

Assignment #2  
CSC 555  
Zehong Zhuang

1)

- a.  
Mapper1 reads the input file;  
Mapper1 produce pair values (Year, Month: Full Name);  
Mapper 1 delivers pair values to Reducer1 through framework;  
Reducer perform count function to generate results Year, Month: count(Full Name);
- b.  
Mapper E reads the employee table;  
Mapper A reads the agent table;  
Mapper E produce pair values (e.First, e.Last: EID, Phone);  
Mapper A produce pair values (a. First, a.First: AID);  
Pair values are delivered to Reducer through framework;  
Reducer Perform join function and produce outcome (a.First, a.Last: EID, AID, Phone);
- c.  
Mapper E reads the employee table;  
Mapper A reads the agent table;  
Mapper E produce pair values (e.ID:E.Age);  
Mapper A produce pair values (a.ID:a.Last);  
Pair values are delivered to Reducer through frameworks;  
Reducer performs join function and output results (a.ID: Age, a.Last);  
Output got passed to another Mapper3 through HDFS;  
Mapper3 reads the inputs;  
Mapper3 produces pair values (Age: a.Last);  
Pair values are delivered to Reducer through framework again;  
Reducer perform distinct count function and produce output (Age: count(distinct a.Last);

2)

- a. Copy in another block
- b.  
    i. 1  
    ii. count
- c.  
i. T  
ii. T  
ii. T

3)

- a. 38 minutes  
b. 2 minues  
c. 1 minutes  
d. 1 minutes  
e. a and b could be faster since there is copy of blocks available in the same rack.

4)

a.

11775010 rows

b.

SELECT MIN(barrels08), AVG(barrels08), MAX(barrels08) FROM VehicleData;

```
02_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-02-03 10:04:26,091 Stage-1 map = 0%, reduce = 0%
2019-02-03 10:04:33,646 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.5 sec
2019-02-03 10:04:40,288 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.75
sec
MapReduce Total cumulative CPU time: 2 seconds 750 msec
Ended Job = job_1549187604482_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.75 sec HDFS Read: 1177741
5 HDFS Write: 37 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 750 msec
OK
0.059892 17.820177449476272 47.06831
Time taken: 24.365 seconds, Fetched: 1 row(s)
```

c.

SELECT (barrels08/city08) FROM VehicleData;

```
117740710/10000710
Time taken: 0.182 seconds, Fetched: 34175 row(s)
```

d.

```
627873 2019-02-03 10:06 /user/ec2-user/ThreeColExtract/000000_0]
```

The size of the file is 627873 bits

e.

The size of the file is 598603 bits

```
ec2-user@ip-172-31-36-25 ~$ hadoop fs -ls /user/ec2-user/ThreeColExtract2
Found 1 items
-rwxr-xr-x 1 ec2-user supergroup 598603 2019-02-03 10:22 /user/ec2-user/ThreeColExtract2/000000_
0
```