CS15-319 / 15-619
Cloud Computing

Recitation 8
March 4th and 6th, 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

• Search before posting

• Post feedback on OLI

• **OLI does not show timer for Quiz! You have to maintain your own timer!**
Project 3, Module 1 Reflections

• Common questions about this module:
  – Why Query 6 and Query 7 got worse performance after indexing
  – SELECT COUNT(*) FROM songs WHERE duration > (SELECT AVG(duration) FROM songs) ;
  – SELECT COUNT(*) FROM songs WHERE duration <= (SELECT AVG(duration) FROM songs) ;
Project 3, Module 1 Reflections

• Common questions about this module:
  – Why Query 6 and Query 7 got worse performance after indexing
    • CREATE INDEX idx_duration ON songs duration, artist_id(255));
    • The index is sorted by the concatenation of duration and artist_id
    • The joint index is stored as a string, instead of number
Unit 3 Quiz, Reflections

Refer to the following Figure for the next two Questions. Assume the following address mappings in a native virtualized system, where each arrow indicates a mapped page:

Question 28
Unit 3 Quiz, Reflections

**Page Table**

<table>
<thead>
<tr>
<th>Virtual Page</th>
<th>Real Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>500</td>
</tr>
<tr>
<td>800</td>
<td>300</td>
</tr>
<tr>
<td>750</td>
<td>No-map</td>
</tr>
<tr>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

**Real Map Table**

<table>
<thead>
<tr>
<th>Real Page</th>
<th>Physical Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>--</td>
</tr>
<tr>
<td>50</td>
<td>--</td>
</tr>
<tr>
<td>100</td>
<td>--</td>
</tr>
<tr>
<td>120</td>
<td>--</td>
</tr>
</tbody>
</table>

**Page Table**

<table>
<thead>
<tr>
<th>Virtual Page</th>
<th>Real Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>500</td>
</tr>
<tr>
<td>800</td>
<td>300</td>
</tr>
<tr>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

**Real Map Table**

<table>
<thead>
<tr>
<th>Virtual Page</th>
<th>Physical Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>--</td>
</tr>
<tr>
<td>50</td>
<td>--</td>
</tr>
<tr>
<td>100</td>
<td>--</td>
</tr>
<tr>
<td>120</td>
<td>--</td>
</tr>
</tbody>
</table>
Unit 3 Quiz, Reflections

Page Table: Mapping between virtual pages and real pages
Real Map Table: Mapping between real pages and physical pages
Unit 3 Quiz

- Average: 84%
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
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
Our Scenario (revisited)

• Online Photo Verification, *PixVerify*

Requests

Throughput
Previous Infrastructure

User

Application Server

Database Server
Current Infrastructure
<table>
<thead>
<tr>
<th>Year</th>
<th>Annual Number of Google Searches</th>
<th>Average Searches Per Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>2,161,530,000,000</td>
<td>5,922,000,000</td>
</tr>
<tr>
<td>2012</td>
<td>1,873,910,000,000</td>
<td>5,134,000,000</td>
</tr>
<tr>
<td>2011</td>
<td>1,722,071,000,000</td>
<td>4,717,000,000</td>
</tr>
<tr>
<td>2010</td>
<td>1,324,670,000,000</td>
<td>3,627,000,000</td>
</tr>
<tr>
<td>2009</td>
<td>953,700,000,000</td>
<td>2,610,000,000</td>
</tr>
<tr>
<td>2008</td>
<td>637,200,000,000</td>
<td>1,745,000,000</td>
</tr>
<tr>
<td>2007</td>
<td>438,000,000,000</td>
<td>1,200,000,000</td>
</tr>
<tr>
<td>2000</td>
<td>22,000,000,000</td>
<td>60,000,000</td>
</tr>
<tr>
<td>1998</td>
<td>3,600,000 <em>Googles official first year</em></td>
<td>9,800</td>
</tr>
</tbody>
</table>

Data from www.statisticbrain.com
Project 3 Module 2 - Vertical Scaling

• Explore the database performance by tweaking 2 parameters
  – Instance Type
    • m1.large
    • m1.xlarge
  – Storage Type
    • RAM Disk
    • Ephemeral Disks
    • Amazon EBS
Different Types of Storage

- Remote copy
  - Asynchronous

- Elastic Block Store (EBS)

- External Storage (Storage Subsystem)

- Availability Zone
  - Region

- Disaster Recovery

- Physical Machine
  - (Server)

- Virtual Machine
  - (Instances)

- Swatches / Routers

- CPU
- Mem
- HDD

- Memory
  - Internal HDD (RAID?)

- RAM Disk
- Ephemeral Disk

- Virtual Machines
- Physical Resources
Different Types of Storage

• **Memory - RAM Disk**
  – Inside the server
  – Usually from several Gigabytes to several hundreds of Gigabytes

• **Internal HDD (Hard Disk Drive)**
  – Inside the server
  – Sometimes employs RAID (Why?)
  – Usually from 100s Gigabytes to several Terabytes
Different Types of Storage

• External Storage Subsystems
  – Outside of the server
  – Connected by cables via switches, routers, directors (Ethernet, Fiber...)
  – Provide extra functionalities (Copy services, concurrent volume accesses, grouping, caching...)
  – Shared by multiple servers
  – Almost always employs RAID
  – Capacity range from dozens of TB to 100s of TB
Different Types of Storage

• **External Storage Subsystems**

IBM 2424-951 DS8800 182TB RAW 129TB
useable w/RAID 5 SYSTEM STORAGE
On eBay: US $899,995.00

EMC SYMTRIX VMAX 40K
Explore the database performance by manipulating 2 parameters:

- Local Access VS. Remote Access
- m1.large VS. m1.xlarge
- RAM Disk / ephemeral disk / ephemeral disk with RAID0 / EBS
- EBS optimized VS. no EBS optimized
Upcoming Deadlines

• **Project 3:**

<table>
<thead>
<tr>
<th>Files vs. Databases</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Gradebook)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Learning Dashboard)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>File vs. Database</td>
<td></td>
<td>Checkpoint</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ended 3/2/14 11:59 PM</td>
</tr>
<tr>
<td><strong>Vertical Scaling in Databases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Gradebook)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Learning Dashboard)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vertical Scaling</strong></td>
<td></td>
<td>Checkpoint</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Available Now</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Due 3/9/14 11:59 PM</td>
</tr>
</tbody>
</table>

• **Unit 4:**

**UNIT 4: Cloud Storage**

**Module 12: Cloud Storage**

**Module 13: Case Studies: Distributed File Systems**
Demo Outline

• Launch Instance with EBS Storage.
• Common Disk Operations in Linux(RAMDISK)
  – Create RAMDISK
  – Mount a Volume
  – Bind a mounted volume to multiple mount-points
  – Running sysbench
Disk Operations Commands

- **mount/umount**
  - attach the file system found on some device to the big file tree

- **dd**
  - Copy and convert file

- **mkfs.ext4**
  - Create an ext4 file system