A Comparative Study On: Nosql, Newsql and Polygot Persistence

Shagufta Praveen, Umesh Chandra

Abstract: ***After a long journey of decades, most of the leading web applications opted for non-relational database. Traditional database exist for so long but data mining application doesn't find relational database as a right choice for it. NoSQL movement was a question mark for the future of SQL. The High Volume, rich heterogeneity and speedy velocity of data generation in entire world is responsible for the Big Data. NoSQL was introduced to us for resolving scalability issues but consistency issue after scalability moved us from NoSQL to NewSQL. This paper emphasizes about NoSQL and NewSQL and it also highlights the reason for recent arrival of Polygot Persistence. Both technologies are distinguished with the help of some parameters (Models, Properties and as per Current Scenario need).***

Keywords: Big Data, Database, Polygot Persistence, NewSQL, NoSQL.

I. INTRODUCTION

Big Data is a big challenge for IT companies whereas, a new hope for researchers. It's a kind of heterogeneous mixture of Data (containing structured, semi-structured and unstructured, sensor, streaming, Continuous data). Survey says that 37.5% of leading companies are unable to analyze big data and it appears to be one of biggest question [1]. The fact is after having major experiences to deal with database, it's really hard to filter and interpret large and complex data in to useful field. Today almost all the organizations are using, discovering new tools and technologies that could enable researchers, marketers and analyzers to improve decision making for future statistics predictions. The reason behind is "any collection of data is so large and complex that it becomes really difficult to process it using on-hand database management tools [2]." In order to store and handle this massive data NoSQL terms evolved by Carlo Strozzi that can deal with the scalability issues [3]. But scalability was just a beginning; there were many more issues to deal with that evolved many more terms after it. This paper is majorly emphasizing on modeling techniques behind different evolved technologies and also reflecting their various characteristics.

II. NOSQL DATABASE: "NOT ONLY SQL"

Scalability was the main reason that introduced NoSQL to world. A Schemaless, open source, distributive and scalable technology that guarantees to offers high performance with fast data delivery. It supports horizontal scaling that means the ability to distribute over many servers with no RAM or Disk [4].

Revised Version Manuscript Received on March 22, 2017.

Shagufta Praveen, Research Scholar, Department of Computer Science, Glocal University, Saharanpur (Uttar Pradesh)-2471001, India.

Dr. Umesh Chandra, Assistant Professor, Department of Computer Science, Glocal University, Saharanpur (Uttar Pradesh)-2471001, India.

SQL dropped ACID (Atomic, Consistency, Isolation and Durability) property and also avoid operations like join, complex queries. In fact, it works on BASE (Basically Available, Soft state and Eventual Consistency) property. It also supports CAP (Consistency, Availability, Partition Tolerance) theorem which says in a distributed system during network failure (Partition tolerance) any one of the property among consistency or availability will be compromised. There are around 100 types of NoSQL databases in the world [5]. It favors horizontal scalability which was poor in RDBMS.

Consistency + Availability + Partition Tolerance = Perfect System

Consistency + availability + Partition Tolerance = Cannot be achieved

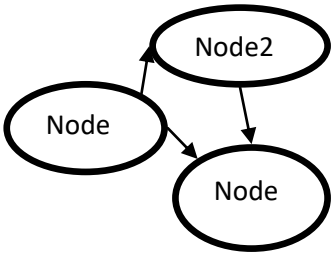
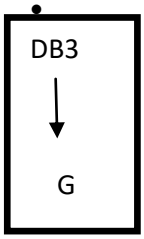
As per CAP Theorem, In a distributed System. If a system is in need of consistency and partition Tolerance then user would not be able to do read and write operations due to unavailability. If a system is in need of availability and partition tolerance then user would not be able to update the operation on the entire database hence consistency cannot be achieved.

A. Classification of NoSQL

NoSQL majorly classified in four groups:

1. Column-Oriented storage
2. Key/Value storage
3. Graphic data management system
4. Document Oriented system

Table 1. NoSQL Classification

Column-Oriented	Key-Value Stores	Graphs	Document Oriented
• It works on row and column model • It stores massive data • In this each row has many number of Columns • It is faster while querying • Each Row has unique Id for Columns • Each column has time stamp that store data while storing • Example→ Cassandra, Hbase • Row Id → Multiple Column	• A data model works on key-value pair [6] • A unique key is associated to a value. Key helps in extracting value from store • Key value data store stores data in allographic form [4] • It can store large amount of data • Replicated data in Key-value stores are stored in ring form • Key Value based data structures has data complexity of O(1). That makes data access easy and fast [7]. • Value can be anything like integer, array, string or object • It works efficiently in distributed data and not for structures and relation [8]. Key 1 → Value 1 • Example → Riak, Redis, Scalaris	• Based on Graph Theory • It consist of Node, edges and properties • It is Scalable and has high performance • Graph database allow all queries on the graph structures •  • In Graph Nodes → Entities Properties→Attributes Edges→Relationship • Example→Neo4j, Hyper Graph DB	• It refers to the loose set of key-value pairs[7] • It primary goal is to store document • Each document has unique Id to access a document from store • Example→ Simple DB, Couch DB • 

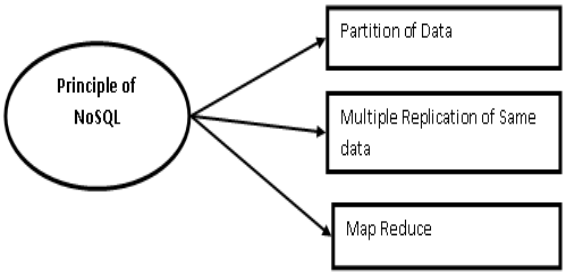


Fig. 1. Principle of NoSQL[9]

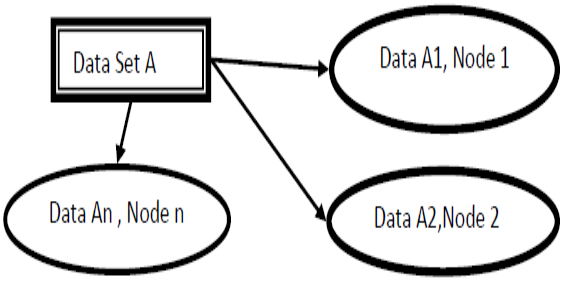


Fig. 2. Partition of Data

1. Partitioning of Data

This distribute data to different among different nodes (clusters).With this amount of load distributes and if any of the node fails then it doesn't stop other nodes to work.

Benefits: - High scalability, high performance, High availability, operational flexibility achieve.

2. Replication

Replication increases the availability. There can be various kinds of replication like intra-cluster and Multi-datacenter replication.

3. **Map Reduce: -** Map Reduce is made of two classes Mapper and Reducer that are designed to process a large amount of data.

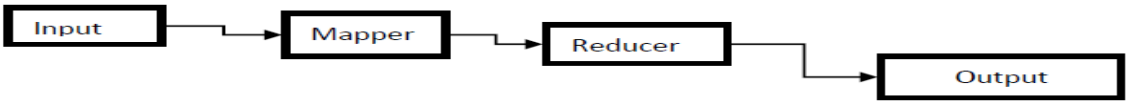


Fig.3. Map Reduce.

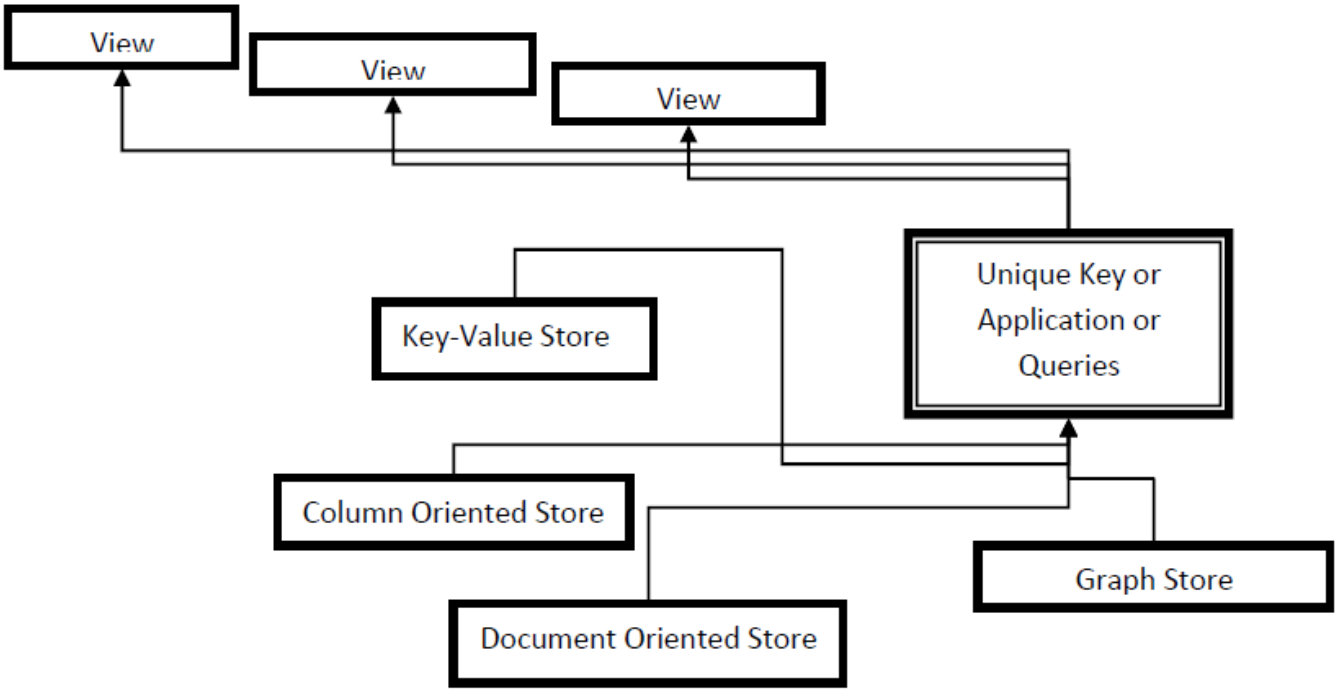


Fig. 4. Pictorial Representation of NoSQL

III. NEWSQL DATABASE

New SQL is a new technology after NoSQL. It is proposed to preserve ACID properties of Relational Database to support OLTP and to achieve scalability factor of NOSQL. Basically it is a combination of NoSQL (Scalable) properties with Relational (ACID) Properties. It helps in executing read-write transactions that are short-lived and operated using index loops and execute same number of queries with different input. It works on lock free concurrency control and share-nothing architecture [10].

A. Some features of New SQL:

Table No.2

S. No.	Features of New SQL
1.	Support Relational Model
2.	Has Schema
3.	Support SQL
4.	Horizontally scalable
5.	Has ACID Properties

B. Classification of New SQL:

New SQL is classified on the base of their implementation

1. Novel Systems
2. Middle ware
3. Database as a Service

➔ **Novel Systems:** Those are built on new architecture. The importance of new architecture is based on distributed architecture that support fault-tolerance, replication and multi node concurrency control. Another part all the systems are optimized for multi-node environment through this NewSQL can send query between nodes [10].DBMS under this systems are written from scratch and include components like flow control and distributed query processor [15].

➔ **Middle Ware:** Database of an organization splits into multiple shards (Shards refers to a small parts of a whole. This is a technique of dividing a very large database into small parts called data shards).Middle ware also supports SQL that achieve OLTP techniques. This way an organization can scale their database across multiple nodes [10].

➔ **Database as a service:** Database as a service is service of cloud that allow users to access database without setting up physical hardware. With its help an organization does not work for private hardware . Users have URL Connection for the DBMS and also an interface provided to them to control a system[10]. Recently, Google announced a new technology named Cloud Spanner that has RDBMS properties with scalability. Products offering SQL properties with NoSQL falls under 'NewSQL' term [11] .Application supported by NewSQL afford very large figure of transactions inspite of using a small part of data retrieving process [12].

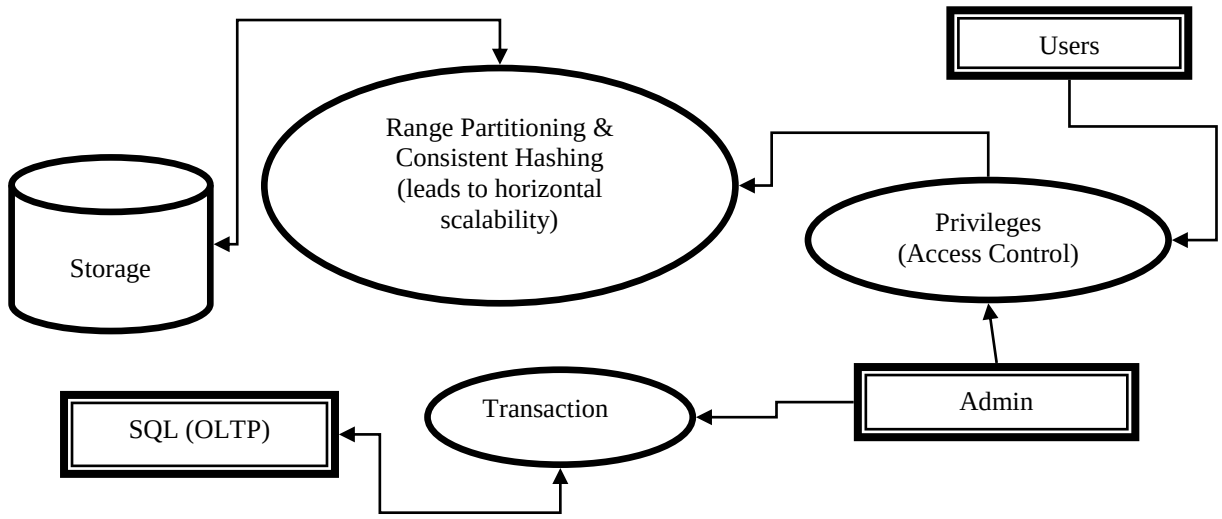


Fig.5. Architecture of New SQL

IV. POLYGOT PERSISTENCE

Ploygot Persistence is a new technology derived from the idea of polygot programming. Polygot programming is a technique to develop an application with the help of various programming languages, same way polygot persistence is a technology of developing an application that supports various database. This technology solves the volume and consistency issues all together. As per the components database can be chosen for better working of an application.

A. Architecture of Polygot Persistence

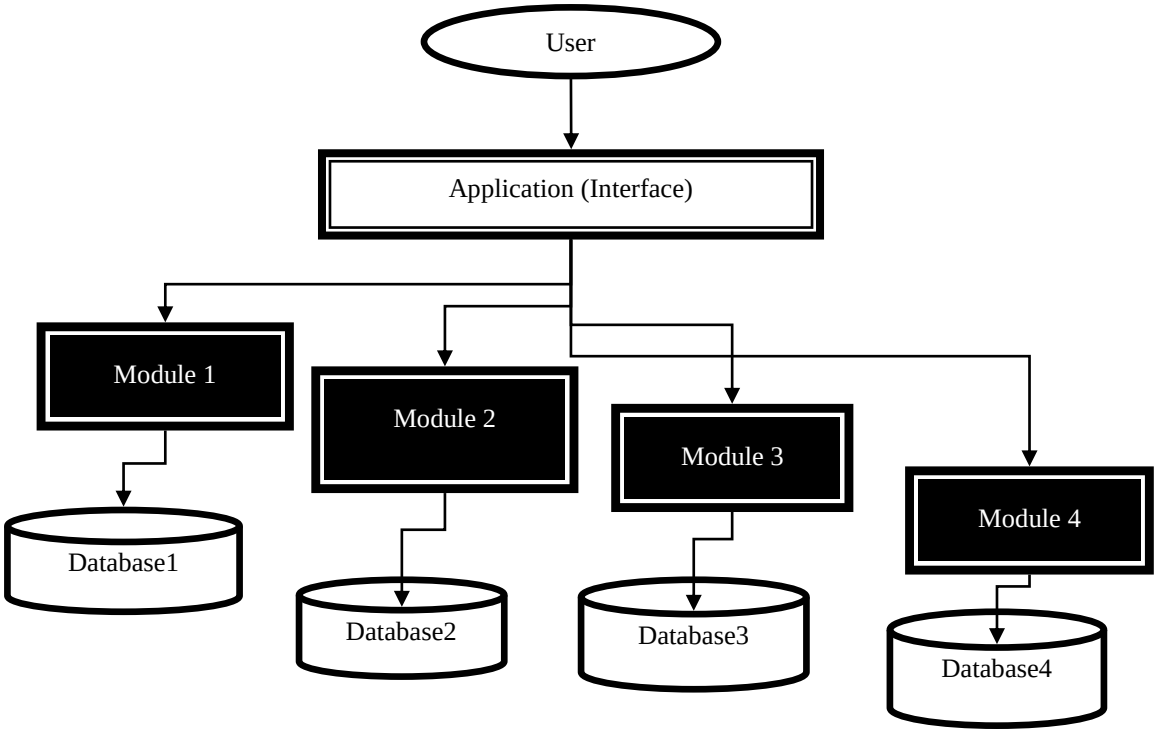


Fig. 6. Architecture of Polygot Persistence

In the above figure a user interacts with interface of the application to access information. Application can be any software that is integrated form of different modules. Every module has a specific function of features. For an example: An E-commerce website which has different modules like a module that handle user session , another module that handle shopping cart and some other module that carry catalog of products and transaction(Financial data). Each module should be linked to a specific database that can make the application best to use. Transaction should be linked with RDBMS for consistent processing and for better OLTP. Catalog products should be linked with Column oriented so that it can store large amount of data and Module like user session should be linked with Key-value store so that instant access can take place.[13]

V. LANGUAGES

Table No. 3. Languages Used

No SQl	New SQL	Polygot	RDBMS
Simple Call level Interface or Protocol [17]	It uses SQL as main Language In order to provide OLTP , SQL is preferred	Every module connected to database has specific logic for data access	It Uses SQL as main Language

VI. CONCURRENCY CONTROL

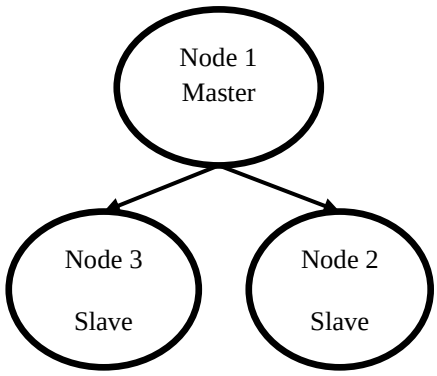
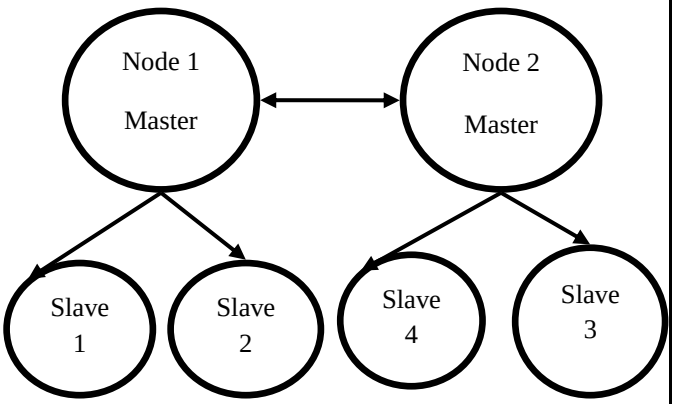
A mechanism that is used to address the conflict during transaction on a database by multiple user at a time[14].

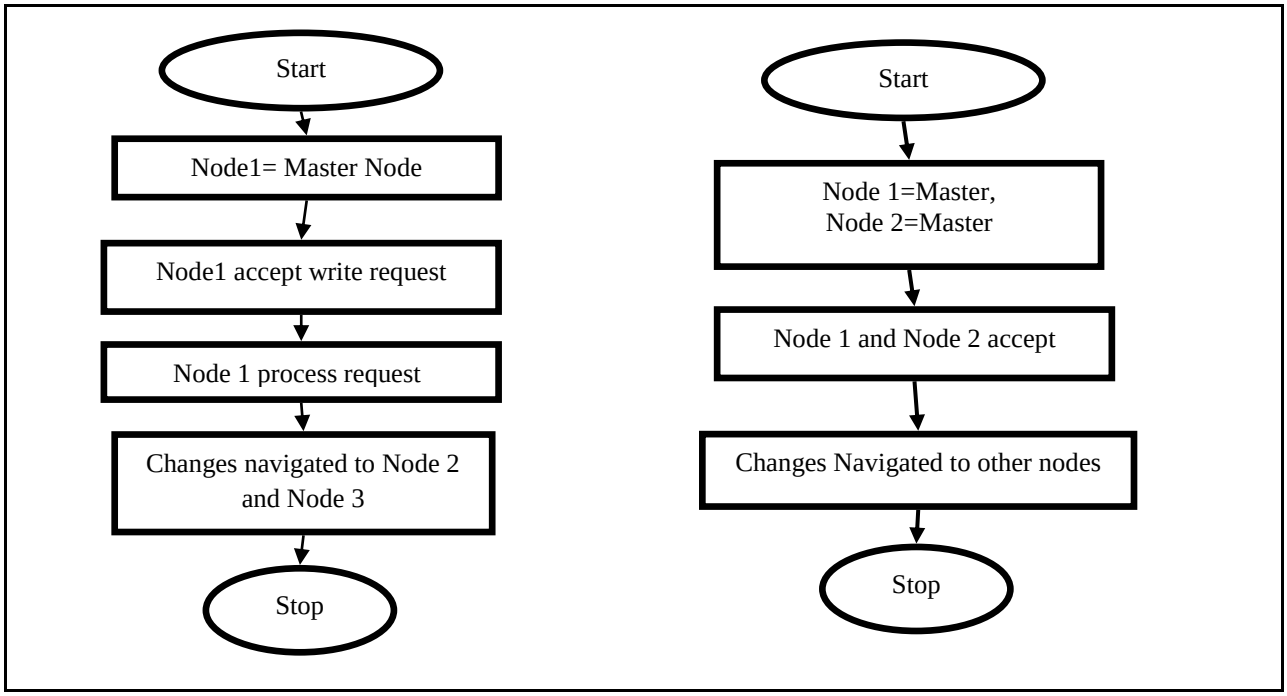
Tableno.4. Categories of Concurrency Control

Pessimistic Concurrency Control	Optimistic Concurrency control	Multiple version concurrency control	Hybrid Concurrency Control
→ When two or more users work on same tuple and try to update it at same time. A lock is used to prevent another user to update so that one user can update it at a time.	→To avoid conflict in database they use roll back command at the end of operation	→Data stored in various version and among all versions, one of the version addressed as current and other marked as old. Updation are made to be done on current version and there is nothing to be done with Lock mechanism	→It is a combination of Lock and Lock free mechanism for avoiding conflict between data of the database [8]
→Used by 1. RDBMS 2. Redis (Key-value, NoSQL) 3. Memcached (Key-value ,NoSQL)	→Used By 1. Amazon SimpleDB 2. CouchBase Server (Document store,NoSQL)	→Used by 1. Voldemort(Key-value, NoSQL) 2. Riak (Key-value, NoSQL) 3. HBase (Column –oriented ,NoSQL) 4. HyperGraphDB(Graph Database ,NoSQL) 5. NuoDB (NewSQL)	→Used by 1. Google Spanner (NewSQL)

VII. REPLICATION

Data Replication is a technique of making various copies of one original copy in order to achieve better performance and high availability. Every technology uses different types of replication. Replication basically is of two types: Master-slave and master-master. Table No. 5.

Master-Slave Mechanism	Master-Master Mechanism
	
Used by **→Redis** **→MongoDB**	**Used By** **→ CouchDB** **→NouDB**



VIII. CONCLUSION

This paper represents different technologies of database. These technologies are described with the help of architecture, classifications and mechanisms. It summarized the recent technologies that gave a new world and new research areas to database. A long journey from RDBMS to NewSQL [16] evolved terms like Google spanner and Polygot persistence and these terms are evidence of new work in database field. The upcoming researches are new hope for database evolution.

REFERENCES

1. S. J. Veloso, 2015, Data Analytics Topic: Big Data [Online]. Available:http://community.mis.temple.edu/sjveloso/data-analytics-topic-big-data/
2. U.Banerjee,21 december 2012, Technology Trend Analysis[Online].Available:https://setandbma.wordpress.com/2012/12/21/definition-of-big-data/
3. [Online].Available:https://en.wikipedia.org/wiki/NoSQL
4. V. Sharma and M. Dave ,SQL and NoSQL Database ,Internation Journal of advance research in computer science and software engineering,2012
5. Jose J, Subramoni H, Miao L, Minjia Z, Jian H, Wasiur M. Memcached design on high performance RDMA capable interconnects[C]. Parallel Processing(ICPP), 2011 IEEE International Conference on:743-752.
6. R. Hetch and S. jablonski ,NoSQL evaluation: A use case oriented survey, proceeding CSC'11 Proceedings of the 2011 International conference cloud and Service computing, 2011
7. C. He,Survey on NoSQL Database technology ,JOURNAL OF APPLIED SCIENCE AND ENGINEERING INNOVATION,2015
8. Grolinger et.al.,Database management in cloud environment: NoSQL and NewSQL DataStore, Journal of Cloud computing, Advances, system and application 2013.
9. [Online].Available:http://natishalom.typepad.com/nati_shaloms_blog/2009/12/the-common-principles-behind-the-nosql-alternatives.html
10. A. Pavlo and M. Asslett What's really new with NewSQL by, SIGMOD Record,June 2016
11. Google-Launches Cloud –spanner-A newSQL databse for enterpeise by Jankiran MSC
12. [Online].Available:www.technopedia.com/29093/newsql
13. [Online].Available:http://www.jamesserra.com/archive/2015/07/what-is-polyglot-persistence/
14. [Online].Available:databasemanagement.wikia.com/wiki/Concurrency_Control
15. ABM Moniruzzamam , New SQL:Towards Next Generation Scalable RDBMS for Online Transaction processing for Big Data Management,2016
16. S. Praveen et.al., A literature review on evolving database, International journal of computer Application, 2017
17. J.M. Monterio et.al ,What comes after NoSQL? NewSQL: A New Era of Challenges in DBMS Scalable Data Processing ,2016