IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 5, Ver. V (Sep. - Oct. 2016), PP 70-74
www.iosrjournals.org

Correlation of NOSQL & SQL Database

Divya Kumawat[1], Aruna Pavate[2]
[1](Computer Dept., Atharva College of Engineering, Malad(W) , India)
[2](Computer Dept., Atharva College of Engineering, Malad(W) , India

***Abstract*:** *Now a day's technologies are moving towards motion user interface, Internet of things, Browser based IDEs etc. These technologies require real time response and large data store. Traditional SQL retrieves and manages database in a tabular form, but in current scenario of distributed large scale database SQL does not perform well. NoSql provides a better solution in this situation. This paper presents survey on NoSQL database; its classification on the basis of CAP theorem and different types of the NoSql databases with their properties. It also performs comparative study of SQL and NoSQL databases.*

Keywords- *Nosql, SQL, CAP theorem, NoSql DB*

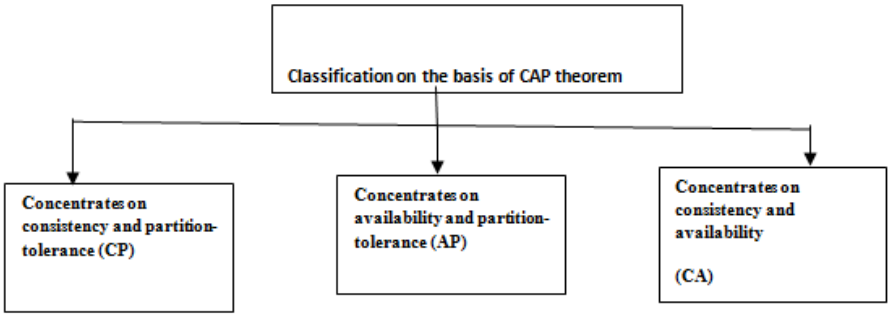
I. Introduction

Big Data handles the data sets of extremely big size like in terabytes, petabytes, exabytes, and zettabytes that are afar the ability of manual techniques. Big data works on large Volume, Variety, Velocity (3 V's) of Structured, Unstructured and Semi structured data. These data can be generated by human or machine. Data owners depend on greatly scalable tools to process and analyses the resulting dataset. Big data must provide quality attributes such as accuracy, consistency, completeness, pedigree, timeliness, precision, relevance while storing data [1]. SQL was introduced in 1970 and has emerged as a standard language for creating and managing Relational Database. The key elements of a SQL language are clause, expression, queries and statements. Many of the vendors have created their own SQL language extensions (example-POSTGRES, Starburst) for their own use. SQL uses MySQL, SQL Server, Access, Oracle, Sybase, DB2, and other database systems to access and manipulate data. The distributed relational database system attempts to resolve the huge problem of volume, but it does not support the segmentation for parallel processing. It supports vertical scaling where a single server is made powerful to handle increasing demand. It is possible to distribute SQL databases over many servers, but core relational features like JOINs, referential integrity and transactions will not be achieved [2].

II. NOSQL & its Databases

NOSQL database was developed after 2009. NOSQL is generally referred as 'not only SQL' [3]. It is an open source, horizontally scalable non-relational and distributed database. It provides flexible schema design, easy replication support, simple API and supports a huge amount of data. [4].

In the year 2000 Eric Brewer has proposed a CAP theorem which states that it is impossible to achieve availability, consistency, partition tolerance (the system continues its operation even in network failures) in a distributed system simultaneously. Only two conditions out of the three can be achieved. If consistency and availability (CA) are considered then that database will use replication approach to ensure data consistency and availability .If Consistency and partition tolerance (CP) are considered then such database system stores data at the distributed nodes, offering consistency on these data, but does not supports availability properly. If availability and partition tolerance (AP) are considered then such systems provides consistency [5]. NoSql databases can be categorized on the basis of CAP theorem and the way they store the data. NoSql database classifications on the basis of CAP theorem are listed below:



- BigTable (Column-oriented)
- Hypertable(Column-oriented)
- HBase (Column-oriented)
- MongoDB (Document)
- Terrastore (Document)
- Redis(Key-value)
- Scalaris (Key-value)
- MemcacheDB(Key-value)
- Berkeley DB (Key-value)
-
- Voldemort (Key-value)
- Tokyo Cabinet (Key-value)
- KAI (Key-value)
- CouchDB (Documentoriented)
- SimpleDB (Document-oriented)
- Riak(Document-oriented)
- Traditional relational DB
- Vertica (Column-oriented)
- Aster Data (Relational)
- Greenplum (Relational)

Figure 1-Classification of different NoSQL databases on the basis of CAP theorem.

III. Types of NOSQL Database

There are different types of NOSql databases having different architecture, query languages and consistency models. Some of the NoSql databases based on the way they store the data are listed as below:

A. **Key value database-** It is implemented in the form of hash table containing unique key and a pointer to a specific set of values; data can be queried with the help of unique key. It supports unstructured data as it imposes schema constraint over key value pair. Key-value databases masters session management, they are mainly used in serving the ad content and in managing user or product profiles. It is also used in the case where the data is represented in many different ways without a careful schema. Key value stores are generally used by Twitter to store tweets where Twitter id is a key and actual message, identity of the user and time of his post is a value. Figure2 demonstrates a key value database.

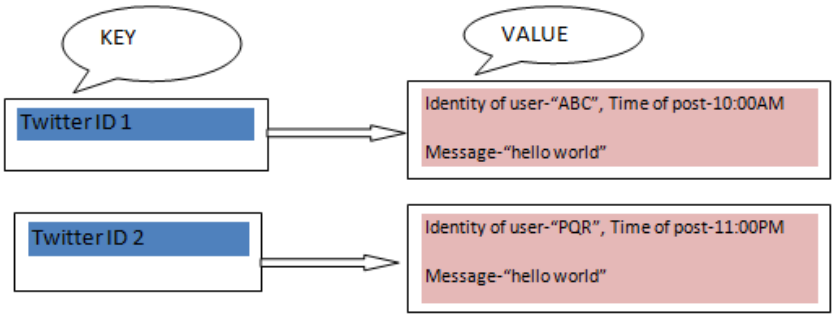


Figure2- Example of Key-value NoSQL database

Some of the leading key value DBMSes are Redis, riak ,Scalaris, MemcacheDB, Berkeley DB, Voldemort, Tokyo Cabinet and KAI [6]. Redis is an open source, BSD-licensed, memory data structure store. It provides LRU eviction, Lua scripting, built-in replication, different levels of on-disk persistence and transactions . It provides high availability through Redis Sentinel and automatic partitioning with Redis Cluster[7]. It supports a lot of data structures such as lists, strings, hashes, , sets, sorted sets for storing key. It facilitates user by providing time duration for set of record. The record will get automatically discarded as time gets elapsed [8]. Redis is best suited for caching, cookie storage, content distribution and management.

Another key-value NoSQL DBMS is Riak from Basho Technologies. Riak is an open source, master less, scalable, fault-tolerant, highly available, distributed multimodel DBMS. Riak is designed to support rapid development, ease of management, multimodal platform, supporting key-value search capabilities and object store all from the same platform.[9]

B. **Column Oriented Databases:** It stores groups of related data into column families and rows where each column of data works as an index of the database. All the entries in columns have indexes, so it is possible to search in only a part of the table. The number of columns stored in each record may differ as shown in table1. This database type is extremely scalable and works well with more complex datasets. It is well-suited for data where write operation is not performed on frequent basis and applications need to access a few columns of any rows all at once. Column databases are suitable for event logging, content management and counting/categorizing for analytics and online Analytical Processing [5]. Ex. Google.

Table 1: Example of Column oriented database

Row ID	Coloumns…..	
1	**Name**	**Address**
	ABC	PSC 450 Box no. 297 APO AP 96206
2	**Name**	**Address**
		C-202,Housing society, Indore-32061.

Some of the column oriented NoSql databases are: BigTable , Hypertable,Vertica and HBase . Hbase is a open-source, horizontally scalable, distributed database built on top of the Hadoop file system. Both Hbase and Big table provide random access. Hbase uses Hadoop's HDFS as storage which is responsible for random access of data.Some of the salient features of HBase are:

- It provides automatic sharding and replicates data across different nodes.
- It provides automatic fault tolerance and load balancing.
- It provides Snapshot support which helps in snapshots of metadata for getting the previous/ correct state form of data.

C. **Document Databases –**It stores the unstructured (text), semi-structured (XML) documents which are generally hierarchal in nature by using a markup language such as JavaScript Object Notation (JSON) or XML. Each record and its associated data are usually stored together in a single document; it not only simplifies the data access but also reduces time consuming joins or complex transactions. In this database, database administrators (DBAs) can modify the schema dynamically without downtime. This is the reason for using column oriented database in online shopping, event logging, content management and in-depth analytical processing. Its schema flexibility makes it suitable for rapid prototyping projects. It comprises of powerful query engines and indexing features that make it easy and fast to execute many different optimized queries.

MongoDB is a well known document database that stores data in a binary form. The key issues with mongo DB are data consistency and upper bound to document size, High-speed access to mass data[10]. When the data exceeds 50GB, MongoDB access speed is 10 times faster than MySQL[11]. It supports automatic data movement over different participants for load balancing. It contains a balancer for deciding when and which data should be migrated without disrupting application. It also provide MongoDB Management Service (MMS) a web tool for tracking databases, machines and backing up data[19]. MongoDB uses Indexing method for resolving queries efficiently. This indexing information is kept in RAM so there exist a index key limit providing a memory constraint on indexing. If a document exceeds index key limit than it will not create an index and will give error. [12]

Couchbase Server is another JSON-based, open source document database using a B-tree for storing key value couples. It provides possible consistency for transactions [13]. It not only offers command line interface (CLI) for administrative task, but also provides Web interface or RESTful API.

MarkLogic Server DB is an enterprise document database platform offering critical enterprise characteristics such as ACID transactions, security and automated failover. [14]. RavenDB, Pivotal GemFire, Apache jena are some other examples of document database [15].

Apache's CouchDB is also a JSON document oriented database providing multi-version concurrency control. It is best suited for devices that goes offline frequently (Ex. smartphones) as it can replicate to devices and will handle data synchronization when the device is again online[16].

D. **Graph Databases:** A graph database can be defined as a structure of nodes, edges, properties to represent and store data in the form of graph. In RDBMS nodes stand for records and edges or relationships are derived from some predefined column on a node but in graph database data is modelled as a network of relationships between the specific key value pairs. Facebook and Twitter are the key users of graph database. It supports HTTP and JSON out-of-the-box. Handling orientDB is quite simple; nodes can be added into it without complex configuration. [17] OrientDB, Neo4j and ArangoDB are some of the popular graph database available.Neo4j is An open source, schema-free NoSQL Graph Database written in java providing database enforced schema and vertical scaling [18]. Orient DB is a open source distributed Graph Database engine with the flexibility of a Document Database. It act as document-based database for schema less content at data level while it act as graph-oriented database for traversing the relationship. It uses OrientDB's SQL, a combination of SQL and graph operations. It uses SQL for basic operations and exploits some graph operators extensions to circumvent SQL joins in order to deal with relationships in data[19].

IV. Comparison of different NOSQL DB

Table 2: Comparison of different NOSQL databases

	Key value DB		Column oriented DB		Document DB		Graph DB	
Attributes/Databas Model	Riak	Redis	Hbase	Vertica	RavenDB [21]	Apache jena [20]	OrientDB	Neo4j
Programming Language	Erlang, C & some JavaScript	C	Java	-	.NET	Java	Java	Java
Operating System	Cross platform	Linux Mac OS Windows	Cross platform	Linux	Cross platform[19]	Cross platform	Cross platform	Cross platform
Query Language	Rest HTTP JavaScript	API Calls	API Calls XML Thrift	SQL	REST API	SPARQL	SQL-like query language (Note: no JOIN, but there are pointers)	Integrated pattern-matching-based query language ("Cypher")
Protocol	HTTP ,REST	Telnet like	Thrift HTTP/REST	UDP +SSL Authentication	HTTP	HTTP	binary, HTTP REST/JSON, or Java API for embedding	HTTP/REST (or embedding in Java)
Replication	Yes	Yes	Yes	Yes	Yes	No	Yes	yes
Replication Mode	Multi Master Replication	Master slave replication	Master-slave replication Master-master replication Cyclic replication[22]	Master-master replication	Master-master replication	None	Multi-master architecture	Master-slave cluster
Horizontal Scalable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

V. Difference between Relational Database and NOSQL

The differences between NoSQL and relational database are immense. It's fast and flexible. Being NoSQL as an open source, you can develop plugins to make it easy to work with. On the other hand, relational databases are more helpful to design more complex database, with relations between tables and a fix structure. It's a reliable database and even thought it moves slower, it is the base for complex structured databases. The main difference between the two databases is that relational Database has relations between the tables. Those relations can be one to one or one to many or many to many. With these relations you can join tables and make complex queries. The main problem with the relational Database is the replication. You cannot copy the database so easily as in NoSQL. You have tools which do that but aren't so fast. Relational database is a much slower database in comparison with NoSQL.

While there are numerous characteristics that differentiate SQL and NOSQL the two most significant are Scaling and Modeling.

- Scaling – Traditionally SQL does not lend itself to massively parallel processing, which lead to larger computers (scale up) vs. distribution to numerous commodity servers, virtual machines or cloud instances (scale out).
- Modeling – SQL databases are highly normalized and require pre-defined data models prior to inserting data into the system. In contrast NOSQL databases do not require (although they support) pre-defined data model(s).

Table 3: Difference between NoSQL and relational database

	SQL	NoSQL
1	Relational Databases(RDBMS)	Non-relational or distributed database
2	Vertically scalable	Horizontally scalable
3	Table based databases	Document based, key-value pairs, graph databases or wide-column stores.
4	Supports predefined schema	Supports dynamic schema
5	SQL (structured query language) for defining and manipulating the data	Uses unstructured Query Language
6	Standard interface for executing complex query	Not good for executing complex query
7	Best suited for huge load and complex transactional applications	Not suited for huge load and complex transactional type applications
8	SQL databases maintains on ACID properties (Atomicity, Consistency, Isolation and Durability)	NoSQL database follows the Brewers CAP theorem/BASE properties
9	Synchronous Inserts & Updates	Asynchronous Inserts & Updates

VI. Conclusion

The comparison study shows that, Nosql is found manageable and scalable. It provides better performance than relational database if database is quite large. NoSQL databases has became a valuable alternative to relational databases. Nosql databases are dynamic data model that are much better for managing large quantities of unstructured data; additionally its schema can be modified without downtime or service disruption.

Some of Nosql databases support ACID while some supports consistency (CASSANDRRA, Dynamo) and even some database such as SimpleDB, does not support transactions. Though a lot of development has occurred in the field of NoSql but it has not reached maturity. A survey conducted by Information Week highlighted that 44% of business IT experts had not heard of NoSQL and only 1% indicated that NoSQL was a part of their strategic direction [23]. Based on above description of databases the companies need to decide whether to use NoSQL or not. The selection will depend on companies business model, ACID transactions demand, cost and other requirements.

References

[1] Cai, L and Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era", *Data Science Journal, 14: 2, pp. 1-10, 2015*

[2] L. Davidson, K. Kline & K. Windisch, "Pro SQL Server 2005 Database Design and Optimization", *2006, p.47.*

[3] H. M. L. Dharmasiri , M. D. J. S. Goonetillake,,A Federated Approach on Heterogeneous NoSQL Data Stores, International Conference on Advances in ICT for Emerging Regions (ICTer): *IEEE Computer Society 234 – 239 , 2013*

[4] NOSQL: Your Ultimate Guide to the Non-Relational Universe! [online] Available at http://nosql-database.org/ [Accessed: 02 Oct 2016].

[5] Jaroslav, NoSQL databases: a step to database scalability in web environment ", *International Journal of Web Information Systems, Vol. 9 Iss: 1, pp.69 – 82,2013*

[6] B.G. Tudorica, C. Bucur, A comparison between several NoSQL databases with comments and notes," *10th Roedunet International Conference, vol., no., pp.1,5, 23-25 June 2011*

[7] Redis news [online] Available at : http://redis.io/ [Accessed: 20 May 2016]

[8] http://www.mongodbspain.com/en/2014/08/17/mongodb-characteristics-future/ [Accessed 14 June 2016]

[9] How to determine which NoSQL DBMS best fits your needs [online] Available at : http://searchdatamanagement.techtarget.com/feature/How-to-determine-which-NoSQL-DBMS-best-fits-your-needs [Accessed: 19 May 2016]

[10] Dileepa Jayathilake, Charith Sooriaarachchi, Thilok Gunawardena, Buddhika Kulasuriya and Thusitha Dayaratne ," A Study Into the Capabilities of NoSQL Databases in Handling a Highly Heterogeneous Tree", *IEEE 6th International Conference on Information and Automation for Sustainability pg 106-111,2012*

[11] Jing Han, Haihong E, Guan Le, Jian Du ,"Survey on NoSQL Database", *IEEE- 6th Int. Conf. On Pervasive Computing and Applications (ICPCA), 2011*

[12] https://docs.mongodb.com/v3.0/reference/limits/[Accessed 20 June 2016]

[13] Jaroslav Pokorny, "NoSQL databases: a step to database scalability in web environment*", International Journal of Web Information Systems Vol. 9 No. 1, 2013 pp. 69-82 , Emerald Group Publishing Limited 1744-0084 DOI 10.1108/17440081311316398,2013*

[14] Marklogic Datasheet, High vailability & disaster recovery, Available at- http://www.marklogic.com/wp-content/uploads/2014/09/MarkLogic-High-Availability-Disaster-Recovery-Interactive-Sept-2014.pdf

[15] 11 Open NOSQL Document Oriented Database ,Available at-https://dzone.com/articles/11-open-nosql-document[Acessed 1 Sept 2016]

[16] Pragati Prakash Srivastava; Saumya Goyal; Anil Kumar (SMIEEE)," Analysis of Various NoSql Database", *International Conference on Green Computing and Internet of Things (ICGCIoT),2015*

[17] OrientDB(R) [online] Available at: http://orientdb.com/ [Accessed: 20/05/16]

[18] What's New in Neo4j 2.3 [Online] available at: http://neo4j.com/whats-new-in-neo4j-2-3/ [Accessed 2 June 2016]

[19] https://en.wikipedia.org/wiki/CouchDB [Accessed 14 June 2016]

[20] Pragati Prakash Srivastava; Saumya Goyal; Anil Kumar (SMIEEE), Analysis of Various NoSql Database, *International Conference on Green Computing and Internet of Things (ICGCIoT),2015*

[21] https://ravendb.net/docs/article-page/3.0/Csharp/server/administration/backup-and-restore [Accessed 2 June 2016]

[22] http://www.cloudera.com/documentation/archive/cdh/4-x/4-2-0/CDH4-Installation-Guide/cdh4ig_topic_20_11.html [Accessed 2 June 2016]

[23] Information week, "Surprise: 44% of Business IT Pros Never Heard of NoSQL." [Online] Available at: http://www.informationweek.com/software/informationmanagement/surprise-44-of-business-it-pros-never-he/227500077 [Accessed: 02 Oct 2016].