have to worry about cleaning up the data.

```
#!/usr/bin/python
import random

def createRec(fileName,sUid,eUid):
    """
        Generates data similar to MovieLens for the range user ids provided
        input: fileName - Ouput file to be generated
               sUid - Starting user id
               eUid - Ending user id
    """
    outFile = open(fileName,"w")
    for uid in range(sUid,eUid):
        movSet = set() # User cannot rate same movie again
        for numMov in range(1,random.randint(10,5000)): # Number of movies generated randomly
            movId = random.randint(1,34210) # MovieId Generated Randomly
            if movId not in movSet:
                rating = random.randint(1,5) # Randomly generated Rating
                movSet.add(movId)
                rec = str(uid)+","+str(movId)+","+str(rating)+"\n"
                outFile.write(rec)
            else:
                continue
    outFile.close()

createRec("1.csv",1,10000) # Generates 300 mb file
createRec("2.csv",1,20000) # Generates 600 mb file
createRec("3.csv",1,30000) # Generates 900 mb file
```

Figure 1 Script to Generate Movie Rating data for Analysis

Part1: Analysis of Performance of various Hadoop Configuration

To evaluate the performance, I used the following configurations:

1. t2.small 3 node Cluster
2. t2.small 2 node Cluster
3. t2.small 1 node Cluster
4. t2.medium 3 node Cluster
5. t2.medium 2 node Cluster
6. t2.medium 1 node Cluster
7. t2.large 3 node Cluster
8. t2.large 2 node Cluster
9. t2.large 1 node Cluster

For each of the cluster configuration I use 1X and 3X replications. I used the following dataset sizes to test the performance: 300 mb, 600 mb, 900 mb, 1.2 G, 1.9G. 1.2 and 1.9 G were only possible in case 3 nodes cluster. For all the executions that were not successful due to size issues or application error, I have given the time of 6500 seconds for analysis. I have used the following commands:

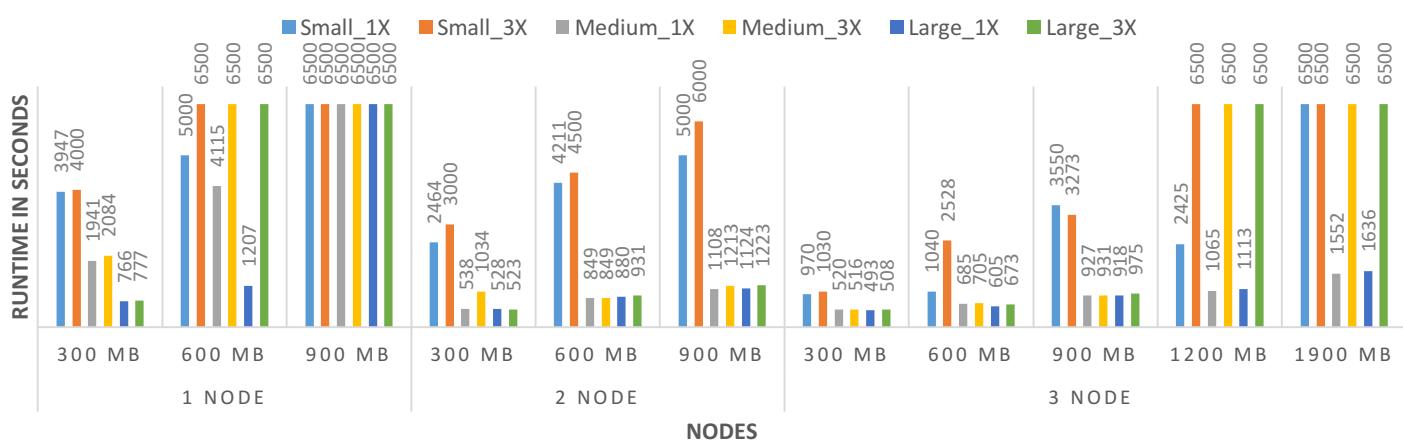
1. bin/mahout splitDataset --input /user/ec2-user/Project/ProjectData/ratings.csv --output /user/ec2-user/Project/ProjectData/step1 --trainingPercentage 0.9 --probePercentage 0.1 --tempDir /user/ec2-user/Project/ProjectData/step1/tmp
2. time bin/mahout parallelALS --input /user/ec2-user/Project/ProjectData/step1/trainingSet/ --output /user/ec2-user/Project/ProjectData/als/out --tempDir /user/ec2-user/Project/ProjectData/als/tmp --numFeatures 20 --numIterations 3 --lambda 0.065
3. time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRatings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData/als/out/U/ --itemFeatures /user/ec2-user/Project/ProjectData/als/out/M/ --numRecommendations 6 --maxRating 5
4. \$HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 | head

I have added the time from step 2 and 3 to calculate the total time of the process.

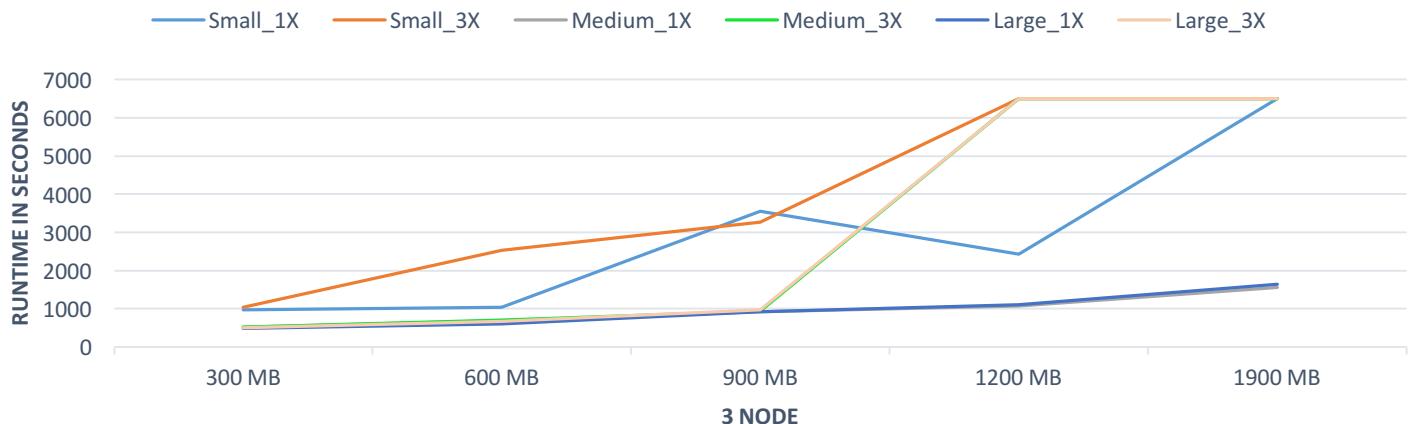
Following is the result in seconds for the configuration:

		t2.small 1X	t2.small 3X	t2.medium 1X	t2.medium 3X	t2.large 1X	t2.large 3X
1 Node	300 mb	3947	4000	1941	2084	766	777
	600 MB	5000	Unable to execute. NullPointerException	4115	Unable to execute. NullPointerException	1207	Unable to execute. NullPointerException
	900 mb	Unable to execute. NullPointerException					
2 Node	300 mb	2464	3000	538	1034	528	523
	600 MB	4211	4500	849	849	880	931
	900 mb	5000	6000	1108	1213	1124	1223
3 Node	300 mb	970	1030	520	516	493	508
	600 MB	1040	2528	685	705	605	673
	900 mb	3550	3273	927	931	918	975
	1.2 GB	2425	Unable to execute	1065	Unable to execute	1113	Unable to execute
	1.9 GB	Unable to execute. NullPointerException	Unable to execute	1552	Unable to execute	1636	Unable to execute

CLUSTER - NODE ANALYSIS



CLUSTER - NODE 3 ANALYSIS



From the above two graphs, we can clearly see t2.large performs well in the single node cluster compared to other configurations. But surprisingly, in a 2-Node cluster t2.medium takes less time compared to t2.large. This might be due to memory requirement might be less than the minimum provided by the t2.medium machines. This can also be confirmed from the results of t2.large and t2.medium; they are almost same except in the case of single node. There is drastic gap between t2.small and other configurations. For most of the jobs on t2.small, I have run the jobs for about an hour. The jobs on other machines complete in about 20 minutes or so. Also we can see the impact of replication by the increase in the time in the jobs.

Part2: Implement of Recommender System using Map-Reduce Paradigm

I have written a set of Map-Reduce codes to process the input ratings file in the following format: userId,movieId,rating

Step1: Process input file and creates user vectors i.e. a vector consisting of all movies rated by the user.

```

1 #!/usr/bin/python
2 import sys
3
4 # Mapper : To Process input file
5 # This Mapper-Reducer pair will create the user vector
6
7 for line in sys.stdin:
8     line = line.strip()
9     if line == "": # Cleaning up any blank lines
10        continue
11     lineContent = line.split(",")
12     uId = lineContent[0].strip()
13     if uId.lower() == "userid":
14        continue
15     movId = lineContent[1].strip()
16     print "%s\t%s"%(uId,movId)
17
18

```

Figure 2 Mapper to process input file and create user vector

```

1 #!/usr/bin/python
2 from __future__ import division
3 import sys
4
5 # Reducer: To Process input file
6 # This Mapper-Reducer pair will create the user vector
7
8 currId = None # To keep track of same userIDs
9 uId = None # Current UserIds
0 movId = None # Current MovieId
1 movList = []
2 for line in sys.stdin:
3     line = line.strip()
4     lineContent = line.split("\t")
5     uId = lineContent[0].strip()
6     movId = lineContent[1].strip()
7     if currId == None:
8         currId = uId
9         movList.append([movId,1])
10    elif currId == uId:
11        movList.append([movId,1])
12    else:
13        print "%s\t%s"%(currId, str([movList,"first"]))
14        movList = []
15        movList.append([movId,1])
16        currId = uId
17    if currId == uId and uId != None:
18        print "%s\t%s"%(uId, str([movList,"first"]))
19        movList = []
0

```

Figure 3 Reducer to create user vectors

Step2: This Map-Reduce job takes the input from step 1 and generates a co-occurrence matrix for the movies

```

1 #!/usr/bin/python
2 import sys
3
4 # Mapper: Takes the output file from the first Map-Reduce step
5 # This Map-Reduce will generate the co-occurrence vector for the movies
6
7
8 for line in sys.stdin:
9     line = line.strip()
10    if line == "":
11        continue
12    lineContent = line.split("\t")
13    con = eval(lineContent[1])
14    movList = con[0]
15    for item1 in movList:
16        for item2 in movList:
17            print "%s\t%s"%(item1[0],item2[0])
18
19

```

Figure 4 Mapper to create Co-occurrence Matrix

```

1 #!/usr/bin/python
2 from __future__ import division
3 import sys
4 currMovId = None
5 movId = None
6 movCount = {}
7
8 # Reducer : Takes the ouput file from the first Map-Reduce step
9 # This Map-Reduce will generate the co-occurece vecor for the the movies
0
1 for line in sys.stdin:
2     line = line.strip()
3     if line == "":
4         continue
5     lineContent = line.split("\t")
6     movId = lineContent[0].strip()
7     otherId = lineContent[1].strip()
8     if currMovId == None:
9         currMovId = movId
10        movCount.setdefault(otherId,0)
11        movCount[otherId] += 1
12     elif currMovId == movId:
13         movCount.setdefault(otherId,0)
14         movCount[otherId] += 1
15     else:
16         mCount = []
17         for item in movCount:
18             mCount.append([item,movCount[item]])
19         print "%s\t%s"%(currMovId,str([mCount,"second"]))
20         mCount = []
21         movCount = {}
22         movCount.setdefault(otherId,0)
23         movCount[otherId] += 1
24         currMovId = movId
25 if currMovId == movId and movId != None:
26     mCount = []
27     for item in movCount:
28         mCount.append([item,movCount[item]])
29     print "%s\t%s"%(currMovId,str([mCount,"second"]))
30

```

Figure 5 Reducer to create Co-occurrence Matrix

Step3: This is first part of Matrix Multiplication. This job takes the input from first mapper and original input calculates partial matrix multiplication

```
1 #!/usr/bin/python
2 import sys
3
4 for line in sys.stdin:
5     line = line.strip()
6     if line == "":
7         continue
8     lineCon = line.split("\t")
9     key = lineCon[0]
10    val = eval(lineCon[1])
11    if val[1] == "first":
12        movLst = val[0]
13        for item in movLst:
14            mov = item[0]
15            pref = item[1]
16            value = [[key,pref],"user"]
17            print "%s\t%s"%(mov,str(value))
18    else:
19        movLst = val[0]
20        val = [movLst,"co-oc"]
21        print "%s\t%s"%(key,str(val))
```

Figure 6 Partial Matrix Multiplication Mapper

```
1 #!/usr/bin/python
2
3 import sys
4 currMovId = None
5 movId = None
6 uid = None
7 pref = None
8 userInfo = []
9 co_oc = None
0
1
2 for line in sys.stdin:
3     line = line.strip()
4     lineCon = line.split("\t")
5     movId = lineCon[0]
6     con = eval(lineCon[1])
7     if currMovId == None:
8         currMovId = movId
9         if con[1] == "co-oc":
0             co_oc = con[0]
1         else:
2             userInfo.append(con[0])
3     elif currMovId == movId :
4         if con[1] == "co-oc":
5             co_oc = con[0]
6         else:
7             userInfo.append(con[0])
8     else:
9         for user in userInfo:
0             uid = user[0]
1             pref = user[1]
2             for item in co_oc:
3                 otherMov = item[0]
4                 count = item[1]
5                 parMul = count * pref
6                 key = uid
7                 print "%s\t%s"%(str(key),str([otherMov,parMul]))
8             co_co = None
9             userInfo = []
0             pref = None
1             uid = None
2             currMovId = movId
3             if con[1] == "co-oc":
4                 co_oc = con[0]
5             else:
6                 userInfo.append(con[0])
7     if currMovId == movId and movId != None:
8         for user in userInfo:
9             uid = user[0]
0             pref = (user[1])
1             for item in co_oc:
2                 otherMov = item[0]
3                 count = item[1]
4                 parMul = count * pref
5                 key = uid
6                 print "%s\t%s"%(str(key),str([otherMov,parMul]))
```

Figure 7 Partial Matrix Multiplication Reducer

Step4 : This Map-Reduce job completes the Matrix Multiplication process. It takes the input from the previous job and completes the addition of vecotors to complete matrix multiplication

```
#!/usr/bin/python
import sys

for line in sys.stdin:
    line = line.strip()
    if line == "":
        continue
    print line
```

Figure 8 Matrix Multiplication Final Step Mapper

```
1 #!/usr/bin/python
2 import sys
3
4 currId = None
5 uid = None
6 mulList = []
7
8 for line in sys.stdin:
9     line = line.strip()
10    if line == "":
11        continue
12    lineCon = line.split("\t")
13    uid = lineCon[0]
14    lst = eval(lineCon[1])
15    if currId == None:
16        currId = uid
17        mulList.append(lst)
18    elif currId == uid:
19        mulList.append(lst)
20    else:
21        mulRes = {}
22        for item in mulList:
23            mov = item[0]
24            parMul = item[1]
25            mulRes.setdefault(mov, 0)
26            mulRes[mov] += parMul
27        val = []
28        for mov in mulRes:
29            val.append([mov, mulRes[mov]])
30        print "%s\t%s"%(currId, str([val, "matrix"]))
31        currId = uid
32        mulList = []
33        mulList.append(lst)
34 if currId == uid and uid != None:
35     mulRes = {}
36     for item in mulList:
37         mov = item[0]
38         parMul = item[1]
39         mulRes.setdefault(mov, 0)
40         mulRes[mov] += parMul
41     val = []
42     for mov in mulRes:
43         val.append([mov, mulRes[mov]])
44     print "%s\t%s"%(currId, str([val, "matrix"]))
45
```

Figure 9 Matrix Multiplication Final Step Reducer

Step5: This is the final step in the process. It takes the original input file and file form previous process to produces recommendations for the user.

```
1 #!/usr/bin/python
2 import sys
3
4 for line in sys.stdin:
5     line = line.strip()
6     if line == "":
7         continue
8     print line
9
```

Figure 10 Recommendation Mapper

```

2 import sys
3 currid = None
4 uid = None
5 matrixList = None
6 movRatedList = None
7
8 for line in sys.stdin:
9     line = line.strip()
10    if line == "":
11        continue
12    lineCon = line.split("\t")
13    uid = lineCon[0]
14    lst = eval(lineCon[1])
15    if currid == None:
16        currid = uid
17        if lst[-1] == "matrix":
18            matrixList = lst[0]
19        else:
20            movRatedList = lst[0]
21    elif currid == uid:
22        if lst[-1] == "matrix":
23            matrixList = lst[0]
24        else:
25            movRatedList = lst[0]
26    else:
27        if matrixList == None or movRatedList == None:
28            print "%s\t%s"%(uid,"Cannot rate user")
29        else:
30            recNoList = {}
31            for item in movRatedList:
32                recNoList[item[0]] = item[1]
33            reclist = {}
34            for item in matrixList:
35                if item[0] not in recNoList:
36                    reclist.setdefault(item[1],[])
37                    reclist[item[1]].append(item[0])
38            scores = reclist.keys()
39            scores.sort()
40            scores.reverse()
41            rec = []
42            for score in scores:
43                for item in reclist[score]:
44                    rec.append(item)
45            print "%s\t%s"%(currid,str(rec[:20]))
46
47
48 matrixList = None
49 movRatedList = None
50 currid = uid
51 if lst[-1] == "matrix":
52     matrixList = lst[0]
53 else:
54     movRatedList = lst[0]
55 if currid == uid and uid != None:
56     if matrixList == None or movRatedList == None:
57         print "%s\t%s"%(uid,"Cannot rate user")
58     else:
59         recNoList = {}
60         for item in movRatedList:
61             recNoList[item[0]] = item[1]
62         reclist = {}
63         for item in matrixList:
64             if item[0] not in recNoList:
65                 reclist.setdefault(item[1],[])
66                 reclist[item[1]].append(item[0])
67         scores = reclist.keys()
68         scores.sort()
69         scores.reverse()
70         rec = []
71         for score in scores:
72             for item in reclist[score]:
73                 rec.append(item)
74         print "%s\t%s"%(currid,str(rec[:20]))

```

Figure 11 Recommendation Reducer

These 5 Map-Reduce jobs worked fine on a very small dataset but when executed on a 300 mb data set on t2.large machine 1X replication, it was not successful. As second and third steps I get a Broken Pipe Error. In all, I was not able to test out the complete application and compare it with the Mahout Recommender application.

```
projectMasterNode >> time $HADOOP_HOME/bin/hadoop jar $HADOOP_HOME/hadoop-streaming-0.20.205.0.jar -D mapred.reduce.tasks=5 -D mapred.map.tasks=10 -input /user/ec2-user/Project/ProjectData/ratings.csv -output /user/ec2-user/Project/first -mapper ~/Project/code/matrixapproach/firstMapper.py -file ~/Project/code/matrixapproach/firstMapper.py -reducer ~/Project/code/matrixapproach/firstReducer.py -file ~/Project/code/matrixapproach/firstReducer.py
Warning: $HADOOP_HOME is deprecated.

packageJobJar: [/home/ec2-user/Project/code/matrixapproach/firstMapper.py, /home/ec2-user/Project/code/matrixapproach/firstReducer.py, /tmp/hadoop-ec2-user/hadoop-unjar802723589594886798/] [] /tmp/streamjob1512513957232569217.jar tmpDir=null
16/06/11 21:58:01 INFO mapred.FileInputFormat: Total input paths to process : 1
16/06/11 21:58:01 INFO net.NetworkTopology: Adding a new node: /default-rack/172.31.55.98:50010
16/06/11 21:58:01 INFO streaming.StreamJob: getLocalDirs(): [/tmp/hadoop-ec2-user/mapred/local]
16/06/11 21:58:01 INFO streaming.StreamJob: Running job: job_201606112134_0004
16/06/11 21:58:01 INFO streaming.StreamJob: To kill this job, run:
16/06/11 21:58:01 INFO streaming.StreamJob: /home/ec2-user/hadoop-0.20.205.0/libexec/../bin/hadoop job -Dmapred.job.tracker=172.31.55.98:9001 -kill job_201606112134_0004
16/06/11 21:58:01 INFO streaming.StreamJob: Tracking URL: http://ip-172-31-55-98.ec2.internal:50030/jobdetails.jsp?jobId=job_201606112134_0004
16/06/11 21:58:02 INFO streaming.StreamJob: map 0% reduce 0%
16/06/11 21:58:19 INFO streaming.StreamJob: map 34% reduce 0%
16/06/11 21:58:22 INFO streaming.StreamJob: map 55% reduce 0%
16/06/11 21:58:25 INFO streaming.StreamJob: map 60% reduce 0%
16/06/11 21:58:34 INFO streaming.StreamJob: map 77% reduce 12%
16/06/11 21:58:37 INFO streaming.StreamJob: map 90% reduce 20%
16/06/11 21:58:40 INFO streaming.StreamJob: map 100% reduce 20%
16/06/11 21:58:49 INFO streaming.StreamJob: map 100% reduce 28%
16/06/11 21:58:52 INFO streaming.StreamJob: map 100% reduce 70%
16/06/11 21:58:55 INFO streaming.StreamJob: map 100% reduce 79%
16/06/11 21:58:58 INFO streaming.StreamJob: map 100% reduce 88%
16/06/11 21:59:01 INFO streaming.StreamJob: map 100% reduce 93%
16/06/11 21:59:04 INFO streaming.StreamJob: map 100% reduce 97%
16/06/11 21:59:07 INFO streaming.StreamJob: map 100% reduce 100%
16/06/11 21:59:13 INFO streaming.StreamJob: Job complete: job_201606112134_0004
16/06/11 21:59:13 INFO streaming.StreamJob: Output: /user/ec2-user/Project/first

real    1m13.217s
user    0m1.520s
sys     0m0.080s
projectMasterNode >> $HADOOP_HOME/bin/hadoop fs -ls /user/ec2-user/Project/ProjectData/first
Warning: $HADOOP_HOME is deprecated.

ls: Cannot access /user/ec2-user/Project/ProjectData/first: No such file or directory.
projectMasterNode >> $HADOOP_HOME/bin/hadoop fs -ls /user/ec2-user/Project/first
Warning: $HADOOP_HOME is deprecated.

found 7 items
-rw-r--r--  1 ec2-user supergroup          0 2016-06-11 21:59 /user/ec2-user/Project/first/_SUCCESS
-rwxr-xr-x  - ec2-user supergroup          0 2016-06-11 21:58 /user/ec2-user/Project/first/_logs
-rw-r--r--  1 ec2-user supergroup 63333114 2016-06-11 21:58 /user/ec2-user/Project/first/part-00000
-rw-r--r--  1 ec2-user supergroup 65363599 2016-06-11 21:58 /user/ec2-user/Project/first/part-00001
-rw-r--r--  1 ec2-user supergroup 65835542 2016-06-11 21:58 /user/ec2-user/Project/first/part-00002
-rw-r--r--  1 ec2-user supergroup 66220301 2016-06-11 21:58 /user/ec2-user/Project/first/part-00003
-rw-r--r--  1 ec2-user supergroup 65654136 2016-06-11 21:58 /user/ec2-user/Project/first/part-00004
```

Figure 12 Run of First Map-Reduce Jobs

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```

packageJobJar: [/home/ec2-user/Project/code/matrixapproach/secondMapper.py, /home/ec2-user/Project/code/matrixapproach/secondReducer.py, /tmp/hadoop-ec2-user/hadoop-unjar799223146563246932/] [] /tmp/streamjob8704060619827761833.jar tmp=/null
5/06/11 22:30:42 INFO mapred.FileInputFormat: Total input paths to process : 5
5/06/11 22:30:43 INFO streaming.StreamJob: getLocalDirs(): [/tmp/hadoop-ec2-user/mapred/local]
5/06/11 22:30:43 INFO streaming.StreamJob: Running job: job_201606112134_0006
5/06/11 22:30:43 INFO streaming.StreamJob: To kill this job, run:
5/06/11 22:30:43 INFO streaming.StreamJob: /home/ec2-user/hadoop-0.20.205.0/libexec/../bin/hadoop job -Dmapred.jobacker=172.31.55.98:9001 -kill job_201606112134_0006
5/06/11 22:30:43 INFO streaming.StreamJob: Tracking URL: http://ip-172-31-55-98.ec2.internal:50030/jobdetails.jsp?job_id=job_201606112134_0006
5/06/11 22:30:44 INFO streaming.StreamJob: map 0% reduce 0%
5/06/11 22:32:36 INFO streaming.StreamJob: map 1% reduce 0%
5/06/11 22:36:36 INFO streaming.StreamJob: map 2% reduce 0%
5/06/11 22:38:58 INFO streaming.StreamJob: map 1% reduce 0%
5/06/11 22:39:21 INFO streaming.StreamJob: map 2% reduce 0%
5/06/11 22:39:27 INFO streaming.StreamJob: map 1% reduce 0%
5/06/11 22:41:31 INFO streaming.StreamJob: map 2% reduce 0%
5/06/11 22:42:10 INFO streaming.StreamJob: map 1% reduce 0%
5/06/11 22:44:39 INFO streaming.StreamJob: map 2% reduce 0%
5/06/11 22:44:49 INFO streaming.StreamJob: map 1% reduce 0%
5/06/11 22:50:09 INFO streaming.StreamJob: map 2% reduce 0%
5/06/11 22:52:24 INFO streaming.StreamJob: map 1% reduce 0%
5/06/11 22:54:45 INFO streaming.StreamJob: map 2% reduce 0%
5/06/11 22:55:15 INFO streaming.StreamJob: map 1% reduce 0%

[...]
2]+ Stopped                  $HADOOP_HOME/bin/hadoop jar $HADOOP_HOME/hadoop-streaming-0.20.205.0.jar -D mapred.reducer.tasks=5 -D mapred.map.tasks=10 -input /user/ec2-user/Project/first/part-00000 /user/ec2-user/Project/first/part-0001 /user/ec2-user/Project/first/part-00002 /user/ec2-user/Project/first/part-00003 /user/ec2-user/Project/first/part-0004 -output /user/ec2-user/Project/second/ -mapper ~/Project/code/matrixapproach/secondMapper.py -file ~/Project/code/matrixapproach/secondMapper.py -reducer ~/Project/code/matrixapproach/secondReducer.py -file ~/Project/code/matrixapproach/secondReducer.py

real    26m46.709s
user    0m0.000s
sys    0m0.000s

```

Figure 13 RUN of Second Map-Reduce Job (Failed)

		ERROR
.ED		<pre> java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 1 at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:311) at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:545) at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:132) at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:57) at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:36) at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:436) at org.apache.hadoop.mapred.MapTask.run(MapTask.java:372) at org.apache.hadoop.mapred.Child\$4.run(Child.java:255) at java.security.AccessController.doPrivileged(Native Method) at javax.security.auth.Subject.doAs(Subject.java:415) at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1059) at org.apache.hadoop.mapred.Child.main(Child.java:249) </pre>
.ED		<pre> java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 1 at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:311) at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:545) at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:132) at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:57) at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:36) at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:436) at org.apache.hadoop.mapred.MapTask.run(MapTask.java:372) </pre>

Figure 14 Error Log for Second Map Reduce Job

To resolve the problem in second job, I increased runtime memory of the jvm to 1G, increased the timeout time to 1000 minutes, increased the number of mappers and reducers and finally tried balancing the nodes but I did not get any favorable output

Appendix:

Screenshots: t2.small

Case1: 600 mb Data, 3 Node, 1X Replication

```

16/06/11 10:11:47 INFO mapred.JobClient: Map input records=34210
16/06/11 10:11:47 INFO mapred.JobClient: Physical memory (bytes) snapshot=546123776
16/06/11 10:11:47 INFO mapred.JobClient: Spilled Records=0
16/06/11 10:11:47 INFO mapred.JobClient: CPU time spent (ms)=98860
16/06/11 10:11:47 INFO mapred.JobClient: Total committed heap usage (bytes)=222494720
16/06/11 10:11:47 INFO mapred.JobClient: Virtual memory (bytes) snapshot=51912691712
16/06/11 10:11:47 INFO mapred.JobClient: Map output records=34210
16/06/11 10:11:47 INFO mapred.JobClient: SPLIT_RAW_BYTES=1204
16/06/11 10:11:47 INFO driver.MahoutDriver: Program took 895167 ms (Minutes: 14.91945)

real    14m59.578s
user    0m6.572s
sys     0m0.956s

```

Figure 15 Alternating Least Squares Recommender

```

real    2m21.114s
user    0m7.084s
sys     0m1.272s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

1      [16883:2.897308,21147:2.8971832,26454:2.8968537,6961:2.8958745,22662:2.8930476,20538:2.8930335]
2      [16883:2.9913094,21147:2.980485,514:2.9800832,17084:2.9798803,26454:2.978278,20538:2.9778905]
3      [26454:2.9127655,20538:2.906761,22662:2.9065602,33783:2.9042242,25316:2.9032238,514:2.9025245]
4      [20538:2.9437103,17084:2.942975,514:2.94297,16883:2.9428954,26454:2.9424512,21147:2.9422479]
5      [21454:2.9211137,20538:2.9185665,21147:2.918446,6961:2.915775,22662:2.9152608,16883:2.9134557]
6      [20538:2.9564772,22662:2.9529617,16883:2.952744,21147:2.9488764,514:2.9472876,26454:2.9461873]
7      [20538:3.0265064,21147:3.0257013,21454:3.0248742,22662:3.0231733,26454:3.0201588,514:3.0200136]
8      [16883:2.9669256,26454:2.9601226,20538:2.959728,6961:2.9590101,21147:2.9587507,20385:2.9579432]
9      [21147:2.9455233,16883:2.9397306,20538:2.9393754,17084:2.9390879,6961:2.9378047,514:2.9359753]
10     [6961:2.994741,32829:2.9892826,26454:2.9880445,21147:2.9855964,20538:2.9840584,17084:2.9810495]
cat: Unable to write to output stream.

```

Figure 16 Recommendation

Case2: 1.2 GB Data, 3 Node, 1X Replication

```

real    5m29.547s
user    0m19.248s
sys     0m3.400s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[19484:2.993691,5730:2.9930685,4850:2.9929245,28720:2.9911282,27750:2.9906702,10401:2.9896917]
[27750:2.8815928,19484:2.8814218,28720:2.879799,3099:2.8795702,4850:2.8789656,21752:2.8779328]
[5730:2.884669,28720:2.884388,27750:2.8830223,3099:2.8827472,10401:2.8826914,22938:2.8825188]
[19484:2.7738068,4850:2.7713723,10401:2.7710197,26634:2.770706,3099:2.7701786,27750:2.7701564]
[27750:2.9245875,5730:2.9245372,19484:2.924072,28720:2.92325,3099:2.9227145,4850:2.922385]
[27750:2.8802104,19484:2.87888,5730:2.876961,26634:2.876821,4850:2.876126,22938:2.8750036]
[19484:2.9465466,27750:2.945019,3099:2.944377,28720:2.9442668,5730:2.9441366,22938:2.942061]
[28720:2.8478103,27750:2.846839,19484:2.8465588,5730:2.8465538,22938:2.8455777,3099:2.8448758]
[19484:2.8992648,27750:2.8979685,3099:2.897735,5730:2.8973083,28720:2.896998,26634:2.8955836]
0      [5730:2.874366,27750:2.8740582,28720:2.8732755,19484:2.8729181,4850:2.872111,3099:2.871516]
cat: Unable to write to output stream.

```

Figure 17 Recommendation

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
16/06/11 11:04:17 INFO mapred.JobClient: Map-Reduce Framework
16/06/11 11:04:17 INFO mapred.JobClient: Map input records=34210
16/06/11 11:04:17 INFO mapred.JobClient: Physical memory (bytes) snapshot=1133711360
16/06/11 11:04:17 INFO mapred.JobClient: Spilled Records=0
16/06/11 11:04:17 INFO mapred.JobClient: CPU time spent (ms)=269120
16/06/11 11:04:17 INFO mapred.JobClient: Total committed heap usage (bytes)=444989440
16/06/11 11:04:17 INFO mapred.JobClient: Virtual memory (bytes) snapshot=103825383424
16/06/11 11:04:17 INFO mapred.JobClient: Map output records=34210
16/06/11 11:04:17 INFO mapred.JobClient: SPLIT_RAW_BYTES=2408
16/06/11 11:04:17 INFO driver.MahoutDriver: Program took 2084535 ms (Minutes: 34.74225)

real    34m56.157s
user    0m18.272s
sys     0m2.812s
```

Figure 18 Alternating Least Squares Recommender

Case3: 1.9 GB Data, 3 Node, 1X Replication

```
6/06/11 12:10:24 INFO mapred.JobClient: Reduce input records=0
Exception in thread "main" java.lang.NullPointerException
        at org.apache.mahout.cf.taste.hadoop.als.ParallelALSFactorizationJob.initializeM(ParallelALSFactorizationJob.java:179)
        at org.apache.mahout.cf.taste.hadoop.als.ParallelALSFactorizationJob.run(ParallelALSFactorizationJob.java:156)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:65)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:79)
        at org.apache.mahout.cf.taste.hadoop.als.ParallelALSFactorizationJob.main(ParallelALSFactorizationJob.java:98)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:606)
        at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:68)
        at org.apache.hadoop.util.ProgramDriver.driver(ProgramDriver.java:139)
        at org.apache.mahout.driver.MahoutDriver.main(MahoutDriver.java:188)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:606)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:156)

real    23m21.938s
user    0m25.864s
```

Figure 19 Null Pointer Exception

Case4: 600 mb Data, 3 Node, 3X Replication

```
6/06/11 13:12:52 INFO mapred.JobClient: Total committed heap usage (bytes)=222494720
6/06/11 13:12:52 INFO mapred.JobClient: Virtual memory (bytes) snapshot=51912691712
6/06/11 13:12:52 INFO mapred.JobClient: Map output records=34210
6/06/11 13:12:52 INFO mapred.JobClient: SPLIT_RAW_BYTES=1204
6/06/11 13:12:52 INFO driver.MahoutDriver: Program took 2180878 ms (Minutes: 36.347966666666665)

real    36m29.066s
user    0m18.760s
sys     0m2.592s
```

Figure 20 ALS Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    5m39.922s
user    0m12.064s
sys     0m2.100s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

1 [8959:2.9128096,28793:2.9038506,16865:2.8990219,13886:2.8986986,13309:2.895759,27176:2.894932]
2 [8959:2.962678,3258:2.9576669,28793:2.9552813,16865:2.95209,13309:2.9495647,13886:2.9448717]
3 [8959:2.9862695,28793:2.9848156,3258:2.9829948,13309:2.9747994,16865:2.972711,12487:2.9706922]
4 [8959:2.950134,3258:2.9491835,13309:2.9472687,28793:2.946138,16865:2.9410067,11263:2.9374306]
5 [28793:2.9854927,3258:2.977944,11263:2.974549,12487:2.971285,27176:2.968208,8959:2.9669428]
6 [28793:2.9782546,16865:2.9675858,13309:2.9675014,11263:2.9661684,26996:2.9646356,12487:2.9643772]
7 [28793:2.9835556,3258:2.9826772,8959:2.98212,16865:2.9755056,13309:2.9709723,13886:2.968703]
8 [8959:2.9224024,3258:2.9213715,16865:2.9095545,11263:2.9085462,26996:2.9065568,27176:2.9061632]
9 [8959:2.9586685,3258:2.9582021,28793:2.9574583,13309:2.953512,16865:2.9500628,13886:2.9474206]
10 [8959:2.9483626,3258:2.944406,28793:2.9417074,11263:2.933194,13309:2.9326165,5878:2.929259]
cat: Unable to write to output stream.
ProjectMasterNodeSmall >>
```

Figure 21 Recommendation

Case5: 300 mb Data, 3 Node, 3X Replication

```
6/06/11 13:58:05 INFO driver.MahoutDriver: Program took 866978 ms (Minutes: 14.44963333333333)
real    14m32.014s
user    0m10.064s
sys     0m1.492s
ProjectMasterNodeSmall >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userR
atings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData
als/out/U/ --itemFeatures /user/ec2-user/Project/ProjectData/als/out/M/ --numRecommendations 6 --maxRating 5
HADOOP_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop using HADOOP_HOME=/home/ec2-user/hadoop-0.20.205.0/
Figure 22 ALS Recommender
```

```
real    2m38.560s
user    0m7.124s
sys     0m1.336s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00
00 | head
Warning: $HADOOP_HOME is deprecated.

[25153:3.0486424,6106:3.0419133,13365:3.0395887,14001:3.0376682,18968:3.0291207,13431:3.027464]
[13431:3.0223808,19001:3.0200534,41:3.018988,33942:3.016822,29719:3.0144284,33845:3.0003457]
[4932:3.02964,3993:3.016143,9514:3.0152595,13431:3.0142524,41:3.0118155,19001:3.010537]
[16313:3.2406087,22803:3.2084131,19966:3.196974,25975:3.1815572,18296:3.179536,22795:3.177533]
[19001:3.0704064,21869:3.0250301,1826:3.023227,41:3.0140195,2405:3.0088258,9581:3.0061867]
[33942:3.127293,3993:3.0985851,27976:3.087953,13502:3.0842605,19001:3.0836396,31393:3.0795507]
[19001:3.0024903,33942:2.9684148,31631:2.9666827,1594:2.9642322,13431:2.9600186,18968:2.9587057]
[33845:3.060449,41:3.049786,13431:3.039566,11132:3.0395656,12586:3.0364287,29719:3.035176]
[13431:3.0393002,19001:3.0224128,33942:3.0201874,11652:3.0176044,1594:3.012035,29719:3.0091438]
[19001:3.0277022,33942:3.0851803,6650:3.0532277,9514:3.047676,13431:3.0464566,2476:3.0421709]
```

Figure 23 Recommendation

Case6: 900 mb Data, 3 Node, 3X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
07/06/11 15:54:01 INFO driver.MahoutDriver: Program took 2906802 ms (Minutes: 49.48005555555555)
real    49m35.316s
user    0m16.392s
sys     0m2.332s
ProjectMasterNodeSmall >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRatings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData/als/out/U/ --itemFeatures /user/ec2-user/Project/ProjectData/als/out/M/ --numRecommendations 6 --maxRating 5
```

Figure 24 ALS Recommender

```
real    4m58.646s
user    0m6.196s
sys     0m1.272s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[5359:2.9055836,33447:2.8958948,11958:2.8946712,14663:2.8908823,2641:2.8854945,16548:2.8847795]
[5359:2.9036863,33447:2.8936753,11958:2.8923507,14663:2.8896227,2641:2.883671,16548:2.8814852]
[5359:2.9130616,33447:2.9047337,11958:2.90298,14663:2.8969047,16548:2.89376,2641:2.8934178]
[5359:2.9768841,33447:2.9658077,11958:2.9654,14663:2.963538,2641:2.957871,5466:2.9541042]
[5359:2.9321754,33447:2.9220774,11958:2.919821,14663:2.9148757,16548:2.9091601,5466:2.9085717]
[33447:2.9420075,11958:2.9418828,14663:2.9381742,16548:2.9309106,5466:2.9296029,29098:2.9284322]
[5359:2.8794355,33447:2.8709571,11958:2.8673847,14663:2.8657343,2641:2.8590217,16548:2.8573089]
[5359:2.9825075,33447:2.9709358,11958:2.9708068,14663:2.9669716,2641:2.9623582,16548:2.959186]
[5359:3.0317833,33447:3.0236402,11958:3.0202649,14663:3.0180533,16548:3.0125415,2641:3.0103078]
[5359:2.896107,11958:2.887579,33447:2.88677,14663:2.8821297,16548:2.876347,5466:2.8752573]
```

Figure 25 Recommendation

Case7: 300 mb Data, 3 Node, 1X Replication

```
real    13m40.834s
user    0m10.228s
sys     0m1.644s
ProjectMasterNodeSmall >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/useRatings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData/als/out/U/ --itemFeatures /user/ec2-user/Project/ProjectData/als/out/M/ --numRecommendations 6 --maxRating 5
MAHOUT_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
Running on hadoop, using HADOOP_HOME=/home/ec2-user/hadoop-0.20.205.0/
No HADOOP_CONF_DIR set, using /home/ec2-user/hadoop-0.20.205.0//conf
MAHOUT-JOB: /home/ec2-user/mahout-distribution-0.6/examples/target/mahout-examples-0.6-job.jar
Warning: $HADOOP_HOME is deprecated.
```

Figure 26 ALS Recommender

```
real    2m30.817s
user    0m15.236s
sys     0m2.720s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[4303:3.045505,1946:3.023472,10363:3.023161,30774:3.0189857,28673:3.01548,27759:3.0149286]
[21427:3.006303,8120:3.0060213,33102:3.001415,25603:2.9990737,4303:2.9977365,10363:2.9917588]
[4303:3.0454772,6877:3.038888,1946:3.0366151,21427:3.0364697,32963:3.0216372,27071:3.0206902]
[28673:3.0197244,6877:3.017614,21427:3.0158615,23147:3.0058143,10363:3.0055835,4303:3.000111]
[4303:3.0322347,6877:3.0007715,23869:2.9922497,8120:2.9851227,32963:2.9835052,25603:2.981367]
[17434:3.0454652,25461:3.037686,23080:3.0356646,6877:3.0350213,10363:3.0310957,23497:3.0296543]
[6877:3.0271697,4303:3.0149517,9359:3.0145144,21427:3.0069206,27071:3.0066926,7858:2.995309]
[21427:3.0415096,1946:3.0386496,8120:3.024666,27071:3.0227578,23147:3.0214255,33854:3.0198984]
[8120:3.0440426,11114:3.0360515,1946:3.0305448,21427:3.0224695,4303:3.0201187,23869:3.0200229]
[4303:3.0821369,6877:3.0619671,32871:3.0540674,23869:3.0536768,25603:3.0482194,27666:3.0422187]
stat: Unable to write to output stream.
ProjectMasterNodeSmall >>
```

Figure 27 Recommendation

Case8: 900 mb Data, 3 Node, 1X Replication

```
5/06/11 18:05:21 INFO mapred.JobClient: Virtual memory (bytes) snapshot=74160988160
5/06/11 18:05:21 INFO mapred.JobClient: Map output records=34210
5/06/11 18:05:21 INFO mapred.JobClient: SPLIT_RAW_BYTES=1720
5/06/11 18:05:21 INFO driver.MahoutDriver: Program took 3046783 ms (Minutes: 50.779716666666666)

real    50m56.290s
user    0m14.112s
sys     0m2.172s
```

Figure 28 ALS Recommender

```
real    8m14.414s
user    0m12.588s
sys     0m2.208s
projectMasterNodeSmall >>
projectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[33622:2.9398515,21787:2.9267523,18034:2.9224033,8868:2.9184935,29309:2.910104,4428:2.9096873]
[33622:2.9752822,21787:2.9618375,18034:2.9572394,4428:2.9460225,29309:2.9442523,33703:2.9427679]
[33622:2.9808607,21787:2.970032,18034:2.9665668,8868:2.9626641,29309:2.9532027,22933:2.95068]
[33622:2.9212925,21787:2.9124055,18034:2.9015489,8868:2.898299,4428:2.8966687,29309:2.894422]
[33622:2.9470675,21787:2.9339926,18034:2.9304738,8868:2.9250114,4428:2.918724,29309:2.915675]
[33622:2.97522,21787:2.9602075,18034:2.957342,8868:2.952917,29309:2.944868,6588:2.9412522]
[33622:2.9015396,21787:2.889623,18034:2.883699,8868:2.8784723,4428:2.877043,29309:2.8722618]
[33622:2.896917,21787:2.8840547,18034:2.876223,8868:2.8756583,33703:2.874721,4428:2.874266]
[33622:2.9551194,21787:2.943971,18034:2.9431803,4428:2.9321096,8868:2.9278903,29309:2.927711]
0 [33622:2.9399028,21787:2.9300427,18034:2.9247856,4428:2.9158697,29309:2.9126716,2546:2.909983]
at: Unable to write to output stream.
```

Figure 29 Recommendation

Case9: 300 mb Data, 2 Node, 1X Replication

```
real    36m39.179s
user    0m16.880s
sys     0m2.444s
projectMasterNodeSmall >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userR
Figure 30 ALS Recommender
```

```
real    4m25.678s
user    0m15.604s
sys     0m2.888s
projectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[4303:3.0197592,8550:3.004526,30774:2.996917,11322:2.9953387,21044:2.9927702,28308:2.9872837]
[28962:3.0403361,27071:3.0290236,32142:3.0173337,444:3.0172634,1604:3.0158603,23428:3.0114756]
[24682:3.0408397,27071:3.0274,24596:3.022912,28962:3.0219862,444:3.020542,1604:3.013453]
[30912:3.0279944,8834:3.0081148,27071:3.0068493,31658:3.0016487,30394:2.997908,21947:2.9970562]
[27071:2.987148,20818:2.9788733,30774:2.9761293,28962:2.975034,8120:2.9722574,25603:2.9680355]
[27071:3.0282488,21044:3.0270545,8120:3.02353,8550:3.020046,8834:3.0183108,11322:3.0177577]
[30584:3.036672,2387:3.0358164,32142:3.0290542,952:3.0247118,21147:3.0230513,13482:3.0229435]
[21044:3.0354989,8120:3.0255554,27071:3.0231068,30912:3.0181062,28962:3.018071,11114:3.0164065]
[30774:3.0672712,25756:3.0635617,8120:3.0464585,31007:3.044232,1946:3.0414956,20818:3.0412414]
0 [27071:3.0152848,5238:3.0131938,30912:3.0113206,29738:3.0093505,20818:3.0089598,30774:3.0066102]
```

Figure 31 Recommendation

Case10: 600 mb Data, 2 Node, 1X Replication

```
real    65m59.738s
user    0m21.588s
sys     0m3.116s
```

Figure 32 ALS Recommender

```
real    4m12.994s
user    0m5.056s
sys     0m0.872s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[15727:2.9691067,23255:2.9628859,34024:2.9580967,32611:2.9554012,15440:2.9553344,5817:2.9537864]
[15440:2.9739912,23255:2.9724324,34024:2.968568,32611:2.9673615,243:2.9670093,5817:2.9665616]
[15727:2.952068,15440:2.946146,34024:2.9422216,32611:2.9344883,13891:2.9335682,11229:2.9332235]
[15727:3.0006533,34024:2.9945154,243:2.9939337,11229:2.9938655,23255:2.9937758,3674:2.9910436]
[23255:2.9446619,34024:2.9446173,15727:2.9438314,32611:2.9406796,5817:2.938296,15440:2.9376774]
[34024:2.9685128,5817:2.9676826,15727:2.9663925,23255:2.9645088,32611:2.9614203,5488:2.958165]
[15727:2.9529214,23255:2.9522836,34024:2.9476292,5817:2.9457994,32611:2.9446173,13891:2.9430525]
[15727:3.0063949,23255:3.0026119,3369:3.0024362,34024:3.0022943,6615:3.00135,15440:2.994694]
[15727:2.930841,15440:2.9180293,13891:2.9171448,34024:2.916575,32611:2.9153168,23255:2.913783]
[15727:2.9851675,34024:2.9832444,15440:2.980455,6615:2.9700172,5817:2.9689965,11229:2.9689314]
```

Figure 33 Recommendation

Case10: 300 mb Data, 1 Node, 1X Replication

```
real    54m30.367s
user    0m17.960s
sys     0m2.596s
```

Figure 34 ALS Recommender

```
real    11m17.607s
user    0m17.484s
sys     0m3.056s
ProjectMasterNodeSmall >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[9981:3.0196548,22124:3.0083253,13029:3.0070317,3080:3.0051422,2737:3.0039415,27666:3.0012689]
[32687:3.0265381,32414:3.0073307,3080:4:3.002429,31678:3.000833,32871:2.998817,19993:2.9984615]
[9981:3.0501208,10520:3.0348558,32687:3.030752,31678:3.0302668,4303:3.0243077,27666:3.0221612]
[23869:3.0049353,12812:2.9952064,20384:2.9923902,20784:2.9901342,11462:2.9880152,4303:2.9877763]
[23869:2.975079,8284:2.968601,6877:2.9642527,23147:2.9621513,2465:2.961777,33975:2.9572241]
[6877:3.0394254,25068:3.0284674,22353:3.0190148,9981:3.0185475,31768:3.0113206,9559:3.011775]
[25068:3.0568383,17418:3.034577,4303:3.0322416,32687:3.0280504,24596:3.0267942,4166:3.0250711]
[6877:3.0373762,4303:3.030922,8284:3.0301843,1946:3.0245488,24593:3.0235677,23869:3.0214555]
[25068:3.0602174,32687:3.042602,23869:3.0422573,2850:3.0359645,16413:3.0344868,27666:3.0322797]
[25068:3.0517669,25048:3.045314,32687:3.0445893,9981:3.0410686,27666:3.0364437,23869:3.029957]
cat: Unable to write to output stream.
```

Figure 35 Recommendation

Screenshots: t2.medium

Case1: 600 mb Data, 3 Node, 1X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
6/06/11 10:02:02 INFO mapred.JobClient: File Input Format Counters
6/06/11 10:02:02 INFO mapred.JobClient: Bytes Read=437780089
6/06/11 10:02:02 INFO mapred.JobClient: Map-Reduce Framework
6/06/11 10:02:02 INFO mapred.JobClient: Map input records=34210
6/06/11 10:02:02 INFO mapred.JobClient: Physical memory (bytes) snapshot=1609502720
6/06/11 10:02:02 INFO mapred.JobClient: Spilled Records=0
6/06/11 10:02:02 INFO mapred.JobClient: CPU time spent (ms)=93890
6/06/11 10:02:02 INFO mapred.JobClient: Total committed heap usage (bytes)=705691648
6/06/11 10:02:02 INFO mapred.JobClient: Virtual memory (bytes) snapshot=17872289792
6/06/11 10:02:02 INFO mapred.JobClient: Map output records=34210
6/06/11 10:02:02 INFO mapred.JobClient: SPLIT_RAW_BYT
6/06/11 10:02:02 INFO driver.MahoutDriver: Program took 617370 ms (Minutes: 10.2895)

real    1m19.440s
user    0m6.092s
sys     0m0.968s
```

Figure 36 Alternating Least Squares Recommender

```
real    1m6.180s
user    0m3.444s
sys     0m0.704s
projectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

[13685:2.9519725,31352:2.951018,21990:2.950889,18770:2.949171,32343:2.9485154,25675:2.9473584]
[26655:2.946212,140:2.942296,25675:2.9397268,18770:2.936872,14219:2.9332821,2385:2.9312134]
[140:2.9699636,25675:2.9691117,2209:2.9648623,13685:2.96473,21990:2.9620004,14219:2.9613943]
[140:2.9163442,25675:2.9151778,13685:2.914088,3520:2.9120357,26655:2.911191,14219:2.910008]
[140:2.9684412,21990:2.9670444,14219:2.9656606,26655:2.9634833,32343:2.9632642,18770:2.960008]
[25675:3.1293411,2209:3.1223698,14219:3.119343,12279:3.1153424,9729:3.1117952,28526:3.1024292]
[26655:2.922896,25675:2.920751,21990:2.9195514,140:2.919325,14219:2.9146743,18770:2.9139981]
[140:2.9298515,26655:2.9280474,25675:2.9279773,2385:2.9229684,14219:2.922623,32343:2.9202266]
[140:2.9521492,21990:2.9504206,26655:2.947815,14219:2.9476962,25675:2.946768,13685:2.9461868]
0 [21990:3.009593,140:3.0095372,26655:3.0063918,25675:3.0061572,13685:3.003489,14219:3.0001526]
at: Unable to write to output stream.
```

Figure 37 Recommendation

Case2: 1.2 GB Data, 3 Node, 1X Replication

```
16/06/11 10:37:25 INFO mapred.JobClient: Physical memory (bytes) snapshot=2963816448
16/06/11 10:37:25 INFO mapred.JobClient: Spilled Records=0
16/06/11 10:37:25 INFO mapred.JobClient: CPU time spent (ms)=263020
16/06/11 10:37:25 INFO mapred.JobClient: Total committed heap usage (bytes)=1314914304
16/06/11 10:37:25 INFO mapred.JobClient: Virtual memory (bytes) snapshot=35744579584
16/06/11 10:37:25 INFO mapred.JobClient: Map output records=34210
16/06/11 10:37:25 INFO mapred.JobClient: SPLIT_RAW_BYT
16/06/11 10:37:25 INFO driver.MahoutDriver: Program took 990510 ms (Minutes: 16.5085)

real    16m32.564s
user    0m6.460s
sys     0m1.064s
```

Figure 38 Alternating Least Squares Recommender

```

real    1m45.205s
user    0m3.396s
sys     0m0.752s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

[28869:2.9245055,25375:2.9214587,32231:2.9143293,11913:2.9049034,6150:2.90477,8239:2.9008245]
[28869:2.972352,25375:2.9630024,32231:2.9597428,11913:2.949742,20993:2.9455957,6150:2.945586]
[28869:2.9128084,25375:2.90615,32231:2.9027367,11913:2.8931863,6150:2.8880608,8239:2.8876567]
[28869:2.929347,25375:2.9187276,32231:2.9163024,11913:2.9054468,6150:2.9010441,8239:2.900645]
[32231:2.9040651,11913:2.8936753,6150:2.8902447,20993:2.8901043,19805:2.8875566,8239:2.8875127]
[25375:2.915309,32231:2.912495,11913:2.9011943,6150:2.8973794,20993:2.8964496,19805:2.8956516]
[28869:2.8893008,25375:2.8843687,32231:2.87821,11913:2.8722982,6150:2.8630543,5943:2.8629322]
[28869:2.944943,25375:2.9388783,32231:2.9356654,11913:2.9247649,6150:2.920838,8239:2.919527]
[28869:2.9317546,32231:2.921363,11913:2.9108434,6150:2.9064863,20993:2.9060125,8239:2.9049096]
[28869:2.8997684,32231:2.8909693,25375:2.8892233,11913:2.879231,20993:2.8744237,5943:2.8734689]

cat: Unable to write to output stream.

```

Figure 39 Recommendation

Case3: 1.9 GB Data, 3 Node, 1X Replication

```

16/06/11 11:16:17 INFO mapred.JobClient: Map-Reduce Framework
16/06/11 11:16:17 INFO mapred.JobClient: Map input records=34210
16/06/11 11:16:17 INFO mapred.JobClient: Physical memory (bytes) snapshot=5485555712
16/06/11 11:16:17 INFO mapred.JobClient: Spilled Records=0
16/06/11 11:16:17 INFO mapred.JobClient: CPU time spent (ms)=399270
16/06/11 11:16:17 INFO mapred.JobClient: Total committed heap usage (bytes)=2580021248
16/06/11 11:16:17 INFO mapred.JobClient: Virtual memory (bytes) snapshot=53616869376
16/06/11 11:16:17 INFO mapred.JobClient: Map output records=34210
16/06/11 11:16:17 INFO mapred.JobClient: SPLIT_RAW_BYTES=3612
16/06/11 11:16:17 INFO driver.MahoutDriver: Program took 1393811 ms (Minutes: 23.230183333333333)

real    23m16.138s
user    0m6.492s
sys     0m1.052s

```

Figure 40 Alternating Least Squares Recommender

```

real    2m36.413s
user    0m3.688s
sys     0m0.592s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

[1088:2.8875303,25619:2.8789296,46:2.8778205,15816:2.8771296,6425:2.875107,7856:2.8744397]
[1088:2.9329202,25619:2.9229774,46:2.9222143,15816:2.9214492,7856:2.9206517,6425:2.9197412]
[1088:2.8913658,25619:2.882606,46:2.8791943,7856:2.8785434,9892:2.8768203,6425:2.8765006]
[1088:2.9022691,25619:2.892643,15816:2.8908942,46:2.890495,7856:2.8896606,6425:2.88921]
[1088:2.8671238,25619:2.857896,15816:2.8553724,46:2.8547475,7856:2.8539538,6425:2.852953]
[1088:2.9031553,25619:2.8945093,15816:2.8931847,46:2.8931458,6425:2.8903873,7856:2.8902617]
[1088:2.7020304,25619:2.6915758,46:2.6900952,15816:2.6894178,9892:2.6869695,6425:2.6864889]
[1088:2.8853555,25619:2.8757946,15816:2.8746395,46:2.874345,7856:2.872661,6425:2.8720484]
[1088:2.9150598,25619:2.905561,15816:2.9042873,46:2.902923,7856:2.902252,9892:2.9012582]
[1088:2.8948913,25619:2.8852801,46:2.8836536,15816:2.8836348,7856:2.8819706,6425:2.8813388]

cat: Unable to write to output stream.

```

Figure 41 Recommendation

Case4: 600 mb Data, 3 Node, 3X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```

real    10m39.195s
user    0m6.596s
sys     0m1.136s
projectMasterNodeMedium >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/ratings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/Project
Figure 42 ALS Recommender

```

```

real    1m6.180s
user    0m3.444s
sys     0m0.704s
projectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[13685:2.9519725,31352:2.951018,21990:2.950889,18770:2.949171,32343:2.9485154,25675:2.9473584]
[26655:2.946212,140:2.942296,25675:2.9397268,18770:2.936872,14219:2.9332821,2385:2.9312134]
[140:2.9699636,25675:2.9691117,2209:2.9648623,13685:2.96473,21990:2.9620004,14219:2.9613943]
[140:2.9163442,25675:2.9151778,13685:2.914088,3520:2.9120357,26655:2.911191,14219:2.910008]
[140:2.9684412,21990:2.9670444,14219:2.9656606,26655:2.9634833,32343:2.9632642,18770:2.960008]
[25675:3.1293411,2209:3.1223698,14219:3.119343,12279:3.1153424,9729:3.1117952,28526:3.1024292]
[26655:2.922896,25675:2.920751,21990:2.9195514,140:2.919325,14219:2.9146743,18770:2.9139981]
[140:2.9298515,26655:2.9280474,25675:2.9279773,2385:2.9229684,14219:2.922623,32343:2.9202266]
[140:2.9521492,21990:2.9504206,26655:2.947815,14219:2.9476962,25675:2.946768,13685:2.94618681]

```

Figure 43 Recommendation

Case5: 300 mb Data, 3 Node, 3X Replication

```

06/06/11 13:23:33 INFO mapred.JobClient: Map input records=34210
06/06/11 13:23:33 INFO mapred.JobClient: Physical memory (bytes) snapshot=962613248
06/06/11 13:23:33 INFO mapred.JobClient: Spilled Records=0
06/06/11 13:23:33 INFO mapred.JobClient: CPU time spent (ms)=54890
06/06/11 13:23:33 INFO mapred.JobClient: Total committed heap usage (bytes)=492306432
06/06/11 13:23:33 INFO mapred.JobClient: Virtual memory (bytes) snapshot=10212737024
06/06/11 13:23:33 INFO mapred.JobClient: Map output records=34210
06/06/11 13:23:33 INFO mapred.JobClient: SPLIT_RAW_BYTES=688
06/06/11 13:23:33 INFO driver.MahoutDriver: Program took 455306 ms (Minutes: 7.588433333333334)

real    7m37.782s
user    0m6.544s
sys     0m1.260s

```

Figure 44 ALS Recommender

```

real    0m59.875s
user    0m3.844s
sys     0m0.848s
projectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[28941:3.097737,16460:3.0397975,6913:3.0199575,22291:3.01497,27966:3.0148118,11291:3.0129766]
[28941:3.0405874,16460:3.0372477,14370:3.0135896,6080:3.0104182,26421:3.0037374,31034:3.001425]
[5176:3.021097,10488:3.0059009,1588:2.9919221,21481:2.9883566,23711:2.987244,13591:2.9863298]
[34025:3.025053,6255:3.011704,2942:2.999338,15447:2.9773455,17862:2.9770465,30097:2.9768398]
[6080:3.0371332,28941:3.0352643,16460:3.0266864,27966:3.0237277,20941:3.0233924,10488:3.0162373]
[16460:3.1834557,27940:3.1638472,32698:3.1636608,33817:3.1605778,1820:3.1490312,4717:3.1483176]
[28941:3.0130234,11613:2.9852765,10488:2.9831457,19680:2.9814293,33817:2.9793954,20941:2.9731297]
[28941:2.9923642,16460:2.9840858,6080:2.9763038,21428:2.9667602,830:2.965261,33817:2.9596598]
[10488:3.0410151,28941:3.0340958,20941:3.0283334,910:3.0266435,6080:3.0231524,32322:3.0201414]
0 [27966:2.9845831,16460:2.978638,6080:2.976997,10488:2.9745815,28941:2.9683204,20941:2.9625418]

```

Figure 45 Recommendation

Case6: 900 mb Data, 3 Node, 3X Replication

```

6/06/11 14:06:33 INFO mapred.JobClient: Map-Reduce Framework
6/06/11 14:06:33 INFO mapred.JobClient: Map input records=34210
6/06/11 14:06:33 INFO mapred.JobClient: Physical memory (bytes) snapshot=2091905024
6/06/11 14:06:33 INFO mapred.JobClient: Spilled Records=0
6/06/11 14:06:33 INFO mapred.JobClient: CPU time spent (ms)=165840
6/06/11 14:06:33 INFO mapred.JobClient: Total committed heap usage (bytes)=878706688
6/06/11 14:06:33 INFO mapred.JobClient: Virtual memory (bytes) snapshot=25531842560
6/06/11 14:06:33 INFO mapred.JobClient: Map output records=34210
6/06/11 14:06:33 INFO mapred.JobClient: SPLIT_RAW_BYT
6/06/11 14:06:33 INFO driver.MahoutDriver: Program took 823279 ms (Minutes: 13.721316666666667)

real    13m45.782s
user    0m6.880s
sys     0m1.336s

```

Figure 46 ALS Recommender

```

real    1m46.778s
user    0m4.036s
sys     0m0.852s
projectMasterNodeMedium >>
projectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[3620:2.9502382,17328:2.9474404,26089:2.9454317,24647:2.94451,20582:2.937287,2948:2.9370563]
[10777:2.8898423,3620:2.8802676,17328:2.8788908,26089:2.8780959,24647:2.878076,10754:2.8697448]
[10777:2.9360657,3620:2.9278085,17328:2.9249218,26089:2.922442,24647:2.9220145,15470:2.9140384]
[3620:2.9611058,17328:2.9582949,26089:2.956174,24647:2.9550872,10754:2.9490335,2948:2.9477944]
[10777:2.9792273,3620:2.9706006,17328:2.9667847,26089:2.9641187,24647:2.961332,10754:2.9562597]
[3620:2.9456484,24647:2.942956,17328:2.9417107,26089:2.9413645,10754:2.936124,2948:2.933772]
[10777:2.9223173,3620:2.910863,17328:2.9090981,26089:2.908168,24647:2.9079127,10754:2.9014292]
[10777:2.9416084,3620:2.929533,17328:2.929226,26089:2.9271607,10754:2.9199119,15470:2.9172554]
[10777:2.9782853,17328:2.9668446,3620:2.9653285,26089:2.964107,24647:2.962992,15470:2.957457]
0      [10777:2.9265633,3620:2.9154158,24647:2.9125433,17328:2.912302,10754:2.9061863,23538:2.9048178]
at: Unable to write to output stream.

```

Figure 47 Recommendation

Case7: 900 mb Data, 3 Node, 1X Replication

```

6/06/11 15:07:31 INFO mapred.JobClient: Map-Reduce Framework
6/06/11 15:07:31 INFO mapred.JobClient: Map input records=34210
6/06/11 15:07:31 INFO mapred.JobClient: Physical memory (bytes) snapshot=1894215680
6/06/11 15:07:31 INFO mapred.JobClient: Spilled Records=0
6/06/11 15:07:31 INFO mapred.JobClient: CPU time spent (ms)=164150
6/06/11 15:07:31 INFO mapred.JobClient: Total committed heap usage (bytes)=837287936
6/06/11 15:07:31 INFO mapred.JobClient: Virtual memory (bytes) snapshot=25531842560
6/06/11 15:07:31 INFO mapred.JobClient: Map output records=34210
6/06/11 15:07:31 INFO mapred.JobClient: SPLIT_RAW_BYT
6/06/11 15:07:31 INFO driver.MahoutDriver: Program took 816588 ms (Minutes: 13.6098)

real    13m39.118s
user    0m7.004s
sys     0m1.164s

```

Figure 48 ALS Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    1m48.712s
user    0m3.976s
sys     0m0.844s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

[3620:2.9356723,12734:2.93322,10754:2.9278142,3858:2.925178,9833:2.919795,21836:2.9195561]
[3620:2.9090133,12734:2.906261,10777:2.9009929,10754:2.8996434,3858:2.898777,9833:2.8928306]
[12734:2.8893948,3620:2.8892543,10777:2.8854704,10754:2.880058,9833:2.8765855,3858:2.8762882]
[3620:2.98864,12734:2.9856942,3858:2.9788723,10754:2.9779258,21836:2.9735982,9833:2.9731123]
[3620:3.0140498,12734:3.007568,3858:3.0058875,10777:3.0048432,10754:3.0045826,9833:3.0022912]
[12734:2.9374115,10754:2.9321084,3858:2.9302244,9833:2.9256325,22692:2.9243643,21836:2.9242342]
[3620:2.9166386,12734:2.9122658,10754:2.9102728,3858:2.9062152,10777:2.9062092,21836:2.9025037]
[3620:2.9371746,12734:2.9330475,10754:2.929635,10777:2.928953,3858:2.9282131,9833:2.9237032]
[3620:2.968053,12734:2.9636877,10754:2.9597013,10777:2.9587007,3858:2.956161,9833:2.952882]
[3620:2.9108472,10754:2.905611,10777:2.9035687,3858:2.8998806,9833:2.896457,22692:2.8960395]
cat: Unable to write to output stream.
```

Figure 49 Recommendation

Case8: 300 mb Data, 3 Node, 1X Replication

```
6/06/11 15:30:50 INFO mapred.JobClient: Total committed heap usage (bytes)=516423680
6/06/11 15:30:50 INFO mapred.JobClient: Virtual memory (bytes) snapshot=10212737024
6/06/11 15:30:50 INFO mapred.JobClient: Map output records=34210
6/06/11 15:30:50 INFO mapred.JobClient: SPLIT_RAW_BYTES=688
6/06/11 15:30:50 INFO driver.MahoutDriver: Program took 459760 ms (Minutes: 7.662666666666666)
real    7m42.250s
user    0m6.608s
sys     0m1.108s
ProjectMasterNodeMedium >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userSettings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData/als/out/U/ --itemFeatures /user/ec2-user/Project/ProjectData/als/out/M/ --numRecommendations 6 --maxRating 5
[...]
```

Figure 50 ALS Recommendation

```
real    0m58.795s
user    0m3.952s
sys     0m0.848s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

[5563:3.0339193,13211:3.0269618,5193:3.0060706,8348:3.0052793,11006:2.9998255,4810:2.9992766]
[18816:3.0518644,33399:3.042015,28517:3.041751,591:3.0391269,4308:3.034952,23632:3.0312862]
[591:3.021284,23527:3.0124018,18816:3.0097582,28517:3.00872,4308:3.0053048,33399:3.005258]
[7200:3.119256,23527:3.0914137,31545:3.0809083,8354:3.0801694,13820:3.0717103,14199:3.063334]
[2658:3.0367076,4308:3.0277996,18816:3.0192013,23527:3.0155044,28517:3.00211,3503:3.002061]
[23632:2.9938443,25541:2.9660423,15773:2.9586103,1764:2.9576185,16062:2.9571111,20683:2.9558]
[591:3.0606685,23527:3.0346384,28517:3.0274198,14199:3.0264711,31890:3.0225422,4308:3.0223527]
[591:3.0299306,4308:3.0227194,8348:3.0225337,13211:3.021287,27853:3.0212066,2658:3.015545]
[28517:3.050474,19985:3.0476155,23632:3.0468013,4308:3.0419362,23527:3.0361288,22643:3.0359135]
[8434:3.0917625,591:3.0861943,18816:3.0852482,33399:3.081373,2658:3.0813243,3503:3.0801635]
```

Figure 51 Recommendation

Case9: 300 mb Data, 2 Node, 1X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
6/06/11 16:31:21 INFO driver.MahoutDriver: Program took 475448 ms (Minutes: 7.92413333333334)
real    7m57.976s
user    0m6.488s
sys     0m1.076s
ProjectMasterNodeMedium >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/ratings/ --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/Project/als/out/U/ --itemFeatures /user/ec2-user/Project/ProjectData/als/out/M/ --numRecommendations 6 --maxRating 5
$HADOOP_LOCAL is not set; adding HADOOP_CONF_DIR to classpath.
```

Figure 52 ALS Recommender

```
real    1m1.764s
user    0m3.896s
sys     0m0.832s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-r-0000 | head
Warning: $HADOOP_HOME is deprecated.

[21064:3.0525608,23527:3.0464213,18984:3.045767,17685:3.0398831,3503:3.022577,13888:3.0211797]
[4308:3.0244799,10160:3.0054464,13820:3.00423,23249:3.0028458,23527:2.9998128,23477:2.998818]
[4308:3.0102313,23527:3.0099509,2658:2.9894724,15325:2.9887836,23249:2.988142,8348:2.9801943]
[31628:3.2258897,23249:3.2221704,29983:3.207863,20626:3.2076588,20346:3.19945,28946:3.1986947]
[13820:3.0032306,14199:2.99492,4308:2.9847207,8348:2.9775858,17329:2.977415,23477:2.9723275]
[31890:3.0016537,27853:3.0015473,28200:2.9896805,17241:2.98383,2658:2.9837604,8348:2.9831948]
[23527:3.1012201,15325:3.060126,4308:3.0588152,2658:3.0492058,23418:3.0424707,14199:3.0391371]
[4308:3.0311422,2658:3.0155714,31890:3.0039206,23418:2.9978824,29779:2.9966452,21453:2.9952776]
[27853:3.0529864,13820:3.0376866,21064:3.0350673,28932:3.0333874,21453:3.032107,25092:3.0288777]
[27853:3.1123638,15325:3.089182,23477:3.0810394,6632:3.0800877,14199:3.0746405,10623:3.0701323]
cat: Unable to write to output stream.
```

Figure 53 Recommendation

Case10: 600 mb Data, 2 Node, 1X Replication

```
6/06/11 17:30:55 INFO mapred.JobClient:      Total committed heap usage (bytes)=691535872
6/06/11 17:30:55 INFO mapred.JobClient:      Virtual memory (bytes) snapshot=17872289792
6/06/11 17:30:55 INFO mapred.JobClient:      Map output records=34210
6/06/11 17:30:55 INFO mapred.JobClient:      SPLIT_RAW_BYTES=1204
6/06/11 17:30:55 INFO driver.MahoutDriver: Program took 741091 ms (Minutes: 12.351516666666667)

real    12m24.196s
user    0m6.812s
sys     0m1.232s
```

Figure 54 ALS Recommender

```
real    1m45.031s
user    0m3.868s
sys     0m0.936s
ProjectMasterNodeMedium >> large_step1_2node_900
bash: large_step1_2node_900: command not found
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-r-0000 | head
Warning: $HADOOP_HOME is deprecated.

[2992:3.005106,31182:3.0033214,20376:3.0018187,28740:2.999167,21795:2.9978278,735:2.9976277]
[31182:2.9535365,28740:2.947707,19652:2.9454596,8838:2.9451284,26993:2.9445066,7517:2.943694]
[32745:2.9358165,26993:2.93131,31182:2.928438,3904:2.9260125,28740:2.9246504,7137:2.922266]
[31182:2.9403214,32745:2.939869,19652:2.9338803,8838:2.9338343,14354:2.930259,251:2.929305]
[32745:2.9441428,27646:2.9315312,19652:2.9306586,28740:2.929154,26993:2.928332,8838:2.9277434]
[32745:2.9192626,31182:2.9152694,26993:2.9097784,19652:2.9092886,3904:2.908143,251:2.9068646]
[32745:2.9251342,19652:2.9132097,31182:2.9123428,3904:2.9115694,28740:2.910927,27646:2.9098897]
[32745:2.9406798,31182:2.9317834,26993:2.9306083,28740:2.92925,3904:2.9273052,18035:2.9272041]
[32745:2.9361954,31182:2.9263391,251:2.9259028,19652:2.9258623,14354:2.9235566,3904:2.920995]
[32745:2.939642,31182:2.927711,19652:2.9244397,26993:2.9242525,7223:2.923651,21149:2.922207]
```

Figure 55 Recommendation

Case11: 900 mb Data, 2 Node, 1X Replication

```

16/06/11 16:57:08 INFO mapred.JobClient: Total committed heap usage (bytes)=865599488
16/06/11 16:57:08 INFO mapred.JobClient: Virtual memory (bytes) snapshot=25531842560
16/06/11 16:57:08 INFO mapred.JobClient: Map output records=34210
16/06/11 16:57:08 INFO mapred.JobClient: SPLIT_RAW_BYTES=1720
16/06/11 16:57:08 INFO driver.MahoutDriver: Program took 971514 ms (Minutes: 16.1919)

real    16m14.009s
user    0m6.980s
sys     0m1.224s

```

Figure 56 Recommender

```

real    2m14.807s
user    0m3.860s
sys     0m0.960s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

[10754:2.9441307,24647:2.9422848,26089:2.940265,22907:2.9395785,27762:2.938199,10667:2.936851]
[10777:2.9210992,24647:2.9032724,10754:2.9030464,26089:2.9014533,22907:2.9002311,27762:2.8977923]
[10777:2.9438782,10754:2.9232829,24647:2.9205787,26089:2.920262,27762:2.9197907,22907:2.919612]
[10754:2.9644828,24647:2.961178,26089:2.9603484,27762:2.960206,22907:2.959268,22661:2.9578974]
[10777:2.997529,24647:2.9820302,10754:2.9815214,26089:2.9814138,22907:2.9775724,27762:2.9742608]
[10754:2.9466739,24647:2.9439757,26089:2.9439526,22907:2.9427998,27762:2.9424803,22661:2.938864]
[10777:2.9257803,10754:2.9106615,24647:2.906961,22907:2.9053783,26089:2.905271,27762:2.9044838]
[10777:2.9485567,10754:2.9298525,26089:2.9290705,27762:2.926058,22661:2.9233518,10667:2.922974]
[10777:2.9675293,10754:2.9473493,24647:2.9460266,26089:2.9454734,22907:2.9429417,27762:2.9421327]
0   [10777:2.925595,10754:2.9102805,24647:2.9081368,22907:2.907881,27762:2.904598,23492:2.9027538]

```

Figure 57 Recommendation

Case12: 300 mb Data, 2 Node, 3X Replication

```

real    14m32.543s
user    0m12.496s
sys     0m2.840s
ProjectMasterNodeMedium >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/voc
Figure 58 ALS Recommender

```

```

real    2m42.035s
user    0m12.060s
sys     0m3.452s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-000 | head
Warning: $HADOOP_HOME is deprecated.

1   [25851:3.0072532,12210:2.9898555,26806:2.9873765,2008:2.9862506,2056:2.9830074,13263:2.9789443]
2   [30082:3.0008214,25851:2.978479,1395:2.9738533,21059:2.9736366,12151:2.9706278,23158:2.9698963]
3   [19516:3.1233063,5545:3.1175966,21988:3.1040258,18116:3.1015894,33120:3.0882285,3171:3.0781415]
4   [12329:3.049338,31729:3.0492244,26423:3.0357873,13263:3.0096798,13424:3.0085413,26806:3.0071404]
5   [30285:3.029399,12151:3.0206609,12329:3.0174012,26806:3.0167322,13263:3.013543,31729:3.0123992]
6   [2056:3.0584753,18144:3.0481186,13424:3.043015,6876:3.0409951,11938:3.0314054,9238:3.0291631]
7   [31729:3.0120463,18144:3.0116003,33120:3.0104778,14547:3.0032032,1395:3.0025995,12151:2.9983597]
8   [26105:3.0141718,23617:3.0013433,13263:3.000806,25851:3.0005336,2056:3.0002005,9238:2.9994695]
9   [30082:3.0270061,19516:3.008079,21843:2.9911678,20020:2.9903016,22588:2.9853394,32127:2.977929]
10  [25851:3.029579,8023:3.0187197,9209:3.0049133,19042:3.0039122,30263:2.9985583,2810:2.9969382]
cat: Unable to write to output stream.

```

Figure 59 Recommendation

Case13: 600 mb Data, 2 Node, 3X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```

6/06/11 18:03:00 INFO mapred.JobClient:      CPU time spent (ms)=99950
6/06/11 18:03:00 INFO mapred.JobClient:      Total committed heap usage (bytes)=709361664
6/06/11 18:03:00 INFO mapred.JobClient:      Virtual memory (bytes) snapshot=17872289792
6/06/11 18:03:00 INFO mapred.JobClient:      Map output records=34210
6/06/11 18:03:00 INFO mapred.JobClient:      SPLIT_RAW_BYTES=1204
6/06/11 18:03:00 INFO driver.MahoutDriver: Program took 743987 ms (Minutes: 12.399783333333334)

real    12m26.439s
user    0m6.712s
sys     0m1.296s
projectMasterNodeMedium >> time_bin/mahout_recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/use
Figure 60 ALS Recommender

```

```

real    1m43.917s
user    0m3.972s
sys     0m0.832s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 | head
Warning: $HADOOP_HOME is deprecated.

[31182:3.0425375,11384:3.0392723,5197:3.0389934,32903:3.0345948,20376:3.0338073,19652:3.0326836]
[31182:2.9578059,5197:2.9499137,6119:2.9475803,33381:2.946273,32903:2.9441612,20376:2.943494]
[31182:2.9406374,32745:2.9375603,5197:2.9302776,18431:2.929906,6119:2.9272451,32903:2.9260437]
[31182:2.9361475,32745:2.9304564,5197:2.9293246,32903:2.9225419,18431:2.9192994,20376:2.9190445]
[5197:2.938222,32745:2.9370863,32903:2.93566,6119:2.9352932,1841:2.9340372,32740:2.9323862]
[31182:2.9257915,32745:2.9216006,5197:2.9171627,14354:2.914136,6119:2.913359,20376:2.9115965]
[31182:2.9345446,32745:2.9291873,14354:2.9218352,18431:2.920923,32903:2.9186237,31314:2.9182296]
[31182:2.9415927,5197:2.936004,32745:2.9349115,6119:2.9344187,32903:2.9287782,32740:2.9258373]
[32745:2.9363315,31182:2.9344323,5197:2.9338512,15733:2.9264717,20376:2.9260075,14354:2.9255145]
10   [31182:2.9142878,32745:2.910867,6119:2.9047217,5197:2.9034846,20376:2.9031868,32903:2.901881]
cat: Unable to write to output stream.
ProjectMasterNodeMedium >> ■
Figure 61 Recommendation

```

Case14: 900 mb Data, 2 Node, 3X Replication

```

6/06/11 18:32:10 INFO mapred.JobClient:      Spilled Records=0
6/06/11 18:32:10 INFO mapred.JobClient:      CPU time spent (ms)=231930
6/06/11 18:32:10 INFO mapred.JobClient:      Total committed heap usage (bytes)=1061158912
6/06/11 18:32:10 INFO mapred.JobClient:      Virtual memory (bytes) snapshot=25531842560
6/06/11 18:32:10 INFO mapred.JobClient:      Map output records=34210
6/06/11 18:32:10 INFO mapred.JobClient:      SPLIT_RAW_BYTES=1720
6/06/11 18:32:10 INFO driver.MahoutDriver: Program took 1023019 ms (Minutes: 17.05033333333334)

real    17m5.633s
user    0m7.240s
sys     0m1.356s
projectMasterNodeMedium >> ■
Figure 62 ALS Recommender

```

```

real    3m8.304s
user    0m6.624s
sys     0m1.616s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 | head
Warning: $HADOOP_HOME is deprecated.

[24647:2.9274426,23492:2.9189265,10754:2.9172182,2402:2.9161208,17990:2.9155748,3620:2.9147835]
[24647:2.913304,23492:2.9047534,2402:2.9023118,17990:2.9007971,10754:2.9007406,3620:2.9000907]
[24647:2.898958,23492:2.8943267,10754:2.8893018,2402:2.8884795,17990:2.8884542,3620:2.888273]
[24647:2.969062,23492:2.961102,3620:2.959171,2402:2.9587677,17990:2.958267,10754:2.9562173]
[24647:2.9777026,10754:2.9677284,2402:2.9675403,3620:2.9661143,23492:2.9656253,17990:2.96511]
[24647:2.95807,23492:2.9472353,17990:2.945457,10754:2.9447582,2402:2.9436305,17348:2.9413452]
[24647:2.9127624,23492:2.9018235,10754:2.900798,17990:2.8995655,2402:2.8983133,3620:2.8980021]
[23492:2.9274454,17990:2.9241748,10754:2.9232764,3620:2.9218276,12734:2.918484,27491:2.9183967]
[24647:2.9761057,10754:2.9666622,23492:2.9652522,2402:2.9628403,3620:2.9620304,17348:2.9619768]
10   [24647:2.9203556,23492:2.912268,2402:2.9099483,10754:2.907753,17990:2.907588,3620:2.907338]
Figure 63 Recommendation

```

Case15: 300 mb Data, 1 Node, 3X Replication

```

16/06/11 19:50:04 INFO mapred.JobClient: Total committed heap usage (bytes)=920649728
16/06/11 19:50:04 INFO mapred.JobClient: Virtual memory (bytes) snapshot=10212737024
16/06/11 19:50:04 INFO mapred.JobClient: Map output records=34210
16/06/11 19:50:04 INFO mapred.JobClient: SPLIT_RAW_BYTES=688
16/06/11 19:50:04 INFO driver.MahoutDriver: Program took 1955798 ms (Minutes: 32.59663333333334)

real    32m41.368s
user    0m18.344s
sys     0m3.968s

```

Figure 64 ALS Recommender

```

real    2m44.943s
user    0m6.636s
sys     0m1.580s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

1 [5025:3.0668354,25851:3.0273683,6861:3.0271487,3725:3.0249147,2917:3.0230806,30549:3.0219193]
2 [25851:3.0066197,13424:2.9923353,12329:2.992327,8744:2.9889772,1395:2.988707,31729:2.984896]
3 [1983:3.152609,8744:3.1250632,3468:3.0986872,7905:3.0970042,16597:3.091307,23182:3.0826802]
4 [29207:3.0025709,25851:3.001975,16215:3.0014956,26423:3.0000198,963:2.9970121,28132:2.9968317]
5 [4021:3.055281,13424:3.0542881,5025:3.0455387,13586:3.0409887,8744:3.0387533,26761:3.0373423]
6 [19128:3.0783486,13424:3.077948,8497:3.0642405,26806:3.061783,19185:3.0542593,19866:3.052964]
7 [31729:3.027727,5672:3.0196118,16033:3.0170956,8667:3.0143542,29618:3.011522,4021:3.0095286]
8 [25851:3.0235865,12329:3.0198948,5672:3.0153043,4839:3.0128422,26806:3.0112295,29618:3.0094182]
9 [6447:3.0254445,32522:3.0195084,25851:3.0106986,12329:3.0079517,31729:3.0056264,32044:3.0055742]
10 [31764:3.1180358,23512:3.1125286,18450:3.1005,24205:3.095264,20988:3.092279,25451:3.0887573]

```

Figure 65 Recommendation

Case15: 300 mb Data, 1 Node, 1X Replication

```

real    29m18.247s
user    0m15.928s
sys     0m3.296s

```

Figure 66 ALS Recommender

```

real    3m3.477s
user    0m8.140s
sys     0m2.036s
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

[6583:3.0673423,26423:3.0539372,3301:3.0304775,8755:3.0304015,2812:3.021269,23617:3.0188358]
[26423:2.9991086,9238:2.9795334,7133:2.9743433,32362:2.9743006,16280:2.9691396,28487:2.9654222]
[31729:3.0961905,4671:3.086367,18450:3.0747948,3996:3.073283,15237:3.0678723,5343:3.0637293]
[30082:3.0316615,19128:3.0288851,28487:3.0192173,32362:3.0051484,10527:2.9996324,19516:2.9990249]
[7085:3.0682912,751:3.0602856,28487:3.03476,4671:3.0271938,15510:3.0242167,23158:3.0235136]
[19128:3.040811,25851:3.025355,26105:3.0163875,751:3.0100749,12835:3.0065968,26709:3.005068]
[13586:3.0031683,24980:3.0016553,32362:2.9888604,31040:2.9880128,14407:2.9855745,23255:2.9848592]
[29618:3.0148985,14547:3.004877,13263:3.002067,26105:3.001686,4717:3.0011692,6962:3.0004873]
[20452:3.02258,769:3.0139883,26709:3.003263,29618:2.99201,26684:2.9893951,7378:2.975934]
0 [8755:3.0988452,18144:3.0749974,24501:3.0722141,24980:3.0721233,20635:3.062197,3301:3.0517662]

```

Figure 67 Recommendation

Case15: 600 mb Data, 1 Node, 1X Replication

```

real    63m18.397s
user    0m22.508s
sys     0m4.160s

```

Figure 68 ALS Recommender

```

real    5m17.423s
user    0m6.992s
sys     0m1.804s
ProjectMasterNodeMedium >> clear

ProjectMasterNodeMedium >> cd
ProjectMasterNodeMedium >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0000 | head
Warning: $HADOOP_HOME is deprecated.

1      [17690:2.8513615,27175:2.8506548,24623:2.8491817,32697:2.8486383,28950:2.8480084,11270:2.8435159]
2      [26248:2.9497707,11975:2.9398718,32697:2.9377353,11270:2.9371731,28950:2.9362674,27253:2.9361854]
3      [26248:2.9726243,11975:2.9641697,11270:2.961294,27253:2.9609854,24623:2.9607506,32697:2.9603114]
4      [10681:2.9831417,12725:2.979645,26248:2.9777124,32697:2.97771488,30170:2.9763553,24623:2.9753823]
5      [28950:2.9501584,32697:2.94453,24623:2.9425206,26248:2.941488,11975:2.940263,6666:2.9377668]
6      [32697:2.8967493,27253:2.8900602,11975:2.8898609,10681:2.884583,30170:2.8832572,11270:2.8821082]
7      [26248:2.9667308,11975:2.9622462,11270:2.9613824,32697:2.9601972,24623:2.9562807,6666:2.9554143]
8      [32807:1.9331316,12192:1.9292378,19844:1.9212518,2850:1.9182612,15563:1.9163477,3274:1.9159428]
9      [26248:2.962951,11270:2.9576058,11975:2.9501624,27253:2.9496057,32697:2.94919,24398:2.9453866]
10     [26248:2.952978,11270:2.9496558,32697:2.945463,27253:2.943605,28950:2.94319,24623:2.941722]
cat: Unable to write to output stream.

```

Figure 69 Recommendation

Screenshots: t2.large

Case1: 600 mb Data, 3 Node, 1X Replication

```

6/06/11 10:01:08 INFO mapred.JobClient: Data-local Map tasks=4
6/06/11 10:01:08 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=0
6/06/11 10:01:08 INFO mapred.JobClient: File Output Format Counters
6/06/11 10:01:08 INFO mapred.JobClient: Bytes Written=6010199
6/06/11 10:01:08 INFO mapred.JobClient: FileSystemCounters
6/06/11 10:01:08 INFO mapred.JobClient: HDFS_BYTES_READ=467446018
6/06/11 10:01:08 INFO mapred.JobClient: FILE_BYTES_WRITTEN=156905
6/06/11 10:01:08 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=6010199
6/06/11 10:01:08 INFO mapred.JobClient: File Input Format Counters
6/06/11 10:01:08 INFO mapred.JobClient: Bytes Read=442771318
6/06/11 10:01:08 INFO mapred.JobClient: Map-Reduce Framework
6/06/11 10:01:08 INFO mapred.JobClient: Map input records=34210
6/06/11 10:01:08 INFO mapred.JobClient: Physical memory (bytes) snapshot=3244208128
6/06/11 10:01:08 INFO mapred.JobClient: Spilled Records=0
6/06/11 10:01:08 INFO mapred.JobClient: CPU time spent (ms)=94340
6/06/11 10:01:08 INFO mapred.JobClient: Total committed heap usage (bytes)=2354053120
6/06/11 10:01:08 INFO mapred.JobClient: Virtual memory (bytes) snapshot=56826241024
6/06/11 10:01:08 INFO mapred.JobClient: Map output records=34210
6/06/11 10:01:08 INFO mapred.JobClient: SPLIT_RAW_BYTES=1204
6/06/11 10:01:08 INFO driver.MahoutDriver: Program took 594109 ms (Minutes: 9.901816666666667)

real    9m56.104s
user    0m5.848s
sys     0m0.976s

```

Figure 70 Alternating Least Squares Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    1m5.072s
user    0m3.440s
sys     0m0.704s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[20996:3.039518,21996:3.0342677,6155:3.0233347,31767:3.0186043,26052:3.0180316,20805:3.0160983]
[20996:2.9645383,21996:2.9588432,6155:2.9463243,30964:2.9415817,26052:2.9414485,31767:2.9394255]
[20996:3.0084803,21996:2.9930353,6155:2.9807646,20805:2.978081,30964:2.9778044,15535:2.9766548]
[21996:3.022399,20996:3.0171201,13047:3.0077677,30964:2.9998674,3728:2.999789,31767:2.9977853]
[20996:2.9704738,21996:2.952211,6155:2.9488754,26052:2.9469318,15535:2.9389203,30964:2.9362006]
[20996:2.9400218,21996:2.9338212,26052:2.9284809,6155:2.9231837,31767:2.9218812,12971:2.91699]
[21996:2.9401298,20996:2.9394565,31767:2.9302921,26052:2.9261906,6155:2.9251213,30964:2.9233325]
[20996:2.943054,21996:2.9311078,26052:2.9302351,31767:2.926107,6155:2.9224527,13047:2.9136071]
[20996:2.9698923,21996:2.9592977,26052:2.953507,6155:2.9512951,12971:2.9407408,13047:2.940076]
[21996:2.936644,20996:2.933355,26052:2.923266,6155:2.920732,31767:2.9186027,30964:2.9173849]
cat: Unable to write to output stream.
```

Figure 6 Recommendation

Case2: 1.2 GB Data, 3 Node, 1X Replication

```
16/06/11 10:36:31 INFO mapred.JobClient:          Total committed heap usage (bytes): 745172946
16/06/11 10:36:31 INFO mapred.JobClient:  Virtual memory (bytes) snapshot=97416413184
16/06/11 10:36:31 INFO mapred.JobClient:  Map output records=39999
16/06/11 10:36:31 INFO mapred.JobClient:  SPLIT_RAW_BYTES=2064
16/06/11 10:36:31 INFO als.ParallelALSFactorizationJob: Recomputing M (iteration 2/3)
16/06/11 10:36:31 INFO input.FileInputFormat: Total input paths to process : 1
16/06/11 10:36:31 INFO mapred.JobClient: Running job: job_201606110929_0025
16/06/11 10:36:32 INFO mapred.JobClient: map 0% reduce 0%
16/06/11 10:36:32 INFO mapred.JobClient: Job complete: job_201606110929_0025
16/06/11 10:36:32 INFO mapred.JobClient: Counters: 0
16/06/11 10:36:32 INFO driver.MahoutDriver: Program took 927446 ms (Minutes: 15.45743333333332)

real    15m29.492s
user    0m6.196s
sys     0m1.212s
```

Figure 71 Alternating Least Squares Recommender

```
real    3m4.179s
user    0m3.464s
sys     0m0.716s
ProjectMasterNode >> https://www.dropbox.com/s/dc4nrllm7c1ep5w/Screenshot%202016-06-11%2006.26.52.png?dl=0
-bash: https://www.dropbox.com/s/dc4nrllm7c1ep5w/Screenshot%202016-06-11%2006.26.52.png?dl=0: No such file or direct
y
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[26172:2.9147668,10467:2.9133258,18298:2.9129195,17554:2.9128878,29979:2.9122787,27321:2.9120197]
[26172:2.8972216,10467:2.8962677,18298:2.895745,17554:2.8955677,29979:2.8949943,27321:2.894441]
[26172:2.851869,18298:2.8504612,10467:2.8500564,17554:2.849539,27321:2.848687,29979:2.8484626]
[26172:2.9042556,10467:2.9028246,29979:2.9017415,17554:2.9017081,13506:2.901405,27321:2.9012632]
[10467:2.89091,18298:2.890588,17554:2.8900406,29979:2.8892655,13506:2.8890421,27321:2.8886962]
[26172:2.9195218,10467:2.9181628,18298:2.9178119,17554:2.9171207,29979:2.9166095,27321:2.9161131]
[26172:2.844431,10467:2.8428319,18298:2.8428054,17554:2.8421092,27321:2.8420453,29979:2.8419268]
[26172:2.9543703,18298:2.9531229,17554:2.9526732,29979:2.951432,13506:2.9508986,27321:2.9508755]
[26172:2.8812687,10467:2.8801394,18298:2.8797624,29979:2.8788288,27321:2.8783293,13506:2.8782156]
[26172:2.8756514,10467:2.8742807,18298:2.8738627,17554:2.873289,29979:2.8727136,27321:2.872466]
```

Figure 72 Recommendation

Case3: 1.9 GB Data, 3 Node, 1X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```

16/06/11 11:18:20 INFO mapred.JobClient: File Input Format counters
16/06/11 11:18:20 INFO mapred.JobClient: Bytes Read=1382997941
16/06/11 11:18:20 INFO mapred.JobClient: Map-Reduce Framework
16/06/11 11:18:20 INFO mapred.JobClient: Map input records=34210
16/06/11 11:18:20 INFO mapred.JobClient: Physical memory (bytes) snapshot=7710461952
16/06/11 11:18:20 INFO mapred.JobClient: Spilled Records=0
16/06/11 11:18:20 INFO mapred.JobClient: CPU time spent (ms)=398540
16/06/11 11:18:20 INFO mapred.JobClient: Total committed heap usage (bytes)=5673844736
16/06/11 11:18:20 INFO mapred.JobClient: Virtual memory (bytes) snapshot=170478723072
16/06/11 11:18:20 INFO mapred.JobClient: Map output records=34210
16/06/11 11:18:20 INFO mapred.JobClient: SPLIT_RAW_BYTES=3612
16/06/11 11:18:20 INFO driver.MahoutDriver: Program took 1405333 ms (Minutes: 23.422233333333333)

real    23m27.418s
user    0m6.784s
sys     0m1.092s

```

Figure 73 Alternating Least Squares Recommender

```

real    3m4.179s
user    0m3.464s
sys     0m0.716s
ProjectMasterNode >> https://www.dropbox.com/s/dc4nrlm7c1ep5w/Screenshot%202016-06-11%2006.26.52.png?dl=0
-bash: https://www.dropbox.com/s/dc4nrlm7c1ep5w/Screenshot%202016-06-11%2006.26.52.png?dl=0: No such file or direct
y
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

1      [26172:2.9147668,10467:2.9133258,18298:2.9129195,17554:2.9128878,29979:2.9122787,27321:2.9120197]
2      [26172:2.8972216,10467:2.8962677,18298:2.895745,17554:2.8955677,29979:2.8949943,27321:2.894441]
3      [26172:2.851869,18298:2.8504612,10467:2.8500564,17554:2.849539,27321:2.848687,29979:2.8484626]
4      [26172:2.9042556,10467:2.9028246,29979:2.9017415,17554:2.9017081,13506:2.901405,27321:2.9012632]
5      [10467:2.89091,18298:2.890588,17554:2.8900406,29979:2.8892655,13506:2.8890421,27321:2.8886962]
6      [26172:2.9195218,10467:2.9181628,18298:2.9178119,17554:2.9171207,29979:2.9166095,27321:2.9161131]
7      [26172:2.844431,10467:2.8428319,18298:2.8428054,17554:2.8421092,27321:2.8420453,29979:2.8419268]
8      [26172:2.9543703,18298:2.9531229,17554:2.9526732,29979:2.951432,13506:2.9508986,27321:2.9508755]
9      [26172:2.8812687,10467:2.8801394,18298:2.8797624,29979:2.8788288,27321:2.8783293,13506:2.8782156]
10     [26172:2.8756514,10467:2.8742807,18298:2.8738627,17554:2.873289,29979:2.8727136,27321:2.872466]

```

Figure 74 Recommendation

Case4: 600 mb Data, 3 Node, 3X Replication

```

16/06/11 11:56:54 INFO mapred.JobClient: Physical memory (bytes) snapshot=5522065550
16/06/11 11:56:54 INFO mapred.JobClient: Spilled Records=0
16/06/11 11:56:54 INFO mapred.JobClient: CPU time spent (ms)=94350
16/06/11 11:56:54 INFO mapred.JobClient: Total committed heap usage (bytes)=2458910720
16/06/11 11:56:54 INFO mapred.JobClient: Virtual memory (bytes) snapshot=56826241024
16/06/11 11:56:54 INFO mapred.JobClient: Map output records=34210
16/06/11 11:56:54 INFO mapred.JobClient: SPLIT_RAW_BYTES=1204
16/06/11 11:56:54 INFO driver.MahoutDriver: Program took 603254 ms (Minutes: 10.05423333333332)

real    10m5.315s
user    0m6.048s
sys     0m1.144s

```

Figure 75 Alternating Least Squares Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
16/06/11 11:58:54 INFO mapred.JobClient:      Map output records=19999
16/06/11 11:58:54 INFO mapred.JobClient:      SPLIT_RAW_BYT
16/06/11 11:58:54 INFO driver.MahoutDriver: Program took 66074 ms (Minutes: 1.101233333333333)
real    1m8.100s
user    0m3.432s
sys     0m0.704s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

1 [14911:2.9895942,9492:2.9884815,7864:2.9872615,28996:2.9855423,28031:2.982319,12015:2.9813244]
2 [28996:2.9490075,14911:2.9459853,12015:2.9442844,9654:2.9441102,9492:2.9428568,33382:2.939939]
3 [9492:2.9567614,9654:2.956703,14911:2.9563005,28996:2.9560053,12015:2.9543347,26844:2.9542809]
4 [9492:2.94451,14911:2.9360564,9654:2.9329221,12015:2.9314854,28996:2.9279518,26844:2.925734]
5 [14911:3.0124297,28996:3.0117252,9492:3.0116425,9654:3.0096877,12015:3.009683,7864:3.0087421]
6 [14911:2.9569185,28996:2.9544713,9492:2.9535587,12015:2.9525454,9654:2.9496734,26844:2.9495456]
7 [28996:2.933884,14911:2.93264,12015:2.931592,9492:2.9297748,28031:2.9271827,33382:2.9268844]
8 [14911:3.0125606,12015:3.0118904,26844:3.010675,28031:3.0099306,28996:3.0097933,9654:3.0063176]
9 [28996:2.9755497,7864:2.9751428,28031:2.9743972,14911:2.9742355,3524:2.9712644,33382:2.9698775]
10 [9654:2.9853234,12015:2.9799166,28996:2.9795434,33382:2.9762576,2123:2.9744375,14911:2.9735916]
```

Figure 76 Recommendation

Case5: 300 mb Data, 3 Node, 3X Replication

```
16/06/11 13:23:19 INFO mapred.JobClient:      Spilled Records=0
16/06/11 13:23:19 INFO mapred.JobClient:      CPU time spent (ms)=51470
16/06/11 13:23:19 INFO mapred.JobClient:      Total committed heap usage (bytes)=909639680
16/06/11 13:23:19 INFO mapred.JobClient:      Virtual memory (bytes) snapshot=32472137728
16/06/11 13:23:19 INFO mapred.JobClient:      Map output records=34210
16/06/11 13:23:19 INFO mapred.JobClient:      SPLIT_RAW_BYT
16/06/11 13:23:19 INFO driver.MahoutDriver: Program took 445594 ms (Minutes: 7.426566666666667)

real    7m27.581s
user    0m5.808s
sys     0m1.024s
ProjectMasterNode >>
```

Figure 77 ALS Recommender

```
real    1m1.113s
user    0m3.460s
sys     0m0.748s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 |
head
Warning: $HADOOP_HOME is deprecated.

1 [27397:3.0267243,29319:2.9970405,6411:2.9931629,31924:2.992212,32850:2.9840126,5330:2.9829319]
2 [29847:3.0209172,944:3.017914,27397:3.0089996,31575:3.006653,2666:2.997159,31434:2.9945765]
3 [27397:3.0227983,2666:3.0196712,29847:3.0151205,28164:3.0031364,5983:2.9980578,32850:2.9961896]
4 [27397:3.0947876,32850:3.0542572,14225:3.0529509,17387:3.0401359,2666:3.0392287,27121:3.0387888]
5 [32850:3.0186102,27397:3.016448,33888:2.9818351,29847:2.9809198,28077:2.9798198,2666:2.9773676]
6 [19473:3.0179563,2666:3.0169952,11577:3.0130723,21104:3.0020077,27397:3.0005167,29319:2.9949641]
7 [27397:3.057837,29319:3.0284414,27121:3.0262787,2666:3.024553,29847:3.0205386,28164:3.0177953]
8 [2666:3.0439782,29847:3.043429,8410:3.0293024,27121:3.0219386,27397:3.0125368,19473:3.0016847]
9 [2666:3.0780013,29847:3.0585747,7544:3.040207,17100:3.0366147,29319:3.0338764,13031:3.0334437]
10 [6801:3.0334086,2666:3.0199645,27282:3.009603,27397:3.007107,29847:2.997906,26686:2.997774]
cat: Unable to write to output stream.
```

Figure 78 Recommendation

Case6: 900 mb Data, 3 Node, 3X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```

6/06/11 14:06:36 INFO mapred.JobClient: Spilled Records=0
6/06/11 14:06:36 INFO mapred.JobClient: CPU time spent (ms)=181600
6/06/11 14:06:36 INFO mapred.JobClient: Total committed heap usage (bytes)=2671771648
6/06/11 14:06:36 INFO mapred.JobClient: Virtual memory (bytes) snapshot=81180344320
6/06/11 14:06:36 INFO mapred.JobClient: Map output records=34210
6/06/11 14:06:36 INFO mapred.JobClient: SPLIT_RAW_BYT
6/06/11 14:06:36 INFO driver.MahoutDriver: Program took 830562 ms (Minutes: 13.8427)

real    13m52.598s
user    0m6.316s
sys     0m1.144s

```

Figure 79 ALS Recommender

```

real    2m23.315s
user    0m3.632s
sys     0m0.676s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[17688:2.9101021,14983:2.9080033,18373:2.904111,32062:2.9032412,17713:2.9011889,29413:2.9009297]
[14983:2.928727,17688:2.9280117,32062:2.9245963,18373:2.9230645,21347:2.9222963,17713:2.9213278]
[14983:3.0229604,17688:3.021016,17713:3.0159502,19852:3.0154157,18373:3.0150862,13135:3.0137389]
[17688:2.8879678,32062:2.8829188,17713:2.8799155,19852:2.8790324,13135:2.8788173,18373:2.8786545]
[17688:2.940017,32062:2.9370315,18373:2.9344318,17713:2.9334214,19852:2.932329,29413:2.931379]
[17688:2.919118,14983:2.9189942,32062:2.9147344,18373:2.9137301,19852:2.9130106,17713:2.9120097]
[17688:2.9470491,14983:2.9465542,18373:2.9387848,13299:2.9375174,32062:2.9374993,17713:2.9372408]
[17688:2.957436,14983:2.9569936,32062:2.9525766,18373:2.9512177,13299:2.9509718,17713:2.9508026]
[17688:2.920926,14983:2.9141505,19852:2.9135633,32062:2.911363,18373:2.9110358,13135:2.9086792]
[17688:2.9416757,14983:2.9370277,17713:2.9335449,29413:2.932859,13299:2.9322126,18373:2.9321494]
cat: Unable to write to output stream.

```

Figure 80 Recommendation

Case7: 900 mb Data, 3 Node, 1X Replication

```

6/06/11 15:08:29 INFO mapred.JobClient: Spilled Records=0
6/06/11 15:08:29 INFO mapred.JobClient: CPU time spent (ms)=183070
6/06/11 15:08:29 INFO mapred.JobClient: Total committed heap usage (bytes)=2534408192
6/06/11 15:08:29 INFO mapred.JobClient: Virtual memory (bytes) snapshot=81180344320
6/06/11 15:08:29 INFO mapred.JobClient: Map output records=34210
6/06/11 15:08:29 INFO mapred.JobClient: SPLIT_RAW_BYT
6/06/11 15:08:29 INFO driver.MahoutDriver: Program took 805822 ms (Minutes: 13.430366666666666)

real    13m27.895s
user    0m6.140s
sys     0m1.036s

```

Figure 81 ALS Recommender

```

real    1m51.162s
user    0m3.372s
sys     0m0.760s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[14983:2.9064815,17713:2.9020796,5770:2.9000068,32062:2.897267,9852:2.8968203,23682:2.8968108]
[14983:2.9257102,17713:2.9207435,5770:2.916361,9852:2.9160974,21347:2.9154143,32062:2.9147935]
[14983:3.0348008,17713:3.0333383,32062:3.0295408,5770:3.0287766,23682:3.0261927,30613:3.0247135]
[17713:2.8800747,32062:2.8755202,9852:2.8747609,23682:2.8746026,5770:2.8739326,30613:2.8715367]
[14983:2.9490726,17713:2.9462776,23682:2.9438798,32062:2.9424427,5770:2.9415452,9852:2.9392374]
[14983:2.9015899,17713:2.8993475,32062:2.8951962,23682:2.8950803,5770:2.894648,21347:2.891122]
[14983:2.9631963,17713:2.95472,32062:2.953699,5770:2.9531505,23682:2.9530761,9852:2.951297]
[14983:2.9397573,17713:2.93715,23682:2.9343157,32062:2.933348,5770:2.9327013,9852:2.9310117]
[14983:2.9058692,17713:2.9036627,9852:2.902829,23682:2.9024117,32062:2.9023998,5770:2.9017875]
[14983:2.936462,17713:2.9333324,32062:2.9303033,5770:2.929708,9852:2.9278476,23682:2.9272077]
cat: Unable to write to output stream.

```

Figure 82 Recommendation

Case8: 300 mb Data, 3 Node, 1X Replication

```
6/06/11 15:30:02 INFO mapred.JobClient: SPLIT_RAW_BYTES=688
6/06/11 15:30:02 INFO driver.MahoutDriver: Program took 428903 ms (Minutes: 7.148383333333333)
real    7m10.916s
user    0m5.760s
sys     0m0.996s
projectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
Figure 83 ALS Recommendation
```

```
real    1m3.151s
user    0m3.472s
sys     0m0.712s
projectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-0
head
Warning: $HADOOP_HOME is deprecated.

[27522:3.003165,8204:2.981991,33217:2.981701,16363:2.9766974,20990:2.9722486,23212:2.9692936]
[24119:3.0688312,23801:3.0311468,30482:3.0308924,4893:3.017572,20990:3.0142484,209:3.0117888]
[24119:3.0892248,23801:3.0872183,33550:3.0709417,13900:3.06898,32452:3.0647943,20990:3.0585623]
[24119:3.0212529,5440:2.9962642,33550:2.993602,23801:2.9928966,4893:2.992315,34060:2.9902177]
[24119:2.9959474,20990:2.994998,33550:2.9912035,209:2.9899988,23240:2.9873023,27522:2.983245]
[16363:3.0869226,4893:3.0772226,24119:3.0711024,12335:3.0311608,23240:3.0216284,18117:3.0210793]
[23801:3.0572724,1114:3.0364668,1581:3.0328753,32452:3.0313637,30482:3.028893,209:3.0281525]
[24119:3.034918,22144:3.0220706,209:3.0152833,16363:3.0104072,4009:3.004164,2296:3.0033293]
[28561:3.1455598,10608:3.1248987,1114:3.1205616,30187:3.1194434,13900:3.117354,1581:3.1108215]
[24119:3.0580096,20990:3.0136871,4893:3.0095265,7982:3.0022435,30482:2.9991474,32452:2.998315]
cat: Unable to write to output stream.
projectMasterNode >>
```

Figure 84 Recommendation

Case9: 300 mb Data, 2 Node, 1X Replication

```
real    7m43.154s
user    0m5.636s
sys     0m1.004s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
--output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData/als/
Figure 85 ALS Recommender
```

```
real    1m5.085s
user    0m3.372s
sys     0m0.700s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 |
head
Warning: $HADOOP_HOME is deprecated.

[30878:2.9799953,24119:2.972371,10762:2.9591956,18117:2.9532,12406:2.9524424,16770:2.951975]
[24119:3.0295637,27522:3.0213604,4433:3.0189414,32452:3.0178525,31367:3.0074348,28655:3.0049996]
[24119:3.109551,28655:3.078495,33550:3.072054,27511:3.0662622,27522:3.0586777,20675:3.05778]
[30878:3.027904,10608:3.0090575,17946:3.0074768,32452:3.006122,23799:2.998518,28561:2.997436]
[27511:3.0023518,24119:3.000735,33550:2.99934,27522:2.9973505,32452:2.9886274,12582:2.9882443]
[24119:3.0503738,23749:3.0186882,23801:3.0147305,8890:3.0133936,23212:3.0130873,28876:3.0058453]
[10608:3.0362709,22755:3.023217,28561:3.022545,26879:3.0181046,3551:3.0179815,29522:3.0179482]
[28655:3.0364256,23801:3.024373,27522:3.0208876,32452:3.0206945,33550:3.0198252,15612:3.0160544]
[24119:3.1970646,25256:3.1505775,20699:3.144074,31367:3.1389532,23799:3.1387126,3752:3.1376598]
[33550:3.0195055,24119:2.9967988,16942:2.9851513,32452:2.9789417,23801:2.9748263,3551:2.9712005]
cat: Unable to write to output stream.
```

Figure 86 Recommendation

Case10: 600 mb Data, 2 Node, 1X Replication

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
6/06/11 17:28:39 INFO mapred.JobClient: Map input records=34210
6/06/11 17:28:39 INFO mapred.JobClient: Physical memory (bytes) snapshot=2686128128
6/06/11 17:28:39 INFO mapred.JobClient: Spilled Records=0
6/06/11 17:28:39 INFO mapred.JobClient: CPU time spent (ms)=95540
6/06/11 17:28:39 INFO mapred.JobClient: Total committed heap usage (bytes)=1798832128
6/06/11 17:28:39 INFO mapred.JobClient: Virtual memory (bytes) snapshot=56826241024
6/06/11 17:28:39 INFO mapred.JobClient: Map output records=34210
6/06/11 17:28:39 INFO mapred.JobClient: SPLIT_RAW_BYTES=1204
6/06/11 17:28:39 INFO driver.MahoutDriver: Program took 718472 ms (Minutes: 11.974533333333333)

real    12m0.550s
user    0m5.996s
sys     0m1.052s
ProjectMasterNode>
```

Figure 87 ALS Recommender

```
real    1m47.110s
user    0m3.424s
sys     0m0.664s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[3280:2.9059482,25923:2.9024813,2458:2.9021406,17901:2.8992834,10731:2.8941815,24547:2.8940938]
[25923:3.0022624,17901:2.9938827,24547:2.9931073,22946:2.9920259,27984:2.9917321,2458:2.990538]
[3280:2.9747536,2458:2.9746265,25923:2.9726088,17901:2.968016,10731:2.9673884,24547:2.9670906]
[25923:3.062246,17901:3.0533674,3280:3.048042,17423:3.0479662,32618:3.0436301,2458:3.0433748]
[3280:3.0355096,2458:3.0340157,25923:3.0324488,27984:3.0291102,17901:3.027604,10731:3.0273972]
[3280:3.0947874,25923:3.0899239,27984:3.0857933,22946:3.0849187,26823:3.0848584,17901:3.084004]
[25923:3.0142732,27984:3.0106556,439:3.0045936,16537:3.0040894,32618:3.001211,2458:2.9969115]
[25923:2.9257557,2458:2.9222703,3280:2.919261,17901:2.9171684,24547:2.9098606,10731:2.909422]
[25923:2.8873267,2458:2.8846223,10731:2.8818142,27984:2.8811038,24547:2.8789408,17901:2.8760402]
[25923:2.9553635,2458:2.9498096,3280:2.9494576,17901:2.9471927,27984:2.9439726,10731:2.9421325]
```

Figure 88 Recommendation

Case11: 900 mb Data, 2 Node, 1X Replication

```
6/06/11 16:57:27 INFO mapred.JobClient: Virtual memory (bytes) snapshot=61160544520
6/06/11 16:57:27 INFO mapred.JobClient: Map output records=34210
6/06/11 16:57:27 INFO mapred.JobClient: SPLIT_RAW_BYTES=1720
6/06/11 16:57:27 INFO driver.MahoutDriver: Program took 995843 ms (Minutes: 16.597383333333333)

real    16m37.996s
user    0m6.324s
sys     0m1.016s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
                           --output /user/ec2-user/Project/ProjectData/recommendations/ --userFeatures /user/ec2-user/Project/ProjectData/als/
                           --itemFeatures /user/ec2-user/Project/ProjectData/items/ --itemFeatures /user/ec2-user/Project/ProjectData/items/
```

Figure 89 ALS Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    2m7.298s
user    0m3.596s
sys     0m0.668s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 | head
warning: $HADOOP_HOME is deprecated.

[32062:2.9145722,18373:2.911623,968:2.9063194,17713:2.906218,9852:2.9003546,30723:2.894319]
[32062:2.9324713,18373:2.9306126,968:2.923289,17713:2.9212694,9852:2.920163,11040:2.9127266]
[32062:3.0041006,17713:3.0005515,18373:2.9954083,968:2.9954062,9852:2.9882247,30723:2.9822557]
[32062:2.879756,18373:2.876595,17713:2.8701775,9852:2.8683126,11040:2.860594,19852:2.860543]
[32062:2.9538147,18373:2.951642,17713:2.9461575,968:2.9447782,9852:2.9425893,30723:2.9329998]
[32062:2.9260926,18373:2.9232185,17713:2.917728,968:2.9123895,19852:2.9050598,30723:2.9045336]
[18373:2.9682467,32062:2.966065,968:2.9577534,17713:2.9568567,9852:2.95558,11040:2.947279]
[32062:2.9631038,18373:2.9592462,17713:2.9535232,968:2.9534276,9852:2.950024,25348:2.9409695]
[32062:2.8951752,18373:2.8944783,968:2.8872821,17713:2.8827345,9852:2.8817008,25348:2.8761523]
[32062:2.9396045,18373:2.9365091,17713:2.92841,9852:2.9275343,30723:2.9191928,19852:2.918813]
at: Unable to write to output stream.
```

Figure 90 Recommendation

Case12: 300 mb Data, 2 Node, 3X Replication

```
real    7m39.000s
user    0m5.716s
sys     0m1.008s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
```

Figure 91 ALS Recommender

```
real    1m4.119s
user    0m3.400s
sys     0m0.804s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 | head
warning: $HADOOP_HOME is deprecated.

[21246:3.0262728,31574:3.0238276,10332:3.0152965,16098:3.0134022,28245:3.013174,12768:3.0078552]
[33867:3.0352721,8375:3.0264907,8631:3.023203,19649:3.0161982,21246:3.015424,16098:3.0128434]
[8375:2.9991353,21246:2.993669,1384:2.9837277,16098:2.9823465,23717:2.9815261,2492:2.979478]
[17250:3.0151787,23717:2.9992511,6407:2.9963722,8631:2.9920106,23542:2.9877703,3287:2.985891]
[16098:3.0864692,21246:3.0758708,8375:3.057363,23717:3.0564742,2745:3.0446527,31574:3.0445917]
[8375:3.0623832,4547:3.052324,4778:3.0510247,22882:3.0489502,21246:3.0452955,12768:3.0434291]
[10332:3.026444,4547:3.0206392,21246:3.0197408,22226:3.0140295,12768:3.0134387,19649:3.0077934]
[8375:3.0624204,21246:3.0524943,22226:3.0400064,1951:3.0374105,10332:3.0313349,732:3.021798]
[22841:3.0386465,12668:3.0127182,22378:3.0072806,27741:3.0013523,13004:2.9959452,8375:2.9904363]
[17250:2.9735458,21246:2.9672637,8631:2.9633489,33867:2.962603,27603:2.9622335,29102:2.9545975]
```

Figure 92 Recommendation

Case13: 600 mb Data, 2 Node, 3X Replication

```
16/06/11 18:04:10 INFO mapred.JobClient:      total committed heap usage (bytes)=2599944192
16/06/11 18:04:10 INFO mapred.JobClient:      Virtual memory (bytes) snapshot=56826241024
16/06/11 18:04:10 INFO mapred.JobClient:      Map output records=34210
16/06/11 18:04:10 INFO mapred.JobClient:      SPLIT_RAW_BYTES=1204
16/06/11 18:04:10 INFO driver.MahoutDriver: Program took 818733 ms (Minutes: 13.64555)

real    13m40.764s
user    0m6.136s
sys     0m1.164s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
```

Figure 93 ALS Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    1m51.267s
user    0m3.496s
sys     0m0.772s
ProjectMasterNode >> https://www.dropbox.com/s/igsu8b2efk0bg2y/Screenshot%202016-06-11%2013.06.51.png?dl=0
-bash: https://www.dropbox.com/s/igsu8b2efk0bg2y/Screenshot%202016-06-11%2013.06.51.png?dl=0: No such file or d
/
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-
head
Warning: $HADOOP_HOME is deprecated.

[2458:2.9221876,3280:2.918794,33582:2.916048,24547:2.9133592,31511:2.9088342,17901:2.9064255]
[33582:2.9999566,2458:2.9978552,17901:2.9904284,24547:2.9871848,31511:2.98505,30503:2.9837673]
[27810:2.9834507,2458:2.9824421,3280:2.9736507,24547:2.9724393,17901:2.972332,28922:2.9645367]
[3280:3.0794144,2458:3.0774963,27810:3.077404,24547:3.0721152,33582:3.070796,31511:3.0655766]
[27810:3.0571325,33582:3.0560162,2458:3.0483413,24547:3.04589,3280:3.0378165,30503:3.0347505]
[33582:3.107291,27810:3.1069174,2458:3.1008463,24547:3.0931735,17901:3.0902069,13247:3.0867927]
[24854:3.0661457,28922:3.0651202,31511:3.0623665,2458:3.059924,30503:3.0459242,17901:3.0435164]
[2458:2.9479113,3280:2.9461586,24547:2.9441116,27810:2.942572,33582:2.9399993,31511:2.9376063]
[27810:2.9205604,2458:2.9189563,33582:2.918223,24547:2.9097347,17901:2.9088771,30503:2.9083285]
[27810:2.965528,2458:2.9629045,3280:2.9622936,31511:2.9564974,17901:2.9555624,24547:2.9528491]
```

Figure 94 Recommendation

Case14: 900 mb Data, 2 Node, 3X Replication

```
16/06/11 18:33:05 INFO mapred.JobClient: Virtual memory (bytes) snapshot=81180344320
16/06/11 18:33:05 INFO mapred.JobClient: Map output records=34210
16/06/11 18:33:05 INFO mapred.JobClient: SPLIT_RAW_BYTES=1720
16/06/11 18:33:05 INFO driver.MahoutDriver: Program took 1080871 ms (Minutes: 18.014516666666665)

real    18m3.085s
user    0m6.576s
sys     0m1.092s
```

Figure 95 ALS Recommender

```
real    2m20.254s
user    0m3.496s
sys     0m0.684s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 |
head
Warning: $HADOOP_HOME is deprecated.

[17688:2.9138703,19852:2.912204,14983:2.9101949,17713:2.9071398,968:2.9049208,32062:2.903673]
[17688:2.9571857,19852:2.9532864,14983:2.9527013,13668:2.9498112,17713:2.9475923,968:2.946327]
[19852:3.03881,17688:3.0361016,13668:3.0321622,32062:3.031428,14983:3.031413,30435:3.0262253]
[19852:2.8743389,17688:2.8740473,17713:2.8665264,22744:2.8656528,13668:2.8648765,32062:2.8646681]
[19852:2.932686,17688:2.931325,17713:2.925128,13668:2.925045,968:2.9232824,32062:2.922659]
[17688:2.9086947,19852:2.907293,17713:2.9036152,14983:2.9035373,13668:2.9033904,968:2.9005413]
[19852:2.972286,17688:2.971691,17713:2.9649277,14983:2.964424,13668:2.9637825,968:2.9632943]
[17688:2.9410462,19852:2.9383895,13668:2.9352288,14983:2.9352043,17713:2.934408,968:2.9318929]
[19852:2.895515,13668:2.8946629,17688:2.8935819,14983:2.8909774,17713:2.888932,32062:2.8871868]
[2.9196074,17688:2.9170156,14983:2.9141407,17713:2.9129753,13668:2.912583,22744:2.9102788]
at: Unable to write to output stream.
ProjectMasterNode >>
```

Figure 96 Recommendation

Case15: 300 mb Data, 1 Node, 3X Replication

```
16/06/11 19:18:30 INFO driver.MahoutDriver: Program took 670380 ms (Minutes: 11.173)

real    11m12.425s
user    0m5.968s
sys     0m0.920s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRatin
```

Figure 97 ALS Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    1m45.172s
user    0m3.480s
sys     0m0.672s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[21246:3.053608,27603:3.0512452,15264:3.028365,22226:3.027472,8375:3.0260396,7239:3.0144858]
[21246:3.038106,8375:3.0347497,7239:3.0242126,22226:3.0126073,27603:3.0118883,15264:3.0014193]
[8375:3.0258493,12768:3.0228133,11227:3.0180707,26666:3.0071633,31502:3.0016427,7525:2.997145]
[33820:3.001459,23400:2.9910786,27603:2.991068,22226:2.987598,11813:2.9870963,7239:2.9865491]
[8375:3.066537,21246:3.0485055,4812:3.047348,12768:3.0440109,31133:3.0387337,7239:3.036455]
[21246:3.067729,8375:3.0561922,22226:3.036852,27603:3.0365696,2492:3.0352864,432:3.0283835]
[14314:3.0183103,21246:3.0161364,22226:3.0156555,17250:3.0101202,9663:3.0083308,20265:3.00724]
[21246:3.044393,432:3.0362887,7525:3.0306149,26666:3.0298922,6407:3.024292,22226:3.023021]
[17843:2.9976518,2457:2.9975064,368:2.9760206,20201:2.974518,3548:2.9742725,8512:2.9726145]
[21246:3.010788,22226:2.9848504,23400:2.9804182,12768:2.9803035,27603:2.976195,432:2.9713473]
```

Figure 98 Recommendation

Case16: 300 mb Data, 1 Node, 1X Replication

```
real    1m3.204s
user    0m5.812s
sys     0m0.980s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
```

Figure 99 ALS Recommender

```
real    1m43.228s
user    0m3.368s
sys     0m0.728s
ProjectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000
head
Warning: $HADOOP_HOME is deprecated.

[13132:2.969977,13942:2.9664378,1056:2.9624553,29020:2.9612548,248:2.959218,1463:2.957534]
[13964:3.077893,17621:3.0713289,19088:3.0581398,22817:3.05652,25698:3.0560913,15789:3.05463]
[28441:3.058447,17162:3.0152104,497:3.0150418,16636:3.0144482,14074:3.0125,1463:3.006795]
[248:3.0783443,21958:3.0678155,28678:3.06354,11904:3.0505872,2800:3.0502467,16793:3.0493293]
[14074:3.016033,16793:3.0028486,28441:2.990309,9564:2.9889326,11345:2.9843802,12570:2.984054]
[28441:3.0306668,22817:3.029168,17162:3.027576,14074:3.0179777,12977:3.0145953,28678:3.0126326]
[18482:3.2718592,933:3.2587657,29234:3.247667,23738:3.247436,20737:3.2382603,24302:3.2237022]
[28441:3.008972,17162:2.998855,21024:2.9910085,1348:2.9905968,4161:2.987379,28855:2.985498]
[16793:2.995788,29335:2.9803834,15803:2.9791548,28441:2.9767406,29282:2.9732063,21074:2.9716318]
[15789:3.0469842,9775:3.0402584,28441:3.0324159,21582:3.031258,17162:3.0304008,497:3.0294309]
```

Figure 100 Recommendation

Case17: 600 mb Data, 1 Node, 1X Replication

```
real    17m38.369s
user    0m6.268s
sys     0m1.156s
ProjectMasterNode >> time bin/mahout recommendfactorized --input /user/ec2-user/Project/ProjectData/als/out/userRating
```

Figure 101 ALS Recommender

HADOOP PERFORMANCE ANALYSIS USING MAHOUT RECOMMENDER SYSTEM

```
real    2m29.185s
user    0m3.456s
sys     0m0.676s
projectMasterNode >> $HADOOP_HOME/bin/hadoop fs -cat /user/ec2-user/Project/ProjectData/recommendations/part-m-00000 | head
Warning: $HADOOP_HOME is deprecated.

[1081:3.0139387,22570:2.9636045,3544:2.9630477,9994:2.9590073,32886:2.9577854,20406:2.9576302]
[20406:2.9804351,22570:2.9777398,3544:2.971975,32886:2.969749,13742:2.9694862,30732:2.9648552]
[1081:2.9370692,22570:2.9071827,3544:2.897339,20406:2.8967776,32886:2.8952246,9994:2.893322]
[1081:3.0043724,3544:2.9616256,20406:2.9613411,9994:2.9574645,32886:2.9554832,13742:2.9554465]
[1081:2.9921741,9994:2.9495866,20406:2.9484987,3544:2.945581,32886:2.9421802,13742:2.9391]
[1081:2.9863982,9994:2.9423852,20406:2.94226,3544:2.9416583,32886:2.9367144,13742:2.9301705]
[1081:2.986892,3544:2.9429893,20406:2.93977,9994:2.939056,32886:2.9382322,13742:2.9379148]
[1081:2.9980578,3544:2.9581687,20406:2.9563665,9994:2.954717,22570:2.951911,32886:2.9491723]
[1081:2.9609087,3544:2.9137058,9994:2.9136636,22570:2.9122355,20406:2.910344,32886:2.9080977]
0   [1081:2.9916809,22570:2.946494,3544:2.9437432,9994:2.9415417,32886:2.9382045,13742:2.9329314]
at: Unable to write to output stream.
projectMasterNode >>
```

Figure 102 Recommendation