Chapter-6

CONCLUSION AND FUTURE WORK

6.1 CHAPTER OVERVIEW

This chapter covers the conclusion of research done and evaluation and analysis of the results generated by the tool TIER with performance analysis of concurrent transaction execution in multiple heterogeneous database using TPC-H benchmark of benchmark factory. It also covers future work and further research to be done in the area of performance improvement of concurrent transactions in multiple heterogeneous databases.

6.2 CONCLUSION

Index is a physical structure which occupies memory [71]. When index is created on the table, it generates index key for each record. This index key is referred when user writes a query on the table which contains index. Because of index, the records are retrieved very fast, if the field on which index is generated referred in the search condition. On the other hand, index needs to be maintained or to be upgraded when new records are inserted, deleted or updated. DML operations reduce efficiency of indexes in order to keep index up to date. Therefore, before index is created on the table, various parameters need to be considered such as total workload, fields which are used frequently in search conditions, read and write operations, CPU cost, variation in workload, no. of concurrent user requests, etc. It becomes more challenging when the physical database is distributed on multiple heterogeneous databases which are connected through network for the applications accessing data from all these databases. Heterogeneity and distribution of data require stronger index to access data faster with reduced cost. Therefore, the tool Table Index Evaluator and Recommender (TIER) is proposed which evaluate current indexes and recommend improved indexes for the inputted workload. Before proposing the tool, experience survey is done on the
basis of questionnaire prepared. The questionnaire is distributed to the people of IT industry. The samples are selected on the basis of age, type of job profile, expertise in DBA, etc. The analysis is done for the collected data to check feasibility of proposed model.

The tool TIER works as an assistant to the DBA. The final decision to implement recommended indexes depend on the DBA. The heuristic used in the implementation of TIER is based on the Parse Matrix which is generated from the mathematical model based on the inputted set of queries. The heuristic is based on the parse matrix, which count frequency of fields referred in WHERE and HAVING clauses and, calculates the ratio from total no. of records in the table and total unique values in the fields. From the collected statistics, TIER recommends three types of indexes – Unique index, Bitmap index and Composite Index. To assess the performance of recommended indexes, the replica of original tables are created with the recommended indexes. Pair of same transactions which are based on tables with recommended indexes and tables with original indexes are executed to record different parameter values such as response time, CPU cost, IO cost and bytes accessed. This task is performed by “Performance Comparer” of TIER. It will generate graphical comparison of pair of similar transactions. DBA can take decision on this basis. To provide more clarity, the Apriori algorithm is also implemented to generated association rules between fields which are frequently used in WHERE and HAVING clauses of the inputted queries. Combining the association rules and recommended indexes, DBA will decide whether or not to implement recommended indexes.

The analysis and evaluation of results generated from TIER is done in great details in the thesis. The comparison of TIER is done with SQLAdvisor of Oracle. SQL Advisor suggests indexes for inputted set of queries, but it doesn’t recommend indexes for the queries which refer remote tables. Therefore, TIER is useful to find indexes for remote tables. Though the “Performance Comparer” module of TIER gives graphical comparison of four parameter values – response time, IO cost, CPU cost and Bytes for inputted pair of same transactions, the benchmark factory is also used to assess effectiveness of indexes recommended from TIER. The TPC-H benchmark of Benchmark Factory is used to load huge volume of data into four different databases. Scalability test is performed on various jobs (different sets of transactions) with different no. of concurrent user loads and up to five iterations. Each job is executed up to 3 runs (times). After each run, the average response time data is generated for each transactions for different user loads by the TPC-H
benchmark. These data are compared to display comparison graphs. The evaluation done by the “Performance Comparer” module and scalability tests agrees on the point that TIER suggests better indexes which help to improve concurrent transaction execution in multiple heterogeneous distributed database system.

### 6.3 FUTURE WORK

The process of index creation demands highly experienced person who should have detailed knowledge of physical database designing. TIER provides help to these persons to take decision regarding index implementation. Hence, TIER should be able to recommend index for any type of inputted query. Presently, TIER takes simple queries with one level of nesting and simple join queries as input. However, user can input “n” number of queries. To strengthen TIER, it could be expanded in terms of variety of inputted queries, more types of index suggestions, runtime enabling or disabling indexes, automated interface to connect all RDBMS which participate in distributed database, etc. In future, TIER could be enhanced to make it recommend function-based indexes. The inputted workload in TIER contains only SELECT statements. No data manipulation commands are considered in the workload. To find out CPU and IO overhead, the DML statements can also be given as inputs in TIER. The distributed tables from remote databases are accessed through Oracle Gateway service in Oracle. The heuristic which could make user access data in reverse fashion could also be developed. The technique to disable index temporarily during DML operations, could also be incorporated to reduce waiting time between user request and acknowledgement. Moreover, TIER could also be extended to get recommendations for other physical structures like materialized view and partitioning method.

### 6.4 CHAPTER SUMMARY

The research work done is discussed in detail in the chapter. The mathematical model of proposed prototype, implementation of this model, functionalities provided by TIER and outputs generated in the form of proposed indexes are important aspects of TIER. The comparison of recommended indexes is done using database standard benchmark named TPC-H. Data in GBs is loaded from TPC-H benchmark to different databases. The evaluation and analysis of recommended indexes is also described in details with scope of future research work.