
SQL Vs NoSQL

By: Blain Bowman

Date: 03/03/2020

Link to Adobe Spark:
https://spark.adobe.com/page/aOjN9UPA0Fdm1/

Differences between SQL and NoSQL

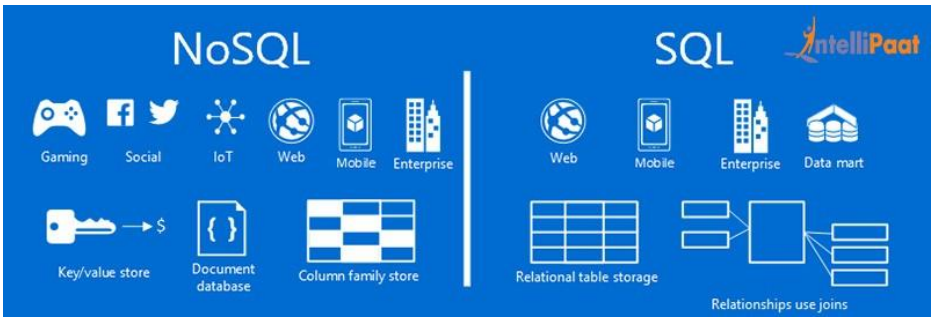
The SQL databases define and manipulate data based on the structured query language, making it extremely powerful and one of the most versatile and mostly preferred available options particularly for significant complex queries (Issac, 2014). On the other hand, a NoSQL database entails dynamic schema for unstructured data providing the opportunity to store information in various ways such as document-oriented, graph-based, or column-oriented.

SQL databases are vertically scalable, meaning that a load can be increased on a single server by advancing features such as RAM, SSD or CPU. NoSQL databases, on the other hand, are horizontally scalable, whereby more traffic can be handled only by adding more servers in your NoSQL or sharding.

SQL databases are table-based while NoSQL databases are either document-based, key-value pairs, wide-column stores or graph databases (Sharma & Dave, 2012). Therefore, this gives SQL databases the upper hand concerning applications that require multi-row transactions like legacy systems or an accounting system that were specifically meant for a relational structure.

SQL databases preferably follow the ACID property, which is Atomicity, Consistency, Isolation and Durability; whereas the NoSQL database strictly follows the Brewers CAP theorem, which is Consistency, Availability and Partition tolerance.

SQL database entails great support since numerous independent consultations are skillful and can handle extremely large-scale relational databases effectively. As for the NoSQL database, and individual has to depend on community support since the availability of the outside experts is limited; hence, making it difficult to set up and deploy large-scale Non-relational database.



Difference between SQL and NoSQL - The Crazy Programmer. (2020). Retrieved 8 March 2020, from https://www.thecrazyprogrammer.com/2018/04/difference-between-sql-and-nosql.html

Three NoSQL Twitter Features

Hadoop: Twitter requires storing more data per day forcing them to store data on the cluster, hence, they use Hadoop distribution to power and enhance its cluster. This feature is significant in running analytics and hit FlockDB in parallel to accumulate social graph aggregates.

Scribe: Twitter applies Scribe to write logs to Hadoop since it made it easy for Twitter to log a significant amount of data, at 80 diverse categories of data (Finley, 2011). Therefore, Scribe proves to be an

important feature in applying the NoSQL database on Twitter.

FlockDB: This is a distribution database feature used by Twitter to perform social graph analysis. FlockDB determines which users to show @ replies to; therefore, this makes it easy for Twitter to send Tweets and trending messages only for the particular followers who follow the person.

Twitter Features Pros and Cons

Hadoop: Hadoop is efficient enough since it saves time as it can obtain valuable data from any form of data within the shortest time possible. The feature also provides other opportunities and functions for derived information such as fraud detection, data warehousing, and market campaign. Hadoop is cost-effective for big companies that value a proper and sustainable data storage process (KnowledgeHut, 2016). The firm can refer to their raw data if in case they decide to change the direction of its processes in the long-term. In addition, organizations can perform their work at a faster and most efficient rate since Hadoop provides huge storage systems. Data is not lost when using Hadoop since it automatically

duplicates the existing data and creates numerous copies.

The security measures often become disabled by default when using Hadoop, and this may cause the data analytics to lose valuable data. Hadoop cannot efficiently perform in a relatively small data environment; therefore, only the large businesses can utilize the feature since they generate big data.

Scribe: Scribe is reviewed as very fast and simple to operate with, and its error handling features enables one to know the accuracy of the data obtained. Scribe is efficient when handling numerous databases with a large number of connections, and the availability of the user interface is reliable.

The response time when debugging using Scribe is reported to take a very long time to step through a map, making it tedious. Scribe is costly; hence, limiting the number of users across the field as well as the amount of data to be handled.

FlockDB: FlockDB is easy to learn and apply, making communication with the team simple since it provides individuals with the intended message. Managing projects within an organization is easier with the

option of applying the integrated tools found in the feature.

The amount of notifications found in this feature can be overwhelming to some users. In addition, finding the message sent or image posted can be hard since not most of the files are easily accessible.

Facebook SQL Feature

The SQL feature that Facebook uses is known as Hive. This is an open-source mega-scale data warehousing framework that is based on Hadoop, mainly to support tables as well as a variant of SQL called the hiveQL (Facebook, 2009). Hive has provided the opportunity for Hadoop users to attain their data processing needs by accessing large data efficiently and within the shortest time.

Facebook Feature Pros and Cons

Pros: Facebook is able to run numerous of jobs on the cluster with thousands of clients using this system for a broad variety of applications. By applying Hive/Hadoop cluster, the social media company is

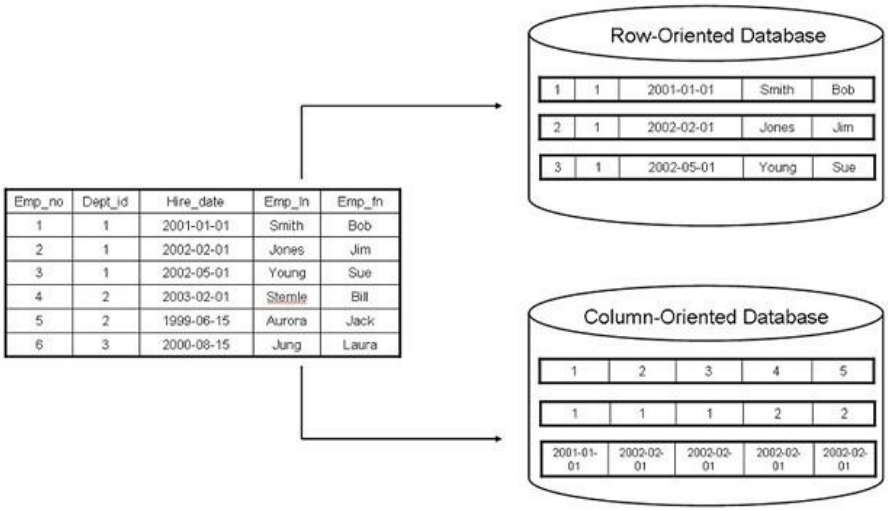
able to store more than 2PB of uncompressed data as well as routinely loads 15TB of data daily. Hive/Hadoop feature is significant for summarization of jobs, machine learning, and business intelligence among many other applications.

Cons: Using Hive on Hadoop is not easy for end users, particularly for those who are not familiar with the map/reduce. More time is taken in writing the popular query languages making it difficult to process and analyze data.

Four NoSQL Database Types

i. Column-Oriented Database

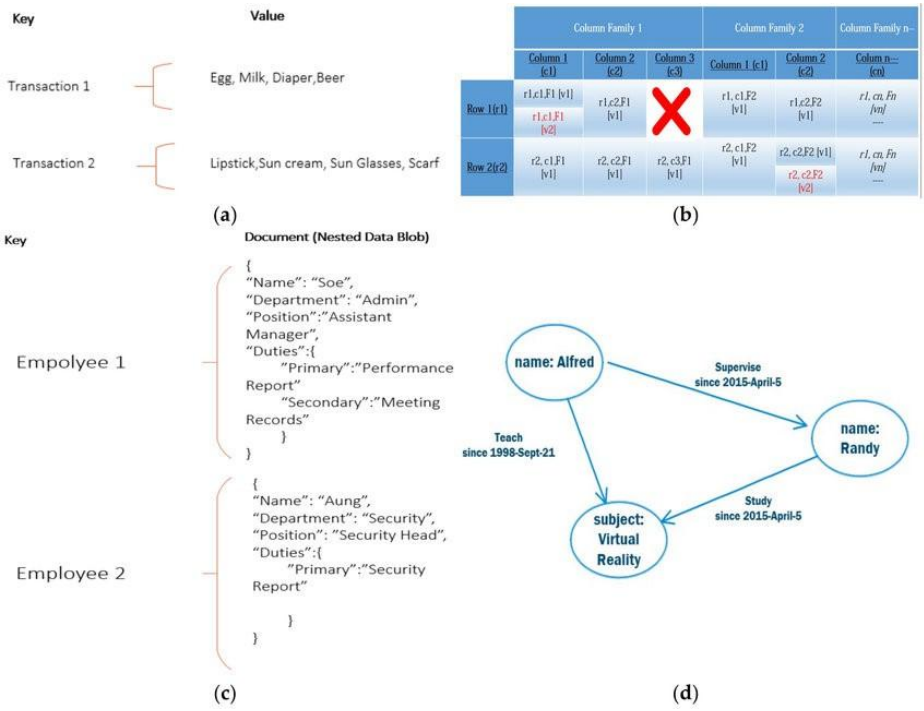
Column database store each column separately to allow quicker scans by mapping the data to the row numbers; hence, performing analytics and reporting summating values and identifying entries efficiently (Nayak, Poriya & Poojary, 2013).



Column-Oriented Database Technologies | DB Best Chronicles. (2020). Retrieved 8 March 2020, from https://www.dbbest.com/blog/column-oriented-database-technologies/

ii. Key-Value Stores

This is the least complex means of a database comprising of a collection of key-value pairs enabling the storage of large amounts of data.



Types-of-NoSQL-data-stores-a-Key-Value-Store-b-Columnar-Store-c-Document-Store. (2020). Retrieved 8 March 2020, from https://www.researchgate.net/figure/Types-of-NoSQL-data-stores-a-Key-Value-Store-b-Columnar-Store-c-Document-Store

iii. Document Stores

Document stores are designed to specifically store daily documents, also allow for the performance of complex calculations on the already aggregated form of data (Meysman, 2016). The documents are stored only once and automatically connecting via foreign keys.

iv. Graph Databases

Graph databases are designed to store relations between entities in the most efficient manner, making it to be the most complex NoSQL database (Meysman, 2016). The feature can perform with both entities of the same nature and two diverse entities.

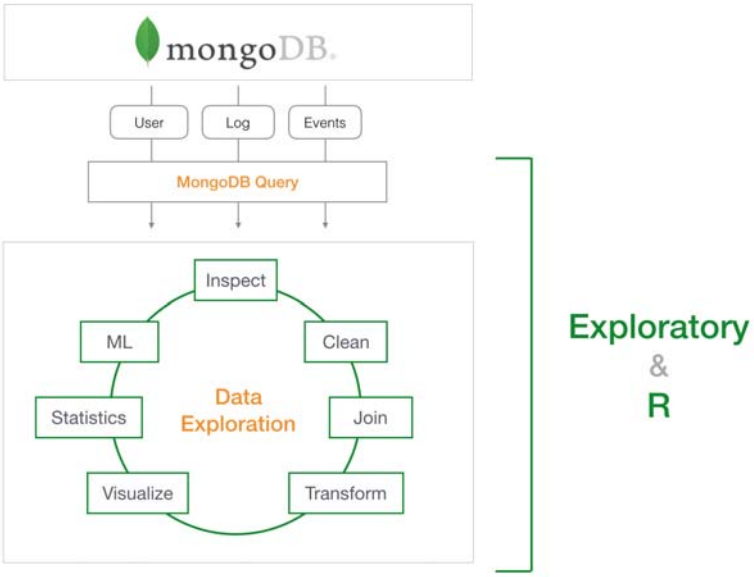
> *Five NoSQL Databases with the Pros and Cons*

1. Apache Cassandra

This distributed database management system is considered extensively scalable. Its ability to scale horizontally makes Cassandra among the few database systems that can process data within the shortest time, give high-performance output, and maintain high reliability. The advantage of using Cassandra is the ability to manage huge amounts of structured, semi-structured, and unstructured data. However, the high costs that are incurred in using the database make it disadvantageous.

2. MongoDB

It is considered an open-source NoSQL database system, making it be the freest of all. The pros of applying MongoDB are that it is highly available, easily scalable, and provides high performance. Its disadvantage is that it is the opposite of the famous MySQL which is read in rows and columns.



An Introduction to MongoDB Query for Beginners. (2020). Retrieved 8 March 2020, from https://blog.exploratory.io/an-introduction-to-mongodb-query-for-beginners-bd463319aa4c?gi=39cb81bdd28a

3. Neo4j

This is an open-source NoSQL of the graph-based database and the most preferred database of the graph-based model. The advantages of using Neo4j are its ability to manage highly connected data proficiently (Hussain, 2014). In addition, it allows the storage of data more naturally from sources such as recommendation engines and social networks.

4. Hadoop

Hadoop is a NoSQL database that provides platforms for distributed data storage and processing that designed to process large amounts of data while minimizing the financial and time taken. Hadoop can store and handle huge data efficiently, only that it is expensive and time-consuming.

5. OrientDB

Orient is designed to combine the flexibility of a document database with the difficulty of a graph database (Nayak, Poriya, & Poojary, 2013). The advantage of using Orient is the fact that it develops various solutions simultaneously, while its disadvantage is that it can be so complex and challenging for some developers.

Two NoSQL Weather App Examples

a. Weather for Windsor

This is a weather app that provides data concerning the most current as well as reliable seven days weather forecasts, reports, and information to the individual's phone or personal computer. Weather for Windsor is an app that is designed to obtain data using Cassandra databases which is a relational database.

b. AccuWeather

The AccuWeather app submits the current date and the N-day weather forecast, meaning that the app can provide its users with any data range concerning the weather forecast (Matteson, 2016). The app applies MongoDB, which is a non-relational database, in collecting and providing data; hence, making it to be reliable and perform efficiently.

References

Facebook. (2009). Hive – A petabyte scale data warehouse using hadoop. Facebook. Retrieved fromhttps://www.facebook.com/notes/facebook-engineering/hive-a-petabyte-scale-data-warehouse-using-hadoop/89508453919/

Finley, K. (2011, January 2). How Twitter uses nosql. Readwrite. Retrieved fromhttps://readwrite.com/2011/01/02/how-twitter-uses-nosql/

Hussain, A. (2014, October 31). Top 5 nosql databases. Packtpub. Retrieved fromhttps://hub.packtpub.com/top-5-nosql-databases/

Issac, L. P. (2014, January 14). Sql vs nosql database differences explained with few example DB. The Geek Stuff. Retrieved fromhttps://www.thegeekstuff.com/2014/01/sql-vs-nosql-db/

KnowledgeHut. (2016). Top pros and cons of hadoop. Knowledgehut. Retrieved fromhttps://www.knowledgehut.com/blog/big-data/top-pros-and-cons-of-hadoop

Matteson, S. (2016, September 8). Nosql is ideal database for mobile apps, says Couchbase's chief mobile architect. TechRepublic. Retrieved fromhttps://www.techrepublic.com/article/nosql-is-ideal-database-for-mobile-apps-says-couchbases-chief-mobile-architect/

Meysman, A. (2016, March 5). Nosql database types. DZone. Retrieved fromhttps://dzone.com/articles/nosql-database-types-1

Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL databases and its comparison with relational databases. International Journal of Applied Information Systems, 5(4), 16-19. https://www.researchgate.net/profile/Dikshay_Poojary/publication/3025[illegible]Type-of-nosql-databases-and-its-comparison-with-relational-databases.pdf

Sharma, V., & Dave, M. (2012). Sql and nosql databases. International Journal of Advanced Research in Computer Science and Software Engineering, 2(8).

CREATED BY
Blain Bowman

Credits:

Created with an image by NASA - "untitled image"