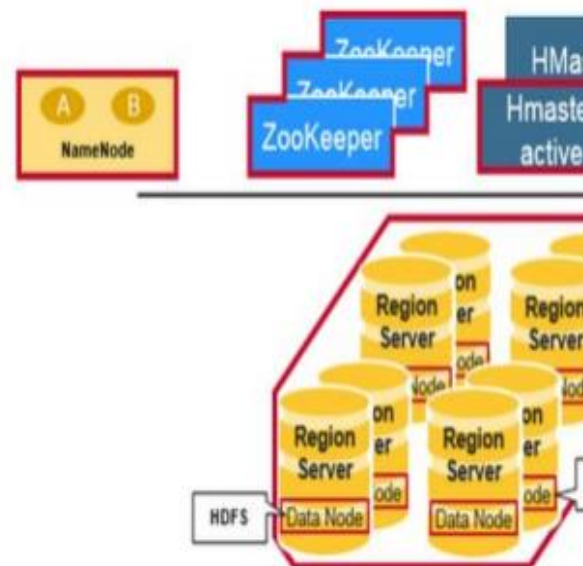


# HBase

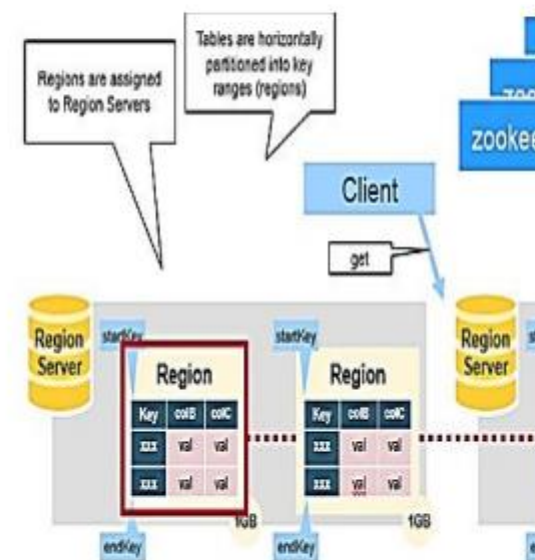
## Architecture

HBase is composed of three types of servers in a master slave type of architecture.

- Region servers serve data for reads and writes.
- HBase Master process handles the Region assignment, DDL (create, delete tables) operations
- Zookeeper 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.
- The NameNode maintains metadata information for all the physical data blocks that comprise the files.



- HBase Tables are divided horizontally by row key range into “Regions.”
- A region contains all rows in the table between the region’s start key and end key.
- Regions are assigned to the nodes in the cluster, called “Region Servers,” and these serve data for reads and writes.
- A region server can serve about 1,000 regions.



## H Master

Region assignment, DDL (create, delete tables) operations are handled by the HBase Master.

A master is responsible for:

- Coordinating the region servers
- Assigning regions on startup
- Re-assigning regions for recovery or load balancing
- Monitoring all RegionServer instances in the cluster (listens for notifications from zookeeper)

#### Admin functions

- Interface for creating, deleting, updating tables

