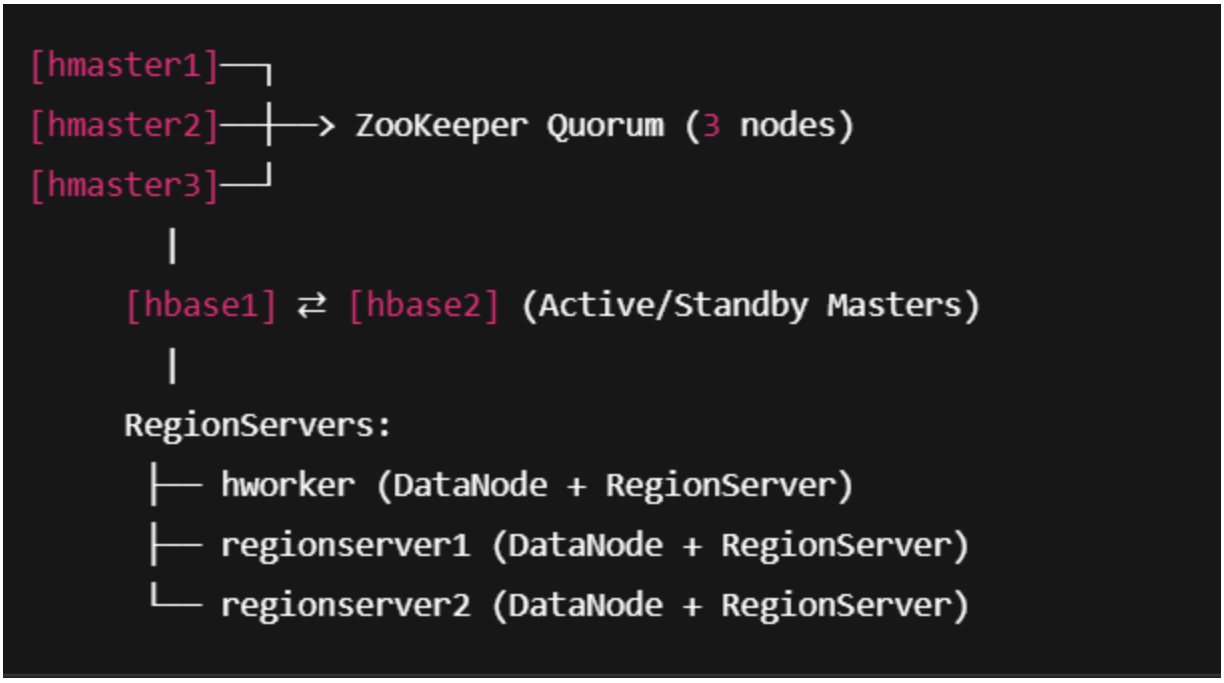

📘 HBase HA Cluster – Mini Documentation

✅ Overview

This project sets up a **high availability HBase cluster** with a highly available Hadoop backend using Docker. The cluster includes:

- Hadoop HA (NameNode + JournalNode + ZooKeeper)
- HBase Masters (active-standby mode)
- 3 RegionServers colocated with HDFS DataNodes
- HBase configured to use ZooKeeper for failover and coordination

---

📐 Architecture Summary

Services:

Node	Roles
`hmaster1-3`	ZooKeeper + JournalNode + NameNode + ResourceManager

Node	Roles
`hbase1`, `hbase2`	HBase Master (Active/Standby via ZooKeeper)
`hworker`	RegionServer + DataNode + NodeManager
`regionserver1/2`	RegionServer + DataNode + NodeManager

⚙️ Setup Steps

1. ✅ Base Docker Image

- Started with a functional **Hadoop Docker image**
- Added HBase 2.4.18 installation to the image (copied HBase binaries and configs)

2. ✅ Configuration Files

- Created and configured:
  - `hbase-site.xml`: defined `hbase.rootdir`, ZK quorum, backup masters, and regionserver handlers
  - `core-site.xml`, `hdfs-site.xml`, `yarn-site.xml` inherited from Hadoop image

> 📝 Configuration allows HBase to connect with the Hadoop NameNodes and ZooKeeper quorum correctly.

3. ✅ Entrypoint Script

- Edited `entrypoint.sh` to:
  - Start SSH
  - Format NameNode and start ZKFC when needed
  - Start HDFS services on `hmaster` nodes
  - Start `datanode + nodemanager + regionserver` on all worker and regionserver nodes
  - Start HBase Master/RegionServer roles based on the container hostname

4. ✅ Docker Compose

- Defined 9 containers:
  - 3 for `hmaster` (HA Hadoop/ZK)
  - 2 for HBase Masters

    - 1 for general worker node
    - 2 dedicated RegionServers (now running DataNode as well)
- Attached all containers to a custom Docker network (`mynetwork`)
- Mounted volumes for HDFS and ZooKeeper data persistence

🔄 Failover Support

✅ HBase Master Failover

- Configured ZooKeeper quorum for master election
- When the active master fails, a standby takes over
- Verified via `status 'detailed'` in HBase shell and `docker logs`

✅ RegionServer Failover

- 3 RegionServers run on independent nodes
- If one fails, HBase master detects the failure and reassigns regions
- Verified using test table + `get` after RegionServer stop
