Swinburne University Of Technology

HD Research Report – Comparing Performance differences in Cloud Databases: NoSQL and SQL

David Long - 102797812

Shaugato Paroi - 103523487

I declare that in submitting all work for this assessment I have read, understood and agree to the content and expectations of the assessment declaration

Project Submitted as COS20015 Fundamentals of Data Management

Due date: 7th April at 23:59

Word Count:

Table of Contents

I. Abstract

Database sizes have recently gotten bigger, and future growth is anticipated. The cost of storage has slowed down over time, and the amount of storage available has increased significantly. In recent years, cloud encroachment has changed the comparisons. The market heavily depends on database performance. New, conventional databases, specialised applications for scalability, dynamic devices, and user-friendliness are the main focuses of cloud databases. Cloud databases are largely utilised by technologies like business intelligence for data storage, retrieval, update, and analysis. These solutions enable the development of new business strategies, show scalability and elasticity, and manage massive amounts of data with dependable, tailored, and affordable services in a variety of applications. An overview of cloud computing, cloud database architecture and types, and database as a service are given in this paper. The performance and functionality of the various SQL and NoSQL cloud database applications and services necessary for evaluating them are also highlighted, as well as the characteristics, deployment, and service model of cloud computing. It concentrates on the various metrics to evaluate their performance, including the ease of software porting, transactional capabilities, and maximum data storage. This paper's main goal is to help companies and consumers better understand how cloud computing could offer them trustworthy, individualised, and affordable services in a variety of applications.

II. Introduction

Regarding business applications, the IT industry has seen a stunning revolution recently. In the organization's IT infrastructure, programs that were formerly hosted on a single server have been migrated or replaced with e-apps. System storage has also assumed the role of specialised storage. The pay-per-view business model, flexibility, and decreased cost are the main factors that have made distributed computing a reality. Programmers, designers, and architects who need to store data for their applications in a flexible and highly accessible manner from the backend when needed are currently seen as having a solution in cloud databases[1]. The Cloud is being used by businesses to migrate their applications. Even national security agencies committed to providing hosting for cloud operations show that prices and flexibility will eventually triumph despite challenges such an economic change (data location limitations and security)[2]. By moving DBMS features like the log applicator to the storage stage, a cloud provider that offers DBMS as-a-service can reduce network traffic. Other advantages include streamlined software updates, node failure tolerance, hot disc management, and an architecture that "decouples" the storage tier from the compute tier. The characteristics, deployment, and service model of cloud computing were covered in this study, along with the performance and utility of the many SQL and NoSQL cloud database applications and services that need to be assessed. To evaluate their performance, it focuses on a number of variables, including the greatest quantity of data that can be stored, transactional capabilities, and software portability. The main objective is to assist businesses and consumers in comprehending how cloud computing may offer reliable, individualised, and affordable services in a variety of applications. A few instances of cloud database applications are provided in

Despite the fact that Section IV offers a simple method to contrast and assess cloud databases, Section III. Section V concludes the essay by offering some conclusions and suggestions based on the evaluation section.

III. Categories of Cloud Databases

Contrary to relational databases, which use mathematical relationships between tables, a non-relational technique was also used to swiftly store and retrieve unstructured data. Carlo Strozzi coined the term NoSQL (Not Only SQL) in 1998. Accessible schema architecture refers to the lack of a clear and well-structured system for joining data from various table structures when it comes to non-relational data storage. Numerous non-relational database management systems, including MongoDB, Cassandra, OrientDB, and Aerospace, are distributed, horizontally scalable, and open-source. Both relational and non-relational databases depend heavily on performance to store massive amounts of data fast and provide quick access to that data[12]. The structure of both types is shown in figure 1.

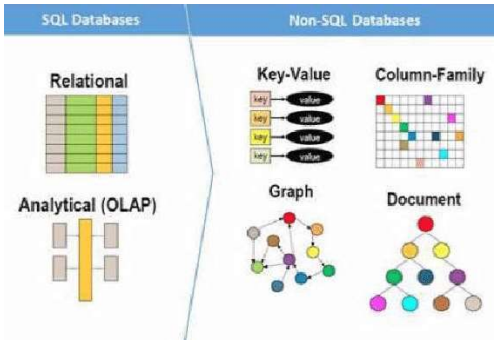

Figure 1 – [2]

IV. Examples of Cloud Databases

There are numerous cloud database companies who offer three different categories of services. An efficient database, a non-rational database, and a virtual computer all employ a local database programme like SQL. DBaaS is a service that is provided by a number of businesses, including Amazon RDS, Google AppEngine Datastore, Microsoft Square Azure, and Amazon SimpleDB[21]. Each service provider differs from the next in terms of the quality and type of services given. The finest service that a business

may offer can be chosen using any criteria, making it non-exclusive to one particular business. However, it does help in deciding which service provider is ideal based on the needs of each firm. The most popular databases in cloud computing are the following:

• PostgreSQL: PostgreSQL is a cloud database that enables management providers and organisations to have highly flexible (DBaaS) while relieving DBAs and application designers from developing and strictly managing current and resilient database conditions. When utilising distributed machines, Postgres Plus Cloud Database reorganises energy efficiency strategies. Additionally, Cloud Database comes with an Oracleperfect DBaaS, which, when combined with Postgres Plus Advanced Server, provides fantastic reserve spending funds and adjustments[26].

• MongoLab: MongoDB is a well-structured, open-source JSON database. It was created by Geir Magnusson and Dwight Merriman. It is not meant to be a high-quality business, but rather a good article database. The data are kept as JSON records of component development. The leading shop and room efficiency are flexible. The lodging is abundant and includes records and a portion of the need for social databases. Additionally offered is the degree of adaptability[9].

• Firebase Real-time Database: This cloud-based NoSQL database enables for offline access and synchronises data in real-time across all users. All related clients share a single instance of the real-time database where data is kept as JSON and automated updates are made to the most recent data[27].

• Google Datastore is a fully managed, highly scalable NoSQL database service provided by Google. Since cloud storage archives can replace physical files, cloud storage "allows users to store data and information in an off-site registry that can be accessed either through the internet or a private network connection," which is highly fiscally practical for ventures. The Cloud Datastore serves as the foundation for Google's Bigtable and Megastore technologies. Users can create databases using Google Cloud Datastore in Native or Datastore Mode. Datastore Mode is designed with new server projects, online applications, and mobile devices in mind[28].

Amazon RDS: Amazon Relational Database Service (or Amazon RDS) is a distributed relational database service by Amazon Web Services (AWS). It is a web service "in the cloud" that aims at simplifying a relationship database installation, operation, and progression in apps. Servicing processes are managed automatically, such as database patching, backup databases, and time-based restoration. A single API allows scaling storage and measuring resources that call on request for the AWS control aircraft. AWS provides no SSH connection as part of a managed service to the underlying virtual machine[1] .

• Amazon SimpleDB: Created by Amazon.com, Amazon SimpleDB is an Erlang-based distributed database. Along with Amazon Elastic Compute Cloud (EC2) and Amazon S3, it is a component of Amazon Web Services and is utilised as a web service. It was declared on December 13th, 2007[29].

• Microsoft SQL Azure: Microsoft Azure SQL Database is a managed cloud database (PaaS) service provided by Microsoft Azure (formerly known as SQL Azure, SQL Server Data Services, SQL Services, and Windows Azure SQL Database).

V. Comparison

In this section, the analysis of the selected publications is discussed. First, pertinent publications addressing problems influencing the performance and capabilities of Cloud Databases are located. The initial assessment and comparison, which are shown in table 1, are then based on some criteria chosen to assess the two most popular types of cloud databases: SQL and NoSQL databases. The requirements specified to evaluate the features and performance of the Cloud Database are further described in depth below.

1. Model: The data model depicts the logical structure of the database. It arranges relevant data and specifies their relationships. Data models are of two sorts Relational and Nonrelational[31] .

2. Data: Each record may have various attributes or identical attributes.

3. Scale: The scalability of the database is the ability to retain increased data quantities without losing performance. Vertical and horizontal scalability are two forms[31]. Only by incorporating additional hardware resources will you compete, including Memory and Processor, to the current computer in vertical scaling via a single node. On the other hand, each node in horizontal scaling includes only a portion of the data, enabling the existing distributed system community to add more machines[32] .

4. Schema: A database schema is a structure that describes how a database is built. It describes the organization of the data and their connection with the relationship between data. Two predefined and dynamic schemes exist—unstructured data in the predefined system. The data structure should be predefined in table form until you start using it. An active schema requires data criteria selected to evaluate the most common cloud databases. Detailed descriptions of the requirements that identified to evaluate functions a performance of Cloud Database are given in **table 2:**

	SQL	NoSQL
Database category	Table base database	Graph, Column, Document and Key Value stored
Scale	Vertical Scalability	Horizontal Scability
Data	Same attribute for every record	Flexible so that each record may have a different attribute
Transactions	ACID transactions	Dependent on different solutions
Performance	Depending on the disk's speed, insert and update performance	Max performance if consistency is reduced
Model	Relational	Non-relational

VI. Discussion

The purpose of this study is to identify previous investigations on Cloud database issues. In light of table 1's comparison, SQL databases are vertically scalable. As a result, vertical scaling costs go up. Because vertical scaling depends on a single computer, another disadvantage is that the programme frequently fails as the server fails. Additionally, SQL databases can be utilised by distributed systems. NoSQL databases, on the other hand, allow for horizontal scaling, which lowers the cost and expedites the process. SQL needs a specified schema in order to process unstructured data. To manipulate data with SQL, you must first build a table-based data structure. For a NoSQL database, however, there is no requirement for a preset structure. NoSQL uses a dynamic method for files that are not structured. Therefore, NoSQL requires no strategies, making the data process faster than SQL if you want real-time data. SQL databases include ACID, which ensures data integrity and consistency. Working with complex queries and reports also makes SQL a better fit for complex query contexts than NoSQL. The majority of contemporary NoSQL databases scale horizontally, allowing you to grow your database by adding additional servers. You can handle an increase in load by sharding your NoSQL database or adding more servers. The BASE consistency paradigm offers "Eventual Consistency" in the majority of NoSQL databases to maintain high scalability and availability. But in SQL databases, the ACID consistency model emphasises "Strong consistency or write consistency." NoSQL is significantly faster at reading and writing data than relational databases, particularly in key-value storage systems. This results in shorter wait times in situations like online transactions. However, a smaller company or business would use a less expensive option, like Amazon SimpleDB. There is much more to go through in terms of challenges before a user if we look at the comparison of table 2, where the DBaaS is the best and has more options as well as better configuration, performance, and features. This essay solely compares the various DBaaS offerings to

determine which is best for a given company or organisation and which offers advantages over the competition. Amazon RDS is our recommendation because it is the superior choice to the others.

VII. Conclusion

This study presented the cloud database structure and looked at some of its important components. For a number of reasons, businesses have come to rely on cloud computing. Data exchange may now be accelerated and improved thanks to cloud computing services, which eliminate the need for each company to build its own data centre. The cloud database has recently created a new (DBaaS), allowing businesses and government organisations to use the resources of DBaaS suppliers and invest in and maintain the hardware and software of data centres containing all of the database's data without any problems. NoSQL databases have shown to be superior for the majority of tasks, according to a thorough study that was presented and comparisons. They are dependable and work faster and more effectively. We advise almost all enterprises to use Amazon RDS since it is the best, has the most options, and has better configuration, performance, and functionality. Additionally, Firebase Realtime Database is perfect for individual or group use, especially if you desire a specific storage location but do not have constant access to the Internet. We can make the following recommendations based on the reviewed articles and our preceding sections:

- NoSQL: NoSQL databases are quick, dependable, and perform better and faster, making them better for the majority of occupations, large corporate facilities, and companies.
- Amazon RDS: We advise almost all enterprises to use Amazon RDS because it is the best and provides more setup, performance, and feature options.
- SimpleDB: Since Amazon SimpleDB is less expensive and better suited for handling lesser amounts of data, small enterprises should adopt it.
- Firebase: For individual or group use, Firebase Real-time Database is best, especially if you don't have constant access to the Internet and need a specific storage place.

VIII. References

https://www.researchgate.net/figure/SQL-and-NoSQL-databases_fig3_299535734