

CS15-319 / 15-619

Cloud Computing

Recitation 11

November 5th and Nov 8th, 2013

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

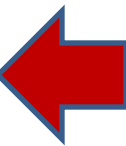
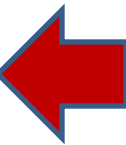
- IAM Role “tagReader”
 - **You only need to assign this IAM role to instances running some specific AMIs** (such as the one in the 619 extra project)
 - For those AMIs, we will specify it in the write up
 - For the others, you only need set the tags
 - Details in [@738](#) on Piazza
- Website problem
 - 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
 - Set up the alarm for AWS costs

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



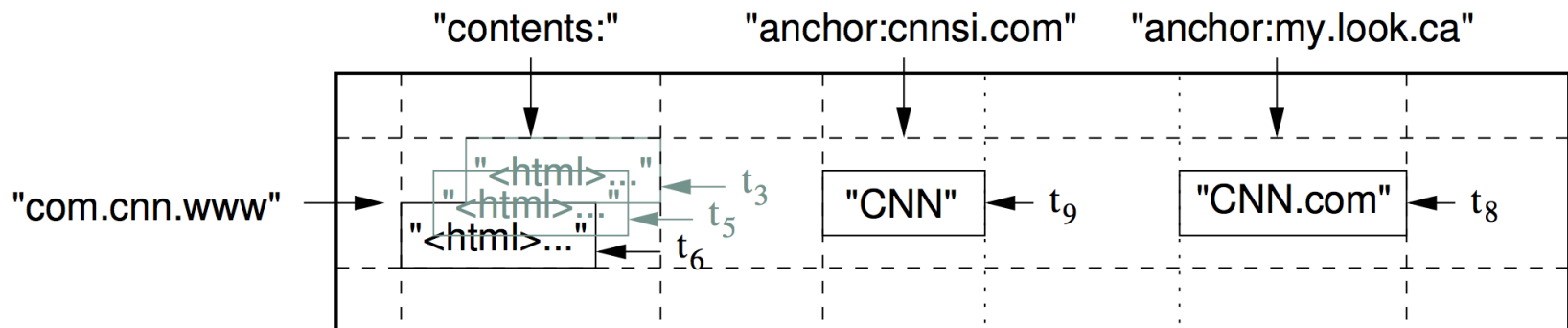
Project 3

- 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



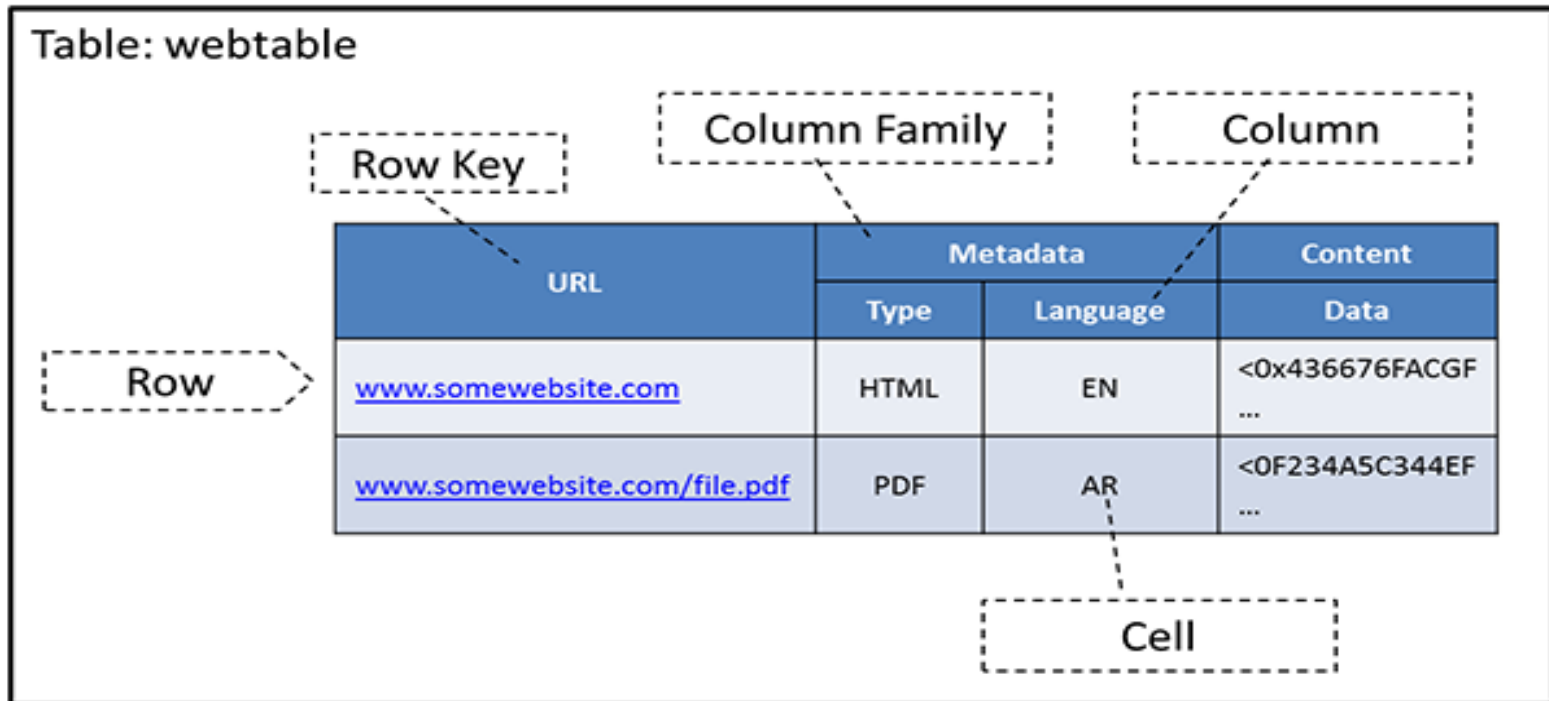
HBase

- HBase
 - NoSQL Database, based on Google's BigTable
 - [Unit 4 Module 14](#)
- Data Model
 - (row:string, column:string, time:int64) -> string
 - 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 upon table creation
- Data can be very “sparse”



DynamoDB vs. HBase

- Data Model
 - Key-value vs. Column oriented Key-value
- Non Hadoop vs. Hadoop
- Proprietary vs. Open source
- Cost
 - DynamoDB: Provisioned Throughput Capacity
 - Hbase: Instance + EMR
- Limitations:
 - DynamoDB:
 - Item size: 64 KB
 - Query result: 1 MB
- See [here](#) for more

YCSB

- Yahoo! Cloud Serving Benchmark
- A standard benchmark for evaluating the performance of different “key-value” and “cloud” serving stores
- In this project, we will use a workload which consists of 50% reads and writes to a database
- [More details](#)
 - How to interpret YCSB summary result

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 done or face a grade penalty.**

Upcoming Deadlines

- Project 3:

[Project 3](#)

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

DynamoDB vs. HBase

[Checkpoint](#)

[Available now](#)

[Due 11/10/13 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 11/7/13 11:59 PM](#)



Upcoming Deadlines

- Quiz 4 Due **THIS WEEK!**
 - Thursday 11/07/2013 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,
 - 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