RESEARCH ARTICLE OPEN ACCESS

Memory Consuption and CPU Time Using Indexing Technique on Large Database

Garima Munjal[1], Chandna[2]
Research Scholar[1], Assistant Professor[2]
Department of Computer Science and Engineering
Jan Nayak Ch. Devilal College of Engineering Sirsa
Haryana- India

ABSTRACT

Biomedical databases are different from traditional data warehouse as it contains non transactional information like bimolecular (protein, RNA, DNA, lipid, carbohydrate). A large amount of data is accumulated which is needed to be accessed in least amount of time when complex queries are executed in present biomedical data warehouses. When the Biological data is compare with the other proteins/structures data, searching takes the lot times for check the any similarities, relation between biological data. There is need of index for access of data in less time. In this paper we are going to Compare and analyze the different indexing techniques on Large Dataset. Different indexing techniques are clustered Index, Non-Clustered Index and Full Text Index.

Keywords:- Indexing, clustered Index, Non-Clustered Index, Full Text Index

I. INTRODUCTION

THe main operations on biomedical databases includes searching of protein and matching of certain patterns of data which is very complex and tedious task to handle with because finding a particular pattern of RNA/DNA in warehouse can even take days, so pattern matching or searching becomes difficult in the systems as it takes considerable amount of time and memory consumption depending upon the queries which are complex and iterative. Consider scenario of forensic science which deals with crime issues which are increasing day by day, a large amount of data is accumulated which is needed to be accessed in least amount of time when complex queries are executed in present biomedical data warehouses which is a major challenge. The ability to answer these complex queries efficiently depends upon a major factor 'Index'. Indexing of a data warehouse is complex and if there are few indexes, the data loads quickly but the query response is slows. If there are many indexes, the data loads slowly and there will be more storage requirements but the query response is good. This is true with large tables and complex queries that involve table joins. Considerable amount of time is taken by the query to be processed is more due to large size of both tables and attributes. Index’s space and time play an important role in choosing an indexing technique in data warehouse. Usually if the space used by an index is large then the results are achieved in short time and on the other side, if the space used by the index space is small then the results are achieved in greater amount of time. So there is a tradeoff between the time consumed and the space used by a particular index. Important factors which are need to be improved:

1. Response time
2. Searching time/Scan time
3. Memory Usage

Different indexing techniques are clustered Index, on-Clustered Index and Full Text Index.

Memory Consumption: Indexes which have been built on character based columns consumed much memory.

CPU Time: CPU time is the sum of Compilation time and Query execution time. There are number of indexes like clustered index, non clustered index and Full text index which having different execution time and corresponding, the different CPU time. Indexes have large CPU time depends on Index.

Working of Indexes: There are different types of indexes which can be implemented in different scenarios and before this, there is need to analyze the working, and memory consumption, etc. When the indexes are created, then the database engines stores and sorted the records in B-Tree Structure which reduces the memory consumption and disk reads during retrieve of data. The B-Tree structure is having the leaf nodes, root nodes and intermediate nodes. The bottom level of nodes is the leaf nodes, top level of nodes are root nodes and levels between the bottom and top nodes are intermediate levels. It contains the data pages and index pages. The index pages stores the records and index rows and different ID's for the specific indexes along with pointers.

II. PROBLEM FORMULATION

Forensic Science deals with crimes issues and it increasing day by day. Due to this, the historical data and current data are growing and fast access of information is required which is not time efficient in present data warehouses. There is need of efficient indexing techniques should build on data warehouse especially DNA Protein datasets. Data warehouse contains data is large in size and increase memory usage, time consumption. In searching over the networks, it requires more bandwidth for processing of data, retrieve of data.

III. METHODOLOGY

There are different indexing techniques has been proposed and need to indentify the best indexing technique in terms of time and memory. The existing techniques can be good for timely searching the records of DNA/RNA structure, protein structure and it is observed that existing techniques takes much time. The proposed work will be implemented in the SQL Server 2008 by generate the dataset of variation records.

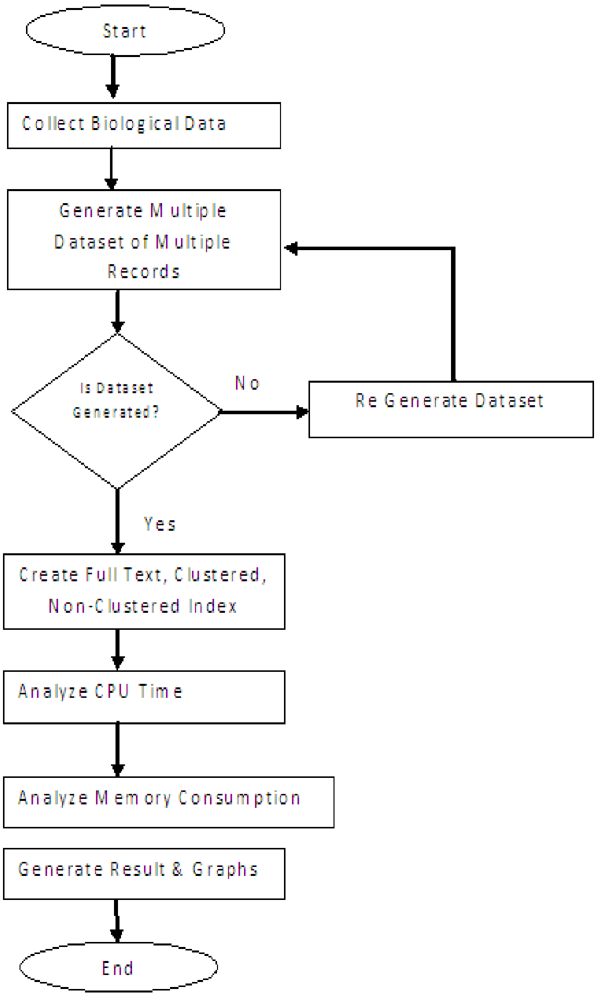
There are many database banks that provide the public database. There are some resources from National Center for Biotechnology Information (NCBI) which gives the data for proteins, Structure information and sequences information.

1. Collection of Protein Data.
2. Generate Dataset of different collection of records.
3. Apply different indexing technique clustered Index, Non-Clustered Index and

Full Text Index.

4. Perform SQL Queries for Calculate the CPU Time and Memory Consumption.
5. Generate the Graphs and Analyze the Results.

Example: GenBank contains the datasets of proteins and structure information.


Start
Collect Biological Data
Generate Multiple Dataset of Multiple Records
Is Dataset Generated?
No
Re Generate Dataset
Yes
Create Full Text, Clustered, Non-Clustered Index
Analyze CPU Time
Analyze Memory Consumption
Generate Result & Graphs
End

IV. OBJECTIVES

When the Biological data is compare with the other proteins/structures data, searching takes the lot times for check the any similarities, relation between biological data. There is need of index for access of data in less time.

a) To Study of the Indexing Techniques
b) To Compare and analyze the different indexing techniques on Large Dataset
c) To analyze the performance of Indexing Techniques
d) To calculate the CPU Time, Elapsed Time.
e) To analyze the Memory Consumption of Indexing Techniques.
f) To predict the efficient Indexing Technique

V. CONCLUSION

The biomedical data warehouse plays a crucial role in order to perform important operations like protein identification, matching but challenge is large size of biomedical data warehouse as it contains RNA/DNA which encapsulate string of large length. Different indexing techniques has been used and analyzed using different types of queries on different size of datasets in biomedical data warehouse in order to perform operation in efficient manner.

ACKNOWLEDGMENT

Special thank you goes to those who contributed to this paper: Mrs Chandna for her valuable comments, sharing her knowledge and for their keen interest, expert guidance, valuable suggestions and persistent encouragement

REFERENCES

[1]Aho A. V., J. E. Hopcroft, and J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley: An Imprint of Addison WesleyLongman, Inc., Reading Massachusetts, 1999.

[2] J. Kratika, I. Ljubic, and D. Tosic, "A genetic algorithm for the index selection problem", Applications of Evolutionary Computing (EvoWorkshops 03), Essex, UK; LNCS, Vol. 2611, Springer, Heidelberg, 2003, pp. 281-291.

[3] Khalid Jaber, Rosni Abdullah and Nur'Aini Abdul Rashid (2009) "Indexing Protein Sequence/Structure Databases Using Decision Tree: A Preliminary Study", Information Technology (ITSim), 2010 International Symposium, IEEE Computer Society Conference.

[4] M. Golfarelli, S. Rizzi, and E. Saltarelli, "Index selection for data warehousing", 4th International Workshop on Design and Management of Data Warehouses (DMDW 02), Toronto, Canada; CEUR Workshop Proceedings, Vol. 58, CEURWS. org, Aachen, 2002, pp. 33-42.

[5] P. O'Neil and D. Quass, "Improved Query Performance with Variant Indexes", SIGMOD, 1997

[6] R. Kimball, L. Reeves, M. Ross and W. Thornthwaite, "The Data Warehouse Lifecycle Toolkit : Expert Methods for Designing, Developing, and Deploying Data Warehouses", John Wiley & Sons, Aug. 1998

[7] Sankalap Arora, Priyanka Anand, Kirandeep Singh," An Efficient Indexing Technique Used In Telemedicine Data Warehouse", (2010)

[8] Safavian, S.R., Landgrebe (1991), "D.A survey of decision tree classifier methodology", Systems, Man and Cybernetics, IEEE Transactions, Page 660-674.

[9] Shawana Jamil and Rashda Ibrahim (2009) "Performance analysis of Indexing Techniques in Data Warehousing", Emerging Technologies, 2009. ICET 2009 International Conference, Page 57-61

[10] Sreerama K. Murthy (1998)," Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey", Association for Computing Machinery

[11] Thabasu Kannan, Dr.K.Iyakutti (2009) "A Clustered Indexing Method for Optimizing the Query for Biological Databases", GCC Conference & Exhibition, 2009 5th IEEE, page 1-6.