15-619
Cloud Computing

15-619 Project Overview

Recitation 9
Mar 18th and Mar 20th, 2014
We are worrying...

• Phase 1 deadline is Mar 25 at 11:59 PM EDT
• By now only 47 teams have registered on the testing system
  – There are 57 teams in total...
  – Less than 20 teams have submitted requests...
• Piazza is so quiet
  – We expect a lot of questions...
619 Project, Phase 1

• What you have
  – A small Twitter dataset (0.8 million tweets, 2.3 GB)
  – Two simple queries
    • Q1 Heartbeat
    • Q2 Database query on userid and tweet_time

• What you have to do
  – Front end system development
  – Two back end databases
    • MySQL
    • HBase
619 Project Evaluation

• Development period
  – Development, evaluation, exploration, optimization

• Live test period
  – A comparison of cost, throughput between all submitted web services
619 Project Phase 1

• What we expect you to learn
  – Hands-on experience with front end systems
  – Working with HBase and MySQL
  – Design of database schema
  – A solid skill set for Phase 2 & Phase 3
  – Cost really matters!!!
    (A shared Google doc might be handy to track cost)

• Start Really Early!!

• A good start leads to success!
More on Front End

• There are numerous options
  – Java servlet, PHP, Go...
  – Read about their performance differences
• Use q1 as a tool to test your front end performance
• Consider compatibility with the back end
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”

<table>
<thead>
<tr>
<th>URL</th>
<th>Metadata</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.somewebsite.com">www.somewebsite.com</a></td>
<td>HTML</td>
<td>&lt;0x436676FACGF ...</td>
</tr>
<tr>
<td><a href="http://www.somewebsite.com/file.pdf">www.somewebsite.com/file.pdf</a></td>
<td>PDF</td>
<td>&lt;0F234A5C344EF ...</td>
</tr>
</tbody>
</table>
Work with HBase

• Install HBase
  – Add a bootstrap action when launching an EMR cluster

• Play with HBase
  – HBase Shell

• Interact with HBase
  – Rest (Project 4 uses this)
  – Thrift
  – Java API
Q & A