**Paper Title (16 Bold)**

Author (14)

 **1** *Computer Science, Swami Devi Dayal Group of Professional Institutes, Haryana, India*

*2Computer Science, Swami Devi Dayal Group of Professional Institutes, Haryana, India*

*Corresponding Author: xxxx (10)*

*--------------------------------------------------------ABSTRACT-----------------------------------------------------------(11 BOLD)*

*The Data Relationships tool is a collection of programs that we can use to manage the consistency and quality of data that is entered in certain master tables. Using this tool to set up data relationships enables us to place tighter controls over our data and helps increase efficiency during data entry.SQL is used to query the RDBMS which generates the specific results according to the needs of the user. Apart from querying the data, we can also update and modify the records in the database. But unfortunately it is not so easy to perform the above mentioned operations. This is because importance of relational databases is going to decrease due to the exponential growth of data as it is difficult to work with large number of joining tables. The best solution to such kind of problem is to make use of graph database for storing data. This pap compares the graph databases Neo4j with relational SQL.*

***KEYWORDS;- (11 BOLD)*** *Graph, Neo4j, MySQL, cypher query language*

---------------------------------------------------------------------------------------------------------------------------------------

Date of Submission: xx-xx-xxxx Date of acceptance: xx-xx-xxxx

---------------------------------------------------------------------------------------------------------------------------------------

1. **INTRODUCTION (11 BOLD)**

A database is an assemblyof meaningful information tabulated in such a way, that we can easily andquickly retrieve the appropriateresult. The database management system is a general purpose software system that promotes the process of defining, constructing and manipulating database for various applications. A relational database management system (RDBMS) is a database management system (DBMS) that is based on the relational model invented by Edgar F. Codd, of IBM's San Jose Research Laboratory. Most databases in widespread use are based on the relational database model.A relational database is a set of tables containing data fitted into predefined categories. Each table (which is sometimes called a relation) contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns. For example, a typical business order entry database would include a table that described a customer with columns for name, address, phone number, and so forth.The relational model offers various levels of refinement of the table relations called database normalization.Relational Databases have been providing the storage spacesupport for many decades now with implementations like Oracle, MySQL, etc.[1].Despite advances in computing, faster processors and high-speed networks, the performance of relational database applications is becoming slower and slower. RDBMS is not the best solution. Database schemas are very strict. When the data gets big, then the traditional SQL join operations may not work. Changes made to a single column will resound and would result in multiple changes. And finally, we need look into the mismatch occurring in different tables due to the changes done. With the intense demand to store huge volumes of information, it was a need to switch from relational databases to graph database.

1. **GRAPH DATABASES (11 BOLD)**

A graph can be formally deﬁned as a set of vertices and edges, eg, G=(V,E). The concept of the graph theory can be attributed to Leonhard Euler, who ﬁrst solved theSeven Bridges of Konigsberg problem in 1736. In a graph database, the objects (entities) are represented by nodes and the relationship between them is represented by edges. Both the nodes and edges contain the properties[2]. The objects are represented by nodes and the edgesrepresent the relationship between the nodes. The nodes and edges both contain many properties that illustrate their particular characteristics. Graph databases examples are: Allegro Graph, Hypergraph DB, Graph Base, Bitsy and Neo4j [3]. Among all of them only Neo4j is examined here.

**Neo4j**

Neo4j is an open source graph database which is implemented in Java and developed by Neo technology [4].It uses graph structure with nodes, edges and properties to store data. It provides index free adjacency, which means that each node is a pointer. It also supports the labelled property graph model. A labelled property graph model has a number of characteristics. It contains nodes and relationships where properties and nodes can be labelled with more than one label[10].Cypher is a declarative graph query language that enables demonstrative and productive querying and updating of the graph database [5]. Cypher is a self explanatory and robust language.Cypheris a language that expresses the logic of a computationwithout describing itscontrol flow. Cypher focuses on what result is to retrieve from graph, not on how to retrieve it [6]. Advantage of this feature is that it promotes the user to target only on their work area. In this model, if we want to model a relationship among friends then we have to create different nodes for friends and connect the friends with edges. In the graph model [7] shown above, there are three nodes and five relationships and a label. User is the label. Ruth, Billy and Harry are the nodes and the property is name for the nodes and ‘follows’ is the p. Here, Ruth follows Billy, which means that there is a connection between these two nodes using a directed edge, and the edge has a property called ‘follows”. In Neo4j, the edges are bidirectional as we can see in the model in Fig.1. Neo4j uses the cypher query language for creating nodes, relationships with properties and also for retrieving, searching and manipulating data.

**Comparison Parameters For Relational And Graph Database**.

The assessment between SQL and Cypher is done on different evaluation parameters [2]. On the basis of these evaluation parameters a decision is taken on the selection of the database [9].

1. **RESULT VIEW (11 BOLD)**

The result view means the view of the result set after thequeries gets executed.In SQL the result set can be viewed in the form of tables as the relational database were originally designed in the tabular form.No Graph view of the result set is available.While in Cypher query language, we can get both views (tabular view and graph view) of the result set. The properties of nodes present in graph database can be directly viewed by just looking at the graph view of the result set which is not possible in case of tabular form.

**Complex Search and Join Operations**

In relational databases the information in one table is related with information present in another table by joining the tables using foreign key

A large number of joins: when queries that join different tables are utilized,there’s an explosion of complexity and computing resource consumption.This results in a corresponding increase in query response times. Numerous self-Joins( or recursive joins):Self-Join statements are common for hierarchy and tree representation of data, but traversing relationships by repeatedly joining tables to themselves is inefficient. In fact, some of the longest SQL queries in the world involve recursive joins. While in Cypher query language, to retrieve the result of complex queries it does not traverse the entire graph. It simply checks for all the nodes that have direct relationship with the current node that satisfies the given condition in the query[9].

**Retrieval Time**

In SQL the time taken by a query to retrieve the result is given in seconds.While in Cypher query language taken by a query to retrieve the result is given in milliseconds. Retrieval time is directly proportional to the complexity and number of relations involved in the query[8].

**Implementation Details**

For comparing the SQL and Cypher query language,we perform a search operation on the number of person’s friend list upto the depth 5. SQL (Structured Query Language) was used to execute queries on MySQL. MySqlsupports relational databases.Cypher Query Language was used to execute queries on Neo4jCommunity version 2.0.3, which supports graph databases. In relational database following relations were made with the given schema shown in tables below:

|  |  |
| --- | --- |
| USER-ID | NAME |
| 1 | RAJ |
| 2 | BISMIL |
| 3 | JHON |
| 4 | HINA |
| 5 | SUHAN |

|  |  |  |
| --- | --- | --- |
| ID | USER-ID |  FRIENDS-ID |
| 1 | 1 | 9130 |
| 2 | 1 | 2503 |
| 3 | 2 | 3038 |
| 4 | 2……… | 7358 |
| 4 | 10000 | 3322 |

Here, user and relationships are shown in two tables and the relationship table represents the relation between users. An id has been assigned to every user in Table 1. Here in table2 we have shown the relationship between the user and its friends.

**The five queries defined were:**

Query 1(Q1): Count friends of a particular user.

Query 2(Q2):Count friends of friends of a particular user.

Query 3(Q3): Count friends of friends of friends of a particular user.

Query 4(Q4): Count friends of friends of friends of friends of a particular user.

Query 5(Q5): Count friends of friends of friends of friends of friends of a particular user

In the tables given below we have shown the query results in milliseconds (ms) for Mysql and Neo4j.

**Graph db performance**

* a sample social graph
* with ~10,000 persons
* average 50 friends per person
* Table For depth 1

|  |  |  |
| --- | --- | --- |
| DATA BASE | #FRIENDS | QUERY TIME |
| SQL | 2204 | 1109ms |
| Neo4j | 2204 | 25 ms |

**Table 1 (For depth 1)**

**Graph db performance**

* a sample social graph
* with ~1,000 persons
* average 50 friends per person
* Table For depth 2

|  |  |  |
| --- | --- | --- |
| DATA BASE | #PERSONS | QUERY TIME |
| SQL | 9999 | 18660ms |
| Neo4j | 9999 | 47 ms |

**Table 2 (For depth 2)**

**Graph db performance**

* a sample social graph
* with ~1,000 persons
* average 50 friends per person



**Table 3(For depth 3)**

**Graph db performance**

* a sample social graph
* with ~1,000 persons
* average 50 friends per person
* Table For depth 4

|  |  |  |
| --- | --- | --- |
| DATA BASE |  #PERSONS | QUERY TIME |
| SQL | 38 | 1784056ms |
| Neo4j | 38 | 15 ms |

 Table 4(for depth 4)

1. **CONCLUSION (11 BOLD)**

The graph databases and relational database both performed well. In general, graph databases performed better when objective tests were performed. The Implementation shows that graph databases retrieve the results of the set of predefines query faster than relational databases.Alsograph databases are more flexible than relational databases as we can add new relationships to graph databases without the need to restructure the schema again.

**REFERENCE (11 BOLD)**

1. Chad Vicknair, Michael Macias, Zhendong Zhao, Xiaofei Nan, Yixin Chen, Dawn Wilkins “A Comparison of a Graph Database andaRelational Database”, ACM Southeast RegionalConference,2010
2. A.Silvescu,D.Caragea, A.Atramentov, “Graph Database”, ArtificialIntelligence Research Laboratory, 2012.
3. M. Buerli, “The current state of graph Databases”, ACM Computing Survey,2012.
4. Learn Graph Database, Available: <http://www.neo4j.org/learn/neo4j>
5. Graph Database Example, Available: <http://en.wikipedia.org/wiki/Graph_database>.
6. C.C. Aggarwal, H. Wang, “Managing and mining graph data”, volume 40,Springer, 2010.
7. [www.neo4j.com](http://www.neo4j.com)
8. Neo4jBlog,Available:[http://blog.neo4j.org/2009/04/current database-delete-and](http://blog.neo4j.org/2009/04/current%20database-delete-and)graph.html
9. Neo4j. The neo database (2006), Available <http://dist.neo4j.org/neo-technology-introduction.pdf>
10. R. Angles and C. Gutierrez,” Survey of graph database moDels”,. ACM Comput. Surv., 40(1):1–39, 2008.