Introduction of HBase

Original Slides By: Hu Yi

2009-3-11

Overview

- HBase is an Apache open source project whose goal is to provide storage for the Hadoop Distributed Computing Environment.

- Data is logically organized into tables, rows and columns.

Outline

- Data Model
- Architecture and Implementation

Conceptual Vi...

- A data row has a sortable row key and an arbitrary number of columns.
- A Time Stamp is designated automatically if not artificially.
- <family>:<label>

Row key	Time Stamp	Column "contents:"	Column "anchor:"	
"com.apache.www"	t12	"<html>…"		
	t11	"<html>…"		
	t10		"anchor:apache.com"	"APACHE"
"com.cnn.www"	t15		"anchor:cnnsi.com"	"CNN"
	t13		"anchor:my.look.ca"	"CNN.com"
	t6	"<html>…"		
	t5	"<html>…"		
	t3	"<html>…"		

Row Ranges:

- Row key/ Colum
  Timestamp desc
- Physically, table
  into row ranges
  from start-key to

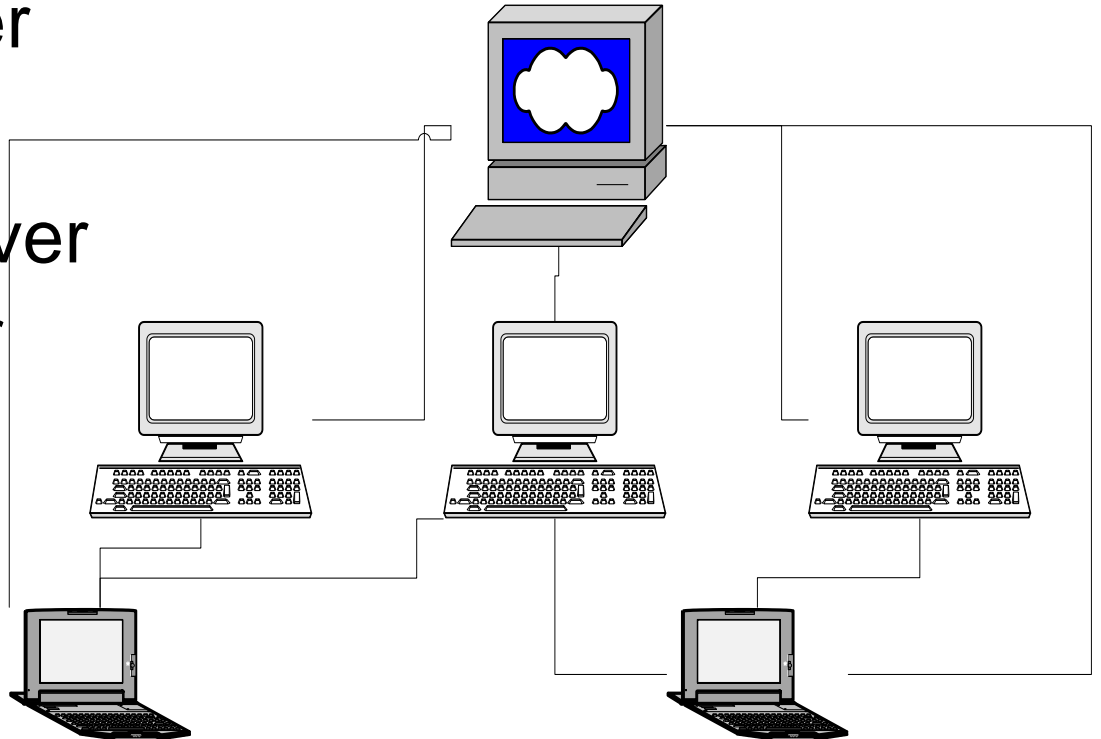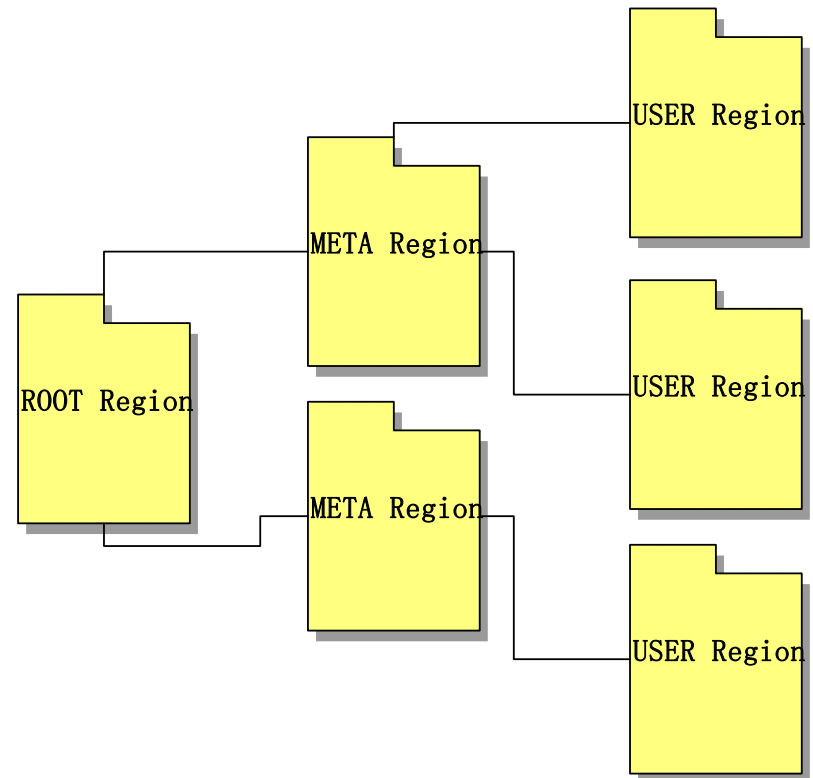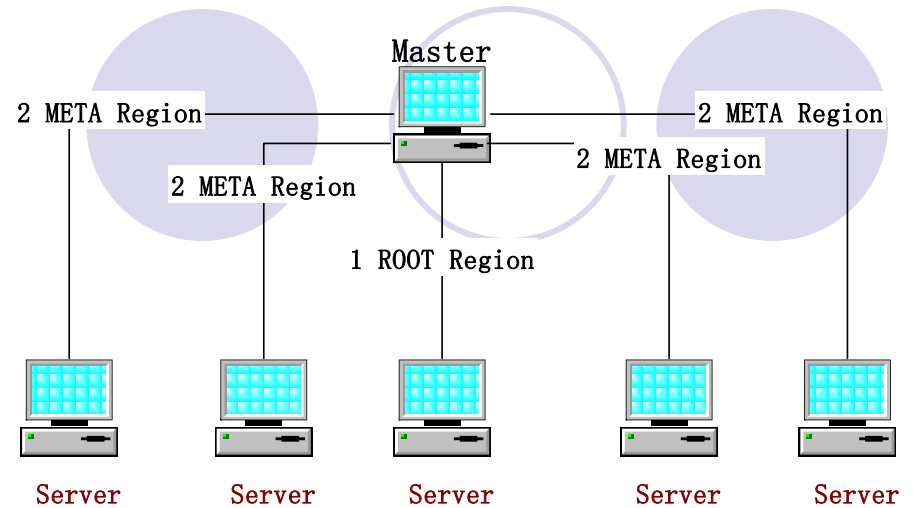| Row key	Time Stamp	Column "contents:"	Column "anchor:"	
aaaa	t15		anchor:cc	value
	t13	ba		
	t12	bb		
	t11		anchor:cd	value
	t10	bc		
aaab	t14			
aaac			anchor:be	value
aaad			anchor:ad	value
aaae	t5	ae		
	t3	af		

Outline

- Data Model
- Architecture and Implementation

Three major components

- ● The HBaseMaster

- ● The HRegionServer
  - ○ Aka Tablet Server

- ● The HBase client

HBaseMaster

- Assign regions to HRegionServers.

1. ROOT region locates all the META regions.

2. META region maps a number of user regions (aka tablet regions).

3. Assign user regions to the HRegionServers.

- Perform all read/write access
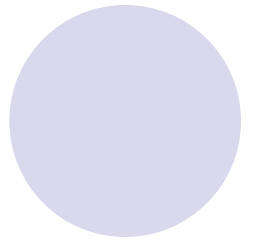
- Monitor the health of each Server

Master

2 META Region ——————————— 2 META Region

2 META Region

2 META Region

1 ROOT Region

Server   Server   Server   Server   Server

USER Region

META Region
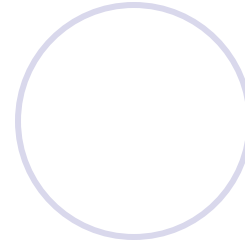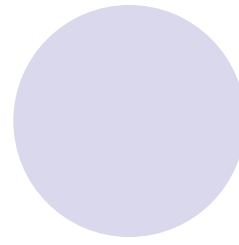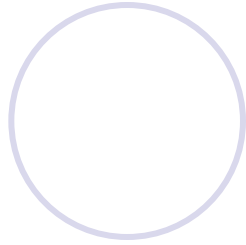
ROOT Region

META Region

USER Region

USER Region

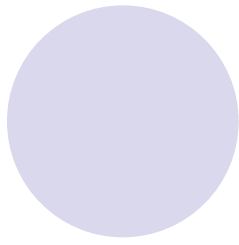# Summary

- Column-oriented modification more flexible.

- Higher performance on row key clusters.

Thank you