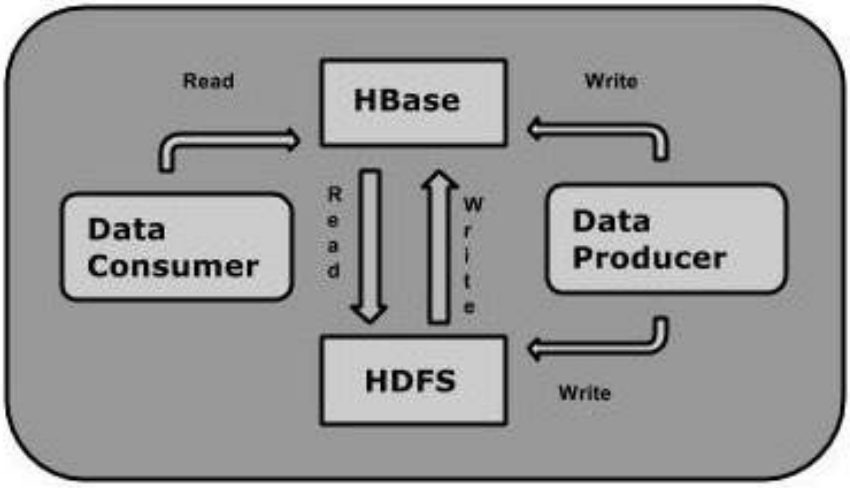
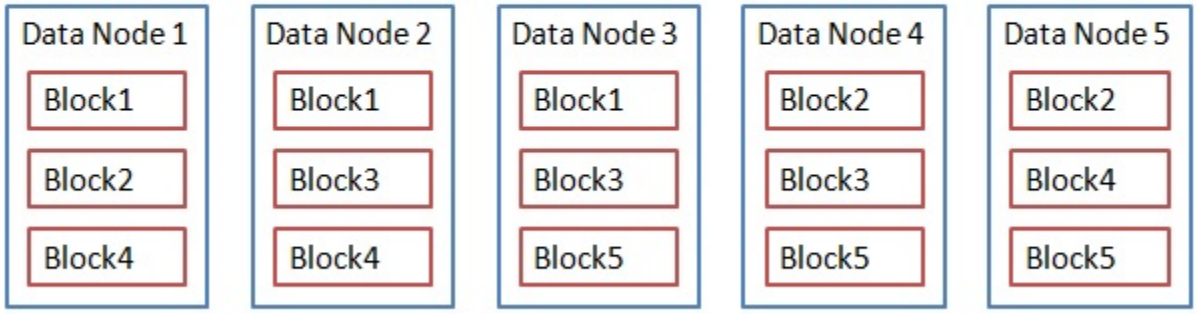


HBase is open-source software that allows to manage data on top of HDFS.

Features :

- Apache project (Open source)
- Provides data replication

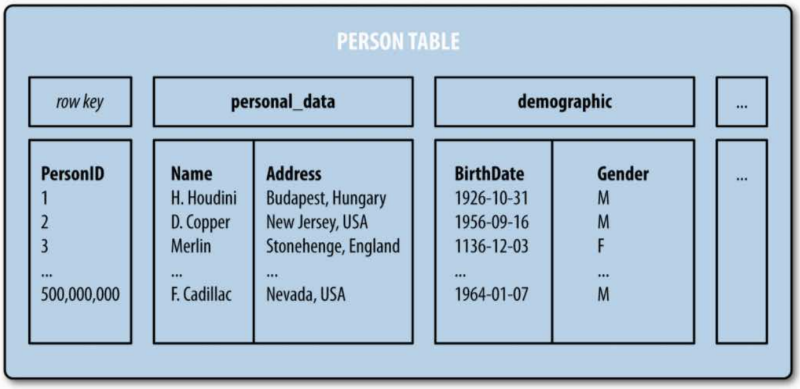


Replication Factor: 3

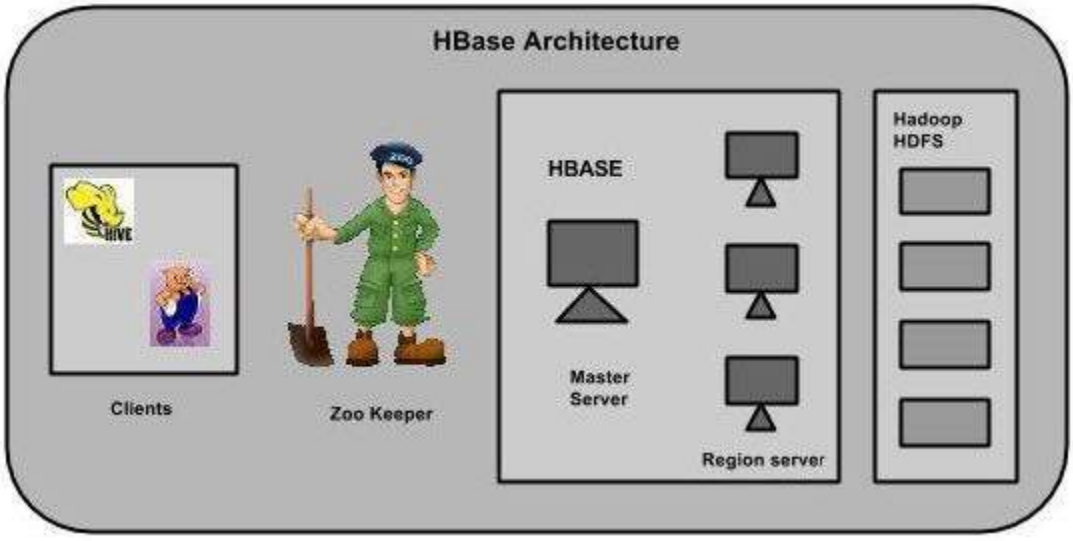
- Use HDFS as a storage mechanism
- Provides fast access on top of HDFS
- Random read-and-write operations
- Provide Rest API to access tables
- Horizontal scalable
- Hbase uses hash tables for random access to provide a fast lookup.
- Low latency to access single records
- Support processing with Hadoop and Spark
- Google BigTable is referenced to implement HBase

- It is a columnar database

Rowid	Column Family			Column Family			Column Family			Column Family		
	col1	col2	col3	col1	col2	col3	col1	col2	col3	col1	col2	col3
1												
2												
3												



HBase Components



- Hadoop HDFS is used to store data
- Hive, Pig, Spark are used to process data
- Rowkey is a key for every wor in HBase. Hence you can search specific data
- Master is used to monitor all region servers (failover, load balancing)
- Region server runs on HDFS datanode to execute all read and write operations
- Zookeeper is used for configuration management.

Create table

create ‘<table name>’,’<column family>’

create 'emp', 'personal data', 'professional data'

list table

list

Disable table

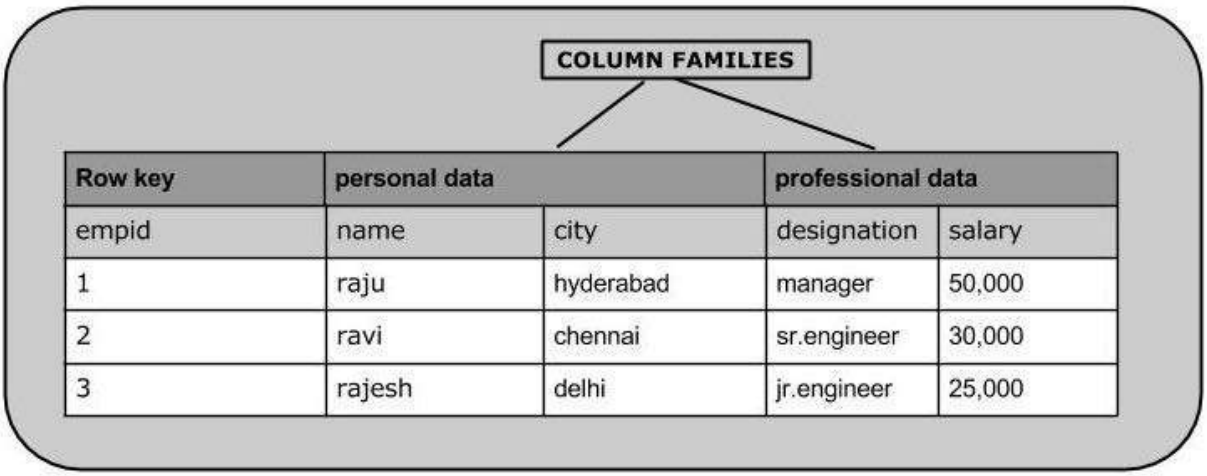
disable ‘emp’

enable ‘emp’

Delete table

drop 'emp'

Create data



Row key	personal data		professional data	
empid	name	city	designation	salary
1	raju	hyderabad	manager	50,000
2	ravi	chennai	sr.engineer	30,000
3	rajesh	delhi	jr.engineer	25,000

put 'emp','1','personal data:name','raju'

put 'emp','1','personal data:city','hyderabad'

put 'emp','1','professional data:designation','manager'

put 'emp','1','professional data:salary','50000'

Update Data

put 'emp','1','personal:city','Delhi'

Read Data

get 'emp', '1'

get 'emp', 'row1', {COLUMN ⇒ 'personal:name'} //read specific column

Delete data

deleteall 'emp','1'

delete 'emp', '1', 'personal data:city'

Scan

```
scan 'emp'
```