

## 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 vs RDBMS

Hadoop and RDBMS are varying concepts of processing, retrieving and storing the data or information. While Hadoop is an open-source Apache project, RDBMS stands for Relational Database Management System. Hadoop framework has been written in Java which makes it scalable and makes it able to support applications that call for high performance standards. Hadoop framework enables the storage of large amounts of data on files systems of multiple computers. Hadoop is configured to allow scalability from a single computer node to several thousands of nodes or independent workstations in a manner that the individual nodes utilize local computer storage CPU processing power and memory.

Traditional RDBMS is utilized to handle relational data while Hadoop works well with structured as well as unstructured data, supporting multiple serialization and data formats such as Text, Json, Xml, Avro and more. The distribution of Hadoop is done from the ground up with the option of adding more nodes in order to boost capacity.