

# NOSQL – New Wave for Big Data Storage and Retrieval Solution and Its Comparison: A Survey

**Neha D. Surati**

*Department of Computer Science IT Engineering  
Shri S'ad Vidya Mandal Institute of Technology, India*

**Bhavin I. Shah**

*Department of Computer Science IT Engineering  
Shri S'ad Vidya Mandal Institute of Technology, India*

## Abstract

The overcome the problem of storing and managing big data in today's web 2.0, new kind of technology is developed which is known as NoSQL. Now a days NoSQL is a very popular as alternative of RDBMS and dealing with large scale data, they also provide high availability and scalability to the distributed systems. In this paper, we will discuss on overview of big data challenges, NoSQL database , characteristics and classification of NoSQL database and survey on comparison on popular NoSQL database.

**Keywords: Big Data Challenges, Nosql Database, Classification of Nosql Database, Comparison**

## I. INTRODUCTION

In recent era, big data is a flood of data which have data as structured, semi- structured and unstructured data. Now Big Data analysis drives nearly every aspect of our modern society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences and social network etc.[1] Today big data are coming through internet and It is estimated that data volume is increasing 40% per year, and will grow 44 times between 2009 and 2020.[2]

Today big data is becoming one of fashionable topic which are given the important of big data in of data in different verity in massive data structure. Today the data has no limit to generate every second through different platform like social networking site, online transaction and so on. So the big data now a days become the part of daily life environment. The characteristics of the big data are Data Velocity, Data Volume and Data Variety. Big Data is not the only about the size of data but also use data variety and data velocity. Hashem introduced the 5's characteristics of big data. The characteristic of big data can be understood better by dividing it into classes. These classes are Data Sources, Content Format, Data Stores, Data Staging and Data Processing (Hashem et.al, 2015).[3]

- 1) Data Sources: social media , ecommerce websites, education world.
- 2) Content Format: Structured (Simple Text), Semi-Structured and Unstructured(Audio, video,call).
- 3) Data Stores: Document-oriented, Column-oriented, Graph based and Key-value.
- 4) Data Staging: Cleaning, Normalization and Transform.
- 5) Data Processing: Batch and Real time.

The Paper is discussed as following scenario. Section I introduce some challenges of big data, Section II discuss the overview on NoSQL database Section III literature survey based on comparison of NoSQL database and last Section IV give the conclusion and future work on it.

## II. CHALLENGES OF BIG DATA

Many challenges occur in big data due to the rapid exponential growth in the volume of data, so many application and tools are developed as the reduction of problem. In big data analysis process every stage contains some problem as data management, store and analyzing data, data retrieval etc. here we focus on one of them that how to storage and analyzing big data. The increasingly massive growth of data causes a problem of how to store and manage such huge heterogeneous datasets with moderate requirements on hardware and software infrastructure.

Nowadays, Internet companies rapidly activate for the acceptance with millions of people to access them as online marketing as well as social media. From the output of these become generation of data are in huge amount. For example, Google processes data of hundreds of Petabyte (PB), Facebook generates log data of over 10 PB per month.[1] Big data are mostly generates from many appliance like industry, education, social network, organization etc.Now those data are collected and managed through data management tool RDBMS. But RDBMS are only work on structured data other than semi and unstructured data and it doesn't handle large amount of data. So the researcher has planned to solution of that and create new tool database as NoSQL. So the daily updating social networking sites like Facebook and twitter, they can handle huge amounts of data with scalability. It is quite helpful in data warehousing.

### III. NOSQL DATABASE

This NoSQL name was first used by Carlo Strozzi in 1998 as the name of the file he was developing for his database. Now a days, NoSQL is becoming so popular due to its high storage and also because its properties avoid the basic features of SQL. NoSQL (non-relational) and also (not only SQL) is comparatively faster and reliable than relational databases (RDBMS). In SQL, we were using Query language to fetch as well as to store data and for NoSQL we store large data entities using documents in XML formats. XML language is basically used to store structured data in a human readable form.

NoSQL databases are based on BASE (Basically Available, Soft State, and Eventually Consistent) principle that is characterized by high availability of data, while sacrificing its consistency.[4] On the other hand, relational databases are represented by ACID (Atomic, Consistent, Isolated, and Durable) principle where all the transactions committed are correct and do not corrupt database, and data is consistent. Both principles come from the CAP theorem– (Consistency, Availability, and Partition Tolerance ). They are described below.

- 1) Consistency: The data present on all machines must be same in all updates to be made on all machines frequently i.e. consistent data.
- 2) Availability: Data should be available permanently and not temporarily i.e. it should be accessible all the time i.e. availability.
- 3) Partition-tolerance: In case of machine failure or any faults occurred in the machines database should work properly without taking any halt i.e. partition tolerance.

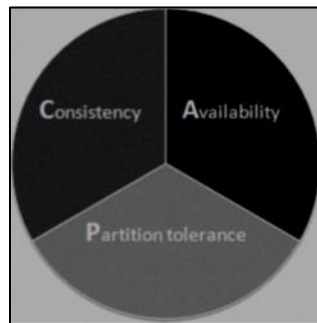


Fig. 1: CAP Theorem

Stefan Edlich provide a comparative study on many NoSQL database through some feature of that database. There are two types of database are categorized and they class have also subclass which are shown in figure.[5]

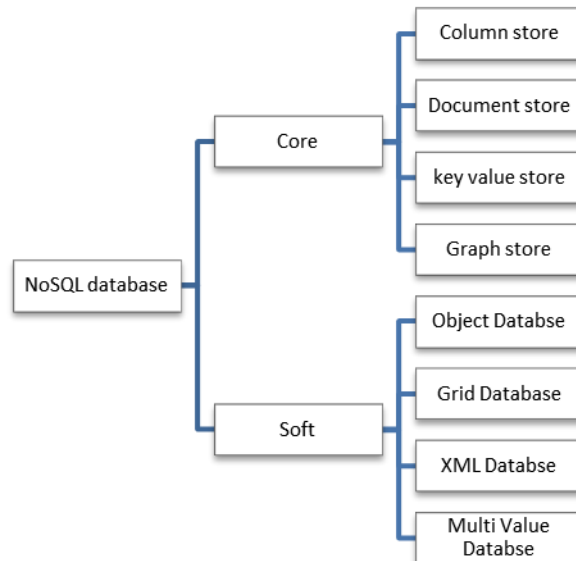


Fig. 2: Types of NoSQL Database

Core database are created as for the web 2.0 technology services when soft database does not related and work with any web 2.0 technology services. Many company use their own storage system to managed millions of big data as Amazon built Dynamo for their internal use, Google developed BigTable – a distributed storage system for managing structured data. Apache Cassandra is a column-family database and was originally developed by Facebook for their own use. Google’s BigTable was the first of the NoSQL databases.[6]

Some of popular NoSQL database are MongoDB, Cassandra, Bigtable, Daynamodb, Hbase as give the availability and scalability on large volume of big data storage and managed.

MongoDB has maintained its top position from last quarter of being responsible for more than 50% of all mentions of NoSQL database in LinkedIn member profiles. As with Q1 there were three changes of position in the rankings in Q2. DynamoDB overtook Riak to claim eighth place, having also gained a place (on MarkLogic) in the previous quarter. Inside the top ten database, DynamoDB had the fastest growth (17.0%).(source: <http://blogs.the451group.com>).

#### IV. RELATED WORK

After the survey on big data issue the big data storage issue is more challenges in the web 2.0 technology so we survey on comparison of NoSQL database through different point of view and review on which database is best from all.

- In [7] they Give the overview and limitation of RDBMS and Comparison made between RDBMS and NoSQL Database.
- In [2] Introduce the concepts of the different NoSQL database types, and provide arguments for and against of the adopting NoSQL database for the new business area.
- In[4] Compute all NoSQL database Cassandra, HBase, MongoDB, OrientDB and Redis based on query performance, read update output for three different workload and conclude that key value redis as a fast response time compare to other.
- In[8] they give overview on bigtable, Cassandra, couchdb, dynamo and mongodb and compare all of them through nonfunctional requirement.
- In[9] they compare NoSQL database based on design, integrity, indexing, distribution and system and provide independent strength and weakness of NoSQL database.
- In[10] compare different NoSQL database against Data model, query model Replication Model and consistency Model.
- In[11] compare NoSQL database in terms of data model, query possibility, concurrency control, partitioning and replication.
- In[12] compare Cassandra, mongodb and Hbase using some parameter as architecture, CAP theorem, performance.
- In[13] Compare SQL and MongoDB based on CRUD (create, read, update, delete) operation for small and large dataset and make out result as for small dataset the retrieval time is less for SQL and for other complex data mongodb is give better result.

#### V. CONCLUSION AND FUTURE WORK

In this paper new NoSQL database are discuss as a big data storage and retrieval solution and These databases have different classes and database store as key value store, document store, column store, graph store. Now a days NoSQL database are widely use in many application as healthcare, education, social network, which contain flood of data in their field so they all are used this database for store and manage big data. Now here we survey on different NoSQL database comparison using different parameter which give the independent strength and weakness of them. From the all NoSQL database the popular once are MongoDB, which are maintain their first leading position on the linkedIn profile. As the limitation of RDBMS the NoSQL are used to build different infrastructure to overcome and useful database. As the familiar with SQL, develop the application which use frontend as SQL and backend as NoSQL. It is mostly use on social media for handle big data storage on cloud. NoSQL database will become more popular in future to develop new application with flexibility and scalability on different platform.

#### REFERENCES

- [1] M.H.Padgavankar, Dr.S.R.Gupta "Big Data Storage and Challenges", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2218-2223
- [2] Jagdev Bhogal, Imran Choksi "Handling Big Data using NoSQL", in 29th International Conference on Advanced Information Networking and Applications Workshops, (2015) IEEE.
- [3] Hakan Özköse, Cevriye Gencer "Yesterday, Today and Tomorrow of Big Data", in World Conference on Technology, Innovation and Entrepreneurship, (2015) 1042 – 1050
- [4] Abramova "Which NoSQL Database? A Performance Overview", in Open Journal of Databases (OJDB), Volume 1, Issue 2, 2014
- [5] Dharmasiri, "A Federated Approach on Heterogeneous NoSQL Data Stores" International Conference on Advances in ICT for Emerging Regions (ICTer), 234 – 239 (2013) IEEE
- [6] Li, Manoharan "A performance comparison of SQL and NoSQL databases", (2013) IEEE.
- [7] Nishtha, Sahil, Mehak, Ishita and Dishant "A Survey and Comparison of Relational and Non-Relational Database" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 6, August – 2012
- [8] Chitra "Study on Basically Available, Scalable and Eventually Consistent NOSQL Databases" in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013
- [9] Moniruzzaman and Hossain "NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison" International Journal of Database Theory and Application, Vol. 6, No. 4, 2013
- [10] Baswanth "A Comparative Analysis of Different NoSQL Databases on Data Model, Query Model and Replication Model" emerging research in computing, information, communication and application eesview publication 2013
- [11] Hecht "NoSQL Evaluation A Use Case Oriented Survey" International Conference on Cloud and Service Computing (IEEE), 2011
- [12] Manoj "COMPARATIVE STUDY OF NOSQL DOCUMENT, COLUMN STORE DATABASES AND EVALUATION OF CASSANDRA" International Journal of Database Management Systems ( IJDBMS ) Vol.6, No.4, August 2014
- [13] Rajat Aghi, Sumeet Mehta, Rahul Chauhan, Siddhant Chaudhary and Navdeep Bohra "A comprehensive comparison of SQL and MongoDB Database" International Journal of Scientific and Research Publications, Volume 5, Issue 2, February 2015

