● Nosql Databases

To define **NoSQL**, it is helpful to start by describing SQL, which is a query language used by RDBMS. Relational **databases** rely on tables, columns, rows, or schemas to organize and retrieve data. In contrast, **NoSQL databases** do not rely on these structures and use more flexible data models.

Examples are like HBase, MongoDB etc.

● Types of Nosql Databases

There are 4 basic types of NoSQL databases:

1. **Key-Value Store** – It has a Big Hash Table of keys & values {Example- Riak, Amazon S3 (Dynamo)}
2. **Document-based** **Store- It**stores documents made up of tagged elements. {Example- CouchDB}
3. **Column-based Store-**Each storage block contains data from only one column, {Example- HBase, Cassandra}
4. **Graph-based**-A network database that uses edges and nodes to represent and store data. {Example- Neo4J}

● CAP Theorem

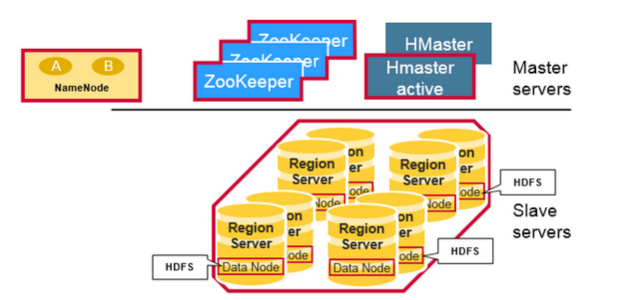
The **CAP theorem**, also named **Brewer's theorem** after computer scientist [Eric Brewer](https://en.wikipedia.org/wiki/Eric_Brewer_(scientist)), states that it is impossible for a [distributed data store](https://en.wikipedia.org/wiki/Distributed_data_store) to simultaneously provide more than two out of the following three guarantees

|  |  |  |
| --- | --- | --- |
| ***Consistency*** | [***Availability***](https://en.wikipedia.org/wiki/Availability) | [***Partition tolerance***](https://en.wikipedia.org/wiki/Network_partitioning) |
| Every read receives the most recent write or an error | Every request receives a (non-error) response – without guarantee that it contains the most recent write | The system continues to operate despite an arbitrary number of messages being dropped (or delayed) by the network between nodes |

In other words, the CAP theorem states that in the presence of a network partition, one has to choose between consistency and availability.

● HBase Architecture

HBase is composed of three types of servers in a master slave type of architecture. Region servers serve data for reads and writes. When accessing data, clients communicate with HBase RegionServers directly. Region assignment, DDL (create, delete tables) operations are handled by the HBase Master process. Zookeeper, which is part of HDFS, maintains a live cluster state.The Hadoop DataNode stores the data that the Region Server is managing. All HBase data is stored in HDFS files. Region Servers are collocated with the HDFS DataNodes, which enable data locality (putting the data close to where it is needed) for the data served by the RegionServers. HBase data is local when it is written, but when a region is moved, it is not local until compaction.The NameNode maintains metadata information for all the physical data blocks that comprise the files.



● HBase vs RDBMS

