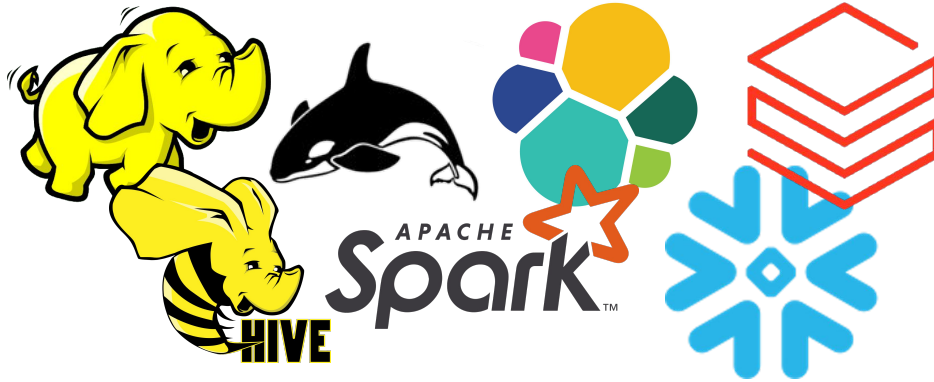


# Big Data



## NoSQL with HBase

# What is NoSQL?

## 1. Key-value:

- 1 key = 1 value
- Examples : Redis, Memcache (in memory)

## 2. Columnar:

- 1 key = X values stored in columns
- Examples : HBase, Cassandra

# What is NoSQL?

## 3. Document stores:

- JSON/XML like objects
- Examples : MongoDB, CouchDB

## 4. Graph:

- Store nodes and relationships between them
- Examples : Neo4j, JanusGraph

# Apache HBase

Is the Hadoop database, a distributed, scalable, big data store - [hbase.apache.org](http://hbase.apache.org)

- Random, realtime read/write access to Big Data
- Inspired from Google BigTable paper (2006)
- Stores data in HDFS

Companies such as Facebook, Twitter, Yahoo, and Adobe use HBase internally.

# The CAP Theorem

No distributed storage system can ensure more than 2 of those properties at the same time:

- **Consistency:** For one given query, all the nodes return the same result
- **Availability:** Every query receives a fast response, without guarantee that this is the latest value
- **Partition tolerance:** The system continues to work even if some nodes are disconnected

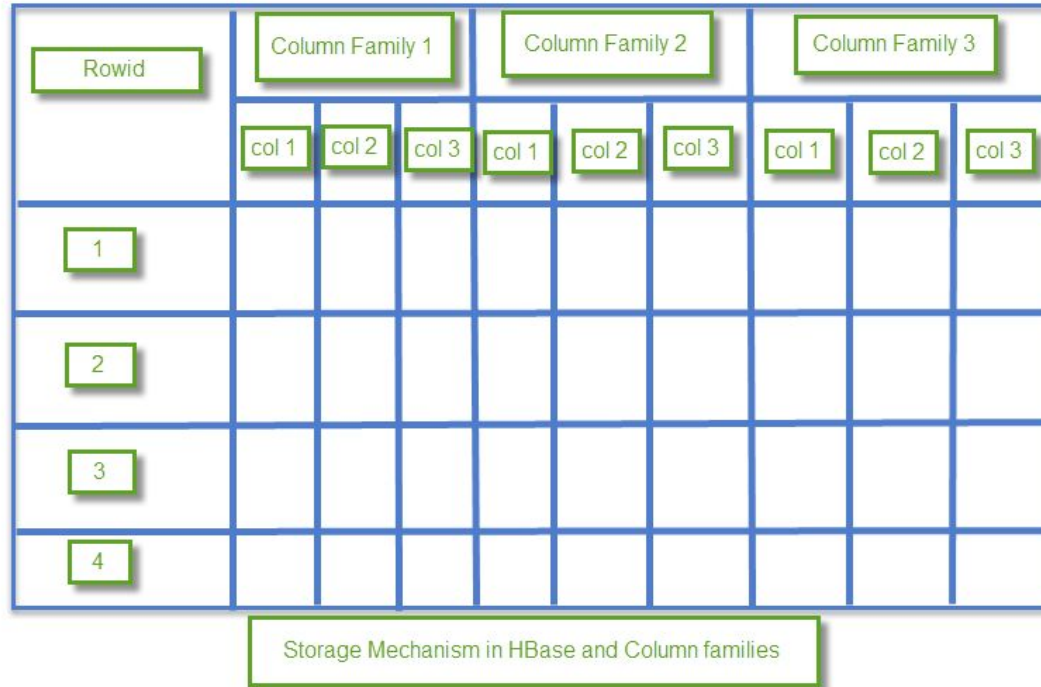
# Apache HBase

- HBase is a **CP columnar** database
- **AP** equivalent = [Apache Cassandra](#)

# HBase: Data structure

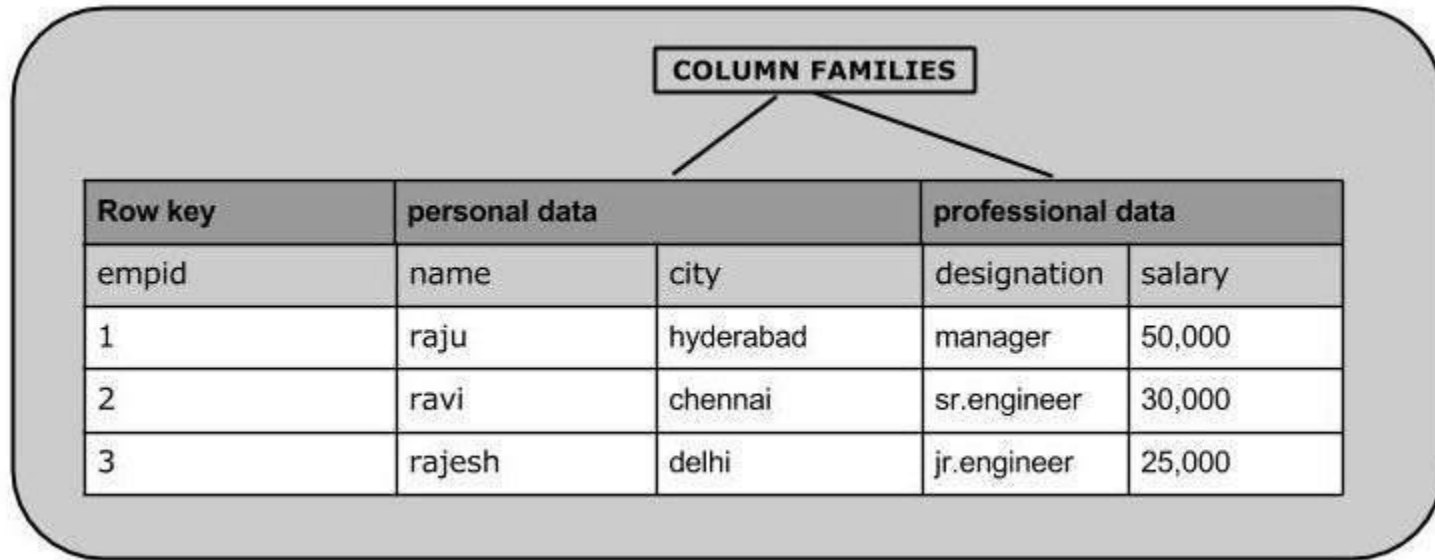
- A **table** is a collection of **rows**
- A **row** is a collection of **column families** defined at table creation
- A **column family** is a collection of **columns**
- A **column** is a **key-value pair**

# HBase: Data structure





# HBase: Data structure example



# HBase: Data storage

- Data is stored:
  - In HDFS in the **HFiles**
  - In RAM in the **Memstore**
- Tables are split into **regions**
  - 1 region = all rows in a certain **range of keys**
  - The number of regions depends on the size of the table

# HBase: Components

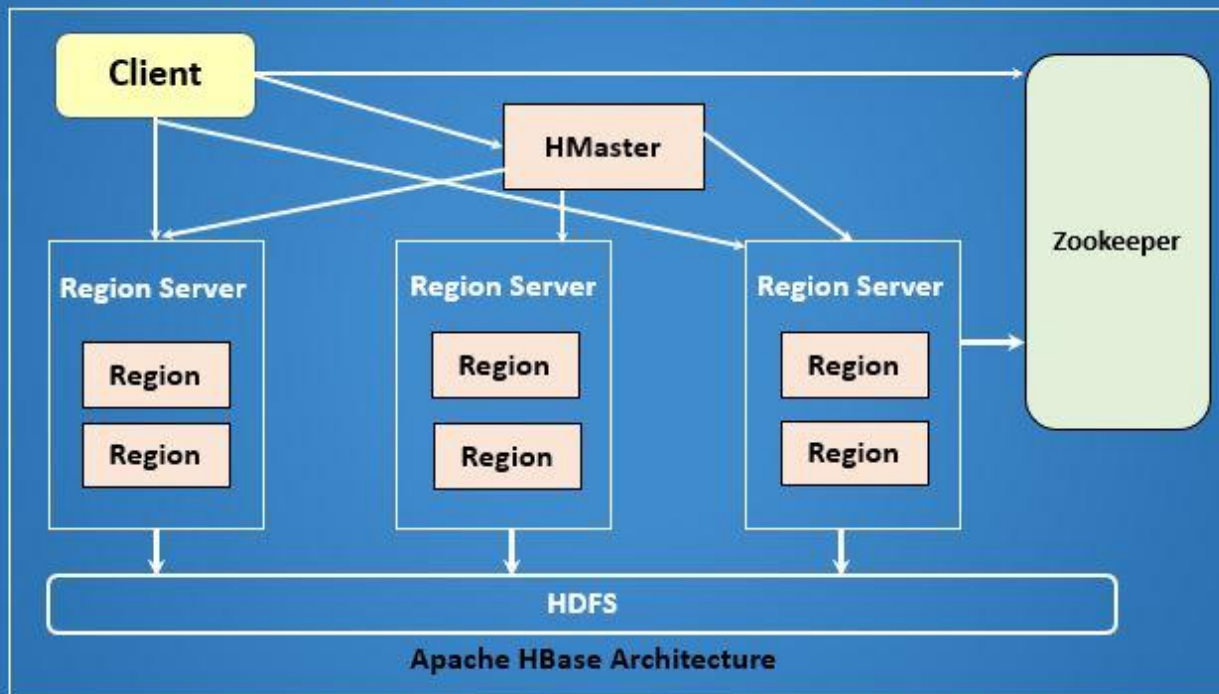
- **HBase Master**

- Handle table creation/deletion queries
- **Assigns regions** to RegionServers and monitor them

- **RegionServer**

- 1 region of a table is managed by 1 RegionServer

# HBase Architecture



[www.educba.com](http://www.educba.com)

# Querying HBase

- HBase client
  - CLI (**hbase shell**)
  - Languages (Java API) and tools (e.g. NiFi)
- Apache Phoenix, SQL on HBase (in memory)
- Apache Hive