**ASSIGNMENT NO.**

**Aim:** To study basic operations in Cassandra.

**Objective:** To implement basic operations on Cassandra using keyspace.

**Theory:**

**Cassandra -**

* Cassandra is an open source distributed database management system.
* Scalability and high availability without compromising performance is its major feature.
* Suitable for applications that cannot afford to lose data, even when an entire data center goes down.
* Every node in a cluster is identical.
* Hence, there is no single point of failure.
* Tables may be created,dropped and altered at runtime without blocking updates and queries.
* Cassandra does not support joins and subqueries.
* It emphasizes denormalization.
* It uses Google's Snappy data compression algorithm.
* Has Cassandra Query Language (CQL).

**Users of Cassandra -**

Cassandra was developed at Facebook for inbox search. It was open-sourced by Facebook in July 2008.Users of Cassandra are:

* Facebook
* Apple
* Wikimedia
* Twitter
* IBM ,etc

**Architecture of Cassandra -**

All the nodes in a cluster play the same role. Each node is independent and at the same time interconnected to each other nodes. Each node in a cluster can accept read and write requests, regardless of where the data is actually located in the cluster. When a node goes down, read/write requests can be served from other nodes in the network.

Each node communicates with each other through the Gossip protocol, which exchange information across the network every second. A commit log is used on each node to capture write activity. Data durability is assured. Data also written to an in-memory structure (memtable) and then to disk once the memory structure if full (an SSTable).

**Internode Communication (Gossip) -**

Gossip is a peer-to-peer communication protocol. Here nodes periodically exchange state information about themselves and about other nodes they know about. It runs every second and exchanges state messages with upto three nodes in the cluster. The nodes exchange information about themselves and about the other nodes they have gossiped about.

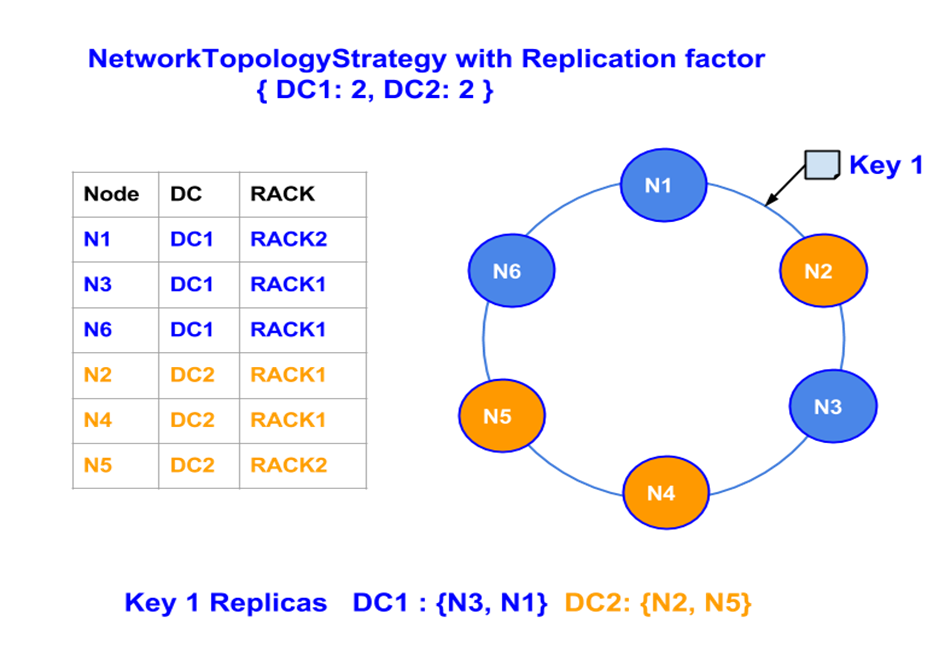
**Data Replication in Cassandra -**

Cassandra stores replicas on multiple nodes to ensure reliability, fault tolerance.

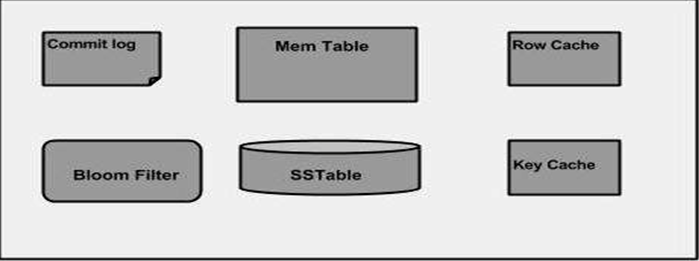
The total number of replicas across the cluster is referred to as the replication factor.

Replication strategies available are:-

1. Simple strategy: Uses for a single data center only.
2. Network Topology Strategy: Highly recommended for most deployments because it is much easier to expand to multiple data centers when required by future expansion.
3. Old Network Topology: This is a legacy replication strategy.



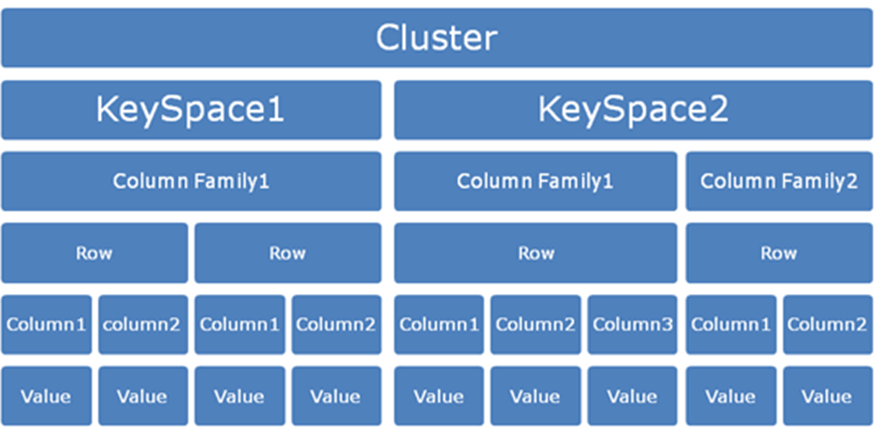
**Components of Cassandra -**



1. Node: location where data is stored.
2. Data center: collection of related nodes.
3. Cluster: a component that contains one or more data centers.
4. Commit log: It is a crash-recovery mechanism in cassandra. Every write operation is written to the commit log.
5. Mem-table: A mem-table is a memory-resident data structure.
6. SSTable: It is a disk file to which the data is flushed from the mem-table when its contents reach a threshold value.
7. Bloom Filter: This is a quick, non-deterministic, algorithms for testing whether an element is a member of set.

**Cassandra Query Language -**

* KEYSPACE:-

A cluster is a container for keyspaces- typically only one. A keyspace is the outermost container for data in Cassandra.

**CREATE a KEYSPACE:-**

Syntax:-

CREATE KEYSPACE <identifier> WITH <properties>

i.e.

CREATE KEYSPACE “KeySpace Name” WITH replication = {'class': ‘Strategy name’, 'replication\_factor' : ‘No.Of replicas’} AND durable\_writes = ‘Boolean value’;

Durable writes:-

By default it is set to 'true'.

**USING a KEYSPACE:-**

Syntax:-

USE <identifier>;

**ALTER a KEYSPACE:-**

Syntax:-

ALTER KEYSPACE <identifier> WITH <properties>;

i.e.

ALTER KEYSPACE “Keyspace Name” WITH replication = {'class': 'Strategy name',replication\_factor':'No. Of replicas'};

**DROP a KEYSPACE:-**

Syntax:-

DROP KEYSPACE <identifier>;

i.e.

DROP KEYSPACE “Keyspace name”;

**CREATING a TABLE-**

Syntax:-

Create (Table|ColumnFamily) <tablename> ('<column\_definition>' , '<column\_definition>') (WITH <option> AND <option>)

**CREATING an INDEX**:-

Syntax:-

CREATE INDEX <identifier> ON <tablename>;

**Input:** CQL commands.

**Output:** Results of the CQL commands and queries.

**Platform:** OS-64 bit Ubuntu 12.04, cqlsh(5.0.1), Cassandra(2.1.9), jdk 8.

**Conclusion:** Thus we have studied and implemented the CQL.