REPORT FOR PROJECT 2

COMPARISON OF QUERY 2 AND QUERY 4

QUERY 2: WITHOUT USING REPICATED JOIN

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MaxReduc
eTime   MinReduceTime   AvgReduceTime   Alias   Feature Outputs
job_201502212346_0038   2       40      1       1       1       11      3       8
1-5,A,B,C       HASH_JOIN
job_201502212346_0039   1       1       0       0       0       9       9       9
D,E     GROUP_BY,COMBINER       /user/o1,
```

Query 2 has two jobs: job_201502212346_0038 and job_201502212346_0038

job_201502212346_0038 finished in 106 seconds

Hadoop job_201502212346_0038 on localhost

User: ubuntu
Job Name: PigLatin:PIG2.pig
Job File: hdfs://localhost:3351/home/ubuntu/Workspace/hadoop-1.1.0/hadoop-data/temp/mapred/staging/ubuntu/.staging/job_201502212346_0038/job.xml
Submit Host: ubuntu-VirtualBox
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Tue Feb 24 03:58:14 CET 2015
Finished at: Tue Feb 24 04:00:01 CET 2015
Finished in: 1mins, 46sec
Job Cleanup: Successful

We also show the screen-shot of its CPU time.

CPU time spent (ms)	540	6,390	6,930

PLAN: 2 Map and 40 Reduce tasks

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	2	0	0	2	0	0 / 0
reduce	100.00%	40	0	0	40	0	0 / 0

job_201502212346_0039 finished in 16 seconds

Hadoop job_201502212346_0039 on localhost

LibreOffice Writer

Job Name: PigLatin:PIG2.pig
Job File: hdfs://localhost:3351/home/ubuntu/Workspace/hadoop-1.1.0/hadoop-data/temp/mapred/staging/ubuntu/.staging/job_201502212346_0039/job.xml
Submit Host: ubuntu-VirtualBox
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Tue Feb 24 04:00:08 CET 2015
Finished at: Tue Feb 24 04:00:24 CET 2015
Finished in: 16sec
Job Cleanup: Successful

We also show the screen-shot of its CPU time.

CPU time spent (ms)	310	480	790

PLAN: 1 Map and 1 Reduce tasks

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	1	0	0	1	0	0 / 0
reduce	100.00%	1	0	0	1	0	0 / 0

QUERY 4: USING REPICATED JOIN

```
Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MaxReduc
eTime   MinReduceTime   AvgReduceTime   Alias   Feature Outputs
job_201502212346_0040   1       0       1       1       1       0       0       0
B       MAP_ONLY
job_201502212346_0041   1       1       1       1       1       10      10      1
0       1-5,A,C,D,E     REPLICATED_JOIN,GROUP_BY,COMBINER       /user/o2,
```

Query 4 has 2 jobs: job_201502212346_0040 and job_201502212346_0041

job_201502212346_0040 finished in 6 seconds

Hadoop job_201502212346_0040 on localhost

User: ubuntu
Job Name: PigLatin:PIG4.pig
Job File: hdfs://localhost:3351/home/ubuntu/Workspace/hadoop-1.1.0/hadoop-data/temp/mapred/staging/ubuntu/.staging/job_201502212346_0040/job.xml
Submit Host: ubuntu-VirtualBox
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Tue Feb 24 04:39:09 CET 2015
Finished at: Tue Feb 24 04:39:15 CET 2015
Finished in: 6sec
Job Cleanup: Successful

We also show the screen-shot of its CPU time.

CPU time spent (ms)	240	0	240

PLAN: 1 Map and 0 Reduce task

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	1	0	0	1	0	0 / 0
reduce	100.00%	0	0	0	0	0	0 / 0

job_201502212346_0041 finished in 16 seconds

Hadoop job_201502212346_0041 on localhost

User: ubuntu
Job Name: PigLatin:PIG4.pig
Job File: hdfs://localhost:3351/home/ubuntu/Workspace/hadoop-1.1.0/hadoop-data/temp/mapred/staging/ubuntu/.staging/job_201502212346_0041/job.xml
Submit Host: ubuntu-VirtualBox
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Tue Feb 24 04:39:25 CET 2015
Finished at: Tue Feb 24 04:39:41 CET 2015
Finished in: 16sec
Job Cleanup: Successful

We also show the screen-shot of its CPU time

CPU time spent (ms)	610	650	1,260

PLAN: 1 Map and 1 Reduce Task

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	1	0	0	1	0	0 / 0
reduce	100.00%	1	0	0	1	0	0 / 0

Conclusion:

1. From the information above, we can find Query5 used almost the same in both their jobs. Meanwhile, there is large time difference between Query2's two jobs.

2. Query5 has a better performance than Query2.

3. The jobs consumed by Query 5 is less compared to Query 2.

COMPARISON OF QUERY6 AND QUERY1

QUERY6

Hadoop job_201502212346_0063 on localhost

User: ubuntu
Job Name: streamjob4902195460289881652.jar
Job File: hdfs://localhost:3351/home/ubuntu/Workspace/hadoop-1.1.0/hadoop-data/temp/mapred/staging/ubuntu/.staging/job_201502212346_0063/job.xml
Submit Host: ubuntu-VirtualBox
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Tue Feb 24 22:17:56 CET 2015
Finished at: Tue Feb 24 22:18:16 CET 2015
Finished in: 20sec
Job Cleanup: Successful

CPU TIME:

CPU time spent (ms)	370	1,150	1,520

PLAN:

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	1	0	0	1	0	0 / 0
reduce	100.00%	3	0	0	3	0	0 / 0

QUERY1:

Hadoop job_201502212346_0064 on localhost

User: ubuntu
Job Name: PigLatin:PIG1.pig
Job File: hdfs://localhost:3351/home/ubuntu/Workspace/hadoop-1.1.0/hadoop-data/temp/mapred/staging/ubuntu/.staging/job_201502212346_0064/job.xml
Submit Host: ubuntu-VirtualBox
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Tue Feb 24 22:25:49 CET 2015
Finished at: Tue Feb 24 22:26:05 CET 2015
Finished in: 15sec
Job Cleanup: Successful

CPU TIME:

CPU time spent (ms)	440	390	830

PLAN:

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	1	0	0	1	0	0 / 0
reduce	100.00%	1	0	0	1	0	0 / 0

Conclusion:

1) Query6 uses less time of Query1.
2) Query6 also requires less map reduces jobs compared to Query1
   So Query6 has a much better performance than Query1