Experimental Calculation VIKOR and ANOVA

Ashis Kumar Samanta*, Nabendu Chaki

Department of Computer Science and Engineering, University of Calcutta, Kolkata, West Bengal, India

1. Analysis of Result of Query-1 using VIKOR

Table 1: The Experimental value of Query-1

Query-1	12.5 Lakh	25 Lakh	37.5 Lakh	50 Lakh
SQL Server 2012	0.01	0.01	0.01	0.1
Cassandra model	0.1	0.1	0.1	0.2
MongoDB model	0.2	0.2	0.2	0.2
Best (b_i^+)	0.01	0.01	0.01	0.1
Worst(ω_i^-)	0.2	0.2	0.2	0.2
Weitage(W_i)	0.25	0.25	0.25	.25

Calculating the value from Table 3 and using (eq) -1, we get the normalized matrix of Table-4. S_i and R_j is computed using (eq)-2 and (eq)-3, respectively.

Table 2: VIKOR Ranking Calculation for Query-1

Query-1	12.5 L	25 L	37.5 L	50L	Si	Rj	Qi	Rank
SQL Server 2012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
Cassandra model	0.118	0.118	0.118	0.250	0.604	0.250	0.802	2
MongoDB model	0.250	0.250	0.250	0.250	1.000	0.250	1.000	3

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq-4). The (eq- 5) is used to calculate Q_i (Table-4) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 1.000$, $\psi_+ = 0.000$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 0.802$ and the value of $\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case). Therefore $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R.

*Corresponding author.

Email addresses: aksdba@caluniv.ac.in (Ashis Kumar Samanta), nabendu@ieee.org (Nabendu Chaki)

March 26, 2022

In case of query-1, the SQL server shows much better performances than that of the Cassandra and MongoDB.

At the time of execution of query-2 it is to be noticed that the query execution did not support Cassandra(on NoSQL Manager for Cassandra) for data sets 3.75 million records and 5 millions records. Therefore, two set of experiments results have been considered here.

Query-2 can be analyzed the said data either for all four sets of data without considering the Cassandra (Case-1) or go for test for all three data models only for 1.25 million and 2.5 million number of records (Case-2), respectively.

The VIKOR analysis of rest of the five queries have done in the same way and due to avoid the same repetition process for rest of the five queries, the analysis process is shown in the gitHub( ). The final outcome of the result of all the six queries is shown in Table 5 in consolidated form.

2. Analysis of Result of Query-2 using VIKOR

The MongoDB data model shows the best response compared to Cassandra and SQL Server in this query observation. It is to be noticed that the query execution did not support Cassandra(on NoSQL Manager for Cassandra) for data sets 37.5 Lakh and 50 Lakh.

Table 3: The Experimental value of Query-2

Query-2	12.5 Lakh	25 Lakh	37.5 Lakh	50 Lakh
SQL Server 2012	20	41	62	79
Cassandra	0.157	2.39	Not Support	Not Support
MongoDB	0.109	0.114	0.158	0.17
Best (b_i^+)	0.019	0.114	0.158	0.17
Worst(ω_i^-)	20	41	62	79
Weitage(W_i)	.25	.25	.25	.25

In this case, Cassandra did not support (NS) executing the query for 37.5L and 5 million records. Therefore, we can analyze the said data either for all four sets of data without considering the Cassandra or go for test for all three data models only for 1.25 million and 2.5 million number of records, respectively.

2.1. Case-1: Without considering Cassandra:

Calculating the value from Table 5 and using (eq) -1, we get the normalized matrix of Table-6. S_i and R_j is computed using (eq)-2 and (eq)-3, respectively

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using The (eq)-4. (eq)-5 is used to calculate Q_i (Table-4) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 1.000$, $\psi_+ = 0.000$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 1.000$ and the value of $\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case). Therefore $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R.

Table 4: VIKOR Ranking Calculation for Query-2 for Case1

Query-2	12.5 L	25 L	37.5 L	50L	Si	Rj	Qi	Rank
SQL Server 2012	0.250	0.250	0.250	0.250	1.000	0.250	1	2
MongoDB model	0.000	0.000	0.000	0.000	0.000	0.000	0	1

2.2. Case-2: Without considering the case 37.5 lakh and 50 Lakh rows:

Calculating the value from Table 5 and using (eq) -1, we get the normalized matrix of Table-7. S_i and R_j is computed using (eq)-2 and (eq)-3, respectively

Table 5: VIKOR Ranking Calculation for Query-2 for Case2

Query-2	12.5 L	25 L	Si	Rj	Qi	Rank
SQL Server 2012	0.250	0.250	0.500	0.250	1	3
Cassandra model	0.001	0.014	0.015	0.014	0.515	2
MongoDB model	0.000	0.000	0.000	0.000	0	1

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq)-4. The (eq)- 5 is used to calculate Q_i (Table-4) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 1.000$, $\psi_+ = 0.500$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 0.515$ and the value of $\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case). Therefore, $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R. In both the cases MongoDB shows the best results.

3. Analysis of Result of Query-3 using VIKOR

Table 6: The Experimental value of Query-3

Query-3	12.5 Lakh	25 Lakh	37.5 Lakh	50 Lakh
SQL Server 2012	1.0	1.0	1.0	1.0
Cassandra	1.72	1.94	2.15	2.31
MongoDB	1.203	1.793	1.914	1.94
Best (b_i^+)	1	1	1	1
Worst(ω_i^-)	1.72	1.94	2.15	2.31
Weitage(Wi)	.25	.25	.25	.25

Calculating the value from Table-8 and (eq), we get the normalized matrix of Table-9. S_i and R_j is calculated using (eq)-2 and(eq)-3, respectively.

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq)-4. The (eq)-5 is used to calculate Q_i (Table-9) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 1.000$, $\psi_+ = 0.000$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 0.752$ and the value of

Table 7: VIKOR Ranking Calculation for Query-3

Query-3	12.5 L	25 L	37.5 L	50L	Si	Rj	Qi	Rank
SQL Server 2012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
Cassandra model	0.250	0.250	0.250	0.250	1.000	0.250	1.000	3
MongoDB model	0.070	0.211	0.199	0.179	0.659	0.211	0.752	2

$\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case).Therefore, $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R.

In the case of query-3, the SQL server shows much better performances than that of MongoDB and Cassandra.

4. Analysis of Query-4 using VIKOR

Table 8: The Experimental value of Query-4

Query-4	12.5 Lakh	25 Lakh	37.5 Lakh	50 Lakh
SQL Server 2012	15.000	30.000	47.000	63.000
Cassandra	1.2	1.9	2.6	3.5
MongoDB	0.128	0.140	0.158	0.178
Best (b_i^+)	.128	0.140	.158	.178
Worst(ω_i^-)	15.0	30	47	63.0
Weitage(W_i)	.25	.25	.25	.25

Calculating the value from Table-10 and (eq), we get the normalized matrix of Table-11. S_i and R_j are calculated using (eq)-2 and (eq)-3, respectively.

Table 9: VIKOR Ranking Calculation for Query-4

Query-4	12.5 L	25 L	37.5 L	50L	Si	Rj	Qi	Rank
SQL Server 2012	0.250	0.250	0.250	0.250	1.000	0.250	1.000	2
Cassandra model	0.018	0.015	0.013	0.018	0.059	0.018	1.000	2
MongoDB model	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq)-4. The (eq)- 5 is used to calculate Q_i (Table-11) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 1.000$, $\psi_+ = 0.000$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 1.000$ and the value of $\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case). Therefore $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R.

In query-4, the MongoDB shows the best performances than the Cassandra and SQL server. In practical experiment the Cassandra shows a much better performance than that of SQL server, but using the VIKOR tools both the Cassandra and SQL server hold the same ranking position 2.

5. Analysis of Query-5 using VIKOR

Table 10: The Experimental value of Query-5

Query-5	12.5 Lakh	25 Lakh	37.5 Lakh	50 Lakh
SQL Server 2012	1.000	1	2	3
Cassandra	3.67	5.14	10.9	17.2
MongoDB	.200	.200	.600	.900
Best ( )	0.2	0.2	0.600	0.900
Worst(	3.67	5.14	10.900	17.200
Weitage(Wi)	.25	.25	.25	.25

Calculating the value from Table-12 and (eq), we get the normalized matrix of Table-13. S_i and R_j are calculated using (eq)-2 and (eq)-3, respectively.

Table 11: VIKOR Ranking Calculation for Query-5

Query-5	12.5 L	25 L	37.5 L	50L	Si	Rj	Qi	Rank
SQL Server 2012	0.058	0.040	0.034	0.032	0.164	0.058	0.223	2
Cassandra model	0.250	0.250	0.250	0.250	0.765	0.250	1.000	3
MongoDB model	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq)-4. The (eq)- 5 is used to calculate Q_i (Table-13) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 0.765$, $\psi_+ = 0.000$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 0.765$ and the value of $\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case). Therefore $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R. In case Query-5, shows that the MongoDB data model shows much best performances than the SQL server and Cassandra data models.

6. Analysis of Query-6 using VIKOR

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq)-4. (eq)- 5 is used to calculate Q_i (Table-15) and generate the Rank.

The calculation of ξ_+, ξ_-, ψ_+ *and* ψ_- are done using (eq)-4. The (eq)- 5 is used to calculate Q_i (Table-15) and generate the Rank. $\xi_+ = 0.000$, $\xi_- = 1.000$, $\psi_+ = 0.000$ *and* $\psi_- = 0.250$The value of $C_1 = Q(r^2) - Q(r^1) = 0.341$ and the value of

Table 12: The Experimental value of Query-6

Query-6	12.5 Lakh	25 Lakh	37.5 Lakh	50 Lakh
SQL Server 2012	2.000	3.000	3.000	3.000
Cassandra	3.62	4.16	7.2	8.9
MongoDB	12.677	14.440	15.721	16.727
Best ( )	2.0	3.000	3.000	3.000
Worst(	12.677	14.440	15.721	16.727
Weitage(Wi)	.25	.25	.25	.25

Table 13: VIKOR Ranking Calculation for Query-6

Query-6	12.5 L	25 L	37.5 L	50L	Si	Rj	Qi	Rank
SQL Server 2012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
Cassandra model	0.038	0.025	0.083	0.107	0.253	0.107	0.341	2
MongoDB model	0.250	0.250	0.250	0.250	1.000	0.250	1.000	3

$\delta_c = \frac{1}{(j-1)} = 0.5$ (as j=3 in this case). Therefore $Q(r^2) - Q(r^1) \geq \delta_c$ is satisfied. i.e C_1 is not satisfied with the compromised solutions. C_2 is also satisfied as a^1 is the best Rank in S or/and R. The rank is considered as C_2 is satisfied. In the case of query-6, the SQL server shows much better performances than that of the Cassandra and MongoDB.

7. Analysis of Result Using Two way ANOVA

The implementation of an analysis of two way ANOVA on all the six queries stated in Table 1, the general hypothesis is considered as below

For row-wise consideration

H_{r0}: No impact due to data model;

H_{r1}: Has an impact due to data model;

For column-wise consideration

H_{c0}: Has no impact due to data volume;

H_{c1}: Has an impact due to data volume;

7.1. Analysis of Query-1 using ANOVA

Therefore the calculated value of F-ratio for row-wise (using Table 16)is F_{cal}= 37.13187 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} > F^{0.001}_{\{2,6\}}$, hence hypothesis H_{r0} is rejected for this query. The data model has an effect on the executed query.

The test value (using Table 16) of F-ratio for column-wise is $F_{cal} = 3.967033$, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hence hypothesis H_{c0} is accepted for this query. The data volume does not affect the insertion of a single record.

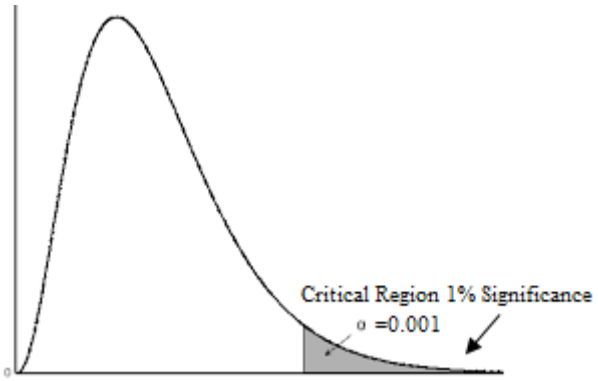

Figure 1: ANOVA 1% Significane Critical Ragion

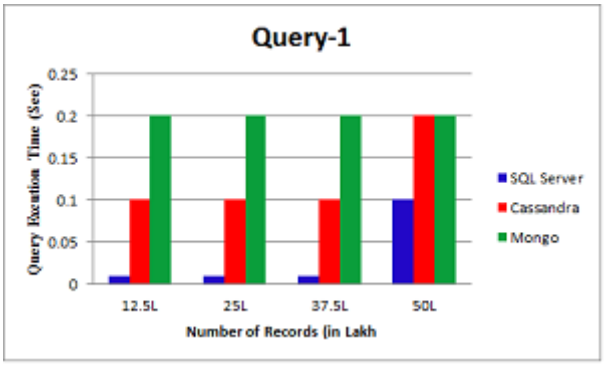

Figure 2: Execution of Query-1

Table 14: Calculation for ANOVA Test for Query-1

Description Q1	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	0.056317	2	0.028158	37.13187
Within columns	0.009025	3	0.003008	3.967033
Error	0.00455	6	0.000758	0.00455
Total	0.069892	11		

7.2. Analysis of Query-2 using ANOVA

7.2.1. Case-1: Without considering Cassandra:

Therefore the calculated value of F-ratio for row-wise (using Table 17)is F_{cal}= 15.52480 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} > F^{0.001}_{\{2,6\}}$, Therefore, , hence hypothesis H_{r0} is rejected for this query Query-2, case-1

The test value (using Table 17) of F-ratio for column-wise is $F_{cal} = 1.00460$, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hence hypothesis H_{c0} is accepted for this query-2 case-1.

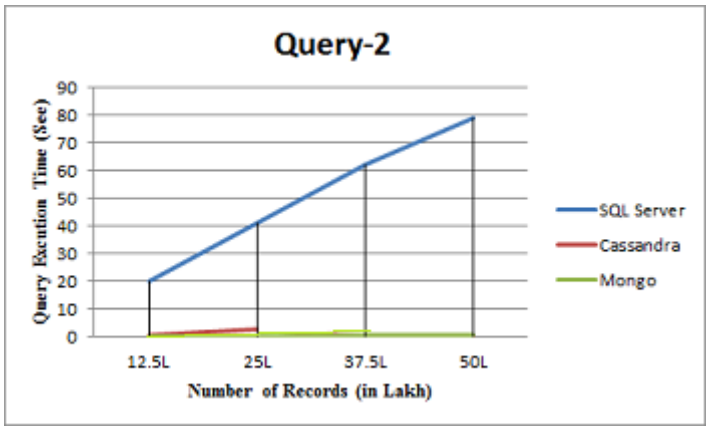

Figure 3: Execution of Query-2

Table 15: Calculation for ANOVA Test for Query-2, Case-1

Description Q2, Case-1	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	5072.712	1	5072.712	15.52480
Within columns	984.7559	3	328.252	1.00460
Error	980.2469	3	326.749	
Total	7037.715	7		

7.2.2. Case-2: Without considering the case 37.5 lakh and 50 Lakh rows:

Table 16: Calculation for ANOVA Test for Query-2, Case-2

Description Q2, Case-2	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	1186	2	592.9998	8.91780
Within columns	90.00077	1	90.00077	1.35347
Error	132.9924	2	66.49619	
Total	1408.993	5		

Therefore the calculated value of F-ratio for row-wise (using Table 18)is F_{cal}= 8.91780 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} < F^{0.001}_{\{2,6\}}$, hence hypothesis H_{r0} is accepted for this Query-2, case-2.

The test value (using Table 18) of F-ratio for column-wise is $F_{cal} = 1.35347$, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hypothesis H_{c0} is accepted for this query Query-2, case-2.

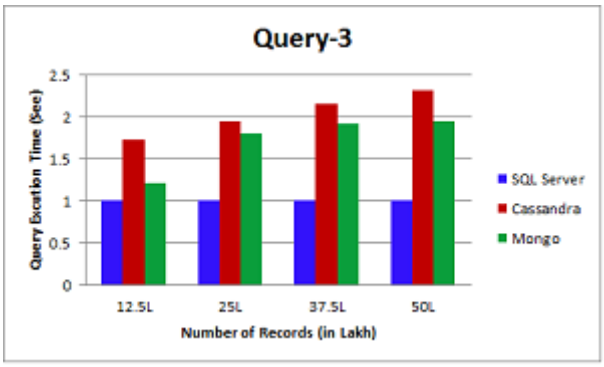

Figure 4:

Table 17: Calculation for ANOVA Test for Query-3

Description Q3	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	2.225817	2	1.112908	31.61182
Within columns	0.344196	3	0.114732	3.258931
Error	0.211233	6	0.035205	0.211233
Total	2.781246	11		

7.3. Analysis of Query-3 using ANOVA

Therefore the calculated value of F-ratio for row-wise (using Table 19)is F_{cal}= 31.61182 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} > F^{0.001}_{\{2,6\}}$, hence hypothesis H_{r0} is rejected for this query. The data model has an effect of reading the records with specific conditions.

The test value (using Table 19) of F-ratio for column-wise is $F_{cal} = 3.967033$, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hence hypothesis H_{c0} is accepted for this query. The data volume does not affect reading the records with specific conditions.

7.4. Analysis of Query-4 using ANOVA

Table 18: Calculation for ANOVA Test for Query- 4

Description Q4	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	3764.138	2	1882.069	13.69295
Within columns	474.9633	3	158.3211	1.151862
Error	824.6882	6	137.448	
Total	5063.789	11		

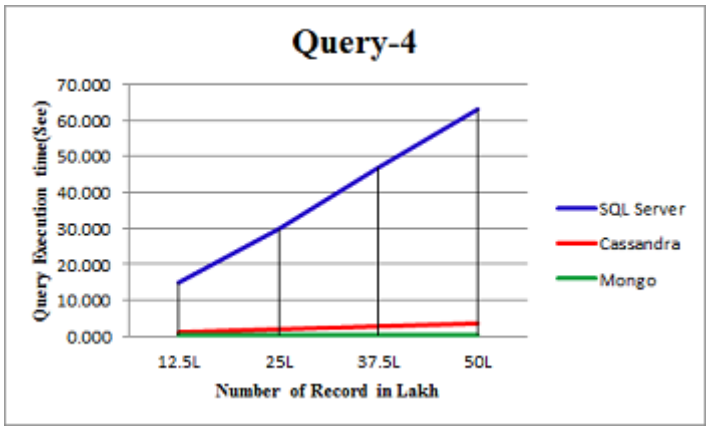

Figure 5: Execution of Query-4

Therefore the calculated value of F-ratio for row-wise (using Table 20)is F_{cal}= 13.69295 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} > F^{0.001}_{\{2,6\}}$, hence hypothesis H_{r0} is rejected for this query-4.

The test value (using Table 20) of F-ratio for column-wise is F_{cal} = 1.151862, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hence hypothesis H_{c0} is accepted for this query-4.

7.5. Analysis of Query-5 using ANOVA

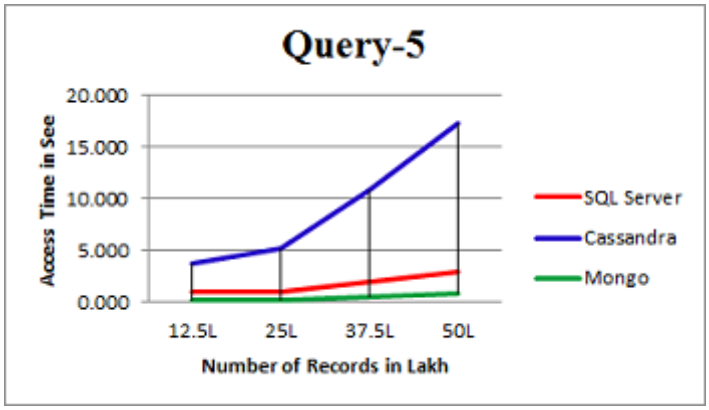

Figure 6: Execution of Query-5

Therefore the calculated value of F-ratio for row-wise (using Table 21)is F_{cal}= 8.728966 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} < F^{0.001}_{\{2,6\}}$, hence hypothesis H_{r0} is accepted for this query, and counting total rows does not affect the data model.

The test value (using Table 21) of F-ratio for column-wise is F_{cal} = 1.808257, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hence hypothesis H_{c0} is accepted for this query. The data volume does not affect the counting of total rows of the data model.

Table 19: Calculation for ANOVA Test for Query-5

Description Q5	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	178.8599	2	89.42993	8.728966
Within columns	55.57783	3	18.52594	1.808257
Error	61.47115	6	10.24519	61.47115
Total	295.9088	11		

7.6. Analysis of Query-6 using ANOVA

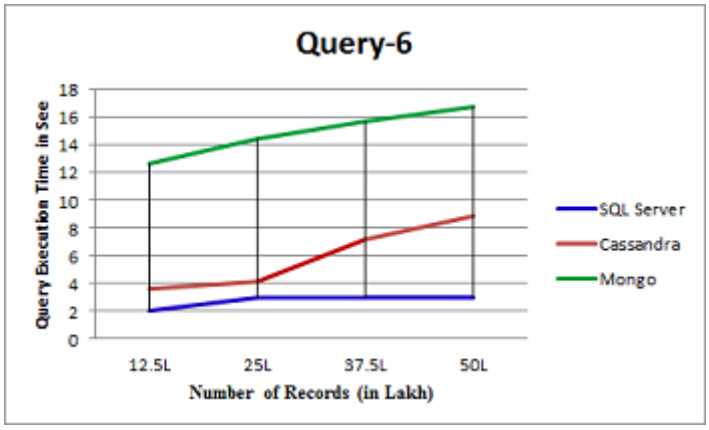

Figure 7: Execution of Query-6

Table 20: Calculation for ANOVA Test for Query-6

Description Q6	Sum of Square	Degrees of freedom	Mean square	F Ratio
within rows	316.4894	2	158.2447	120.4139
Within columns	20.92636	3	6.975453	5.307867
Error	7.885035	6	1.314173	
Total	345.3008	11		

Therefore the calculated value of F-ratio for row-wise (using Table 22)is F_{cal}= 120.4139 and the critical value for 1% significance is $F^{0.001}_{\{2,6\}} = 10.93$ Therefore, $F_{cal} > F^{0.001}_{\{2,6\}}$, hence hypothesis H_{r0} is rejected for this query query-6. The deletion of same number of records has an effect on data model.

The test value (using Table 22) of F-ratio for column-wise is $F_{cal} = 5.307867$, and the critical value for 1% significance is $F^{0.001}_{\{3,6\}} = 9.78$ Therefore, $F_{cal} < F^{0.001}_{\{3,6\}}$, hence hypothesis H_{c0} is accepted for this query Q6. The deletion of the same number of records affects data volume of the data model.