Performance Evaluation of NewSQL Databases

Karambir Kaur
Department of Computer Science & Engineering
Punjab Institute of Technology (PTU Main Campus)
Kapurthala-144601(Punjab), INDIA
Karamkaur64@yahoo.com

Dr. Monika Sachdeva
Associate professor
Department of Computer Science & Engineering
Punjab Institute of Technology (PTU Main Campus)
Kapurthala-144601(Punjab), INDIA
Monasach1975@gmail.com

***Abstract*-For over forty years, relational databases have been the leading model for data storage, retrieval and management. However, due to increasing needs for scalability and performance, alternative systems have emerged, namely NewSQL technology. NewSQL is a class of modern relational database management systems (RDBMS) that provide the same scalable performance of NoSQL systems for online transaction processing (OLTP) read-write workloads while still maintaining the ACID guarantees of a traditional database system. The rising interest in NewSQL technology, over the last few years resulted in an increasing number of evaluations and comparisons among competing NewSQL technologies. Some of the NewSQL databases currently used in various popular web applications are VoltDB, Google Spanner, MemSQL, SAP HANA, NuoDB, and TokuDB.**
This work is trying to comment on the various NewSQL database systems which describes benefits, characteristics, and classification of NewSQL databases for online transaction processing (OLTP) for Big data management. It also provides the list of popular NewSQL databases in categorized tables. This dissertation work mainly covers evaluation comparison between four NewSQL databases: NuoDB, VoltDB, MemSQL, and Cockroach DB on the basis of various parameters like read latency, write latency, update latency, and execution time. The experiments do not cover only performance (latency and execution time) but also focus on ease of use and flexibility of the used NewSQL databases.

Keywords: Big Data, Distributed Databases, NewSQL databases, NuoDB, VoltDB, MemSQL, Cockroach DB.

I. INTRODUCTION TO NEWSQL DATABASES

The first use of the term was in 2011 in report of 451 group senior analyst, Matthew Aslett. Advances in Web technology and the proliferation of mobile devices and sensors connected to the Internet have resulted in immense processing and storage requirements. Traditional relational databases were designed in a different hardware and software era and are facing challenges in meeting the performance and scale requirements of Big Data. For Big Data environment, supporting more clients or higher throughput required an upgrade to a larger server. Until recently this meant that implementing a scale-out architecture either required a non-SQL programming model or relying on sharding and explicit replication. There were no solutions that provided complete ACID semantics. This tension is what inspired the NewSQL movement.

NewSQL is a class of modern relational database management systems (RDBMS) that provide the same scalable performance of NoSQL systems for online transaction processing (OLTP) read-write workloads while still maintaining the ACID guarantees of a traditional database system [1]. NewSQL databases address scalability, performance and availability issues of relational DBMS and provide full or partial SQL query capability. Most of them follow relational data model. New SQL should be considered as an alternative to NoSQL or Old SQL for New OLTP applications.

A. Technical characteristics of NewSQL

- SQL as the primary mechanism for application interaction.
- ACID support for transactions.
- A non-locking concurrency control mechanism so real-time reads will not conflict with writes, and thus cause them to stall.
- An architecture providing much higher per-node performance than available from traditional RDBMS solutions.
- A scale-out, shared-nothing architecture, capable of running on a large number of nodes without suffering bottlenecks [2].

II. TESTED DATABASES

This section gives a description about the particular databases that are to be tested. Their performance will be recorded and compared under different parameters.

A. NuoDB

Launched in 2010 by industry-renowned database architect Jim Starkey and accomplished software CEO Barry Morris, the company is based in Cambridge, MA [3]. NuoDB is a distributed, peer to peer system that provides an in-memory database service with **ACID** transactions. A NuoDB database appears to the developer and operator as a single, logical system. In practice, a NuoDB database can be running in multiple locations with hosts added and removed according to demand [4].

978-1-5090-4715-4/17/$31.00 ©2017 IEEE

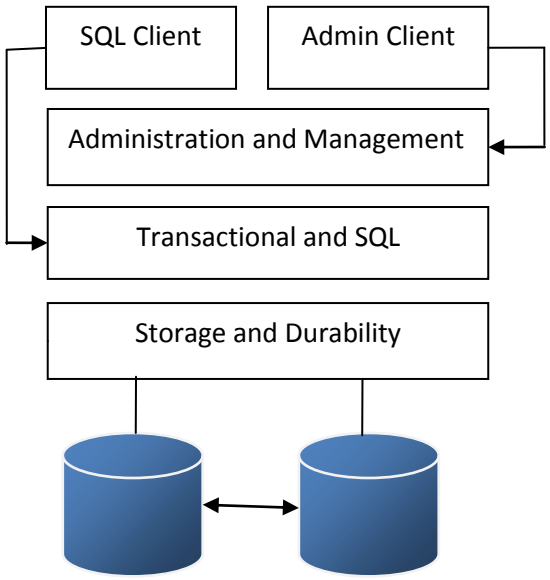

Fig I: Architecture of NewSQL Database Nuodb [5]

B. VoltDB

VoltDB is an in-memory, scale-out SQL database purpose- built to power a new generation of applications that thrive on fast, smart data. Tapping the lightning speed and real time analytics of VoltDB, organizations are able to add context and intelligence to data-the instant it arrives-to make real-time transactional decisions that maximize business value. First, it's the latest database designed by Michael Stonebraker, the database pioneer best known for Ingres, PostgreSQL, Illustra, Streambase, and more recently, Vertica. But interestingly, Stonebraker declared that he has thrown "all previous databases Architecture out the window" and "started over with a complete rewrite." What's resulted is something totally different from every other database—including all the column- and table-oriented NoSQL systems. Moreover, VoltDB claims a 50 to 100x speed improvement over other relational database management systems (RDBMSs) and NoSQL systems.

C. Cockroach DB

CockroachDB is a distributed relational database that is scalable, survivable, and strongly consistent - hence the name. It combines the rich functionality of SQL with the horizontal scalability common to NoSQL offerings, and provides enterprise-grade disaster recovery. The company was founded in 2015 with a mission to Make Data Easy. CockroachDB is a distributed SQL database built on a transactional and strongly-consistent key-value store. It scales horizontally; survives disk, machine, rack, and even datacenter failures with minimal latency disruption and no manual intervention; supports strongly-consistent ACID transactions; and provides a familiar SQL API for structuring, manipulating, and querying data.

Cockroach Labs is the company behind CockroachDB, an open source, scalable SQL database headquartered in New York City. CockroachDB was launched as an open source project on Github in 2014. When Cockroach Labs was founded in 2015 to develop the project full-time, the founders made open source a core principle of the company.

D. MemSQL

MemSQL is a distributed, relational database that handles mixed transactions and real-time analytics at scale. It is accessible through standard SQL drivers and syntax and supports a broad ecosystem of drivers and applications.

MemSQL has a two-tiered architecture that provides high throughput. It is a distributed system that can scale horizontally on commodity hardware, and is very compatible with other technologies in the modern data processing ecosystem (e.g. orchestration platforms, developer IDEs, and BI tools). It features an in-memory rowstore, and an on-disk columnstore. It also features Streamliner, a tool that can efficiently stream data into the MemSQL rowstore and columnstore. MemSQL is the database platform for real-time analytics. Querying is done through standard SQL drivers and syntax, leveraging a broad ecosystem of drivers and applications [6].

III. QUALITATIVE COMPARISON

TABLE I: QUALITATIVE COMPARISON TABLE

Databases / Features	VoltDB	NuoDB	CockroachDB	MemSQL
Category	New Architecture Databases	New Architecture Databases	New MySQL Storage Engine	New Architecture Databases
Availability	Open source. Commercial enterprise edition	Closed source. Developer's edition. Available as a service in AWS marketplace	Open source. Community edition	Open source. Community edition
Developed by	VoltDB Inc. Michael Stonebraker, 2010	Jim starkey in 2012	Spencer Kimball at cockroach labs in 2014	MemSQL Inc. in 2013

Who is using?	Nokia, Ericsson, Flipkart, hp	UAExchange, Forbes, Kodiak	No one yet	Intel, Microsoft Azure
Consistency	Strong	Eventual	Strong	Strong
Storage type	In-memory	In-memory	In-memory	In-memory rowstore and on-disk column store
Concurrency control	No, single threaded model	Yes, MVCC	Yes, optimistic concurrency control	Yes, MVCC
Scalability	Bi-directional scale-up, scale-out	Scale-out	Horizontal, automated scaling	Highly scalable horizontally and durable

IV. PROPOSED WORK

A. Problem Statement

There are many NewSQL databases being used in big data era, for real time web and big data applications. Every database is having a particular data format for storing its data. Hence a new customer faces the problem of selecting the appropriate NewSQL database that can be used for his business requirements, while moving from relational database and NoSQL, to NewSQL. So the problem is how to compare and find the suitable NewSQL system as solution for better performance, evaluating them on a set of common workloads of different types or applications. To explore the new features of NewSQL, which it adopt from NoSQL and SQL relational system. Interesting databases can be analyzed to find out if they can improve the performance of business analytics.

B. Objectives

The reason behind choosing this topic for thesis research work is that, why NewSQL databases derived when there are NoSQL databases present. NoSQL databases have capability to provide scalable solutions by compromising performance of processing. NewSQL systems provide same solution like NoSQL with additional features which improve the scalability, transaction processing and relational databases. The main objectives are as follows:

- To study NewSQL Databases: Nuo DB, Volt DB, MemSQL and Cockroach DB.
- To analyze the performance of each NewSQL database by applying operations on the data sets: create, insert, update, and select and conclude which database performs the best.
- To compare the NewSQL Databases i.e. NuoDB, VoltDB, MemSQL and Cockroach DB (qualitative and quantitative comparison), on the basis of parameters like Execution Time, Read Latency, Write Latency and Update Latency.

C. Requirements

TABLE II: SOFTWARE AND HARDWARE REQUIREMENTS

Requirements	
Software	• Oracle VirtualBox-5.0.16-105871-Win • Operating System: Ubuntu 14.04 LTS • Java (JDK 1.8) Oracle • cockroach.linux-amd64.tgz • memsql-ops-4.1.11.tar.gz • voltdb-ent-6.1.tar.gz • nuodb-ce-2.4.1.2.x86_64.exe
Hardware	• CPU: 8 physical cores • RAM: 8 GB (4 GB to virtual machine) • Disk: 60 GB • Processor: Intel(R) Core™ i7-3770 CPU @ 3.40 GHz • System Type: 64-bit Operating System, x64-based processor

D. Methodology

NewSQL solutions aim to bring the relational data model into the world of NoSQL. Therefore, a comparison among different NewSQL databases is essential to understand this new class of data stores. Methodology is devised which includes detailed comparison among various NewSQL databases by using different parameters. NewSQL solutions differ greatly in their characteristics; moreover, changes in this area are rapid, with frequent releases of new features and options.

This work was developed to answer the following research question: “Is there presently sufficient knowledge on quality attributes of NewSQL systems to assist a software engineer’s decision process”? Thus, a methodology is devised to develop this work and answer our original research question.

Step 1- Install Ubuntu 14.04 LTS OS on virtual Box (single node cluster).

Step 2- First check all Prerequisites like Java, SSH (secure shell), if Prerequisites are not install then install it.

Step 3- Install the NewSQL databases (VoltDB, CockroachDB, MemSQL) on machine and after configuring some configuration files, start the databases.
Step 4- Install NuoDB and java (JDK 1.8) on windows.
Steps 5- Before running the test, tables are created in all databases and load files into them.
Step 6- Run the loaded workload and get the results.
Step 7- Performance metrics such as creation, insertion, updation, selection are calculated from output of all database local host's screen.

V. RESULTS

A. Comparison Graph

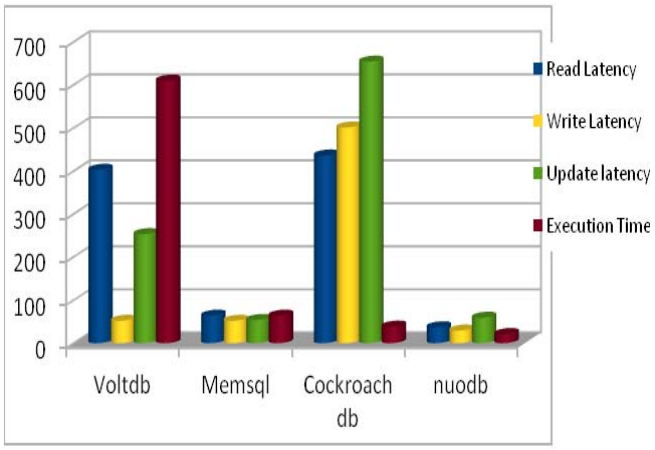

Fig. II: Final Graph comparing all databases

Almost similar results are noticed while plotting the write latency curve for various target NewSQL systems, except in the case of cockroach DB, which takes maximum time for inserting rows in table. Overall performance of Nuo DB is good in all parameters, which takes minimum time to perform all operations. Over the next set of tests, MemSQL again has slight edges over Nuo DB as it demonstrated in write, read and execution time. While Cockroach DB almost takes maximum time for all operations, except in execution time.

TABLE III: COMPARISON TABLE

Parameters / Databases	Read latency	Write latency	Update latency	Execution Time
Volt DB	40.3	52.02	253.36	610.3
MemSQL	64	52	54	64
Cockroach DB	436.2	501.21	654.31	38.53
NuoDB	36.9	29.3	59.1	21.2

VI. CONCLUSION AND FUTURE WORK

A. *Conclusion*

NewSQL solutions aim to bring the relational data model into the world of NoSQL. Since there are a wide variety of NewSQL databases available now days, a new customer wishing to switch from the traditional physically hosted database server design to using NewSQL on the cloud, faces the problem of selecting the appropriate NewSQL database that can be used for his business needs without drastically changing his existing application. Therefore, a comparison among different NewSQL databases is essential to understand this new class of data stores.
In this experiment, NewSQL technology and databases are analysed with the objectives of providing a perspective on the field, while approaching different aspects that highly contribute to the use of those systems. A comparison among the most prominent solutions is performed on a number of dimensions.
Performance comparison of the NewSQL databases (NuoDB, VoltDB, MemSQL, and Cockroach DB) takes place using qualitative and quantitative comparison methods. According to the experiments performed, Results show that NuoDB takes minimum time to perform all the operations except in updation for which memSQl takes minimum time. So, the NuoDB performed better in most test cases.

B. Future Work

As we have seen, many business applications and industries in big data era, are adapting to NewSQL databases. NewSQL is still an under-development field, with many questions and a shortage of definite answers. Its technology is ever-increasing and ever-changing, rendering even recent benchmarks and performance evaluations obsolete. There is a much scope and opportunities that can be explored to compare their aspects and performances against each other. Some of the possible future scope of this work can be:

- Security was not taken as criteria while studying different products. This was because it is a huge field of study in itself. So, security features and vulnerabilities of chosen products can be investigated further.
- Deploying multiple nodes of each NewSQL systems: The experiments were performed in this work were on single node. By increasing memory and other aspects of NewSQL used could vary the results observed. So, increasing the number of client's performance tests and judging them would be a great concern.
- There is also a lack of study material or research papers which focus on use-case oriented scenarios or software engineering quality attributes. There is a current need for a broad study of quality attributes in order to better understand the NewSQL ecosystem, and it would be interesting to conduct research in this domain. When more studies and with more consistent results have been performed, a more thorough survey of the literature can be done, and with clearer, more concise results.

REFERENCES

[1] http://cacm.acm.org/blogs/blog-cacm/109710-new-sql-an-alternative-to-nosql- and-old-sql-for-new-oltp-apps/fulltext

[2] Piekos John,"*Sql Vs. Nosql Vs. Newsql: Finding The Right Solution*", August 2015 http://dataconomy.com/sql-vs-nosql-vs-newsql-finding-the-right-solution/

[3] Jim Starkey, Barry Morris, "The architecture and motivation for NuoDB", January 2013, 215 First Street Cambridge, MA 02142, www.nuodb.com

[4] http://doc.nuodb.com/Latest/Default.htm#Introduction-to-NuoDB.html

[5] Moniruzzaman, A. B. M. *"Newsql: towards next-generation scalable rdbms for online transaction processing (oltp) for big data management." arXiv preprint arXiv:1411.7343* (2014).

[6] http://docs.memsql.com/docs/

[7] Grolinger, Katarina, et al. "Data management in cloud environments: NoSQL and NewSQL data stores." *Journal of Cloud Computing: Advances, Systems and Applications* 2.1 (2013): 1.

[8] Murugan, Perumal, and Ananda Sentraya. *"A Study of NoSQL and NewSQL databases for data aggregation on Big Data."* (2013).

[9] Doshi, Kshitij A., et al. "Blending SQL and NewSQL approaches: reference architectures for enterprise big data challenges." *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2013 International Conference on*. IEEE, 2013.

[10] Stonebraker, Michael, et al. "The end of an architectural era:(it's time for a complete rewrite)." *Proceedings of the 33rd international conference on Very large data bases*. VLDB Endowment, 2007.

[11] Kumar, Rakesh, et al. "Critical Analysis of Database Management Using NewSQL." *International Journal of Computer Science and Mobile Computing* 3 (2014): 434-438.

[12] Shamshi Jawwad,"*NoSQL vs NewSQL*", *https://sites.google.com/a/nu.edu.pk/syslab/projects/nosql-vs-newsql*.

[13] Aslett, Matthew. "How will the database incumbents respond to NoSQL and NewSQL." *San Francisco, The* 451 (2011): 1-5.

[14] Binani, Sneha, Ajinkya Gutti, and Shivam Upadhyay. "SQL vs. NoSQL vs. NewSQL-A Comparative Study." *database* 6.1 (2016).