

SQL vs NoSQL

A study on SQL and NoSQL Databases

By: Ciera Workman

Date: February 7th, 2019

Link to Page: <https://spark.adobe.com/page/ba6tREeqUFyen/>

The Differences Between SQL & NoSQL Databases

There are a variety of differences between a SQL and a NoSQL database. The dissimilarities include: language, scalability, and structure. A relational database is purely structured by tables, while a non-relational database has more options like document-oriented, graph-based, column-oriented, and KeyValue pairs. A SQL database must have a defined schema to create the structure of the data before someone can work with it. NoSQL does not require a predefined schema. Any document in a non-relational database may have any structure, may have fields added to it at any time, and each document may have its own syntax with its own structure (Xplenty, 2017).

A NoSQL database may be scaled by adding more servers to the database to handle the increased traffic. A SQL database server usually has to be scaled by increasing the physical capabilities of the system it is running on such as the CPU, RAM, or SSD. One can copy a relational database across more than one node as well. The ease of scaling has made a non-relational database popular in data sets that change often or are very large (Xplenty, 2017). A NoSQL approach is good for taking advantage of cloud storage to store large amounts of unstructured data and for rapid development (Wodehouse, n.d.).



Twitter Features

The tweet is a feature of Twitter that could be included in a NoSQL database. A tweet is the main way of communicating within the social media platform. It can contain a combination of text, images, links, or videos. Carey Wodehouse of UpWork describes a NoSQL database like a file folder on one's computer that brings together related data of all types. By using this way of thinking for a non-relational database, it would make sense that Twitter could store each tweet as its own document within a folder with the poster's username as the unique ID to retrieve them.

Pros: A tweet is an easily sharable feature. It allows information to be passed around very quickly to a wide audience. A user can see tweets from anyone in the world with an account. A user may add multiple photos to their tweet. Twitter allows protected tweets so users can control who sees their tweets.

Cons: The character count for a tweet is limited to 140. This causes users to have to reply to their own tweets in order to tell a story or share related information in an orderly fashion. A user cannot add multiple .gifs to their tweet. A protected tweet's images can be relinked in order to be shared in an unprotected way.



A reply to a tweet is very similar to creating an initial tweet. It has the same style capabilities such as text, image, .gif, video and adding a poll. When a user replies to someone they create a thread of responses in chronological order. The difference being that a reply is identified both to the user who makes it and the original tweet. A NoSQL database could create a separate folder using the parent tweet as the identifier with all the replies contained inside it. If the parent tweet is deleted then the replies are as well, but the user would still exist.

Pros: A user can share their thoughts and reactions to a tweet with ease. No functionality is lost between a main tweet and a reply. Both options allow text,

images, .gifs etc.

Cons: Once a parent tweet is deleted, all replies are too. A reply is still limited by a specific amount of characters.



Reposting a tweet could be handled by a NoSQL database. Since a retweet is a copy of the original tweet or retweet with the possibility of additional commentary, it would be best handled by a database that can create relationships on short notice. Someone who uses this feature would create a link between their repost and the original one. Each tweet that is reposted would then carry a connection between each user and their copy.

Pros: A retweet allows users to have a direct reference to information. There is no limit to how many times a tweet could be retweeted. A reposted tweet contains the entire thread of replies to each original tweet. A user can retweet a reply.

Cons: A retweet may not have any additional images or .gifs added to it.

Emoticons are the only images a user may use.

Email or Phone

Password

Log In

Forgot account?

Create a New Account

It's free and always will be.

Facebook Feature

A feature of the social media site Facebook that could be included in a relational database is the login and registration process. Logging in requires a hundred percent match to a user's email and of their password. The registration procedure would insert a new user into the database after filling out the required fields. With an organized SQL database a query such as determining if a login was successful or not is concluded quickly. A SQL database would be helpful in remembering specific devices a user utilizes.

Pros: Facebook allows two-step authentication. A user may customize their settings to get alerted when a login or attempt to login has occurred on an unrecognized device. Registration for a Facebook account is simple and takes very little time.

Four Types of NoSQL Databases

The Key-Value Pair is a type of NoSQL Database. Information is stored as a key/value pair which is retrievable by the unique key. The value can be anything, like user preferences or session data. The database does not have to understand the data being stored. This type can be scaled with ease. It uses a hash table to store the keys. This database lacks any column type relations as well (Vishwakarma, 2017).

A Document NoSQL Database is incredibly similar to the key-value type as data is a value which is identified with a unique key. However, the data document stored is semi-structured or structured. Often these documents are in XML, JSON, or BSON formatting (Vishwakarma, 2017).

A Column Store is a database which stores data as columns instead of rows. They are organized into column families. Each family contains rows that have a

unique identifier and these rows contain columns. Every row can have a differing number of columns and none of the columns have to match the other rows. The columns have a key-value structure that is paired with a timestamp (Database.Guide, 2016).

A graph based database is made of nodes and edges. A node is data while an edge is the relationship between two nodes. This type shows all the relationships between all the stored data. A graph database can create relationships between two nodes on short notice (Vishwakarma, 2017).

Five NoSQL Databases

1. Cassandra is considered a column store database. It boasts that anyone who uses it will not lose data during outages. It uses linear scalability. The database promises no bottlenecking as all data is copied across all nodes. There are familiar names that utilize Cassandra including Netflix, Hulu, Instagram and Comcast (Cassandra Apache, n.d.).

2. MongoDB is a document store database. Documents are a lot like JSON, which allows each document to be unique. MongoDB has support for over ten languages not counting the community created ones. This database scales

horizontally. This type of database allows for cloud storage to be utilized. MongoDB is free and open source (MongoDB. 2019).

3. Neo4j is the most popular graph based database available. It is a great database for social media platforms, network management, and fraud detection as graph databases focus on relationships between data. Neo4j has a sandbox feature to allow users to interact and get a feel for their type of database. It has support for several popular language such as Java, C#, and Python. This database uses Cypher Query Language. Neo4j provides free and discounted prices for startup businesses. They have a variety of bundles to choose from to meet a business' needs. Neo4j is ACID compliant. This database boasts short read time even with large amounts of data. Some familiar names that use Neo4j are Walmart, eBay, Adobe, and NASA! (Neo4j, n.d.)

4. Oracle NoSQL Database has a variety of features. Data can be modeled like as relational tables, JSON documents, or key-value pairs. According to the official Oracle website, "Oracle NoSQL Database is a sharded (shared-nothing) system which distributes the data uniformly across the multiple shards in the cluster, based on the hashed value of the primary key." The nodes within each shard are duplicated to assure availability. This database does have a cloud service. That service does allow a user to bring their own license. (Oracle, n.d.)

5. Azure is a document store database. It is also considered a multi-model database that can handle graph, key-value, and column models. It uses an auto indexing feature. A user only pays for what they need to use. This database scales horizontally. Pricing includes \$0.25 a month per GB used, \$5.84 a month starting single region provision writes per 100R/s, and \$11.68 a month for multiple region provision writes per 100R/s. A user may reserve capacity for one to three years for a one time fee at a significant discount. The familiar names that use Azure are Bank of America, Intel, and Ford.

NoSQL Weather App Implementation

A database type that a weather app could utilize would be column based. Each day's weather could be its own column within a column family based on area. The rows contained in the column could store information on what kind of weather, how much rain or snow, and temperature range. Since a column based store does not require the rows contained in columns to be the same type, it is easy to contain varying data.

Another type a weather app could implement would be key-value based. This would be an excellent option to handle the visuals of the weather app. Since weather situations can be accessed based on a key, like rain, the database

could retrieve a raining animated background to go alongside the other information.

References

C. (2019, February 09). Test Tweet. Retrieved February 09, 2019, from https://twitter.com/_Maroon_Rose/status/1094163624736817153

Database.Guide. (2016, June 23). What is a Column Store Database? Retrieved February 10, 2019, from <https://database.guide/what-is-a-column-store-database/>

Facebook. (n.d.). Facebook Login. Retrieved from <https://www.facebook.com>

Microsoft. (n.d.). Microsoft Azure. Retrieved from <https://azure.microsoft.com/en-us/>

MongoDB. (n.d.). Open Source Document Database. Retrieved February 10, 2019, from <https://www.mongodb.com/>

Neo4j. (n.d.). Neo4j Graph Platform – The Leader in Graph Databases.
Retrieved February 10, 2019, from <https://neo4j.com/>

Orcale. (n.d.). Orcale. Retrieved February 10, 2019, from
<https://www.oracle.com/technetwork/database/database-technologies/nosqldb/overview/index.html>

The Apache Software Foundation. (n.d.). Apache Cassandra. Retrieved February 10, 2019, from <https://cassandra.apache.org/>. (2019, February 10).

Titan's Sky #Space #Art by Fred Gambino pic.twitter.com/kPzY34be2V.
Retrieved February 10, 2019, from
<https://twitter.com/GhostPlanetTV/status/1094774348341264389> Vishwakarma,
R. (2017, May 16).

The Different Types of NoSQL Databases. Retrieved February 10, 2019, from
<https://opensourceforu.com/2017/05/different-types-nosql-databases/>

Wodehouse, C. (n.d.). SQL vs. NoSQL Databases: What's the Difference?
Retrieved February 9, 2019, from <https://www.upwork.com/hiring/data/sql-vs-nosql-databases/>

Xplenty. (2017, September 28). The SQL vs NoSQL Difference: MySQL vs MongoDB – Xplenty Blog – Medium. Retrieved February 9, 2019, from <https://medium.com/xplenty-blog/the-sql-vs-nosql-difference-mysql-vs-mongodb-32c9980e67b2>



CREATED BY
Ciera Workman

Credits:

Created with an image by geralt - "binary binary system data"