

Parallel Programming & Performance on Clusters
Using the C Message Passing Interface
Performance Graph

Rachit Magon & Rishi Raj Sahu,
M.E. – Software Systems,
Bits Pilani, Pilani.



MPI – Partitioned Model – Select Query

We queried the cluster set up for the above mentioned datasets. The query we sent as user input was

```
SELECT PERSON.PERSONID,PERSON.FIRSTNAME FROM PERSON WHERE  
PERSON.FIRSTNAME=Rachit
```

The database we created had a set of 50 names which were added through pseudo random number generator; hence the probability of the name being “Rachit” was 1 in every 50 names.

Curves with fixed table size

Table Size – 10K	
Cluster Size	Time Taken in Seconds
2	0.00
4	0.01
6	0.02
8	0.01

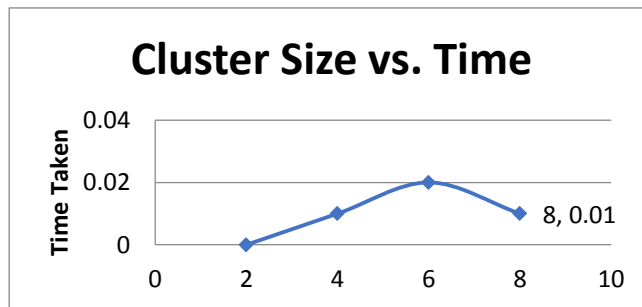


Table Size – 100K	
Cluster Size	Time Taken in Seconds
2	0.03
4	0.02
6	0.01
8	0.01

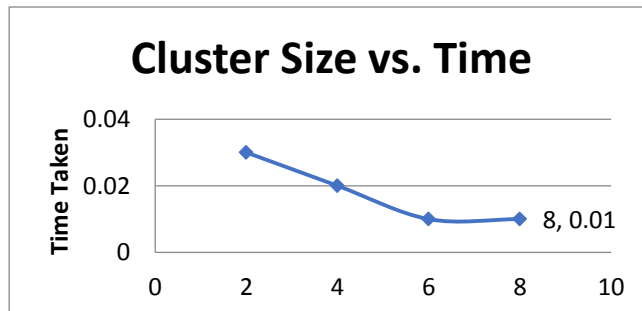


Table Size – 1000K	
Cluster Size	Time Taken in Seconds
2	0.32
4	0.24
6	0.18
8	0.17

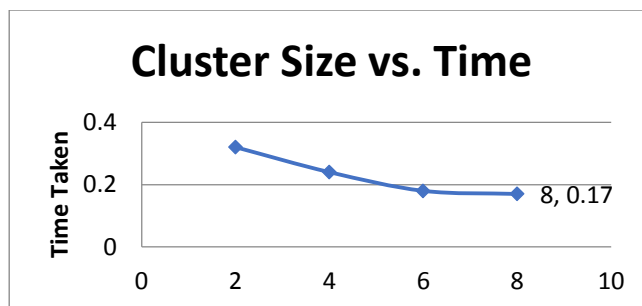


Table Size – 6000K	
Cluster Size	Time Taken in Seconds
2	1.73
4	1.76
6	1.2
8	0.9

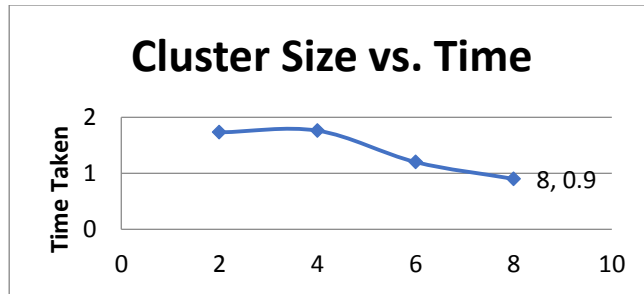
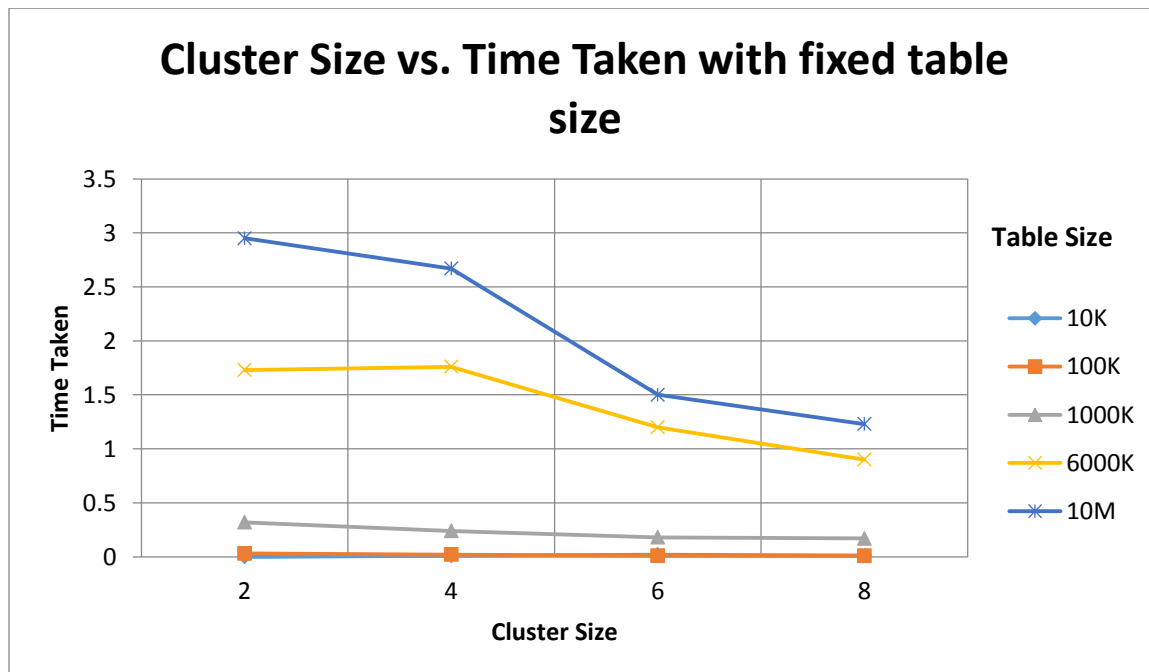
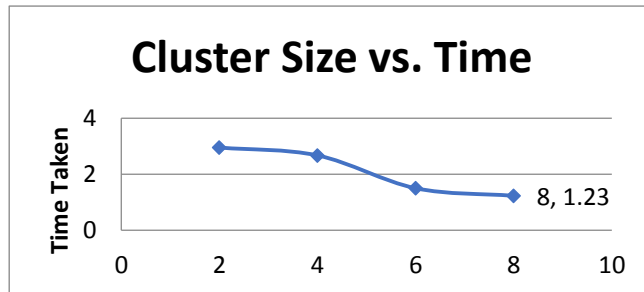
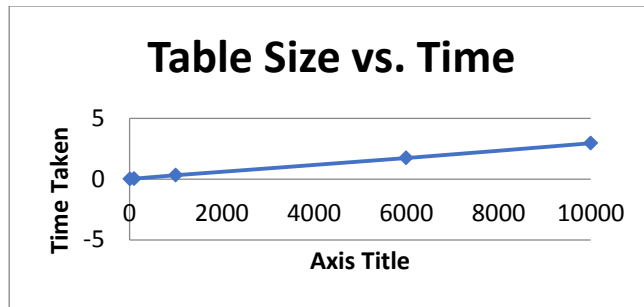


Table Size – 10M	
Cluster Size	Time Taken in Seconds
2	2.95
4	2.67
6	1.5
8	1.23

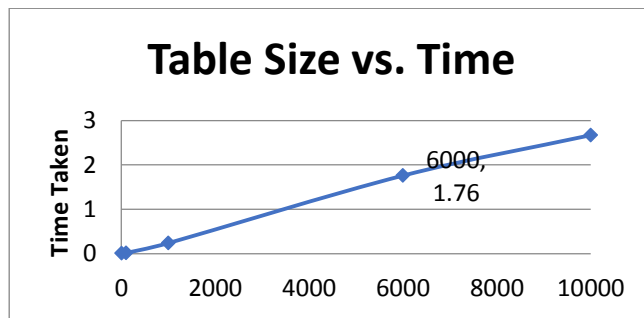


Curves with fixed cluster size for Partitioned Model SELECT Query

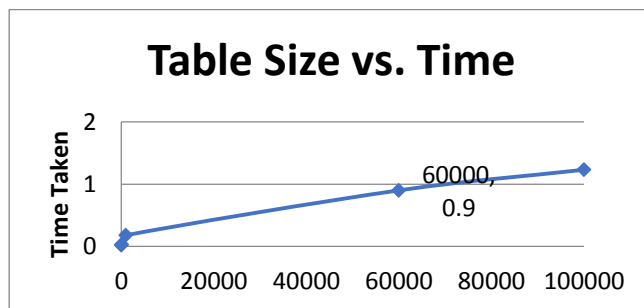
Cluster Size – 2	
Table Size	Time Taken in Seconds
10K	0
100K	0.03
1000K	0.32
6000K	1.73
10M	2.95



Cluster Size – 4	
Table Size	Time Taken in Seconds
10K	0.01
100K	0.02
1000K	0.24
6000K	1.76
10M	2.67

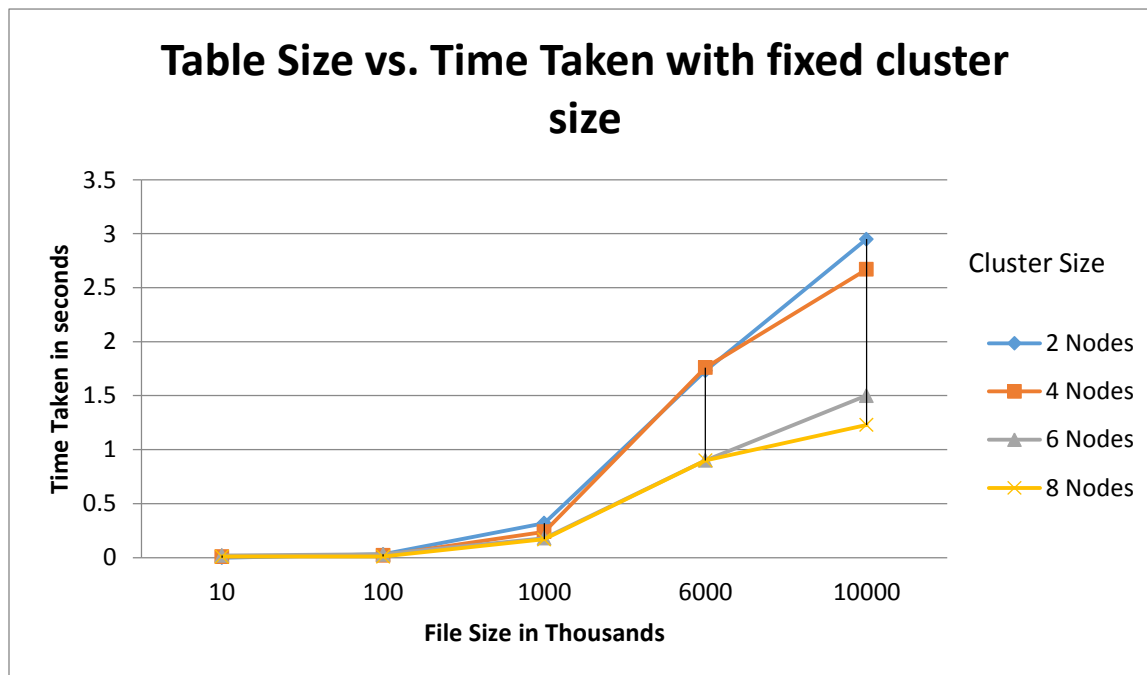
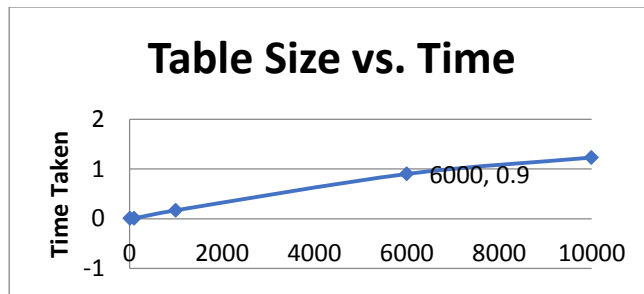


Cluster Size – 6	
Table Size	Time Taken in Seconds
10K	0.02
100K	0.03
1000K	0.18
6000K	0.9
10M	1.5



Cluster Size – 8	
------------------	--

Table Size	Time Taken in Seconds
10K	0.01
100K	0.01
1000K	0.17
6000K	0.9
10M	1.23



Partitioned Model – Select Query

MPI – Partitioned Model – Join Query

The query given for join was:

```
SELECT * FROM PERSON,PARKING WHERE PERSON.PERSONID=PARKING.PERSONID
```

Curves with fixed table size

Table Size – 10K	
Cluster Size	Time Taken in Seconds
2	0.02
4	0.01
6	0.01
8	0.00

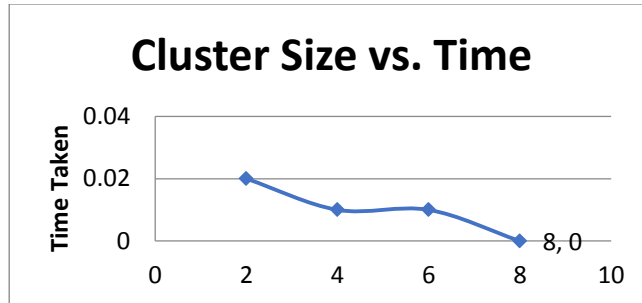


Table Size – 100K	
Cluster Size	Time Taken in Seconds
2	0.17
4	0.09
6	0.06
8	0.05

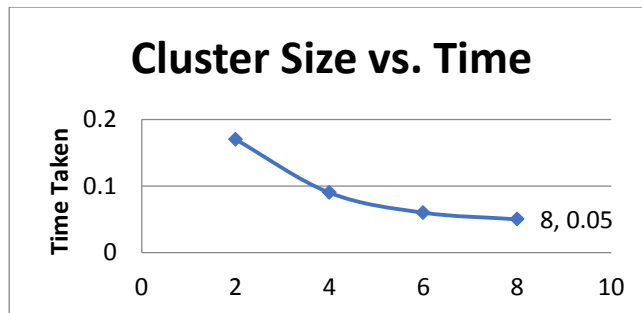


Table Size – 1000K	
Cluster Size	Time Taken in Seconds
2	1.87
4	1.56
6	0.54
8	0.74

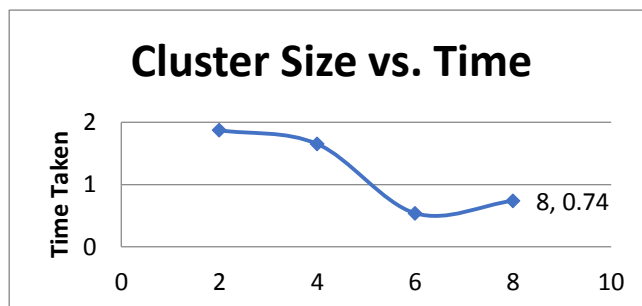


Table Size – 6000K	
Cluster Size	Time Taken in Seconds
2	10.5
4	9.74
6	5.79
8	4.42

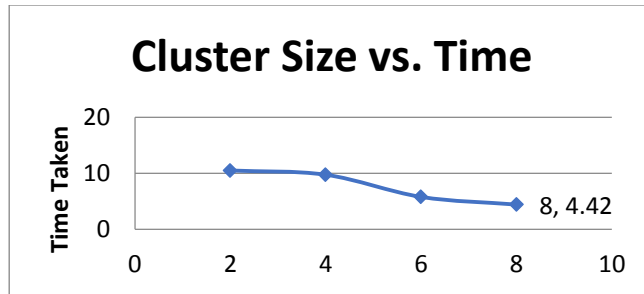
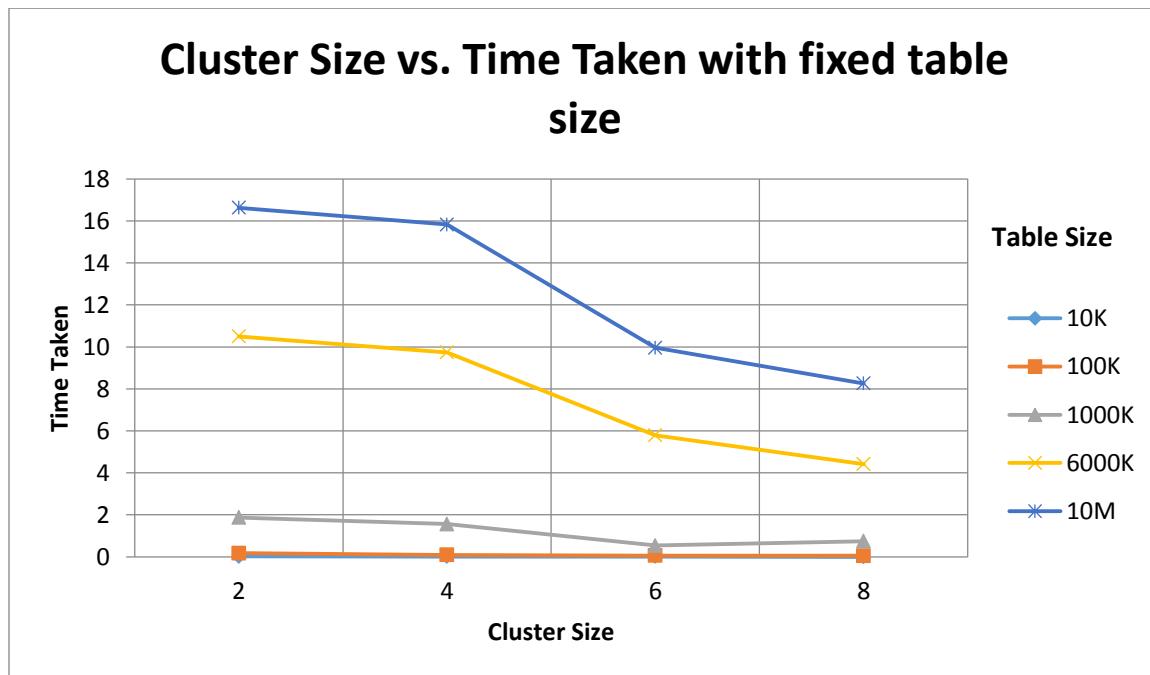
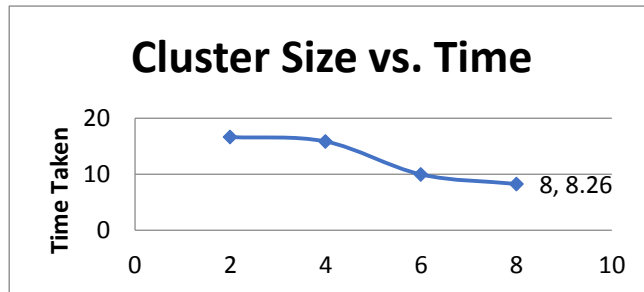
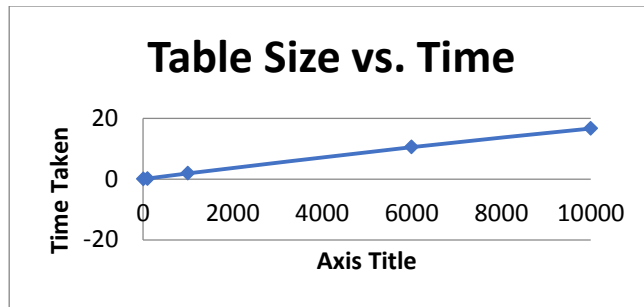


Table Size – 10M	
Cluster Size	Time Taken in Seconds
2	16.63
4	15.83
6	9.96
8	8.26

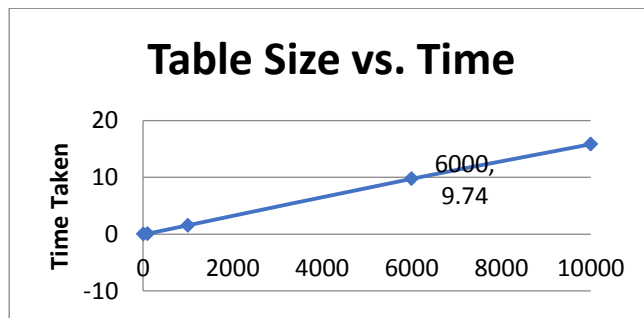


Curves with fixed cluster size for Partitioned Model JOIN Query

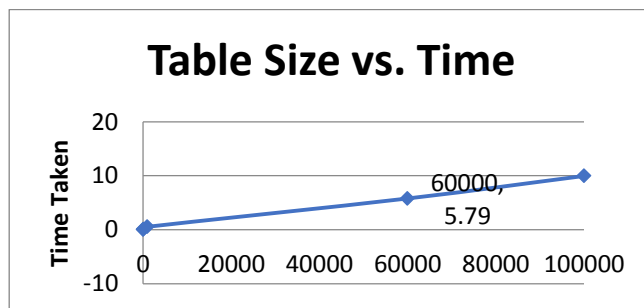
Cluster Size – 2	
Table Size	Time Taken in Seconds
10K	0.02
100K	0.17
1000K	01.87
6000K	10.5
10M	16.63



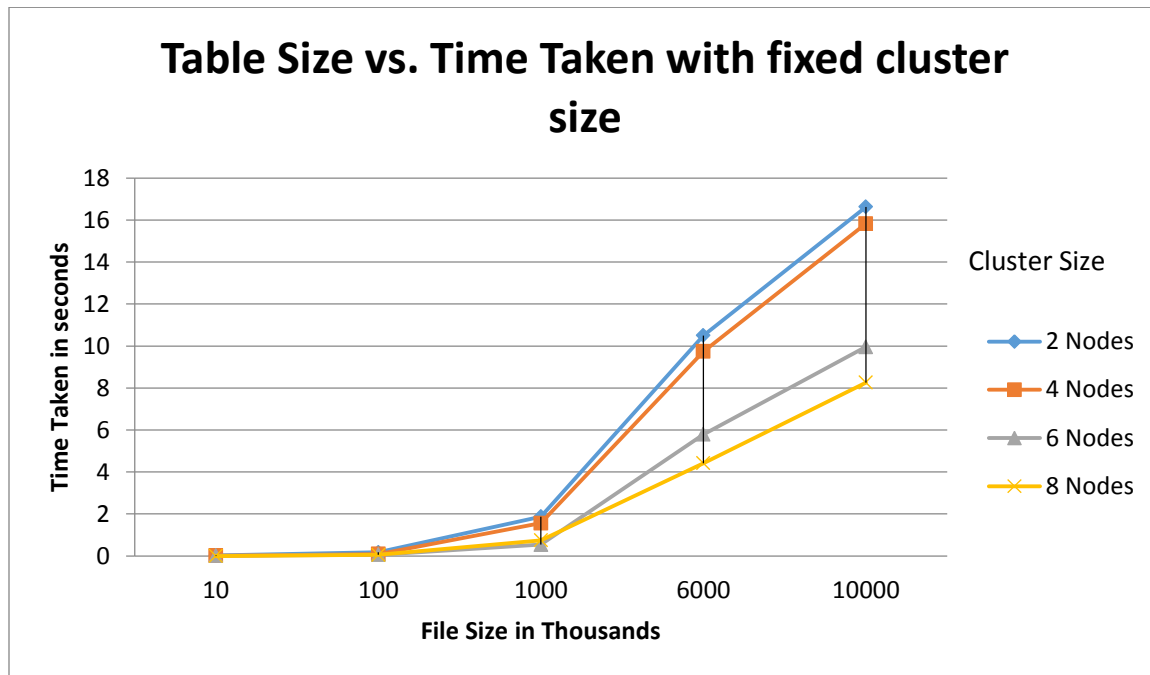
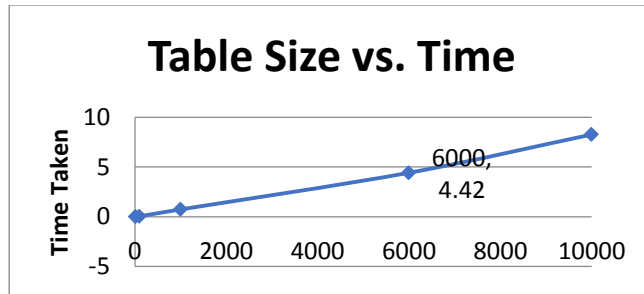
Cluster Size – 4	
Table Size	Time Taken in Seconds
10K	0.01
100K	0.09
1000K	1.56
6000K	9.74
10M	15.83



Cluster Size – 6	
Table Size	Time Taken in Seconds
10K	0.01
100K	0.06
1000K	0.54
6000K	5.79
10M	9.96



Cluster Size – 8	
Table Size	Time Taken in Seconds
10K	0.0
100K	0.05
1000K	0.74
6000K	4.42
10M	8.26



Partitioned Model – JOIN Query

MPI – Replicated Model

Replicated Model – Select Query

We queried the cluster set up for the previously mentioned datasets. The query we sent as user input was

```
SELECT PERSON.PERSONID,PERSON.FIRSTNAME FROM PERSON WHERE  
PERSON.FIRSTNAME=Rachit
```

Curves with fixed table size

Table Size – 10K		
Cluster Size	Replication Factor	Time Taken in Seconds
2	0	0.01
4	3	0.01
6	4	0.01
7	6	0.02

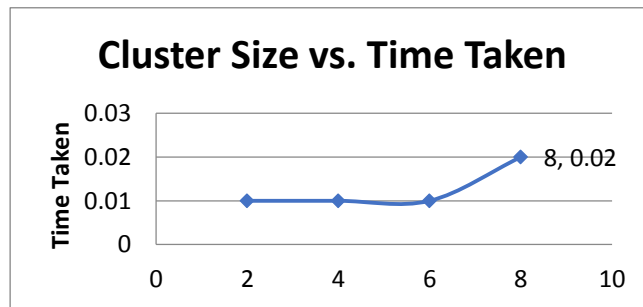


Table Size – 100K		
Cluster Size	Replication Factor	Time Taken in Seconds
2	0	0.08
4	3	0.05
6	4	0.04
7	6	0.04

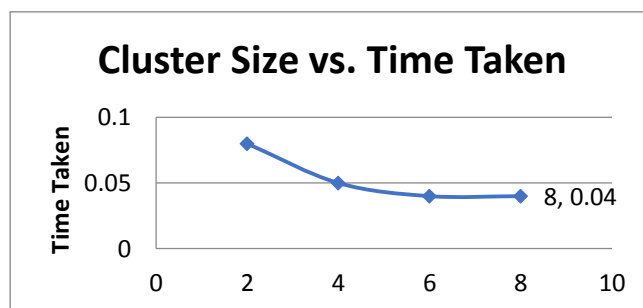


Table Size – 1000K		
Cluster Size	Replication Factor	Time Taken in Seconds
2	0	0.78
4	3	0.51
6	4	0.50
7	6	0.47

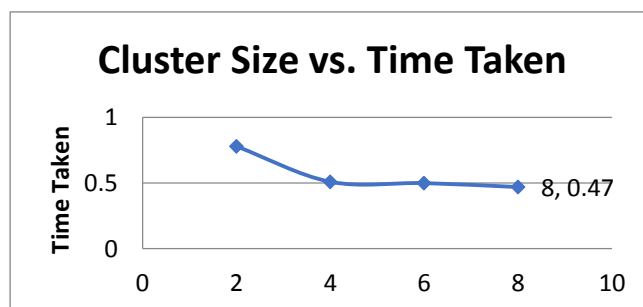


Table Size – 6000K		
Cluster Size	Replication Factor	Time Taken in Seconds
2	0	4.72
4	3	3.16
6	4	3.09
7	6	3.06

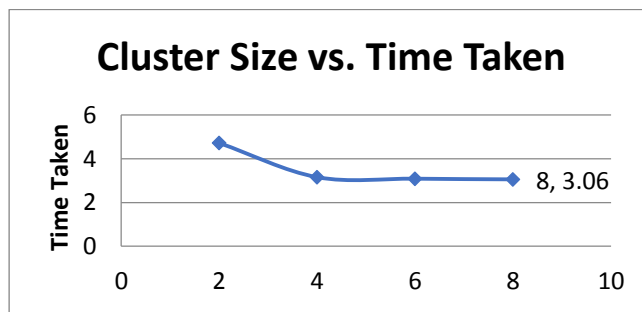
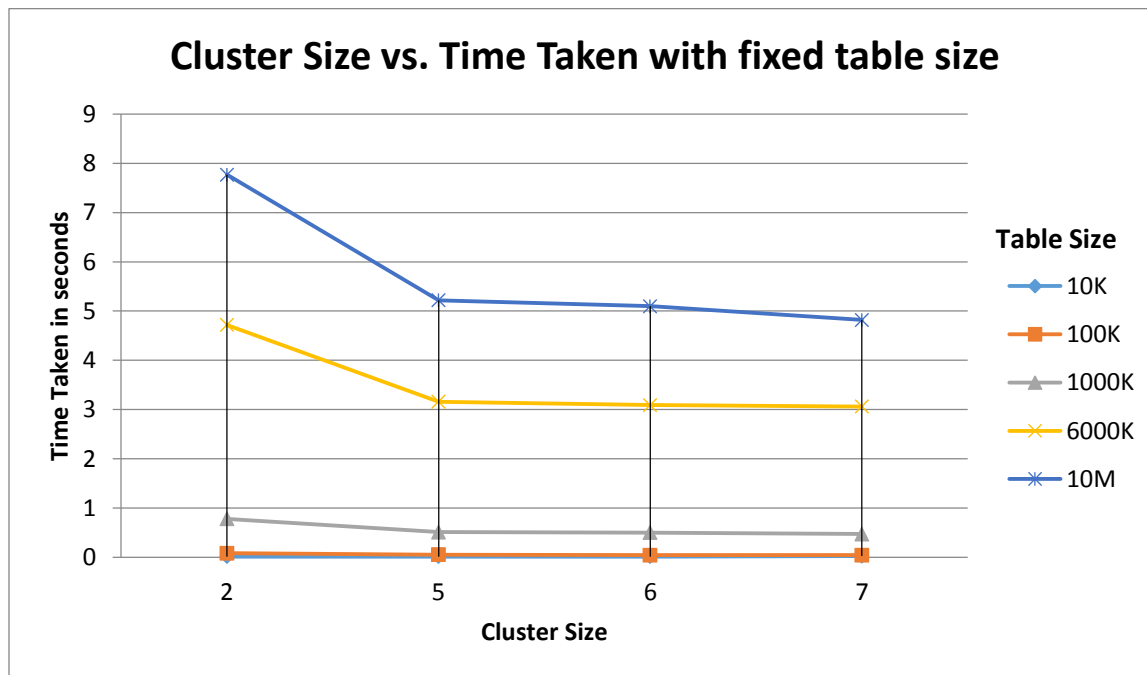
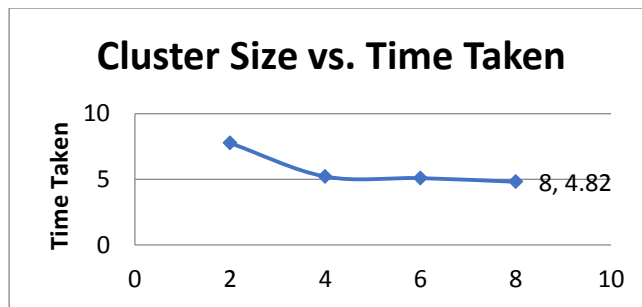
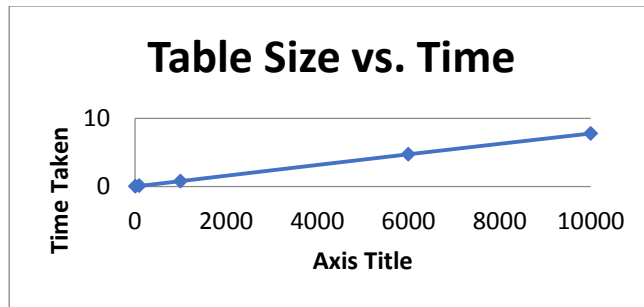


Table Size – 10M		
Cluster Size	Replication Factor	Time Taken in Seconds
2	0	7.77
4	3	5.22
6	4	5.1
7	6	4.82

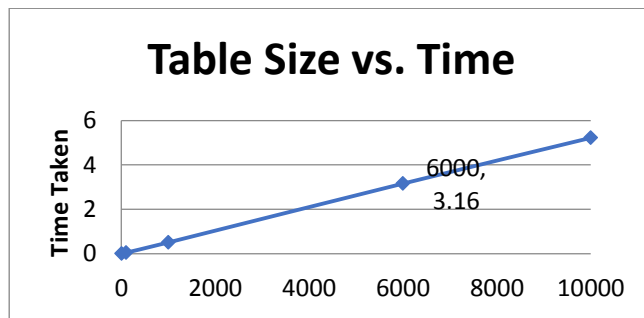


Curves with fixed cluster size for Replicated Model SELECT Query

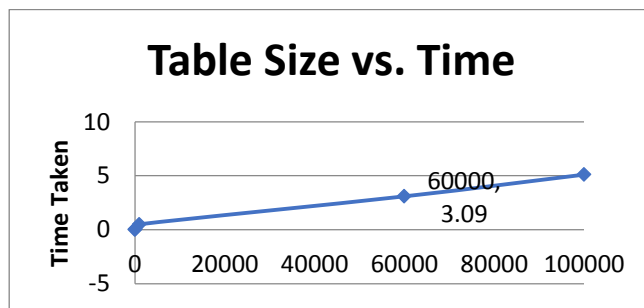
Cluster Size – 2	
Table Size	Time Taken in Seconds
10K	0.01
100K	0.08
1000K	0.78
6000K	4.72
10M	7.77



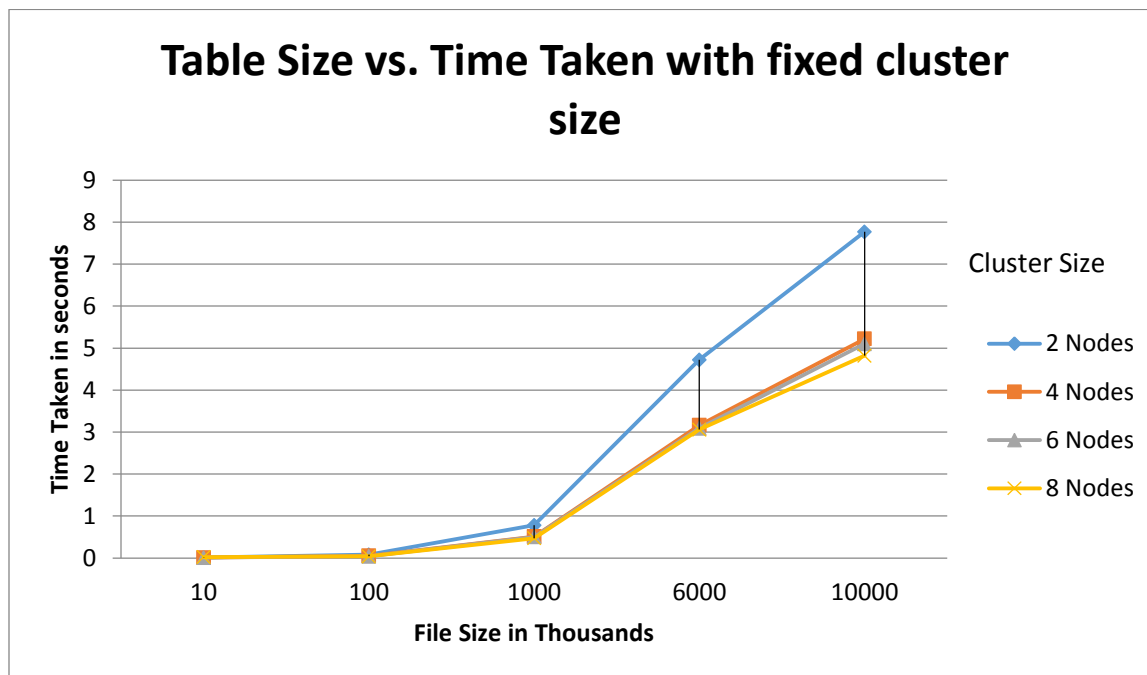
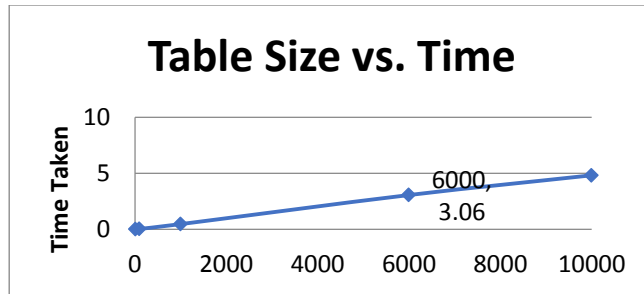
Cluster Size – 4	
Table Size	Time Taken in Seconds
10K	0.01
100K	0.05
1000K	0.51
6000K	3.16
10M	5.22



Cluster Size – 6	
Table Size	Time Taken in Seconds
10K	0.01
100K	0.04
1000K	0.5
6000K	3.09
10M	5.1

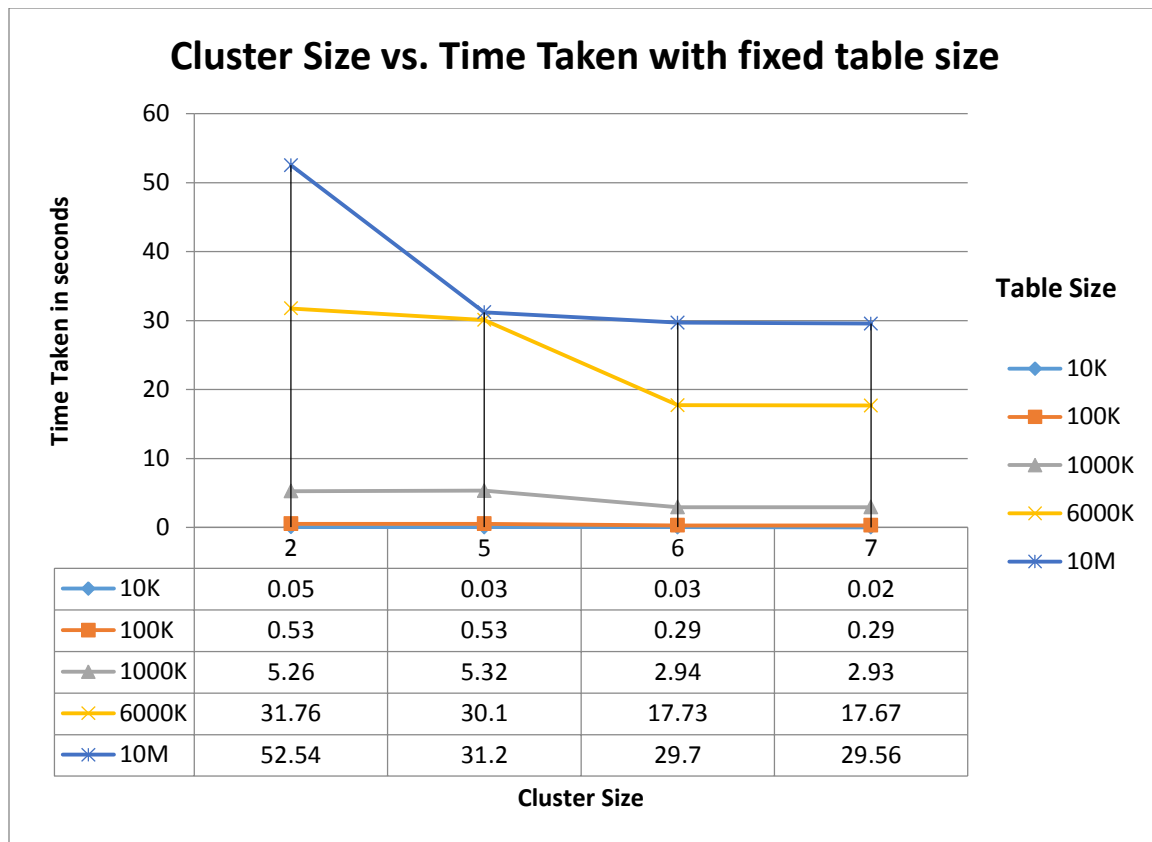


Cluster Size – 8	
Table Size	Time Taken in Seconds
10K	0.02
100K	0.04
1000K	0.47
6000K	3.06
10M	4.82



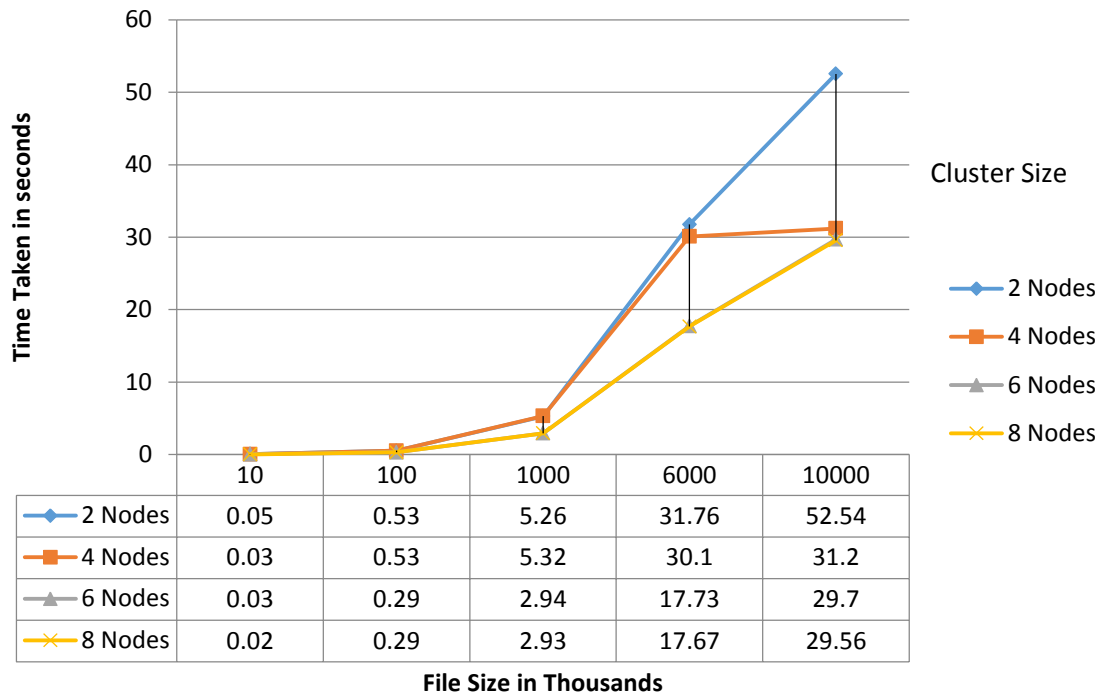
Replicated Model – SELECT Query

MPI – Replicated Model – JOIN Query



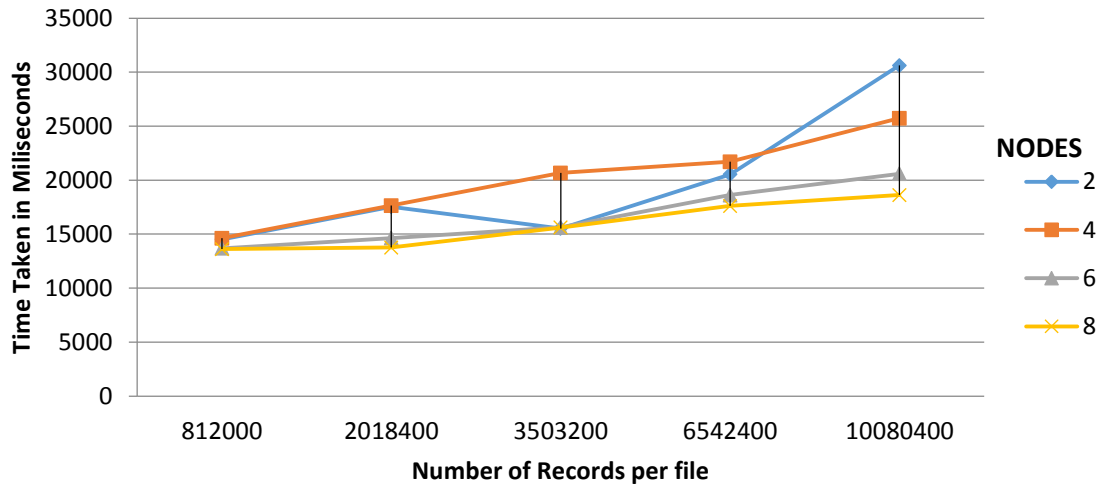
MPI – Replicated Model – JOIN Query

Table Size vs. Time Taken with fixed cluster size

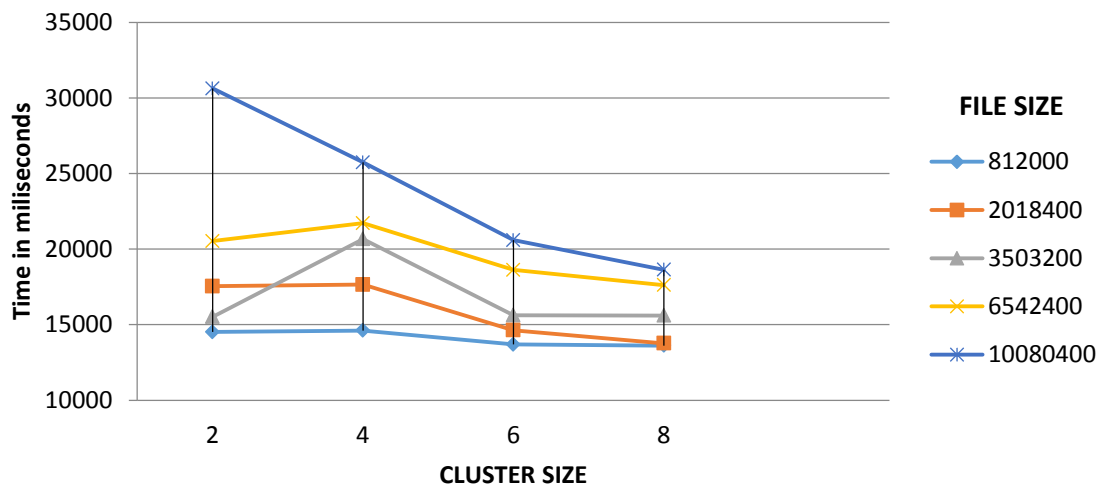


HADOOP CLUSTER

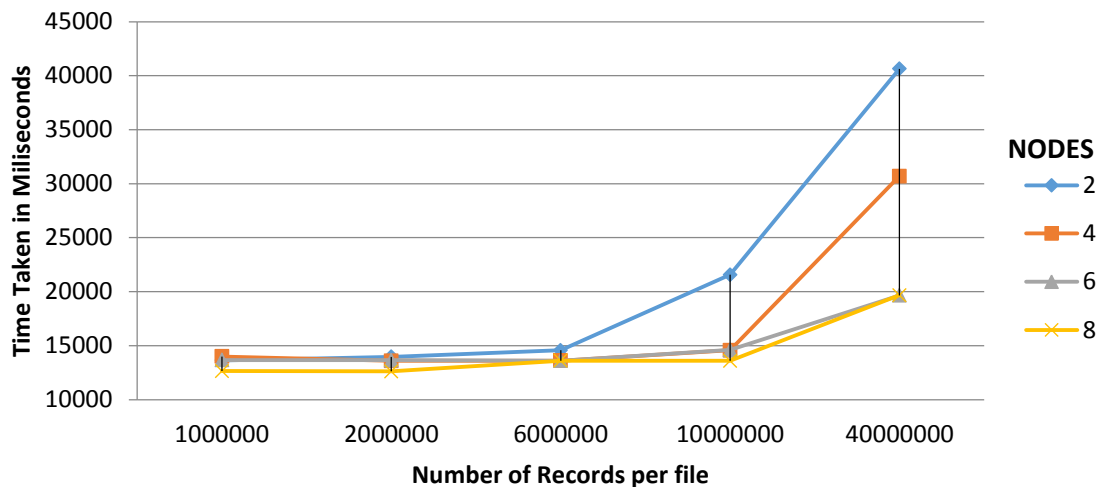
SELECT QUERY, KEEPING NUMBER OF NODES CONSTANT, VARYING THE FILE SIZE



SELECT QUERY, KEEPING THE FILE SIZE CONSTANT, VARYING THE CLUSTER SIZE



JOIN QUERY, KEEPING NUMBER OF NODES CONSTANT, VARYING THE FILE SIZE



JOIN QUERY, KEEPING FILE SIZE CONSTANT, VARYING THE NUMBER OF NODES

