**ASSIGNMENT 12.3**

**1.Explain the uses of NoSQL Databases in brief and list the current NoSQL** **databases which are available.**

When compared to RDBMS, NoSQL databases are more scalable and provide superior performance, and their data model addresses several issues that the relational model is not designed to address:

* Large volumes of rapidly changing structured, semi-structured, and unstructured data
* Agile sprints, quick schema iteration, and frequent code pushes
* Object-oriented programming that is easy to use and flexible
* Geographically distributed scale-out architecture instead of expensive, monolithic architecture

The following are situations when nosql is used

* Need to handle large volumes of structured, semi-structured, and unstructured data
* Follow modern development practices such as agile sprints, quick iterations, and frequent code pushes
* Prefer object-oriented programming that is easy to use and flexible
* Want to leverage efficient, scale-out architecture instead of expensive, monolithic architecture

Current NoSQL databases which are available:

Scylla

Accumulo

Cassandra

Druid

Apache Flink

Hadoop/HBase

Vertica

Hypertable

Cloudera

SAP HANA and etc.

**2.Explain the Differences between Hive and HBase in Brief with examples.**

Hive should be used for analytical querying of data collected over a period of time - for instance, to calculate trends or website logs. Hive should not be used for real-time querying since it could take a while before any results are returned.

HBase is perfect for real-time querying of Big Data. Facebook use it for [messaging](https://www.facebook.com/note.php?note_id=454991608919) and real-time analytics. They may even be using it to count Facebook likes.

Hive is an SQL-like engine that runs MapReduce jobs, and HBase is a NoSQL key/value database on Hadoop.

Hive can be used for analytical queries while HBase for real-time querying.

Apache Hive is a data warehouse infrastructure built on top of Hadoop. It allows for querying data stored on HDFS for analysis via HQL, an SQL-like language that gets translated to MapReduce jobs. Despite providing SQL functionality, Hive does not provide interactive querying yet - it only runs batch processes on Hadoop.

Apache HBase is a NoSQL key/value store which runs on top of HDFS. Unlike Hive, HBase operations run in real-time on its database rather than MapReduce jobs. HBase is partitioned to tables, and tables are further split into column families. Column families, which must be declared in the schema, group together a certain set of columns (columns don’t require schema definition). For example, the "message" column family may include the columns: "to", "from", "date", "subject", and "body". Each key/value pair in HBase is defined as a cell, and each key consists of row-key, column family, column, and time-stamp. A row in HBase is a grouping of key/value mappings identified by the row-key. HBase enjoys Hadoop’s infrastructure and scales horizontally using off the shelf servers.