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:2181

Table of Contents

I. Abstract 2
II. Introduction 2
III. Examples of Cloud Databases 3
IV. Comparison 4
V. Discussion 6
VI. Conclusion 7

I. Abstract

Database sizes have recently gotten bigger, and future growth is quite expected. 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. There is now a new type of database called Cloud Database. Cloud databases main focuses are conventional databases, specialized applications for use of dynamic devices, scalability and user-friendliness. Cloud databases are largely utilized 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. In this paper, we'll provide an overview of databases, cloud database architecture, and cloud database types. Additionally, performance and the functionality of the various SQL and NoSQL cloud database applications and services are required for evaluating them, as well as the deployment, characteristics, 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 aim is to help users to have better understand how cloud computing could offer them trustworthy, individualized, 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. There is also system storage which also has assumed the role of specialised storage. The pay-per-view business model, versatility, and decreased price are the main factors that have made distributed computing a reality. Designers, architects, and programmers that need to retain data for their applications in a flexible and easily available from the backend as desired are presently seen as having a solution in cloud databases. The Cloud is being used by businesses to migrate their applications. Even the 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). By moving we can also see that Database Management Services features like the log applicator to the storage stage, a cloud provider that offers DBMS as-a-service can reduce network traffic In addition, hot disc management, quicker software upgrades, and an architecture that "decouples" the storage tier from the computation tier are benefits. The functionality and performance of the many SQL and NoSQL cloud database applications and services that need to be evaluated were examined in this study, along with the features, implementation, and service model of cloud computing. It focuses on a number of factors to assess their performance, such as the maximum amount of data that can be stored, transactional capabilities, and software portability. The main objective is to help businesses and consumers comprehend how cloud comp can offer reliable, individualized, 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 to store and retrieve unstructured data, a non-relational technique was also used to swiftly save and retrieve unstructured data. Carlo Strozier coined the term Not Only SQL, now NoSQL in 1998. Accessible schema architecture refers to the lack of a lucid and properly organized system for joining data from various table structures when it comes to non-relational data storage. Numerous non-relational database management systems, including Cassandra, OrientDB are horizontally scalable, distributed, 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. The structure depicted below shows both types in Figure 1.

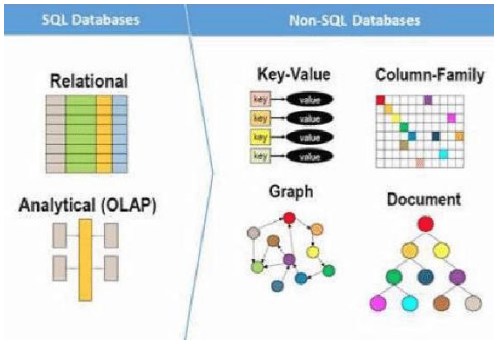

Figure 1

IV. Examples of Cloud Databases

There are numerous cloud database companies who offer three different categories of services. An efficient database, a non-relational database, and a VM computer which employ a local database program like SQL. DBaaS is a service that is provided by a number of businesses, including Microsoft Square Azure, Amazon RDS, Microsoft Azure, Google AppEngine , and Amazon SimpleDB. Each service

provider differs from the 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 organizations to have highly flexible (DBaaS) while relieving DBAs . Besides,when utilizing distributed machines, Postgres Plus Cloud Database reorganizes 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.

• MongoDB: 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. Additionally, it offered very degree of adaptability.

• Google Datastore is a fully managed, highly scalable NoSQL database service provided by Google. Since cloud storage archives can replace physical files, this 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 practical for ventures. This Cloud Datastore serves as the foundation for Google's Bigtable and Megastore technologies. Datastore Mode is designed with new server projects, online applications, and mobile devices in mind.[We use google cloud database as a foundation of creating our database in our custom database]

- Amazon RDS: A distributed relational database service provided by Amazon Web Services is known as Amazon Relational Database Service (or Amazon RDS) (AWS). It is a web service "in the cloud" that seeks to make it easier to instal, run, and advance relationship databases in apps. Automatic management is used for maintenance procedures such database patching, database backups, and time-based restoration. Scaling storage and monitoring resources that the AWS control plane calls for are made possible by a single API. As a managed service, AWS does not offer an SSH connection to the underlying virtual machine.

• 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). [We can also migrate data from google cloud Microsoft cloud database very easily, we also demonstrate in in our custom project]

V. Comparison

In this section, we firstly address problems influencing the performance and capabilities of Cloud Databases . 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.

2. Data: Each record could have unique or repeated properties.

1. Model: The logical layout of the database is represented by the data model. It lists the relationships between pertinent facts and arranges them. Two different types of database models exist. Both of these have relationships and do not.

3. Schema: A database schema is a structure that outlines the construction of a database. It specifies how the data are organised and how those connections relate to the relationships between the data. Before you begin utilising it, the data structure should be predefined in table form. Data criteria must be chosen in order to compare the most popular cloud databases, which requires an active schema.

4. Scale: The database's scalability refers to its capacity to maintain larger amounts of data without suffering performance degradation. It comes in two forms: vertical and horizontal scalability. You won't be able to compete with the present computer's memory and processing power while scaling vertically via a single node without adding more hardware resources. However, with horizontal scaling, each node only contains a fraction of the data, allowing the distributed system community to expand with new hardware.

Table 2 provides a detailed description of the requirements that were identified to evaluate the Cloud Database's performance and functionality.

	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

Table – 1

Cloud Database Comparison

	Microsoft SQL Azure	Mango Lab	Amazon RDS	Google Datastore
Type	SQL	NoSQL	SQL	NoSQL
Maximum data stored	50GB	3TB	1TB	1MB per subset 9f data.
Configurability and tunability of database	5/10	4/10	8/10	3/10
Cost Per Month	Price based on single database	monthly pay with three possible plans with dedicated and shared instances.	Free trial, pay as you go	No initial cost, subsequent cost based on store, read and write.
Accessibility	AD (Active Directory) authentication	AC (Access control) is enabled	Amazon Web Services identity and Access Management	Using access & identity management.
Transaction capabilities	Yes	Yes	Yes	Yes

Table 2

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 utilized 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 sharing 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 emphasizes "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. If we look at the comparison of table 2, the DBaaS is the best and has more alternatives as well as superior configuration, performance, and functionality. On the other hand, a smaller corporation or business would utilise a less expensive option, like Amazon SimpleDB. This essay solely compares the various DBaaS offerings to determine which is best for a given company or organization 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 center. 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

https://passer.garmian.edu.krd/article_144858_955875dc5e9526f02691c53e3aaa95d9.pdf

https://www.mongodb.com/

https://www.couchbase.com/products/capella/start-today?utm_source=bing&utm_medium=search&utm_campaign=BNG+-+APAC+-+Australia+-+Cloud+-+NonBrand&utm_keyword=nosql%20vs%20sql&kpid=bi_cmp-415783426_adg-1212761812231509_ad-75797742565641_kwd-75797970112831:loc-9_dev-c_ext-_prd-&msclkid=ad980d3e6358183527b06e569c418ff1