

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

A SURVEY ON NOSQL AND NEWSQL DATA STORES FOR BIG DATA MANAGEMENT

Esha Awasthi¹, Shikha Agrawal² & Rajeev Pandey³

^{1,2&3}Department of Computer Science and Engineering, UIT- RGPV, Bhopal-402033, India

ABSTRACT

In this era of technologies, where due to the advancement in several web technologies and frequent growth of portable devices, and sensors linked over the web are resulting to the huge amount of data. Due to this rapid increase of well-structured, semi-structured and other types of unstructured data called Big Data, traditional database systems are facing several difficulties. Since forty years, traditional databases are the leading model for several data manipulation tasks such as data storing, data retrieving and managing the data. However, because of growing requirements for better scalability and higher performance, other alternative database technologies, namely NOSQL and NewSQL technology have emerged. In order to overcome the challenges faced by traditional database system, there are numerous NOSQL and NEWSQL databases in the industry. It becomes challenging to select appropriate database solution for Big Data Management. This research work will present the comparative analysis and performance evaluation of most popular NOSQL and NEWSQL databases on the basis of several criteria. In this survey paper, there are mainly five sections: first section comprises of basic introduction and background of research in this field. In further sections, there is a brief introduction about adopting nosql and newsql technologies. Finally, there is a literature survey and survey report of this field.

Keywords: NoSQL, NewSQL, ACID, BASE, CAP, Big Data, Database, RDBMS.

I. INTRODUCTION

In today's world, the data is increasing tremendously with the number of users over the internet. The advancement in most of the Web Technologies and rapid increase of portable devices connected over the network are resulting to the production of large amount of data sets that needs to be organised and processed. For instance, there are numerous of web sites such as Facebook where 2.4 billion of data[1], either structured or unstructured are shared among friends on a daily basis. Presently, around 2 billion populations across the globe are connected to the (Web) internet, and around 5 billion peoples were having mobile phones, also by 2020, around 50 billion devices are expected to be connected over the Internet[2]. Also, today's business has been generating huge volume of data that is too huge to be handled and processed by RDMS tools.

1.1 Big data

It is considered as voluminous amount of data, either it is well-structured or it is not in a structured form. Big Data is defined by mainly three concepts which are: variety, velocity and volume[3]. Further, veracity and volume has been come into consideration[3]. Further, veracity and volume has been come into consideration[3]. It is huge amount of data set which is very complex to handle with relational database system. Even though traditional database system have several advantages over file approach such as simplicity, flexibility, security, robustness and consistency, but this approach is facing several difficulties regarding several issues like scalability issue and performance requisites of Big Data. The requirement for Big Data generated from the industries such as Facebook, Google, yahoo, amazon, YouTube etc. The datasets that are generated differ from structured data. There are several features of Big Data:

1. Data is distributed at different locations
2. It is very much complicated in nature, thus difficult to manage
3. The data is dynamic in nature i.e., it keeps on changing
4. It is very large in volume, variety and velocity

In order to handle this data, cloud computing technology has been come into consideration.

1.2 Cloud computing

Cloud Computing is one of the trending technology that is growing very rapidly day-by-day. It is the most prominent and cost-effective solution in the field of computation and storage of data. It provides various computational services like storing the data, at very attractive cost. It is no longer dependent upon a server or a number of hardware machines, as it is a virtualsystem. This technology includes five characteristics which are: on-demand self-service, broad network access, rapid elasticity, resource pooling and measured services[4]. These make this technology significant. This is the reason of several industries and corporate sectors to move their works and services towards cloud computing. There are various service models of cloud computing: Software-as-a-service (SAAS), Platform-as-a-service (PAAS) and Infrastructure-as-a-service (IAAS)[4] .

Despite of its benefits, few limitations are also there such as: It can be very time consuming if someone wants to transit large amount of data in or out of cloud storage. Therefore, the “Nosql” and “Newsql” technologies come into consideration and become alternatives to RDBMS which are able to provide the required performance and scalable solutions. In the next sections, we are going to describe about basics of Nosql and newsql databases.

II. NOSQL DATABASES

NoSQL has been emerged as one of the most prominent data store to resolve storage problems related to Big Data. It is also referred as “Not Only SQL” to focus that they may also able to support SQL-like query language. NoSQL databases are flexible in nature and considered as an alternative to RDBMS. The main objective of the rapid and future growth of this technology is that it allows better storing and retrieving capabilities of data, without bothering about the structure and content.

2.1 Features of nosql data stores

1. NoSQL data stores are not allowed to use relational data model.
2. NoSQL databases can store huge amount of data.
3. In distributed system, NoSQL can be used easily without any inconsistency.
4. If there is any fault or failure present in any machine, in that case there will be no interruption in any work.
5. The source code of NoSQL is openly and freely available to everyone i.e., it’s an open-source database.
6. NoSQL is a schema less and allows storage of data without having any rigid schema.
7. There is no support for ACID property of RDBMS.
8. NoSQL databases are more flexible.
9. NoSQL is horizontally scalable database.

2.2 Fundamentals of nosql

2.2.1 No acid support

ACID property is widely supported by Relational Database System, but in NoSQL, there is no support for ACID properties which stands for Atomicity, Consistency, Isolation, and Durability. Due to the consistency feature, it is not supported by NoSQL.

2.2.2 Base

NoSQL is a system that supports BASE rather than ACID properties. **BASE** stands for **B**asically **A**vailable, **S**oft **S**tate, and **E**ventually **C**onsistent.

- **Basically Available:** The data stores are having high availability even when its subsets of data are not available for a short period of time. They are available in all time period whenever they are being accessed.
- **Soft State:** It can tolerate and handle inconsistency for a particular time interval, as it does not required to be consistent all the time.
- **Eventually Consistent:** The data stores become consistent after a particular period of time. They come into the consistent state.

2.2.3 The cap theorem

In order to deal with the huge amount of datasets, it is essential to follow some strategy to store and process that dataset. The commonly known technique is to partition the dataset across various distributed servers. In addition, the copies can be made and kept in different machines, so that in case of any server's failure, we can easily use that data as it is still present.

These methods of partitioning the huge dataset and replicate them is having an essential restriction, which was produced by this theorem. This Theorem states that any two CAP properties out of three should be fulfilled by distributed system simultaneously.

NoSQL is capable to give high availability and therefore it supports AP (Available/ Partition Tolerance).

- **Consistency:** The ability of managing the database in a consistent manner is usually referred to as Consistency, in CAP.
- **Availability:** There should be high availability of data in order to serve a request whenever it is required.
- **Partition Tolerance:** It is the ability of distributed network to handle partition and to tolerate it.

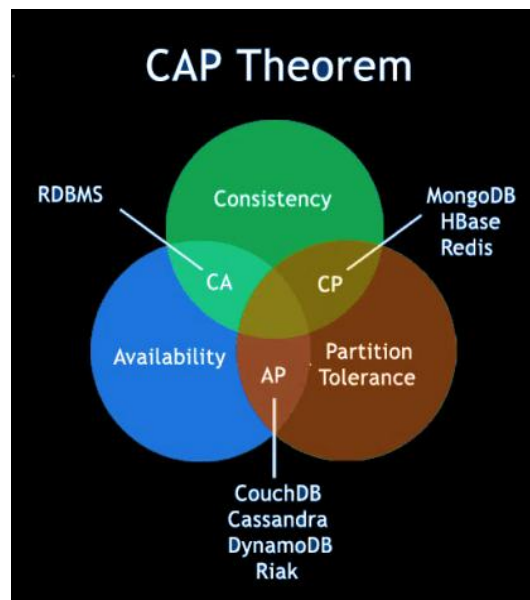


Figure1: cap theorem

2.3 Classification of nosql data stores

2.3.1 Key value store:

In this, the storage of data is done in a key-value pairs such as key is used to identify data uniquely. Each data or value is associated with a unique key. This model is simple and schema-free in nature. An instance is Amazon's SimpleDB.

2.3.2 Column-family data stores:

This data store comprises of several rows (row-key), where each row consist of several column- family (different row can have different no. of column- family). Different column-family may have different number of columns in it. This data store provides highly effective capabilities to index and query the data rather than above mentioned data stores. An instance is Cassandra, Hbase, and Google's Big Table.

2.3.3 Document- based stores:

This data store provides another type of derivation of the key- value model and allowed to use keys to identify several document. This mainly focuses on organising the data and stored them as a collection of documents. Examples are MongoDB, Apache CouchDB.

2.3.4 Graph data stores:

These allow putting several query upon the graph based structure. The implementation of these data stores can hold up such query effectively. An example includes Neo4j.

2.4 Disadvantages of nosql databases

- NoSQL Databases does not provide effective support such applications that are already developed and created by using traditional RDBMS.
- The migration of existing applications in order to adopt new patterns of data; to create emerging applications on OLTP system that is highly scalable.

The next section comprises about NewSQL databases that overcome the above limitations and also provide equal amount of performance.

III. NEWSQL DATABASES

NewSQL is an upcoming generation's modern and scalable relational database for OLTP which is able to give the equal scalable performances of NoSQL Data stores, and also maintaining the ACID properties of RDBMS.

3.1. Features of newsql data stores

1. A primary mechanism of this data stores: SQL.
2. It fully supports ACID properties of traditional RDBMS.
3. It supports non-locking concurrent techniques and mechanism.
4. Architecture that provides very high per-node performance as compared to available from old SQL.
5. NewSQL supports distributed, scale-out, parallel, shared-nothing architecture, that have ability to run on huge number of nodes.
6. NewSQL databases are approximately 50 times faster as compared to traditional RDBMS.

3.2 Classification of newsql databases

3.2.1 New architecture databases:

The very first category of NewSQL Data stores are totally latest platforms of databases. These are developed in order to process in a cluster of distributed environment where individual node is having its own subset of the information. These are written with a intention of distributed environment in mind from scratch. Examples are VoltDB, NuoDB, and MemSQL.

3.2.2 New mysql storing engines:

Another category includes MySQL engines that are having high optimisation capabilities. The interface for programming is similar to that of a SQL. It provides better scalability as compared to built-in engines. Examples are TokudB, MySQL Cluster, and MyRocks.

IV. LITERATURE SURVEY

This section comprises of several number of research papers in which several researchers have shown the analysis they have made on NOSQL and NEWSQL Databases. Researchers have done analytical study on various databases suitable to handle large volume of data which are as follows:

In this paper, four types of NewSQL databases are evaluated and compared on the basis of their performance. **NuoDB, VoltDB, MemSQL and CockroachDB** are analyzed based on several parameters such as read, write, update latencies and execution time. Along with Quantitative analysis in [5], qualitative analysis has been made by the author which covers availability, consistency, storage type, concurrency control and scalability. It not only describes the comparison based on performance but also covers other qualitative parameters. This mainly focuses on various NewSQL data stores in order to describes its advantages, features and classification. Finally, this paper concludes that NuoDB is better in performance than other databases in several test cases.[5].

The author of [6] paper has been noticed that there are huge amount of data generation from different locations in a rapid manner through Internet of Things [6]. Though RDBMS is efficient enough for storage of data, but when it comes to large amount of sensors data, it lacks some of its features. Thus, NoSQL and most trending NewSQL databases performed far better than traditional databases. In this paper, a comparison is made between MySQL, **NoSQL'sMongoDB** and **NewSQL'sVoltDB** databases and these are tested for several parameters: single (write, read and delete) and multi-write operations. It concluded that NewSQL(VoltDB) performs exceptionally well as compared to other databses.[6].

The [7] paper cover the basic feature, architectures and categories of NewSQL Databases for Online Transaction Processing for handling huge amount of Data. It provides the comparative study of NewSQL, NoSQL and SQL (RDBMS) database systems based on certain parameters. Also, it gives the categorised description of several Databases of NewSQL and NoSQL. The paper shows that NoSQL and NewSQL are having more security differences than RDBMS due to the performance criteria as top metrics[7].

In this paper, it has been described that advancement in the technologies related to web and the generation of portable devices linked to the internet results in a huge volume of several types of data. Traditional database systems are having several difficulties in order to meet the performance of Big Data. An alternative to RDBMS have been emerged to solve the problems which are: NoSQL and NewSQL. This paper comprises of the various characteristics of NoSQL and NewSQL databases. NewSQL is considered as a most promising technology to give huge data OLTP concerns, and also secure ACID properties [8].

In this paper, NoSQL and NewSQL databases have been reviewed in order to: 1. provide a view and perspective in these areas, 2. help to select the suitable data store among them, and 3. identify difficulties, challenges and future perspectives in the area. In order to find out the optimum database, the comparison has been made on the basis of data models, querying properties, scalability, and security. Characteristics that drive the capabilities of scaling read and write requests have been investigated, specifically in partitioning, consistency, and concurrency control. Also, the NoSQL and NewSQL use cases and scenarios have been studied [9].

This paper mainly comprises of two NewSQL databases that are very popular. The first database is MemSQL and other is the VoltDB. The advantages of these databases have been shown using one of the benchmark called as TPC-H benchmark. The ability of both the databases has been shown by this experiment of executing the TPC-H benchmark queries. MemSQL performed better than VoltDB for 1GB Data. [10].

NoSQL is one of the very effective database management systems for Big Data services. But, apart from its advantages, there are several drawbacks also. NoSQL have no support to ACID, thus due to this, it is unable to support several existing and earlier applications of SQL databases. This is the reason of motivation for NewSQL databases. It is the latest and most prominent solution that comes into the world. NewSQL provides same performance just like NoSQL. In this paper, several databases have been reviews to find the suitable solution for huge amount of data [11].

In this paper, the four NoSQL databases have been compared to understand the perspective and overview. This paper comprises of comparative study of 4 Data models of NoSQL and also compare with SQL. In this comparison which is based on performance for huge amount of data, NoSQL database performed far better for industry situations which require several features [12].

The comparative study is based on the several criteria which are as follows: schema, data models, scaling model, transactional capabilities. This paper reviews the NoSQL concepts that are required in this era of voluminous data of several types, which is in distributed form. It also compared the NoSQL and traditional databases. It also compared the NoSQL with traditional databases, explains about ACID properties, BASE and CAP theorem [13].

This paper reviews about the complete necessity of NewSQL databases in this era. It discusses about history of NewSQL to know that why NewSQL come into existence. This paper also covers detailed description of NewSQL and its several categories such as main memory storage, partitioning, concurrency control and replication [14].

This paper mainly covers about the concept and background of NewSQL. It also provides the description of LDDB (Liberty Database Connectivity) and New Architecture and describes how it works. Also, it covers about main categories and features of NewSQL databases. The main objective of [15] is to cover significance of NewSQL data stores and provide better solutions for appliances [15].

This paper mainly covers the approaches, data models for voluminous data, introducing to batch, serving, and speed layer. Also, it covers the Lambda and Couch reference architecture and describes how to deal with big data by using NewSQL databases. It also covers future opportunities of NewSQL databases [16].

In this paper, an approach has been described that comprises of two imperatives which are address by blending scale-up and scale-out for processing the data. In [17], it categorizes the growth of data into following classes: Chronological, Horizontal, and Vertical; takes several different techniques for blending traditional database, and NewSQL databases for every class. Further, this paper gives description about various instances of the real-world in which these techniques have been used [17].

In this paper, several databases of SQL (**MySQL**), NoSQL (**MongoDB, Cassandra**) and NewSQL (**VoltDB**) have been discussed on the basis of several parameters. Further, it includes data models, features and comparison has been made between different databases [18]. Comparative analysis has been made between the databases based on several parameters such as Partitioning, OLTP, Map Reduce, Transaction, Replication, Querying. It has been observed that every database is having its own advantages and disadvantages with performance taken as a high priority. The result has been shown that VoltDB has consistent performance with less latency [18].

V. CONCLUSION

In today's scenarios, where due to the advancement in several web technologies, corporate industries and frequent growth of portable devices, and sensors linked over the web are resulting to the huge amount of data. Due to the rapid increase of data and information in several industries, not only in websites such as social networking and e-commerce but also in medical field, agriculture field etc, traditional database systems are facing several difficulties. Therefore, there is a high need of handling and managing the large amount of data with the help of growing technologies such as NoSQL and NewSQL databases to overcome the limitation of relational databases. In order to provide better solutions to manage the information, the requirement and scope for NoSQL and NewSQL has been increased day by day. In this research, it has been clearly observed that the evaluation of NewSQL after NoSQL is worth enough due to its several features like ACID support. Though NewSQL and NoSQL have delivered same scalable performance but still there is a need of NewSQL in today's era of Big Data for most OLTP solutions. It has been observed that it is essential to make a clear comparison between both the data stores on the basis of several factors due to its increasing demand in the industry. It is necessary to provide a comparative study between NoSQL and NewSQL databases for the future scope.

AUTHOR	TITLE	PARAMETERS	RESULT
Karambirkaur, Dr. Monica Sachdeva[5].	Performance Evaluation of NewSQL Databases (2017- IEEE).	Quantitative Analysis: 1. Read, write, update Latency 2. Execution Time Qualitative : 1. Availability 2. Consistency 3. Storage Type 4. Concurrency Control 5. Scalability	NuoDB is better in performance than other databases in several test cases.
Haleemunnisa Fatima, Prof. KumudWasnik[6].	Comparison of SQL, NoSQL and NewSQL Databases for Internet of Things (2016- IEEE).	1. Single Client with Single operations 2. Single client with Multi operations 3. Multi-client - single write and read	NewSQL (VoltDB) performs exceptionally well as compared to other databases.
A. B. M. Moniruzzaman[7].	NewSQL: Towards Next-Generation Scalable RDBMS for Online Transaction Processing (OLTP) for Big Data Management (2014- IJDTA).	Quantitative parameters such as Security, cloud support, OLTP, Big Volume, Scalability, Storage etc.	NoSQL and NewSQL are having more security differences than RDBMS due to the performance criteria as top metrics.
Deepika Aggarwal, Roopam, Sonika[8]	Emerging Technologies For Big Data Processing: NoSQL and New SQL Data Stores (2016- IJECS).	Qualitative Analysis 1. Data Models 2. Interface 3. Storage type 4. Query method 5. Replication 6. Concurrency Control etc.	NewSQL is considered as a most promising technology to give huge data OLTP concerns, and also secure ACID properties.
Katarina Grolinger, Wilson A Higashino, Abhinav Tiwari and Miriam AM Capretz[9].	Data Management in Cloud Environments: NoSQL and NewSQL Data Stores (2013 – Springer).	1. Data Models 2. Querying 3. Scaling 4. Partitioning & replication 6. Consistency 7. Concurrency Control 8. Security	Comparative Analysis between different NoSQL and NewSQL Databases.
João Oliveira ¹ and Jorge Bernardino[10].	NewSQL Database: MemSQL and VoltDB Experimental Evaluation (IC3K 2017).	The performance analysis has been performed with TPC-H Scale factor (1 GB Data).	MemSQL performed better than VoltDB for 1GB Data.
Sneha Binani, Ajinkya Gutti, Shivam Upadhyay[11].	SQL vs. NoSQL vs. NewSQL – A Comparative Study (CAE-2016).	1. Scaling 2. ACID 3. SQL 3. OLTP 4. Querying Complexity 5. Distributed	It has been found that NewSQL is the best choice for OLTP.

SitalakshmiVenkatraman, Kiran Fahd, Samuel Kaspi, RamanathanVenkatraman[12].	SQL Versus NoSQL Movement with Big Data Analytics (2016-MECS).	1. Schema 2. Data Models 3. Scaling Model 4. Transactional Capabilities	NoSQL shows better results in the context of Big Data as compared to SQL.
OmpalSingh[13].	A Comparative Study Of Nosql Data Storage Models For Big Data (2015-IJARSE).	1. CAP 2. BASE 3.ACID 4. Data Models	This paper reviews the NoSQL concepts that are required in this era of voluminous data of several types.
Andrew Pavlo, Matthew Aslett[14].	What's Really New with NewSQL?	1. Main Memory Storage 2. Partitioning 3. Concurrency Control 4. Replication	This paper reviews about the complete necessity of NewSQL databases in this era
Rakesh Kumar, Neha Gupta, Hitesh Maharwal, ShilpiCharu, KusumYadav[16].	Critical Analysis of Database Management Using NewSQL (IJCSMC- 2014).	Liberty Database Connectivity (LDBC) and New Architecture	NewSQL databases are better options for use.
Rakesh Kumar, Neha Gupta, ShilpiCharu, Sunil Kumar Jangir[17].	Manage Big Data Through NewSQL.	Lambda Architecture	NewSQL databases should be chosen as an alternative solution to NoSQL for OLTP.
Haleemunnisa Fatima, KumudWasnik[18].	Comparison Of SQL, NOSQL And New SQL Databases In Light Of Internet Of Things – A Survey.	Partitioning, OLTP, Map Reduce, Transaction , Replication, Querying	The consistent performance has been shown by VoltDB with less latency.

REFERENCES

1. "A New Data Center for Iowa | Facebook Newsroom", *Newsroom.fb.com*, 2018. [Online]. Available: <http://newsroom.fb.com/News/606/A-News-Data-Center-for-Iowa>. [Accessed: 22- Jun- 2018].
2. Nawsher Khan, IbrarYaqoob, Ibrahim AbakerTargioHashem, ZakiraInayat, WaleedKamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani "Big Data: Survey, Technologies, Opportunities, and Challenges"*The Scientific World Journal*, vol. 2014, Article ID712826, 17 July 2014.
3. "Big data", *En.wikipedia.org*, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Big_data. [Accessed: 22- Jun- 2018].
4. "Cloud computing", *En.wikipedia.org*, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Cloud_computing#Characteristics. [Accessed: 22- Jun- 2018].
5. K. Kaur and D. Sachdeva, "Performance Evaluation of NewSQL Databases", *International Conference on Inventive Systems and Control (ICISC-2017)*, 2017.
6. H. Fatima and P. Wasnik, "Comparison of SQL, NoSQL and NewSQL Databases for Internet of Things", *IEEE Bombay Section Symposium (IBSS)*, 2016.
7. Moniruzzaman, "NewSQL: Towards Next-Generation Scalable RDBMS for Online Transaction Processing (OLTP) for Big Data Management.", *International Journal of Database Theory and Application*, vol. 7, no. 6, pp. 121-130, 2014.
8. D. Aggarwal and R. Sonika, "Emerging Technologies For Big Data Processing:
9. NOSQL And NEWSQL Data Stores", *International Journal Of Engineering And Computer Science*, vol. 5, no. 12016, pp. 15598-15604, 2016.

10. K. Grolinger, W. Higashino, A. Tiwari and M. Capretz, "Data management in cloud environments: NoSQL and NewSQL data stores", *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 2, no. 1, p. 22, 2013.
11. J. Oliveira and J. Bernardino, "NewSQL Databases MemSQL and VoltDB Experimental Evaluation", In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2017)*, vol. 2, pp. 276-281, 2017.
12. S. Binani, A. Gutti and S. Upadhyay, "SQL vs. NoSQL vs. NewSQL- A Comparative Study", *Communications on Applied Electronics*, vol. 6, no. 1, pp. 43-46, 2016.
13. S. Venkatraman, K. Kaspi and R. Venkatraman, "SQL VersusNoSQL Movement with Big Data Analytics", *International Journal of Information Technology and Computer Science*, vol. 8, no. 12, pp. 59-66, 2016.
14. O. Singh, "A COMPARATIVE STUDY OF NOSQL DATA STORAGE MODELS FOR BIG DATA", *International Journal of Advance Research in Science and Engineering*, vol. 4, pp. 31-39, 2015.
15. A. Pavlo and M. Aslett, "What's Really New with NewSQL?", *ACM SIGMOD Record*, vol. 45, no. 2, pp. 45-55, 2016.
16. R. Kumar, N. Gupta, H. Maharwal, S. Charu and K. Yadav, "Critical Analysis of Database Management Using NewSQL", *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 5, pp. 434-438, 2014.
17. R. Kumar, N. Gupta, S. Charu and S. Jangir, "Manage Big Data ThroughNewSQL", *International Conference on Innovation in Wireless Communication and Networking Technology*, 2014.
18. Kshitij A Doshi, Tao Zhong, Zhongyan Lu, Xi Tang, Ting Lou, Gang Deng "Blending SQL and NewSQL Approaches Reference Architectures for Enterprise Big Data Challenges." *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2013.
19. H. Fatima and K. Wasnik, "Comparison Of SQL, NOSQL And New SQL Databases In Light Of Internet Of Things – A Survey", *International Journal of Advances in Electronics and Computer Science*, vol. 3, no. 3, pp.31-34,2016