

CS15-319 / 15-619

Cloud Computing

Recitation 11

April 1st and April 3rd, 2014

Announcements

- Encounter a general bug:
 - Post on Piazza
- Encounter a grading bug:
 - Post Privately on Piazza
- Don't ask if my answer is correct
- Don't post code on Piazza
- Search before posting
- Post feedback on OLI

Piazza Questions

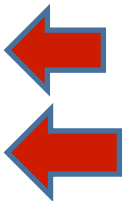
- Website problem
 - Open default port 80
 - Permission to access DynamoDB service

Project 3, Module 4 Reflections

- DynamoDB vs. MySQL (Checkpoint question)
 - Queries on any column other than key will require expensive scan operations in DynamoDB.
 - DynamoDB can return only 1MB max for a query, which is a limitation for certain queries that could potentially return many rows.
- Learn to use DynamoDB's API
- Learn how DynamoDB can be configured for a web server
- Be cost conscious about using DynamoDB

Module to Read

- UNIT 4: Cloud Storage
 - Module 12: Cloud Storage
 - Module 13: Case Studies: Distributed File Systems
 - Module 14: Case Studies: NoSQL Databases
 - Module 15: Case Studies: Cloud Object Storage
 - Quiz 4: Cloud Storage
- UNIT 5: Distributed Programming and Analytics Engines for the Cloud
 - Module 16: Introduction to Distributed Programming for the Cloud



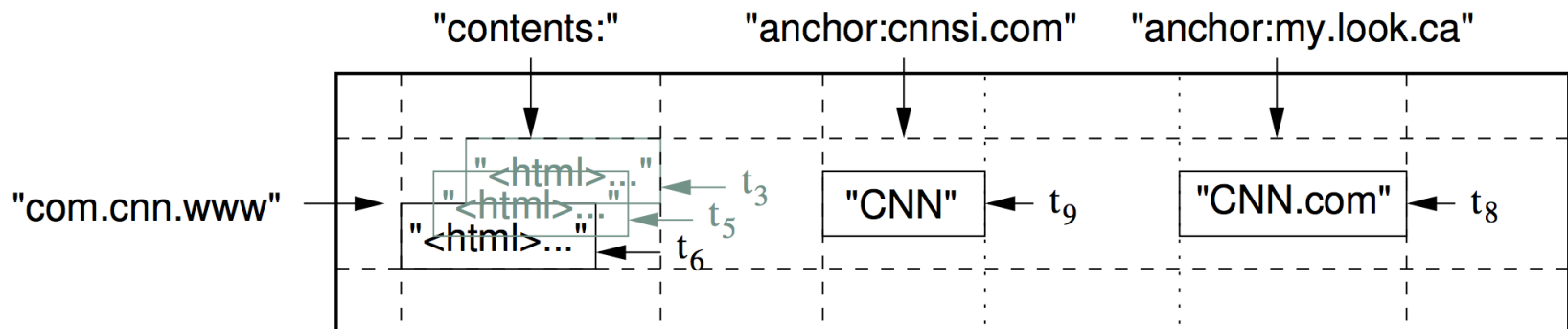
Project 3

- Files vs. Databases
 - File vs. Database
- Vertical Scaling in Databases
 - Vertical Scaling
- Horizontal Scaling in Databases
 - Horizontal Scaling
- Working with NoSQL: DynamoDB / Hbase
 - Amazon DynamoDB
 - DynamoDB vs. HBase



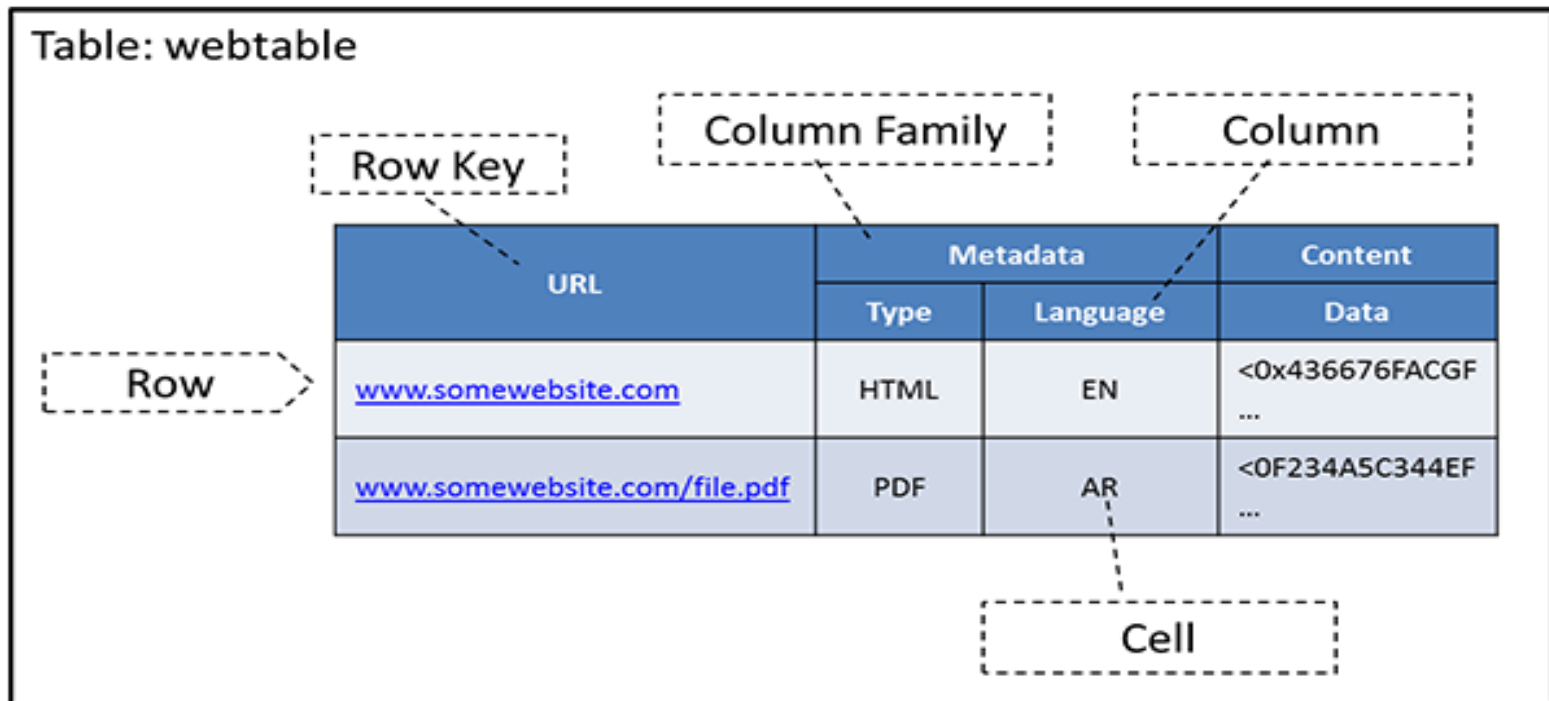
HBase

- HBase
 - NoSQL Database, based on Google's BigTable
 - [Unit 4 Module 14](#)
- Data Model
 - stored as key-value pairs
 - Rows are ordered and accessed by row key
 - Columns are grouped into sets called **column families**, which form the basic unit for access control



HBase example

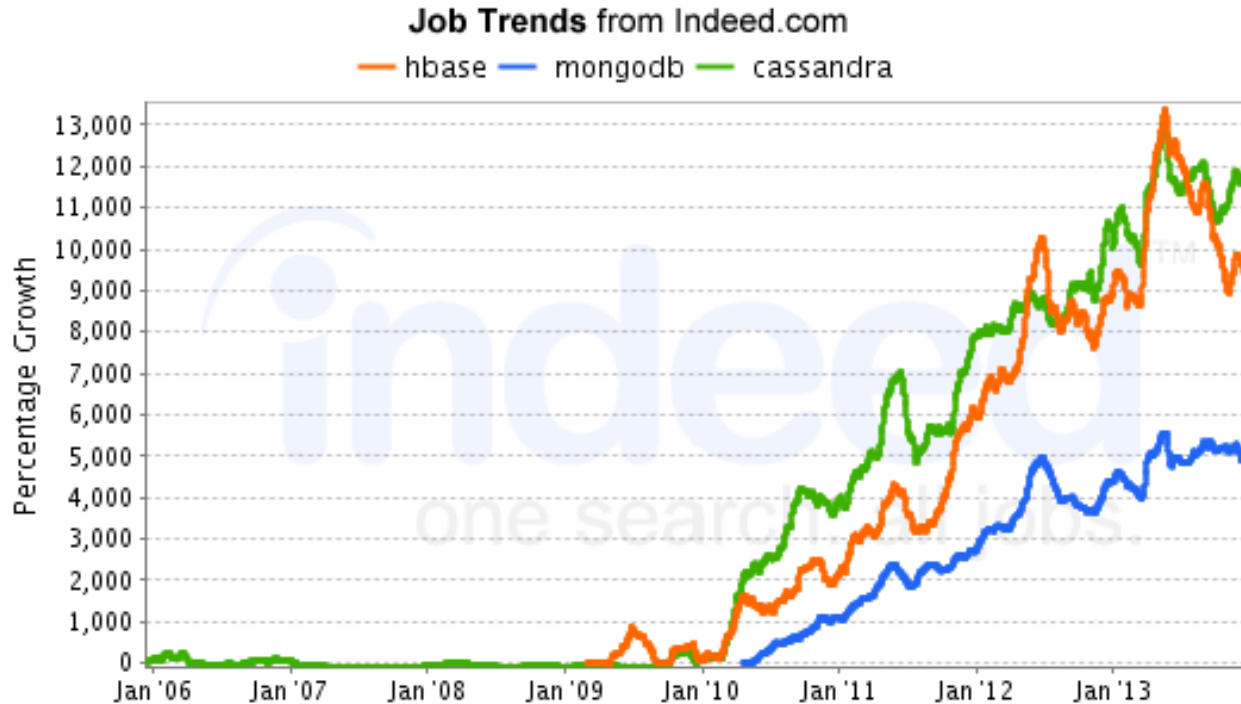
- Column family defined when table created
- Data can be very “sparse”



DynamoDB vs. HBase

- Data model
 - Key-value vs. column oriented key-value
- Non-Hadoop compatible vs. Hadoop compatible
- Proprietary vs. open source
- Cost
 - DynamoDB: Provisioned Throughput Capacity
 - Hbase: Instance(s) cost + EMR cost
- Limitations
 - DynamoDB:
 - Item size: 64 KB
 - Query result: 1 MB
- See [here](#) for more

HBase, MongoDB, Cassandra Job Trends



Indeed.com searches millions of jobs from thousands of job sites.
This job trends graph shows relative growth for jobs we find matching your search terms.

- One possible reason: the highly-publicized adoption of HBase by Facebook for a variety of production workloads

YCSB

- Yahoo! Cloud Serving Benchmark
- A standard benchmark for evaluating the performance of your database
- In this project, we will use a workload which consists of 50% reads and writes to a database
- [More details](#)
 - Learn how to interpret YCSB summary result on your won

Project 3 Module 5

- Your Task
 - Run YCSB Benchmark against DynamoDB and HBase
 - Compare performance and cost
- **Start early** (Many experiments required)
- **Terminate DynamoDB and HBase Cluster when you finish project tasks**

Upcoming Deadlines

- Project 3:

[Project 3](#)

[Working with NoSQL: DynamoDB / HBase](#)

DynamoDB vs. HBase

[Checkpoint](#)

[Available now](#)

[Due 04/06/14 11:59 PM](#)



- Unit 4:

[UNIT 4: Cloud Storage](#)

[Module 12: Cloud Storage](#)

[Module 13: Case Studies: Distributed File Systems](#)

[Module 14: Case Studies: NoSQL Databases](#)

[Module 15: Case Studies: Cloud Object Storage](#)

Quiz 4: Cloud Storage

[Checkpoint](#)

[Available now](#)

[Due 04/03/13 11:59 PM](#)



Upcoming Deadlines

- Quiz 4 Due **THIS WEEK!**
 - Thursday 04/03/2014 11:59PM Pittsburgh
 - Late submissions are NOT accepted
- Timed
 - 180 minutes once started
 - Remember to click SUBMIT before the deadline

Demo Outline

- 1. DynamoDB (no demo)
 - Remember to **delete the table** after finishing the experiments
- 2. Hbase
 - Provision EMR cluster with HBase
 - Modify Security Group
 - Connect to EMR cluster nodes
 - Create table
 - Run YCSB from launchpad (no demo)

Recommendations

- Not a difficult module, however,
 - It is time consuming
 - Can be **very expensive and jeopardize your grade** if you are not being careful
 - You need to do some reading and thinking to answer the checkpoint quiz questions