

Comparative analysis of relational and non- relational databases

[Science](#), [Computer Science](#)



Ever since 1970s, Relational database has been the foundation of enterprise applications and it has been popular and inexpensive since the release of MySQL in 1995. Yet in recent years, an increased number of companies have adopted different types of non-relational database (MongoDB, Cassandra, Hypertable, Hbase/Hadoop, CouchDB etc.), commonly referred to as NoSQL database. A NoSQL technology like MongoDB is not only used for new applications but also to augment or replace existing relational databases. This Research mainly focuses on one of the new technology of NoSQL database i. e. MongoDB, and makes a comparison study with one of the relational database i. e. MySQL and thus justifies why MongoDB is liked over MySQL. I will also describe the advantages and disadvantages of using relational & non-relational databases. A comparison criterion includes theoretical differences, characteristics, limitation, integrity, distribution, system requirements, and architecture, query and insertion times.

Motivation

With the explosion in volume and variety of data due to increased mobile and web applications, the popularity of using an efficient database system to cater the needs of business has become an essential part. With that in mind, “ Public Transport Victoria (PTV)” has made a decision to rebuild their Public transport timetable which is used to cater their website – “ ptv. vic. gov. au” as well as mobile applications.

Public Transport Victoria is a Government website of Victoria State, Australia, which provides their users with the ability to access the daily time table of Buses, Trains and Trams throughout the State. With millions of users using

this facility, PTV processes one million time tabling queries per day and forecasted a steady growth of 10% per year for next 5 years.

This Research is done to determine the best suitable technology, which caters the need to handle this large amount of data more efficiently, between Relational databases (MySQL) and NoSQL (MongoDB) databases.

Introduction

Databases are defined as “ collections of data”. Although when using the term database we refer to the complete database system, it refers only to the collection and data. “ The system which handles Big data, transactions, problems, database engines, database schemas is called the Database Management System (DBMS)”.

In order to satisfy the need of storing and retrieving data, Databases were created. Different types of databases have invented their inception in the 1960's, each using its own data representation and different technology for handling queries & transactions. “ They began with navigational databases which were based on linked-lists, moved on to relational databases with joins, afterwards object-oriented and without joins in the late 2000s NoSQL (MongoDB, Cassandra, Hypertable, Hbase/Hadoop, CouchDB etc.) emerged and has become a popular trend”.

Relational databases are widely used in most of the applications and they exhibit great performance when they deal with a limited amount of data. For data with large volume like internet, multimedia and social media, traditional relational databases is ineffective. To overcome this problem the “ NO SQL”

term was introduced meaning, namely “ Not Only SQL”, which is a lenient variant of the term, compared to its previous significance, the anti-relational. NoSQL is a methodology and not a tool, composed of many interdependent tools. The primary benefit of a NoSQL database is that, unlike a relational database, it can handle unstructured data such as documents, email, multimedia and social media efficiently. Non-relational databases do not use the RDBMS principles (Relational Database Management System). There are four strategies for storing data in a non-relational database as follows:

1. Key-Value
2. Document: e. g. MongoDB
3. Column/Field
4. Graph-Oriented.

Also, non-relational databases provide high flexibility for insertion or deletion of an attribute from the database because of the fact that they don't have a fixed database schema. In this research we concentrate on one of the NoSQL technologies, namely MongoDB, and make a comparison with MySQL to highlight why MongoDB is more capable than MySQL to cater the needs of “ Public Transport Victoria (PTV)”.

Research & Analysis

There are many differences between relational databases and NoSQL, all of them are important to understand before making a decision about best data management system. These include differences among:

Language

SQL databases use SQL (Structured Query Language) for defining and manipulating data. This allows SQL to be extremely versatile and widely-used; however, it also makes it more restrictive. It requires that you use pre-defined schemas to determine the structure of your data before you even begin to work with it. A NoSQL database has a dynamic schema for unstructured data and the data can be stored in many different ways as stated in Introduction. This flexibility allows you to create documents without having to carefully plan and define the data structure or schema and add fields as you go. Scalability: Most SQL databases are vertically scalable, which means that you can increase the load on a single server by increasing components like CPU, SSD or RAM. On the other hand, NoSQL databases are horizontally scalable, which means that they can handle more traffic/load simply by adding more servers to the database. They have the ability to become larger and much more powerful, which makes them the preferred choice for constantly evolving or large data sets. This caters the need of PTV to handle queries efficiently and powerfully with the expected annual increase over next five years without failing.

Community

Due to SQL being in market for more than 40 years, it has a much larger, stronger and more developed community compared to NoSQL. There are thousands of chats and forums available where experts can share knowledge and discuss SQL best practices, which continuously enhance the skills.

Although NoSQL is growing rapidly, its community is not as good as SQL due to the fact that it is still relatively new.

Structure

SQL database are table-based which puts them as a better option for applications that require multi-row transactions. For Example, accounting systems or even legacy systems that were originally built for a relational structure. As stated earlier, NoSQL databases can be key-value pairs, wide-column stores, graph databases, or document-based.

NoSQL**Advantages**

- Non-Relational means table-less
- Mostly Open Source and Low-Cost: An appealing solution for smaller organizations with limited budgets.
- Easier scalability through support for Map Reduce: NoSQL databases are designed to function on full throttle even with low-cost hardware.
- No need to develop a detailed database model: This saves a lot of development time.
- Speed: It's high-performing for simple queries, as all the related data are in single document which eliminates the join operations.

Disadvantages

- Lack of reporting tools for analysis and performance testing.
- Lack of standardization in Query Language.

Properties of Both:

SQL databases follow ACID properties (Atomicity, Consistency, Isolation and Durability) while the NoSQL database emphasizes on the Brewers CAP theorem (Consistency, Availability and Partition tolerance).

To Sum Up:

Relational Databases or MySQL is a good choice for any business who has pre-defined data structures and schemas. Relational databases are generally better for ACID level Transactions and for the systems whose schema doesn't change often. MongoDB or other NoSQL databases are used in the business with rapid growth or with the databases with no unambiguous schema definitions or if your application has large amount of queries to be performed in efficient way.

Conclusions

“ If your business is not going to experience significant growth in the near future and your data is structured, SQL is the right choice for your business. But if you desire rapid processing of data and you don't have transactional data to protect, NoSQL is your go-to solution”.

Keeping in mind the case study and performance evaluation for both paradigms, comparing above research & analysis and summary, as the “ Public Transport Victoria (PTV)” don't deal with any Transactional level scenarios and it is expecting to have a significant growth in the future, and that needs a faster and efficient processing of large amount of data, I would recommend MongoDB as their backend database server.