

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

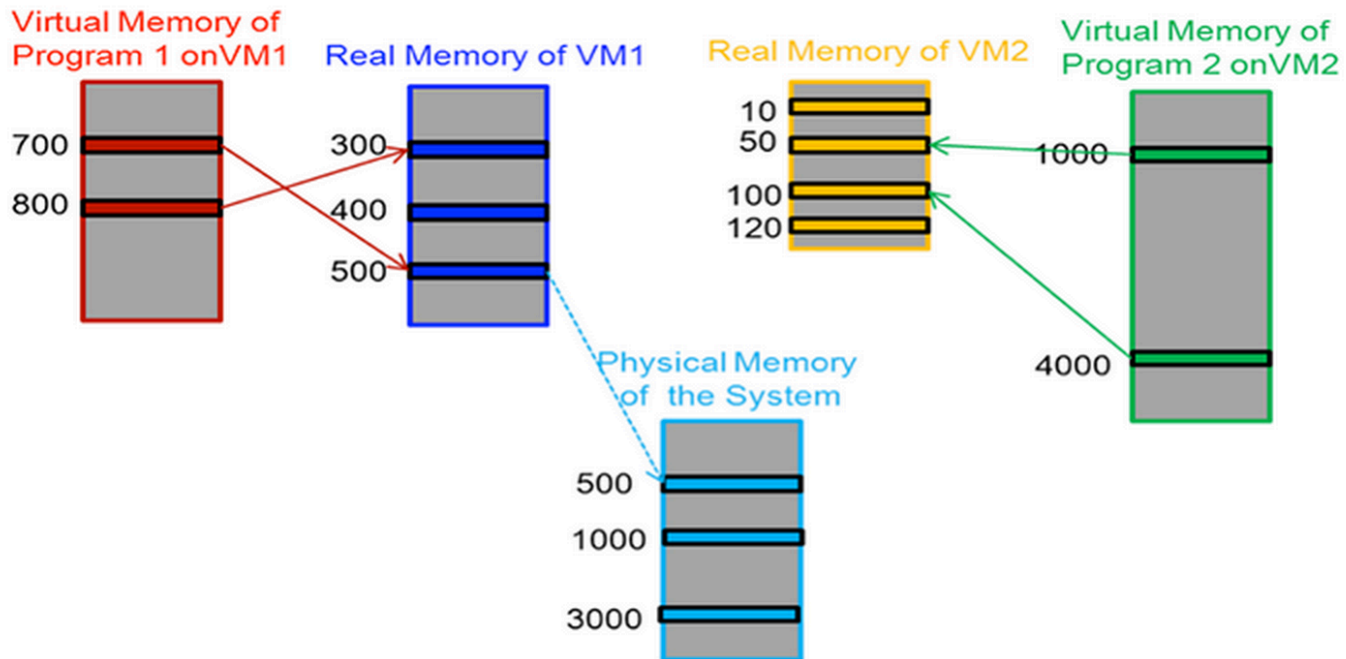
- Common questions about this module:
 - Why Query 6 and Query 7 got worse performance after indexing
 - `SELECT COUNT(*) FROM songs WHERE duration > (SELECT AVG(duration) FROM songs) ;`
 - `SELECT COUNT(*) FROM songs WHERE duration <= (SELECT AVG(duration) FROM songs) ;`

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

Virtual Page	Real Page
700	500
800	300
750	No-map
--	--

Page Table

Real Page	Physical Page
10	--
50	--
100	--
120	--

Real Map Table

Virtual Page	Real Page
700	500
800	300
--	--
--	--

Page Table

Virtual Page	Physical Page
10	--
50	--
100	--
120	--

Real Map Table

Unit 3 Quiz, Reflections

Virtual Page	Real Page
700	500
800	300
750	No-map
--	--

Page Table

Real Page	Physical Page
10	--
50	--
100	--
120	--

Real Map Table

Virtual Page	Real Page
700	500
800	300
--	--
--	--

Page Table

Virtual Page	Physical Page
10	--
50	--
100	--
120	--

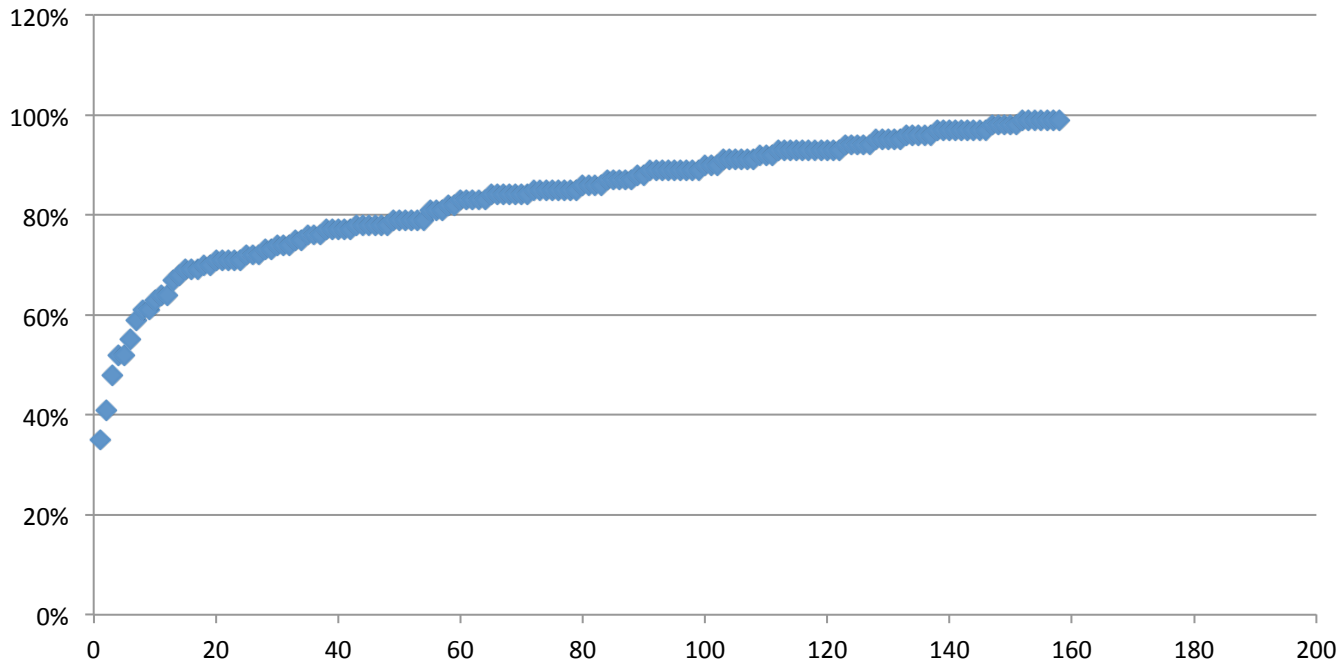
Real Map Table

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*

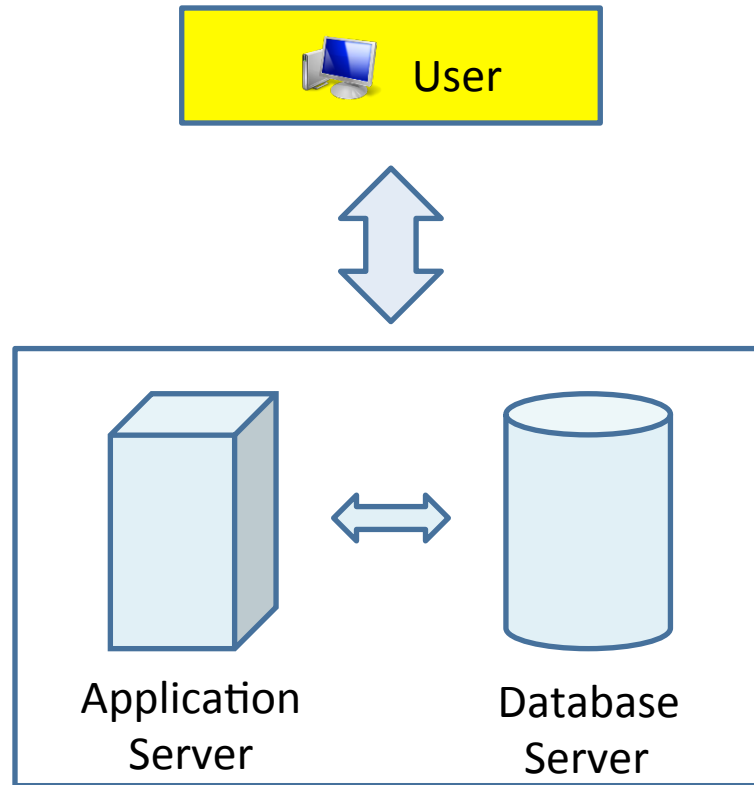


Requests

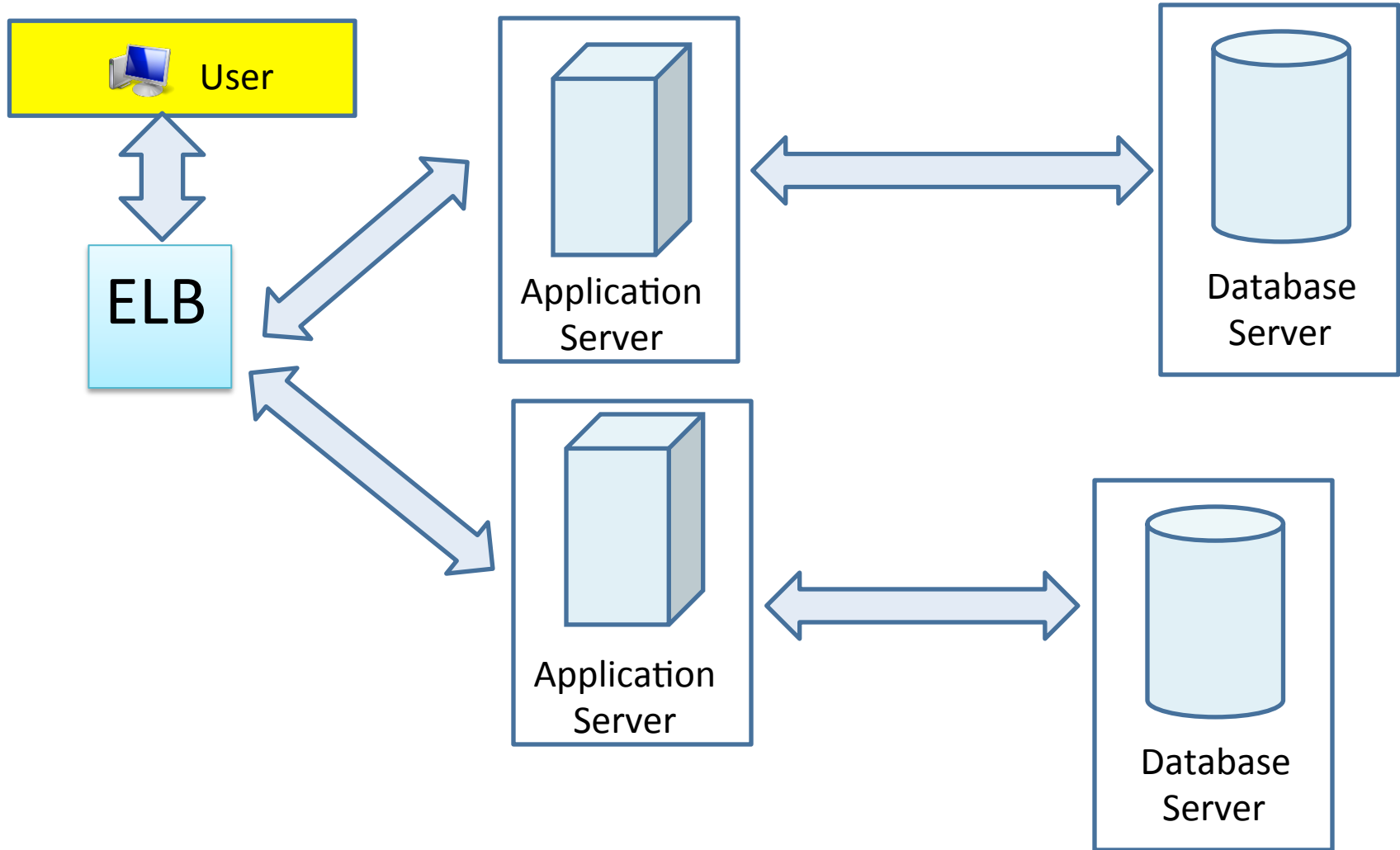


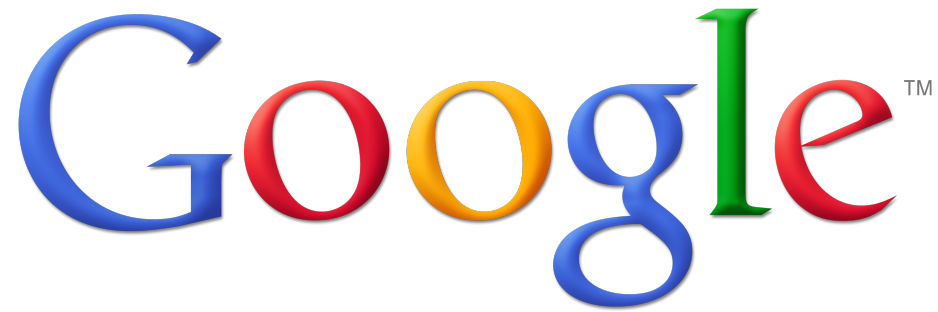
Throughput

Previous Infrastructure



Current Infrastructure



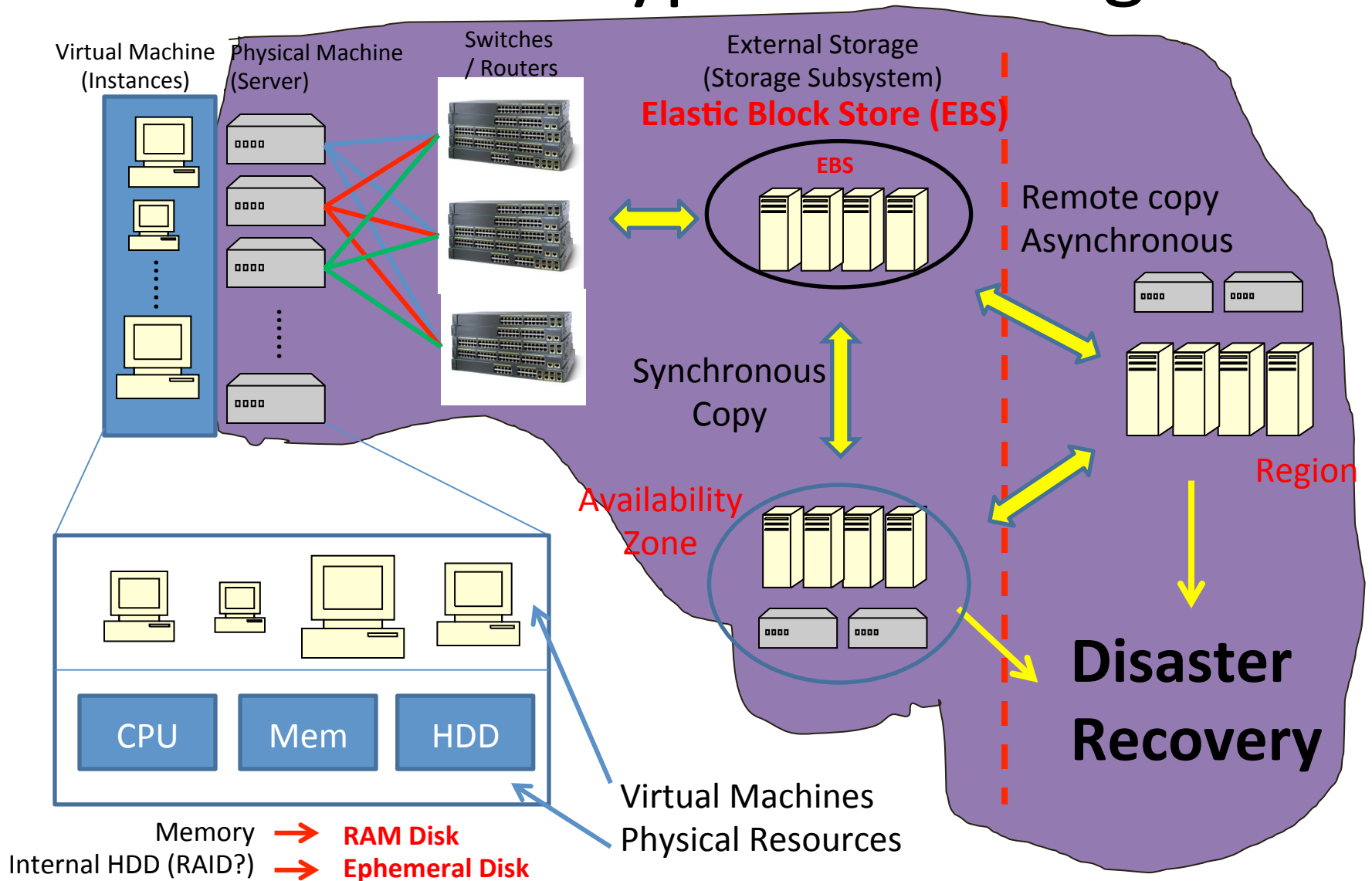


Year	Annual Number of Google Searches	Average Searches Per Day
2013	2,161,530,000,000	5,922,000,000
2012	1,873,910,000,000	5,134,000,000
2011	1,722,071,000,000	4,717,000,000
2010	1,324,670,000,000	3,627,000,000
2009	953,700,000,000	2,610,000,000
2008	637,200,000,000	1,745,000,000
2007	438,000,000,000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

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

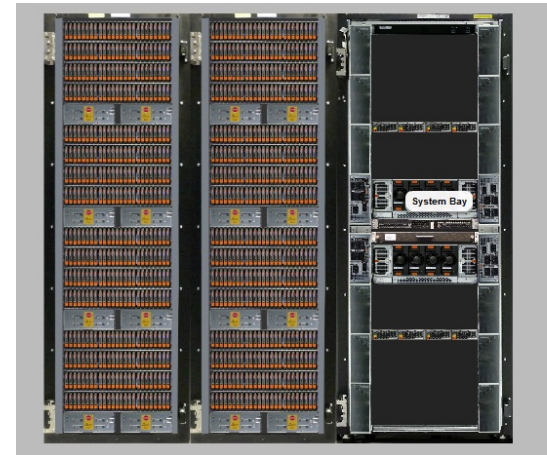
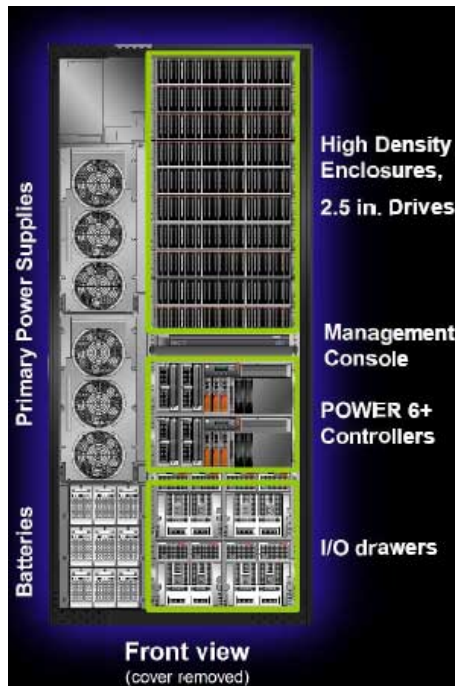
- **Memory - RAM Disk**
 - Inside the server
 - Usually from several Gigabytes to several hundreds of Gigabytes
- **Internal HDD (Hard Disk Drive)**
 - Inside the server
 - Sometimes employs RAID (Why?)
 - Usually from 100s Gigabytes to several Terabytes

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

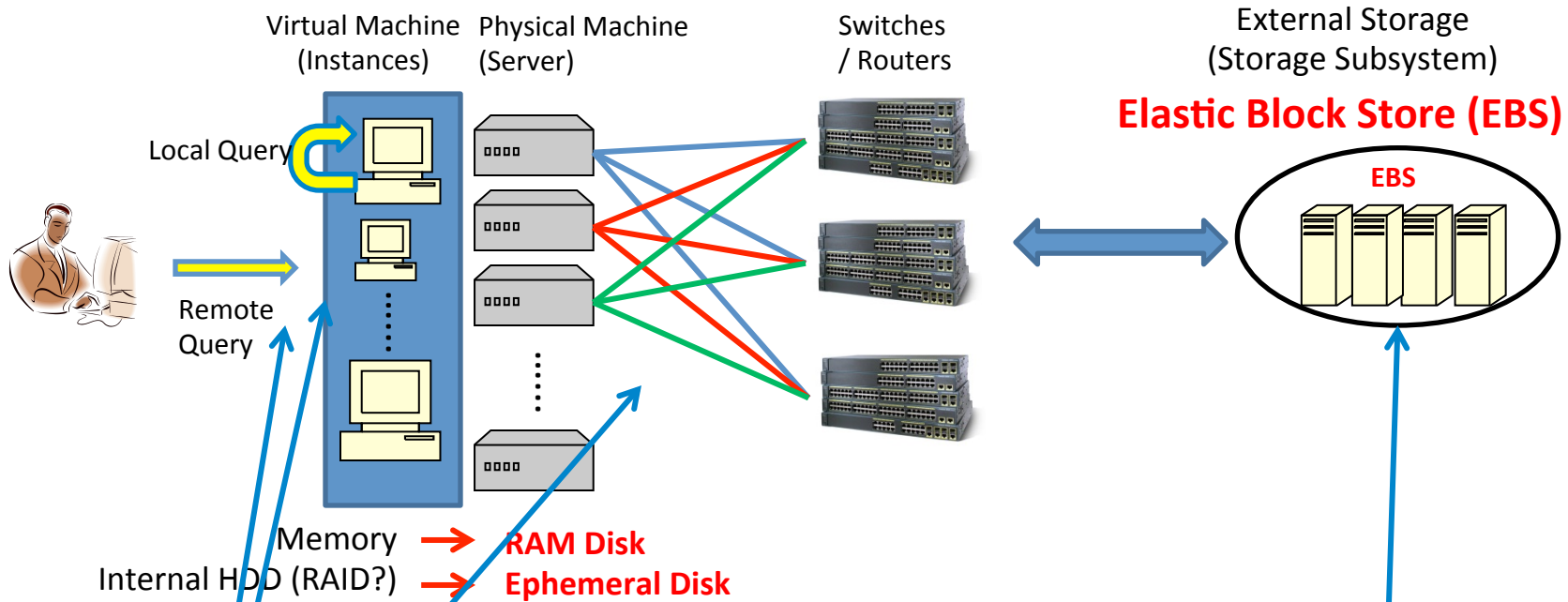
- External Storage Subsystems



**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		
Files vs. Databases (Gradebook) (Learning Dashboard)		
File vs. Database	Checkpoint	Ended 3/2/14 11:59 PM
Vertical Scaling in Databases (Gradebook) (Learning Dashboard)		
Vertical Scaling	Checkpoint	Available Now Due 3/9/14 11:59 PM



- 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