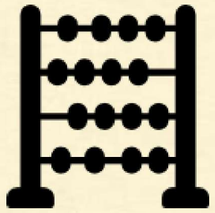
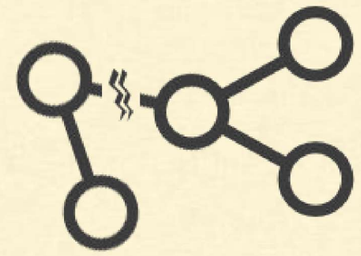
Welcome to HBase

HBase - Quick Definition



Consistent

Partition Tolerant

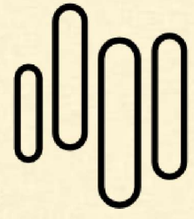


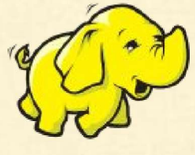
	Average	
	Height	Weight
	1.9	0.003

Column Family Oriented

Good for hundreds of millions or billions of rows

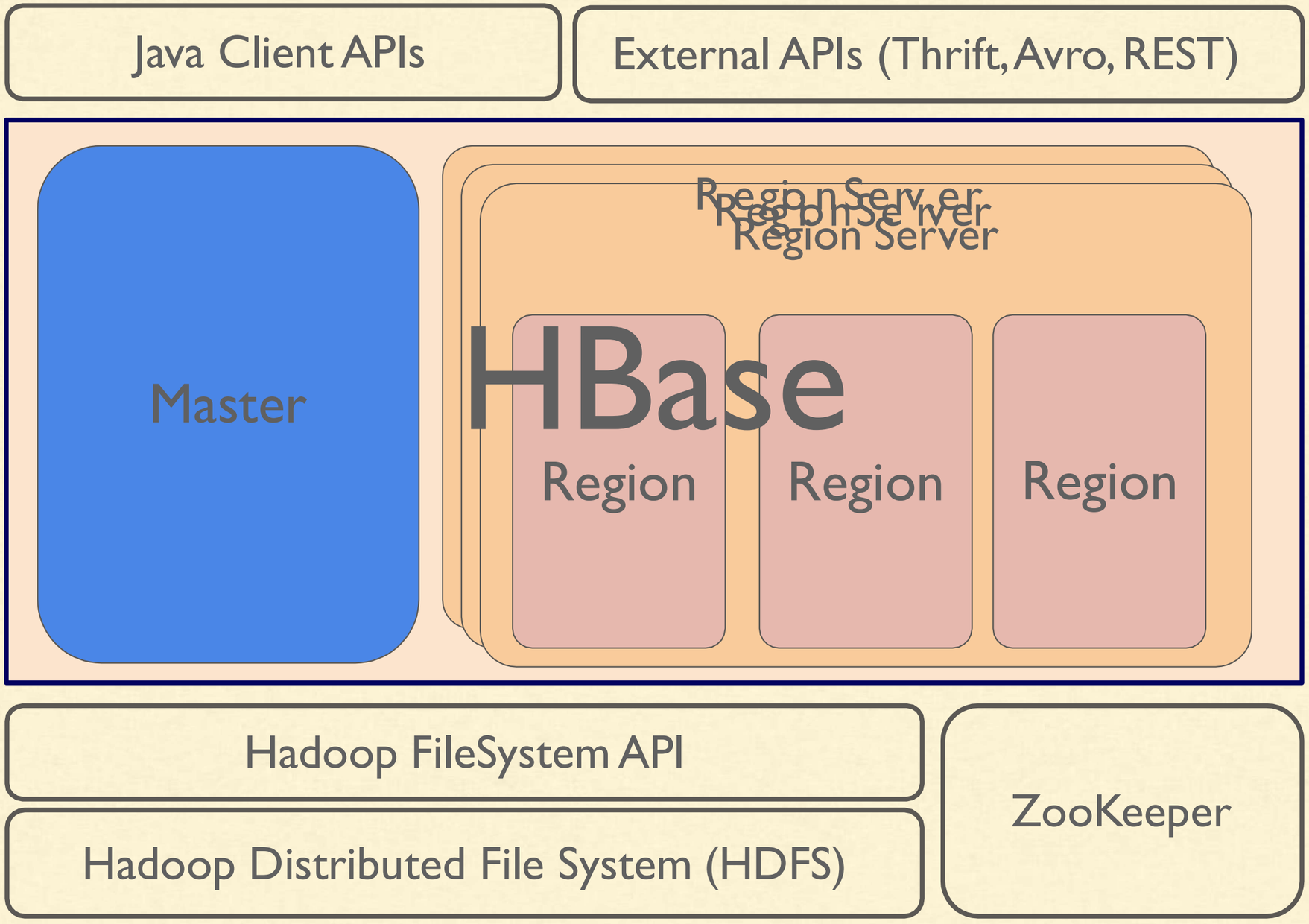
Based on Google's Big Table



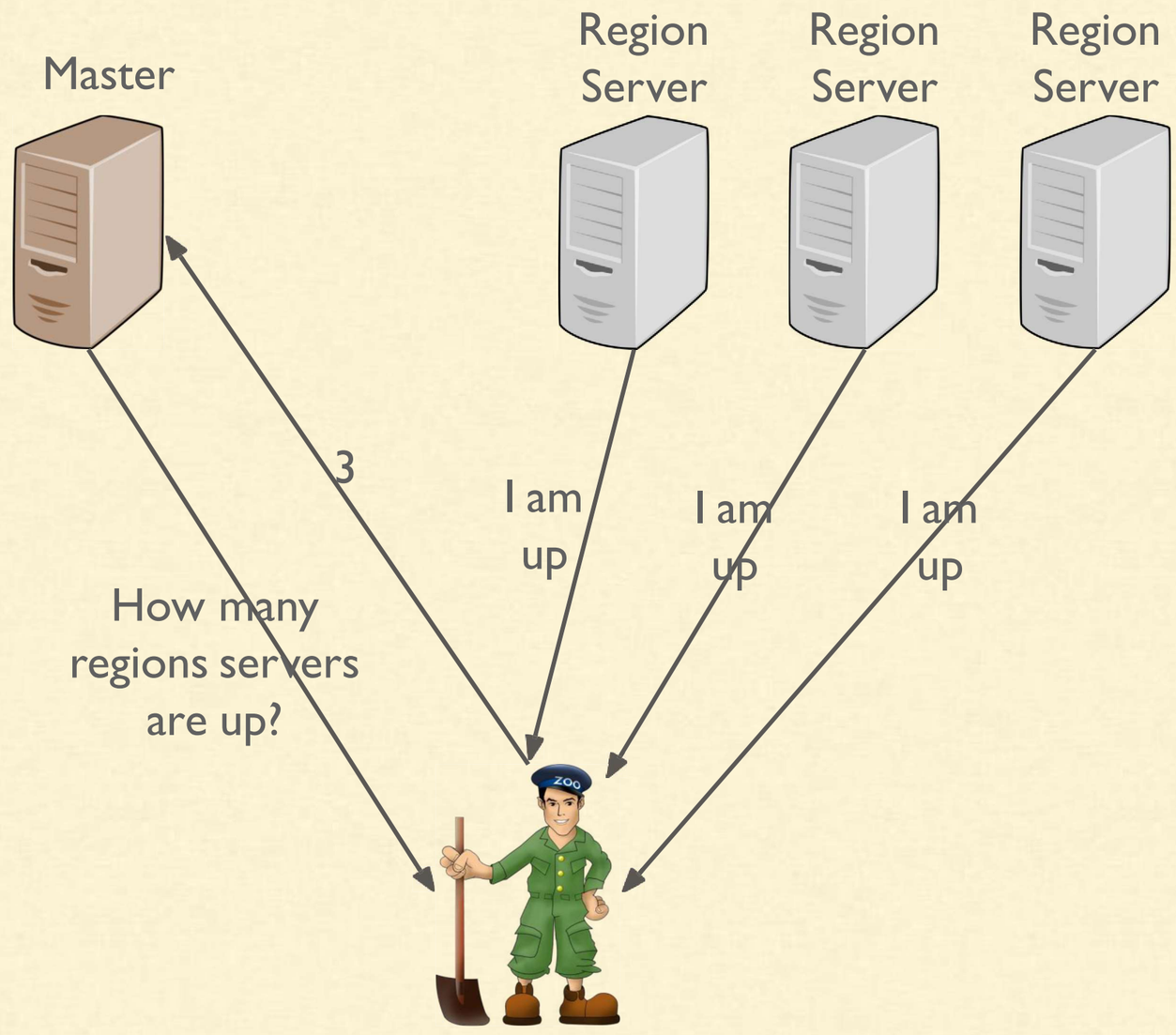


Runs on Hadoop / HDFS

HBase - Architecture - Overview



HBase - Architecture



HBase is

- An analytics engine capable of performing processing structured data
- A data store good for storing data in structured format having millions of rows
- A file system like hdfs that can store data in append only format
- A client library to interact with Hive

What is not true about HBase:

- ○ HBase provides immediate consistency
- ○ It is partition tolerant
- ○ It is column family oriented
- ○ HBase provides eventual consistency
- ○ You can store data having billions of row and millions of columns
- ○ HBase runs on top of Hadoop

What is true about Hbase

- HBase has single master and has only single region server
- HBase can have multiple masters and has single region server
- HBase has single master and can have multiple region servers

How does HBase master keep track of region servers?

- While installing HBase we configure region servers in in config file
- HBase discovers the region servers by using ZooKeeper

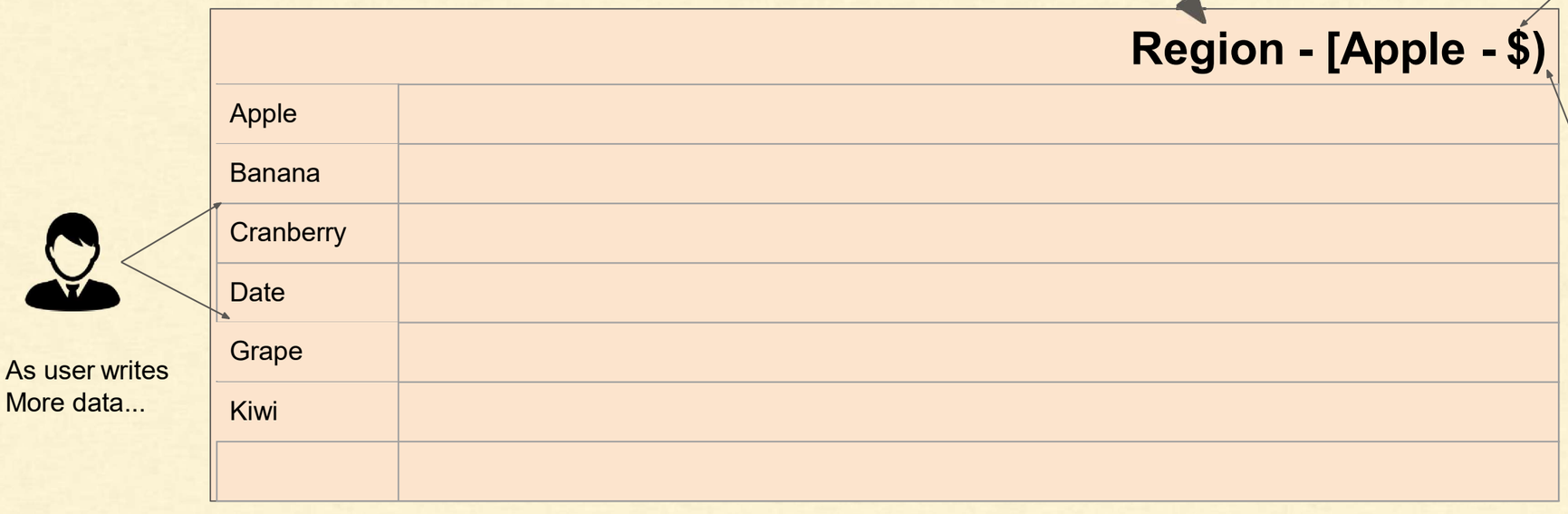
HBase - Architecture

HBase - Architecture

HBase - Architecture

- Tables are automatically partitioned horizontally into regions.
- Each region comprises a subset of a table's rows.
- Initially, a table comprises a single region,

The $ - dollar denotes the end of the row keys. Every key will smaller than $.



- A region is denoted by
  - Table it belongs to,
  - Its first row, inclusive
  - Last row, exclusive

HBase - Architecture

As user writes
More data...

Region - [Apple - $)

Apple
Banana
Bilberry
Cranberry
Cucumber
Date
Grape
Kiwi

HBase - Architecture

Region 1 - [Apple - Cucumber)

Apple	
Banana	
Bilberry	
Cranberry	

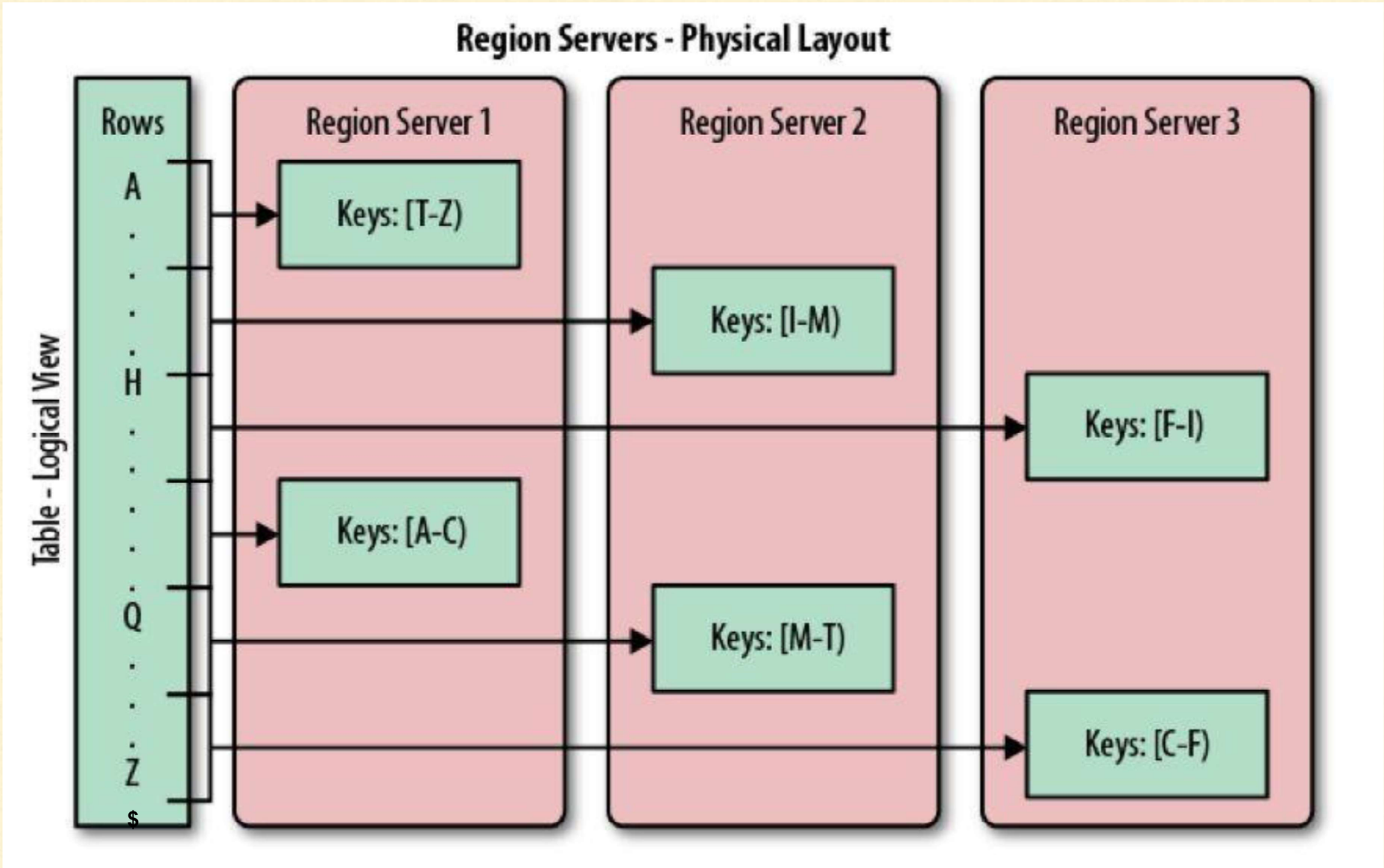
Region 2 - [Cucumber - $)

Cucumber	
Date	
Grape	
Kiwi	

As user writes
More data...

- As size of the region grows beyond threshold, it splits into 2 halves
- As the table grows, the number of its regions grows.

HBase - Region Servers



Every row in a Hbase table must have a row key?

- Yes
- No

Data in table is ordered by row key?

- ○ Yes
- ○ No, it depends on the sorting order

A subset of rows is called a region?

- ○ True
- ○ False

If we have stored four records in HBase with the keys as Apple, Banana, Cherry, Apricot then,

- Apple and Cherry will be adjacent records
- Apple and Apricot will be next to each other

As the data grows, the number of regions increase

- ○ True
- ○ False

Every region server can contain only one region?

- ○ True
- ○ False

HBase - Data Model

Data Modeling?

Process of structuring your data using the constructs provided by a datastore to solve your problem.

Requirements

Technology

- Database features
- Limitations