

NoSQL Databases: A Survey on Schema Less Databases

Rachapalli Devendra Reddy, Asst. Professor,
Department of CSE
Srikalahasteeswara Institute of Technology (SKIT) Srikalahasti,
Chittoor (DT), Andhra Pradesh, India
e-mail : csdreddy@gmail.com

Abstract- NoSQL (Not only SQL) databases are called as non-relational or unstructured databases. These databases are open source databases. Basically horizontally scalable feature is useful in most of the cloud computing applications. These are to be able to store and process mass data effectively. Traditional relational databases are facing many problems in solving mass data storage and retrieving due to their structured schema. Unstructured databases does not follow the specific schema structure. This survey will realize the necessity in moving towards NoSQL databases. The survey on different types of unstructured databases gives a clear cut list and comparison on their nativity.

Index Terms- SQL schema, mass data , unstructured databases.

I. INTRODUCTION

In olden days, the schema level fitting of data stores has been argued by medical and internet based applications, that is lead to the revolution on a variety of alternate schema less databases. The movement of alternate data stores falls under the category of NoSQL. Not Only SQL(NoSQL) emphasizes that hey may support SQL like languages. Many NoSQL data bases compromise in consistence in favor of high availability and horizontal scalability than traditional data bases. The most important result of the rise of NoSQL is as follows

- No schemas and mostly open source
- No join operations
- Not using the Relational Databases

By using NoSQL databases developers can convert in-memory structures to relational structures. NoSQL databases were developed to execute efficiently on clusters. The rising of web as a platform, it also created a measurable factor change in data storage as to support massive volume of data by running on clusters. NoSQL DBases store each component separatly with a unique key. Some NoSQL databases provide high data throughput when compared with traditional RDBMSs. For example, the column-store hyper table, that pursues Google's Big table approach allows the local search engine Zvent for storing one million data cells in one day. NoSQL data bases better than rational databases in velocity, variety and volume but former is complex.



Figure1: Feature of NoSQL DBses

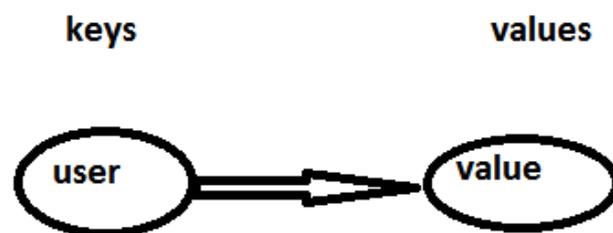


Figure2: Data in unstructured manner

II. TYPES OF DATA STORED IN NoSQL DATABASES

Structured Data

The data can be stored in database SQL in table with tuples and fields. They have unique key and it can be mapped to pre-designed fields. Now a days, these data are the most processed in development and the simplest way to manage information in small volume of store. Structured data is managed by using Structured Query Language (SQL) – it is a programming language developed for managing and querying data in RDBMS.

Unstructured data

Unstructured data can't be so readily classified and fitted into a neat box. Generally it can be used in photo, images, videos, streaming data, web pages, pdf files, PowerPoint presentations, emails, and word processing documents. Unstructured data is everywhere. In fact, most individuals and organizations conduct their lives around unstructured data. Unstructured data is either machine generated or human generated.

Semistructured data

Semistructured data are intermediate between the unstructured and structured. This is a type of structured data, but lagging in the strict data model structure. Tags or other types of markers are used to identify certain elements within the semi structured data, but the data does not have a rigid structure. For example, word processing software now can include metadata showing the author's name and the date created, with the bulk of the document just being unstructured text. Electronic mails have the sender, receiver, date with time and other fixed default fields added to the unstructured data of the e-mail message content and any attachments. Photos or other graphic images can be tailed with keywords such as the creator, date, location and etc. Extended Markup Language(XML) are often used to manage semistructured data.

III. TYPES OF NoSQL DATABASES

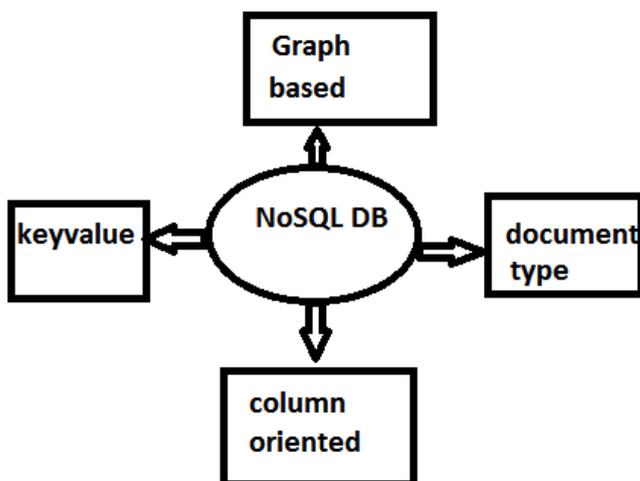


Figure3: types of NoSQL databases

There have been four approaches to classify NoSQL databases, each with different subcategories. Because of different types of approaches and overlaps it is difficult to maintain an overview of non-relational databases. However their basic classification is depends on data models.

A. Column-oriented :

These serializes all the values of a column together and next column and so on. Data is stored in cells grouped in columns of data rather than as rows of data.

Ex: smith:001,jhan:002,5500:003,4400:004,etc
 . It is a tuple consisting of three elements:^[1]

- Column : name of the column.
- Value : content of the column.
- Timestamp: Timestamp used to determine the valid content.
- Cassandra, Druid, HBase, are the examples of Column oriented storages.

B . Document store :

These are one of the main categories of NoSQL databases and the popularity of the term "document-oriented database" (or "document store"). Values stored as a document by compressed can provide encoded management structure. XML is one of the common encoding structure. These systems are designed around an abstract notion of a "document"[2].

Ex: Lotus Notes, Cluster point, Apache CouchDB, Couchbase, MarkLogic, MongoDB, OrientDB, Qizx e.t.c.,

C. Key-value store:

These databases work by matching keys with values, as similar to dictionary. These uses big hash table of keys and values.

Key	Value
"A.P"	{"P-16, Sector-28, India – 517640"
"Telangana"	{"T-17,sector-29,india-044-666"}

Figure4. Key-Value data store

- Get (key), value associated with the provided key.
- Put (key, value), associates the value with the key.
- Multi-get (key1, key2, ..., key N), list of values associated with the list of keys.
- Delete (key), removes the entry for the key from the data store.

Theses offer considerable flexibility and follows modern concepts like object-oriented programming.[3]

Ex: Amazon’s Dynamo, FoundationDB, MemcacheDB, Redis, Riak, etc.

D. Graph based:

There are three basic building blocks i.e node, relationship and property.

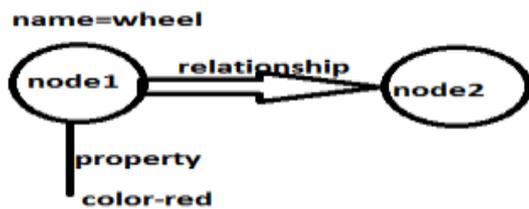


Figure4. Terminology of graph based data store

A graph database is any storage system that provides index-free adjacency [1]. These are well used in which clear boundaries are established. Data can be transformed from one model to other model easily. These use tree like structures. Graph databases are based on graph theory. Graph databases employ nodes, properties, and edges. These are mainly used in handling complex relational information.

Ex: Allegro, Neo4J, Infinite Graph, OrientDB, etc.

IV. ADVANTAGES

By comparing NoSQL databases with traditional SQL support databases, the following key benefits are addressed[4].

1. These are designed to address horizontal scalability
2. Supporting large volumes database management with dynamic schema of different types of data.
3. Easy to use and flexible due to object oriented programming.
4. These allow us to insert data into database without predefined schema.
5. These support auto-sharding, that automatic spread of data among arbitrary servers in the network in case of server down.
6. These are open source data bases.
7. Support high availability.
8. These uses separate cache layer for data integration.

V. DISADVANTAGES

Although NoSQL databases having several advantages to solve large volumes of data, there are some limitations[5].

1. Depending upon the database these are eventually consistence.
2. Many of these databases does not support transactions.
3. SQL can support queries beyond what NoSQL can do.
4. Very few people are familiar with operating them.

5. Support conditional atomicity.

VI. CONCLUSION

SQL data bases follows ACID properties where as NoSQL data bases follows BASE properties. NoSQL databases have becoming most popular over the recent years, due to the emergence of big data and cloud computing applications. In this paper it is shown that some of the benefits and disadvantages of NoSQL databases over RDBMS. Developer has to keep in touch with both the technologies to solve the real time problems, Although NoSQL databases are having some key advantages in particular to solve massive data problems, developers has to aware about the limitations of NoSQL databases.

REFERENCES

- [1] Chaker Nakhli. "Cassandra's data model cheatsheet:Data model elements: Column". Chaker Nakhli's Blog Yet another technical blog.Retrieved2011-03-29. A name value pair [consisting of a name and a value].
- [2] DB-Engines Ranking per database model category.
- [3] "Glossary of Big Data Terminology". Retrieved 5 august 2013
- [4] www.e-dba.com/mongodb.com
- [5] <http://www.bigdatacompanies.com>



Rachapalli Devendra Reddy, Asst. Professor, Department of CSE, Srikalahasteeswara Institute of Technology (SKIT) Srikalahasti, Chittoor (DT), Andhra Pradesh, India

e-mail : csdreddy@gmail.com