**ASSIGNMENT NO. 2**

**Aim**

Study of NoSQL Databases such as Hive/Hbase/Cassendra/DynamoDB

**Objective**

* Understand NoSQL Database concepts
* To understand the purpose of NoSQL databases
* To Learn the CURD Operation in NoSQL Databases

**Theory**

**Introduction**

Efficient Storage and retrieval of data with availability and scalability is the main purpose of NoSQL databases. NoSQL does not stand for no to SQL; it means “NOT ONLY SQL”. NoSQL database is just an alternative to traditional relational database. The industry of database has seen an introduction of many non relational databases such as MongoDB, Hbase , Neo4j in last few years. Depending upon the business requirement and strategy a cloud vendor can go with any of the database type.

**Importance Of NoSQL**

For last couple of years, SQL vs. NoSQL has been emerged as a heated argument over the Internet. The argument “SQL vs NoSQL,” actually talks about relational versus nonrelational databases. Because of normalized data model and enforcement of strict ACID properties, traditional relational database is considered to be a schema based transaction oriented database. It requires a strict predefined schema prior to storing data into it. Redefining a schema in case of a future change, once after data got inserted into the database is disruptive. Whereas in the era of Big Data, there is a constant need for adding new types of data to enrich the applications. Again the storage solution of relational database can make a big impact on speed and scalability. Web services like Amazon and Google have terabytes and petabytes of data stored in their big data centers and have to respond to massive read-write requests without a noticeable latency. To scale a relational database, data needs to get distributed on multiple servers. Before providing to the application the desired information has to be collected from many tables and combined. Similarly, while writing data also; it has to be performed on many tables in a coordinated manner. For any application, it could be a bottleneck to handle tables across multiple servers. In relational databases “join” operation slow-down the system to a crawl, especially when millions of users are doing lookups against tables with millions of rows of data. Large scale web services such as Google, Amazon, Yahoo, Facebook found these to be the cases to develop their own non-relational database in order to meet the scalability and performance needs.

**Features of NoSQL**

NoSQL databases may not require a predefined table schema, typically scale horizontally and usually avoid join operations. Because of schema less nature and involvement of smaller subset analysis of NoSQL system, this database can be better described as structured data stores. Three important basic features of NoSQL databases are scale-out, flexible data structure and replication, which are explained as follows.

* Scale-out: Scaling out refers to achieve high performance in a distributed environment by using many general-purpose machines. NoSQL databases allow the distribution of the data over a large number of machines with a distributed processing load. Many NoSQL databases allow automatic distribution of data to new machines when they are added to the cluster. Scale-out is evaluated in terms of scalability and elasticity.
* Flexibility: Flexibility in terms of data structure says that there is no need to define a schema for databases. NoSQL databases do not require a predefined schema. This allows the users to store data of various structures in the same database table. However, support for high-level query languages such as SQL is not supported by most of the NoSQL databases.
* Data Replication: One of the features of NoSQL databases is data replication. In this process a copy of the data is distributed to different systems in order to achieve redundancy and load distribution. However there is a chance of losing data consistency among the replicas. But it is believed that sometimes this consistency may be achieved eventually. Consistence and availability are the factors for evaluating replication.

**NoSQL Data Models**

These are some categories of NoSQL database models discussed as follows.

1. Key-Value Data Stores
2. Document Oriented Data Stores
3. Column Family Data Stores
4. .Graph Database

**Comparision of Different Nosql Databases**

