Web: Inceptez.com Mail: info@inceptez.com Call: 7871299810, 7871299817

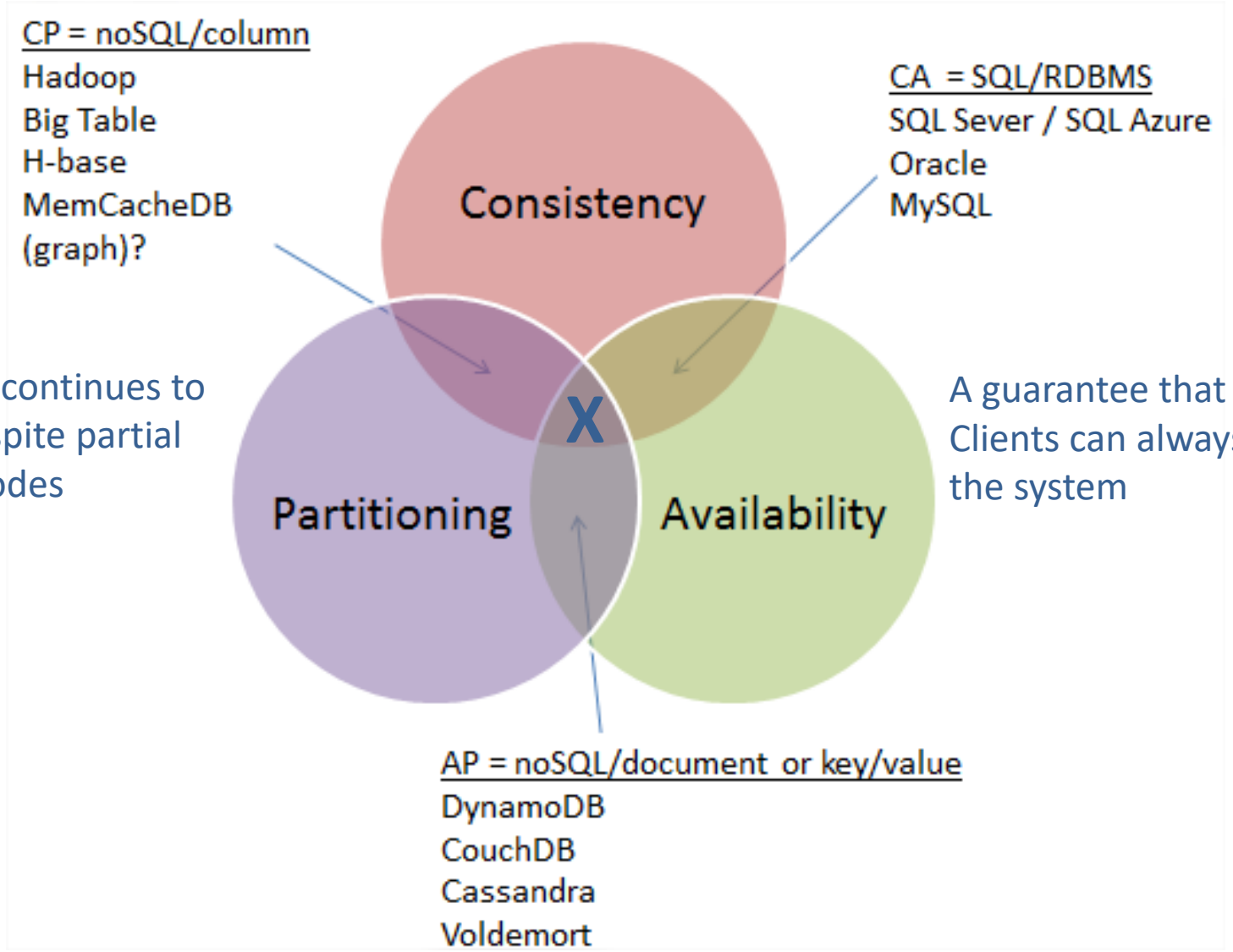
APACHE - HBASE

NOSQL

- Intro to NOSQL
- Types of NOSQL
- Characteristics of NoSQL
  - Dynamic Schemas
  - Auto Sharding
  - Replication
  - Integrated Caching
  - No joins, low cost.

CAP Theorem

All nodes see the same data at the same time

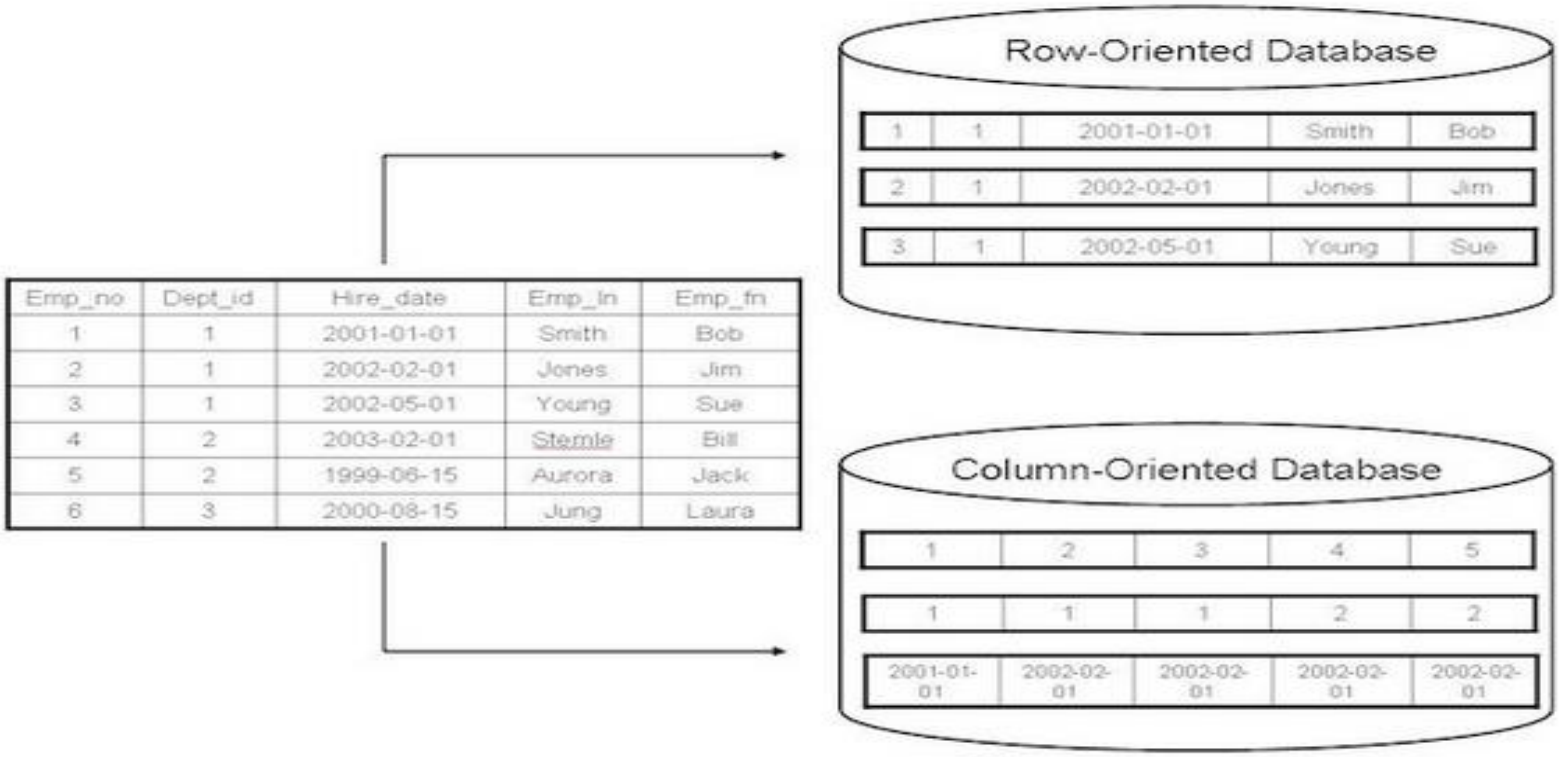


The system continues to operate despite partial failure of nodes

A guarantee that each Clients can always access the system

HBASE

- What is Hbase
- Brief History
- RDBMS vs HBASE



Storage Hierarchy

- **Column family** - Collection of columns.
- **Table** - Collection of Column Families.
- **Row** - Spread across column families.
- **Row Key** - Rows identified by unique ID.
- **Column** - Collection of key value pairs.
- **Cell** – Each value of the column.
- **Timestamp** – Versions of a cell.

Characteristics

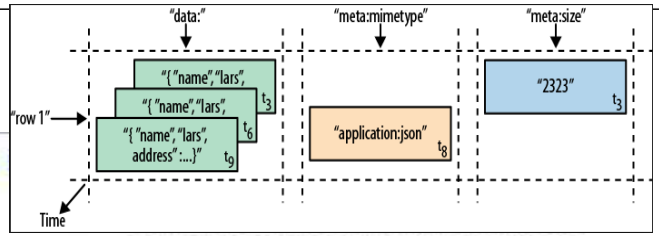
- **Isolation – Row level**
- **Lexographic Sorting** (1,10,11,2,20,21)

Row Key	Customer		Sales	
Customer Id	Name	City	Product	Amount
101	John White	Los Angeles, CA	Chairs	$400.00
102	Jane Brown	Atlanta, GA	Lamps	$200.00
103	Bill Green	Pittsburgh, PA	Desk	$500.00
104	Jack Black	St. Louis, MO	Bed	$1600.00

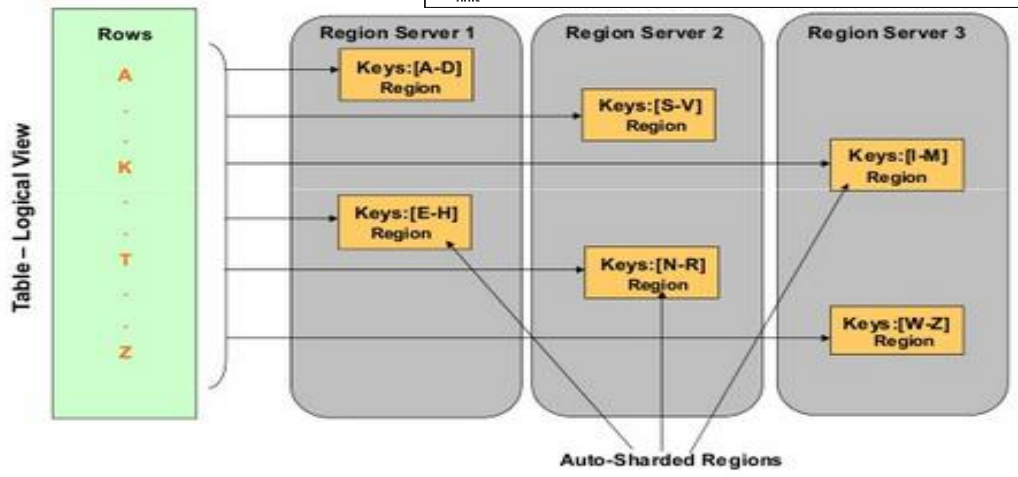
Row Oriented

Column Oriented

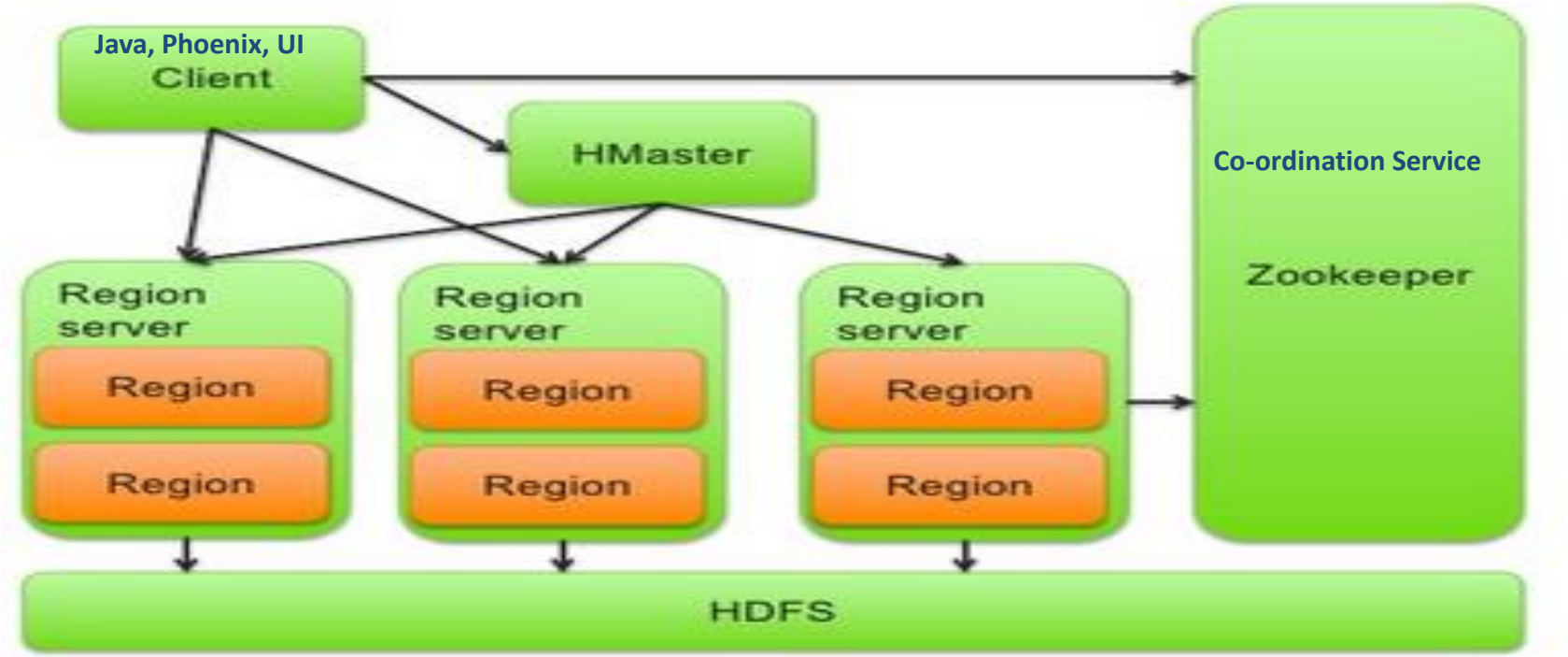
Column Families



How are Tables Stored

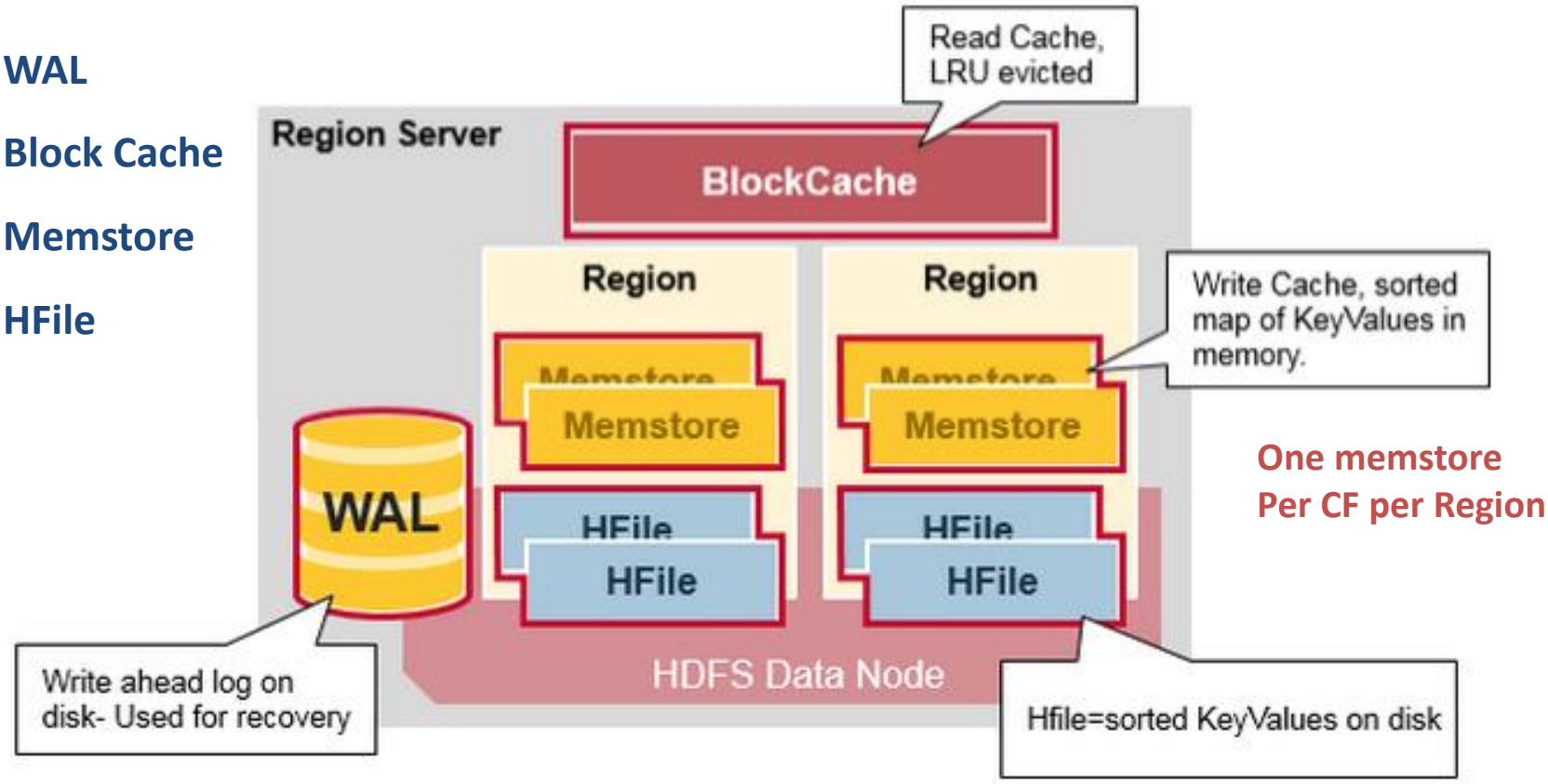


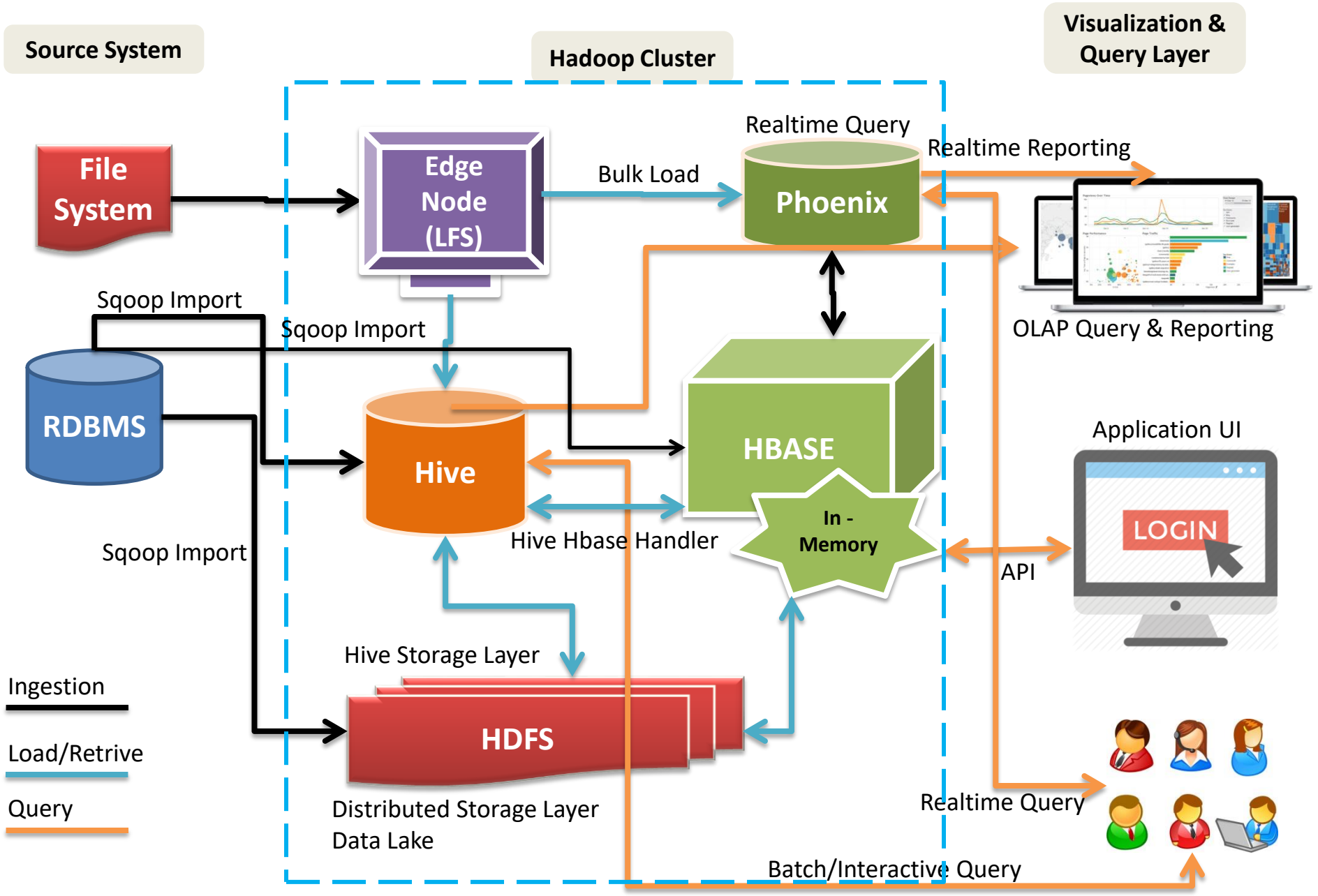
Architecture Overview



Inside RegionServer

- WAL
- Block Cache
- Memstore
- HFile




Source System
Hadoop Cluster
Visualization & Query Layer
File System
Edge Node (LFS)
Bulk Load
Phoenix
Realtime Query
Realtime Reporting
OLAP Query & Reporting
Sqoop Import
Sqoop Import
RDBMS
Hive
HBASE
In - Memory
Hive Hbase Handler
Application UI
LOGIN
API
Sqoop Import
Hive Storage Layer
HDFS
Distributed Storage Layer
Data Lake
Realtime Query
Batch/Interactive Query
Ingestion
Load/Retrive
Query

- **Installation**
- **Workouts**

Web: Inceptez.com Mail: info@inceptez.com Call: 7871299810, 7871299817

Apache Phoenix

What is Apache Phoenix?

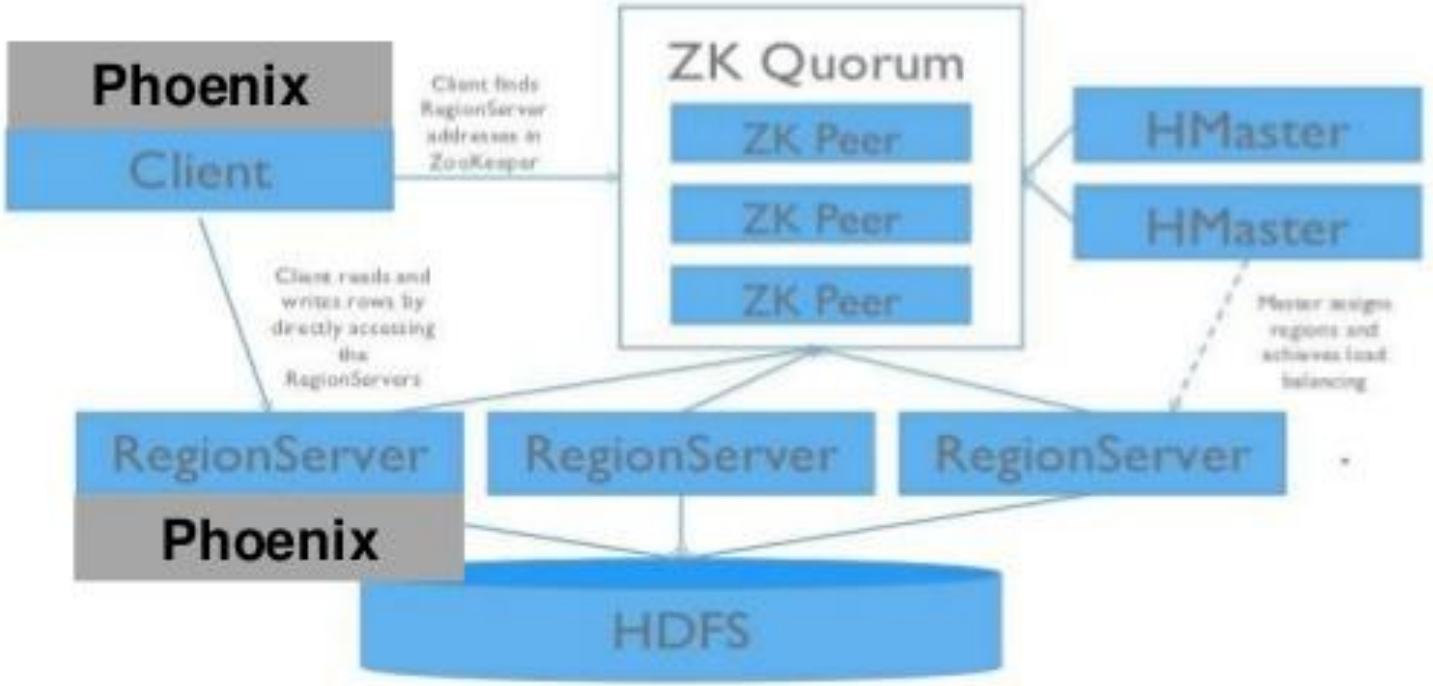
1. Turns HBase into a SQL database
   - Query Engine
   - MetaData Repository
   - Embedded JDBC driver

SELECT * FROM t WHERE k IN (?,?,?)

Phoenix	Stinger (Hive 0.11)
0.04 sec	280 sec

* 110M row table

HBase Cluster Architecture



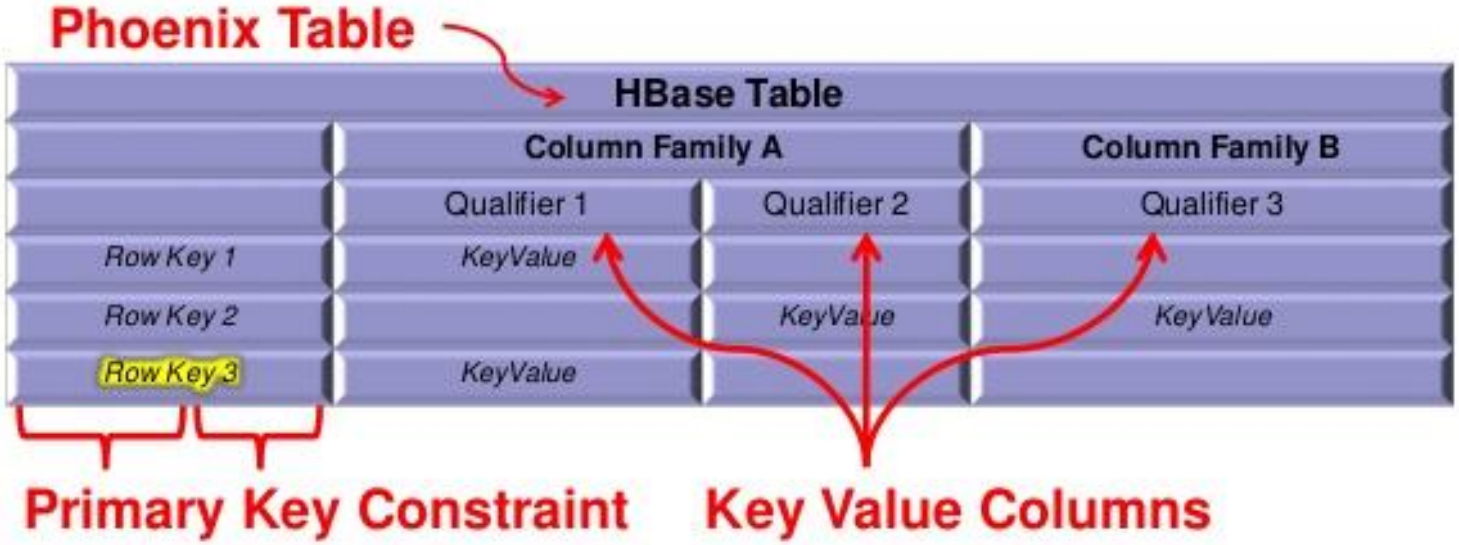
Example

DDL command looks like this:

```
CREATE TABLE SERVER_METRICS (
        HOST                VARCHAR,
        DATE                DATE,
        RESPONSE_TIME       INTEGER,
        GC_TIME             INTEGER,
        CPU_TIME            INTEGER,
        IO_TIME             INTEGER,
        CONSTRAINT pk PRIMARY KEY (HOST, DATE))
```

Phoenix Data Model

Phoenix maps HBase data model to the relational world



Example

With data that looks like this:

SERVER METRICS			
HOST + DATE		RESPONSE_TIME	GC_TIME
SF1	1396743589	1234	
SF1	1396743589		8012
…			
SF3	1396002345	2345	
SF3	1396002345		2340
SF7	1396552341	5002	1234
…			

Row Key