CS15-319 / 15-619 Cloud Computing

Recitation 1

Course Overview and Introduction

Aug 29 and 31, 2017

http://www.cs.cmu.edu/~msakr/15619-f17/

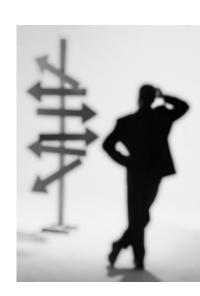
Outline

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

So What is Cloud Computing?

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

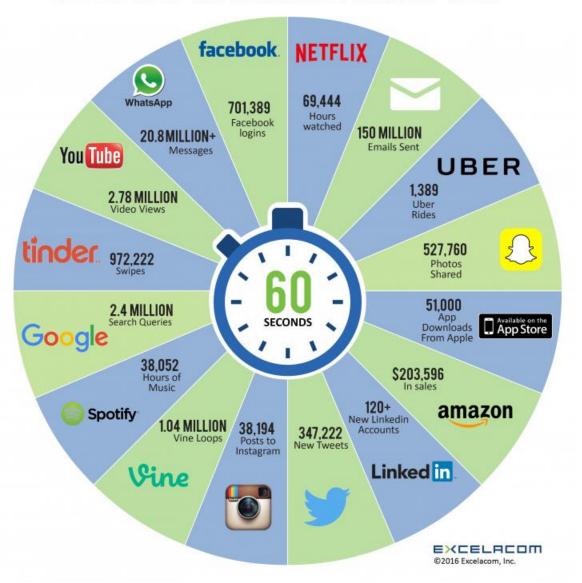
– ...

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?



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



Evolution of Computing

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

Innovation

Product

Service

Evolution of Electricity



Innovation

New Disruptive
Technology



ProductBuy 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



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

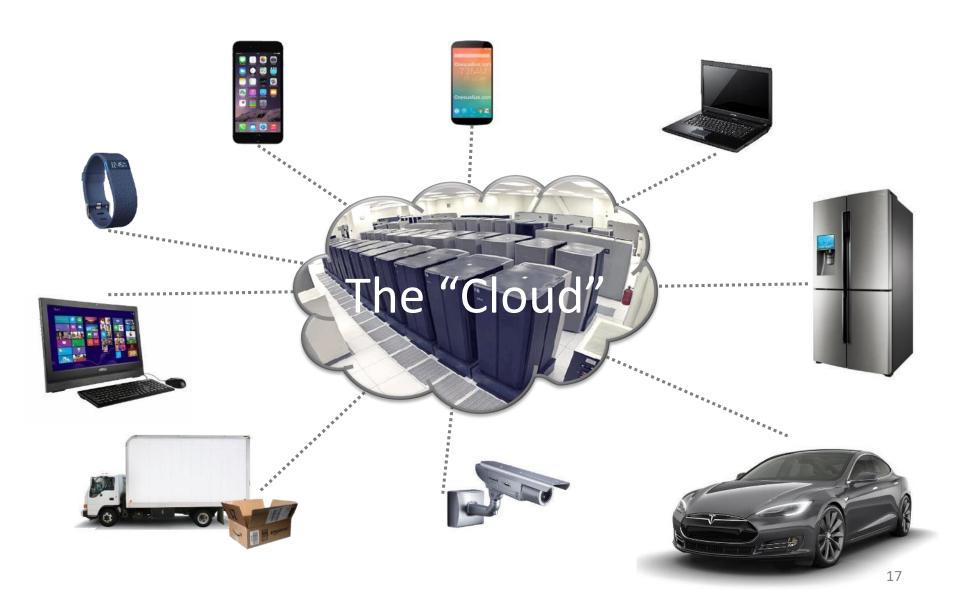








The Cloud



... for a more complete definition!



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

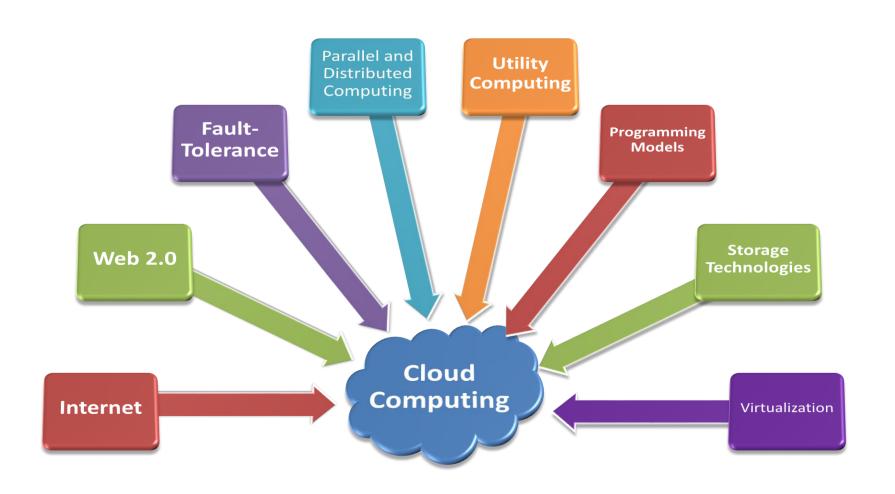
whereby shared resources, software, and information are provided to computers and other devices,





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

Enabled by Maturing Technologies



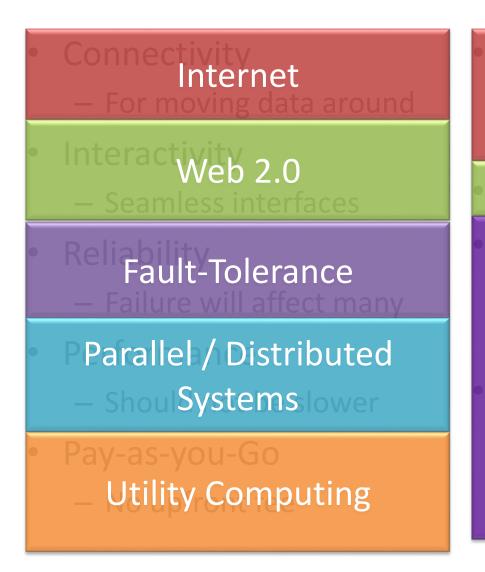
So... how would you transform information technology into a Service?

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?

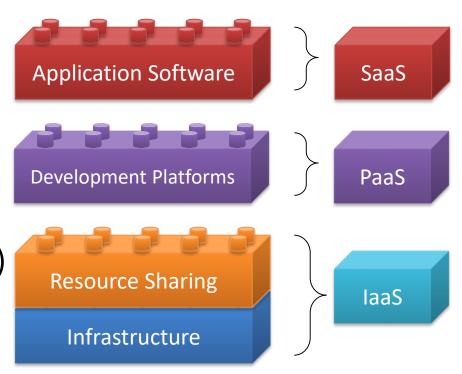


Programming Model Storage Technologies Virtualization and **Resource Sharing Technologies**

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 (laaS)



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











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



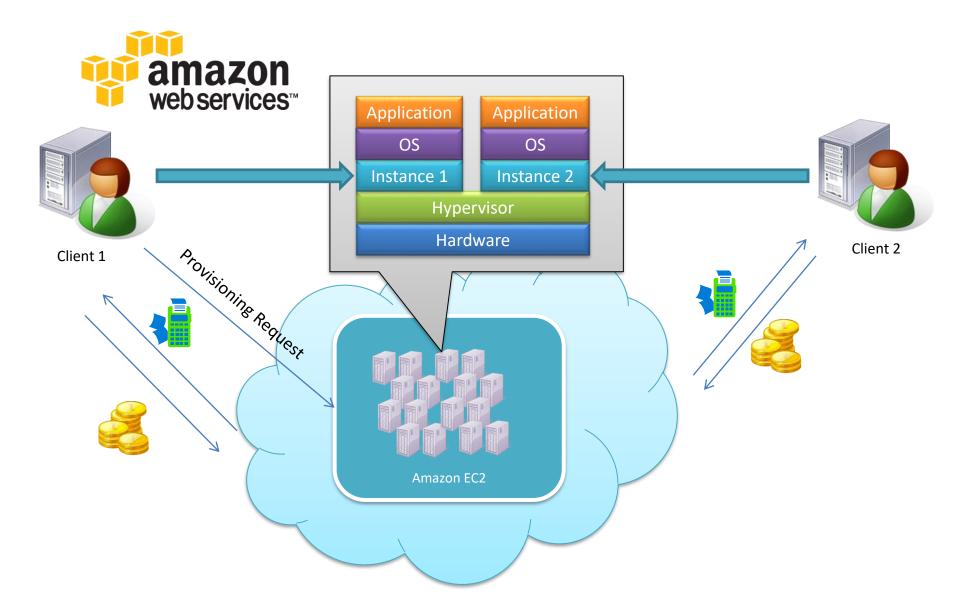








Infrastructure as a Service



Benefits of Cloud Computing



Risks and Challenges



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

Cloud Computing Stack

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



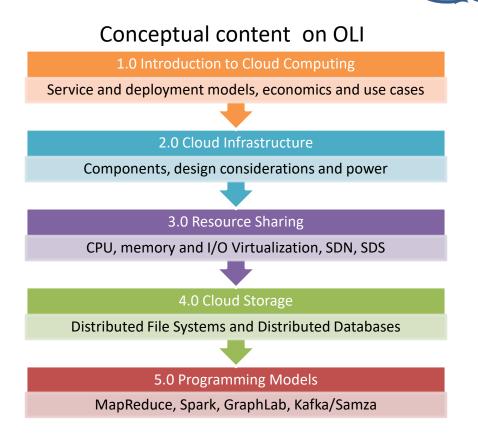
What is this course about?

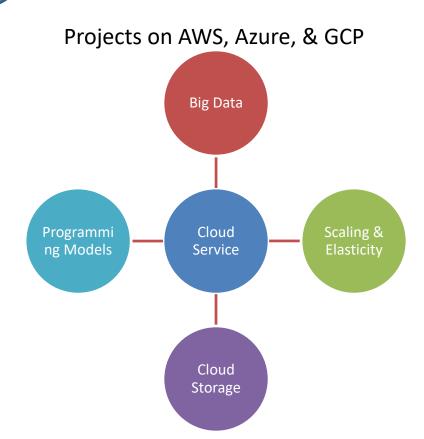
- Applied aspects of cloud computing
 - Between systems and services

Cloud Systems and Infrastructures

Course Content

Cloud Services and Applications





Course Objectives

Students will learn:

- the fundamental ideas behind Cloud Computing;
- the basic ideas and principles in cloud infrastructure, data center design and management; cloud software stack and software deployment on the cloud
- the resource sharing and virtualization techniques that serve in offering software, computation and storage services on the cloud; Software Defined Networks (SDN) and Software Defined Storage (SDS);
- about cloud storage technologies and relevant distributed file systems, NoSQL databases and object storage;
- the variety of programming models and develop working experience in three of them.

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	Quiz 1, Sep 8, 2017
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)		Swift, Ceph)
6	Programming Models	Distributed Programming for the Cloud	
		Data-Parallel Analytics with Hadoop MapReduce (YARN)	
		Iterative Data-Parallel Analytics with Apache S	
		Stream Processing with Apache Kafka and Sam	nza

Projects on AWS/Azure/GCP Clouds

Due 9/3/2017

O. AWS/Azure Account Setup & Tool Primer

• Benchmarking VMs, SSH, Authentication, Billing, Security Groups, Vertical Scaling

1. Big Data Analytics

 Amazon EC2, Amazon EMR, Azure HDInsight

2. Scaling, Elasticity and Failure

 Auto Scaling, Load Balancing, Monitoring, Containers, Kubernetes, Docker

3. Cloud Storage

 Amazon EBS, MySQL, HBase, Amazon RDS, MongoDB, Impala, Redshift

4. Analytics Engines for the Cloud

 Amazon EMR, MapReduce, Spark, GraphLab, Kafka/Samza

5. A Complete Web Service (Team project, extra 3-units)

- No restrictions on tools or AWS services
- Evaluated based on cost and performance

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

Projects

- Four Individual Projects (all students):
 - O. Primers and PO (Due Sunday, Sep 3, 2017)
 - 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, extra 3-units)
 - One multi-week team project to build a complete web service

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

Outline

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

Carnegie Mellon Global Course

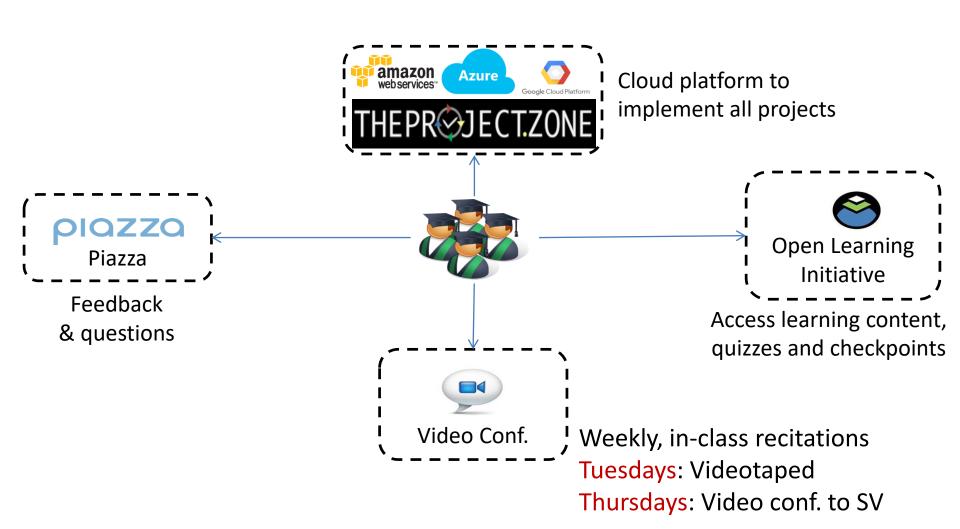
Carnegie Mellon University



Location	Sections	Students	Teaching Staff
CMU Pittsburgh	A & B	139	14
CMU Silicon Valley	С	72	7
CMU Rwanda	D	2	0

Please move to the section for your campus ASAP

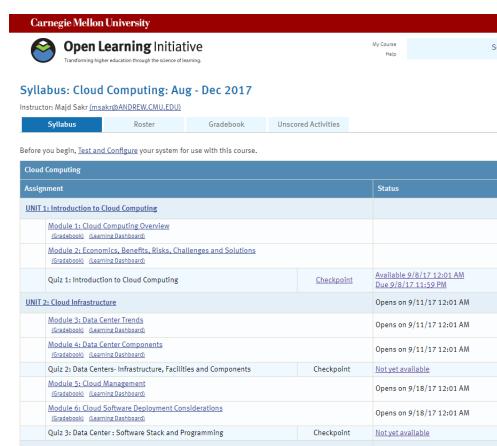
Online Course Engagement Model



Online Course Content - OLI

Course content is on the Open Learning Initiative:

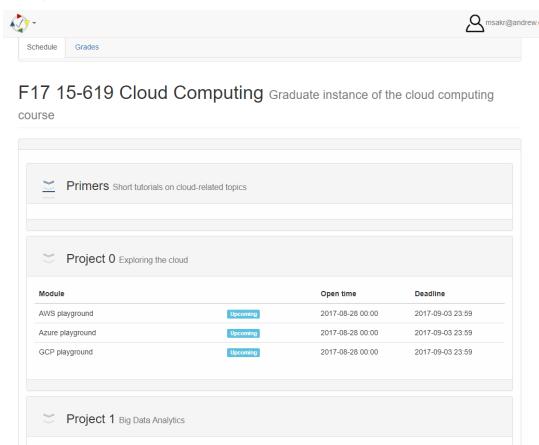
- Students are automatically registered
- Access to OLI is through Canvas
 - canvas.cmu.edu
- Check if Flash is installed
- Provide feedback on OLI
 - Bottom of each page
 - End of each module
- Do not copy or share content



The Project. 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 feedbackon 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 FALL 2017

1. OVERVIEW

Title: Cloud Computing

Units: 15-319 is 9 units and 15-619 is 12 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

Piazza: http://piazza.com/cmu/fall2017/1531915619/home

Recitation:

- Tuesday, 8:00 AM 8:50 AM, GHC 4307 (Videotaped)
- 2. Thursday, 4:30 PM 5:20 PM, GHC 4307 (VC to SV) (First three weeks and when needed)

Teaching Staff:

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

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:

Prof. Seth Copen Goldstein seth@cs.cmu.edu GHC 7111, +1-412-268-3828

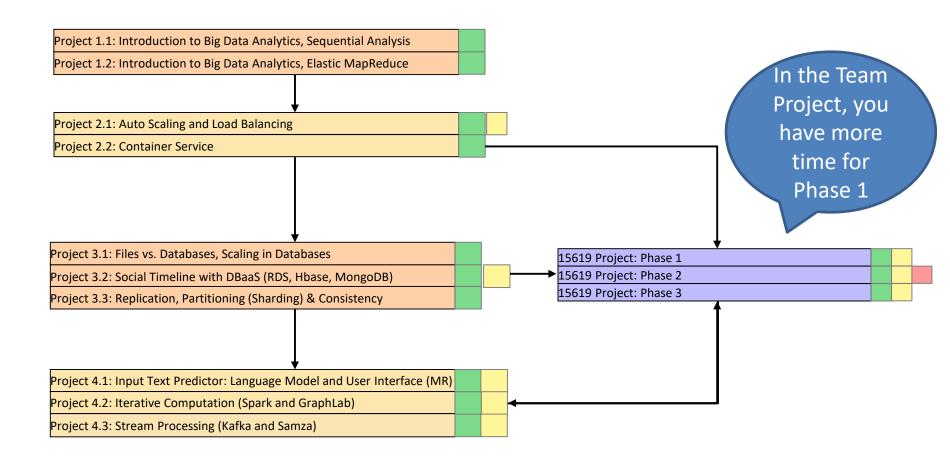
Office hours: TBA

Tentative Schedule

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

Week	Monday	OLI Content	Individual Projects	Team Project	Quizzes
1	8/28/2017	Unit 1, Module 1, 2	Primers/P0 (Sep 3)		Q0 (Ac. Integ.)
2	9/4/2017	Unit 1, Module 1, 2	P1.1 (Sep 10)		Q1 (Sep 8)
3	9/11/2017	Unit 2, Module 3, 4	P1.2 (Sep 17)		Q2 (Sep 15)
4	9/18/2017	Unit 2, Module 5, 6	P2.1 (Sep 24)		Q3 (Sep 22)
5	9/25/2017	Unit 3, Module 7, 8, 9	P2.2 (Oct 1)		Q4 (Sep 29)
6	10/2/2017	Unit 3, Module 10, 11, 12	P3.1 (Oct 8)		Q5 (Oct 6)
7	10/9/2017	Unit 3, Module 13	P3.2 (Oct 15)	Project Out (Oct 9)	Q6 (Oct 13)
8	10/16/2017	Unit 4, Module 14	P3.3 (Oct 22)		Q7 (Oct 19)
9	10/23/2017	Unit 4, Module 15		Phase 1 Due (Oct 29)	Q8 (Oct 27)
10	10/30/2017	Unit 4, Module 16, 17	P4.1 (Nov 5)		Q9 (Nov 3)
11	11/6/2017	Unit 5, Module 18		Phase 2 Due (Nov 12)	Q10 (Nov 9)
12	11/13/2017	Unit 5, Module 19, 20	P4.2 (Nov 19)		Q11 (Nov 17)
13	11/20/2017	Thanksgiving			
14	11/27/2017	Unit 5, Module 21, 22		Phase 3 Due (Dec 3)	Q12 (Dec 1)
15	12/4/2017		P4.3 (Dec 8)		

Projects: Timeline and Dependencies



Grading

Course Elements	#	Weight
Projects	4 or 5	80%
OLI Unit Checkpoint Quizzes	12	20%

- Projects weights
 - -15-319
 - 80 %, 10 individual project modules, each 8%
 - -15-619
 - 60%, 10 individual project modules, each 6%
 - 20%, 1 team project, three phases
- Weekly quizzes (12 in total)
 - equal weight

Pass/Fail option is no longer available for this course

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



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



- 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







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



Azure Subscription

- Microsoft Azure provides the cloud computing course with one coupon per student
 - \$100 for this course
 - 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 (\$100) for a given project
 - You will have to pay-as-you-go

GCP Coupon



- Google Cloud Platform provides the cloud computing course with one coupon per student
 - \$100 for this course
 - Students will access a \$50 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

This Week

- Academic Integrity Module on OLI
 - Sunday, Sep 3, 2017
- Become familiar with conceptual content on OLI
 - Start reading Unit 1, Module 1 & Module 2
 - Quiz 1: Unit 1, Module 1 & 2, Friday, Sep 8, 2017
- 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 PO, due Sunday, Sep 3, 2017
- Check that you were enrolled on Piazza

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!

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 enough 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"!

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

The Penalties are Severe

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

LET IT NOT BE YOU!

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 P1	Cheating on P1
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

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 September 3, 2017

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 SV 118)
 - 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)

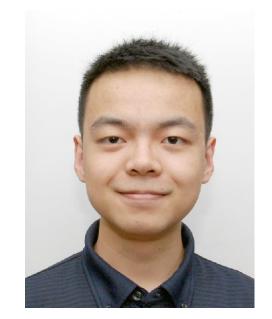


- Seth Goldstein
 - GHC 7111
 - <u>seth@cs.cmu.edu</u>
 - Office Hours
 - TBD

Teaching Staff



Cameron Dashti



MarshallAn



ChaskielGrundman

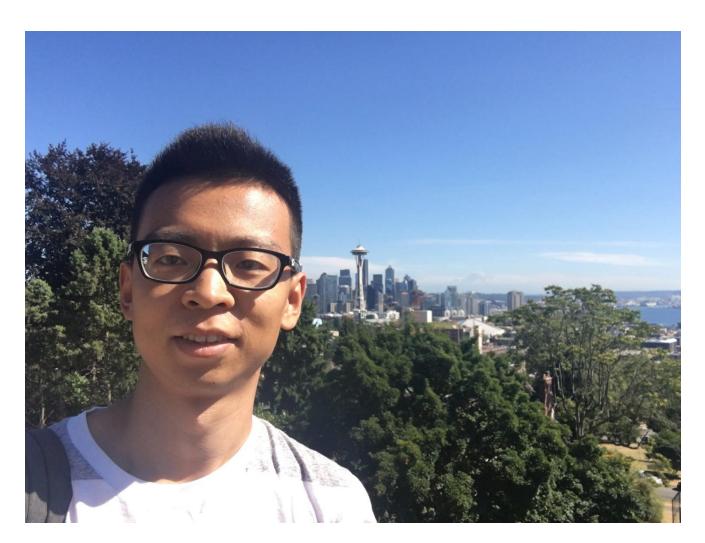
Mao Cai



Jimesh Chokshi



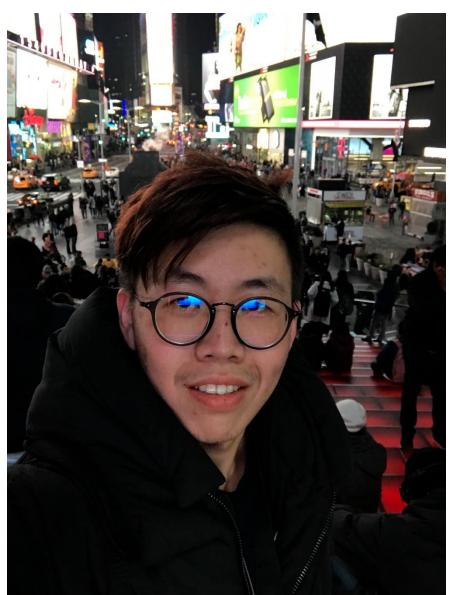
David Dai



Amanda Eng



Nick Guan



Yihan He



Ankit Jain



Harsha Konda



Zebing Lin



Yifang Liu



Sachin Menezes



Yiqun Ouyang



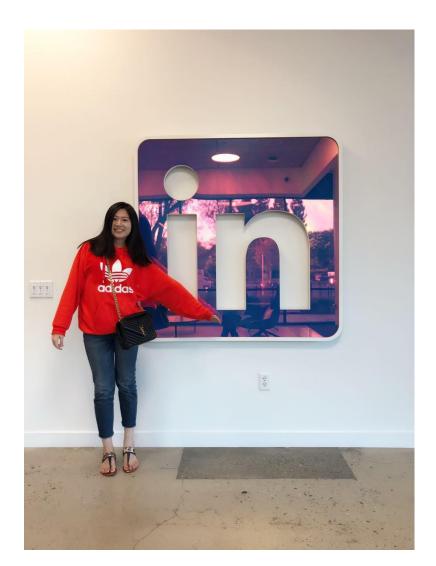
Peng Wang



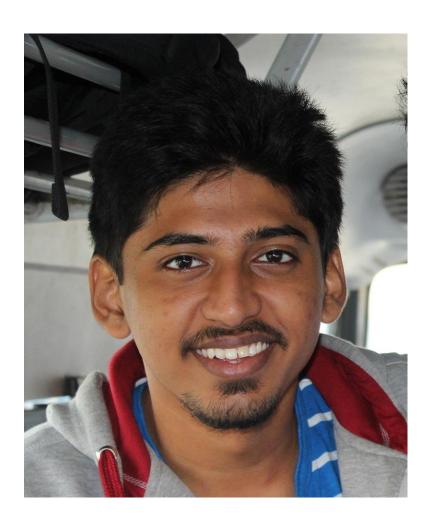
• Edam Li



Shiyue Liu



Shardul Mahadik



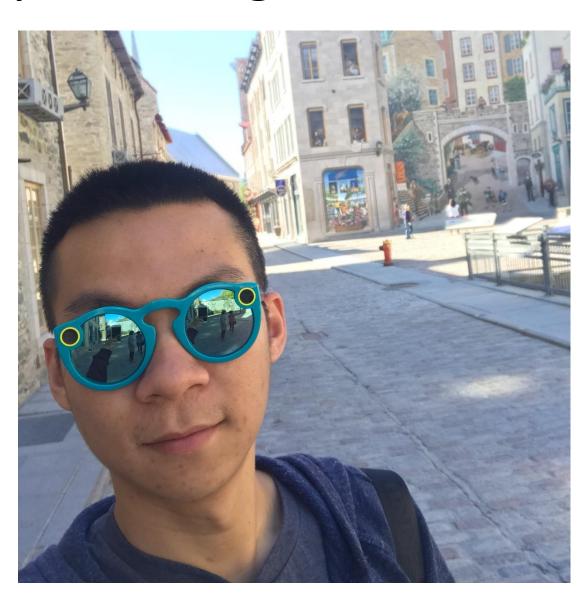
Imre Nagi



Jiupeng Sun



Henry Yang



Era of Globalization

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





Questions?

