Comparative Study of SQL & NOSQL Databases
Prof. Ashwini M. Bhugul
1Department of Computer Engineering
1MMCOE, Pune-52

Abstract—In last few decades, rapid growth in using mobile applications, Internet leads to unstructured data. We cannot use SQL to store large amount of unstructured data. For that we have NOSQL databases.today NOSQL database is widely used to for big data. This paper gives comparative study of SQL and NOSQL databases with advantages of NOSQL databases over SQL.

Key words: NOSQL, SQL

I. INTRODUCTION

Database is organized collection of data. Someone wants to create database or someone want to add some information in database.today, there is tremendous use of mobile applications, blogs, internet surfing etc by the people. These all things lead to creation of unstructured data. Hence we required databases that can store unstructured data. Previously SQL is widely used to store structured data, using which we can maintain data. But to maintain unstructured data we toward NOSQL databases.databases are of two types relational & non-relational. Relational databases also called as SQL database & Non-relational also called as NOSQL databases.

II. SQL DATABASES

It is a relational database, collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Relational databases are designed to run on a single machine, it is based on relational model. This model organizes data into one or more tables (or "relations") of rows and columns, with a unique key for each row. SQL is a database computer language designed for the retrieval and management of data in relational database. SQL stands for Structured Query Language. SQL, is a standardized computer language that was originally developed by IBM for querying, altering and defining relational databases, using declarative statements. SQL can be used to create, insert, delete, update records. Also, to create stored procedures, to create views in database. Some SQL databases are oracle, MySQL, MYSQL SERVER, MS ACCESS, etc.

SQL databases uses ACID properties i.e. Atomicity, Consistency, isolation and durability.

Advantages of SQL Databases:
1) Reduced data redundancy
2) Reduced updating errors and increased consistency
3) Greater data integrity and independence from applications programs
4) Improved data access to users through use of host and query languages
5) Improved data security
6) Reduced data entry, storage, and retrieval costs
7) Facilitated development of new applications program.

Disadvantages
1) Complexity
2) Size.
3) Cost of database.
4) Cost of conversion.

A. Types of SQL Databases:

1) MYSQL:

It is an open source database, the most popular Open Source SQL database management system, is developed, distributed, and supported by Oracle Corporation. MySQL is a very powerful program in its own right. It handles a large subset of the functionality of the most expensive and powerful database packages. MySQL uses a standard form of the well-known SQL data language. MySQL works on many operating systems and with many languages including PHP, PERL, C, C++, JAVA, etc. MySQL works very quickly and works well even with large data sets. MySQL is very friendly to PHP, the most appreciated language for web development.

MySQL supports large databases, up to 50 million rows or more in a table. The default file size limit for a table is 4GB, but you can increase this (if your operating system can handle it) to a theoretical limit of 8 million terabytes (TB). MySQL is customizable. The open-source GPL license allows programmers to modify the MySQL software to fit their own specific environments.

2) ORACLE:

An Oracle database is a collection of data treated as a unit. The purpose of a database is to store and retrieve related information. Oracle Database is the first database designed for enterprise grid computing, the most flexible and cost effective way to manage information and applications. Oracle introduced many innovative technical features to the database as computing and deployment models changed. Oracle Database 10g Database Server describes the most recent major version of the Oracle Relational Database Management System.

Features of Oracle:
1) Database application development features.
2) Database connection features • Distributed database features
3) Data movement features
4) Performance features
5) Database management features

3) MS-ACCESS:

Microsoft Access, also known as Microsoft Office Access, is a database management system from Microsoft that combines the Relational Microsoft Jet Database Engine with a graphical user interface and software development tools.

The database component of MS Office is designed to act as an "intelligent" electronic filing cabinet. Like a manual filing cabinet, it enables you to

− collect sets of related data and keep the data organized;
Update the data once you’ve collected them, by adding or deleting records and by changing existing records. MS-Access is the most efficient database wherein user can create tables, can fire queries easily.

III. NOSQL DATABASES

NOSQL is not only SQL database. It is non-relational database, used to work on large datasets. It is open source, distributed high performance database. Today’s industries using NOSQL database to manage unstructured database. It is schema less, does not use row-column structure, and joins like SQL. It uses BASE property and CAP theorem. BASE stands for Basically, Available, Soft state, and Eventual consistency. CAP stands for Consistency, Availability and Partition tolerance. Due to these properties NOSQL gives high performance and hence it is popular now a day’s. Some of the companies like facebook, google, twitter uses NOSQL database. Some cores of NOSQL database are:

A. Key-Value Stores:

Key Value (KV) databases uses It is one of the simple data model among all (we will discuss later) as it uses only key and a value. It handles huge data load. It scales to large volume of data.

B. Column Family Stores:

Here are two types of column oriented databases whose detail is as given below:

1) Wide-Column data stores: I Wide Column data stores are those databases that are used for processing of web, streaming of data and documents.

2) Column oriented databases: these databases are column oriented in which in which all the values containing columns are put together. Example is shown below:

C. Document Databases:

Document Stores databases are those NoSQL databases which use records as documents. This type of database store unstructured (text) or semi-structured (XML) documents which are usually hierarchal in nature. Here each document consists of a set of keys and values which are almost same as there in the Key Value databases. Characteristics of Document Stores Database

Documents are addressed in the database using key (unique) that represents that document.

1) There are number of varieties to organize data that is collections, tags, non-visible metadata and directory hierarchies.

2) In this we can use a key-value lookup to retrieve a document.

3) Example of Document oriented database:

{officeName:”3Pillar Noida”,
{Street: “B-25, City:”Noida”, State:”UP”, Pincode:”201301”}
}

{officeName:”3Pillar Timisoara”,
{Boulevard:”Coriolan Brediceanu No. 10”, Block:”B, 1st Floor”, City: “Timisoara”, Pincode: 300011”}
}

4) Graph Databases

Graph databases are based on the graph theory. In general, we see that graph usually consists of nodes, properties and edges. NoSQL Graph database consists of:

- Nodes represent entities
- Properties represent attributes
- Edges represent relationships.

Characteristics of Graph databases:

- Graph traversals are executed with constant speed independent of total size of the graph. There are no set operations involved that decrease performance as seen with join operations in RDBMS. Graph databases are having high performance in context to their deep traversals.
- These are used for shortest path calculations. These are scalable. But its complexity increases.

Figure shows an example of graph database.
A. **MongoDB:**

MongoDB is a document database that provides high performance, high availability, and easy scalability. It is an open source database widely used to work on large data. It gives high performance as it is using indexing, aggregation, load balancing, etc.

The basic idea is to replace the concept of a “row” with a more flexible model, the “document”. Features MongoDB support Bson data structures to store complex data types; supports powerful and complex query language; high speed access to mass data; stores and distribute large binary files like images and videos. MongoDB is optimized for CRUD operations.

- **Features of MongoDB:**
  1) **Document-Oriented:** Instead of taking a business subject and breaking it up into multiple relational structures, MongoDB can store the business subject in the minimal number of documents.
  2) **Ad Hoc Queries:** MongoDB supports search by field, range queries, regular expression searches. Queries can return specific fields of documents and also include user-defined JavaScript functions.
  3) **Indexing:** Any field in a MongoDB document can be indexed. Indexing provides efficiency in searching of data.
  4) **Replication:** MongoDB provides high availability with replica sets. A replica set consists of two or more copies of the data.
  5) **Load Balancing:** MongoDB scales horizontally using sharding. The user chooses a shard key, which determines how the data in a collection will be distributed.

Today NOSQL databases are playing model role in industries or socially.

B. **CouchDB:**

CouchDB is open source database widely used. It is document oriented database that uses JSON to store data. CouchDB does not store data and relationships in tables. Instead, each database is a collection of independent documents. Each document maintains its own data and self-contained schema. CouchDB also supports for mapReduce replication, security, etc features like MongoDB.

Documents have the advantage of data being already conveniently packaged for storage rather than split out across numerous tables and rows in most database systems. When documents are committed to disk, the document fields and metadata are packed into buffers, sequentially one document after another. When CouchDB documents are updated, all data and associated indexes are flushed to disk and the transactional commit always leaves the database in a completely consistent state.

We use NOSQL databases when we have to work on unstructured data sets as, using relational database we are not able to read, write, update unstructured data such as images, blogs, videos, we have NOSQL databases for that.

1) **SQL versus NOSQL**

SQL databases having Schema where as NoSQL is Schemaless it is not possible to read images, videos, etc. by using SQL. Scaling of SQL is Vertical where as for NOSQL it is horizontal. SQL uses ACID properties and NOSQL uses BASE properties. Today’s world is the world of Computer, internet, etc. to access big data we require to study NoSQL databases. Using that we can achieve the good performance.

IV. **CONCLUSION**

This paper reviews Comparative study of SQL and NOSQL databases with examples. It is observed that NOSQL databases are more useful for big data applications.

Google’s core AdWords business is now running completely on F1. F1 provides the SQL database functionality.

To work with big data sets NOSQL database like MongoDB is good. Using such databases we can achieve good performance in database arena.

**REFERENCES**


[3] Bogdan George Tudorica, Cristian Bucur, “A Comparison between several NoSQL Databases with comments and notes”.


[6] www.mongodb.org.in