International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 603-610 www.ijaem.net ISSN: 2395-5252

Review Paper OnComplexity Reduction In Relational Database Using Neo4j Graph Database

T.P.Deshmukh[1], B.R.Bamnote[3], S.W.Ahmad[3]

[1]*(Computer Science And Engineering, Prof.RamMeghe Institute Of Technology And Research/ SantGadge Baba Amravati University, Badnera Maharashtra)*

[2]*(Computer Science And Engineering, Prof.RamMeghe Institute Of Technology And Research/ SantGadge Baba Amravati University, Badnera Maharashtra)*

[3]*(Computer Science And Engineering, Prof.RamMeghe Institute Of Technology And Research/ SantGadge Baba Amravati University, Badnera Maharashtra)*

Date of Submission: 15-11-2020 Date of Acceptance: 30-11-2020

ABSTRACT: Now a days, data is increasing in the volume because of digital process .more online transaction and new social network. There is large amount of data generated everyday in the field of manufacturing, business, science and our personal lives. Proper processing of the data could reveal new knowledge about market, society and environment, and enable us to react to emerging opportunities and changes in a timely manner. ,The ratio of increasing data amount in our digital world seems to outspeed the advance of our computing infrastructure. Databases are an indispenable part of a computing system and users heavily depend on the services they provide. When coact with a computing system, we expect that data be stored for future use, that the data is able to be looked up fastly, and we can perform complex queries and get the required data easily which is stored in the database. Many different types database types available for use such as relational databases, object databases, keyvalue databases, graph databases, and RDF databases. Each type of database has it's unique qualities that have applications in certain domains. It'swork aims to investigate and compare the performance and scalability of relational databases to graph databases in terms of handling multilevel queries such as finding the impact of a particular disease with the working area of pass out students. MySQL was selected as the relational database, Neo4j as the graph database.

KEYWORDS –Cypher Query Language, Graph DataBase, Relational Database, NoSql, Neo4j

I. INTRODUCTION

The relational model is most influential database in the computer industry since the 1980s mainly for storing and retrieving data. Later, relational database has been losing its importance due to its reliance on a strict schema which makes it difficult to add new relationships between the objects. Another reason of its failure is that as the available data is growing manifolds, it is becoming complicated to work with relational model as joining a large number of tables is not working efficiently.

Chief information officer and chief technical officer don't just need to manage larger volumes of data – they need to generate insight from their existing data. So, the relationships between data points matter more than the individual points themselves. . One of the suggested solutions is to transfer to the Graph databases as they aspire to overcome such type of problems. In order to maintain data relationships, organizations need a database technology that stores relationship information as a first-class entity. That technology is a GraphDatabase.

Relational databases like Oracle and MySql excel when it comes to capturing repetitive, tabular data. The word "relational" in their name, relational database are much less effective at storing or showing relationships between stored data elements. A relational database, a graph database is structured fully around data relationships. Graph databases consider relationships not as a schema structure but as data.

The graph database queries are domain-specific ,user-friendly and can be treated as "SQL for graphs". The similarity to SQL is intentional and makes transition much easier for developers. When SQL execute query on the RDBMS is as long as half, the Cypher Query equivalent is much shorter. The traverser Application Programming Interfacce in RDBMS is highly resource intensive, since each step to neighboring node has to be depicted with JOIN. In the graph database

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 603-610 www.ijaem.net ISSN: 2395-5252

hypergraph property allows direct access to adjacent nodes by eliminating the edge attribute.
There are several graph databases like Neo4j Graph Database, OrientDB ,ArangoDB , MarkLogic and Allegro Graph. Neo4j is most popular and open source graph database.Its native and elegant solution and ease of data remodeling.

To exlore the concept of Neo4j Graph Database proposed the system todevelope the application for medical data analysis. Medical data analysis is important for application to find relationship between health data. Other than normal relational database, a novel and popular graph database, Neo4j uses graph related concepts to describe data models, the data can easily be nodes, edges and their associated attributes. By focusing on medical data. Firstly, Neo4j and Cypher Query Language are introduced. Then Neo4j is applied to analyze the associations among key objects in Medical data which are Doctor, Patient etc. Neo4j database is good at dealing with complex and multi-connection data, using Neo4j database to store and manage medical data makes it convenient for Medical data analysis.

Graph Database:

Databases are an integral a part of a computer system and users heavily believe the services they provide . When interact with a system , we expect that data be stored for future use, that the info is in a position to be searched fastly, and that we can perform complex queries against the info stored within the database. many various emerging database types available to be used like relational databases, object databases, key-value databases, graph databases, and RDF databases. Each sort of database provides unique qualities that have applications in certain domains. Our work aims to research and compare the performance and scalability of relational databases to graph databases in terms of handling multilevel queries like finding the impact of a specific subject with the working area of pass out students. MySQL was chosen as relational database , Neo4j as thegraph database.

Very simply, a graph database is a database designed to treat the relationships between data as equally important to the info itself. it's intended to carry data without constricting it to a pre-defined model. Instead, the info is stored like we first draw it out – showing how each individual entity connects with or is said to others. Accessing nodes and relationships during a native graph database is an efficient, constant-time operation and allows you to quickly traverse many connections per second per core.

Independent of the complete size of your dataset, graph databases excel at managing highly-connected data and complex queries. With solely a pattern and a group of beginning points, graph data bases explore the neighboring knowledge around those initial beginning points— grouping and aggregating info from various nodes and relationships — and going any data outside the search perimeter untouched.

To solve the problem of data storage in computer filed, we need to use different storage technologies. In many storage technologies, relational databases have long been dominant. With the application of Wed 2.0, the rise of medical sector, the internal data dependence and complexity increase gradually, more and more problems arise in relational databases. Then, graph database appeared. In recent years there have been a number of high-performance graphics database for the product environment, such as Neo4j, Infinite, Graph, DEX, InfoGrid, HyperGraphDB, Trinity and so on. Among them, Neo4j is the mainsream of a Java based open source software currently, its kernel is a very fast graphics engine, with the recovery, the two phase of the submission, support for XA transactions and other database product features.

What is Neo4j

Neo4j is an open-source, NoSQL, native graph database that offers an ACID-compliant transactional backend for your applications. The source code, written in Java and Scala. Neo4j has every a Community Edition and Enterprise Edition of the database. The Enterprise Edition includes all that Community Edition has to produce, and extra enterprise necessities like backups, clustering, and failover skills.

Neo4j is a network-oriented database—that is, an embedded, disk-based, fully transactional Java persistence engine that stores data structured in networks instead of in tables. What makes Neo4j interesting is that the use of the so called "network oriented database". during this model, domain data is expressed in an exceedingly "node space" - a network of nodes, relationships and properties (key value pairs), compared to the relational model's tables, rows &columns. Relationships are first class objects and should even be annotated with properties, revealing the context within which nodes interact. The network model is compatible to problem domains that are naturally hierarchically.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 603-610 www.ijaem.net ISSN: 2395-5252

Neo4j is stated as a native graph database because it efficiently implements the property graph model all the way down to the storage level. this implies that the information is stored exactly as you whiteboard it, and also the database uses pointers to navigate and traverse the graph. In contrast to graph processing or in-memory libraries, Neo4j also provides full database characteristics, including ACID transaction compliance, cluster support, and runtime failover – making it suitable to use graphs for data in production scenarios.

Fig:1 simple graph database.

Neo4j providing the following feature.

Need OF Neo4j Graph Database:

o Cypher, a declarative query language quite like SQL, but optimized for graphs. now employed by other databases like SAP HANA Graph and Red is graph via the open Cypher project.

o Constant time traversals in big graphs for both depth and breadth because of efficient representation of nodes and relationships. Makes possible to measure billions of nodes on moderate hardware.

o Flexible property graph schema that will adapt over time, making enables materialize and add new relationships later to shortcut and speed up the domain data when the business needs change.

Today's administator don't just need to manage larger volumes of data – they need to generate insight from their existing data. In this case, the relationships between data points matter more than the individual points themselves.

So, as to use information connections, associations need an information base innovation that stores relationship data as a top notch substance.Tech giants like Google, Facebook, LinkedIn and PayPal all tapped into the power of graph databases to create booming businesses. Their secret. They each used graph database technology to harness the power of data connections.

A graph database is purpose-built to handle highly connected data, and the increase in the volume and connectedness of today's data presents a tremendous opportunity for sustained competitive advantage.

Graph databases have three other key advantages:

1) Performance

For serious information relationship dealing with, diagram data sets improve execution by a few significant degrees. With customary information bases, relationship questions will go to a pounding stop as the number and profundity of connections increment. Conversely, chart information base execution remains consistent even as your information develops year over year.

2) Flexibility

With grapg bases, IT and information draftsman groups move at the speed of business in light of the fact that the structure and diagram of a diagram model flexes as applications and ventures change. As opposed to comprehensively displaying an area early, information groups can add to the current diagram structure without jeopardizing current usefulness.

3) Agility

Creating with chart information bases adjusts impeccably with the present dexterous, test-driven improvement works on, permitting your diagram information base to advance in sync with the remainder of the application and any changing business prerequisites. Present day diagram information bases are prepared for frictionless turn of events and smooth frameworks support.

Utilization of the neo4j diagram information base:

The present endeavor associations use diagram information base innovation in a decent variety of ways:

- o Fraud discovery
- o Real-time proposal motors
- o Master information the executives (MDM)
- o Network and IT tasks
- o Identity and access the executives (IAM)

From undertakings like Walmart, eBay and the adidas Group to new companies like Cobrain, Zephyr Health and Wanderu and even non-benefits like the ICIJ and the World Economic Forum – contextual investigations with chart

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 603-610 www.ijaem.net ISSN: 2395-5252

information bases overflow with assorted variety and profundity of utilization.

Importance Of Neo4j Graph Database

In graph database each node represents to a substance (an individual, place, thing, classification or other bit of information), and every relationship shows how two nodes are related. This broadly useful structure permits you to show a wide range of situations – from an arrangement of streets, to an organization of gadgets, to a populace's clinical history or whatever else characterized by connections.

A graph database is an online data set administration framework with Create, Read, Update and Delete (CRUD) activities chipping away at a diagram information model. In contrast to different information bases, connections take primary goal in diagram data sets. This implies your application doesn't need to surmise information associations utilizing things like unfamiliar keys or out-of-band handling, for example, MapReduce. The information model for a diagram information base is additionally essentially easier and more expressive than those of social or other NoSQL information bases. Diagram information bases are worked for use with value-based (OLTP) frameworks and are built in view of conditional respectability and operational accessibility.

Graph Storage: Some graph databases utilize local chart stockpiling that is explicitly intended to store and oversee graph, while others utilize social or item situated information bases. Non-local stockpiling is frequently substantially more inert.

Graph Processing Engine: Native diagram preparing (a.k.a. "file free contiguousness") is the most proficient methods for handling diagram information since associated hubs genuinely "point" to one another in the data set. Non-local chart handling utilizes different intends to deal with CRUD tasks.

II. RELATED WORK

Concerning the starting points of chart information bases, K. Nagi.built up a review on chart information base models proposed before the year 2002.. The authors synthesized the notion of a “graph database model” and compare the proposals available at the moment. It is important to emphasize that most of the works reviewed by the authors followed a theoretical interest more than practical developments. Hence, a practical evaluation of the models is not available. With respect to the recent developments in the area. Pere Burton [9] reviewed six graph databases (Neo4j, HyperGraphDB, DEX, InfoGrid, Sones and VertexDB) and published a comparison-matrix that included information like software features (e.g., license), schema features (e.g., types of nodes and edges), query features (e.g., language and traversals), general database features (e.g., transactions, indexing), database operation utilities (e.g., protocols), language bindings and operating systems. This work summarizes the features but does not include major discussion nor analysis. Another informal review Pere Burton included an interesting questionnaire about desirable features for graph databases.

J. Webber [3] assessed the exhibition of three graph databases (Neo4j, HypergraphDB and DEX) and a RDF Database (Jena). The tests, that incorporated the assessment of a few ordinary graph activities over various graph sizes, indicated that DEX and Neo4j were the most effective usage.

The thought learning system considering foundation divulgence is Graph Based Induction (GBI) (Yoshida, Motoda, 1995). It utilizes hued digraph as the portrayal system where hues joined to the hubs speak to the qualities of the realities. GBI looks at each associated pair of hubs, and unions the successive regular ones. The last consolidated bases are named as ideas. Domingues-Sal etal.evaluated the execution of three diagram information bases (Neo4j, HypergraphDB and DEX) and a RDF Database (Jena). The tests, that joined the appraisal of a couple of normal diagram activities over different chart sizes, shown that DEX and Neo4j were the most effective executions.

B.T. Messmer et al. presents a choice tree approach for ordering models for isomorphism and subgraph Isomorphism. This technique makes answers in polynomial time, at the expense of a record which is exponential in measure concerning information base size.

Emil Eifrem, CEO of Neo Technology has ran a couple of tests in which he differentiated the speed of social with diagram information bases. He made a "companions of companions" question and found that when the question of connections went three levels significant that the diagram information base beat the social one by a factor of 150, and when the request significance was extended to four the graph data set outclassed the social one by a factor of 1000

Points, R.; et al presents an overview of prior work (pre-NOSQL) in diagram information bases. for example before 2002,particularly topographical, spatial and semi organized

information base models. More established information models zeroed in vigorously on semi organized and XML information in a conventional data set. The creators blended the idea of a "diagram information base model" and think about recommendations accessible right now. Edges, R.; performs examination and execution investigation of various chart information base models analyzes current diagram data sets focusing on their information model highlights, that is information structures, question offices, and trustworthiness requirements. Creator shows that most diagram information base models offer an inborn help for various chart structures, inquiry offices as APIs (a large portion of the models) and question and essential ideas of trustworthiness constraints.dialects (a couple of them),

Philip Howard, inspector at Bloor Research itemized that Graph information bases are essential when the degree of division [ie, I know x who knows y who is related to z who used to live in an indistinct house from w etc.between substances ends up being exorbitantly mind boggling, making it difficult to manage using conventional development. Prophet or DB2, for example, can reasonably manage up to three degrees of separation yet not the six or seven. Howard commented on the limitations of diagram information bases, saying that "The genuine obstacle is that while these are really NoSQL data sets, essentially they can't be executed over an insignificant exertion gathering (at any rate not a present) but instead need to continue running on a singular machine, the explanation being that execution taints rapidly over a framework. Another potential drawback is that conceivably you have to make your own inquiries using Java or whatever — which infers using expensive programming engineers — or you use SparcQL or one of the other request tongues that have been made to help chart information bases, anyway this suggests taking in another skill

Jouili, S.;et al precisely contemplates chart information bases ie shows Graph Database Benchmark, to see four diagram data sets: Neo4j, DEX, Titan (BerkeleyDB and Cassandra) and OrientDB (neighborhood) on different kinds of remaining burdens, each time perceiving which information base was the best and the less balanced. Considering measure, the information base that got the best results with crossing outstanding burdens is unquestionably Neo4j: it outmaneuvers the different contenders, regardless the remaining task at hand or the boundaries used[6]. General systems for making built records for benchmarking and testing record linkage approaches are depicted in (Christen and Vatsalan, 2013; Ioannou, Rassadko, and Velegrakis, 2013; Talburt, Zhou, and Shivaiah, 2009). The information age device utilized in this work is specific for genealogical populace structures, and the worldwide setup boundaries that are upheld mirror this.

Healthcare system in United States was generating more data and they required new technology to handle data analytics effectively. Data driven approach is used to handle data analytics in healthcare systems by using two independent tasks, data management and data services. Here, data management means storing the data with minimal redundant structure and error free. Information administrations portray different examination questions, for example, join, search and measurable quiries. The problem appeared due to the gap between data management and data services in relational databases. To overcome this problem, they presented an approach to convert third normal form (3NF) of relational databases in equivalent graph of Graph database. A graph database does not require creating more tables and replicating them unlike relational databases. For example, Neo4j is suitable in OLTP (online transaction processing) environment. Pregel is utilized where high idleness and high through put have high need. The experiments have shown that Graph database performed better than relational database (MySQL) in the heterogeneous environment of healthcare systems of United States in OLTP.

The drawback of the existing system is that it is very difficult to manage the data in database level (i.e. Relational Database). It is hard to deal with the entire framework physically and it is less exact and to save the information in connection for future reference since it might get mind boggling on data set level. Moreover it is very difficult to retrieve data. Redundancy of data may occur and this may lead to the inconsistency.

In Relational database-management systems (RDBMS) only model data as a set of tables and columns, carrying out complex joins and self-joins when the dataset turns out to be more between related. Such questions are in fact complex to build and costly to run. And, making them work in real time while end users wait is not easy, with performance faltering as the total dataset size increases. Plus, making them work in real time while end users wait is not easy, with performance faltering as the total dataset size increases.It tends to the less speed and accuracy.

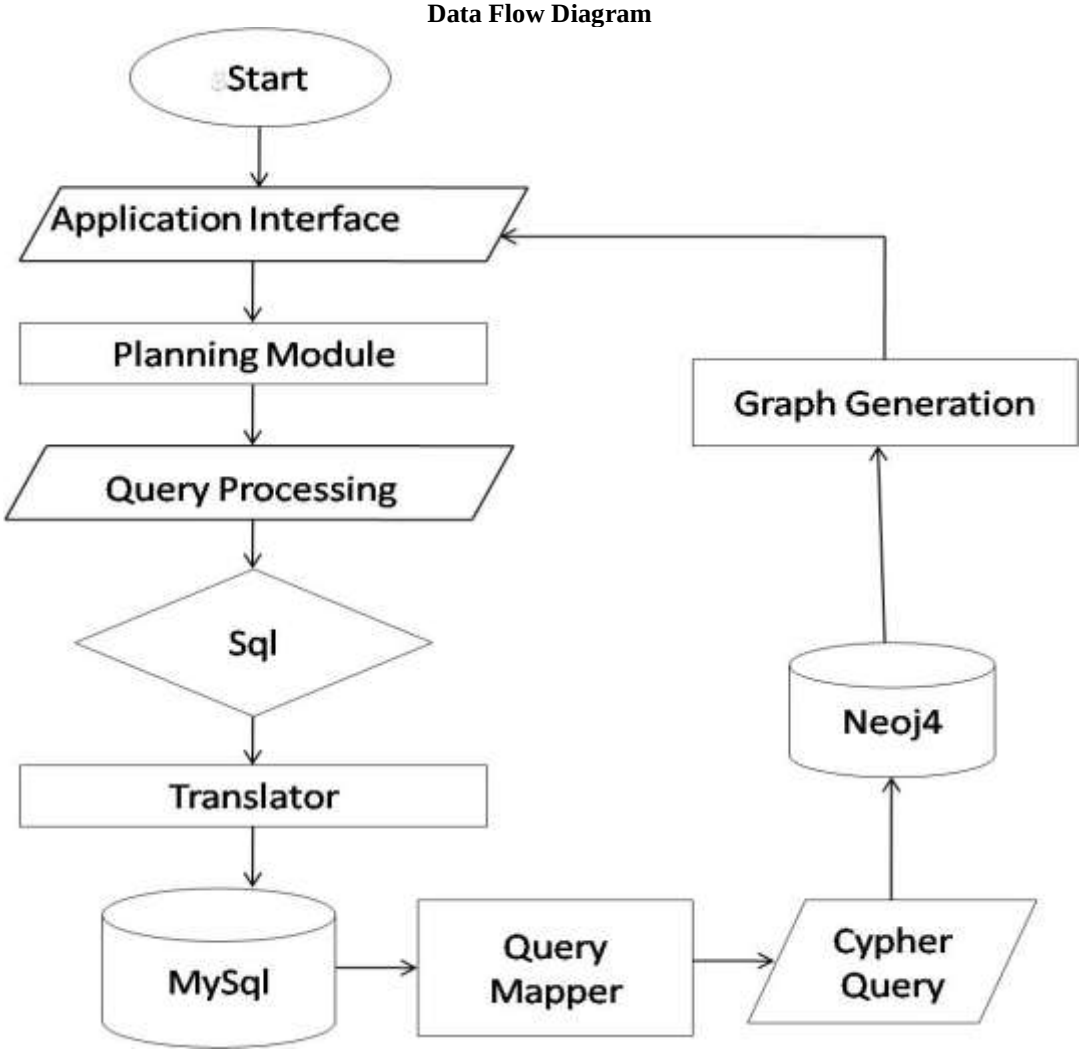
In proposed system Relational database and Graph database is use for Speed and accuracy

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 603-610 www.ijaem.net ISSN: 2395-5252

this is main advantages of proposed system. There is no redundancy of data. The data are stored in relational database and then it can be use for later manipulation which use by Graph Database. Data is receive from relational database and its converted into cypher query language. self-joins when the dataset turns out to be more between related. Such questions are in fact complex to build and costly to run. Cypher, a definitive inquiry language like SQL, yet enhanced for charts. The proposed frameworks kill the disadvantages of the current framework generally.

III. PROPOSED WORK

Data Flow Diagram



Neo4j, in the same way as other graph bases, expands upon the property chart model; marked hubs (for educational substances) are associated by means of coordinated, composed connections. The two hubs and connections hold self-assertive properties (key-esteem sets). There is no unbending pattern, yet with nodelabels and relationship-types we can have as much meta-data as we like. When bringing information into a chart information base, the connections are treated with as much incentive as the information base records themselves. This permits the motor to explore your associations between hubs in steady time. That thinks about well to the exponential stoppage of many-JOIN SQLqueries in a social information base.

Proposed System comprises of examination and correlation of two information bases, for example, Neo4j and MySQL data sets. A chart information base stores information in a diagram, the most nonexclusive of information structures, prepared to do richly speaking to any sort of information in an exceptionally available manner.

MySQL has huge market infiltration in the scholastic and logical fields. Besides MySQL has

critical help, both from the producers and from the client network. It is an unadulterated social information base, instead of an item social information base like Oracle and SQL Server. Neo4j is open hotspot for every noncommercial use. It has been underway for more than five years. It is rapidly getting one of the chief chart information base frameworks. As per the Neo4j site, Neo4j is "an inserted, plate based, completely conditional Java constancy motor that stores information organized in diagrams as opposed to in tables". The engineers guarantee it is astoundingly adaptable (a few billion hubs on a solitary machine), has an API that is anything but difficult to utilize, and underpins productive crossings. Neo4j is fabricated utilizing Apache's Lucene for ordering and search. Lucene is a book web crawler, written in Java, designed for superior

Objectives:To Reduce The Complexity Of Relational Database.

1. DBA Will Be Able To Monitor The Operation With Ease.
2. Allows You To Model All Kinds Of Scenarios – From A System Of Roads, To A Network Of Devices, To A Population's Medical History Or Anything Else Defined By Relationships.
3. To Handle Highly Connected Data, And The Increase In The Volume And Connectedness Of Today's Data Presents A Tremendous Opportunity For Sustainable Competitive Advantage.
4. To Provide Minutes-to-Milliseconds Performance: Query performance and responsiveness are at the top of many organizations' concerns with regard to their data platforms. Online value-based system – huge web applications specifically – must react to end clients in milliseconds in the event that they are to be fruitful.

IV. CONCLUSION

Neo4j is suitable for database creation, query, update, etc, particularly suitable for handling a large number of complex, dynamic, interactive, low structured data. It is an effective method to solve the problem of mass data storage in the fields of medical data and information visualization. Now a day's medical data is growing exponentially, the relationship between data become complicated more and more. Most of time we need to focus not only on the data itself, the hidden relationship between perceived data is also needed to focus on. Traditional relational database cannot meet this demand, Neo4j database is good at dealing with complex and multi-connection data, using Neo4j database to store and manage medical data makes it convenient

REFERENCES

[1]. S. Sencer and K. Eren. "Graph Database for Agent Based Emergency Response Model," Proceedings of the 2014 international conference on advances in big data analytics: July 21-24, 2014.

[2]. K. Nagi. "A New Representation of WordNet using Graph Databases" The fifth international conference on advances in databases, knowledge, and data applications. 2013:1-8

[3]. J. Webber. "A Programmatic Introduction to Neo4j," 2012 ACM conference on systems, programming, and applications: software for humanity, October 19-26, 2012, Tucson, Arizona.

[4]. Y. L. Wang. "A comparative study of graphic database NEO4J and relational database," Modern Electronic Technology, vol. 35, pp. 77-7, 2012.

[5]. L. Wang. "Construction of Document Resources Association Network Based on graph database technology," Digital Library Forum, pp. 59-65, 2014.

[6]. Y. L. Wang. "Research on embedded application of graphic database," Neo4j modern electronic technology, vol. 35, pp. 22, 2012

[7]. B.T. Messmer et al, D. Montag, "Understanding Neo4j Scalability,"NeoTechnology, USA, 2013.

[8]. [8] H. Darwen, An Introduction to RelationalDatabase Theory, United Kingdom: Bookboon. com, 2010.

[9]. J. Zhang, Z. Li and S. Liu, " Graph Search with Neo4j," Department of Computer Science, Georgia State University, Atlanta,USA, 2013.

[10]. Pere Burton , R. Angles, A. Prat-Pérez, D. Dominguez-Sal and J.-L. Larriba-Pey, "Benchmarking database systems for social network applications," in First International Workshop on Graph Data Management Experience and Systems, NewYork, USA, 2013

[11]. C. Vicknair, M. Macias, Z. Zhao, X. Nan, Y.Chen and D. Wilkins, "A Comparison of aGraph Database and a Relational Database," inACMSE 10', Oxford, MS, USA, 2010.

[12]. H. Florian and P. René, "Performance of Graph Query Languages: Comparison of

Cypher,Gremlin and Native Access in Neo4j," ACM, pp.18-22, 2013.

[13]. WisalKhan,Ejazahmed,WaseemShahzad, "Predictive Performance Comparison Analysis of Relational &NoSQL Graph Databases, "National University of Computer and Emerging Sciences,Islamabad, Pakistan, IJACSA, Vol**8,** No.**5**,2017

[14]. Mahmut, Yusuf Kavurucua, Alev Mutlub Mahmut Idea, Yusuf Kavurucua, Alev Mutlub, "Graph Representation of Relational Database for Concept Discovery", aTurkish Naval Academy ,Computer Engineering, Istanbul-34940, Türkiyeb Kocaeli University, Computer Engineering, Kocaeli-41000, Türkiye, Science Direct Procedia - Social and Behavioral Sciences **195 (2015) 1981 –1989**

[15]. Jaroslav pokorny, michalvalanta, jirikovacic, " Integrity constraints in graph databases", Faculty of Mathematics and Physics, Charles University, Malostranskénám. 25, 118 00 Prague, Czech Republic bFaculty of Information Technology, Czech Technical University, Thákurova 9, 160 00 Prague, Czech Republic, ScienceDirectProcedia Computer Science **109C(2017)975-981**

[16]. David W. Williams, Jun Huan, Wei Wang, "Graph Database Indexing Using Structured Graph Decomposition", 1Department of Computer Science University of North Carolina, Chapel Hill 2Department of Electrical Engineering and Computer Science University of Kansas.

[17]. ShaliniBatra, CharuTyagi, "Comparative Analysis of Relational And Graph Databases", IJCSE ISSN: **2231-2307**, Volume-**2**,Issue-**2,**may**2012**.