CS15-319 / 15-619
Cloud Computing

Recitation 8

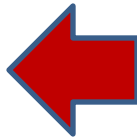
October 14th and 16th, 2014

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!**

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

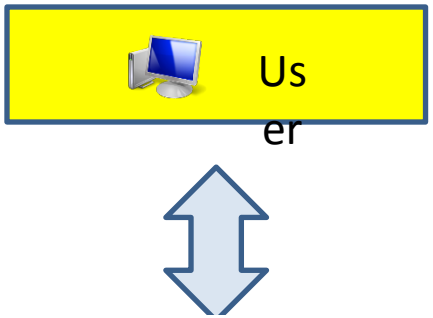


Project 3

- Files vs. Databases
  - File vs. Database
- Vertical Scaling in Databases
  - Vertical Scaling
- Horizontal Scaling in Databases
  - Horizontal Scaling
- Sharding Databases
- Provisioned Databases

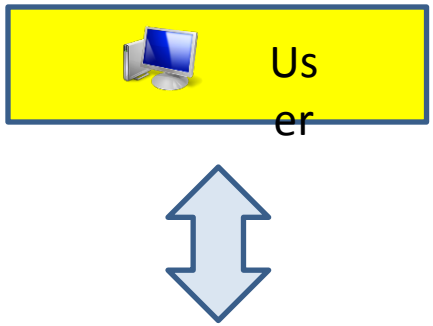
Our Scenario

- User has to open a photo on his local machine



HDD

VS



SSD

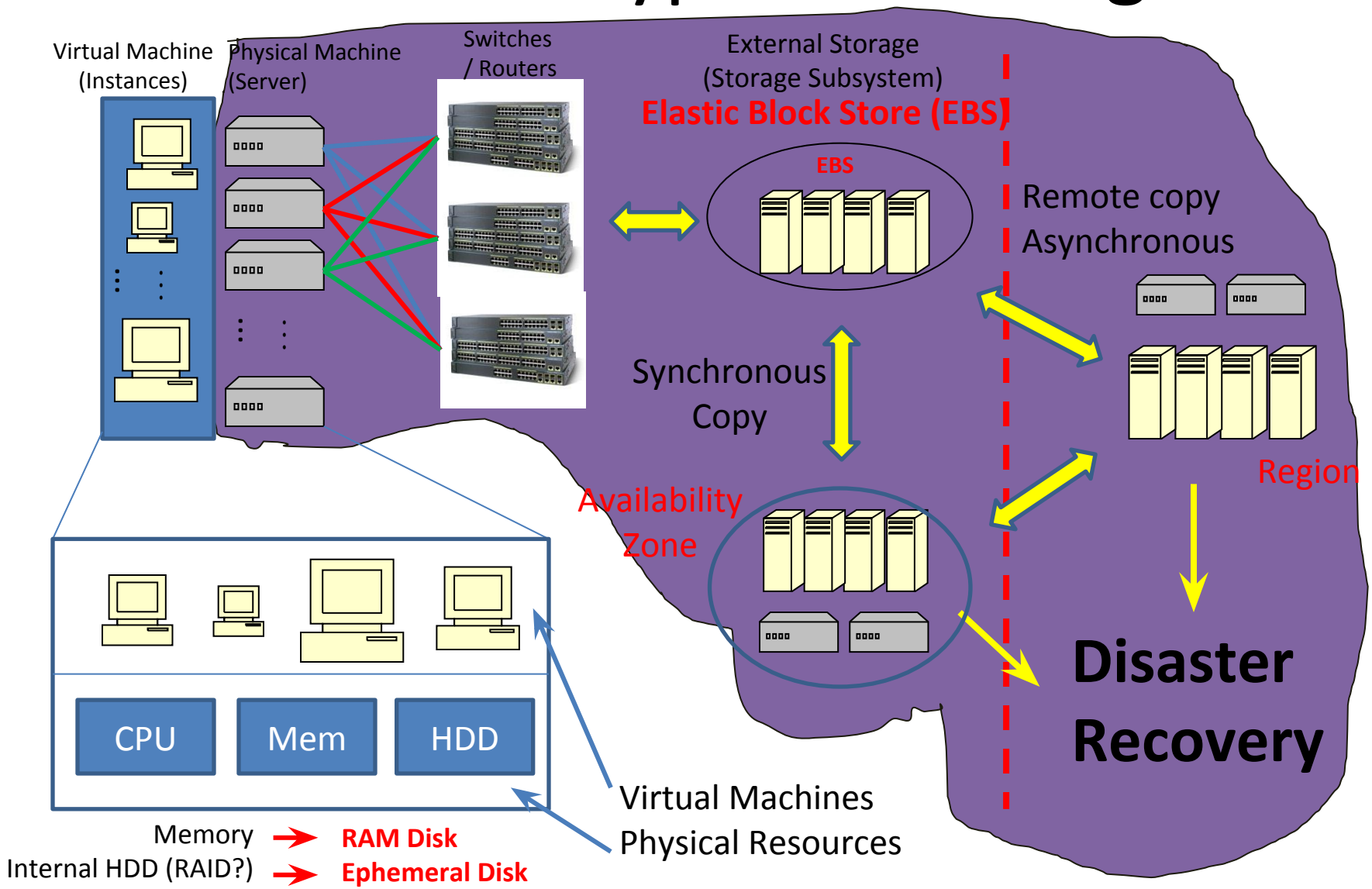
Our Scenario

- Which storage resource offers an improvement in performance and for what type of applications?
- Why?
- Which one would you utilize/receommend?

Brief overview of Project 3.2

- Vertical scaling in storage technologies
- Compare between EBS Magnetic disks, and Solid State Drive
- Measure performance for three distinct scenarios
  - FileIO
  - MySQL Database
  - HBase cluster
- Aim is to understand the benefits of employing faster storage systems.

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)
  - Mechanical Disk
  - Usually from 100s Gigabytes to several Terabytes
  - Work best with large files

Different Types of Storage

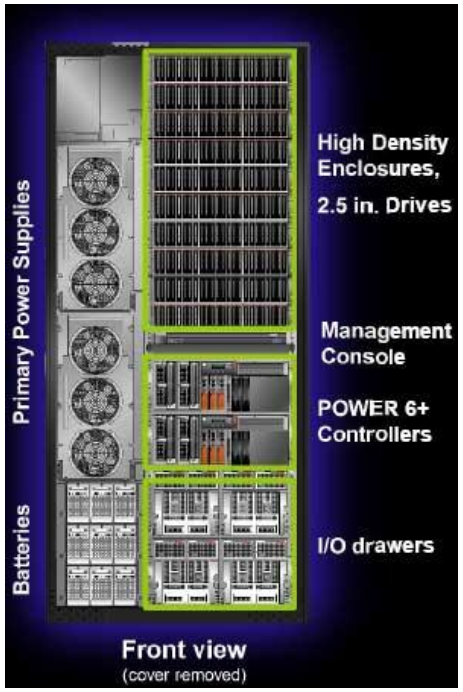
- Internal SSD (Solid State Drive)
  - Data is stored on chips
  - Much faster access
  - Storage capacity is not as high as HDD, but it is slowly catching up

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

Upcoming Deadlines

- Project 3 :

Files vs. Databases (Gradebook) (Learning Dashboard)		
File vs. Database	Checkpoint	Ended 10/12/14 11:59 PM
Vertical Scaling in Databases (Gradebook) (Learning Dashboard)		
Vertical Scaling	Checkpoint	**Available Now** Due 10/19/14 11:59 PM

- Unit 4:

UNIT 4: Cloud Storage

- Module 12: Cloud Storage
- Module 13: Case Studies: Distributed File Systems

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

Potential roadblocks

- Read the Vertical scaling section first and use “Common disk operations in Linux”
- Before running sysbench , bind your volume to /var/lib/mysql

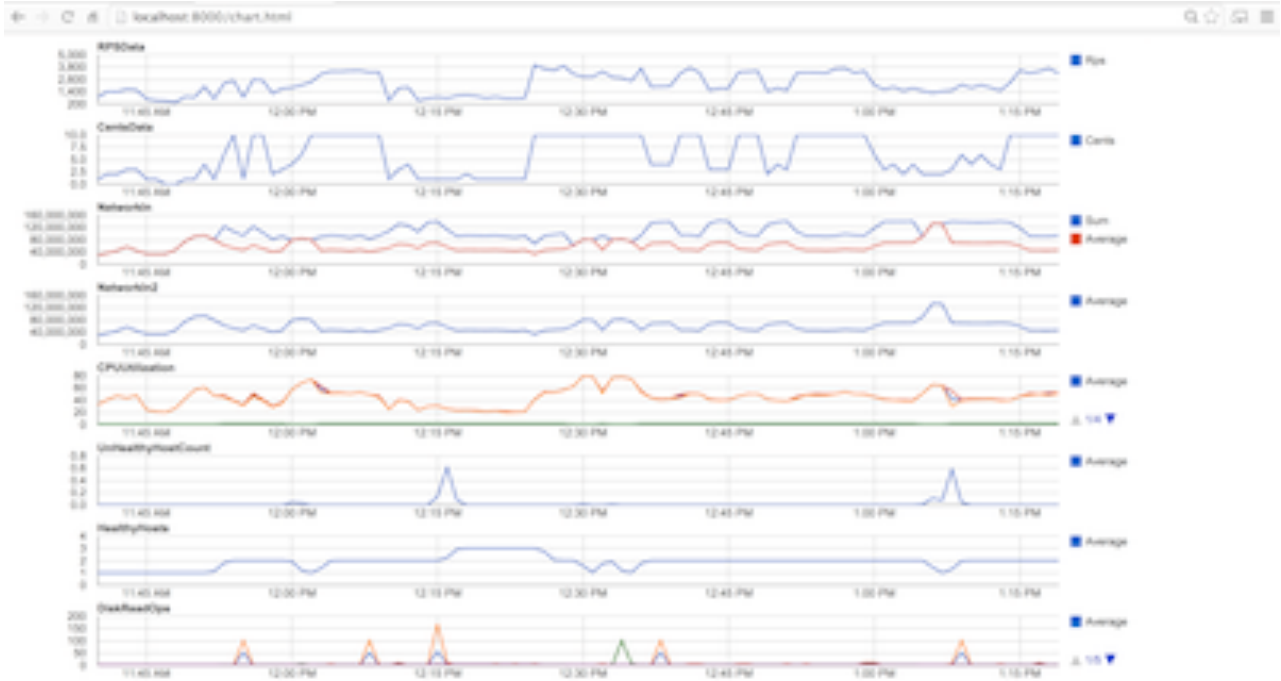
Project 2 : Scaling

Oh, no

Not this again!!!

Motivations behind P2

2.2	2.3
Importance of warm-up	More warmup = More dead instances
Rapid scaling	Scale down!!!
Identifying traffic pattern using CW	Heterogeneous resources
Scaling too frequently (oscillating)	Failure detection and handling
Measurements	Trade-offs



TWITTER ANALYTICS:
THE 619 PROJECT

Congratulations Team FDU

- Building a performant web server is non-trivial
- Building a performant web server that gets huge amounts of data from MySQL is hard
- Building a performant web server that gets huge amounts of data from HBase is harder

- **START NOW!!!**

START

NOW

Q1 : Semiprime Factorization

- Easy to implement?
- Can you optimize? How much impact does the calculation have on throughput?
- Semiprime factorization used in:
  - RSA
  - Blum-Blum-Shlub

Q2A : Naïve Sentiment Analysis

Amazingly, despite the nice,cloudy weather, the BEST Hope for us to enjoy is to study CLOUD COMPUTING.Cloud is supper interesting.

amazing	4		interesting	3	
best	3		enjoy	1	
nice	2		super	7	
hope	2		study	-100	

Score:??

Q2A : Naïve Sentiment Analysis

Amazingly, despite the nice,cloudy weather, the BEST Hope for us to enjoy is to study CLOUD COMPUTING.Cloud is supper-interesting.

amazing	4
best	3
nice	2
hope	2

interesting	3
enjoy	1
super	7
study	-100

Score:-89

Q2B : Text Censorship

Amazingly, despite the nice,cloudy weather, the BEST Hope for us to enjoy is to study CLOUD COMPUTING.Cloud is supper-interesting.

Banned words
pybhq
vagrerfgvat

Q2B : Text Censorship

Amazingly, despite the nice,cloudy weather, the BEST Hope for us to enjoy is to study CLOUD COMPUTING.Cloud is supper-interesting.

Banned words
cloud
interesting

Guess the output

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study CLOUD COMPUTING.Cloud
is supper-interesting.
```

```
Amazingly, despite the nice,[redacted]weather, the BEST
Hope for us to enjoy is to study [redacted]COMPUTING.C[redacted]
is [redacted]
```

a	cloudy
b	c***dy
c	c****y
d	

Guess the output

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study CLOUD COMPUTING.Cloud
is supper-interesting.
```

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study C ? COMPUTING.C
is
```

a	CLOUD
b	C***D
c	cloud
d	c***d

Guess the output

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study CLOUD COMPUTING.Cloud
is supper-interesting.
```

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study C***D COMPUTING.C ?
is
```

a	Cloud
b	C***D
c	cloud
d	C***d

Guess the output

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study CLOUD COMPUTING.Cloud
is supper-interesting.
```

```
Amazingly, despite the nice,cloudy weather, the BEST
Hope for us to enjoy is to study C***D COMPUTING.C***d
is ?
```

a	s******************g
b	supper-i*********g
c	supper-interesting
d	supper-i**********g

Guess the output

الحوسبة السحابية

बादल कंप्यूटिंग

云计算

クラウドコ

ンピューティング

ಕೌಲ್ಡ್ ಕಂಪೂಯ್ಟಿಂ

ווּאָלקן קאַמפּיוטינג

облачныхвычислений

This is the output

124142412412:0:الحوسبة السحابية\n
बादल कंप्यूटिंग\n
云计算\n
クラウドコ\n
ンピューティング\n
ಕೌಲ್ಡ್ ಕಂಪ್ಯೂಟಿಂ\n
וואָלקן קאַמפּיוטינಗ್\n
גоблачныхвычислений\n
\n

Tagging

All instances used for 619 work should be tagged with: *"**15619project**" and value "**phase1**".*

All instances in your HBase cluster should be tagged with: **"15619backend"** and value **"hbase"**.

All instances with MySQL installed should be tagged with: **"15619backend"** and value **"mysql"**.

Read Piazza

@1316

@1336

@1313

@1302

@1326

Questions?