RELATIONAL DATABASE WITH SQL AND GRAPH DATABASE

[1]Reena, [2]Renu

[1,2] Computer Science & Engineering Department, M.D.U, (India)

ABSTRACT

This paper represents the study of various theory based on database model. In this paper a study of various papers is done, and in the reviewed paper graph database and SQL is done. This huge repository of unstructured data has resulted in making the data search and knowledge extraction, a very cumbersome task if one continues using the legacy relational database.

Keywords: *DBMS, flexibility, Relational DBMS, SQL, security*

I. INTRODUCTION

A database management system (DBMS) is the software that allows a computer to perform database functions of storing, retrieving, adding, deleting and modifying data. It provides facilities for controlling data access, enforcing data integrity, managing concurrency control, recovering the database after failures and restoring it from backup files, as well as maintaining database security.

1.1 Relational DBMS

The relational database model was modelled by E. F. Codd in 1969, a researcher at IBM. The model is based on set theory and predicate logic. The basic idea behind the relational model is that a database consists of a series of unordered tables (or relations) that can be manipulated using non-procedural operations that return tables. This model was in contrast to the more traditional database theories of the time that were much more complicated, less flexible and dependent on the physical storage methods of the data.

The word relational has its roots in the terminology that Codd used to define the relational model. The table in Codd's model was actually referred to as a relation (a related set of information). In fact, Codd (and other relational database theorists) used. Relational database has been providing support with implementation of many decades like oracle, MySql etc [1].The pages of WWW which are connecting to network of hyperlink is highly complex and not efficient in relational database [2].similarly these issues involved in the modeling of social network like facebook, orkut etc. To implement such issues is very expensive to calculate it[3].

II .LITERATURE REVIEW

➢ Information Storage has been a challenge throughout human history and existed long before modern computer systems. With examples like Dewey Decimal Classification (also called the Dewey Decimal System) , a proprietary system of library classification developed by Melvil Dewey in 1876 made information retrieval and indexing more efficient. However it still required vast amounts of physical volume to store data and relied on the human intellect to process trivial relations in that data.

- In 1964, Charles Bachmann developed the first commercial database management system (DBMS), IDS - Integrated Data Store based upon an early network data model while working at Honeywell. IDS maintained a single set of shared files on disk. Programs responsible for particular tasks, such as billing or inventory updates, retrieved and updated these files by sending requests to IDS. IDS provided application programmers with a set of powerful commands to manipulate data.

- In the late 1960s, IBM and North American Aviation (later Rockwell International) developed the first commercial hierarchal DBMS, IMS - Information Management System, and its DL/1-language. These models lack a good abstraction level which implies that it is difficult to separate the db-model from the actual implementation. The data structures provided are not apt for modelling non traditional applications. They permit database navigation at the record level by providing low-level operations that can be used to derive more abstract structures. The solution to these problems was given by Codd. In 1970, Codd released ―A Relational Model of Data for Large Shared Data Banks.‖ which provided the first definition of Relational Model.

- Codd's relational model suggested that all data in a database could be represented as a tabular structure (tables with columns and rows, which he called relations) and that these relations could be accessed using a high-level non-procedural (or declarative) language [reference]. Edgar F. Codd published a series of papers about relational databases from 1970 to 1972, in which he focused on the concept of abstraction levels by introducing a separation between the physical and logical levels. It was a major development because it gave the data modelling discipline a mathematical foundation. It is based on the simple notion of relation, which together with its logic and algebra, made it the most widely used model and its standard query language SQL, became a paradigmatic language for querying. But with time, database community realized that the relational model was inadequate for data intensive domains (knowledge bases, engineering applications) involving complex data objects and object interactions such as CAD/CAM software, computer graphics, and information retrieval.

- In order to provide additional semantics necessary to model these new applications Object-oriented (O-O) db-models appeared in the eighties. According to Rao in 1994, "The object-oriented database (OODB) paradigm is the combination of object-oriented programming language (OOPL) systems and persistent systems‖. O-O db-models view the world as a set of complex objects having certain states (data), where interaction is via method passing. Although O-O db-models permit much richer structures than the relational db-model, they still require that all data conform to a predefined schema. One of the promising solutions to this problem was given by document databases such as Lucene, which is able to fully index large document collections and support queries that rank the documents according to information retrieval measures.

- But what makes document databases really different, is the fact that documents are usually retrieved through dynamic and unpredictable queries. Thus document databases can usually associate any number of fields of any length to a document. Document databases are usually schema-less, i.e. there is no predefined data model. But with the vast increase in the usage of internet, these document models were not proving to be efficient in performing many operations for example the World Wide Web exhibits far more complicated networks of relationships than were expected . The network of hyperlinks connecting all the pages on the World Wide Web is highly complex and almost impossible to model efficiently in a

document database. Similar issues are involved in modeling the social network like Twitter, Facebook, etc. Implementing such a large network of relationships in the form of these conventional databases was an impossible task. So there was a need of the kind of model where both data manipulation and representation are graph-based.

➤ The database community became aware of this need and proposed graph databases. A survey of the literature on community structure, link prediction and social influence shows that graphs are the predominant model for social networks. Adrian Silvescu in his paper on graph databases showed how graph databases easily handle changes in the schema. A survey on the literature of graph databases shows that although activities around graph databases flourished in the first half of the nineties, the topic almost disappeared in later years. The reason behind it was the emergence of XML that captured all the attention of those working on hypertext which forced people working on graph databases to move to particular applications like spatial data, Web, and documents; the tree-like structure was enough for most applications at that time.

➤ A number of papers have been published on graph databases during the 70's and 80's in early approach, Roussopoulos and Mylopoulos proposed a semantic network to store data about the database in 1975. Shipman in 1981 presented an implicit structure of graphs for the data itself in the Functional Data Model, whose goal was to provide a ―conceptually natural‖ database interface. Logical Data Model (LDM) developed by Kuper and Vardi in 1984, was proposed on a different approach according to which an explicit graph db-model intended to generalize the relational, hierarchical and network models. Later Kunii in 1987 proposed a graph db-model for representing complex structures of knowledge called G-Base.

➤ In the late eighties, Lecluse in 1988 introduced O2, an object-oriented db-model based on a graph structure. Along the same lines, GOOD developed by Gyssens in 1990 was a system in which manipulations as well as representation are transparently graph based. Among the subsequent developments based on GOOD are: GMOD by Andries in 1992, which proposes a number of concepts for graph-oriented database user interfaces; Gram by Amann and Scholl in 1992, which is an explicit graph db-model for hypertext data; PaMaL by Gemis and Paredaens in 1993, which extends GOOD with explicit representation of tuples and sets; GOAL developed by Hidders and Paredaens in 1993, which introduces the notion of association nodes; G-Log in 1995, proposed a declarative query language for graphs; and GDM developed by Hidders in 2005 which incorporates representation of n-ary symmetric relationships.

➤ There were proposals that used generalization of graphs with data modeling purposes. Levene and Poulovassilis in 1990 introduced a db-model based on nested graphs, called the Hypernode Model, on which subsequent work was developed by Poulovassilis and Levene in 1994 and Levene and Loizou in 1995. The same idea was used for modelling multi-scaled networks and genome data [Graves et al. 1995a]. GROOVY is an object-oriented db-model which is formalized using hypergraphs. There are several other proposals that deal with graph data models. Guting in 1994 proposed GraphDB, intended for modeling and querying graphs in object-oriented databases and motivated by managing information in transport networks. Database Graph Views [Guti´errez et al. 1994] proposed an abstraction mechanism to define and manipulate graphs stored in either relational object-oriented or file systems. The project GRAS [Kiesel et al. 1996] uses attributed graphs for modeling complex information from software engineering projects. The well known OEM

[Papakonstantinou et al. 1995] model aims at providing integrated access to heterogeneous information sources, focusing on information exchange.

➢ One more problem that arose with the evolution of Internet was the need of searching and accessing information from the WWW. A search engine is a document retrieval system designed to help find information stored in a computer system, such as on the World Wide Web, inside a corporate or proprietary network, or in a personal computer. But the presence of huge amount of resources on the Web poses a serious problem of accurate search. This is mainly because today's Web is a human-readable Web where information cannot be easily processed by machine. Highly sophisticated, efficient keyword based search engines that have evolved today have not been able to bridge this gap. This huge repository of unstructured data has resulted in making the data search and knowledge extraction, a very cumbersome task if one continues using the legacy relational databases. One of the proposed solutions is to shift to the graph databases as they not only makes it easier to store and retrieve data for the world wide web but also provide relevant and streamlined search results from the abundant information present on the internet.

III. GRAPH DATABASE

Graph database models can be defined as those in which data structures for the schema and instances are modelled as graphs or generalizations of them, and data manipulation is expressed by graph-oriented operations and type constructors. These models took off in the eighties and early nineties alongside object oriented models. Recently, the need to manage information with graph-like nature has re-established the relevance of this area.

In graph database entity ,object have equal importance with object themselves[5]. With the exponential rapid development of Internet and its advent as an indispensable medium for electronic sharing of data, it has emerged as a huge repository of unstructured data. This has resulted in making the data search and knowledge extraction, a very cumbersome task if one continues using the legacy relational databases. The retrieval of relevant information from internet in minimum possible time and with maximum efficiency has been possible with the help of graph databases. Easily retrieving the relevant search results from the vast ocean of information requires the search to be semantic and not syntactic.

Some examples of are Infogrid, HypergraphDB, Jena, DEx, FlockdB,sQl and Neo4j[6]-[8].

Out of these sql is explained below

IV MYSQL

To implement relational databases, MySQL version 5.1.41 was used. The database was queried using PHP scripting language. Queries designed to analyze the performance difference between a relational database and a graph database are:

S0: Find all friends of Esha.

S1: Find the favorite movies of Esha's friends.

S2: Find the lead actors of favorite movies Esha's friends.

The first step is to connect to the MySQL server. Once the connection has been established to MySQL server, a new database is created.

Next step is to create tables within the database. Schema for relational database includes the following tables

1) User: user_id, user_name

2) Friends: user_id, friend_id

3) Fav_movies: user_id , movie_name

4) Actors: movie_name, actor_name

To retrieve the query results and note down the retrieval times of each query for hundred and five hundred nodes respectively. Unlike Neo4j, MySQL gives theretrieval time in seconds and not milliseconds. To get the retrieval time in milliseconds, we need to explicitly do the coding.

V. SECURITY

The database is queried using PHP scripting language. Retrieval times for the above query as calculated by MySQL for hundred and five hundred MySQL has extensive multi user support. However Neo4j does not have any built in mechanisms for managing security restrictions and multiple users. It presumes a trusted environment. Although there is Access Control List security mechanisms but even Access Control List management is handled at application layer. On the other hand, there is extensive support for ACL based security in MySQL users.

VI.FLEXIBILITY

Although relational databases are more mature and secure as compared to graph databases, but its schema is fixed, which makes it difficult to extend these databases and less suitable to manage ad-hoc schemas that evolve over time.

VII. CONCLUSION

This paper represents the graph database as well as MYSQL. In general, graph databases retrieve the results of the set of predefines query faster than relational databases. Not only this, graph databases are more flexible than relational databases as new relationships can be added to graph databases without the need to restructure the schema again . Graph databases are more scalable as well as increasing number of nodes from one hundred to five hundred does not increase the retrieval time by much as can be visualized.

VIII. FUTURE SCOPE

My SQL has extensive multiuser support. Although there is Access Control List security mechanisms but even Access Control List management is handled at application layer. On the other hand, there is extensive support for ACL based security in MySQL.

IX. ACKNOWLEDGEMENT

We express thanks to all department and personnel.We heartidly thanks to all members who contribute in the working of this paper

REFERENCES

[1] Tony Marston, ―The Relational Data Model, Normalisation and effective Database Design‖, Internet: http://www.tonymarston.net/php-mysql/database-design.html, 12th August 2005

[2] Paul Litwin, ―Fundamentals of Relational Database Design‖, Internet: http://www.deeptraining.com/litwin/dbdesign/FundamentalsOfRelationalDatabaseDesign.aspx, November2011

[3] Marko A. Rodriguez, July 2007, ―Problem-Solving using Graph Traversals‖, AT&Ti Technical Talk - Glendale, California. [4] Marko A. Rodriguez, April 8, 2009, ―Graph Databases and the Future of Large-Scale Knowledge Management‖, Los Alamos National Laboratory.

[5] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, , 2008: ―Bigtable: A distributed storage system for structured data‖. *ACM Trans. Comput. Syst.*, 26(2):1–26.

[6] G. DeCandia, D. Hastorun, M. Jampani,G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, 2007, ‖Dynamo: amazon's highly available key-value store‖. *SIGOPS Oper. Syst. Rev.*, 41(6):205–220,

[7] A. Lakshman., ‖Cassandra - a structured storage system on a p2p network‖, Internet: http: //www.facebook.com/note.php?note_id=24413138919, 2008.

[8] Charles Bachman, 1965. "Integrated Data Store." in: *DPMA Quarterly*.

[9] Codd, E. F., 1971 : "Normalized Data Base Structure: A Brief Tutorial", Proc. ACM-SIGFIDET Workshop on Data Description, Access and Control, San Diego, available from ACM, New York. 49

[10] Codd, E. F, 1971: "A Data Base Sublanguage Founded on the Relational Calculus" Proc. ACM-SIGFIDET Workshop on Data Description, Access and Control, San Diego, available from ACM, New York.

[11] Codd, E. F., May 24-25, 1971: "Further Normalization of the Data Base Relational Model", Courant Computer Science Symposia 6,New York City, Prentice-Hall.

[12] Fausto Rabitti, Ellsa Bertino, Won Kim, And Darrell Woelk , 1994, ―A Model of Authorization for Next-Generation Database Systems‖,Microelectronics and Computer Technology Corporation.

[13] A. S. Maiya and T. Y. Berger-Wolf, 2010, ―Sampling community structure‖ . *www*, pages 701–710, New York, NY, USA,. ACM.

[14] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, 2008,‖ Statistical properties of community structure in large social and information networks‖.. *WWW*, pages 695–704, New York, NY, USA, . ACM .

[15] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, 2006,‖Group formation in large social networks: membership, growth, and evolution‖ ,*SIGKDD*, pages 44–54, New York, USA,. ACM.

[16] H. Kashima and N. Abe,2007 ― A parameterized probabilistic model of network evolution for supervised link prediction‖. *ICDM*.

[17] D. Liben-Nowell and J. M. Kleinberg, 2007,‖The link-prediction problem for social networks‖ *JASIST*, 58(7):1019–1031.

[18] B. Taskar, M. fai Wong, P. Abbeel, and D. Koller, 2003 ―Link prediction in relational data‖. In in Neural Information Processing Systems.

[19] A. Anagnostopoulos, R. Kumar, and M. Mahdian, 2008 ― Influence and correlation in social networks‖. *SIGKDD*, pages 7–15, New York, NY, USA,. ACM.

[20] J. Tang, J. Sun, C. Wang, and Z. Yang,2009 ―Social influence analysis in large-scale networks‖, *SIGKDD*, pages 807–816, New York, NY, USA, ACM

[21] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proc. WSDM*, pages 241–250, New York, NY, USA, 2010. ACM.

[22] Roussopoulos, N., Mylopoulos, J., September 1975, "Using Semantic Networks for Database Management", Proceedings of the First International Conference on Very Large Databases, , 144-172. **[25]** Chrrstophe Lbcluse, Phllrppe Rwhard and Fernando Velez GIP Altcur, 1988 ―O2, an object-oriented data modell, SIGMOD '88 Proceedings of the 1988 ACM SIGMOD international conference on Management of data.

[26] Amann, Bernd, Scholl, Michel,1992 "Gram: A Graph Data Model and Query Language", Proceedings of the 1992 European Conference on Hypertext, ACM.

[27] Jan Hidders, Jan Paredaens , 1993, ―GOAL, A Graph-Based Object and Association Languagel. CISM - Advances in Database Systems 1993: 247-265

[28] D. Mukhopadhya, A. Banik, S. Mukherjee, and J. Bhattacharya, ―A Domain Specific Ontology Based Semantic Web Search Enginel, Arxiv preprint, arXiv, 2011, Available: arxiv.org. 51

[29] Neo4j. Home. http://neo4j.org, 2012.

[30] D. Dominguez-Sal, P. Urb´on-Bayes, A. Gim´enez-Va˜n´o, S. G´omez Villamor, N. Mart´ınez-Baz´an, and J.L. Larriba-Pey, 2010―Survey of Graph Database Performance on the HPC Scalable Graph Analysis Benchmarkl, Proceeding WAIM'10 Proceedings of the 2010 international conference on Web-age information management.

[31] Chad Vicknair, Michael Macias, Zhendong Zhao, Xiaofei Nan, Yixin Chen, Dawn Wilkins,2010, ―A Comparison of a Graph Database and a Relational Databasel, ACM Southeast Regional Conference.

[32] Database Trends And Applications, Internet: http://www.dbta.com/Articles/Columns/Notes-on-NoSQL/Graph-Databases-and-the-Value-They-Provide-74544.aspx,2012

[33] Chang, F., Dean, J., Ghemawat, 2008.‖Bigtable: A distributed storage system for structured datal. ACM Trans. Comput. Syst. 26(2)

[34] Mart´ınez-Bazan, N., Munt´es-Mulero, V., G´omez-Villamor,2007,―Dex: high performance exploration on large graphs for information retrievall. In: CIKM pp.573–582.

[35]Jena-RDF. Jena documentation. Internet: http://jena.sourceforge.net/documentation. html , 2010.

.