

CS15-319 / 15-619

Cloud Computing

Recitation 4

September 17th & 20th, 2013

Announcements

- Encounter a general bug:
 - Post on Piazza
- Encounter a grading bug:
 - Post Privately on Piazza
- Post feedback on OLI

Unit 2: Data Centers

- Data Centers Unit
 - Module 3: Data Center Trends
 - Module 4: Data Center Components
- Read and complete:
 - Module 5: Design Considerations
 - Unit 2: Checkpoint Quiz
 - Timed Quiz (120 minutes), due 19th September, 2013

[UNIT 2: Data Centers](#)

[Module 3: Data Center Trends](#)

[Module 4: Data Center Components](#)

[Module 5: Design Considerations](#)

Quiz 2: Data Centers

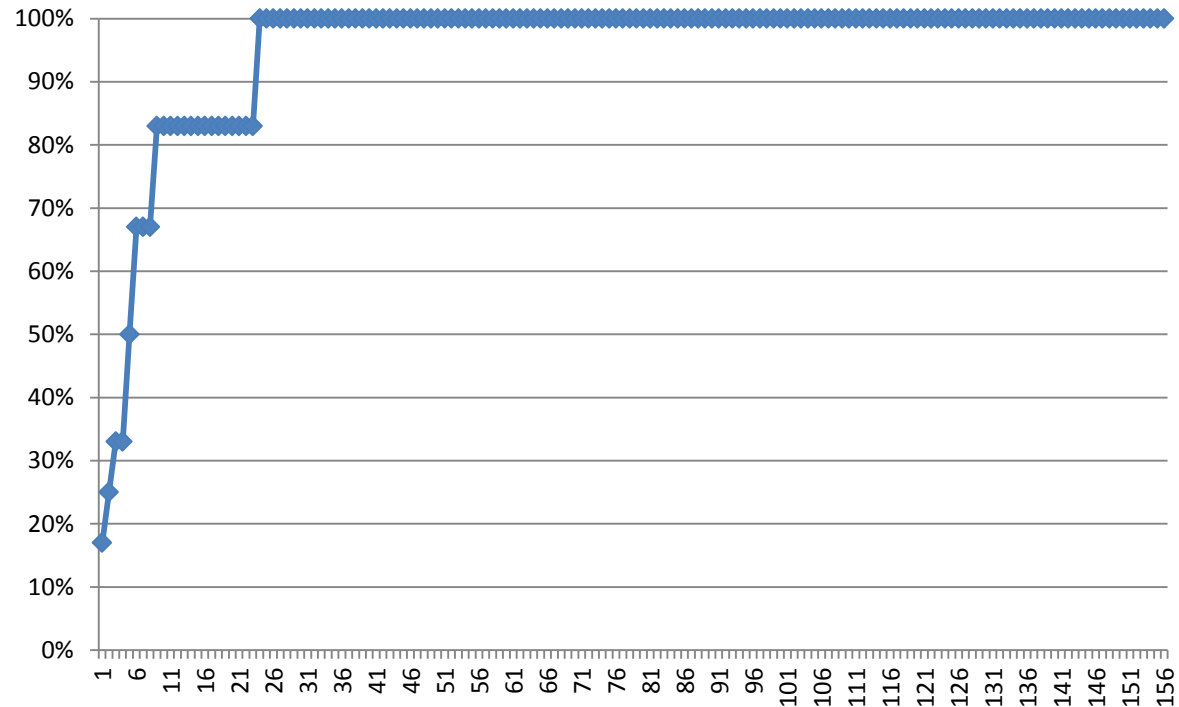
[Checkpoint](#)

[Available 9/16/13 12:01 AM](#)
[Due 9/19/13 11:59 PM](#)



Project 1 Student Progress

- Introduction to Big Data:
 - Sequential Analysis: Average is: **98%**
 - Elastic MapReduce: **96%**



Piazza Questions

- Elastic MapReduce Billing Question
 - [Normalized Hours \(Elastic MapReduce\)](#)

Date	Elapsed Time	Normalized Instance Hours
1 11:59 EDT	1 hour 46 minutes	40

- 1 hour of m1.small = 1 hour normalized compute time
- 1 hour of m1.medium = 2 hours normalized compute time
- 1 hour of m1.large = 4 hours normalized compute time
- 1 hour of m1.xlarge = 8 hours normalized compute time
- 1 hour of c1.medium = 2 hours normalized compute time

Piazza Questions

- Elastic MapReduce Billing Question
 - [Elastic MapReduce Pricing](#)

Region: <input type="text" value="US East (N. Virginia)"/>		
	Amazon EC2 Price	Amazon Elastic MapReduce Price
Standard On-Demand Instances		
Small (Default)	\$0.06 per hour	\$0.015 per hour
Medium	\$0.12 per hour	\$0.03 per hour
Large	\$0.24 per hour	\$0.06 per hour
Extra Large	\$0.48 per hour	\$0.12 per hour


Piazza Questions

- Elastic MapReduce Debugging


Debug a Job Flow



Close X

Job Flow: My Job Flow (j-36Q4Q4B3WQYNF)

 Job flow failed with reason: Shut down as step failed

[Steps](#) → [Jobs](#) → [Tasks](#) → [Task Attempts](#)

 Refresh List

Step	Name	State	Start Time	Log Files	Actions
1	Setup Hadoop Debugging	 COMPLETED	2013-09-10 09:57 EDT	controller stderr stdout syslog	View Jobs
2	Streaming Job	 FAILED	2013-09-10 09:57 EDT	controller stderr stdout syslog	View Jobs


Debug a Job Flow


Close X

Job Flow: My Job Flow (j-36Q4Q4B3WQYNF)

View logs for steps, Hadoop jobs, tasks, and task attempts.

[Steps](#) → [Jobs](#) → [Tasks](#) → [Task Attempts](#)

 Refresh List

Job	Step	State	Start Time	Actions
job_201309101355_0001	2	 FAILED	2013-09-10 09:58 EDT	View Tasks


Piazza Questions

- Elastic MapReduce Debugging

Hadoop Job: job_201309101355_0001

Task Summary: 33 Total Tasks - 0 Completed, 0 Running, 5 Failed, 0 Pending, 28 Cancelled.


Steps → Jobs → **Tasks** → Task Attempts

 Refresh List

Task	Type	Job	State	Start Time	Elapsed Time	Actions
m_000009	map	201309101355_0001	● FAILED	2013-09-10 09:58 EDT	0 hours 1 minute	View Attempts
m_000008	map	201309101355_0001	● FAILED	2013-09-10 09:58 EDT	0 hours 1 minute	View Attempts
m_000002	map	201309101355_0001	● FAILED	2013-09-10 09:58 EDT	0 hours 1 minute	View Attempts
m_000001	map	201309101355_0001	● FAILED	2013-09-10 09:58 EDT	0 hours 1 minute	View Attempts

view logs for steps, hadoop jobs, tasks, and task attempts.

Steps → Jobs → **Tasks** → Task Attempts

 Refresh List

Attempt	Job	Task	Type	State	Start Time	Log Files
3	201309101355_0001	m_000000	map	● FAILED	2013-09-10 09:59 EDT	stderr stdout syslog
2	201309101355_0001	m_000000	map	● FAILED	2013-09-10 09:59 EDT	stderr stdout syslog
1	201309101355_0001	m_000000	map	● FAILED	2013-09-10 09:58 EDT	stderr stdout syslog
0	201309101355_0001	m_000000	map	● FAILED	2013-09-10 09:58 EDT	stderr stdout syslog

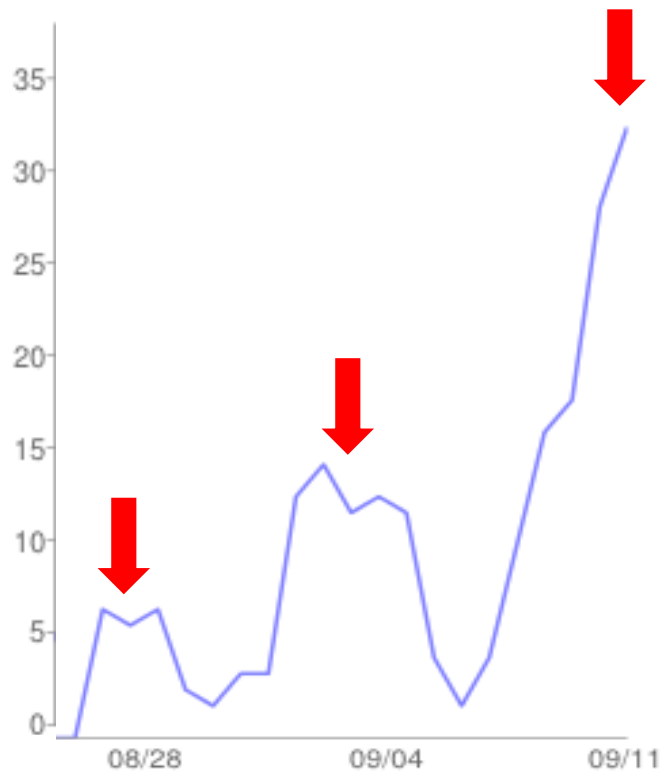
Piazza Questions

- Elastic MapReduce Debugging

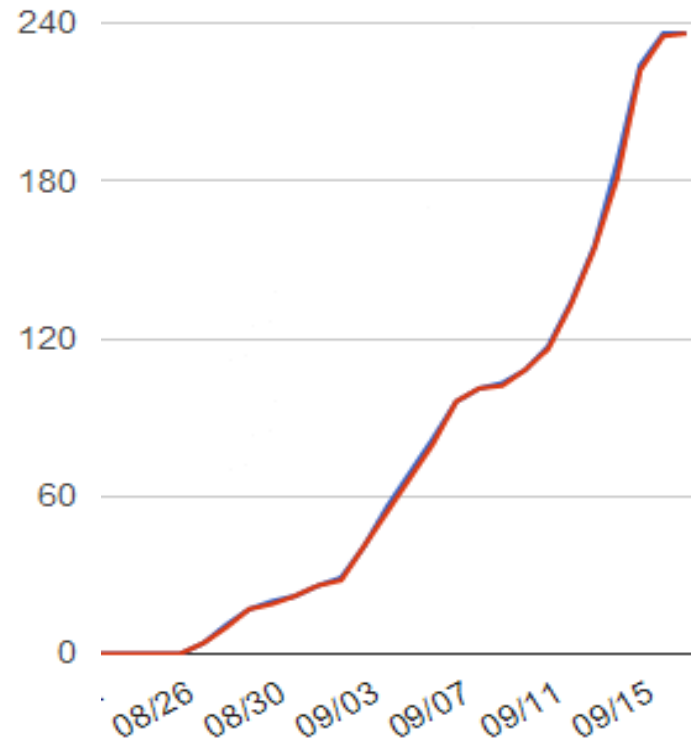
```
/mnt/var/lib/hadoop/mapred/taskTracker/hadoop/jobcache/job_201309101355_0001/attempt_201309101355
/mnt/var/lib/hadoop/mapred/taskTracker/hadoop/jobcache/job_201309101355_0001/attempt_201309101355
unexpected token `('
/mnt/var/lib/hadoop/mapred/taskTracker/hadoop/jobcache/job_201309101355_0001/attempt_201309101355
is_in_filtered_title(line_item_list):'
java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 2
    at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:372)
    at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:586)
    at org.apache.hadoop.streaming.PipeMapper.map(PipeMapper.java:125)
    at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:50)
    at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:36)
```

Piazza Traffic

Posts Per Day



Total Questions Asked



OLI Feedback

- Students want to know more about MapReduce

Scenario

- Online Photo Verification, *PixVerify*

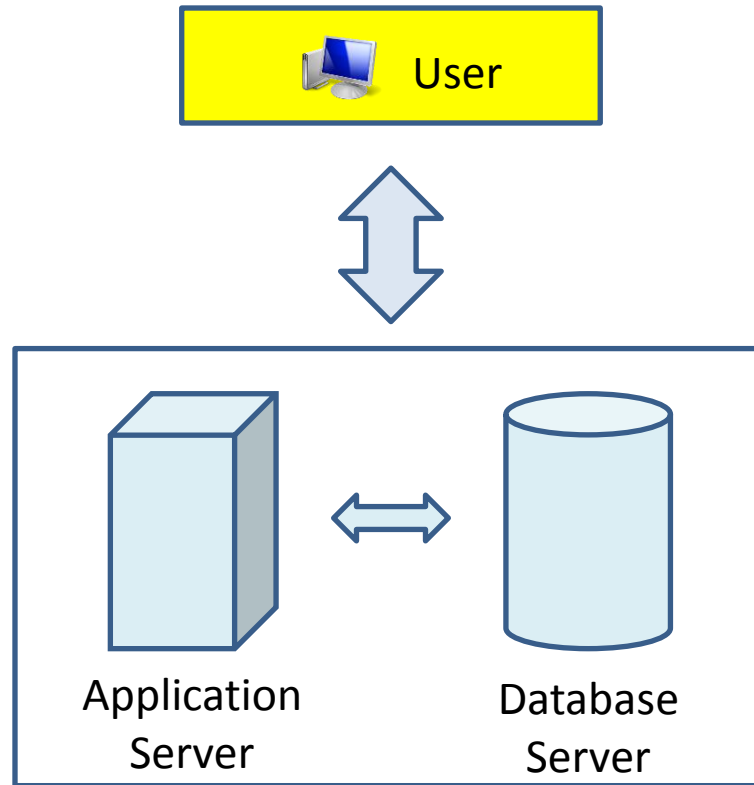


Sales

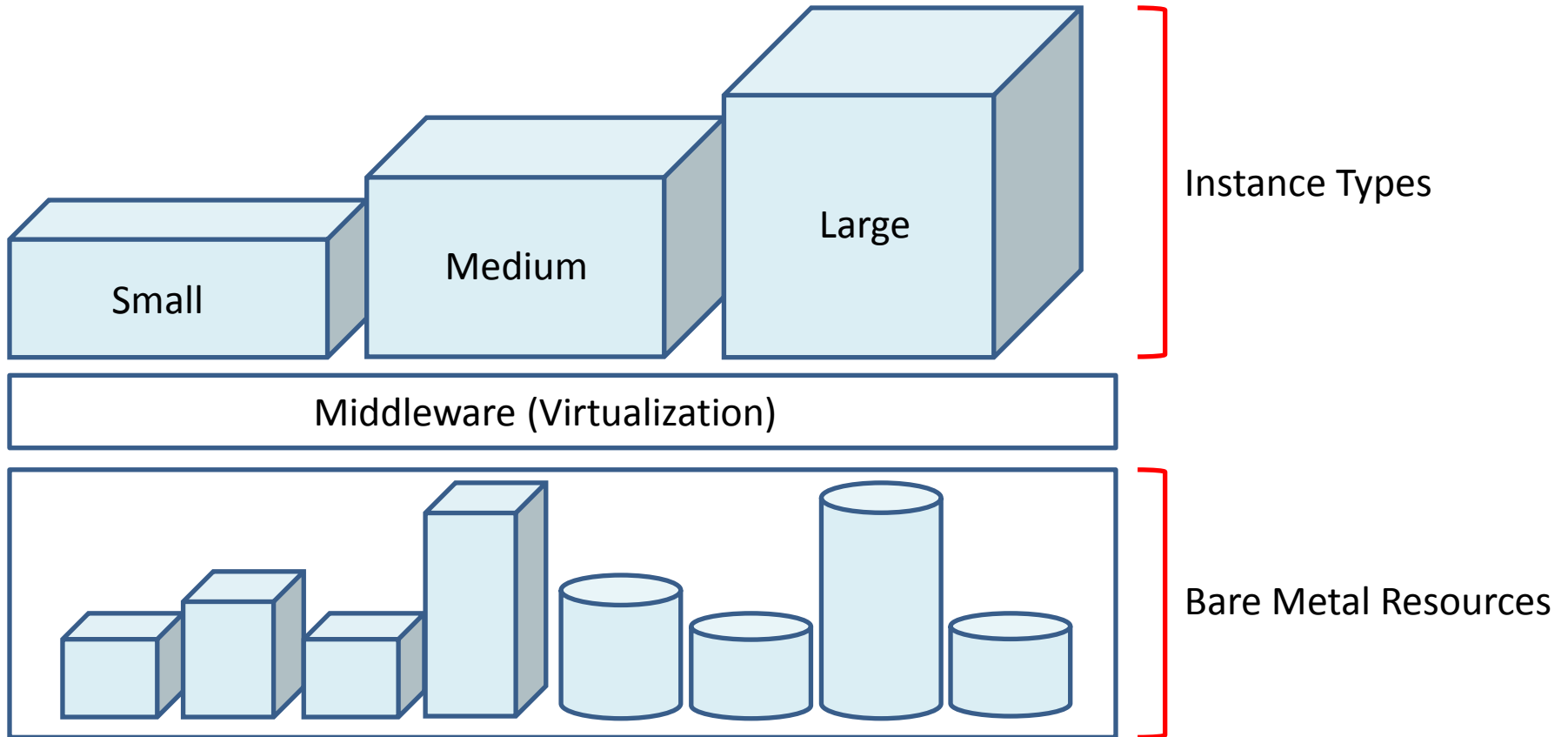


Reliability

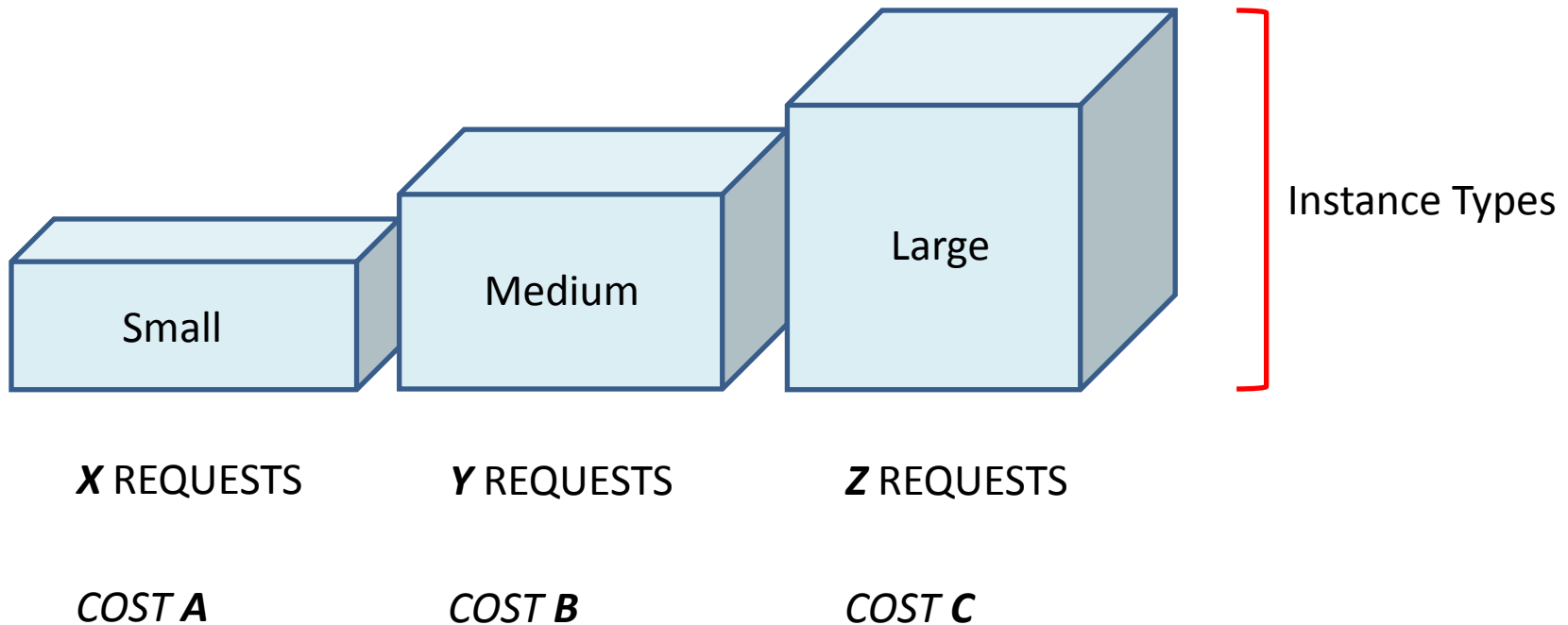
Current Infrastructure



Cloud Infrastructure

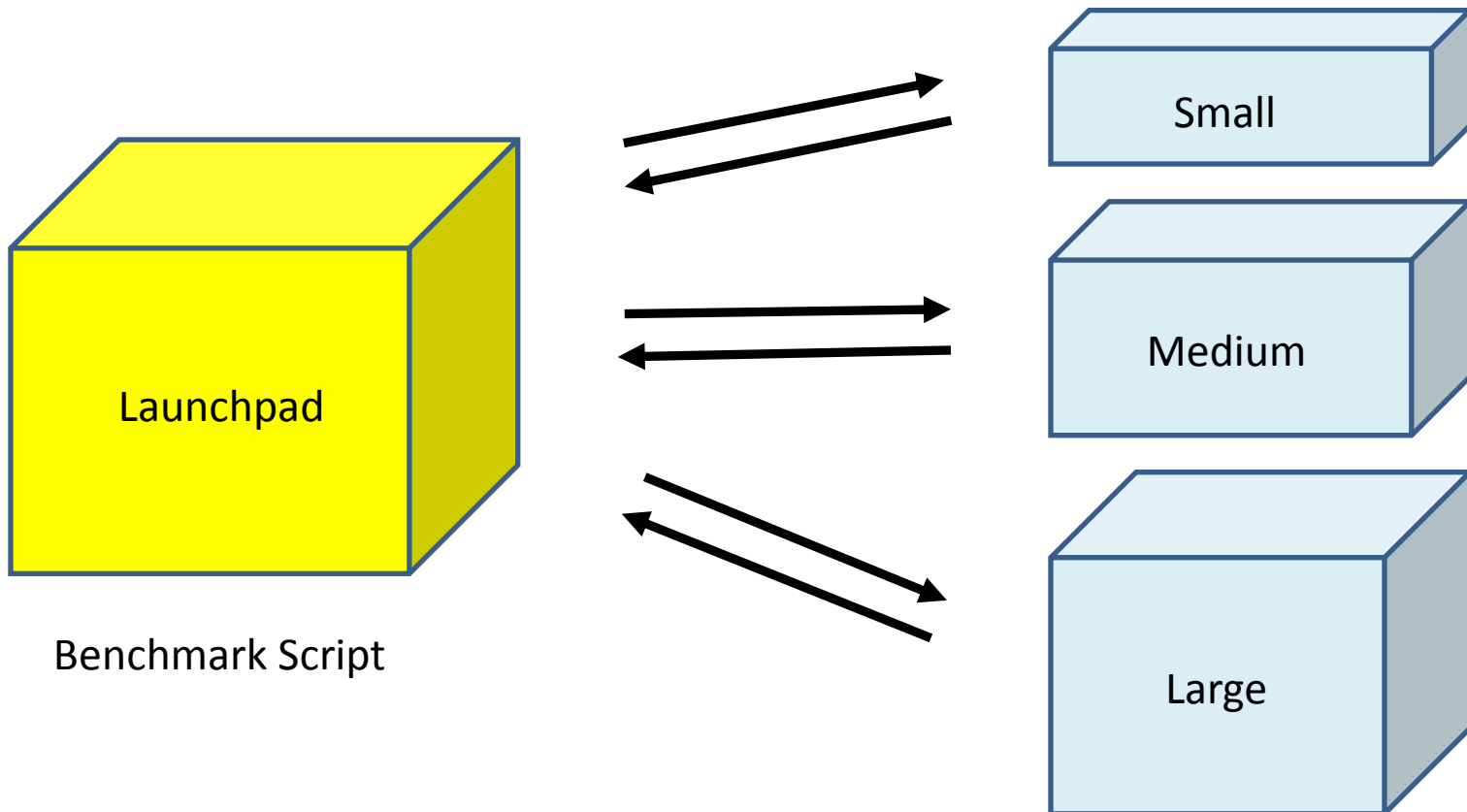


Maximize Requests per Dollar



GOAL: MAXIMIZE (REQUESTS/DOLLAR)

Load Testing Request & Response Flow



Provisioning Instances

- Manual Provisioning: Online Interface

Request Instances Wizard Cancel X

CHOOSE AN AMI **INSTANCE DETAILS** CREATE KEY PAIR CONFIGURE FIREWALL REVIEW

Provide the details for your instance(s). You may also decide whether you want to launch your instances as "on-demand" or "spot" instances.

Number of Instances: **Instance Type:**

Launch as an EBS-Optimized instance (additional charges apply): Not supported for this instance type

Launch Instances

EC2 Instances let you pay for compute capacity by the hour with no long term commitments. This transforms what are commonly large fixed costs into much smaller variable costs.

Launch into: EC2-Classical EC2-VPC

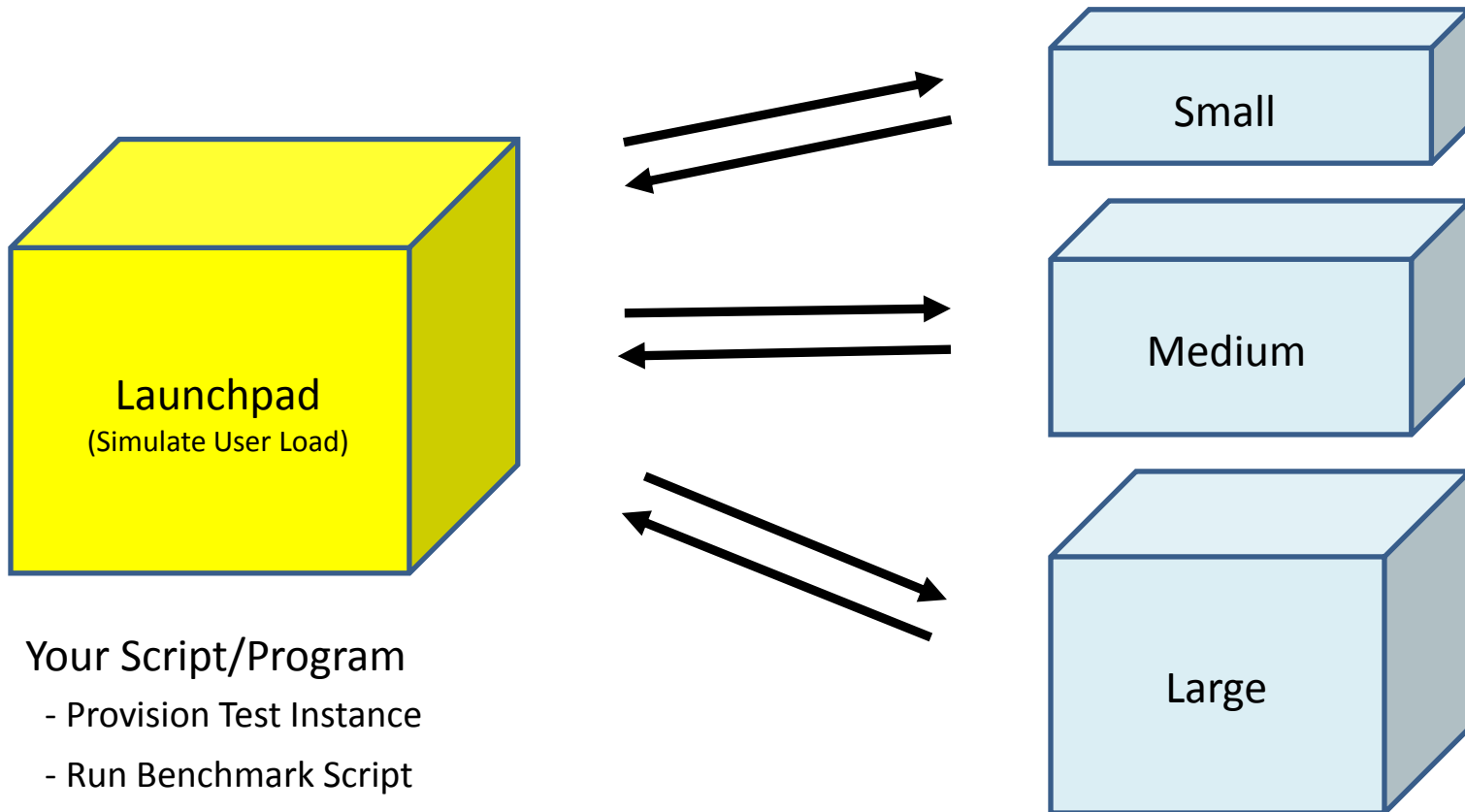
Availability Zone:

Request Spot Instances

Amazon APIs

- Provisioning & Monitoring:
 - EC2 API
 - CloudWatch API
- Supported APIs
 - Command Line Interface API Tools
 - AWS SDK for Java
 - AWS SDK for Python

Provisioning and Monitoring



Your Script/Program

- Provision Test Instance
- Run Benchmark Script
- CloudWatch to retrieve CPUUtilization
- Store Benchmark Output into a file
- Terminate Test Instance
- Repeat for Other Test Instances

Demo Outline

- 1. Launch an instance using command line
 - Set up command line
 - Launch an instance
 - CloudWatch command line
- 2. Launch an instance using Java API
 - Set up in eclipse
 - Amazon EC2 API
 - CloudWatch class

Setup ec2 command line tool

- Set JAVA_HOME environment variable
 - Slightly different in Mac OS if you would like to use the command line tool locally.
- Set EC2_HOME & system PATH variable
- Tell system your access key and secret key
- Set private key & certificate
- Verify
 - ec2-describe-regions
- http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/SettingUp_CommandLine.html

Launch an instance

- `ec2-run-instances ami_id [-n instance count] [-k keypair] [--instance-type instance_type] [--availability-zone zone]`
- <http://docs.aws.amazon.com/AWSEC2/latest/CommandLineReference/ApiReference-cmd-RunInstances.html>

Cloud Watch Command Line

- Download cloud watch
 - <http://aws.amazon.com/developertools/2534>
- Fill your access key and secret key in
\$AWS_CLOUDWATCH_HOME/credential-file-
path.template
- You will have to find out how to use this
command line
 - Hint: mon-get-stats

Set up AWS Toolkit for Eclipse

- Plug-in for the Eclipse Java IDE
- Help -> Install New Software
- Enter '<http://aws.amazon.com/eclipse>' in 'Work with'
- Select 'AWS Toolkit for Eclipse'
- <http://aws.amazon.com/eclipse/>

Amazon EC2 API

- Load AWS Credentials
 - BasicAWSCredentials
- Create Amazon EC2 Client
 - AmazonEC2Client
- Create and Configure Instance Request
 - RunInstancesRequest
- Launch Instance
 - RunInstanceResult

Amazon CloudWatch API

- Load AWS Credentials
 - BasicAWSCredentials
- Create Amazon CloudWatch Client
 - AmazonCloudWatchClient
- Create and Configure Metric Request
 - GetMetricStatisticsRequest
- Get Metric Result
 - GetMetricStatisticsResult
- <http://aws.amazon.com/sdkforjava/>

Discussion

- Your questions...

Upcoming Deadlines

- Unit 2: **Timed Quiz (120 minutes)**

UNIT 2: Data Centers

Module 3: Data Center Trends

Module 4: Data Center Components

Module 5: Design Considerations

Quiz 2: Data Centers

Checkpoint

Available 9/16/13 12:01 AM

Due 9/19/13 11:59 PM



- Project 2:

Project 2

Introduction and APIs

Single Instance Benchmarks

Checkpoint

Available Now

Due 9/22/13 11:59 PM

