YCSB / CA 3
Name: Lalit Pathak

Objective:

You have been provided with two sets of output files that were generated by running workloads against a MySql database instance and a MongoDB database instance.

Four workloads were run against each of the database systems. Workload parameters are as follows:

Workload 1	Workload 2	Workload 3	Workload 4
recordcount=12500	recordcount=25000	recordcount=50000	recordcount=100000
operationcount=12500	operationcount=25000	operationcount=50000	operationcount=100000

The following parameters were also specified for all workloads:

readproportion=0.5
updateproportion=0.5
scanproportion=0
insertproportion=0
workload=com.yahoo.ycsb.workloads.CoreWorkload
requestdistribution=zipfian

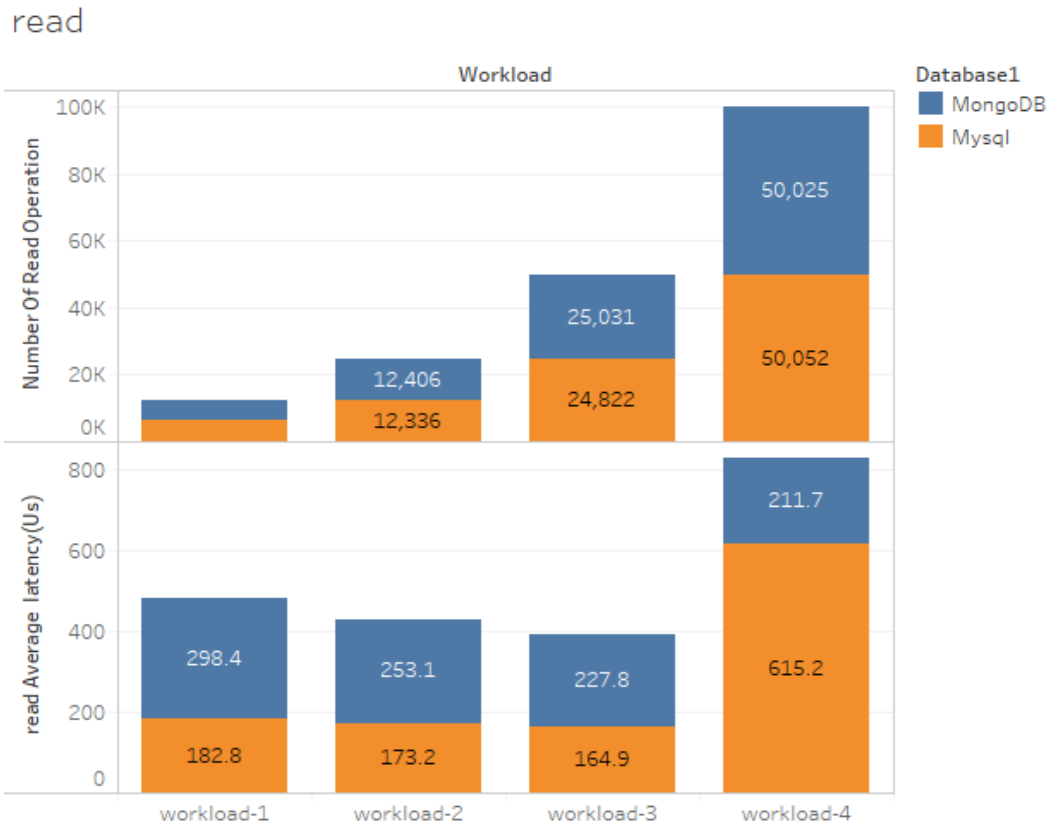
Requirements and Result:

1. **Requirement:** Compare and analysis of performance of the NoSQL databases namely MongoDB database instance and Mysql Database Instance for the given workload.

Question 1:

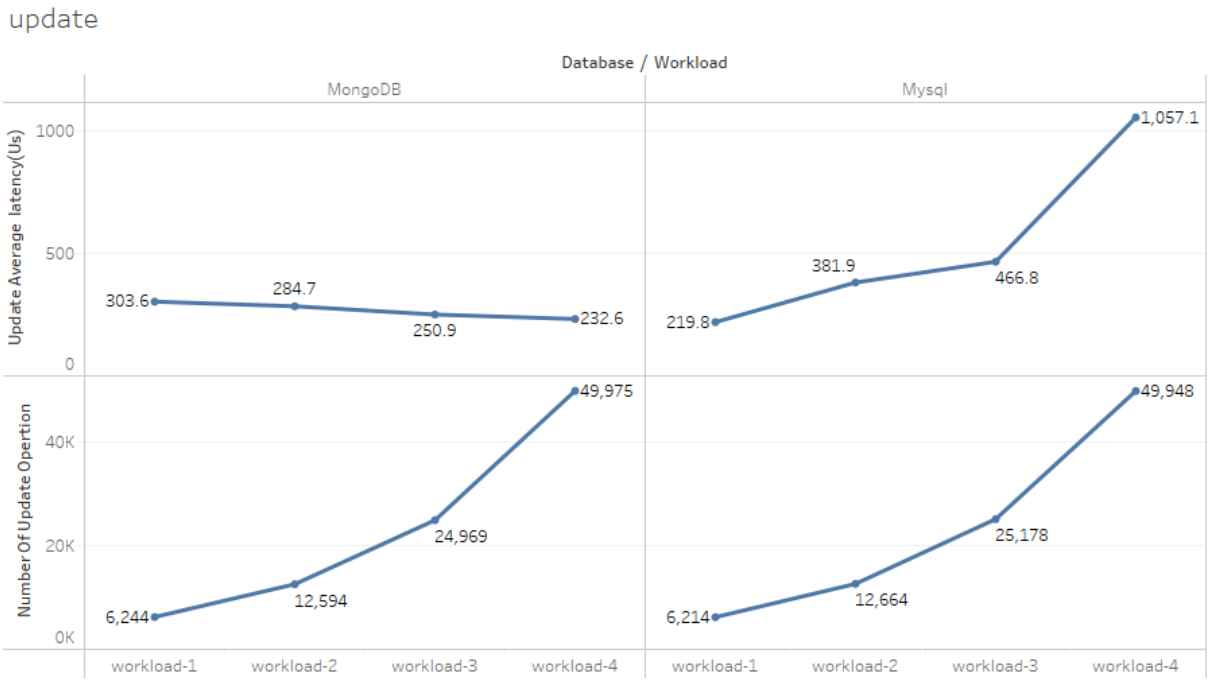
A graphical representation of the recorded Average Latency against the number of record read operations for MySql and MongoDB over all 4 workloads.

Graph shows that MongoDB is more efficient than Mysql for the high workload and mysql performance compare to MongoDB is better when the workload is less.



Question 2:

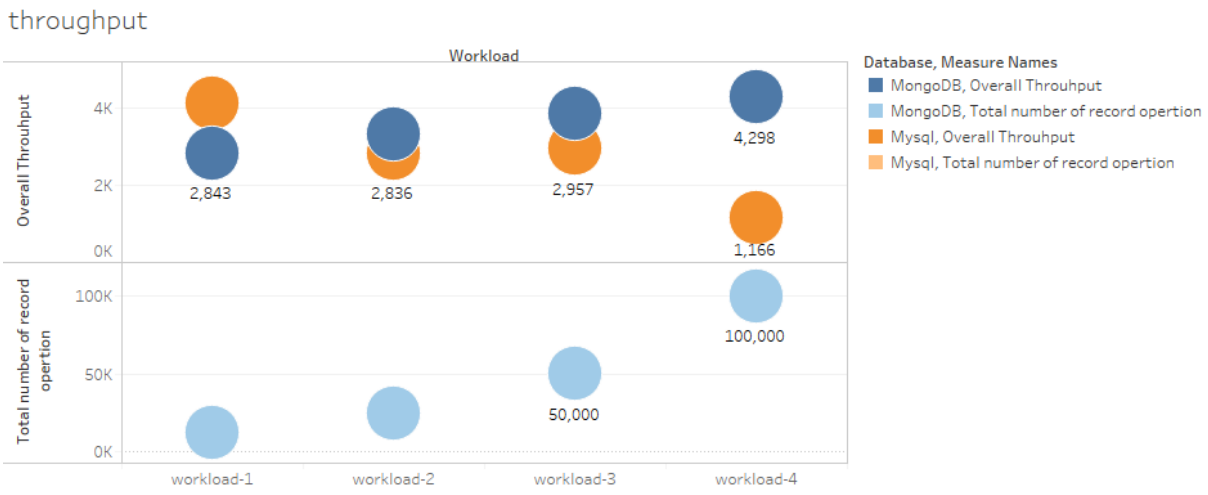
A graphical representation of the recorded Average Latency against the number of record Update operations for MySql and MongoDB over all 4 workloads



The trends of sum of Update Average latency(Us) and sum of Number Of Update Opertion for Workload broken down by Database.

Performance of MongoDB increases as workload of updated operation increases graph also shows that Mysql performance decreases with increase in workload.

Question: 3



Overall Throuhput and Total number of record opertion for each Workload. Color shows details about Database, Overall Throuhput and Total number of record opertion.

Graph clearly shows that performance of MOngoDB is Better compare to MySql
For the given Workload.

Question 4:

Did any of the attempted database operations across any of the workloads fail? Justify your answer by indicating how this can be determined from the output files.

Answer

All the files of YCSB test have same count of Read and return value for all the operations.
Which shows that all operations are successful .So we can say that Non of the workload failed.

```
[READ], Operations, 50052
[READ], AverageLatency(us), 615.2045672500599
[READ], MinLatency(us), 97
[READ], MaxLatency(us), 1603583
[READ], 95thPercentileLatency(us), 293
[READ], 99thPercentileLatency(us), 425
[READ], Return=OK, 50052
[CLEANUP], Operations, 1
[CLEANUP], AverageLatency(us), 1171968.0
[CLEANUP], MinLatency(us), 1171456
[CLEANUP], MaxLatency(us), 1172479
[CLEANUP], 95thPercentileLatency(us), 1172479
[CLEANUP], 99thPercentileLatency(us), 1172479
[UPDATE], Operations, 49948
[UPDATE], AverageLatency(us), 1057.0799030992232
[UPDATE], MinLatency(us), 101
[UPDATE], MaxLatency(us), 2795519
[UPDATE], 95thPercentileLatency(us), 306
[UPDATE], 99thPercentileLatency(us), 484
[UPDATE], Return=OK, 49948
```

Question 5:

What other environment/OS details would be important to know or measure whilst performing tests using the YCSB tool?

Answer

YCSB is important to improve the performance of the database and useful while selecting the databases according to the particular requirement, which includes the data volume as well as time require to execute the task. And also useful to setup the environment. The performance of network and the host system hardware details are also important factors.