IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Graph Database: A Contemporary Storage Mechanism for Connected Data

N.M.A. Mohamed Ali[1], Dr. T. Padma[2]

Research Scholar, Bharathiar University, Coimbatore[1]

Professor, Department of Computer Applications, Sona College of Technology, Salem[2]

Abstract: The Relational databases have been used for several decades as the database technology for most traditional data-centric storage applications. Recently there has been much attention in graph data stores as an effective technology for highly connected data. This paper reports the basic concepts and technologies relating to NoSQL graph data stores and compares graph database with common relational database system. It also describes different types of NoSQL data stores and its applications.

Keywords: Graph Database, Relational Database, Data Model, No SQL.

I. INTRODUCTION

This article enlightens the basic concepts and technologies relating to graph database. The graph databases are rapidly calling the attention of the database community as an effective technology for deciphering meaningful information from graph structured data. The new phenomenon, big data is being produced by every digital process and transmitted by computer systems, mobile devices and sensors. Currently, high-volume of data is generated from millions of internet users via social network services like, Twitter, Facebook, online dating websites, LinkedIn and so on. The social networks maintain data relationships and get more interest on data interconnectivity than data itself. The new technology, Internet of Things (IoT) will connect millions of devices to the Internet and each device will be able to identify themselves to other devices. Here, we do not only need to manage high volumes of data, the relationships between data are more concern than the individual data points. The graph database architecture is well appropriate for this situation due to harnessing capability for inter connectedness of data and also flexible to add new kinds of relationships in easy way.

II. NEXT GENERATION DATABASES

The Relational databases have been used for several decades as the data-centric storage technology for applications such as: banking, payroll, inventory control systems and etc. These applications need small amount of data with a simple flat structure and perform straight forward operations. In the past few years database industries want to store, search and process unprecedented level of massive amounts of complex data whose structure is either hierarchical or a more general graph structure. Currently database industries are moving towards Next Generation Databases technologies in order to manage data which cannot be easily modelled by relational approach.

NoSQL (non SQL) databases are stated as Next Generation Databases and growing rapidly in recent years because of their high scalability and performance. NoSQL databases are, also known as cloud databases, non-relational databases or Big Data databases, designed to store analyse high volume of user-generated and machine-generated data. These databases address a wide variety of different techniques for storage and retrieval of semi structure, complex large volume of data and mostly non-relational and horizontally scalable.

The NoSQL databases are more scalable and give better performance, and tackle several issues that are not addressed by the relational databases. First, the data set that most enterprises handle today is huge volume in size and complex, which cannot fit into relational databases. Secondly, the relational databases require schemas that are to be defined before adding data. The schemas need to change often in order to fulfil the requirements of applications. The NoSQL databases are designed to add data without a predefined schema which allows developers to make significant application changes in real-time and makes software development faster and needs less database administrator time. Thirdly, the relational databases are not designed to manage the scalability and agility challenges of modern database applications.

The business enterprises like Google, Amazon, LinkedIn, and Twitter, struggle to cope with high volume of data and operation under tight latency constraints, use the NoSQL databases [1]. There are an extremely large number of NoSQL databases being developed and used in database industry. The most famous NoSQL databases include Google Bigtable, Apache Cassandra, MongoDB and Apache HBase. The NoSQL databases are classified into four basic categories that are Key-value stores, Wide-Column family stores, Document database and Graph database.

A. Key-value stores

A Key-value data store works in a very different approach to store data in the database as an attribute name, or key together with its value. In contrast to relational database

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

system, key-value systems handle entirely opaque collection of data which may have different fields for every record. This flexibility offers the application developer to store schema-less, polymorphic and unstructured data. This schema-less data model allows for easy scaling [2]. The key-value stores provide a single ways of retrieving data by primary key. The prominent key-value databases include: Dyanmo DB, Azure Table Storage (ATS), Riak andBerkeley DB.

B. Column store

A Column store database stores data tables as sections of columns of data for each row, rather than as rows of data by the predefined table structure. Columns can be grouped together for access in column families, or columns can be spread across multiple column families. Data is retrieved by primary key per column family. The Column store databases offer very high performance and a highly scalable architecture [3]. The most famous column store database is Bigtable and Cassandra. The Bigtable is a proprietary database of Google. Google uses Bigtable along with MapReduce computation engine for working with high volume of datasets.

C. Document database

A Document database is used for storing, managing and retrieving a complex data structure known as document which contains semi-structured data. The each document is assigned a unique key which is used to retrieve the document from database. The widely used document databases are MongoDB and Couch DB.

D. Graph database

A Graph database is based on the general-purpose data structure known as graph. The graph database is based on graph theory. These databases are designed for data that is structured as a graph and has elements which are interconnected, with an undetermined number of relations between them. Examples include: Neo4j and Titan.

III. COMPARISON OF RELATIONAL AND GRAPH DATA MODEL

A data model is a collection of conceptual tools that is used to model real-world entities and the relationships among them. A data model consists of three important components such as data structure types, query operators and integrity rules. The growth of large networks such as geographical systems, biological networks or social networks [4] needs to manage information with inherent graph data model. The graph data model implements graph-like data structure. The graph data model is more appropriate in the environment where data interconnectivity is at least as important as the data itself. This model supports more natural modelling of data and allows easy to accomplish graph queries, such as finding paths or graph pattern mining, which require SQL statements with complex join operations [5]. This model also eases better visualizing of the data.

The fundamental difference between graph data model and relational data model is that the graph data model stores the relationships at the individual record level, whereas the relational data model defines the relationship at the conceptual level. The relational data model imposes rigid schema which provides safety, but makes difficult to add a new relationship between two entities. Since the graph data model stores all the relationships of data along with data, there is no additional computing process required to join the data to extract useful information. Because of this type data modelling approach, the graph database is much faster than relational databases for highly connected data [7]. The new graph database technology is well suitable for search intensive approach, rather than data-centric approach, for huge volume of data [6]. There are two types of graph database: property graph database and RDF database.

IV. PROPERTY GRAPH DATABASE

A property graph database can store any kind of data using few simple fundamental units that are nodes and relationships [8]. The nodes are used to represent entities that can be named with zero or more labels. The labels are useful graph construct that is used to group the same kind of nodes. In addition to nodes, connections between nodes are called relationships. The relationships organize nodes into richly inter-connected arbitrary data structure and allow for finding related data. In property graph database, both relationships and nodes can have properties which are in the form of arbitrary key-value pairs. The keys are names as string and the values are arbitrary data types. The capability to define properties to relationships is helpful for adding extra metadata information for graph. The following figure 1 shows the each component of property graph [12].

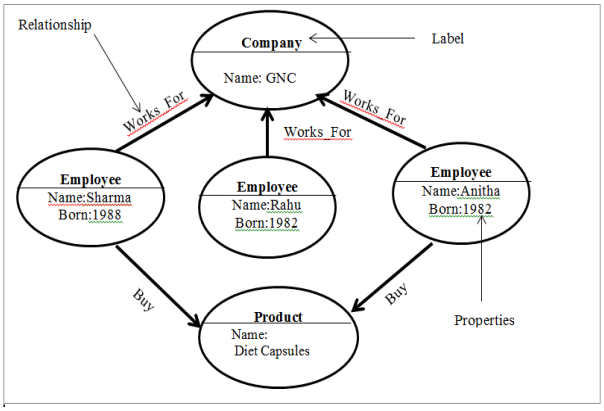

Fig 1.A Sample Property Graph

V. RDF DATA MODEL SEMANTIC GRAPH

A Resource Description Framework (RDF)is a semantic graph data store in Oracle database for highly connected data, Social Networks applications, media, life sciences and intelligence communities.

The RDF data model is based upon the idea of describing information in the form of RDF triples. The RDF triples represent information as subject–predicate–object model [9]. The RDF stores can be called as a sub class of graph

database system. Because the subject represents the resource, and the predicate represents behaviours of the resource and makes a relationship between the subject and the object. Moreover, most of the RDF data stores support SPARQL, and SQL-like query language for RDF data [10].

VI. APPLICATIONS OF GRAPH DATABASE

A. Master Data Management

Master data management (MDM**)** is a comprehensive process in which business intelligence analysts and data architects' work together to maintain the core entities of an enterprise. For example an ERP system for an enterprise includes a customer master, a suppliers master, an item master and an account master. This master data is one of the essential assets of an enterprise to identify the most critical information within an organization. The application developers are moving their master data in relational database to graph database to extract real-time business insights from master data and their relationship. The graph database is used to easily store and represent master data as a graph and perfectly develop the model to incorporate new data sources and types [11]. Using graph database, an enterprise can serve better product and service to their customers. They can get and retain more customers.

B. Graph-Based Search

A graph-based search offers a powerful approach to explore and discover information in enterprises that have high amount of products or content. Modern social network data, the spatial and temporal context data of a user and large knowledge resources data are highly structured in terms of relations and entities. So it needs huge multidimensional graph with flexible search techniques. Graph-based search has several competitive advantages to search and discover most meaningful and relevant information, whereas conventional relational database search approach retrieves diluted, random and lower quality information [11]. Already, Facebook's graph search and Google's knowledge graph have demonstrated better way to make this data available to users from huge volumes of content. Since the graph databases are more flexible and schema less, the companies can easily change their data structure, add new kinds of attributes and rework search query model.

C. IT and Network Operations

IT and Network operations management requires a database to manage the day-to-day activities of IT and Network infrastructure. In an organization, data is a complex entity in terms of size, structure and connectedness and network infrastructure is constantly changing in components and topology. Using a graph database, the enterprises easily manage the availability and performance of components and services in IT and networking environment. Since the graph database is a flexible graph model, it accurately represents highly interconnected physical and virtual network resources along with users and services. Apart from this, the graph database offers elegant simulation techniques to model network infrastructure and real-time suggestions to replace in case of any components failure in the environment. So managers can effectively analyse and manage the IT and network resources with the help of graph database.

D. Real-Time Recommendation System

Graph databases is the best revolutionized tool for real-time recommendations in areas such as general jobs, entertainment , movies, retail, industrial spare parts, even online dating and restaurants. A recommendation system is formulating recommendations based on multiple factors and multilevel of relationships between data in real time. For example, real-time recommendation system for online purchase needs to understand the customer's past purchase history, relationship between purchases, relationship between customers, and match customer with social-economic groups to look buying patterns. The graph database is most appropriate for this type of analysing relationships because it keeps both the data and relationships between them [8]. This gives semantically rich context for the data and avoids unnecessary join operations which are time consuming process in relational data model. Already graph database is used by many leading companies such as Walmart, eBay, Hewlett-Packard, Cisco, and National Geographic.

E. Social Data Analysis

As social networks are already graphs, a graph database is the sophisticated way to modelling, extracting different sub graph patterns for further cluster analysis, characterization, classification, discrimination and classification from this network [8]. Graph based analytics are efficient and inexpensive tool for the challenging issues such as creating small sample graphs from huge social graph, analysing the community structure of a network and gathering information from networks.

VII. CONCLUSION

Since relational databases do not strongly store relationships between data elements, they are not well appropriate for massive and highly connected data. The recent widespread of graph database technology is also a testament to the fact that graph data stores are one significant solution to consider in the management of complex and high volume of datasets. Graph databases use a flexible schema-less data model that allows modelling various domains in an easy way.

REFERENCES

[1] An Oracle White Paper, "Oracle NoSQL Database", September2011,http://www.oracle.com/technetwork/database/nosqldb/learnmore/nosql-database-498041.pdf.

[2] Giuseppe DeCandia, DenizHastorun, Madan Jampani, Guna vardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall and Werner Vogels , Dynamo: Amazon's Highly Available Key-value Store, Amazon.com (2007).

[3] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber, Bigtable: A Distributed Storage System for Structured Data, Googleinc.

[4] R. Angles, C. Gutierrez, Survey of graph database models, ACM Computer. Survey. 40 (2008) 1–39.
[5] R. Angles, A comparison of current graph database models, in: The 28th IEEE International Conference on Data Engineering Workshops, Arlington, Virginia, USA, 1–5 April, 2012.
[6] Chad Vicknair, Michael Macias, Zhendong Zhao, A Comparison of a Graph Database and a Relational Database, ACM (2010).
[7] ShaliniBatra, CharuTyagi, Comparative Analysis of Relational and Graph Databases, International Journal of Soft Computing and Engineering, May 2012.
[8] Introduction to Graph Databases.-L. Larriba-Pey, N. Martínez-Bazan, D. Domínguez-Sal. (2014) Introduction to Graph Databases, Reasoning Web 2014: 171-194.
[9] D.D. Abreu, et al., Choosing between graph databases and RDF engines for consuming and mining linked data, in: 4th International Workshop on Consuming Linked Data, Sydney, Australia, 22 October, 2013.
[10] J. Huang, D.J. Abadi, K. Ren, Scalable SPARQL querying of large graphs, in: The 37th International Conference on Very Large Data Bases, Seattle, Washington, USA, 29 August–3 September, 2011.
[11] JosepLluisLarribaPey, David Domonguez, Graph Databases and their applications (2014).
[12] Rik Van Bruggen, Learning Neo4J, PACKT publishing(2014).

DOI 10.17148/IJARCCE.2016.53220