

Big Data with Column Oriented NOSQL Database to Overcome the Drawbacks of Relational Databases

NaglaaSaeedShehata

Faculty of Computers & artificial intelligence, Helwan University, Cairo, Egypt
Email: nagla_sd@yahoo.com

Amira Hassan Abed

Egyptian Organization for Standardisation & Quality, Cairo, Egypt
Email: Mirahassan61286@gmail.com

ABSTRACT

Due to the Era of Big Data with the large amount of distributed databases in the web and the rapid growth in the smart systems a rapid growth happening in database models and the relational database fails to dealing with such a big amount of data and have many limitations the need to new technologies comes up, which makes DBMS developers move towards column oriented NOSQL database. The main goal of this paper is to provide a survey on NOSQL Model especially a column oriented NOSQL database, providing the user with the benefit of using NOSQL database, Instead of using the (row database) relational to overcome the drawbacks of the relational database Model.

Keywords - Relational Databases, NoSQL, Columnar Database, BASE properties.

Date of Submission: Feb 16, 2020

Date of Acceptance: May 08, 2020

1. INTRODUCTION

The rapid growth of the web technologies and cloud applications that changes the nature of stored data which included social media information's, transactions, online purchases and because of the relational database model scalability issues, the need of new and easier approach arises to overcome those problems researchers provides NOSQL Model [1], that provides a new data stores techniques rather than the relational database tabular data store.

1.1 Relational Database

The relational Database model presented by coded in the 1960s and "it's the model that deals with data and organizes it into one or more tables, or we could call it a relations, which consists of columns and rows, defining that relation with a unique key"[2][3] identifying each row "Primary Key " we also called the Rows in this kind of databases as records or tuples and Columns called "attributes", each relation represents one entity type and rows represent instances of that entity and the columns representing values attributed to the instance that connect between two or more tables is called a relation and it has some characteristic as follows[4] :

- Optional attributes (NULLs),
- Depends on defined schema.
- Use joins to aggregate related data
- Dealing with large data VOLUME and high rate of READ (scalability)

And it has a number of advantages [5]:

- Based on ACID
- Depends on Strong consistency, concurrency, and recovery.
- Mathematical background
- Using The Standard Query language (SQL)
- Vertical scaling (up scaling)

But in the other hand it has some drawbacks as it can't deal easily with a huge amount of data and the distributed databases that contains a variety of data type (semi structured – unstructured).

1.2 Big Data

The Big data is a new technology that deal with a huge number of data (terabytes – zettabytes) and it can be defined in different ways, but the 3 V's namely volume, variety, and Velocity are sufficient to represent the most general characteristics of big data [6]: fig.1 showing that the three characteristics of big data.

- Volume: which is refers to the magnitude of data.
- Variety: that refers to data come from a number of variety sources.
- Velocity: and it's refers to the streamed data collected on real time [7].

There is great interest in deploying big data technology in the healthcare industry to manage massive sets of diverse health datasets such as electronic health records and sensor data, which are increasing in magnitude and variety due to the commoditization of electronic devices

such as mobilephones and wireless sensors. The newfangled medical andhealthcare systems have to be augmented with new “bigdata” computing and analysis capabilities.[8]

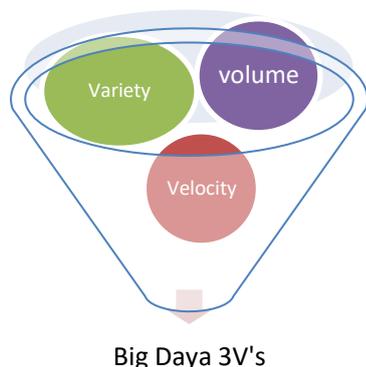


Fig .1 Big Data 3v's

1.3 NOSQL Database Model

NoSQL is a non-relational databasemanagement system that neither uses SQL query languagefor operation data nor is based on tabular relations that areextremely good in dealing with the large amount of datainvolved in big data. The main concept upon whichNoSQL is based on is the notion of distributed storage ofdata alongside to the handling of parallel processing [27].

NoSQL is based mainly on horizontal scalability and there are a lot of different implementations, different systems and techniques in building a NoSQL database system. NoSQL databases mainly differ in the way data is stored and accessed they can be classified into many different types for example, wide-column store, document store, and value store each of which has its own characters and these three categories cover most of the techniques involved.

NoSQLis “Class of database management systems (DBMS)” [1] which Stands for "Not Only SQL“

Which Characteristics according to[4] [20] as follows:

- No fixed schema (schema less)
- No joins (typical in databases operated with SQL)
- Does not use SQL as querying language
- Distributed, Partition-tolerant architecture

Characteristics Of NOSQL Databases [4][22]:

NOSQL have a number of Characteristics like it's an open free source,didn't depended on a schema which make it easy to use and it could be considered as The most cluster friendly[2]

Also the researchers say that it provides the users [9] with the ability to add frequent changes to DB, some good Solution to Impedance mismatch, also NOSQL depends on BASE usability (Basically available, soft state, Eventual consistency) and CAP (consistency, Availability, Partition tolerance) .

NoSQL databases types are classified into four major data models are [25]: which are showed in Fig.2

- Key-value model
 - Document model
 - Column family model (our point of scope)
 - Graph model
- (Each DB Model in NOSQL has its own query language)

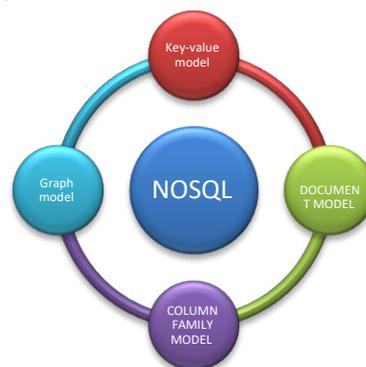


Fig.2 NOSQL Models

1.4 Column-Oriented Database Systems

It's called (Column - stores) [21] that referring to it stores data in column rather than rows as in the relational database because the column is the smallest and lowest instance of data.It contains the data name – value, also in NOSQL, database stores each table with different columns, with many values that are belongs to the same column stored contiguously, compressed, and densely packed, unlike the traditional database systems that store main records (rows) after each other in a continually manner.

In this category of NoSQL database, the columns are realized and determined in relevant to each row in state of predefined by the table organization owned uniform sized columns for each tuple. Such these stores introduce a two dimension gross/aggregate organization, a key and a row gross that is defined as a set of columns. This allows any column to be added to any particular row, and in this case the rows can own a lot of various columns. In other words, each row possesses a number of different columns that were maintained and stored. It also is able to maintain datain tables like segments of data columns.

2. State of The Art

In (Row Database) relational the data stored as tables containing entities which relate to each other relationally depends on a structured data type to fit them In the relational tables but due to the huge evolving in the web based application [27] nowadays which contains a different types of data "semi structured and un-structured data" (social media – on line purchasing process and any kind of online activity and according to that there are significant researches to solve the problem of scale that

occurred in the relational databases and connect with the big data using NOSQL databases to overcome the drawbacks in the (Row Database) relational which focused especially on who use the "Column-oriented database systems" and the result of the survey in this filed listed in the next few lines [11]:

2.1 NOSQL Vs. Relational Database

Researchers in [22] discuss the NOSQL database concepts and explained how it differ from relational database, also they mention why NOSQL database is needed through the era of the big data also provides the features of NOSQL data base and the no SQL model types then focus on the consistency methods for NOSQL database, but the paper did not mention the advantages and disadvantage of NOSQL database or the relational database throw their work.

[19] Presented a comparative study which compares between relational and NOSQL databases they focus in the presented work on the processes and (features and characteristics) constraints for each two types of databases. This paper could be considered as a qualitative research paper that based on deeply analysis and detailed, they produced their work on the latest researches that published during the last few years, but the researcher didn't answer the important question through their work, which is the best database solution?, also the researcher didn't take in consideration a number of key points such as "flexibility, scalability, performance, query language, security, and availability which considered as strong and important issues".

Researchers provide in [16] strength comparison between NOSQL and SQL databases. Focuses on the overhead, the conclusion of their work is that the "overhead is not related to SQL, but to other components defined main four components such as buffer management, locking, logging, and latching". The researcher examined their work through the previous four components. Getting their comparative with that avoidance of the overhead, specialized with one or more of these components, can provide a speedup of two of them.

[23] Presents a Survey with more focuses on the difference between Relational and Non-Relational Database, shown the main differences between each databases, take into consideration the advantages and disadvantages of each database and the tools used through each types of them but the work didn't give a clear results that could help the user which database type to choose to apply their needs.

2.2 NOSQL Models

In [29] provides comparative study for NOSQL and They explained the main concepts and the analysis of the NOSQL database's architecture like "Mongo DB, Cassandra, and HBase", the researcher focuses on Cassandra as a case study and explain briefly the performance evaluation of the other databases in the aspects of read and write performance. But the researcher

ignores or they didn't mention during his work to the result of their analytical study, which should be mentioned together as results of the study they made.

Researcher in [14] compared between relational and NOSQL data base considering well the performance of each databases and some technologies that is used during his research, the researcher found that NOSQL databases perform better than relational databases due to the good facility in non-relational databases, the researcher also compared the different types of NOSQL databases, testing or checking some operations which are read, write, delete and instantiate that are considered the main operations.

2.3 Column-Oriented Database Systems

The study in [28] provides extensive solutions to the problem of relational DB migration to HBase. The researcher uses MySQL which is considered relational DB as input to the model output with column oriented database HBase. They extract the features of objects using semantic enrich, encompassing inheritance, aggregation and composition which are represented in a New Optimized Data Model (NODM). The goal of the proposed model to store data in a column oriented database through novelty method focusing on Map structure. The model ignores the details of the relational database limitations; they focus only on the goals only and how to achieve it.

[15] Presents a Tutorial for Column-oriented where some open research problems were discussed as by authors column-store systems includes physical database design, indexing techniques, parallel query execution, replication, and load balancing, as a conclusion of this work the authors compared between the column-stored systems as a commercial products which exists in the real market.

Authors in [17] provide fine comparison between Column-Oriented Database Systems and XML their work explained the relationships between XML compressors and column-stores. They illustrated that a permuting XML compressor, called XSAQCT with the DBMS back-end has essentially the same functionality as a column-store (while ignoring things such as SQL Joins), including in their work a specific kind of compression, Also they test the compression ratio achieved with the compressor they made, experiments were performed on an XML corpus and the test showed a very good results that make their work strong and applicable to use instead of the XML they also describes the existing XML compressor showing the similar inherent between its compression technique and column-stores.

3. Column-Oriented Database Systems

The main purpose of using the columnar database is to reach to high performance in the operation of read and write data from and to stored data and get speed up in the

process returns of a query results, the column – oriented has so many uses as in customer relationship management (CRM) systems, data warehouses, and in any other inquiry systems.

3.1 Advantages Of Column-Oriented DBMS[12][21]

- Uses less disk space through using the self-indexing.
- Highly Data compressed.
- Use lower disk space due to the compressed schemes
- Don't read unnecessary columns.
- Faster to perform operations avoiding decompression costs.

3.2 NOSQL database (column oriented) vs. relational database (row oriented)[13][24]

Table (1) (Column Oriented) Vs. (Row Oriented)

Category	Column oriented	Row oriented
Description	Is a direction system (DBMS) that stores information tables by column instead of by row	Relational DBMS that Stores the data in two-dimensional table, of columns and rows
Stored systems	A column-oriented info serializes all of the values of a column along, then the values of the following column, and so on	A common technique of storing a table is to arrange every row of information

Benefits	suited for OLAP	suited for OLTP
	like workloads (e.g., Data warehouses) (INSERTs) should be separated into columns and compressed as they're keep, creating it less suited to OLTP workloads	like workloads that are a lot of heavily loaded with interactive transactions
Compression	Allow compression through using uniform data type	Due to the lake of its uniform data type not available in row-oriented data.

Table 2 differentiate between relational database and column data base [19].

Table.2 Difference between Column Data Store and Row Store [10]

Table (2): shows how the data stored in relational (row) differ from to be stored in column data base. The example assumes that we have three variables sales, product and country; we need to store values and store them in the database. In the row store each row contains values about product and sales and country but in column store seriously all country then all products then sales.

4. Survey Results

After presenting and surveying a number of related papers and researches that related to our research scop we could say that now:

- Column-arranged associations are progressively productive when a total should be registered over numerous lines yet just for a remarkably littler subset of all segments of information, since perusing that littler subset of information can be quicker than perusing all information.
- Column-oriented organizations are a lot of economical once new values of a column are equipped for all rows without delay, as a result of that column information will be written expeditiously and replace previous column information while not touching the other columns for the rows.
- Row-oriented organizations are additional economical once several columns of one row are needed at the identical time, and once row-size is comparatively little, because the entire row will be retrieved with one disk request.[18]
- Row-oriented organizations are additional economical once writing a new row if all of the column information is provided at the identical time, because the entire row will be written with one disk.

Table

	Country	Product	Sales
Row 1	India	Chocolate	1000
Row 2	India	Ice-cream	2000
Row 3	Germany	Chocolate	4000
Row 4	US	Noodle	500

Row Store

Row 1	India Chocolate 1000
Row 2	India Ice-cream 2000
Row 3	Germany Chocolate 4000
Row 4	US Noodle 500

Column Store

Country	India India Germany US
Product	Chocolate Ice-cream Chocolate Noodle
Sales	1000 2000 4000 500

And as result of the survey done in the filed we will compare between NOSQL database (column oriented) and the relational database (row oriented) some other perspective as follows in table (3):

Table 3The Difference Between Column Oriented and Row Oriented

Category	Column oriented	Row oriented
1-Transaction reliability	Transaction can occurred when Column range from BASE to ACID.	high transaction reliability
2- Data Model	Many models	Depend on the mathematics
3- Scalability	Horizontal scalability.	Vertical scalability
4- Cloud	suitable with cloud	Not suitable
5- Big data handling	Used mainly for big data	Face difficult to dealing with big data
6-Data warehouse	Can serve datawarehouse	Difficult to manage increased data
7-Complexity	Support differentData types	High complexity only table/row formula
8- Crash Recovery	Recovery achieved through replication	achieved through log files and ARIS algorithm
9- Security	Need many solution to be secure	Very secure also can use different tools to be more secured

Security in NOSQL database (column oriented) and the relational database (row oriented): in table 4

Table 4Column Oriented and Row Oriented

Category	column oriented	row oriented
1.Authentication	Could be achieved using some methods and techniques	Authentication mechanisms applied to All relational databases
2- Data Integrity	data integrity is not always achieved	achieved using ACID properties
3.Confidentiality	Difficult to be achieved	using encryption techniques
4. Auditing	don't provide auditing	mechanisms to audit using advanced language are Provided
5. Client communication	Client communication Security is missed.	provide secure client communication mechanism via using SSL protocols and encryption

Table (4) illustrated a brief comparison in the security issues between the column and row databases as a result of the survey taking in consideration the main 5 categories Authentication, data integrity, confidentiality, auditing, and the client communication.

5. Conclusion

Finally concluding of the work done here that column oriented systems are used when a new data are bringing in the data set while those data are un or semi –structured data, and the consistency could be relaxed for a while in the situation that the performance will come first, in the case we choose column a huge number of user requests can be answered with eventually consistent unlike the row DB which is focuses on having a strong consistency but at the cost of scale and performance speed which makes the column oriented systems a good choice in so many fields but we still have a great problem, that we still need distributed DBMS that having the four main properties:

- High availability
- Consistency
- Scalability
- Fault tolerance

Which founds here is no way to achieve them together according to the CAP theorem which makes the column also not the best solution but the future will be using a combination of (SQL and NOSQL)the researchers named it a NEWSQL and it could considered as future point of research .

REFERENCES

[1] Raj, P., &Deka, G. “A Deep Dive into NoSQL Databases”. San Diego: Elsevier Science &Technology., 2018.

[2] M. State, “Relational Database Management Systems Semester - III,” *Management*, no. 9038, pp. 1–8.

[3] M. State, “Relational Database Management Systems Semester - III,” *Management*, no. 9038, pp. 1–8.

[4] Radoev, M. (2017). “A Comparison between Characteristics of NoSQL Databases and Traditional Databases”. *Computer Science And Information Technology*, 5(5), 149-153. doi: 10.13189/csit.2017.050501

[5] M. A. Mohamed and O. G. Altrafi, “Relational vs .NoSQL Databases : A Survey,” vol. 03, no. 03, pp. 598–601, 2014.

[6] A. Corbellini, C. Mateos, A. Zunino, D. Godoy and S. Schiaffino. “Persisting Big Data: The NoSQL landscape”. *Information Systems*. Vol. 63, pp. 1-23. Elsevier Science, 2017.

[7] A. E. YOUNESS KHOURDIFI, MOHAMED BAHAJ, “A new approach for migration of a relational Database into column-oriented A NEW APPROACH FOR

MIGRATION OF A RELATIONAL DATABASE INTO COLUMN-ORIENTED NOSQL DATABASE ON HADOOP,” no. November, 2018.

[8] Amira Hassan. Mona Nasr, Walaa Saber “The Future of Internet of Things for Anomalies Detection using Thermography”, International Journal of Advanced Networking and Applications (IJANA), Volume 11 Issue 1, pp. Pages: 4142-4149 (2019).

[9] S. Gajendran, “A Survey on NoSQL Databases,” *IBM J. Res. Dev.*, vol. 5, no. 8, 2014.

[10] HANA .s “row store vs column store Internet”: <http://www.saphanacentral.com/p/rowstore-vs-column-store.html>, 2016 accessed on [Jan. 19, 2019].

[11] H.R.V.yawahere, Drp.pkard&DrV.M.Thakare, "Nosql database". International journal of evoloutionary scientific research and technology ,Conference Paper · October 2017

[12] N. D. Karande, “A Survey Paper on NoSQL Databases : Key-Value Data Stores and Document Stores,” vol. 6, no. 2, pp. 153–157, 2018.

[13] K.Dwivedi, A., S. Lamba, C., &Shukla, S. “Performance Analysis of Column Oriented Database Vs Row Oriented Database”. International Journal Of Computer Applications, 50(14), 31-34, 2012.

[14] Y. Li and S. Manoharan, “A performance comparison of SQL and NoSQL databases” in IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Canada, Aug. 2013, pp. 15- 19.

[15]D. J. Abadi and S. Harizopoulos, “Column-Oriented Database Systems (Tutorial) Column-oriented Database Systems,” no. August 2009.

[16] M. Stonebraker, “SQL databases v. NoSQL databases,” *Commun.ACM* Journal, vol. 53, no. 10–12, 2010.

[17]T. Corbin, T. Müldner, and J. K. Miziołek, “Column-Oriented Database Systems and XML Compression Column-Oriented Database Systems and XML Compression,” no. November, 2014.

[18] M. Abdellatif, M. Salah, and N. Saeed, “ScienceDirect Overcoming business process reengineering obstacles using ontology-based knowledge map methodology,” *Futur.Comput.Informatics J.*, vol. 3, no. 1, pp. 7–28, 2018.

[19]K. Sahatqija, J. Ajdari, X. Zenuni, B. Raufi, and F. Ismaili, “Comparison between relational and NOSQL databases,” no. May, 2018.

[20]M. Radoev, “A Comparison between Characteristics of NoSQL Databases and Traditional Databases,” vol. 5, no. 5, pp. 149–153, 2017.

[21]N. C. Ug and J. Steemann, “Column-oriented databases,” pp. 1–34, 2012.

[22]H. Vyawahare, “NoSql Database,” no. June, 2017.

[23]C. Györödi, R. Györödi, and R. Sotoc, “A Comparative Study of Relational and Non- Relational Database Models in a Web- Based Application,” vol. 6, no. 11, pp. 78–83, 2015.

[24]N. Jatana, “A Survey and Comparison of Relational and Non-Relational Database,” vol. 1, no. 6, pp. 1–6, 2012.

[25]A. Nayak, “Type of NOSQL Databases and its Comparison with Relational Databases,” vol. 5, no. 4, pp. 16–19, 2013.

[26] Amira Hassan. Mona Nasr, “Diabetes Disease Detection through Data Mining Techniques”, International Journal of Advanced Networking and Applications (IJANA), Volume 11 Issue 1, pp. Pages: 4142-4149 (2019).

[27]Amira H. A., “Recovery and Concurrency Challenging in Big Data and NoSQL Database Systems”, International Journal of Advanced Networking and Applications (IJANA), Volume 11 Issue 04, pp. Pages: 4321-4329 (2020).

[28] Oliveira, F., Oliveira, A., & Alturas, B..Migration of Relational Databases to NoSQL - Methods of Analysis. Mediterranean Journal Of Social Sciences, 9(2), 227-235, 2018.

[29]T. Of and V. Of, “COMPARATIVE STUDY OF NOSQL DOCUMENT , COLUMN S TORE D ATABASES A ND EVALUATION OF CASSANDRA,” vol. 6, no. 4, pp. 11–26, 2014.