CHAPTER-7

CONCLUSION AND FUTURE SCOPE OF WORK

This chapter presents a brief summary of the main findings and contributions of the thesis. An attempt has been made to draw some broad conclusions and to indicate some further directions of the work.

7.1 Conclusion

The allocation of data among the different sites over the network plays an important role in performance of the distributed database system. The main aim of the data allocation problem is to minimize communication cost, increase reliability and availability of the data, maximize performance and minimize storage cost. Therefore, data allocation in distributed database is a combinatorial optimization problem with an objective function that has to be minimized or maximized with respect to a set of constraints.

The data allocation problem in distributed database systems has been investigated for static as well as dynamic environment. To calculate the cost of allocation of data fragments, a data allocation model has been described for static distributed database environment. Two separate cost functions have been defined one for non-replicated static allocation and another for replicated static allocation. The major contributions of the proposed work are outlined as follows:

1. Two frameworks have been proposed for both non-replicated and replicated static allocation of data. The first proposed framework is based on the biogeography-based optimization (BBO). Biogeography-based optimization (BBO) is a newly developed population-based evolutionary technique. The second proposed framework is based on the simplified biogeography-based optimization (SBBO). Simplified biogeography-based optimization (SBBO) is a modified version of the biogeography-based optimization. The performance of these new algorithms has been compared with genetic algorithm for data allocation in distributed database systems. All the three algorithms are tested on the same data set for each experiment and the results are obtained after running
each algorithm 20 times independently for each experiment. Comparison is done on the basis of quality of the solution provided by all the algorithms and average running time taken by the algorithms. From the experimental results, it is observed that the performance of SBBO is much better than BBO and GA for non-replicated as well as for replicated allocation of data. SBBO is providing quality solution within shorter period of time as compare to BBO and GA. In almost all the experiments, the proposed SBBO algorithm is providing fragments allocation schema having cost of allocation less than that of BBO and GA. BBO is also providing good solutions as compare to GA when the numbers of fragments are relatively less in numbers. The performance of BBO degrades as the number of fragments increases. Therefore BBO can be used as an alternative technique for data allocation when the number of fragments and the number of sites are relatively less in number. Overall it has been observed that the SBBO algorithm outperformed BBO and GA in terms of solution quality and computational speed.

2. A new heuristic named Threshold and Time Constraint Algorithm (TTCA) has been proposed for non-replicated dynamic allocation of data and a framework for non-replicated dynamic allocation of data in distributed database has been developed using TTCA algorithm. The proposed algorithm re-allocates data with respect to the changing data access patterns with time constraint. Results show that the TTCA algorithm is an effective method for re-allocation of data as compare to Threshold algorithms for distributed database system where the frequency of access pattern changes rapidly. TTCA algorithm decreases the total cost of re-allocation and the number of migrations of fragments from one site to another site as compare to Threshold algorithm. TTCA algorithm improves the performance of the distributed database system significantly by decreasing the network traffic and reduce the data transfer cost compared to Threshold algorithm. TTCA gives the ownership of fragments to the site having maximum access rate but on the other hand Threshold algorithm the ownership of the fragment to last accessing site which may or may not be the frequent user of the fragment. TTCA ensures that the migration rate of the fragments decreases or remains same as the threshold value increases. But in case of Threshold algorithm migration rate can be increased as the threshold value
increase. By minimizing the number of migrations, TTCA algorithm further improves the overall performance of the distributed database system.

7.2 Future Scope of Work

The results of the proposed techniques are quite promising and contributing the significant improvement over the existing techniques. However, some directions for future work are given below to further improve the overall performance of the system:

➢ Since the main objective of database distribution design is to improve system performance and throughput. Concurrency control mechanism has to be incorporated in the cost model. The problem here is how to incorporate the information related to concurrent queries within the cost model.

➢ Security issues should be included in future improvements since distributed databases are frequently used for electronic commerce and encryption of the data is applied during the transactions of electronic commerce. The problem is how to calculate cost a query when the data associated with the query is encrypted.

➢ Static data allocation cost model can be extended for more complex queries.

➢ The performance of SBBO and BBO for data allocation can be enhanced by incorporating features of other evolutionary algorithms into SSBO and BBO.

➢ The performance of BBO for data allocation can be evaluated by considering non-linear curves for immigration and emigration rates.

➢ Threshold and Time Constraint Algorithm (TTCA) can further be extended for replicated dynamic allocation of data.