15-319 / 15-619
Cloud Computing

Recitation 4

February 3rd & 5th, 2015

Overview

- **Administrative issues**
  OH hours, Piazza guidelines
- **Last week's reflection**
  Project 1.2, OLI unit 2 module 3 and module 4
- **This week's schedule**
  OLI Unit 2 module 5 out - Feb 2nd
  Quiz 2 - Feb 6th
  Project 2.1 - Feb 2nd
- **Demo**

Administrative

- TA office hours are posted on Piazza and Google calendar.
- Suggestions for using Piazza
  - Discussion forum, contribute questions and answers
  - Read the Piazza Post Guidelines (@20) before asking
  - Read Piazza questions & answers carefully to avoid duplicate ones
  - Don’t ask a public question about a quiz question
  - Try to ask a public question if possible
  - Encounter a grading bug, post privately on Piazza

For Piazza to be effective, everyone should contribute!!!

Code Submission Announcements

In your submissions:

Do include:

- readme.txt
- References in a file called ‘references’

Do not include:

- Packages (jar files)
- Folder structures
- Input or output data
- Pem files

Last Week's Reflection

- Until now you have completed
  - Sequential Analysis
  - Elastic MapReduce
- You should have learned
  - Why MapReduce for big data
  - How MapReduce works
  - How to program a Mapper and Reducer
  - Performance/cost tradeoff
  - How to narrow down bugs by using logs

Project 1.2 Checkpoint

- We will manually grade Question 1 in P1.2
  - Always make sure that your code is readable
    - Follow style presented in Recitation 2

Project Grading Penalties

The following table outlines the violations of the project rules and their corresponding grade penalties for this project.

Violation	Penalty of the project grade
Failing to tag your all your instances (including all of the instances in the EMR cluster) for this project	-10%
Using any instance type in your cluster that has an on-demand pricing higher than m1.large	-10%
Attempting to hack/tamper the autograder in any way	-100%
Using more than **$15** to complete this project	-10%
Using more than **$30** to complete this project	-100%

Project 1.2 Questions - 1

- Late policy
  - We do not have a late policy!
    - All deadlines are hard. No exceptions.
    - No excuses, and no special cases are allowed.
  - Please **start early**!
    - We are working with a public cloud infrastructure, things are bound to take more time or break. One of the reasons why this experience is so valuable.
    - You get more TA support during the week.

Project 1.2 Questions - 2

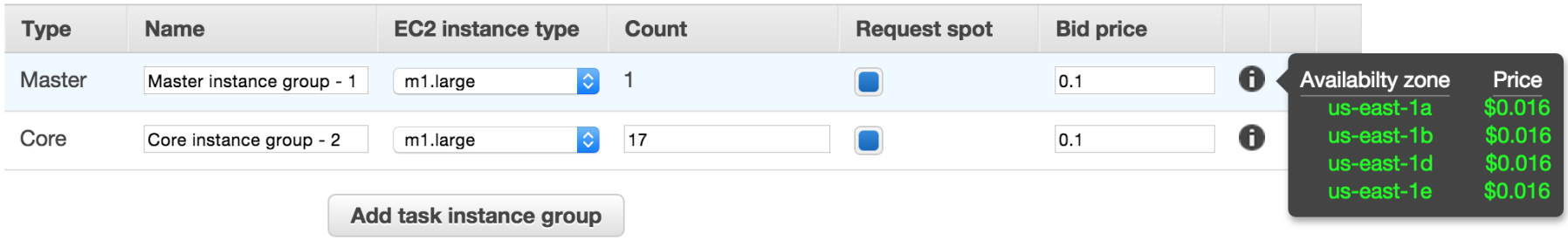
Elastic MapReduce Billing Question

- EC2 cost + EMR cost
- Elastic MapReduce Pricing (On demand)
- for US East (N. Virginia)

Region: US East (N. Virginia)

	Amazon EC2 Price	Amazon Elastic MapReduce Price
General Purpose - Current Generation		
m3.xlarge	$0.280 per Hour	$0.070 per Hour
m3.2xlarge	$0.560 per Hour	$0.140 per Hour
General Purpose - Previous Generation		
m1.small	$0.044 per Hour	$0.011 per Hour
m1.medium	$0.087 per Hour	$0.022 per Hour
m1.large	$0.175 per Hour	$0.044 per Hour
m1.xlarge	$0.350 per Hour	$0.088 per Hour

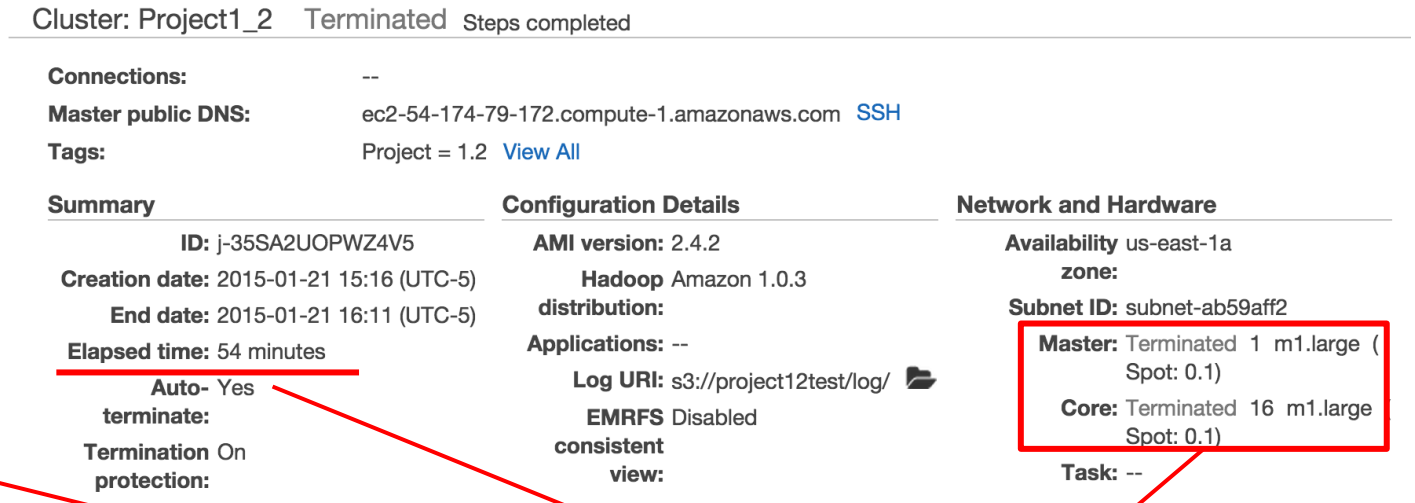
Project 1.2 Questions - 3



Type	Name	EC2 instance type	Count	Request spot	Bid price
Master	Master instance group - 1	m1.large	1	☑	0.1
Core	Core instance group - 2	m1.large	17	☑	0.1

Add task instance group

Availabilty zone	Price
us-east-1a	$0.016
us-east-1b	$0.016
us-east-1d	$0.016
us-east-1e	$0.016



Cluster: Project1_2 Terminated Steps completed

Connections: --
Master public DNS: ec2-54-174-79-172.compute-1.amazonaws.com SSH
Tags: Project = 1.2 View All

Summary
ID: j-35SA2UOPWZ4V5
Creation date: 2015-01-21 15:16 (UTC-5)
End date: 2015-01-21 16:11 (UTC-5)
Elapsed time: 54 minutes
Auto-terminate: Yes
Termination protection: On

Configuration Details
AMI version: 2.4.2
Hadoop distribution: Amazon 1.0.3
Applications: --
Log URI: s3://project12test/log/
EMRFS consistent view: Disabled

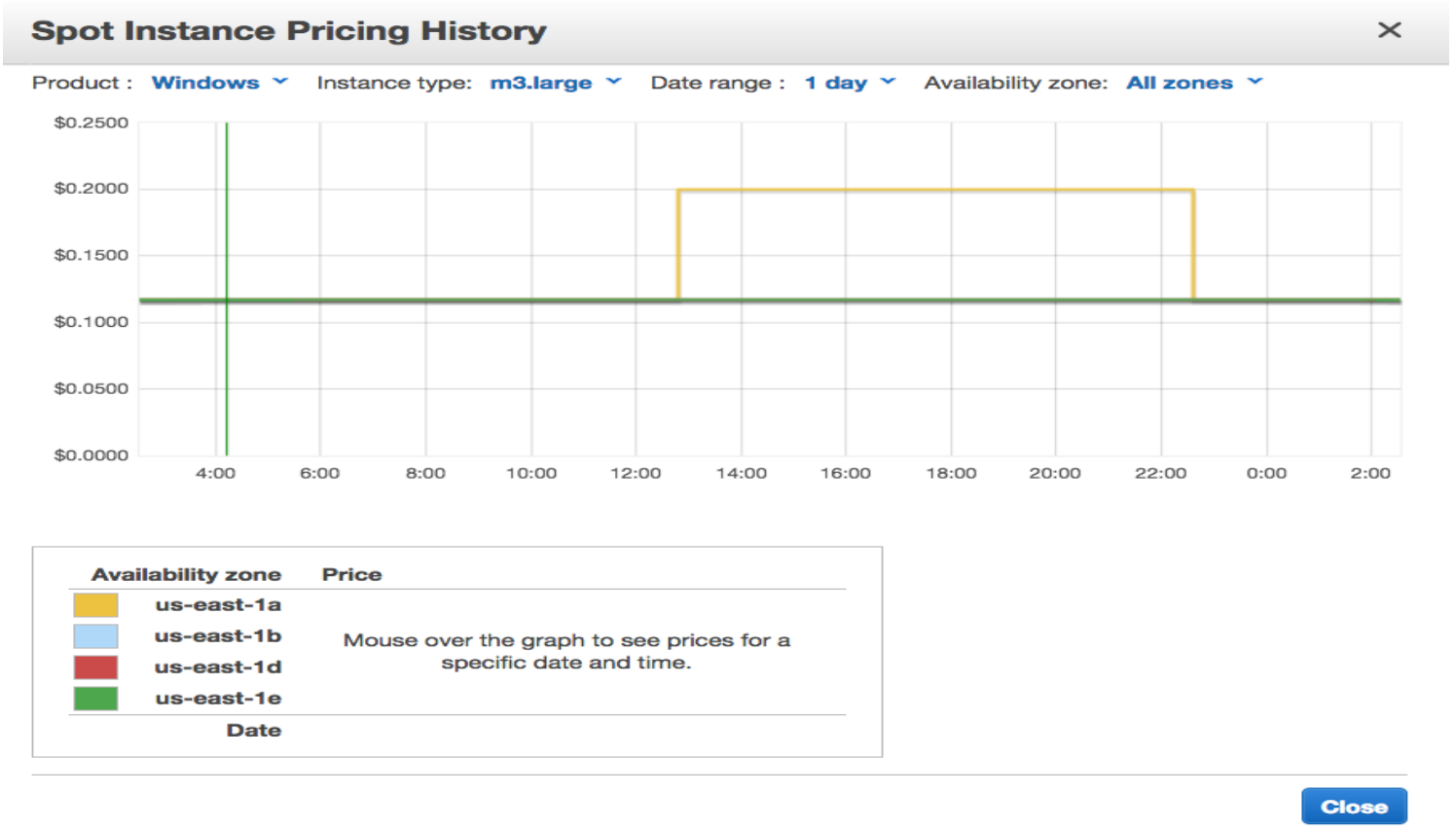
Network and Hardware
Availability zone: us-east-1a
Subnet ID: subnet-ab59aff2
Master: Terminated 1 m1.large (Spot: 0.1)
Core: Terminated 16 m1.large Spot: 0.1)
Task: --

EMR cost

On demand = ($0.175 + $0.044) * 1 hour * 17 instances = $3.723

Spot = ($0.016 + $0.044) * 1 hour * 17 instances = $1.02

Project 1.2 Questions - 4



Please remember to double check the tag if you use spot instances.

What happens this week

- Unit 2: Data Centers
  - Module 3: Data Center Trends
  - Module 4: Data Center Components

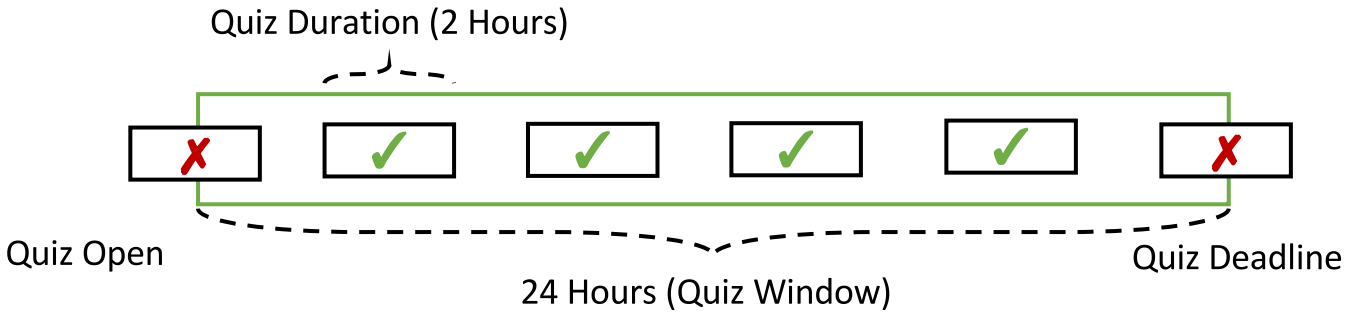
- Read and complete:
  - Module 5: Design Considerations
  - Unit 2: Checkpoint Quiz
    - 150 minutes, due February 6th, 2015, Pittsburgh

Module 5: Design Considerations

- Design considerations
  - Requirements and Geographic Location
  - Costs
  - Power and Efficiency
  - Redundancy
- Challenges and Requirements
  - Scalability
  - Network Topologies
  - Utilizations & Resiliency
- Software Defined Data Centers and Networking
- Software Define Storage

Amazon data center

Quiz 2



- Quiz 2 will be open for 24 hours, Friday, Feb 6
  - Quiz 2 becomes available on **Feb 6, 00:01 AM ET.**
  - Deadline for submission is **Feb 6, 11:59 PM ET.**
  - Once open, you have **150 min** to complete the quiz.
  - Late submissions are NOT accepted.
  - You may not start the quiz after the deadline has passed.
  - **Maintain your own timer from when you start the quiz.**
  - **Click submit before deadline passes. No Exceptions!**

Location	Silicon Valley	Pittsburgh	Rwanda	Adelaide
Open	Feb 5 09:01 PM	Feb 6, 00:01 AM	Feb 6 07:01 AM	Feb 6 03:31 PM
Deadline	Feb 6 , 08:59 PM	Feb 6, 11:59 AM	Feb 6 06:59 AM	Feb 7 03:29 PM

Pinterest

http://www.pinterest.com/

- Started with 12 engineers, now 300+ employees
- Monthly active users: 53 Million
- Growth in web traffic from 9/12-9/13: 66.52%

From Unit 1, we know that cloud computing provides several advantages, including:

- Elasticity
- Reduced upfront cost
- Reduced maintenance cost

Pinterest

- Utilizes AWS
  - S3
    - File storage
  - Elastic MapReduce
    - Data analysis
  - Auto Scaling
    - Scale up and down
  - Elastic Load Balancer
    - Distribute traffic

Highlights of this Week's Project

Objectives:

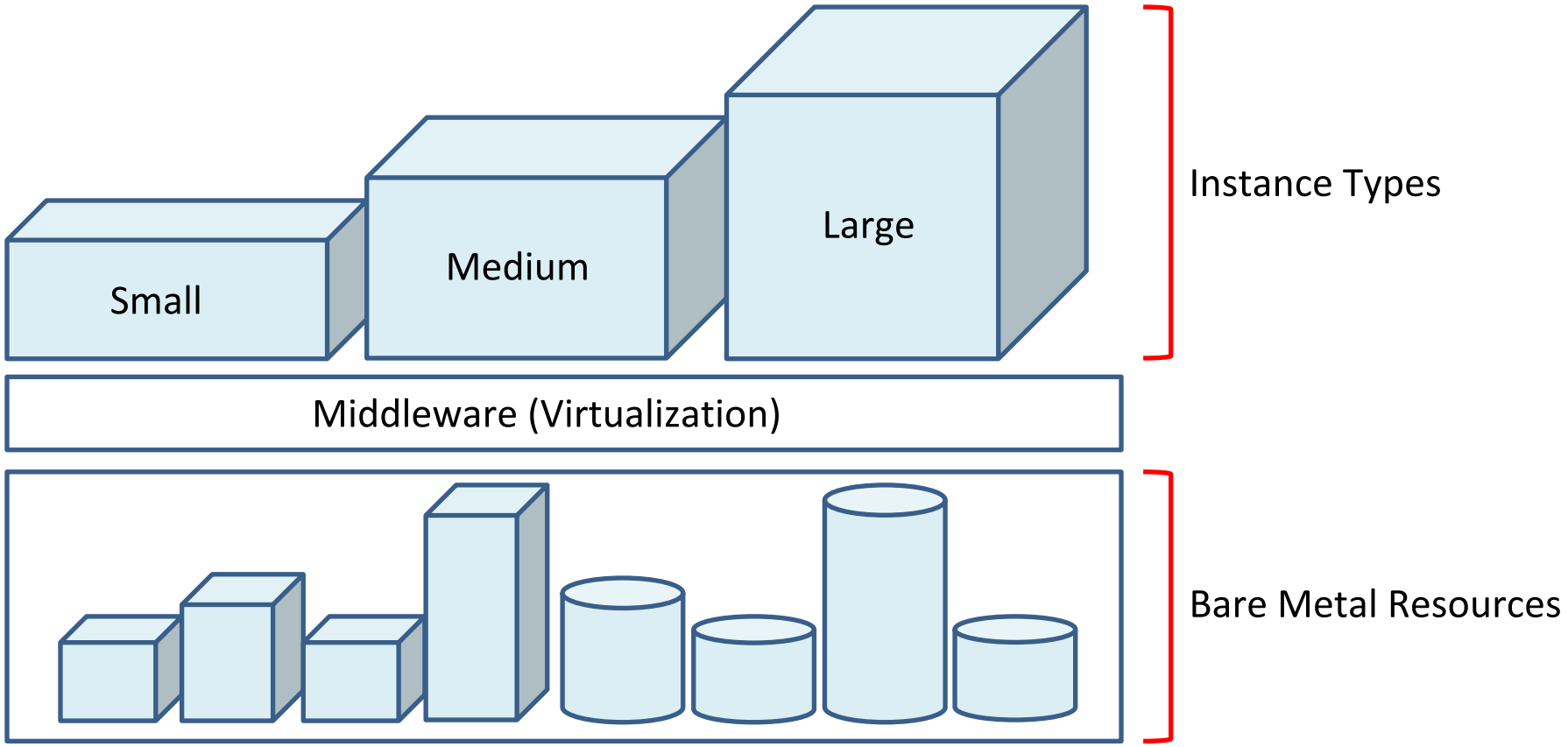
- What is scaling?

  Horizontal vs Vertical
- Understand how to use AWS APIs and use them to automate the scaling out of servers to handle network traffic.
- Learn how to distribute load uniformly among your servers using Elastic Load Balancing.

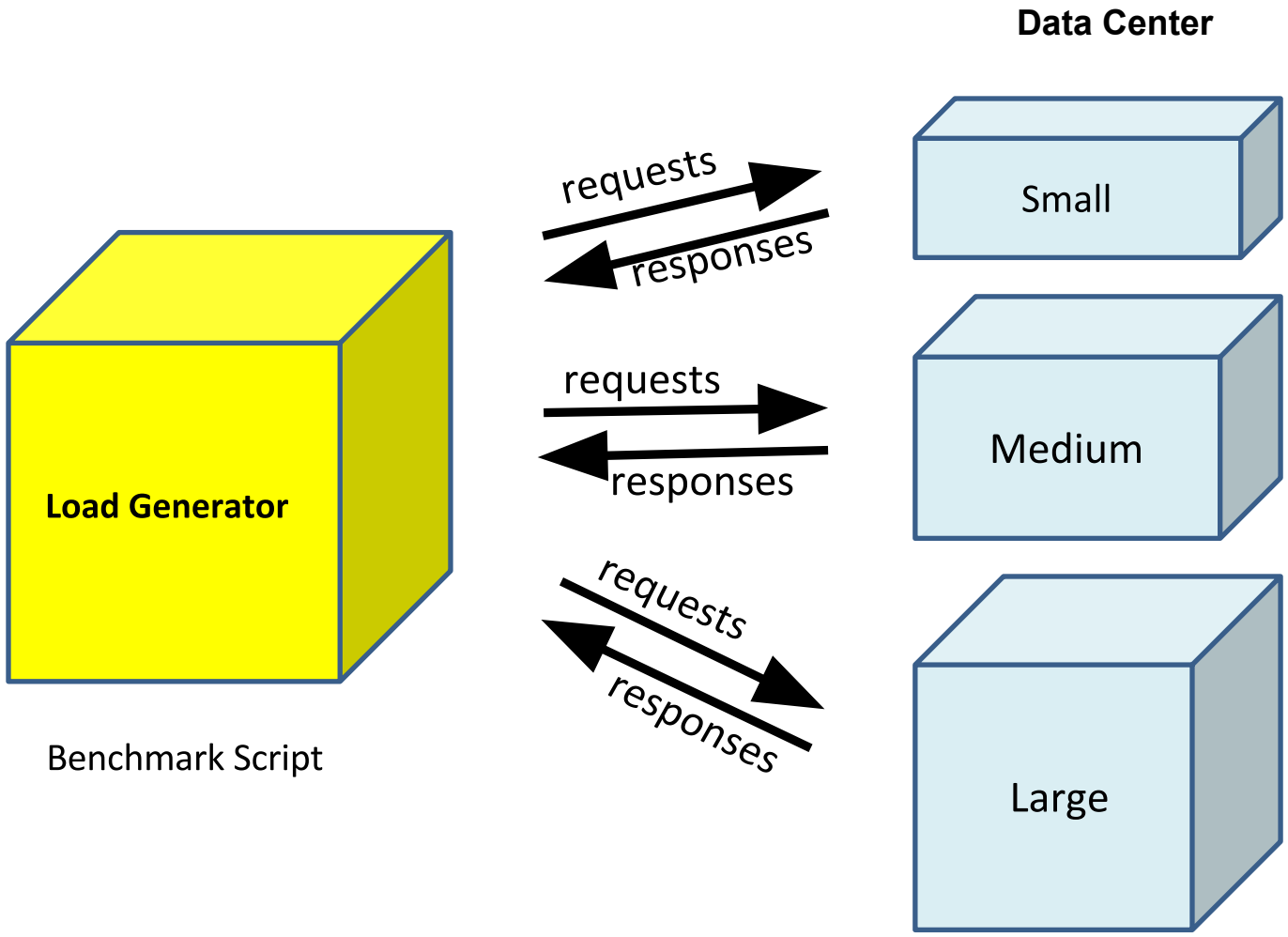
Deadline:

- Project 2.1, due on Feb 8, 2015 at 11:59 pm ET

Resources in Cloud Infrastructure

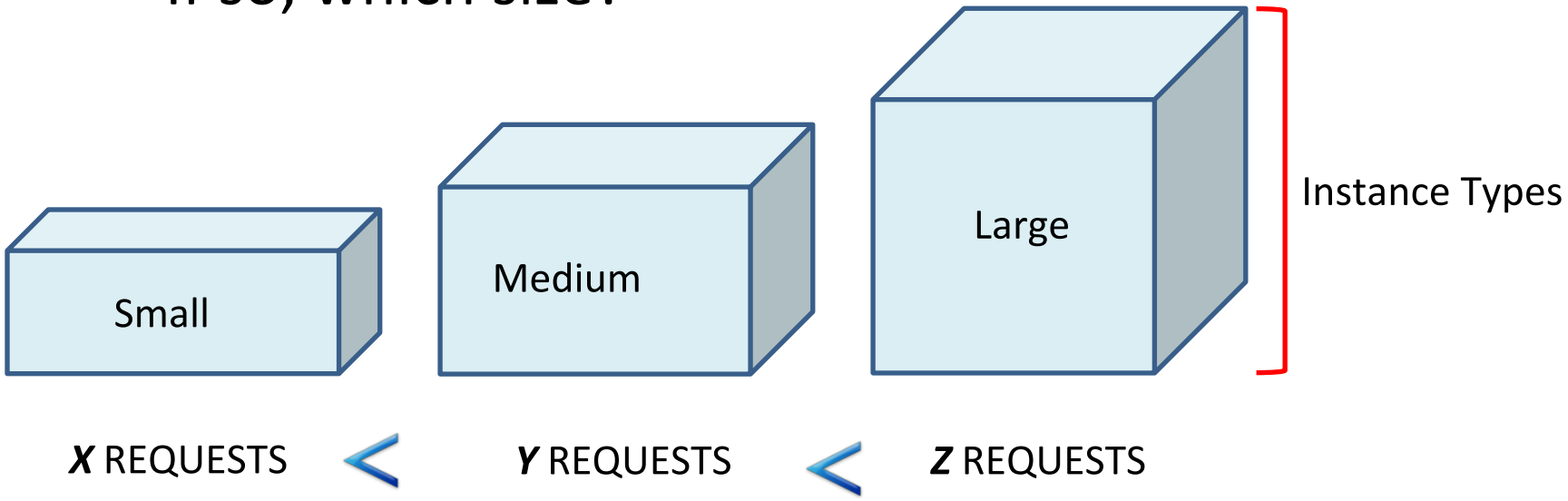


Load Testing Request & Response Flow



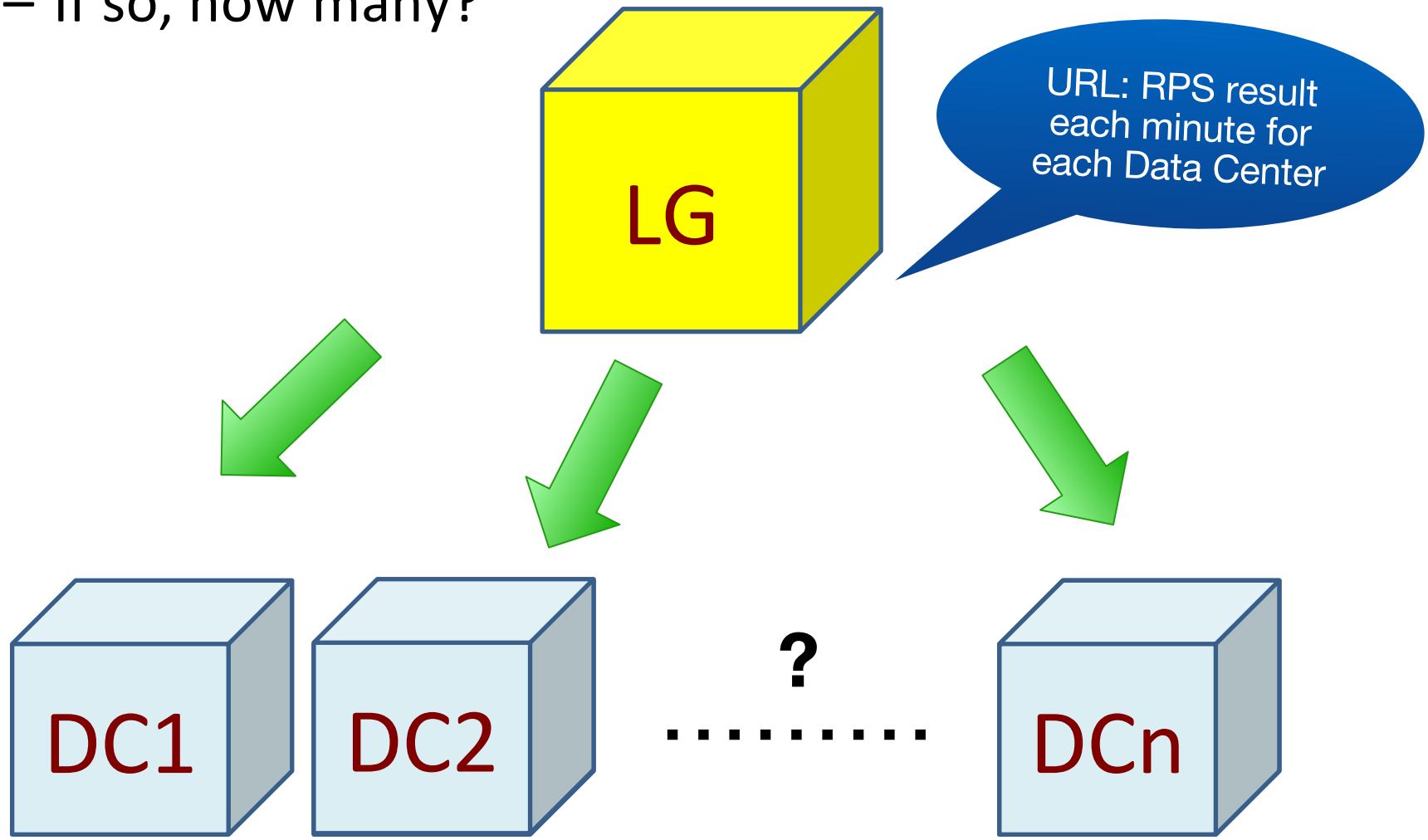
Vertical Scaling

- Would a larger instance suffice?
  - If so, which size?



Horizontal Scaling

- Would more instances suffice?
  - If so, how many?



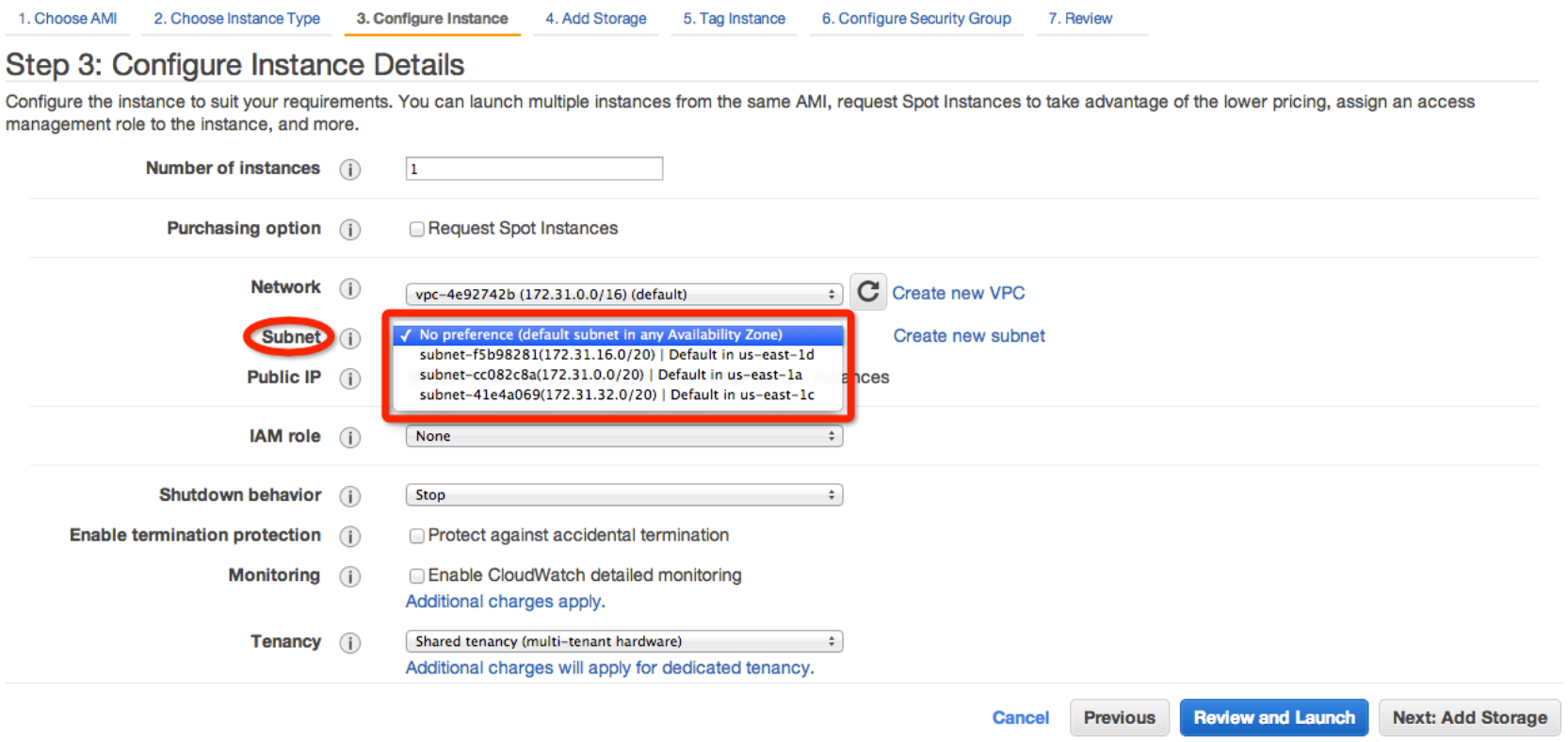
Amazon APIs

Supported APIs

- Command Line Interface API Tools
- AWS SDK for Java
- AWS SDK for Python

Reminders for Project 2.1

- Make sure the Load Generator and Data Center VMs are in the same subnet (availability zone)



Reminders for Project 2.1 (cont'd)

- Terminate instances vs. Stop instances
  - Stop will still charge for VM storage (EBS volumes)
    - Stop is a good idea when you need a break
- DO NOT add your credentials in your submitted code

Penalties for 2.1

Project Grading Penalties

Besides the penalties mentioned in recitation and/or on Piazza, penalties accrue for the following:

Violation	Penalty of the project grade
Spending more than $8 for this project phase	-10%
Spending more than $16 for this project phase	-100%
Failing to tag all your instances for this project	-10%
Submitting your AWS credentials in your code for grading	-10%

Upcoming Deadlines

- Quiz 2: Data Centers
  - Due Friday 02/06/15 11:59 PM ET

- Project 2.1: Introduction to AWS APIs
  - Due Sunday 02/08/15 11:59 PM ET