Name: Shweta Malhotra
Supervisor: Prof. M. N. Doja
Co-supervisor: Dr. Bashir Alam
Department: Department of Computer Engineering, Faculty of Engineering.
Title of Thesis: Developing a Framework for Cloud Database Management System.

Abstract

Big Data, Cloud computing, Data Science and many more fantasizing words are the future of IT industries. Cloud Database Management System is a new emerging concept recently introduced in the world. In Cloud the concept of Standard architecture of Cloud Database Management System is not yet been implemented. A framework has been proposed for a 5-layered architecture in cloud database management system to provide more levels of abstraction. First layer introduced is the External Layer, this layer is closest to the user, in which manageability, providing transparency and security are the important issue that should be considered. Second layer is the Conceptual Middleware Layer, as there are heterogeneous databases and clouds are available in the market, so here interoperability is the major issue. Third layer is the Conceptual Layer in which programming techniques, transaction management, query processing and optimization are the issues that should be considered. Fourth layer is the Physical Middleware Layer, as there are various platforms available so here also, interoperability between various platforms are the biggest issue and the last layer is the Physical Layer in which how data can be stored so that it can be easily accessible without so much overhead so here data security, storage, backup, load balancing, partitioning, scaling, elasticity, fault tolerance and replication are the important issues that should be considered.

For all the new techniques one common thing is that they deal with Data, not just Data but the Big Data. Users store their various kinds of data on cloud repositories. Cloud Database Management System deals with such large sets of data. Cloud Database service provider deals with many obstacles while providing the service. Among all the challenges processing of large amount of data, interoperability and security are the major ones. Enhanced Generalized Query Processing through MapReduce (E-GENMR), a prototype model defined and implemented provides solution for these problems. Firstly, for processing such gigantic amount of data, traditional approaches are not suitable because these approaches are not able to handle such amount of data. To handle this, various solutions have been developed such as Hadoop, MapReduce Programming codes, HIVE, PIG etc. but these technologies don't provide solution for these problems at the same time, and moreover users are not compatible with these latest technologies like MapReduce codes. E-GENMR provides interoperability as it takes queries written in various RDBMS forms like SQL Server, ORACLE, DB2, MYSQL and convert into MapReduce codes as MapReduce codes are considered to be the efficient way for processing large data. Secondly, Client's data is being stored in encrypted form and processing is being done on an encrypted data hence it ensures the security aspects. A comparison of various queries on unencrypted data has been done to evaluate the performance of E-GENMR with latest techniques like HIVE and PIG and it has been concluded that E-GENMR shows better performance as compare to both the techniques. Also, we compared the processing time of encrypted data with unencrypted data. With the help of regression analysis we proved that the prototype model is showing precise time. We also introduce an optimization technique for mapper placement problems to perceive the effect of parallelism which improves the performance of such Amalgam approach.