15-319 / 15-619
Cloud Computing

Recitation 1

Course Overview and Introduction

January 14 and 16, 2020

http://www.cs.cmu.edu/~msakr/15619-s20/

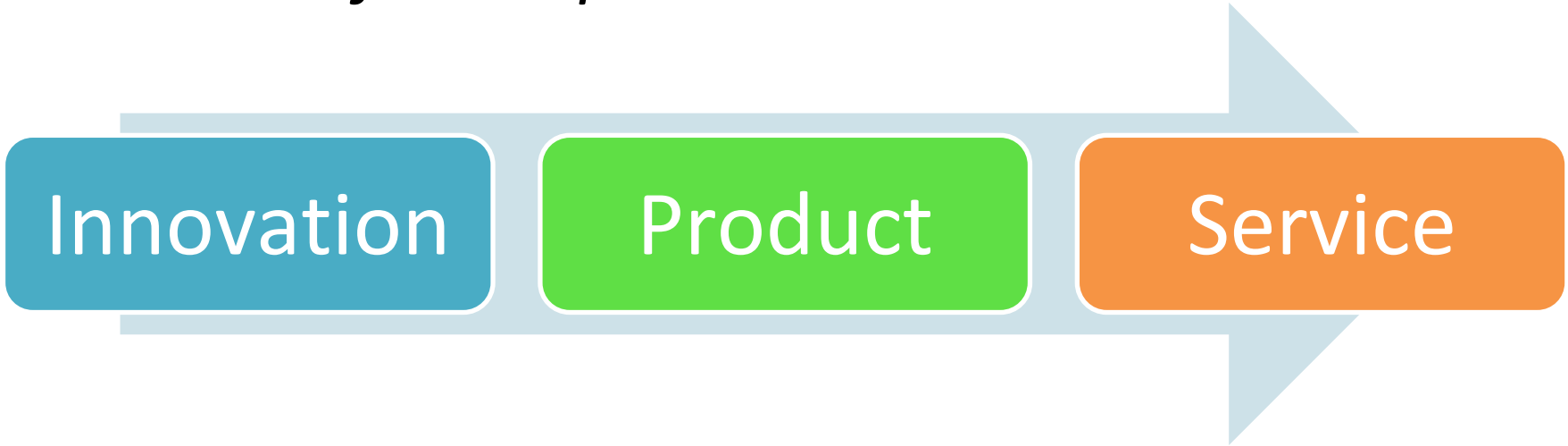
Outline

- **What is the course about?**
- What is an online course?
- Administrivia

So What is Cloud Computing?

Evolution of Computing

Cloud Computing is the transformation of“
”IT from a product to a service



Evolution of Electricity

Innovation

New Disruptive Technology

Product

Buy and Maintain the Technology

Service

Electric Grid, pay for what you use

A Cloud is …

- Datacenter hardware and software that the vendors use to offer the computing resources and services

Cloud-enabling Data Centers

- Large warehouse scale data centers
- Growing at a rapid rate
- Next is an example from Microsoft Azure
  - Azure US-East2 (Boydton, VA)
  - Azure Expansion 1
  - Azure Expansion 2

Microsoft Azure

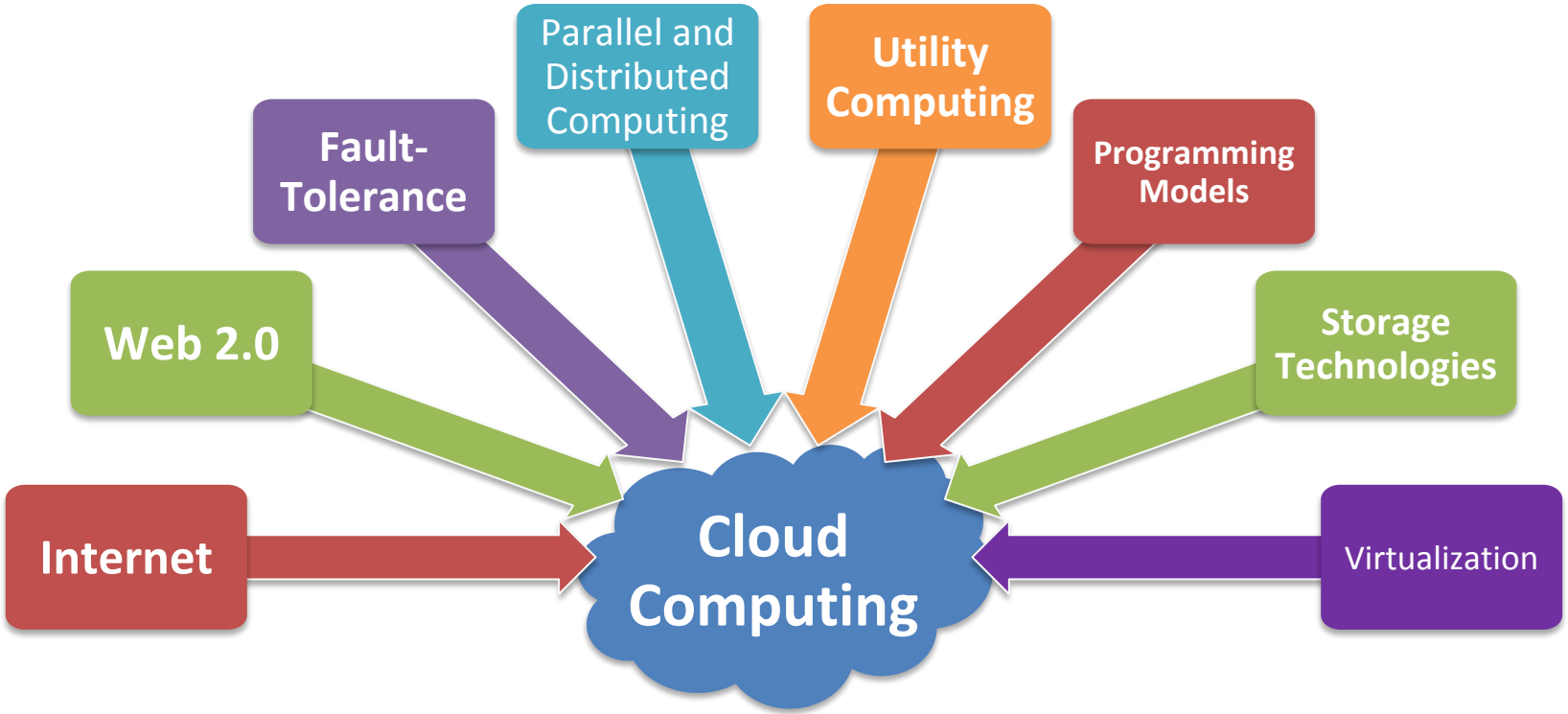

US-East2 (Boydton, VA)


US-East2 (Boydton, VA)


US-East2 Hub Expansion

The Cloud

Enabled by Maturing Technologies



So… how would you transform information technology into a Service?

How to Transform IT to a Service?

- Connectivity
  - For moving data around
- Interactivity
  - Seamless interfaces
- Reliability
  - Failure will affect many
- Performance
  - Should not be slower
- Pay-as-you-Go
  - No upfront fee
- Ease of Programmability
  - Ease of development of complex services
- Manage Big Data
- Efficiency
  - Cost
  - Power
- Scalability & Elasticity
  - Flexible and rapid response to changing user needs

How to Transform IT to a Service?

Internet

Web 2.0

Fault-Tolerance

Parallel / Distributed Systems

Utility Computing

Programming Model

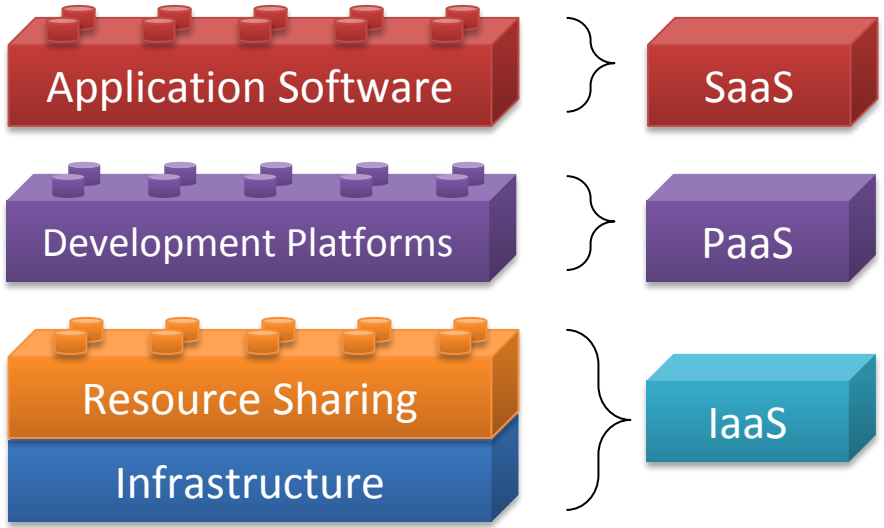
Storage Technologies

Virtualization and Resource Sharing Technologies

Cloud Building Blocks

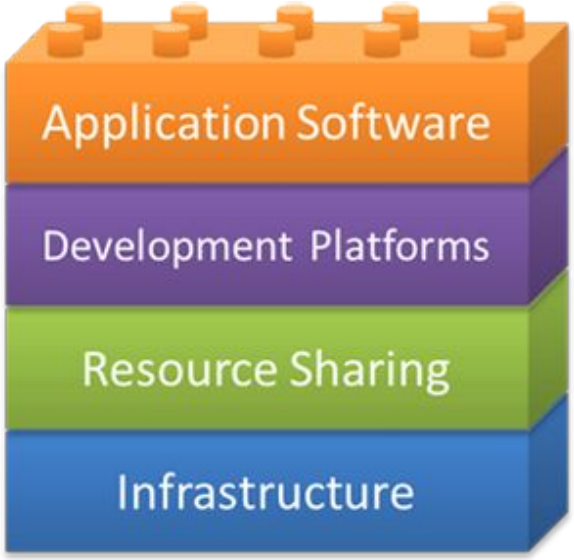
Cloud services are available in various forms, corresponding to the layer of abstraction desired by the user

- Software as a Service (**SaaS**)
- Platform as a Service (**PaaS**)
- Infrastructure as a Service (**IaaS**)



Cloud Computing Stack

- Applications
- Development Platforms
- Elasticity
  - APIs to enable automation,
    Alarms, protocols, triggers, etc…
- Sharing mechanisms
  - Virtualization, Containers, …
- Distributed systems
  - Programming models
  - Storage
- Data centers



What is this course about?

- Applied aspects of cloud computing
  - Between systems and services

Conceptual content on OLI

1.0 Introduction to Cloud Computing
Service and deployment models, economics and use cases

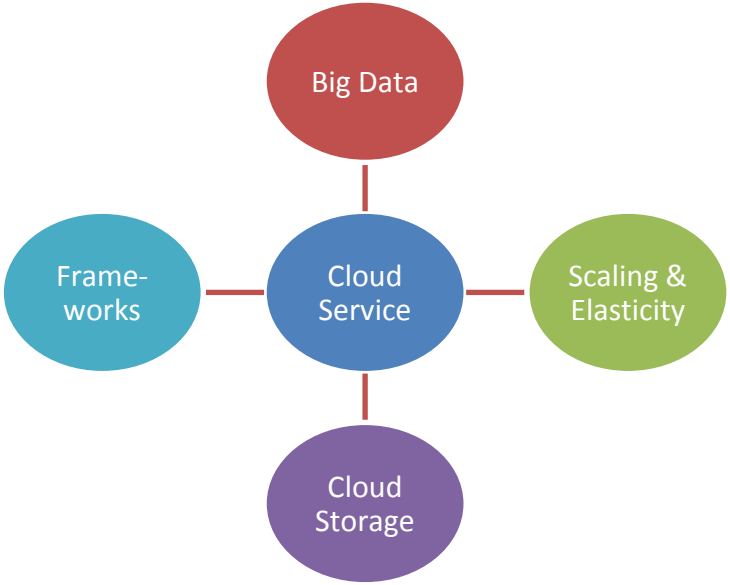
2.0 Cloud Infrastructure
Components, design considerations and power

3.0 Resource Sharing
CPU, memory and I/O Virtualization, SDN, SDS

4.0 Cloud Storage
Distributed File Systems and Distributed Databases

5.0 Programming Models
MapReduce, Spark, GraphLab, Kafka/Samza

Projects on AWS, Azure, & GCP



Course Goals

Students gain hands-on experience solving real world problems by completing projects in the areas of cloud **analytics**, **compute and elasticity**, **storage** and **frameworks**, which utilize existing public cloud tools and services. Students are exposed to real-world data, scenarios, infrastructure and budgets in order to learn how to:

1. Design, architect, implement, test, deploy, monitor and maintain cloud-based applications;
2. Identify the appropriate tools and architectures to implement a cloud-based design;
3. Analyze the tradeoffs between different tools and cloud offerings to meet real-world constraints;
4. Evaluate performance characteristics of cloud-based services to implement optimizations;
5. [15-619 only] Collaborate with a team on an open-ended project to incrementally realize an optimized end-to-end cloud-based solution.

Conceptual Content on OLI

Unit #	Title	Modules and Description
1	Introduction	Definition and evolution of Cloud Computing Enabling Technologies Service and Deployment Models Popular Cloud Stacks and Use Cases Benefits, Risks, and Challenges of Cloud Computing Economic Models and SLAs Topics in Cloud Security
2	Cloud Infrastructures	Historical Perspective of Data Centers Datacenter Components: IT Equipment and Facilities Design Considerations: Requirements, Power, Efficiency, & Redundancy Power Calculations and PUE Challenges in Cloud Data Centers Cloud Management and Software Deployment Considerations
3	Virtualization	Virtualization: CPU, Memory, I/O) Case Study: Amazon EC2 Software Defined Networks (SDN) Software Defined Storage (SDS)
4	Cloud Storage	Introduction to Storage Systems Cloud Storage Concepts Distributed File Systems (HDFS, Ceph FS) Cloud Databases (HBase, MongoDB, Cassandra, DynamoDB) Cloud Object Storage (Amazon S3, OpenStack Swift, Ceph)
6	Programming Models	Distributed Programming for the Cloud Data-Parallel Analytics with Hadoop MapReduce (YARN) Iterative Data-Parallel Analytics with Apache Spark & Graph-Parallel with GraphLab Stream Processing with Apache Kafka and Samza

Read the Syllabus

Quiz 1, Jan24, 2020

Projects on AWS/Azure/GCP Clouds

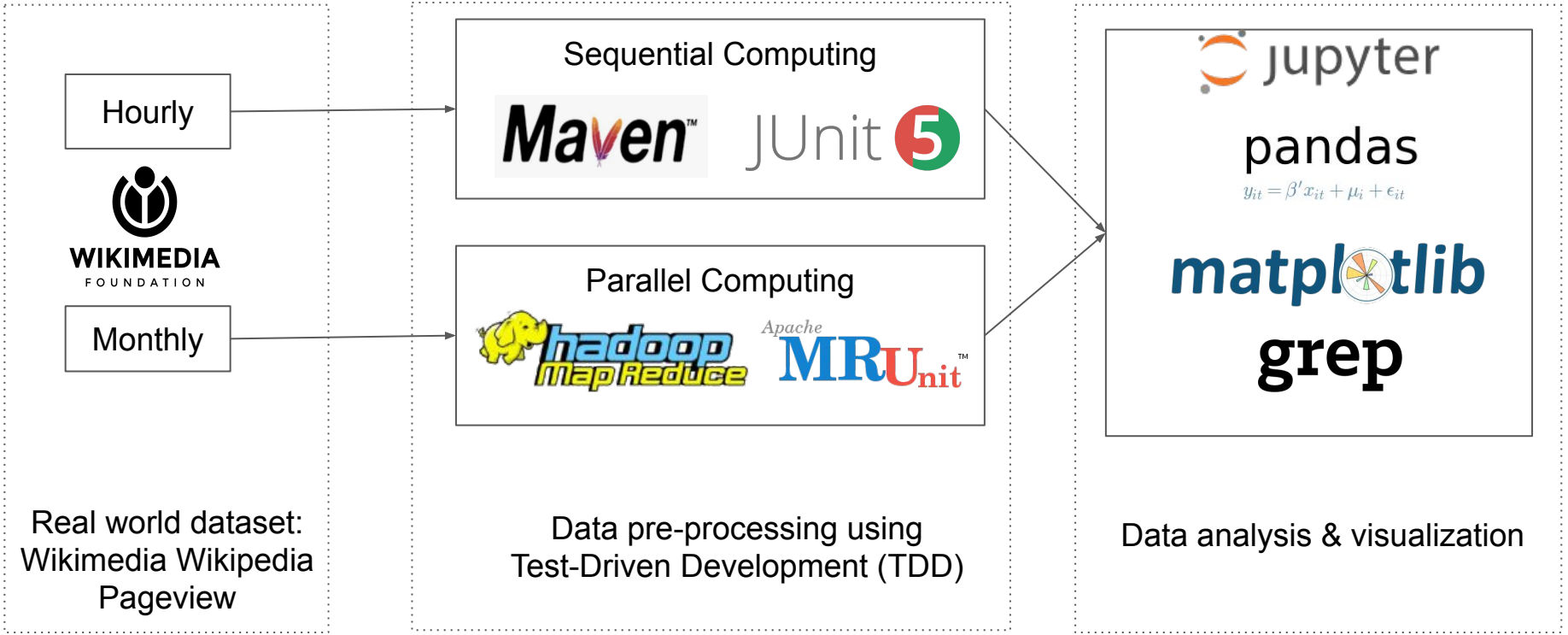
Due 01/19/2020

0. AWS/Azure/GCP Account Setup & Tool Primers	• Benchmarking VMs, SSH, Authentication, Billing, Security Groups, Vertical Scaling
1. Big Data Analytics	• Amazon EC2, Amazon EMR
2. Scaling, Elasticity and Failure	• Auto Scaling, Load Balancing, Monitoring, Docker Containers, Kubernetes, Functions
3. Cloud Storage	• MySQL, HBase, GCP Bigtable, GCP SQL DB, MongoDB
4. Analytics Engines for the Cloud	• Spark, DataBricks, Cloud ML Frameworks, Kafka/Samza
5. A Complete Web Service (**Team project**)	• No restrictions on tools or AWS services • Evaluated based on cost and performance

Project Learning Objectives

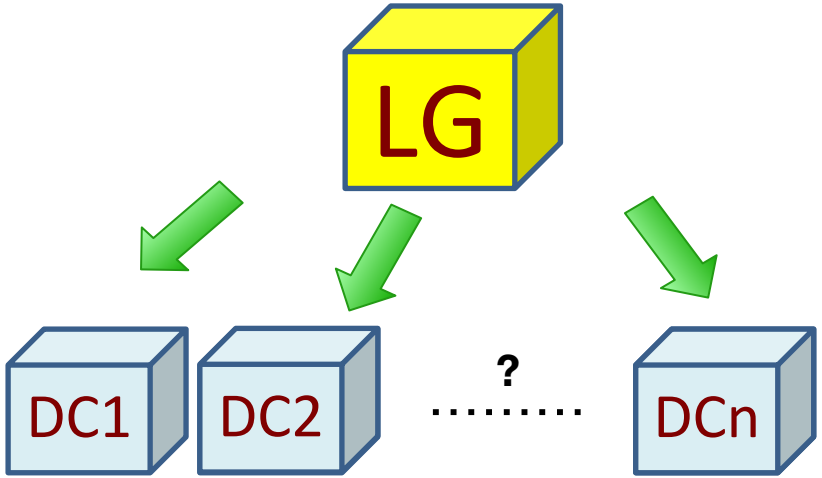
Data Analytics	1) Analyze and visualize small and large data sets on the cloud using interactive computing as well as data manipulation and analysis libraries.
Compute & Elasticity	2) Design, implement, test, package, deploy and monitor cloud applications using Virtual Machines (VMs), Containers and Serverless cloud computing services.
Cloud Storage	3) Explore and experiment with different distributed cloud-storage abstractions and compare their features, capabilities and approaches. 4) Orchestrate, deploy and optimize a unified application that integrates heterogeneous SQL and NoSQL database systems. 5) Implement and compare consistency models to recognize the tradeoff between consistency and performance in replicated and distributed cloud storage systems.
Frameworks	6) Design, implement, test and debug applications using interactive, batch and stream processing frameworks and compare their suitability to different problem domains. 7) Illustrate and explain the execution workflow, overhead, fault-tolerance and logical flow of interactive, batch and stream processing frameworks. 8) Train and deploy a machine learning model using a cloud-based framework. 9) Analyze and identify potential sources of bottlenecks in programming frameworks to optimize their performance.
Team Project	10) [15-619 only] Design, build, and deploy a performant, reliable, scalable and fault-tolerant web service on the cloud within a specified budget. 11) [15-619 only] Perform extract, transform and load (ETL) on a large data set. 12) [15-619 only] Design schema as well as configure and optimize cloud-based databases to deal with scale and improve the throughput of a web service. 13) [15-619 only] Explore methods to identify the potential bottlenecks in a cloud-based web service and implement methods to improve system performance.
Overall	14) Practice gathering, cleaning and preparing data for analysis on the cloud. 15) Practice Test-driven Development (TDD) in the software development process. 16) Orchestrate and automate the process of managing and provisioning cloud resources through machine-readable definition files. 17) Make informed decisions about choosing an appropriate cloud tool that will satisfy a set of specified requirements.

P1.1 and P1.2: Big Data Analytics



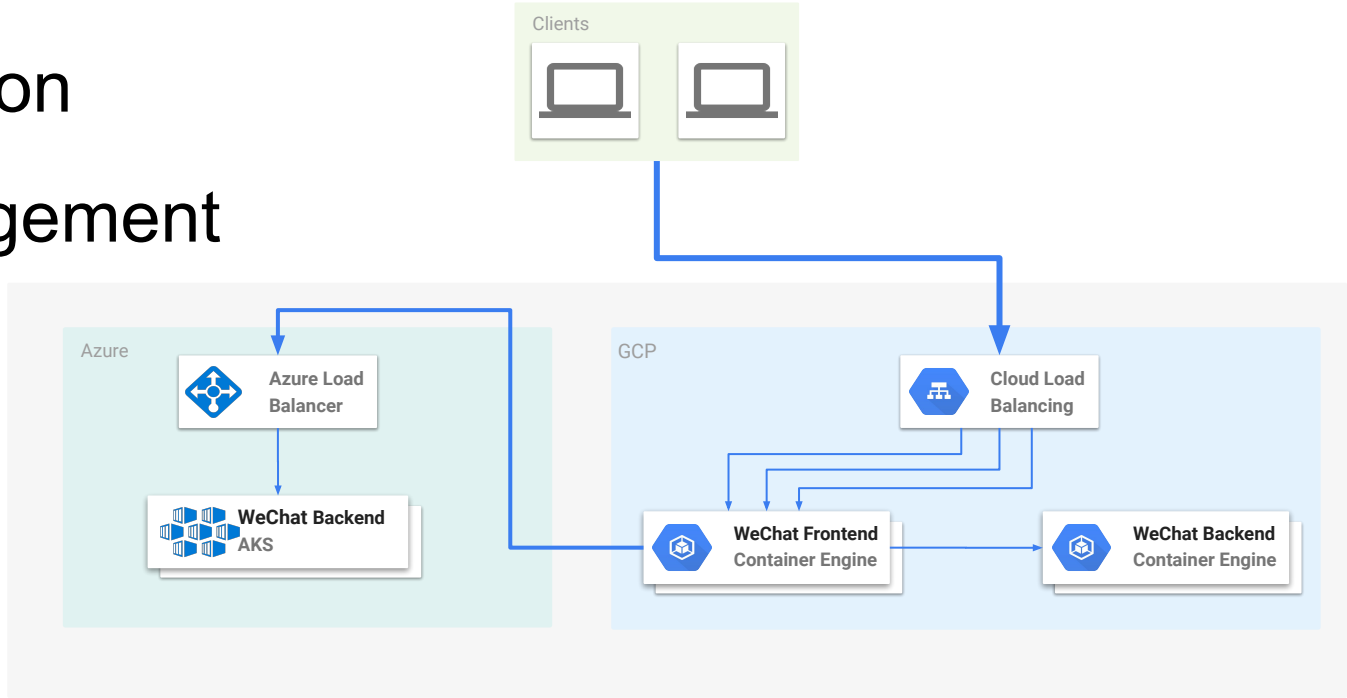
P2.1: VM Scaling, Elasticity & Failure

- Closed VMs
  - Load Generators
    - Dynamic load
  - Data Center Instances
    - Performance & failure
- Scale out & scale in DCIs to achieve desired RPS within budget
  - Auto-scale groups, elastic load balancers, monitoring, etc.



P2.2: Containers and Kubernetes

- Containerization
- Cluster Management
- Multi-Cloud Deployment

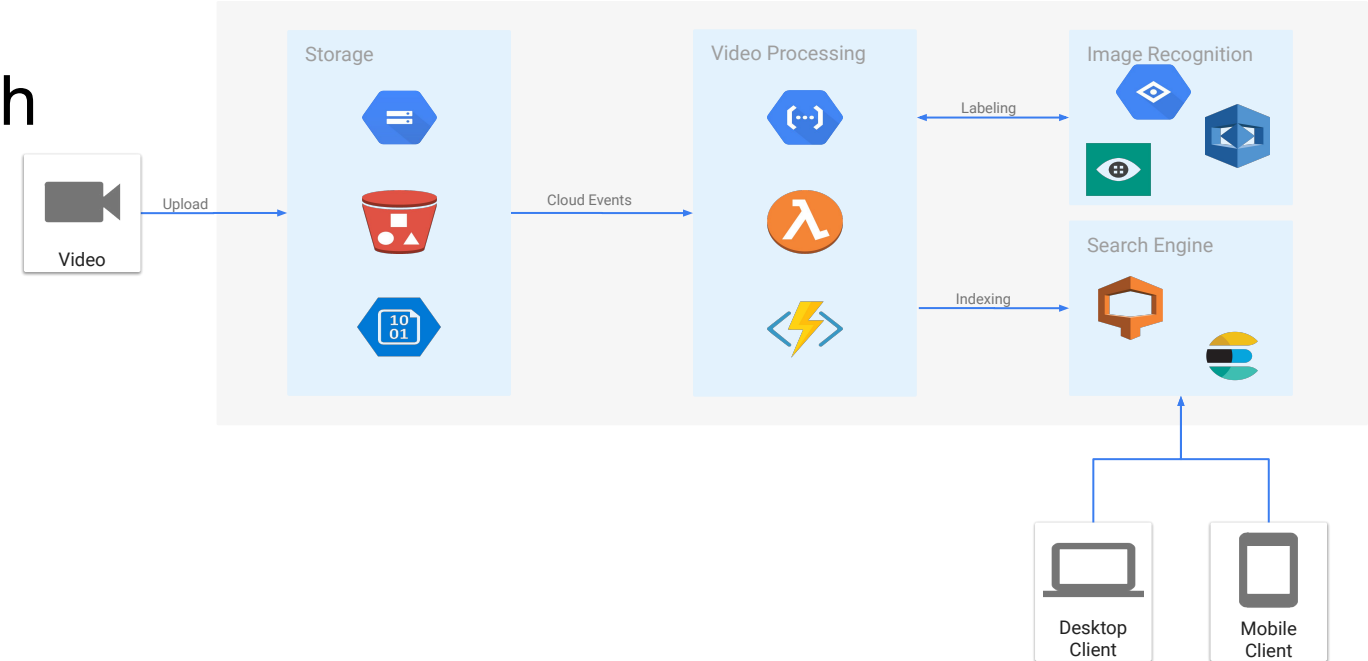


Azure

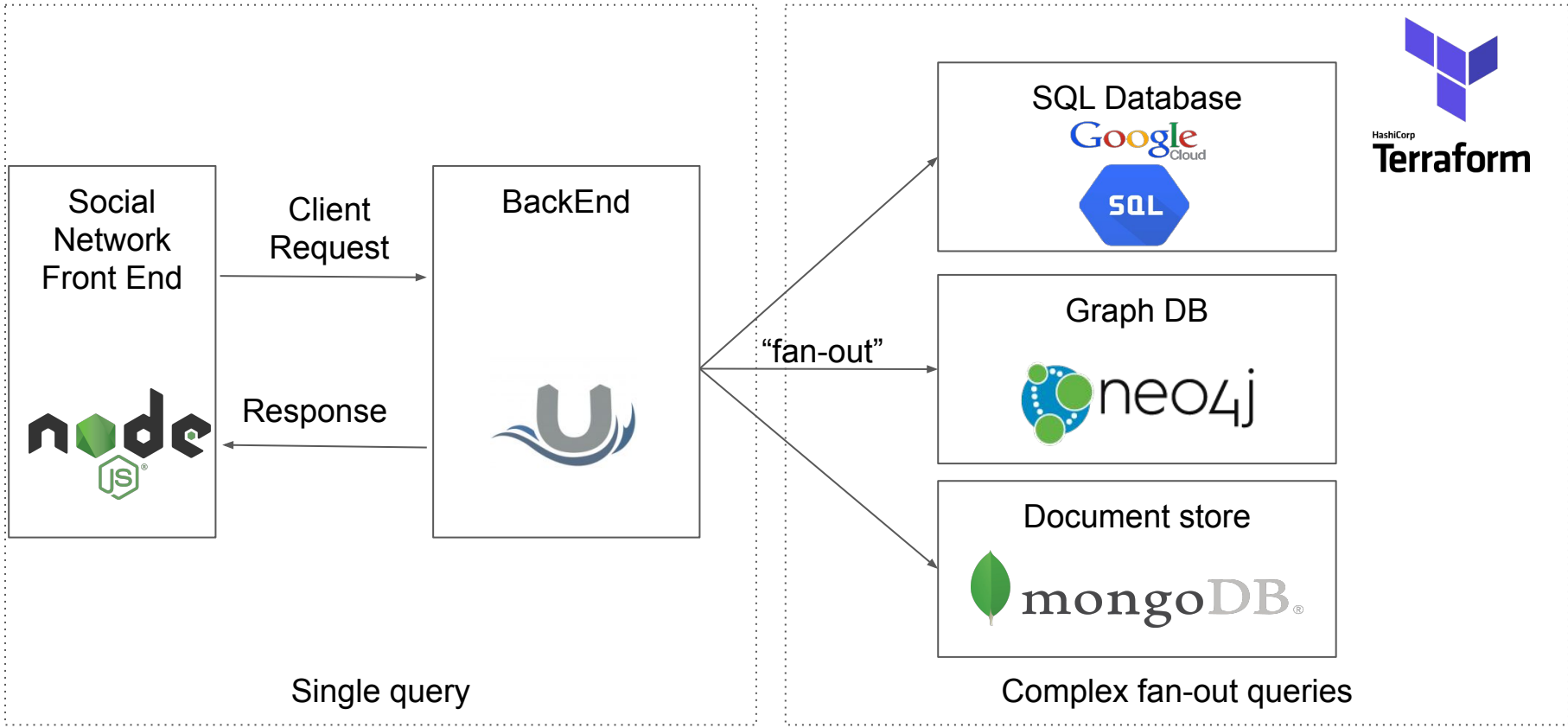
GCP

P2.3: Video Processing Pipeline - Serverless

- Functions and FFmpeg to process videos
- Use a cloud ML API for image labeling
- Azure Search to index videos

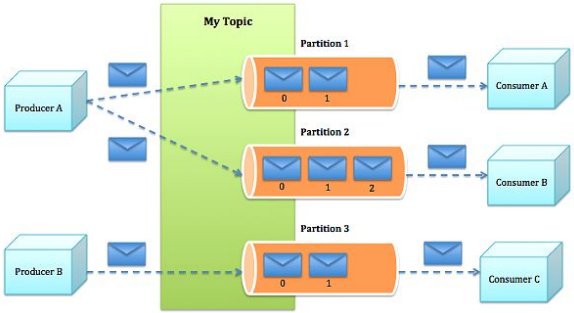


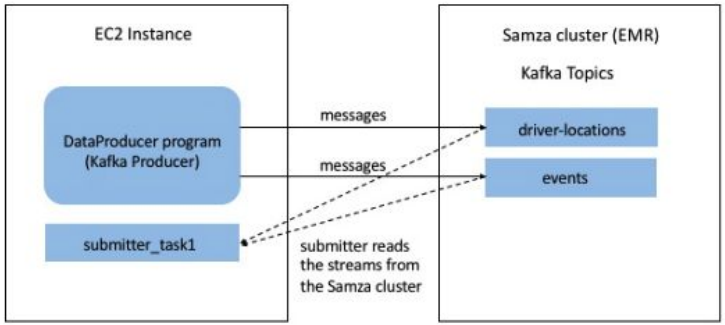
P3.2: Social Networking with Heterogeneous Backends



P4.3: Uber-like Application for NYC

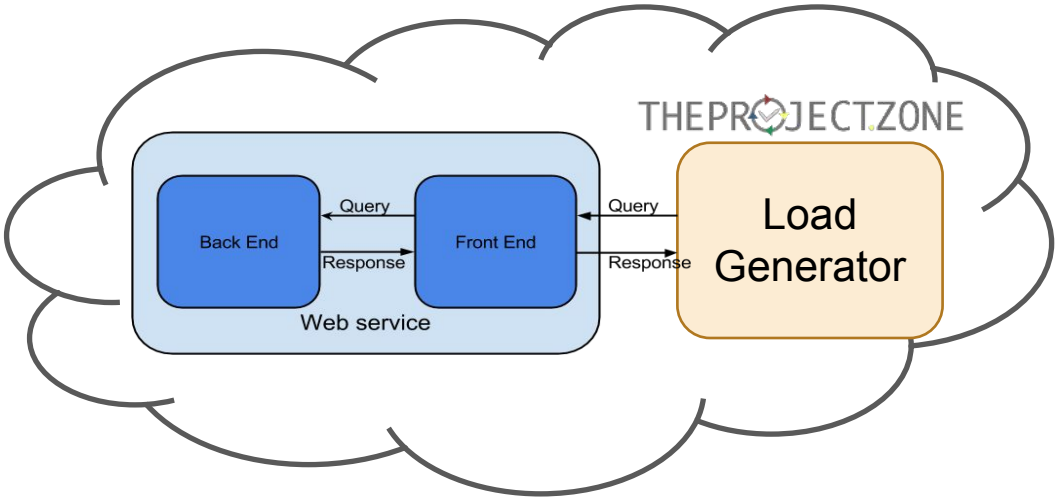
- Stream Processing with Kafka/Samza
  - Stream 1: Car GPS coordinates
  - Stream 2: Customers
- Task:
  - Match customers with cars to minimize travel time & other constraints





Team Project: Web Service

- Team-based
- 1.2 TB of raw data
- Specified queries
- Constraints
  - Correctness
  - Throughput
  - Budget
  - Time



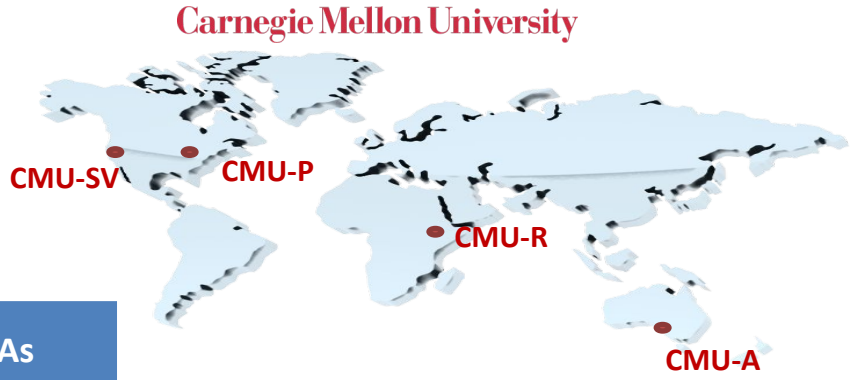
Phases	Duration	Query Type
Phase 1	2 weeks	Q1, Q2
Phase 2	2 weeks	Q1, Q2, Q3
Phase 2 Live Test	6 hours	Q1, Q2, Q3, mix-Q1Q2Q3
Phase 3	2 weeks	Q1, Q2, Q3
Phase 3 Live Test	6 hours	Q1, Q2, Q3, MIX-Q1Q2Q3

Outline

- What is the course about?
- **What is an online course?**
- Administrivia

Carnegie Mellon Global Course

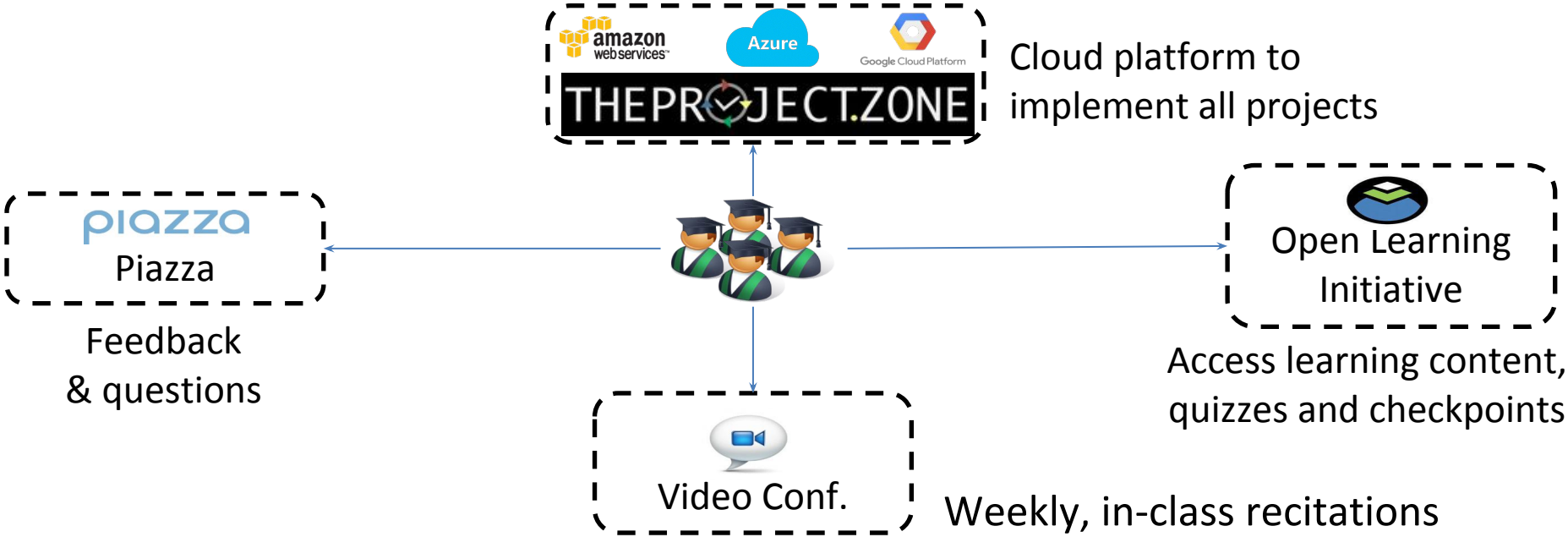
15-319 - 12 units
15-619 - 15 units



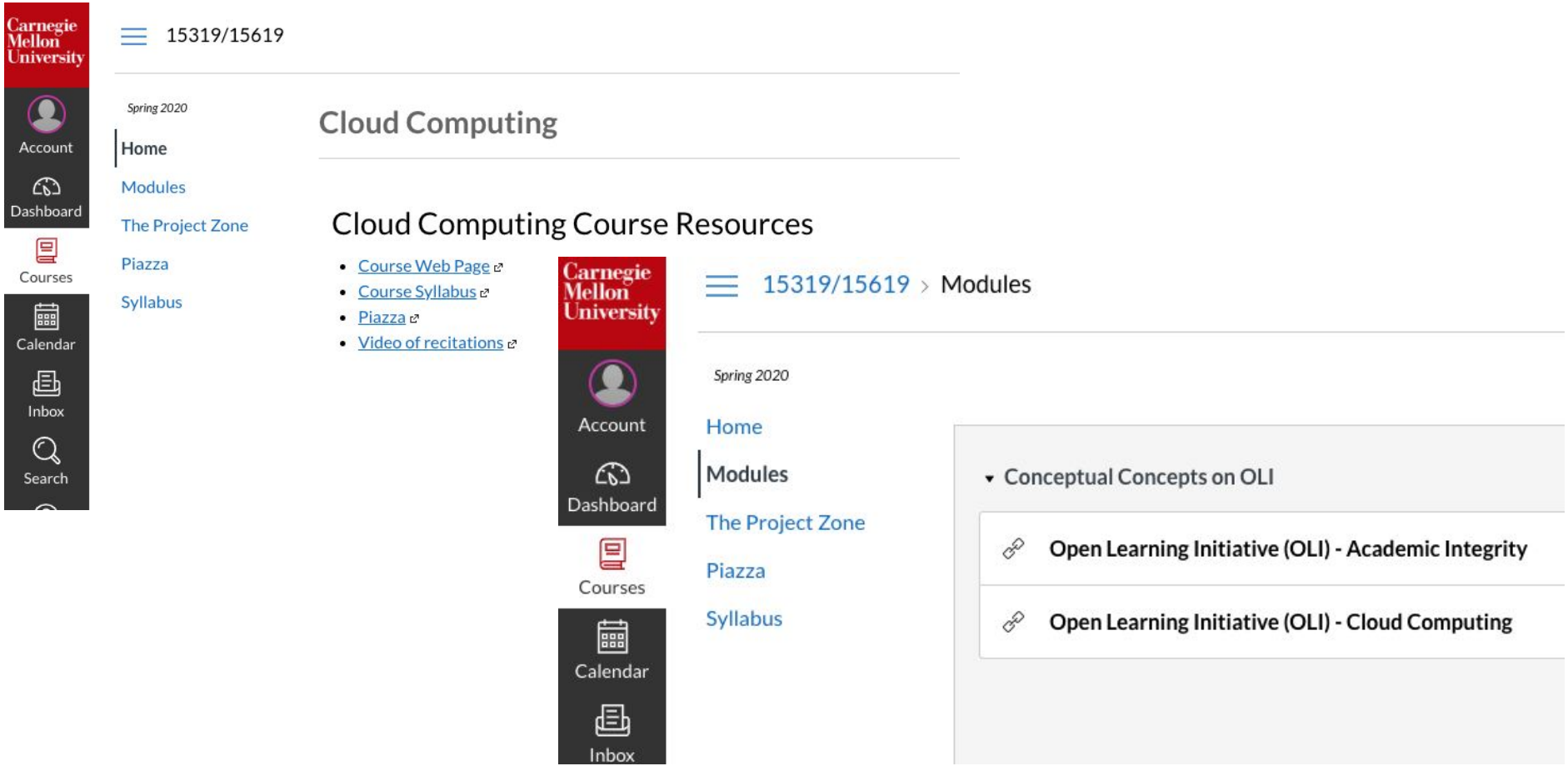
Location	Sections	Students	TAs
CMU Pittsburgh	A & B	138	13
CMU Rwanda	C	4	
CMU Adelaide	D	8	
CMU Silicon Valley	E	45	2

Please move to the section for your campus ASAP

Online Course Engagement Model



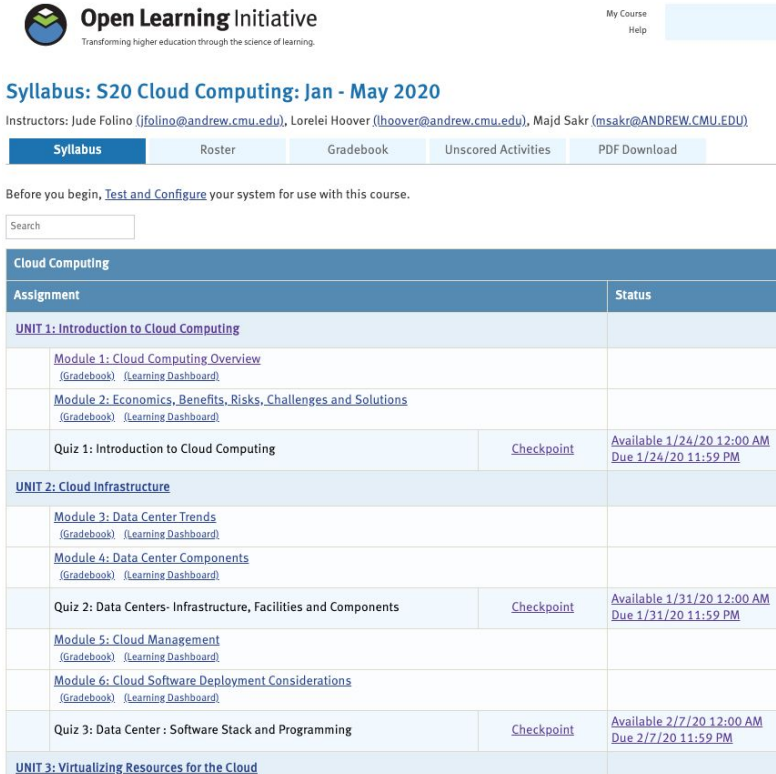
Canvas


Carnegie Mellon University
Account
Dashboard
Courses
Calendar
Inbox
Search
15319/15619
Spring 2020
Home
Modules
The Project Zone
Piazza
Syllabus
Cloud Computing
Cloud Computing Course Resources
Course Web Page
Course Syllabus
Piazza
Video of recitations
15319/15619 › Modules
Conceptual Concepts on OLI
Open Learning Initiative (OLI) - Academic Integrity
Open Learning Initiative (OLI) - Cloud Computing

Online Course Content - OLI

Conceptual content is on the Open Learning Initiative:

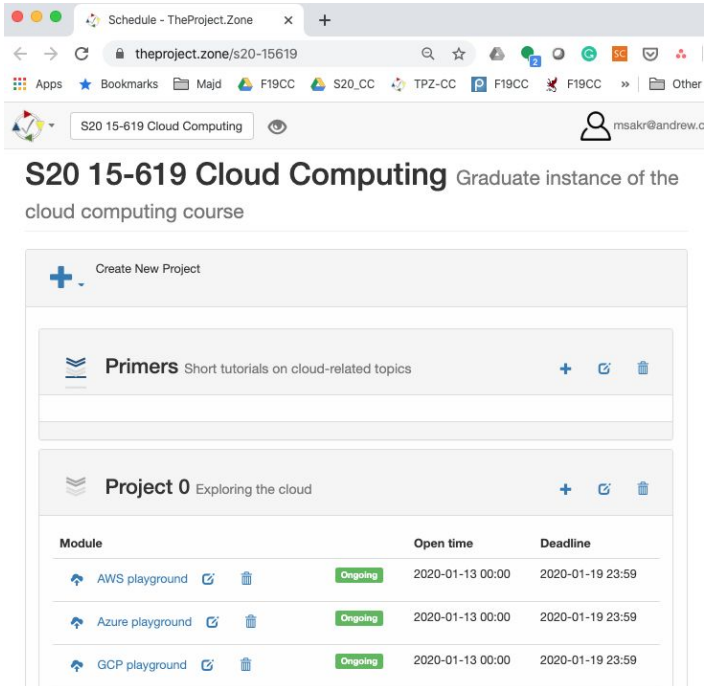
- Students are automatically registered
- Access to OLI is through Canvas
  - canvas.andrew.cmu.edu
- Check if Flash is installed
- Provide feedback on OLI
  - Bottom of each page
  - End of each module
- Do not copy or share content



TheProject.Zone

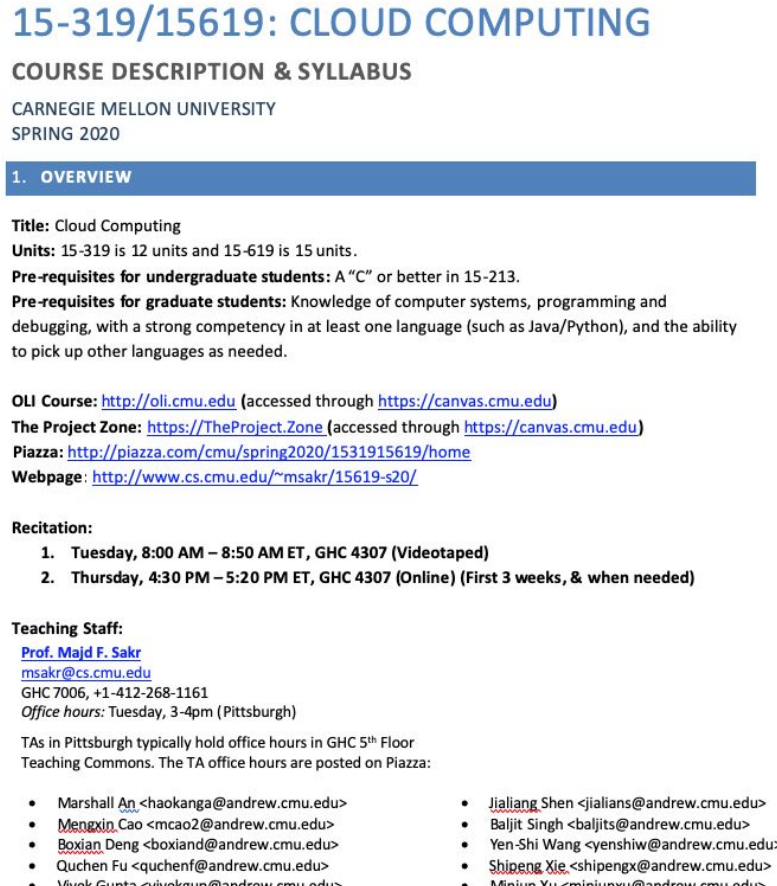
Course projects are on https://TheProject.Zone:

- Learn through repetitive attempts and feedback
- Students are automatically registered
- Access through browser
  - Not mobile friendly yet
- Work in progress
  - We will encounter bugs
  - Provide feedback on Piazza
  - Please be patient



Syllabus

- Updated on [webpage]
- Provides details on:
  - Course Objectives
  - Learning Outcomes
  - Policies
  - Grading
  - Tentative Schedule



15-319/15619: CLOUD COMPUTING

COURSE DESCRIPTION & SYLLABUS

CARNEGIE MELLON UNIVERSITY
SPRING 2020

1. OVERVIEW

Title: Cloud Computing
Units: 15-319 is 12 units and 15-619 is 15 units.
Pre-requisites for undergraduate students: A "C" or better in 15-213.
Pre-requisites for graduate students: Knowledge of computer systems, programming and debugging, with a strong competency in at least one language (such as Java/Python), and the ability to pick up other languages as needed.

OLI Course: http://oli.cmu.edu (accessed through https://canvas.cmu.edu)
The Project Zone: https://TheProject.Zone (accessed through https://canvas.cmu.edu)
Piazza: http://piazza.com/cmu/spring2020/1531915619/home
Webpage: http://www.cs.cmu.edu/~msakr/15619-s20/

Recitation:

1. **Tuesday, 8:00 AM – 8:50 AM ET, GHC 4307 (Videotaped)**
2. **Thursday, 4:30 PM – 5:20 PM ET, GHC 4307 (Online) (First 3 weeks, & when needed)**

Teaching Staff:

Prof. Majd F. Sakr
msakr@cs.cmu.edu
GHC 7006, +1-412-268-1161
Office hours: Tuesday, 3-4pm (Pittsburgh)

TAs in Pittsburgh typically hold office hours in GHC 5th Floor Teaching Commons. The TA office hours are posted on Piazza:

- Marshall An <haokanga@andrew.cmu.edu>
- Mengxin Cao <mcao2@andrew.cmu.edu>
- Boxian Deng <boxiand@andrew.cmu.edu>
- Quchen Fu <quchenf@andrew.cmu.edu>
- Jialiang Shen <jialians@andrew.cmu.edu>
- Baljit Singh <baljits@andrew.cmu.edu>
- Yen-Shi Wang <yenshiw@andrew.cmu.edu>
- Shipeng Xie <shipengx@andrew.cmu.edu>

Tentative Schedule

- Schedules:
  - Quizzes on OLI
  - Projects on TheProject.Zone

Week	Monday	OLI Content	Individual Projects	Team Project	Quizzes
1	1/13/2020	Unit 1, Module 1, 2	Primers/P0 (Jan 19)		Q0 (Ac. Integ.)
2	1/20/2020	Unit 1, Module 1, 2	P1.1 (Jan 26)		Q1 (Jan 24)
3	1/27/2020	Unit 2, Module 3, 4	P1.2 (Feb 2)		Q2 (Jan 31)
4	2/3/2020	Unit 2, Module 5, 6	P2.1 (Feb 9)		Q3 (Feb 7)
5	2/10/2020	Unit 3, Module 7, 8, 9	P2.2 (Feb 16)		Q4 (Feb 14)
6	2/17/2020	Unit 3, Module 10, 11, 12	P2.3 (Feb 23)		Q5 (Feb 21)
7	2/24/2020	Unit 3, Module 13	P3.1 (Mar 1)	Project Out (Feb 24)	Q6 (Feb 28)
8	3/2/2020	Unit 4, Module 14	P3.2 (Mar 8)		Q7 (Mar 5)
9	3/9/2020	Spring Break			
10	3/16/2020	Unit 4, Module 15, 16, 17	P3.3 (Mar 22)		Q8 (Mar 20)
11	3/23/2020	Unit 5, Module 18		Phase 1 Due (Mar 29)	Q9 (Mar 27)
12	3/30/2020	Unit 5, Module 19, 20	P4.1 (Apr 5)		Q10 (Apr 3)
13	4/6/2020	Unit 5, Module 21, 22		Phase 2 Due (Apr 12)	Q11 (Apr 10)
14	4/13/2020		P4.2 (Apr 19)		
15	4/20/2020			Phase 3 Due (Apr 26)	
16	4/27/2020		P4.3 (May 1)		

Grading

Course Elements	#	Weight
Projects	4 or 5	80%
OLI Unit Checkpoint Quizzes	11	20%

- Projects weights
  - 15-319
    - 80 %, 10 out of 11 individual project modules, each 8%
  - 15-619
    - 60%, 10 out of 11 individual project modules, each 6%
    - 20%, 1 team project, three phases
- Weekly quizzes (11 in total)
  - 10 out of 11, 2% equal weight

Audit & Pass/Fail option is not available for this course

Outline

- What is the course about?
- What is an online course?
- **Administrivia**

Target Audience

- Technical Majors
- Undergraduate Juniors / Seniors
  - Pre-requisites:
    - 15213 – Introduction to Computer Systems
- Graduate Students
  - Experience:
    - Unix, scripting, python, & java

Course Administration

- Students are automatically registered on OLI through canvas.cmu.edu
- A *single* Piazza course page is created
  - We manually register students to Piazza
- Schedule of units and quizzes is on OLI
  - Content weekly quizzes are due on Fridays
- Schedule of weekly projects is on TheProject.Zone
  - Weekly project modules are due on Sundays

Public Cloud Infrastructure

- Paid Cloud Service
  - billed by the hour/minute
- Start a resource only when you need it
- To explore, use inexpensive instances
- Terminate all other resources as soon as you are done with them
- Students will be penalized for over usage
  - We have a fixed budget, do not abuse the resources!
  - Intentional or unintentional abuse ➔ grade penalties
  - Resources need to be tagged, otherwise ➔ penalties

This Week

- Check that you were enrolled on Canvas and Piazza
- Academic Integrity Module on OLI
  - **Sunday, January 19, 2020**
- Become familiar with conceptual content on OLI
  - Start reading Unit 1, Module 1 & Module 2
  - Quiz 1: Unit 1, Module 1 & 2, **Friday, January 24, 2020**
- Create an account on AWS, Azure and GCP (ASAP)
  - Submit your AWS account info using the link provided in the primers on TheProject.Zone
- Projects on TheProject.Zone
  - Primer and P0, due **Sunday, January 19, 2020**

Diverse Technical Preparation

- Students come from diverse backgrounds and technical preparation
  - We offer primers to get you started.
  - If your programming skills are rusty, take the first two weeks to improve.
  - If you don’t think you have the skills required, allocate more time each week for the projects.
  - The first couple of weeks are less demanding, take advantage of them.

Perfect Conditions Do Not Exist

- Don’t ask to be trained under perfect conditions
  - We will not provide a sanitized sandbox for you to learn
- You will encounter
  - Badly formed data, inaccurate documentation, intermittent services, insufficient information, etc.
  - Learn how to deal with all these issues
  - Very valuable experience for your career

We are NOT special!

- We are as good as what skills, and hard work we bring to the table.
- Don’t ask for special circumstance due to drama.
  - Find out the source of the drama and make adjustments.
- …

Academic Integrity

It is the responsibility of each student to produce her/his own original academic work.

- Individual work:
  - Weekly Project Modules
  - Unit Checkpoint Quizzes
- Team work:
  - 15-619 Project

Read the university policy on Academic Integrity.

Disciplinary Policies

- First offense:
  - Minimum: worse than not doing the work.
  - Maximum: immediate expulsion.

- Second offense results in expulsion. Always.
  - Previously undiscovered offenses can count as “first offense”!

The Penalties are Severe

- Cheating leads to several students being dismissed from the university every semester

LET IT NOT BE YOU!

Academic Integrity Module on OLI

- Required for all students
- Process
  - Pretest Quiz
    - Please take this without looking at the modules
  - Page 1, Overview
  - Page 2, Policies
  - Page 3, Methods of Prevention
  - Quiz
    - Complete this quiz this week
      - By January 19, 2020

Working within Budgets

- Design is a critical element to success
- Develop a budget for
  - Development
  - Testing
  - Drama
- If funds are left over in the budget, feel free to explore and learn!

Tagging is painful, why the penalty?

- Your boss has a budget and a boss
- The budget is allocated among the team
- Your boss has to keep track of how the resources are being spent in order to
  - Re-allocate budget or ask for more resources
- On the cloud, the only way to keep track is through tagging
  - Learn how to tag correctly, don’t complain about penalties!

Getting Help

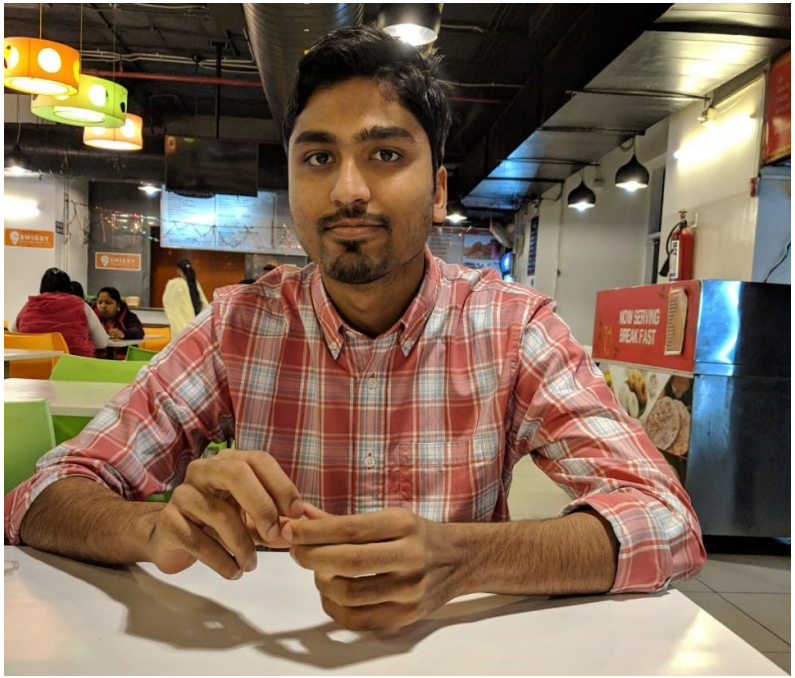
- TAs in Pittsburgh, Rwanda, Silicon Valley and Adelaide
- Piazza
  - Email does not scale
  - Discussion forum to support each other
- Recitations
  - Tuesdays (recorded)
    - At 8AM in **GHC 4307** (GHC 4401 for the first recitation **only**)
  - Thursdays (video conferenced to SV)
    - At 4:30PM in **GHC 4307** (1:30PM in online)
    - First 3 weeks only, afterwards only when needed
- Office Hours
  - Check Piazza for Office Hour schedule

Teaching Staff

- Majd Sakr
  - GHC 7006
  - msakr@cs.cmu.edu
  - Office Hours
    - Tuesdays, 3-4pm (Pittsburgh)

Teaching Staff

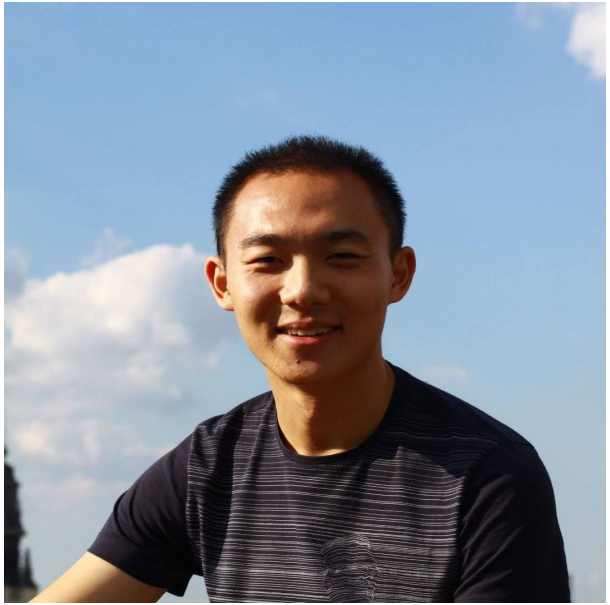
Haokang (Marshall) An



- Siddharth Kandimalla

Teaching Staff

- Aaron Madved



- Adam Zhang

Pittsburgh: Teaching Assistants

- Mengxin Cao

Pittsburgh: Teaching Assistants

- Boxian Deng

Silicon Valley: Teaching Assistant

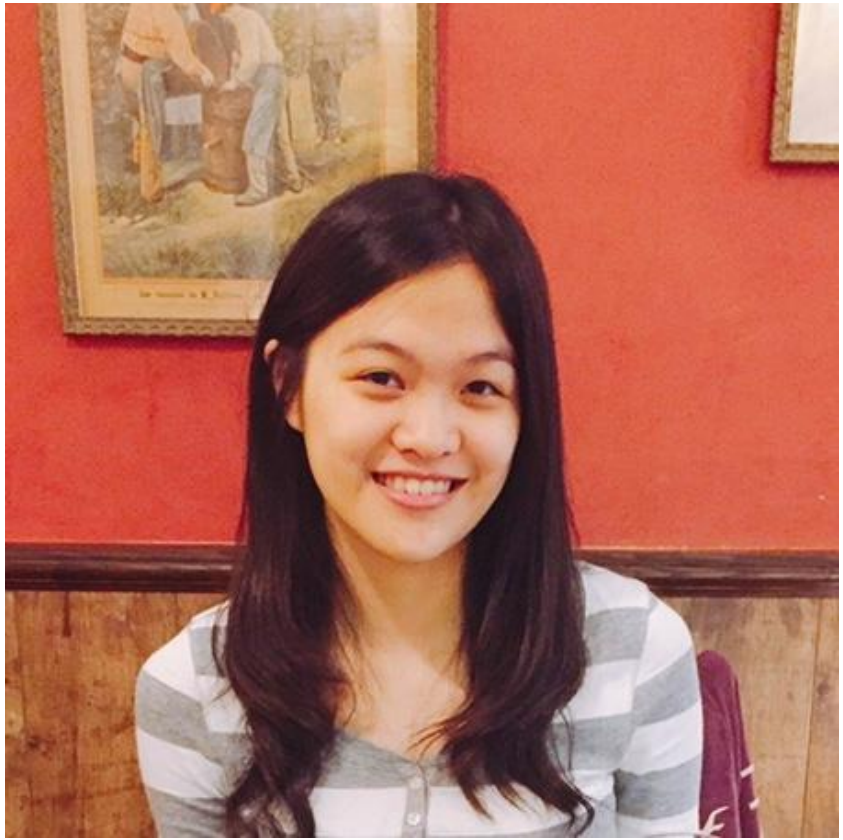
- Quchen Fu

Silicon Valley: Teaching Assistant

- Zecheng (Ron) He

Pittsburgh: Teaching Assistants

- Yu-Ning (Rebecca) Huang



Pittsburgh: Teaching Assistants

- Anmol Jagetia

Pittsburgh: Teaching Assistants

- Zian Ke

Pittsburgh: Teaching Assistants

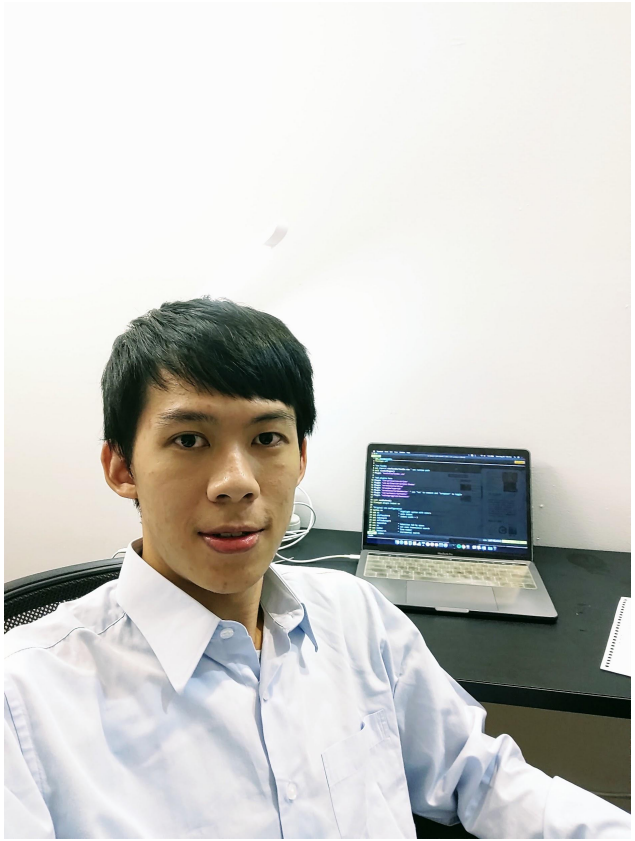
- Ignacio Maronna

Pittsburgh: Teaching Assistants

- Baljit Singh

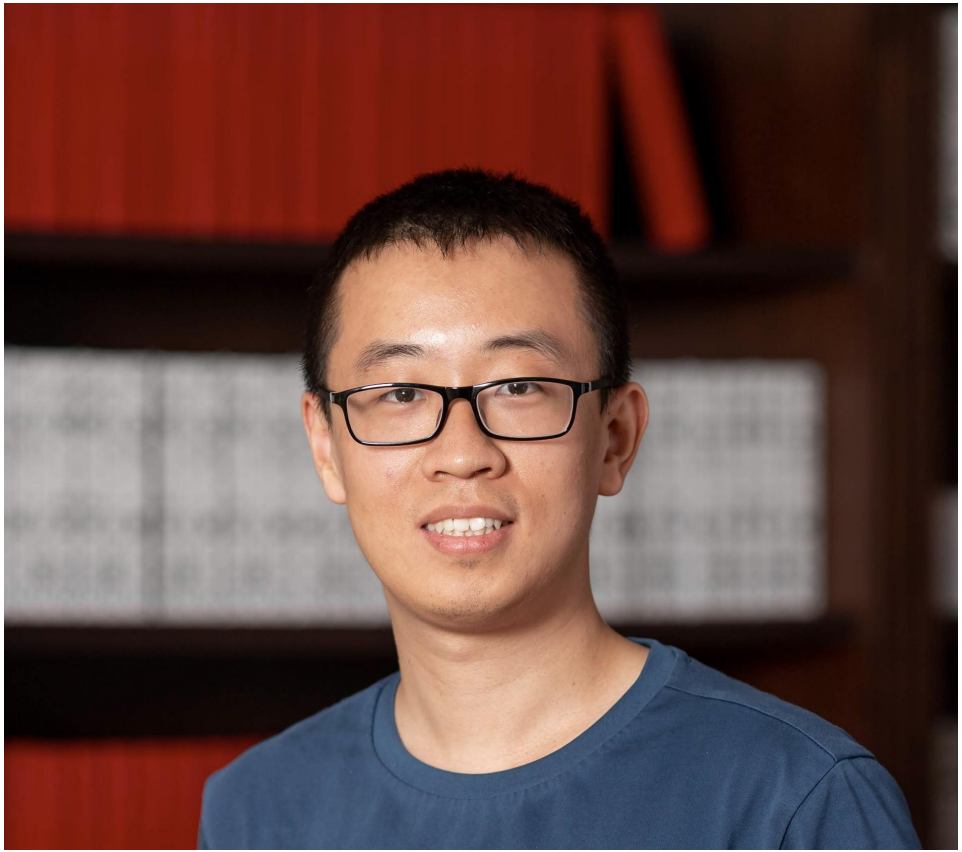
Pittsburgh: Teaching Assistants

- Yen-Shi Wang



Pittsburgh: Teaching Assistants

- Shipeng Xie



Pittsburgh: Teaching Assistants

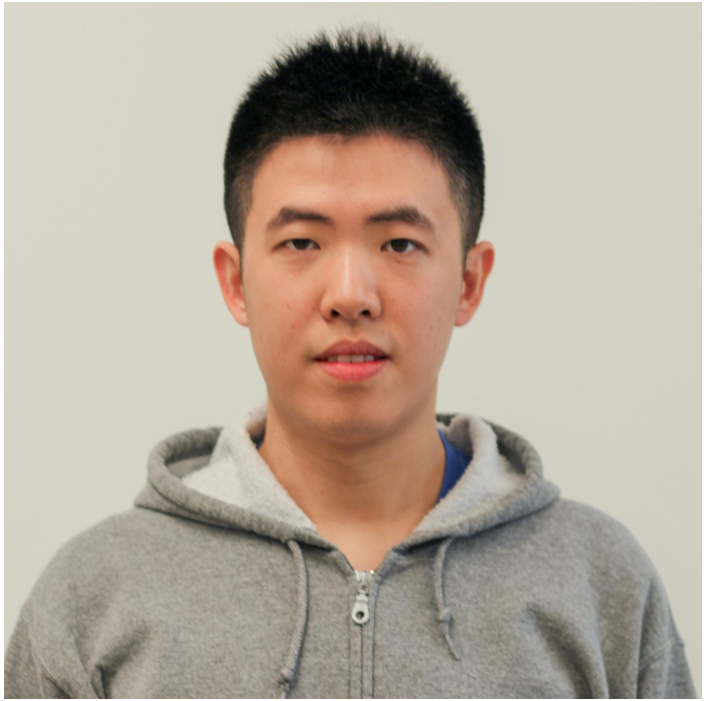
- Minjun Xu

Pittsburgh: Teaching Assistants

- Fengying Yang

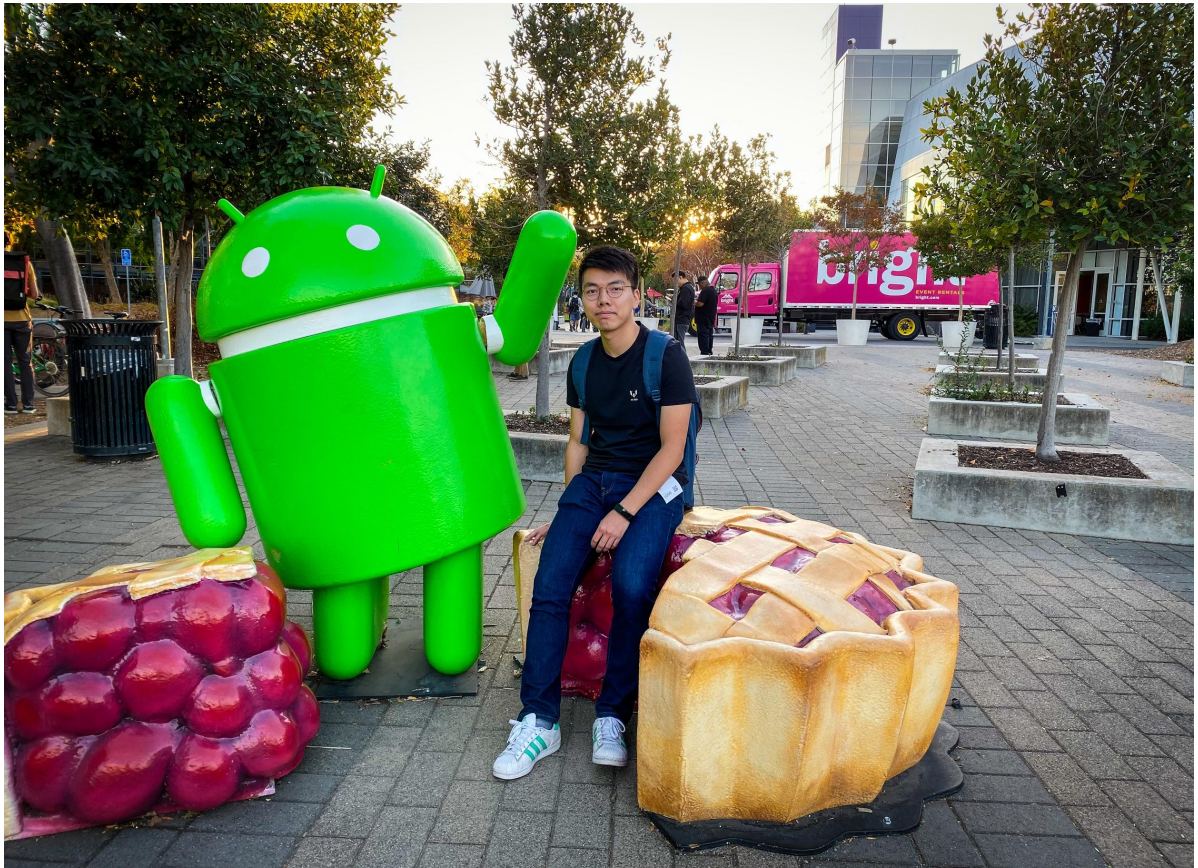
Pittsburgh: Teaching Assistants

- Zixuan (Tony) Zhu



Silicon Valley: Teaching Assistant

- Jialiang Shen

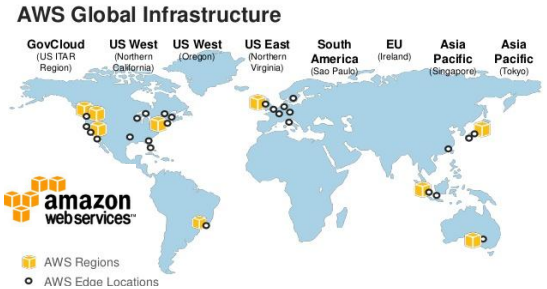


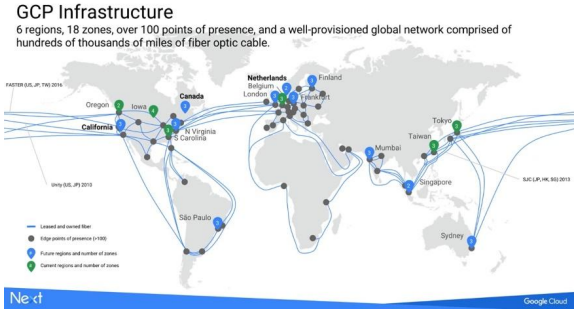
Silicon Valley: Teaching Assistant

- Hanya Zhang

Era of Globalization

- Economics
- Communication
- Entertainment
- Sports
- Education
- Compute Services 😊
  - You're programming the global computer.





Questions?

Data and Decision Making

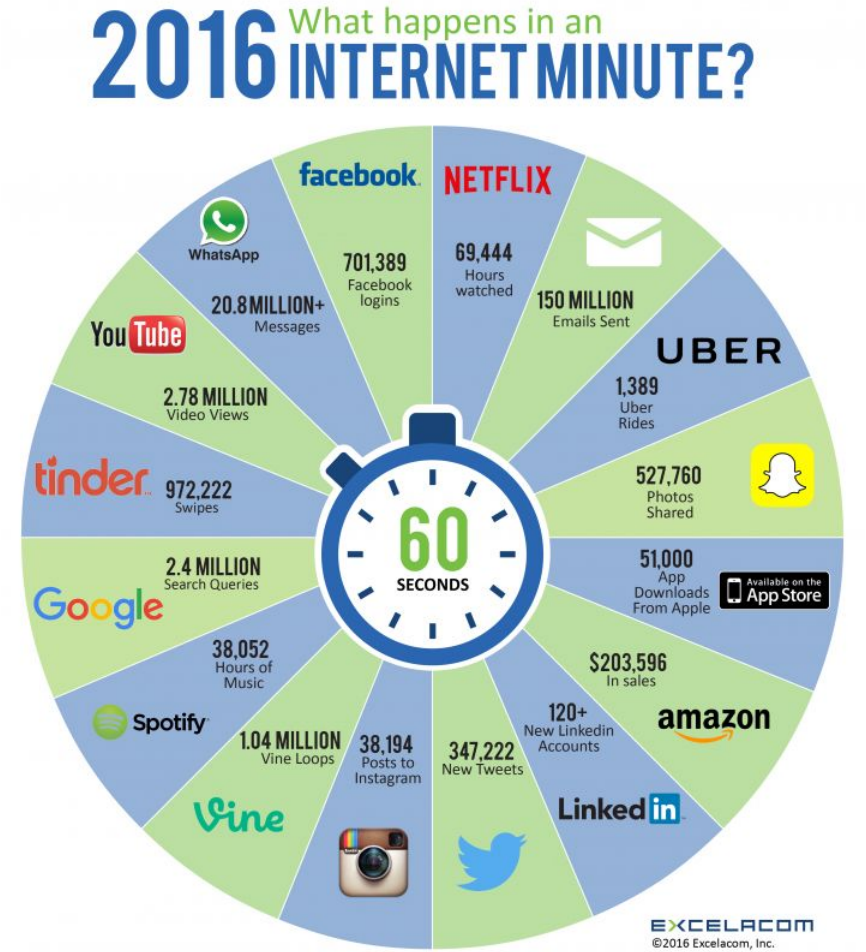
- Analyzing data reflects reality
- Walmart: hurricane warning
  - Stock beer and strawberry pop-tarts
  - 7x increase in sales during large storms
- Government: resource allocation decisions
  - Data mining in Maryland ➔ crime hotspots
  - Shuffle resource allocation, more to hotspots
    - violent crime down by 25%
    - $20 million saved in the city of Baltimore

Data Science

- Data Science is the science of learning from data
- Employs statistical, machine learning and data mining techniques
  - Look for trends, patterns or anomalies in the data
- Affects research in many domains
  - Business, Economics and Finance
  - Biological Sciences and Bioinformatics
  - Social Sciences and Humanities
  - …

An Increase in Data Capture

- Physical Sensors and Sensor Networks
  - Environmental, safety, transportation
- Social Media Interactions
  - Facebook, Twitter, Instagram
- Public Video and Image Capture
  - Surveillance, mobile phones, …
- Customer Spending Habits
  - Loyalty programs and purchase data


2016 What happens in an INTERNET MINUTE?
facebook
701,389 Facebook logins
NETFLIX
69,444 Hours watched
150 MILLION Emails Sent
UBER
1,389 Uber Rides
527,760 Photos Shared
51,000 App Downloads From Apple
Available on the App Store
$203,596 In sales
amazon
120+ New Linkedin Accounts
Linked in
347,222 New Tweets
38,194 Posts to Instagram
1.04 MILLION Vine Loops
Vine
38,052 Hours of Music
Spotify
2.4 MILLION Search Queries
Google
972,222 Swipes
tinder
2.78 MILLION Video Views
YouTube
20.8 MILLION+ Messages
WhatsApp
60 SECONDS
EXCELACOM
©2016 Excelacom, Inc.

What is Big Data?

- Big Data
  - Volume, Velocity, Variety, Veracity
  - Data of next year >> data of this year
- Many Challenges
  - Store, share, analyze, search, transfer, visualize, and secure

we need…

Large Scale Systems

Large Scale System Challenges

- Lengthy procurement cycles
- Lengthy deployment effort
- Costly power and cooling
- Costly systems administration
- Low utilization
- Costly disaster recovery

Software as a Service (SaaS)

- Software is delivered through the internet over a browser or mobile application
- Replace desktop software with cloud-based versions
- Webmail, Productivity Software, ERP, CRM etc.
- Centrally managed, globally available, automatically updated

Adobe Creative Cloud

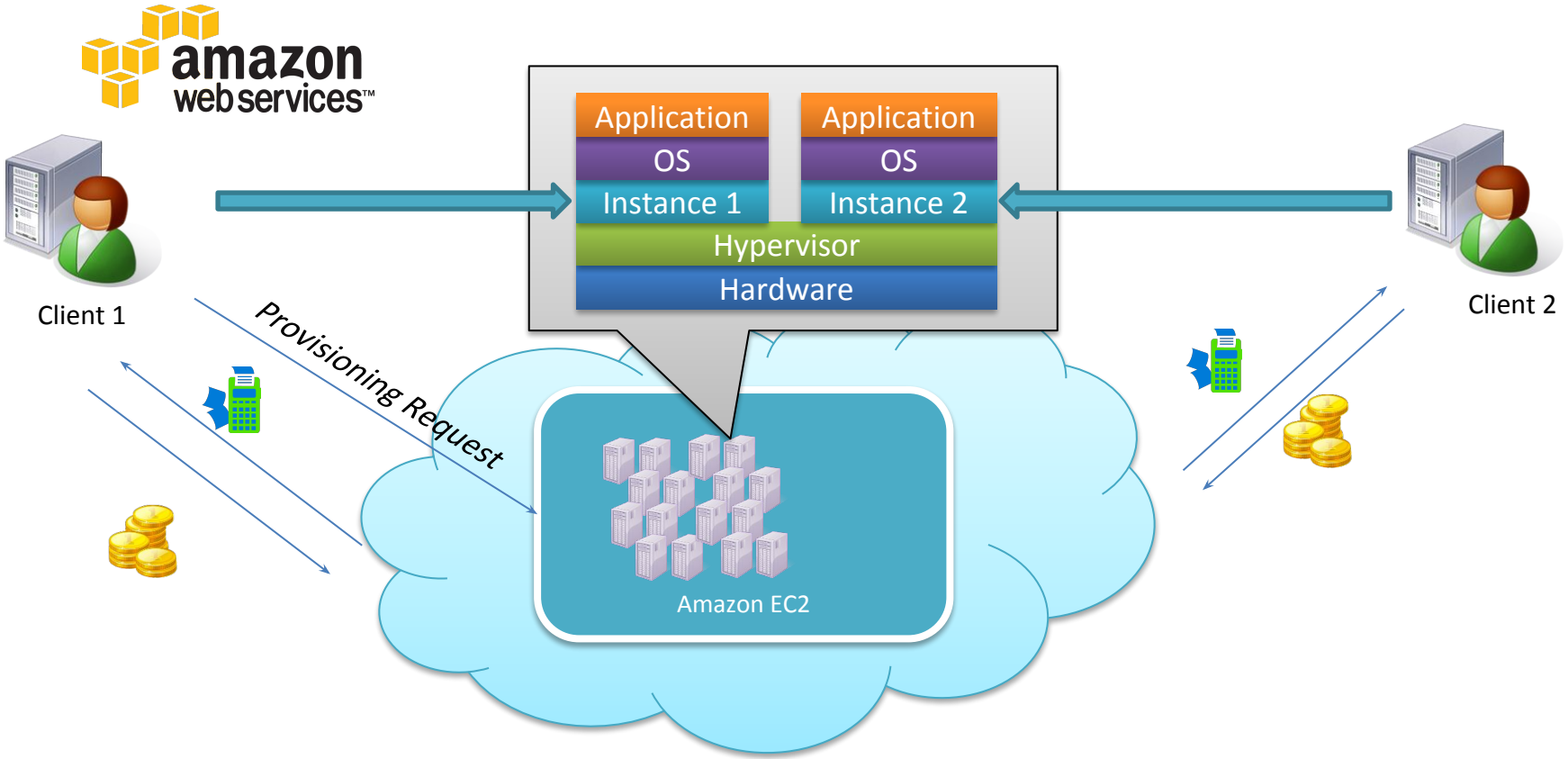
Platform as a Service (PaaS)

- Tools and APIs to develop and deploy cloud-based applications
- Create customized SaaS in the form of Web or mobile applications

Infrastructure as a Service (IaaS)

- Compute, storage and network resources bundled in the form of virtual machines
- Fully flexible in terms of software and environment

Infrastructure as a Service



Benefits of Cloud Computing

Pay-as-You-Go economic model

Simplified IT management

Elasticity
Scale quickly & effortlessly

Customization
Flexible options

Carbon Footprint decreased

Risks and Challenges

Migration

Security & Privacy

Vendor Lock-In

Legal

Internet Dependence

Service Level Agreements and Objectives (SLA/SLO)

- SLA: Contract between cloud providers and users to define expected service
  - Service availability and delivery
  - Payment terms, bonuses and penalties for service
- SLO: Individual performance/service metrics regarding service delivery defined in the SLA
- Auditing: monitor resources to enforce SLOs and SLAs

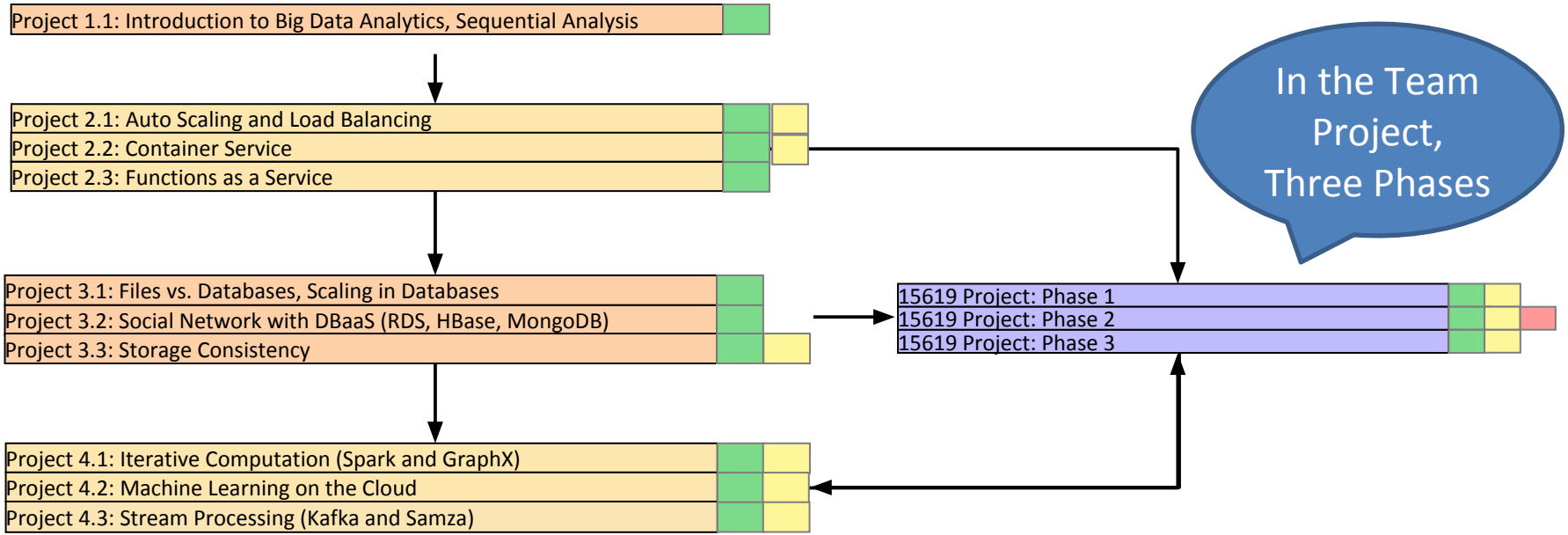
Cloud Use Cases: Start-ups

- Infrastructure on demand
- Save money on data center real estate, servers, power and cooling
- Saving in capital expenditure which could be used to drive other areas of business growth
- Scale infrastructure as the business grows
- Levels the infrastructure playing field with established companies

Projects

- Four Individual Projects (all students):
  0. Primers and P0 (Due Sunday, September 02, 2018)
  1. Big Data Analytics
  2. Scalability, Elasticity and Failure
  3. Cloud Storage
  4. Analytics Engines for the Cloud
- One Team Project, Twitter Analytics Web Service (15-619 students)
  - One multi-week team project to build a complete web service

Projects: Timeline and Dependencies



Expectations

- Real world practical experience
  - Learn on your own
    - Languages, API, debugging
  - Overcome challenges
  - Deal with uncertainty
- Self paced learning
- Using experimental tools
  - Bleeding edge comes with risks

Amazon Web Services

- Paid Cloud Service – billed by the hour
- Start a resource only when you need it
- To explore, use a micro instance
  - You can keep one micro instance running 24x7
- Terminate all other resources as soon as you are done with them
- Students will be penalized for over usage
  - We have a fixed budget, do not abuse the resources!
  - Intentional or unintentional abuse ➔ grade penalties
  - Resources need to be tagged, otherwise ➔ penalties

Microsoft Azure

Azure Subscription

- Microsoft Azure provides the cloud computing course with one coupon per student
  - Students will provide an MSA on theproject.zone
  - We will create a subscription on Azure
- If you are careless in spending
  - Spending penalty
  - You will run out of money for a given project
    - Make a private Piazza post

GCP Coupon

- Google Cloud Platform provides the cloud computing course with one coupon per student
  - Students will access to a GCP coupon through theproject.zone
- If you are careless in spending
  - Spending penalty
  - You will run out of money for a given project
    - You will have to pay-as-you-go

Typical Reasons

- I didn’t know that this would be considered an AIV at Carnegie Mellon.
- I started late, ran out of time, looked online to see if there are any solutions that will help me regain some time.
- I took too many demanding courses at the same time. I had no choice.

Outcomes of Committing an AIV

- Positives
  - It's a terrible way to learn an important lesson
- Negatives
  - Grade Penalty
    - Ranges from significant penalty to failure
    - Lose time and change plans since you have to retake course
  - Dismissal from program/university
    - Time, money and emotion
    - Job offers indicate
      - "Upon successful completion of your program"
    - Having to go home and tell everyone

What is Cheating

- Sharing code or other electronic files either by copying, retyping, looking at, or supplying a copy of any file.
  - Other students, github, stackoverflow, anywhere on the internet,…
- Copying answers to any checkpoint quiz from another individual, published or unpublished written sources, and electronic sources.
- Collaborating with another student or another individual on Unit Checkpoint Quizzes or Project Modules.
- Sharing written work, looking at, copying, or supplying work from another individual, published or unpublished written sources, and electronic sources.
- Collaboration in team projects is strictly limited to the members of the team.
- …(read the syllabus and the university policy)

Minimum Cheating Penalty

- Must be worse than not submitting anything
  - Example impact of a -100% penalty on a project

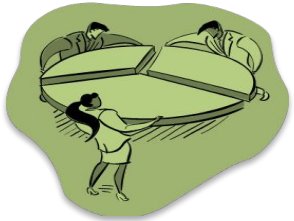
	Perfect Score	Not submitting	Cheating
Assessment #1	20%	20%	20%
Assessment #2	20%	20%	20%
Assessment #3	20%	20%	20%
Assessment #4	20%	20%	20%
Assessment #5	20%	0%	-20%
Total	100%	80%	60%

  - Typical penalty, “R” in the course

… for a more complete definition!

Cloud Computing is the delivery of computing as a **service** rather than a **product**,

whereby **shared resources, software**, and **information** are provided to computers and other devices,



as a **metered service** over a **network.**

Hands-on Skill Building Course

- Experience with
  - Elasticity, scalability, monitoring, load balancing
  - Cloud storage (HDFS, SQL, NoSQL, etc.)
  - Frameworks (MapReduce, Spark, Kafka/Samza)
  - Latest cloud services (AWS, Azure and GCP)
- Real world
  - data, projects, infrastructure within a budget

What this course is *not* about

- Building Cloud Stack Modules
  - OpenStack
- Cloud Software Development
  - SaaS software engineering
- Distributed Systems
  - Synchronization, Consistency, …
- Operating Systems
  - Developing a hypervisor
- Networks
  - Routing and switching protocols

Skill training to help with your goals

- Student has goals
  - Internship, job, graduate school, research project
- Student hires Carnegie Mellon to help achieve goals
- Student and Carnegie Mellon are partners in Learning effectively
- But, how do we learn?

Deeper Processing Leads to Deeper Learning

- Studying a lot
  - You will forget it soon
- Infrequent testing
  - You will still forget most of it soon
- Frequent testing
  - Helps improve retention a great deal
- Project work using/exploring/attempting to solve a problem
  - Helps even more

The deeper you think, the more you practice, the more you will learn and retain.

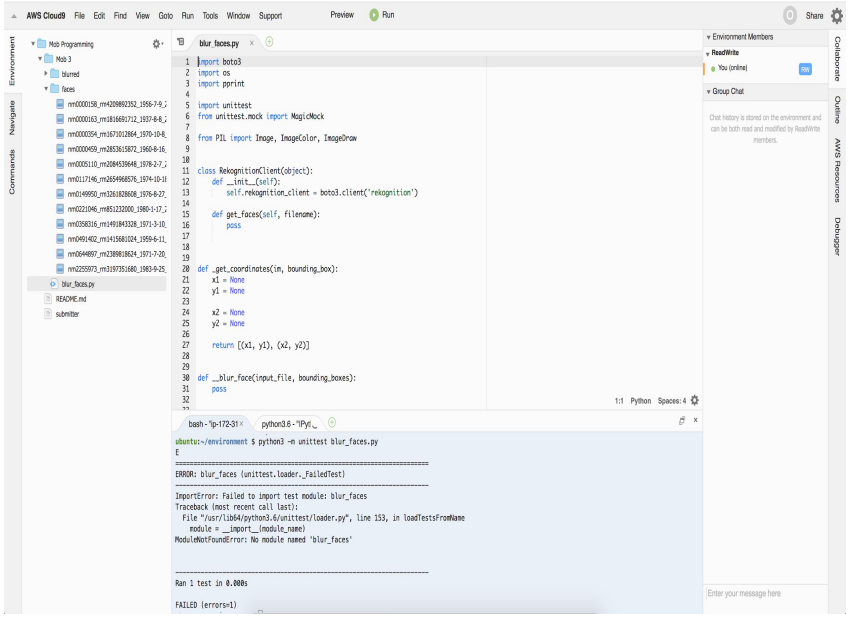
Working on Projects

- Understand the context and tasks/problem
- Choose a path towards a solution
- Attempt the path
- Face obstacles
  - Ask for help?
  - Look for another path on Stackoverflow and try?
  - Choose another path?
  - Find the reason behind the obstacle?
- Which is the hardest?

Performance versus Mastery

- If you have high performance but low mastery
  - ➔Means high grades but low skills & confidence
    - ➔Get an interview but will not pass
- The grades will not get you the job
  - Companies care about what you know and what you can do
- When you have high mastery, you will be in very high demand
  - Seek mastery!

Participate in Online Programming Exercise Sessions



Cloud IDE: AWS Cloud9