H-Base

- **Distributed Column-Oriented** database on top of Hadoop/HDFS
- Provides low-latency access to single rows from billions of records
- Column oriented:
  - OLAP
  - Best for aggregation
  - High compression rate: Few distinct values
- Do not have a Schema or Data type
- Built for Wide tables: Millions of columns
  Billions of rows
- Denormalized data
- Master-Slave architecture
HBase System Overview

- Java Client APIs
- External APIs (Thrift, Avro, REST)
- Master
  - Region Server
    - Region
      - MemStore
      - HFile
    - Region
      - MemStore
      - HFile
    - Region
      - MemStore
      - HFile
  - Write-Ahead Log (WAL)
- Hadoop FileSystem API
- ZooKeeper
- Hadoop Distributed FileSystem (HDFS)
H-Base Architecture
HMaster Server

- Like Name Node in HDFS
- Manages and Monitors HBase Cluster Operations
- Assign Region to Region Servers
- Handling Load-Balancing and Splitting
Region Server

- Like Data Node in HDFS
- Highly Scalable
- Handle Read/Write Requests
- Direct Communication with Clients
Internal Architecture

- Tables → Regions
- Store
  - MemStore
  - FileStore → Blocks
- Column Families
Apache HBase Architecture

- HBase is composed of three main components in a master slave type of architecture.
- Region servers serve data for reads and writes.
- 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.
Contd…

- Regions are assigned to Region Servers.
- Tables are horizontally partitioned into key ranges (regions).
- HMaster coordinates region servers.
- ZooKeeper is a distributed coordination service.
HBase consists of:

- Set of tables
- Each table with column families and rows
- Row key acts as a Primary key in HBase.
- Any access to HBase tables uses this Primary Key
- Each column qualifier present in HBase denotes attribute corresponding to the object which resides in the cell.

Apache HBase storage structure
HBase HFile and Indexing

**HBASE HFILE**

Data is stored in an HFile which contains sorted key/values. When the MemStore accumulates enough data, the entire sorted KeyValue set is written to a new HFile in HDFS. This is a sequential write. It is very fast, as it avoids moving the disk drive head.

**HFILE INDEX**

The index, which we just discussed, is loaded when the HFile is opened and kept in memory. This allows lookups to be performed with a single disk seek.
Characteristics of HBase

- Fault tolerant
  - Replication across the data center
  - Atomic and strongly consistent row-level operations
  - High availability through automatic failover
  - Automatic sharding and load balancing of tables
Characteristics of HBase

- Fast
  - Near real time lookups
  - In-memory caching via block cache and bloom filters
  - Server side processing via filters and co-processors
Applications

- Adobe
- Airbnb uses HBase as part of its Aistream real-time stream computation framework
- Facebook uses HBase for its messaging platform.
- Flurry
- Imgur uses HBase to power its notifications system
- Netflix
- Rocket Fuel
- Spotify uses HBase as base for Hadoop and machine learning jobs.
- Sears
- Yahoo!
Apache ZooKeeper
ZooKeeper

- What is ZooKeeper?
  - Distributed coordination service for distributed applications
  - Like a Centralized Repository
- Challenges for Distributed Applications
  - Coordination
  - Race Condition
  - Dead-locks
  - Partial Failure
  - Inconsistency
- ZooKeeper Goals
  - Serialization
  - Atomicity
  - Reliability
  - Simple API
ZooKeeper Architecture

**Leader**
- Writes are quorum based, handled by leader and committed to followers
- New leader elected if existing fails

**Follower**
- Forwards writes to Leader
- Leader syncs state with followers
- Maintains quorum through voting

**Observer**
- Non-voting members
- Followers without the write observed of actual follower (voting)
Introduction to Zookeeper

- **Zookeeper**: A software service for a distributed environment that coordinates and configures different machines in a centralized way.
- A change is not considered successful until it has been written to a quorum
- A leader is elected within the ensemble for conflicts
- In HBase, ZooKeeper coordinates and shares state between the Masters and RegionServers.
- **Tagline**: Enables highly reliable distributed coordination
ZooKeeper Architecture

- Always Odd number of nodes.
- Leader is elected by voting.
- Leader and Follower can get connected to Clients and Perform Read Operations
- Write Operation is done only by the Leader.
- Observer nodes to address scaling problems
ZooKeeper Data Model

```
/  /app1
/   /app1/p_1
     /app1/p_2
        /app1/p_3

/ /app2
```

ZooKeeper Data Model

- **Z Nodes:**
  - Similar to Directory in File system
  - Container for data and other nodes
  - Stores Statistical information and User data up to 1MB
  - Used to store and share configuration information between applications
Z Node Types

- Persistent Nodes
- Ephemeral Nodes
- Sequential Nodes
- Watch: Event system for client notification
Projects & Tools on Hadoop

- HBase
- Hive
- Pig
- Jaql
- ZooKeeper
- AVRO
- UIMA
- Sqoop
References


[7] Jiong Xie, Shu Yin, Xiaojun Ruan, Zhiyang Ding, Yun Tian, James Majors, Adam Manzanares, Xiao Qin, "Improving MapReduce Performance through Data Placement in Heterogeneous Hadoop Clusters", 19th International Heterogeneity in Computing Workshop, Atlanta, Georgia, April 2010
References


[9] "Apache Hadoop",
http://en.wikipedia.org/wiki/Apache_Hadoop

[10] "Hadoop Overview",
http://www.revelytix.com/?q=content/hadoop-overview

References
