

INTRODUCTION TO BIG DATA

ECAP456

Dr. Rajni Bhalla
Associate Professor

Learning Outcomes



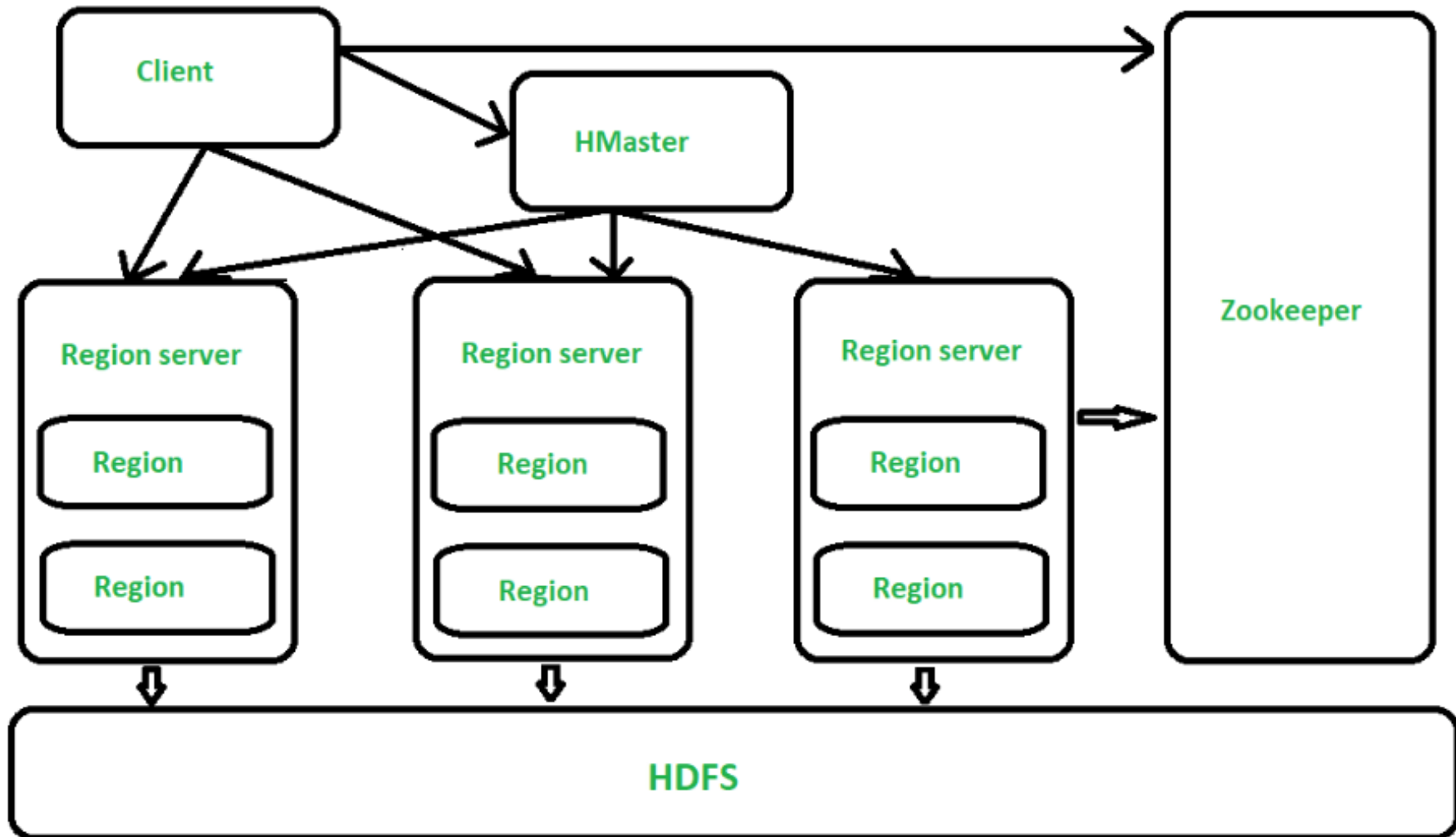
After this lecture, you will be able to

- learn fundamentals of Hbase,
- explore concepts of ZooKeeper.

Introduction



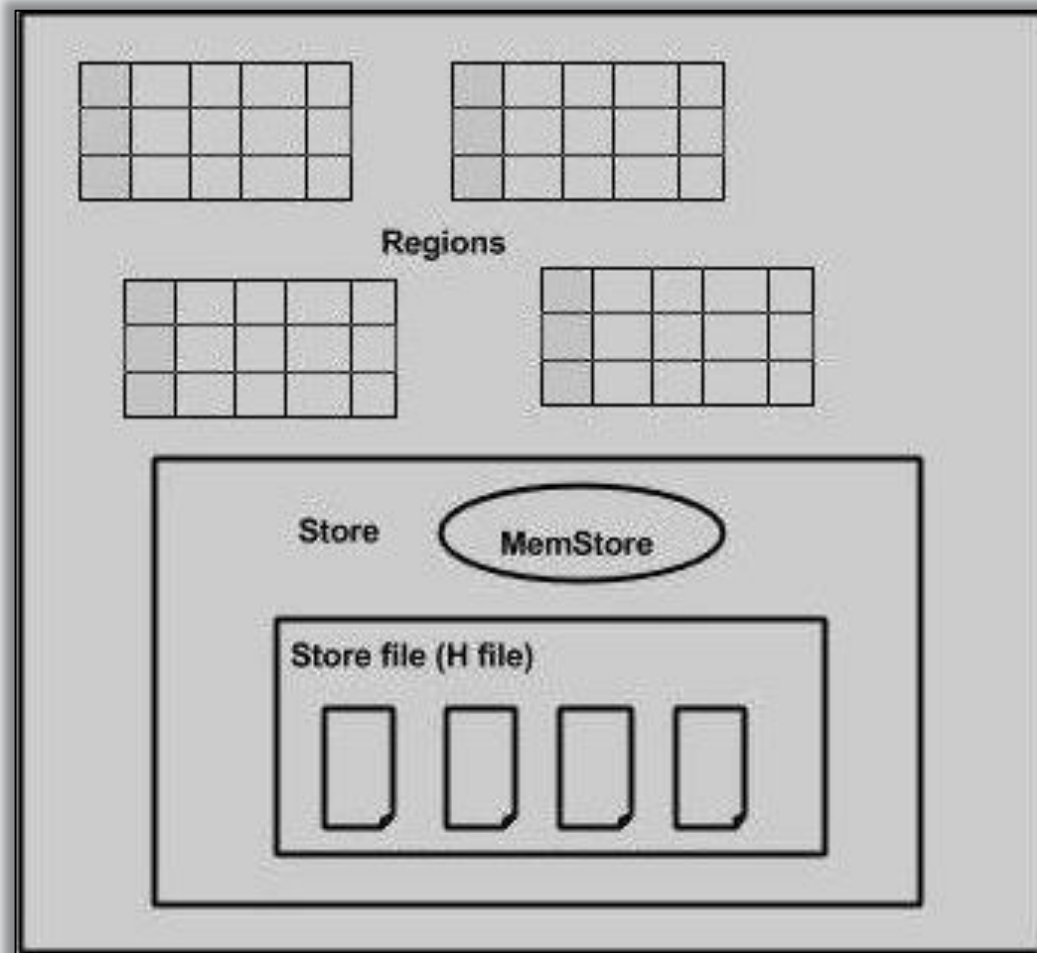
HBase - Architecture



Master Server

- Assigns regions to the region servers.
- Handles load balancing of the regions across region servers.
- Maintains the state of the cluster
- Is responsible for schema changes and other metadata operations

Regions



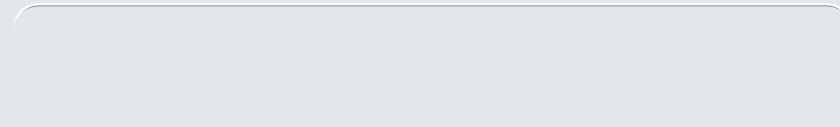
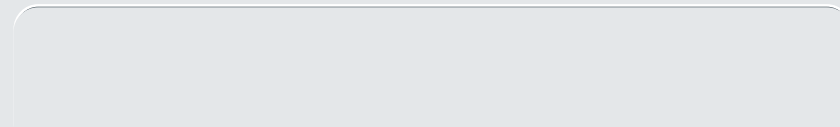
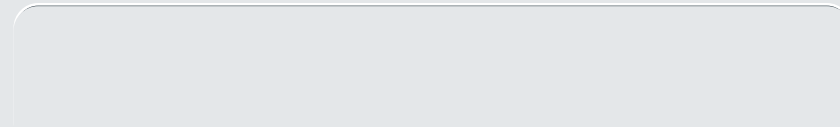
HBase Data Model

HBase Tables

HBase Data Model

HBase Tables

HBase Row



HBase Data Model

HBase Tables

HBase Row

RowKey

HBase Data Model

HBase Tables

HBase Row

RowKey

Columns

HBase Data Model

HBase Tables

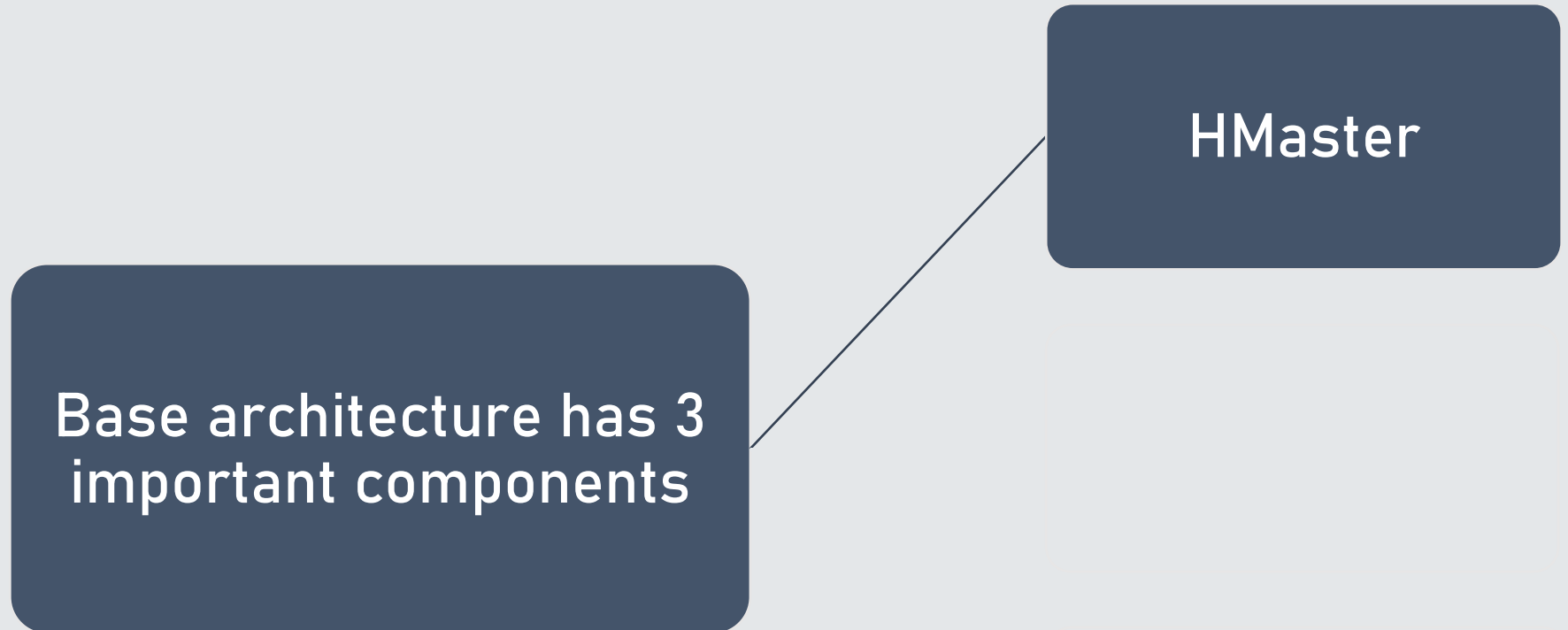
HBase Row

RowKey

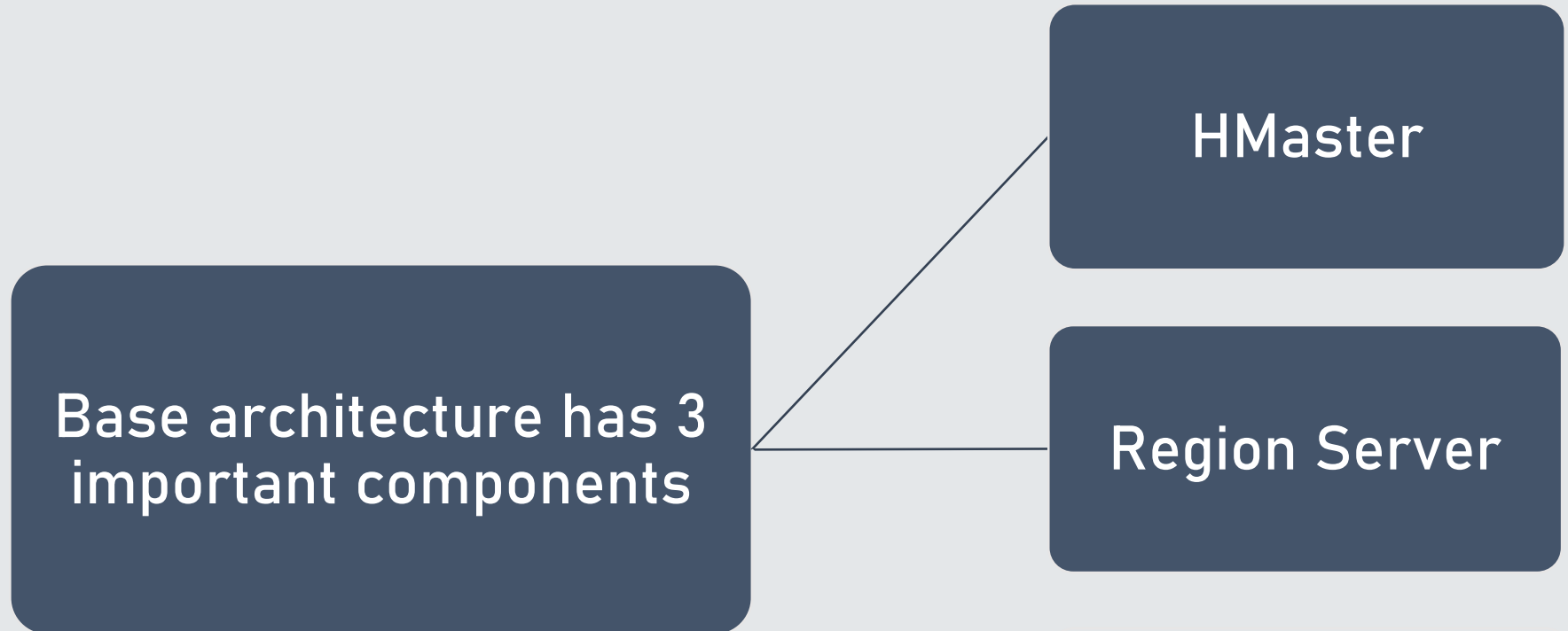
Columns

Column Family

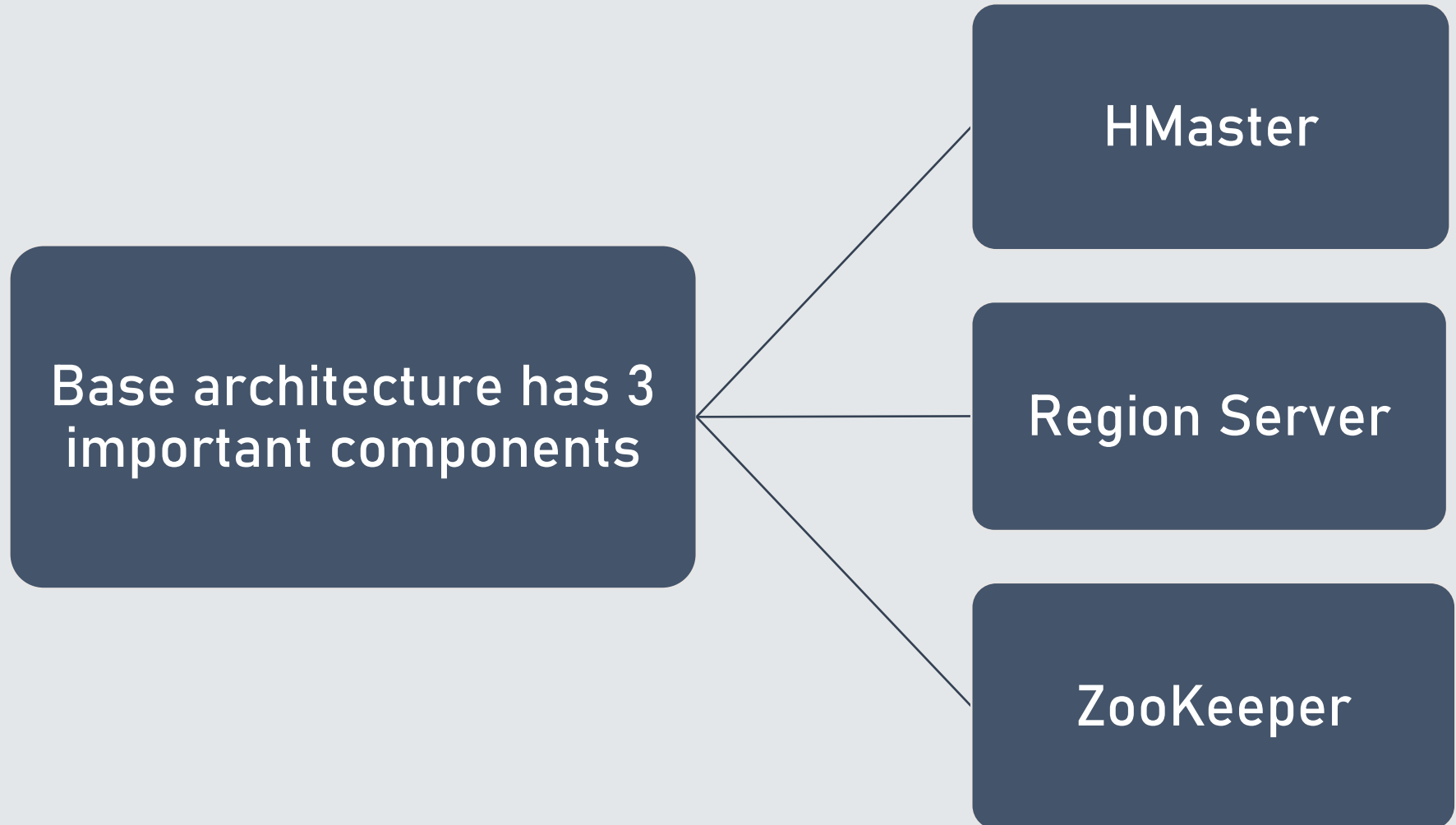
Components of Apache HBase Architecture



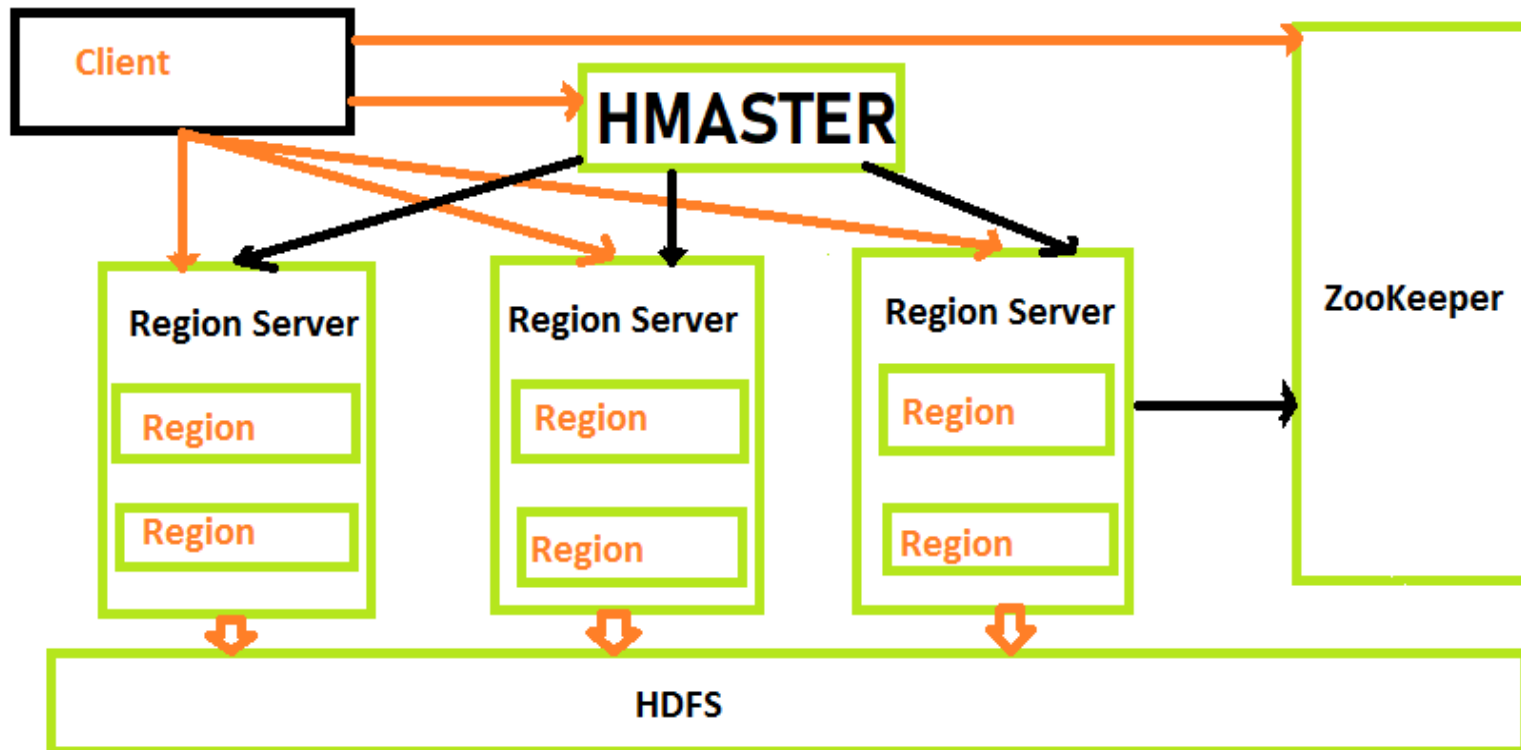
Components of Apache HBase Architecture



Components of Apache HBase Architecture



HMaster



Apache Hbase Architecture

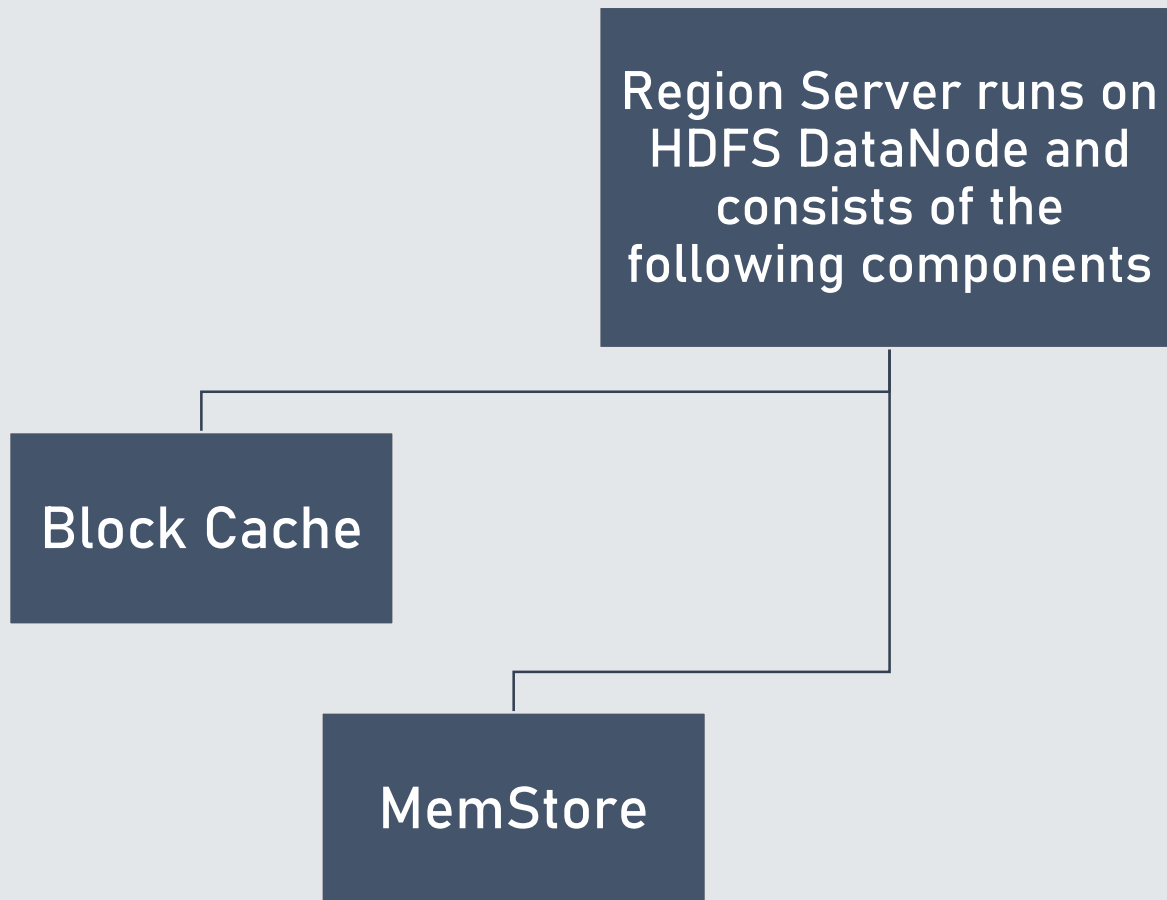
Region Server

Region Server runs on
HDFS DataNode and
consists of the
following components

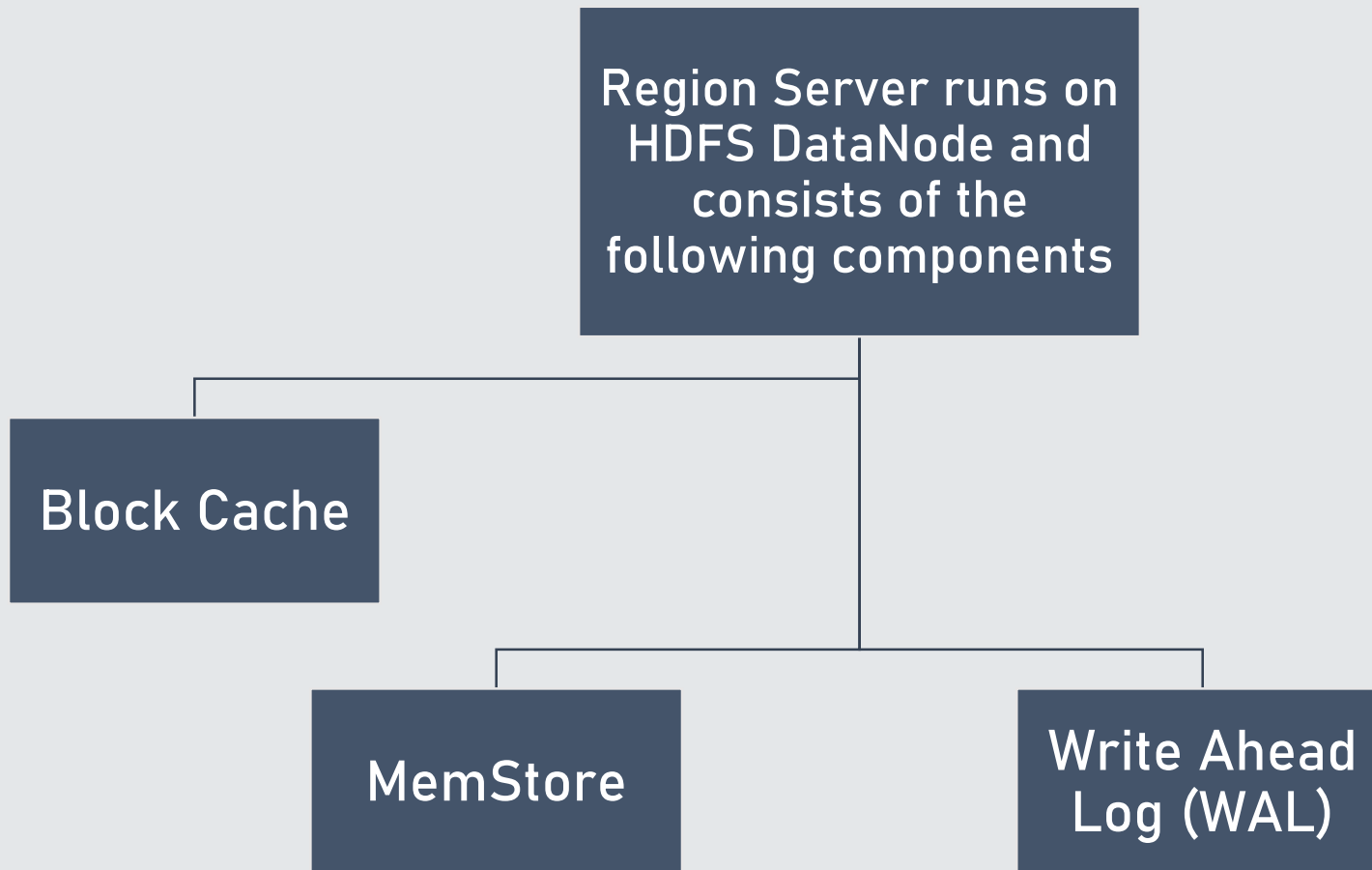
```
graph TD; A[Region Server runs on HDFS DataNode and consists of the following components] --- B[Block Cache]
```

Block Cache

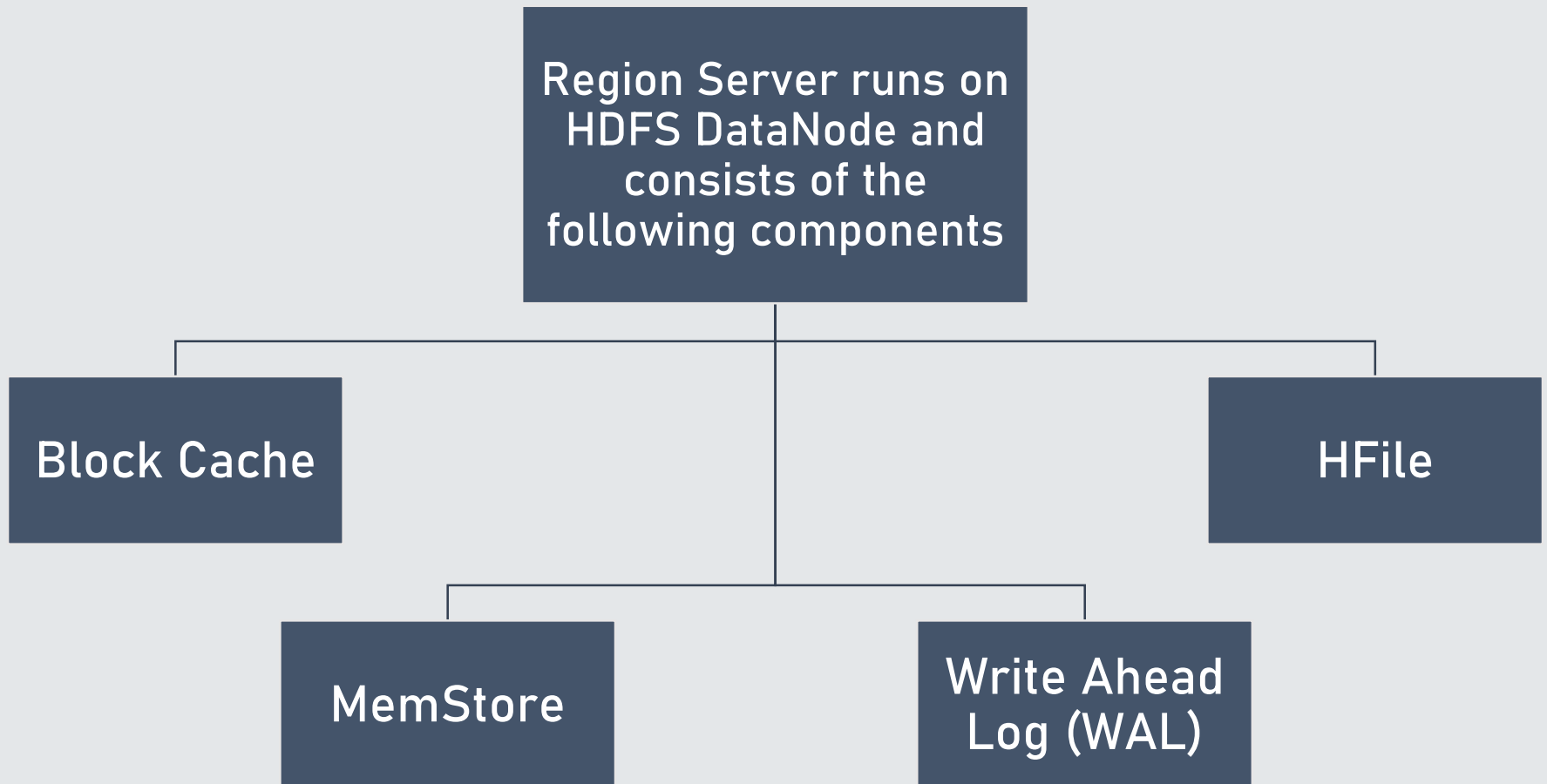
Region Server



Region Server



Region Server



Advantages of HBase

Can store large data sets

Database can be shared

Cost-effective from gigabytes to petabytes

High availability through failover and replication

Disadvantages of HBase –

No support SQL structure

No transaction support

Sorted only on key

Memory issues on the cluster

Comparison between HBase and HDFS

Comparison between HBase and HDFS:

latency access

Comparison between HBase and HDFS:

latency access

Random reads
and writes

Comparison between HBase and HDFS:

latency access

Random reads
and writes

shell
commands

Introduction



APACHE

ZooKeeperTM

Introduction

- Zookeeper is an open-source project
- Zookeeper has ephemeral nodes representing different region servers
- the nodes are also used to track server failures or network partitions.
- Clients communicate with region servers.
- HBase itself will take care of zookeeper.



That's all for now...