Goel, Ajay: Nu Id (001897443)

INFO 7250
Engineering Big-Data Systems
Summer Full 2019
Assignment 1

Comparing NoSQL MongoDB to an SQL DB

NoSQL database solutions are becoming more and more prevalent in a world currently dominated by SQL relational databases.

- NoSQL databases were designed to provide database solutions for large volumes of data that is not structured.
- There are not many studies that compare the performance of processing a modest amount of structured data in a NoSQL database with a traditional relational database.

The data is typically non-structured, complex and does not fit well into the relational model.

- Data in the relational model is usually represented by a database schema, in order to capture the semantics of the database.
- Objects in the database with the same number of characteristics, type and format are grouped together, making it structured data.
- The relational model is built on this assumption of structured data, with its data stored in the rows and columns of a table, whereby each row has the same number and type of data columns.
- Tables in relational databases are typically normalized, which results in the creation of multiple tables.
- Querying those tables requires fetching and combining information from the many different tables.
- Combining information based on a matching value for a primary key and foreign key across multiple tables in the relational model requires using a join operation.
- The larger the schema and the more tables that need to be joined, the longer it takes for the relational database to fetch the data.
- NoSQL can help deal with data that is not structured.
- Data can be semi-structured, such that similar data objects can be grouped together, but the objects may have different characteristics.
- Schema information may also be mixed in with data values in semi-structured data, such as found in XML data. Unstructured data can be of any type and may have no format. This data cannot be represented by any type of schema, such as web pages in HTML.
- The typical features of SQL databases, such as the ACID properties, require a certain amount of overhead and are relaxed or eliminated in NoSQL databases to maximize performance.
- Many NoSQL databases organize the data into key-value pairs.
- The key is used to uniquely identify a particular data item and the value can be a simple word, number or a complex structure with unique semantics.
- The development of queries is more complex, there is no standard query language, and there are limits to the operations. Specifically, there is no join operation. However, in general processing is simpler, more affordable and more flexible.

GIST:

- SQL Database is a Relational Database and a structured one whereas NoSQL is Non-relational database likely to be more document and distributed than structured.
- Relational database strictly adheres relations where it is divided into the set of rows and columns to store data often named as tables, but the Non-relational database has a document-oriented or distributed storage which doesn't require any table structure.
- NoSQL databases have a dynamic schema for document type or unstructured data whereas SQL Databases have a well-designed pre-defined schema.
- SQL databases are vertically scalable whereas NoSQL databases are horizontally scalable. You can scale the SQL databases by expanding the strength of its hardware. Similarly, when it comes to NoSQL databases you can scale it by expanding the databases servers in the pool of assets to lessen the heap.

- While SQL uses structured query language for defining data, NoSQL uses the collection of documents which is also known as UnQL (Unstructured Query Language).
- NoSQL uses hierarchical data storage, and there is no hierarchical data storage for SQL.
- We can easily add the new data in NoSQL without requiring prior steps whereas SQL might require doing some changes like backfilling data, altering schemas.
- Since SQL has a standard interface for handling complex queries, it would be great to deal with complex queries, NOSQL don't have any standard interface, so it's quite difficult to handle complex queries in NoSQL.