

CS15-319 / 15-619

Cloud Computing

Recitation 8

October 15th and 18th, 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

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

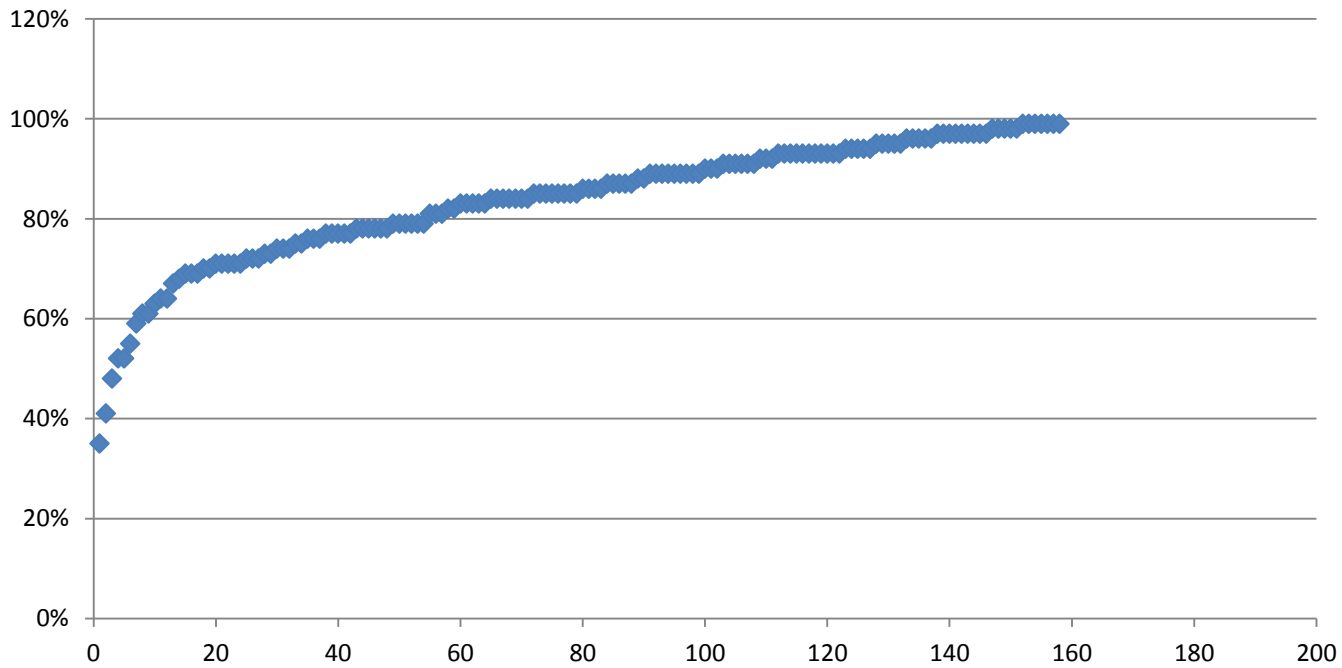
- Common questions about this module:
 - Why Query 6 and Query 7 got worse performance after indexing
 - `CREATE INDEX idx_duration ON songs duration, artist_id(255));`
 - The index is sorted by the **concatenation of duration and artist_id**
 - Binary search can be used for searching
 - `Average()` will not benefit from binary search
 - `Count(*)` also seems have some negative effects

Project 3, Module 1 Reflections

- Common questions about this module:
 - Why did Query 6 and Query 7 get worse performance after indexing
 - Indexes and real raw data are not residing together
 - For an average: 2 disk reads happen, 1 for index, 1 for real data, which is slow.
 - Different databases have different implementations

Unit 3 Quiz

- Average: **78%**



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

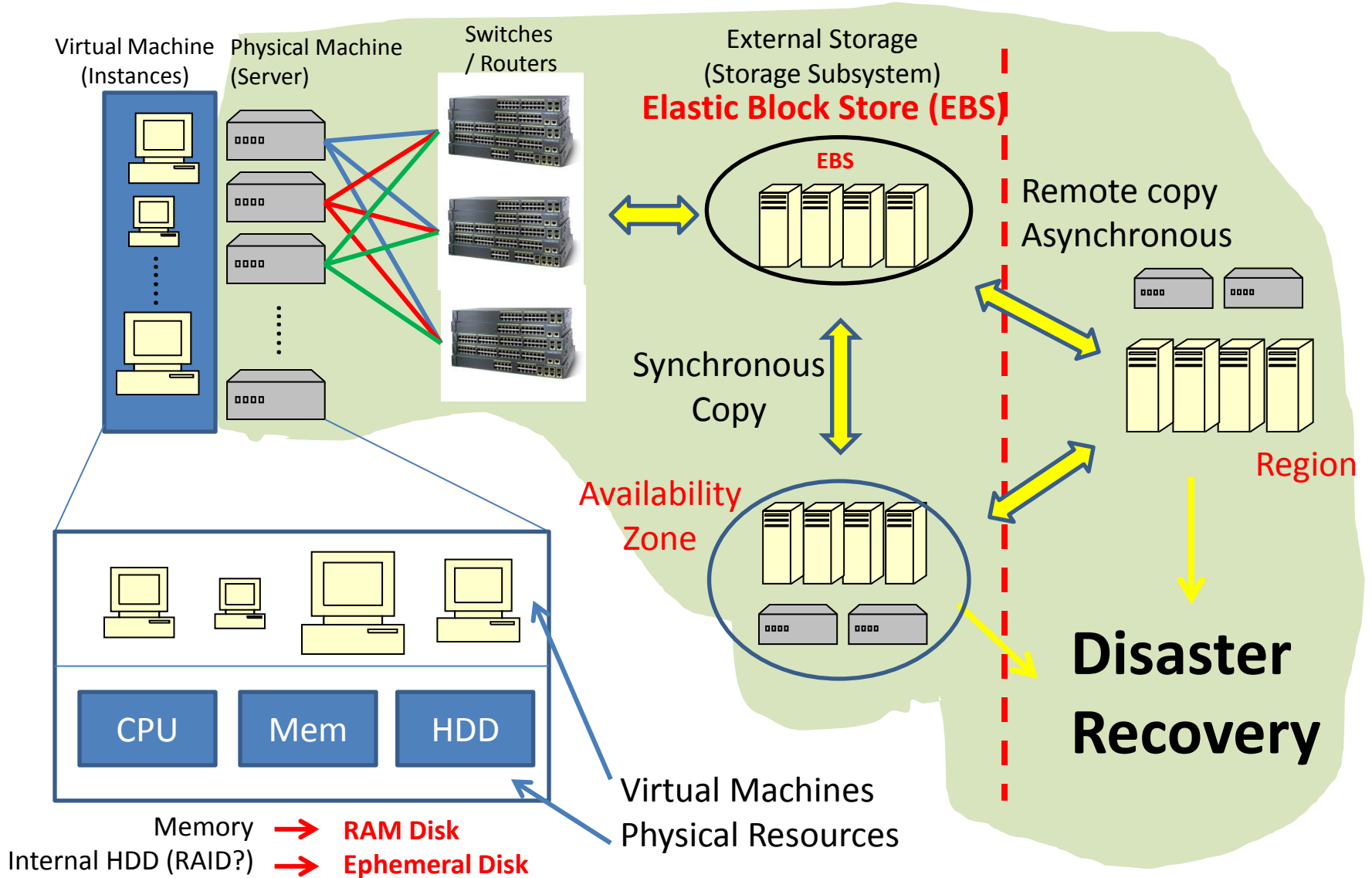
- Files vs. Databases
 - File vs. Database
- Vertical Scaling in Databases
 - Vertical Scaling
- Horizontal Scaling in Databases
 - Horizontal Scaling
- Working with NoSQL: DynamoDB / Hbase
 - Amazon DynamoDB
 - DynamoDB vs. HBase



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



Different Types of Storage

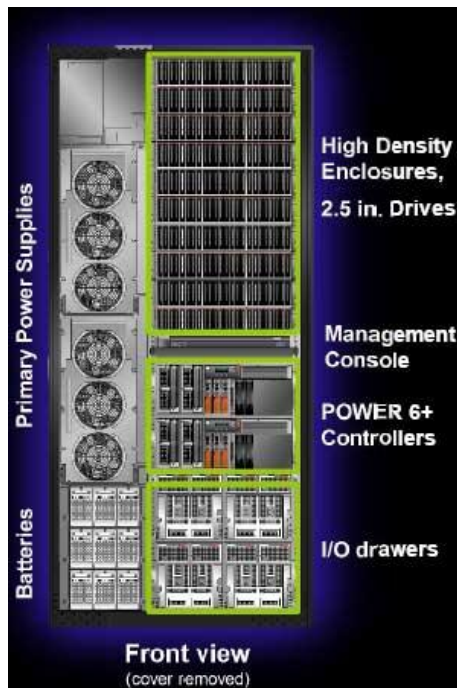
- **Memory - RAM Disk**
 - Inside the server
 - Usually from several Gigabytes to several hundreds of Gigabytes
- **Internal HDD (Hard Disk Drive)**
 - Inside the server
 - Sometimes employs RAID (Why?)
 - Usually from 100s Gigabytes to several Terabytes

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

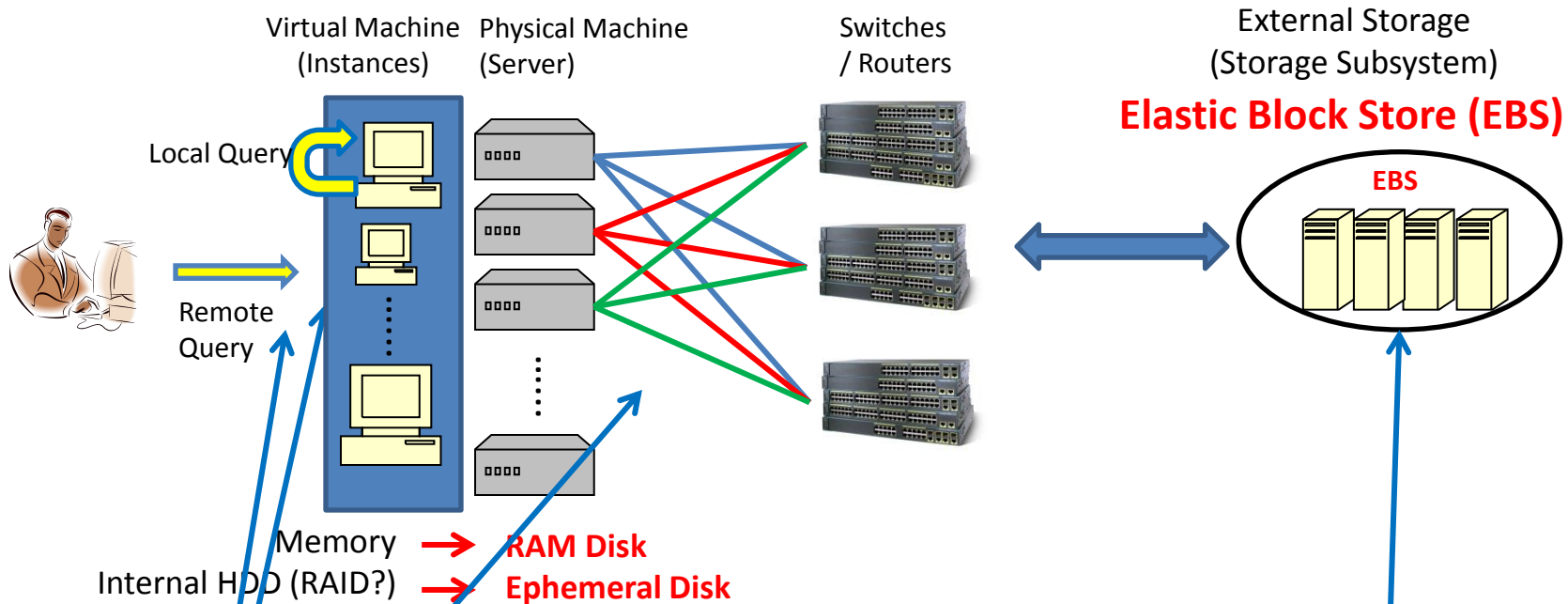
- External Storage Subsystems



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

EMC SYMITRIX VMAX 40K

Project 3 Module 2 – Vertical Scaling



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:

[Project 3](#)

[Vertical Scaling in Databases](#)

Vertical Scaling

[Checkpoint](#)

[Available Now](#)

[Due 10/20/13 11:59 PM](#)



- Unit 4:

[UNIT 4: Cloud Storage](#)

[Module 12: Cloud Storage](#)

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



Demo Outline

- Provisioning with spot instances
- Running sysbench
 - RAM disk
 - Ephemeral disk