CS15-319 / 15-619
Cloud Computing

Recitation 1

Course Overview and Introduction

September 1 & 3, 2015

http://www.cs.cmu.edu/~msakr/15619-f15/

Outline

- **What is the course about?**
- What is an online course?
- Administrivia
- A couple of demos

So What is Cloud Computing?

Data and Decision Making

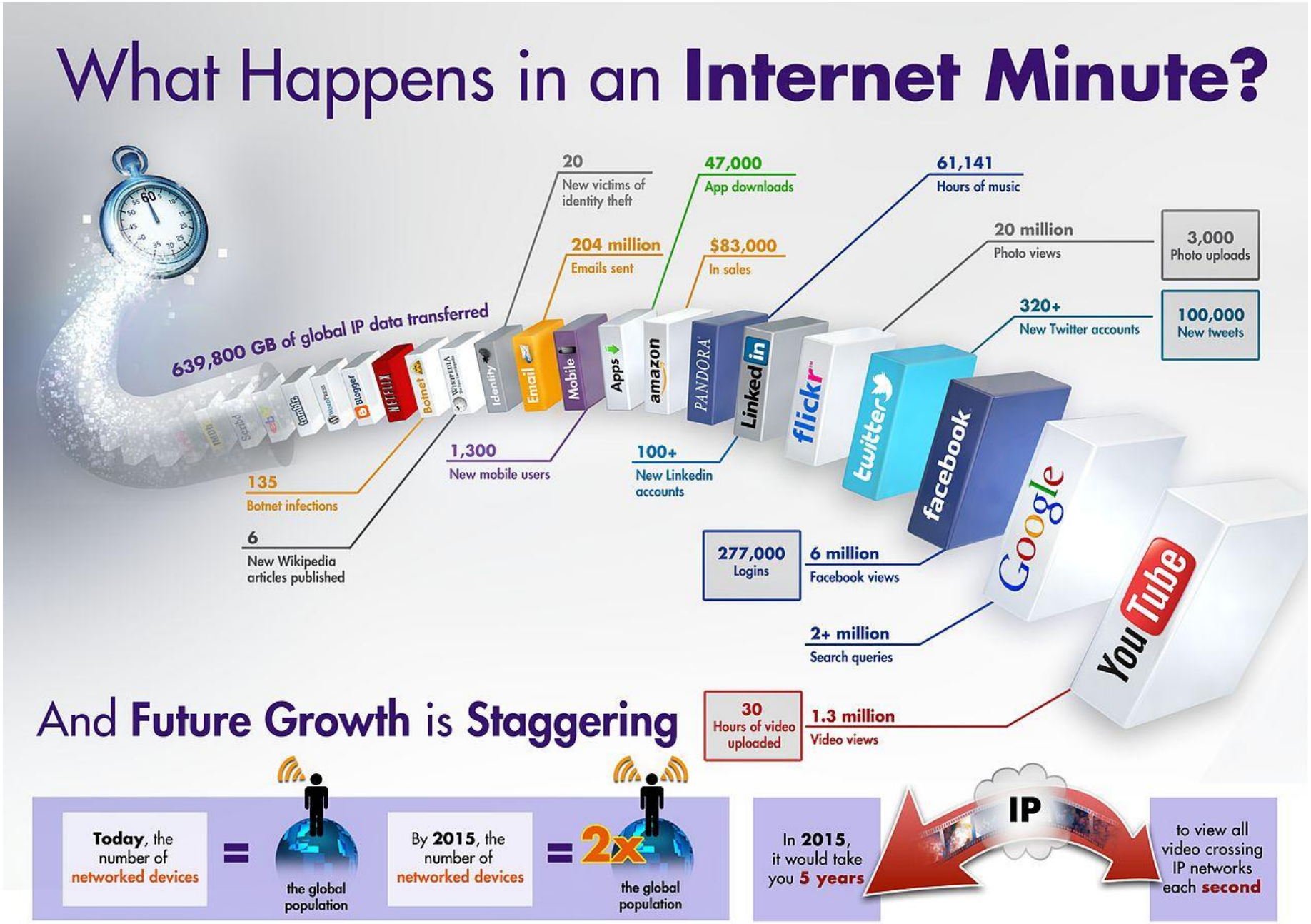
- Analyzing data reflects reality
- Walmart: hurricane warning
  - Stock beer and strawberry pop-tarts
  - 7x increase in sales during large storms
- Government: resource allocation decisions
  - Data mining in Maryland → crime hotspots
  - Shuffle resource allocation, more to hotspots
    - violent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

Data Science

- Extraction of knowledge 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



Source: IntelFreePress

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

– Traditional IT systems are insufficient

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"

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

The Cloud

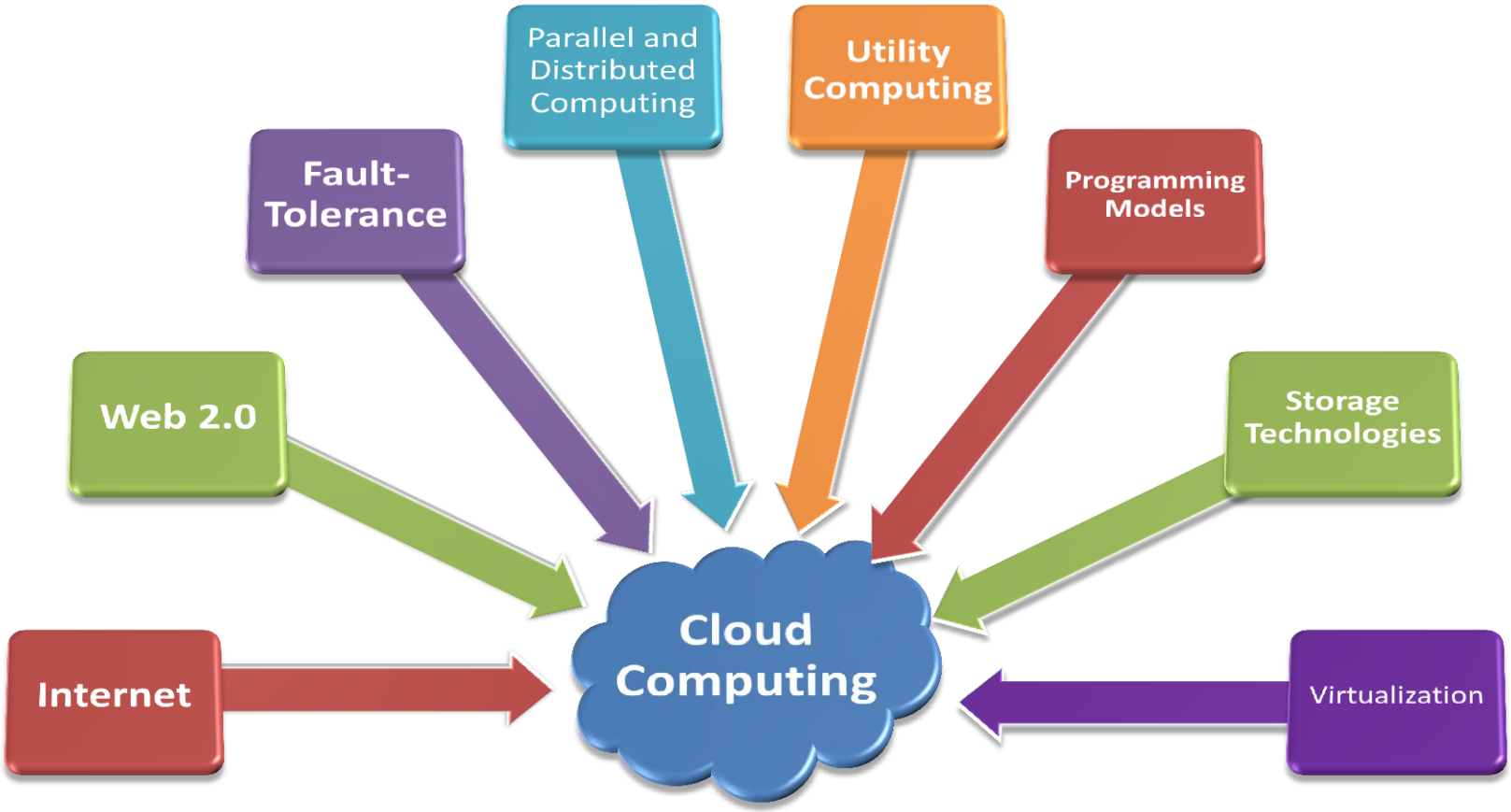
… for a more complete definition!

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

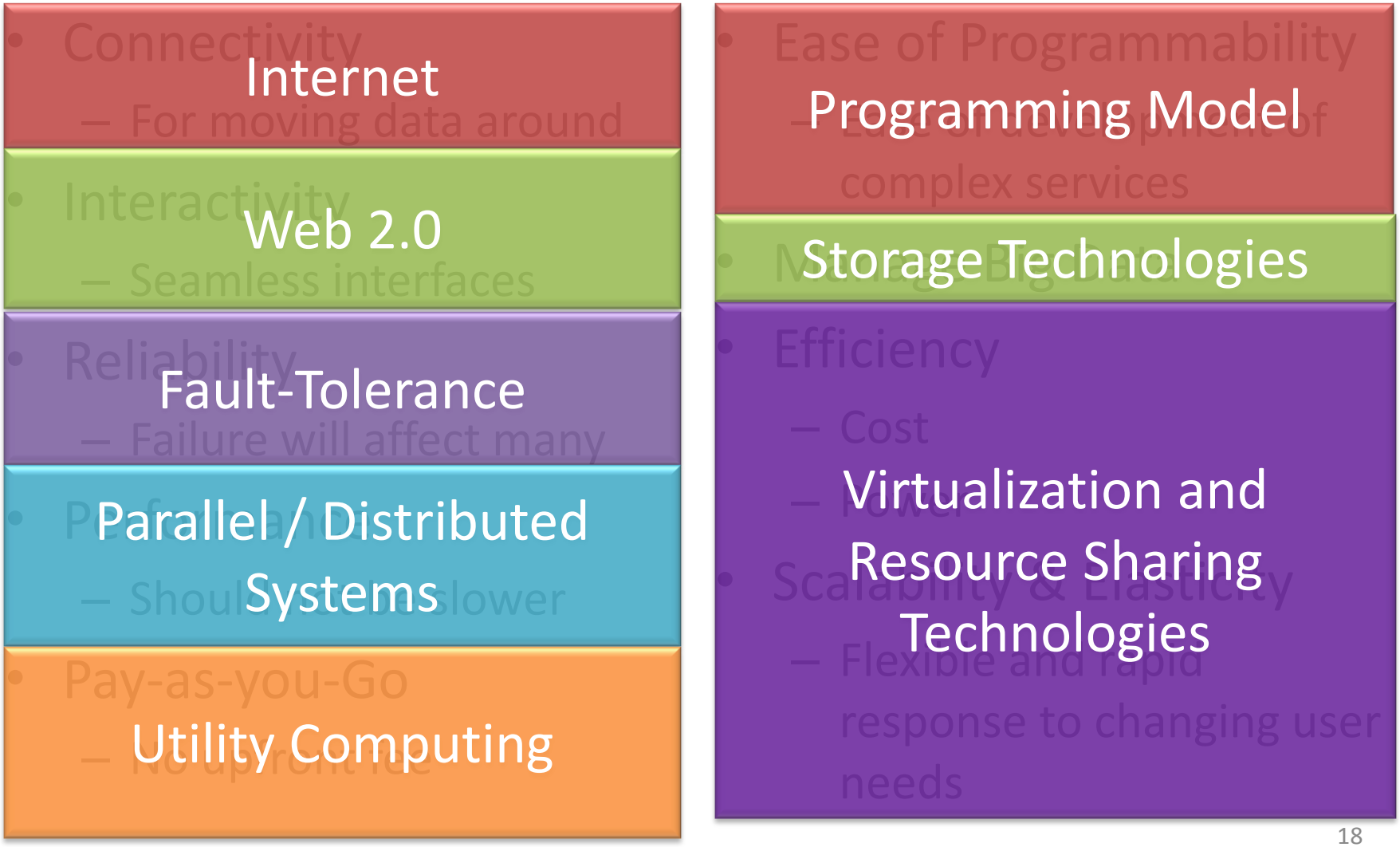


So… how would you transform information technology into a Service?

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

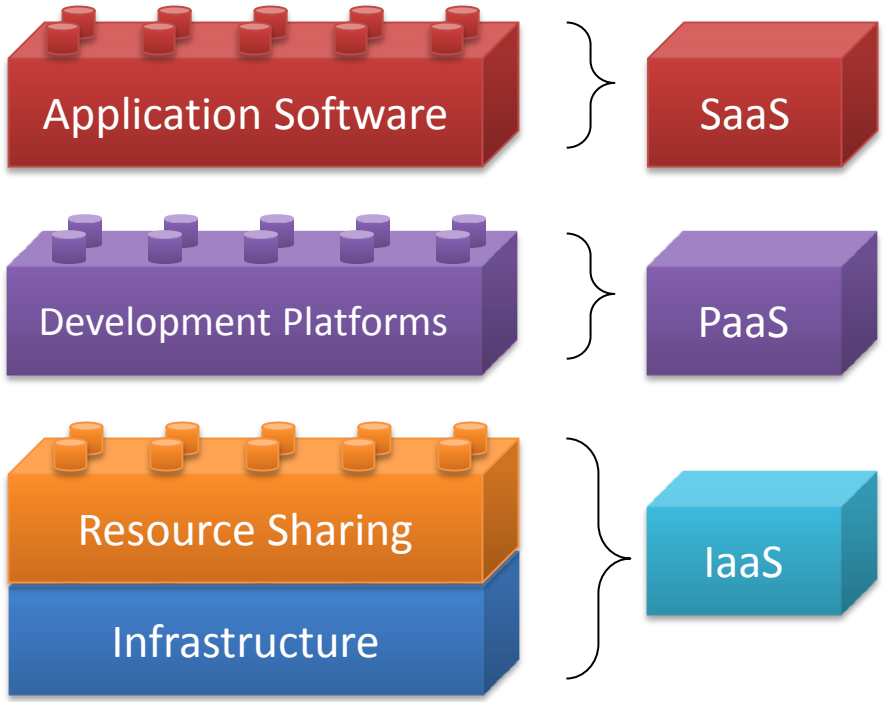
How to Transform IT to a Service?



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**)



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

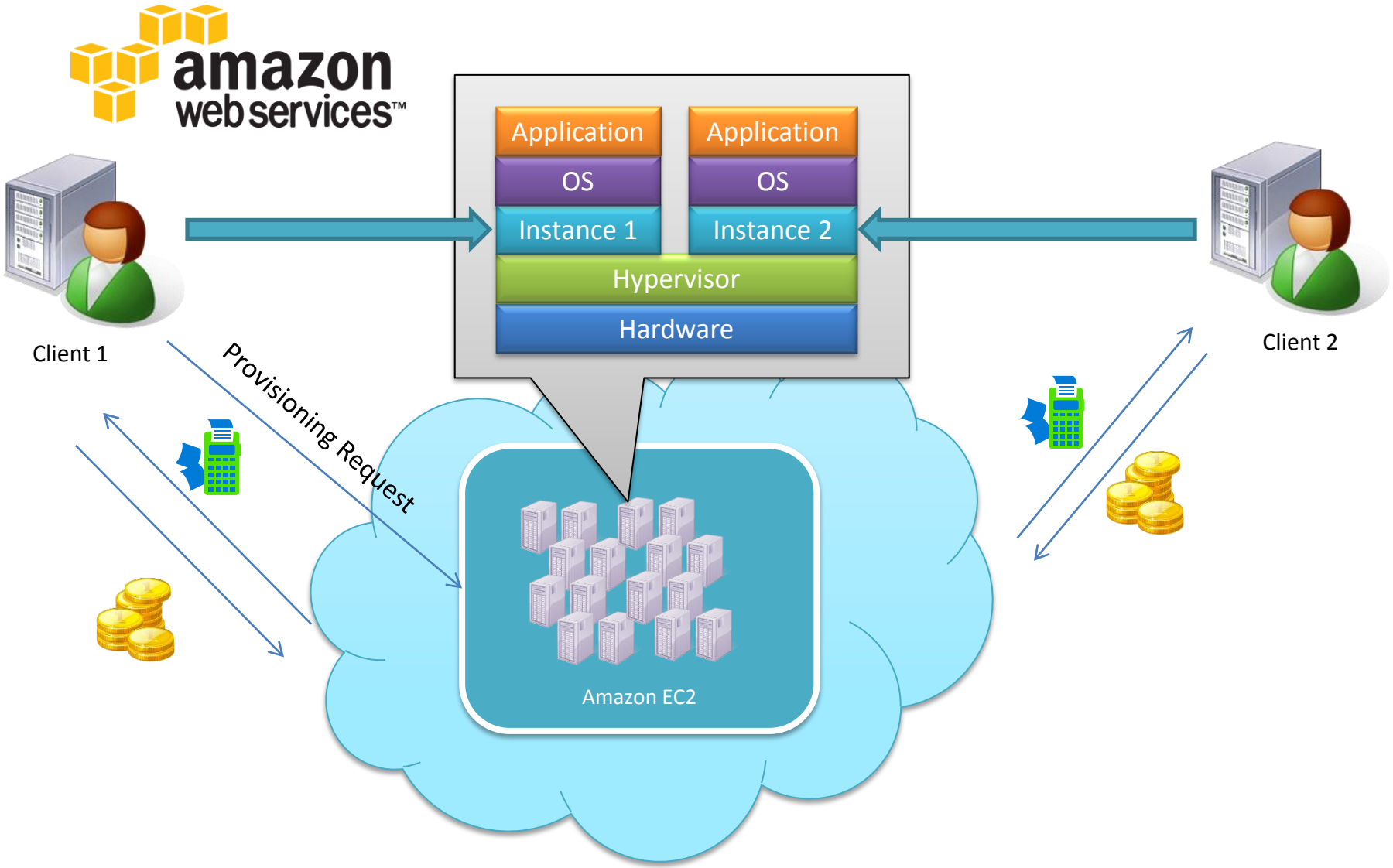
Platform as a Service (PaaS)

- 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** **o**ptions

Carbon **Footprint** decreased

Risks and Challenges

Migration

Security & Privacy

Vendor Lock-In

Legal

Internet Dependence

Service Level Agreements and Objectives (SLA/SLO)

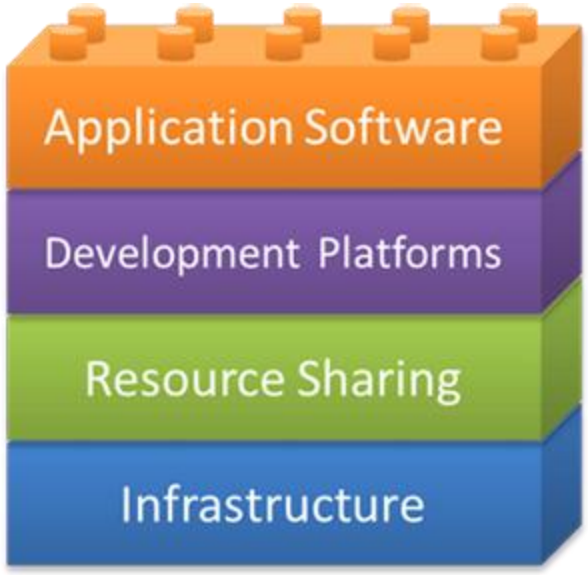
- SLA: Contract between cloud providers and users to define expected service
  - Service availability and delivery
  - Payment terms, bonuses and penalties for service
- SLO: Individual performance/service metrics regarding service delivery defined in the SLA
- Auditing: monitor resources to enforce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

Cloud Computing

- 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

Online content on OLI

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

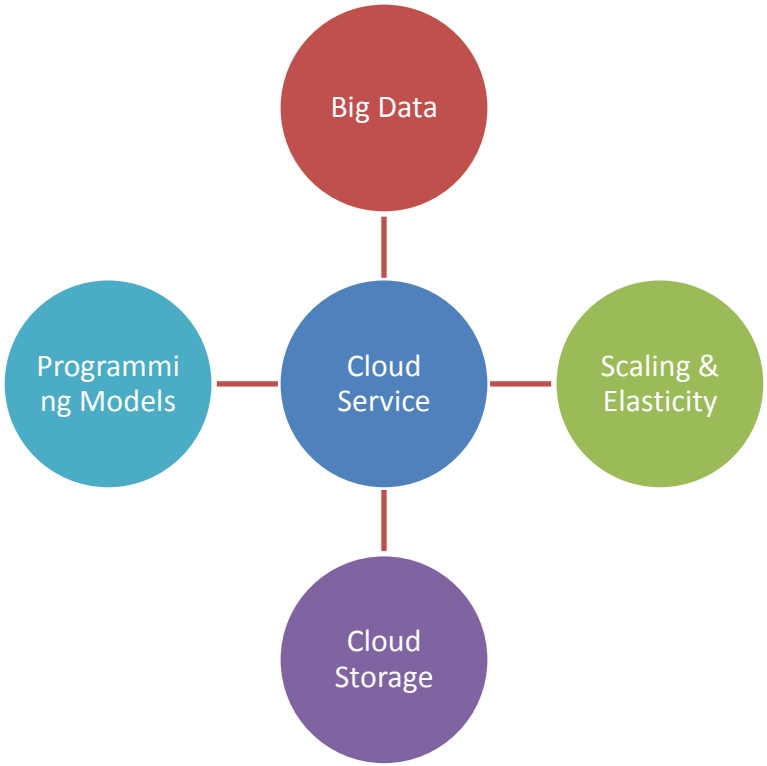
2.0 Data Centers
Components, design considerations and power

3.0 Resource Sharing
CPU, memory and I/O Virtualization

4.0 Cloud Storage
Distributed File Systems and Distributed Databases

5.0 Programming Models
MapReduce, Spark and GraphLab

Projects on AWS



Course Objectives

Students will learn:

- the fundamental ideas behind **Cloud Computing**;
- the basic ideas and principles in **data center** design and management; cloud software stack and 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.

Units 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 11th, 2015
2	Data centers	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 Analytics with GraphLab 2.0 (PowerGraph)

Projects

- Four Individual Projects (all students):
  0. Primers and P0 (Due Sunday, September 6, 2015)
  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

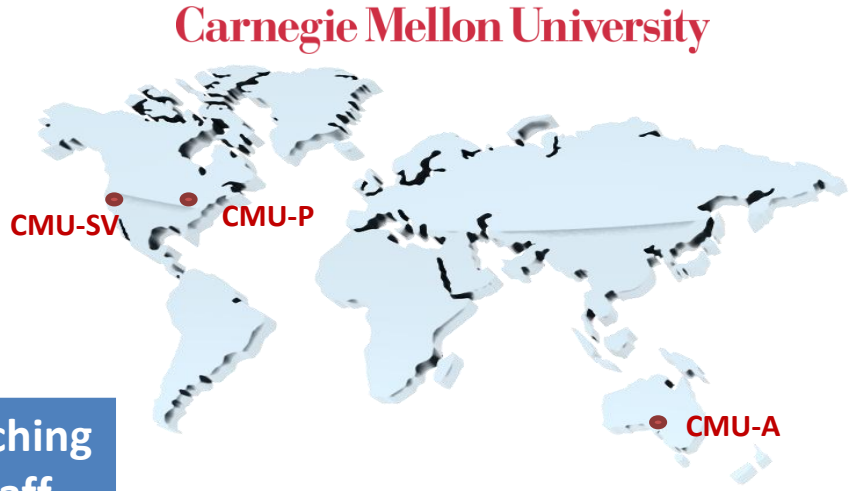
What this course is *not* about

- 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

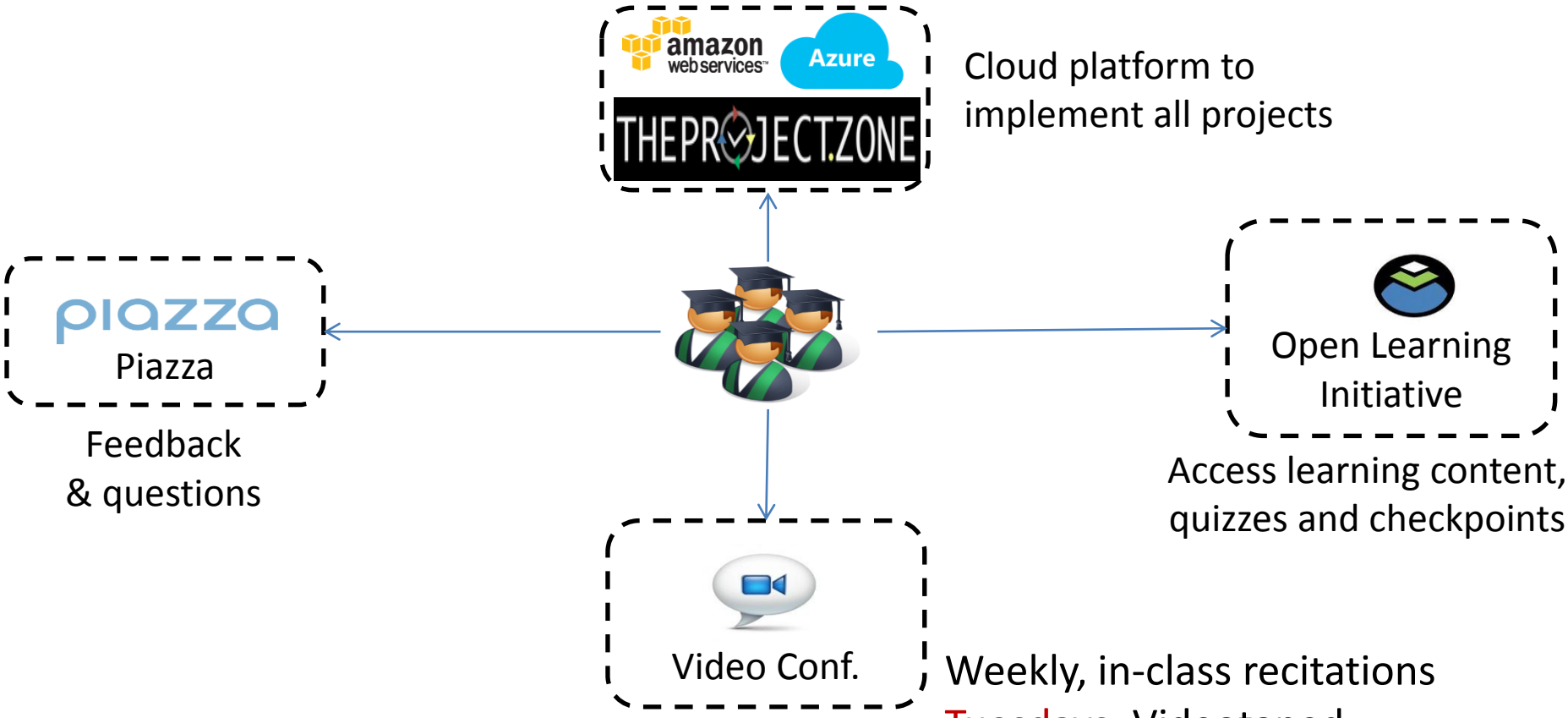
- What is the course about?
- **What is an online course?**
- Administrivia

Carnegie Mellon Global Course



Location	Sections	Students	Teaching Staff
CMU Pittsburgh	A & B	293	22
CMU Silicon Valley	C	84	7
CMU Adelaide	E	12	1

Online Course Engagement Model



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

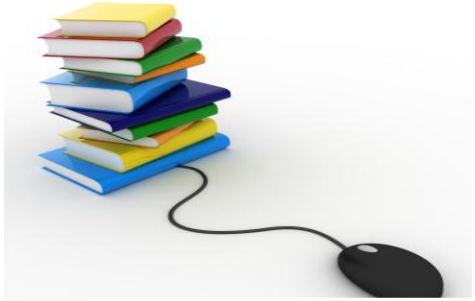
Outline

- 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 Organization



Course Units

Course Projects

Weekly Recitations

Office Hours

Discussions on Piazza

Getting Help

- TAs in Adelaide, Pittsburgh & Silicon Valley
- Piazza
  - Email does not scale
  - Discussion forum to support each other
- Recitations
  - Tuesdays (recorded)
    - At 8AM in GHC 4307 (GHC 4401 for first few weeks)
  - Thursdays (video conferenced to SV)
    - At 4:30PM in GHC 4307 (1:30PM in SV 212)
- Office Hours
  - Check Piazza for Office Hour schedule

Teaching Staff

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

Pittsburgh: Teaching Assistants

- Aaron Hsu

Pittsburgh: Teaching Assistants

- Chao Zhang

Pittsburgh: Teaching Assistants

- Daryl Zhang

Pittsburgh: Teaching Assistants

- Diane Zhang

Pittsburgh: Teaching Assistants

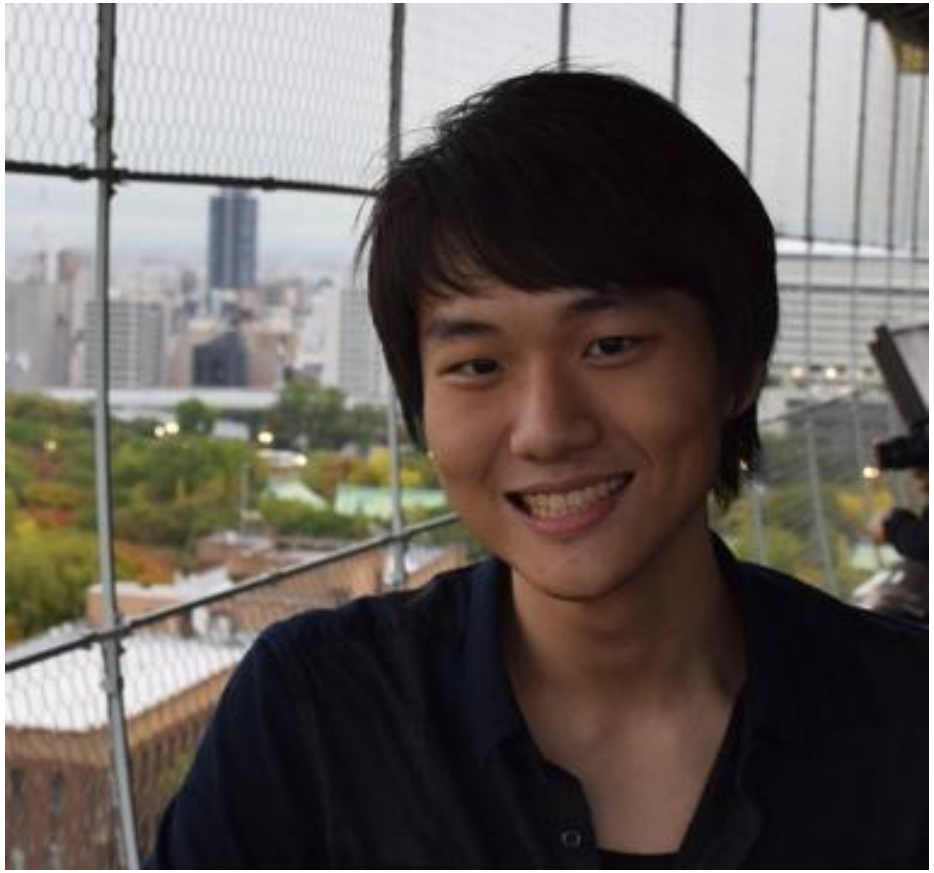
- Eryue Chen

Pittsburgh: Teaching Assistants

- Haoliang Quan

Pittsburgh: Teaching Assistants

- Jingbang Liu



Pittsburgh: Teaching Assistants

- Lee Yu

Pittsburgh: Teaching Assistants

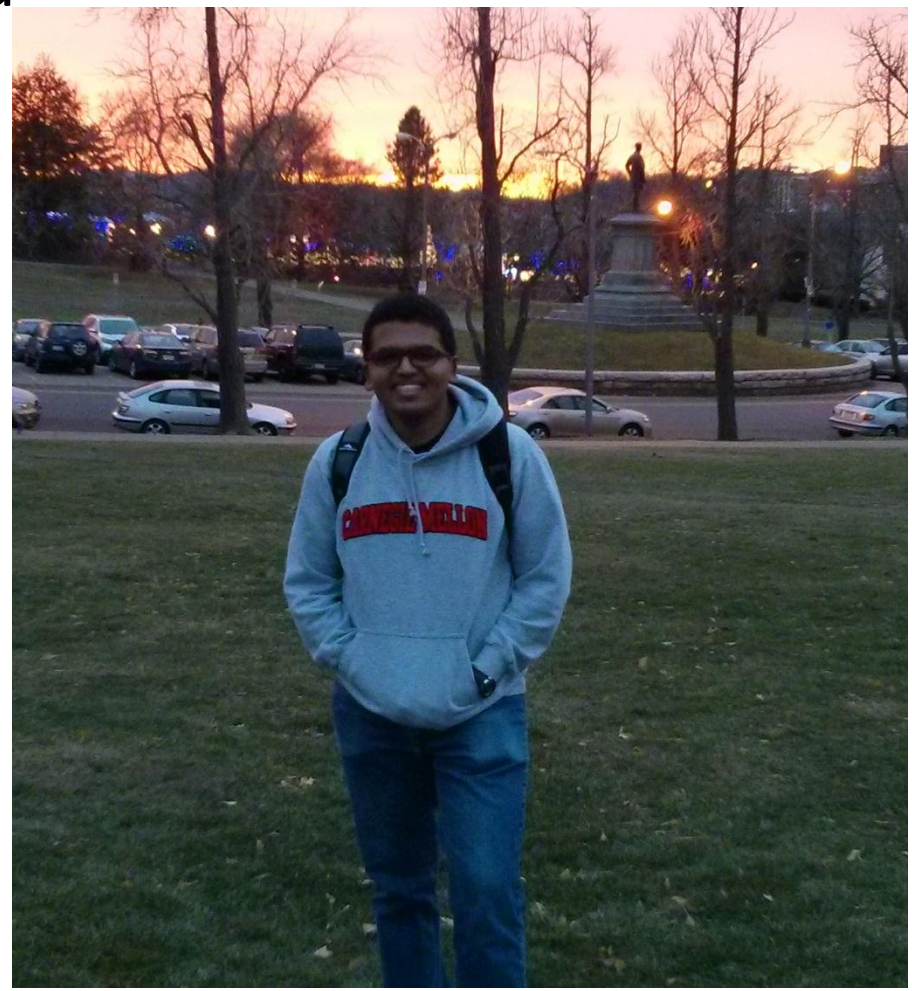
- Mengyu Yang (Rainy)

Pittsburgh: Teaching Assistants

- Mrigesh Kalvani

Pittsburgh: Teaching Assistants

- Prajwal Yadapadithaya



Pittsburgh: Teaching Assistants

- Rohit Upadhyaya

Pittsburgh: Teaching Assistants

- Ru Jia

Pittsburgh: Teaching Assistants

- Samarth Jain

Pittsburgh: Teaching Assistants

- Suhail Rehman

Pittsburgh: Teaching Assistants

- Tianqi Wen

Pittsburgh: Teaching Assistants

- Vikram Nair

Pittsburgh: Teaching Assistants

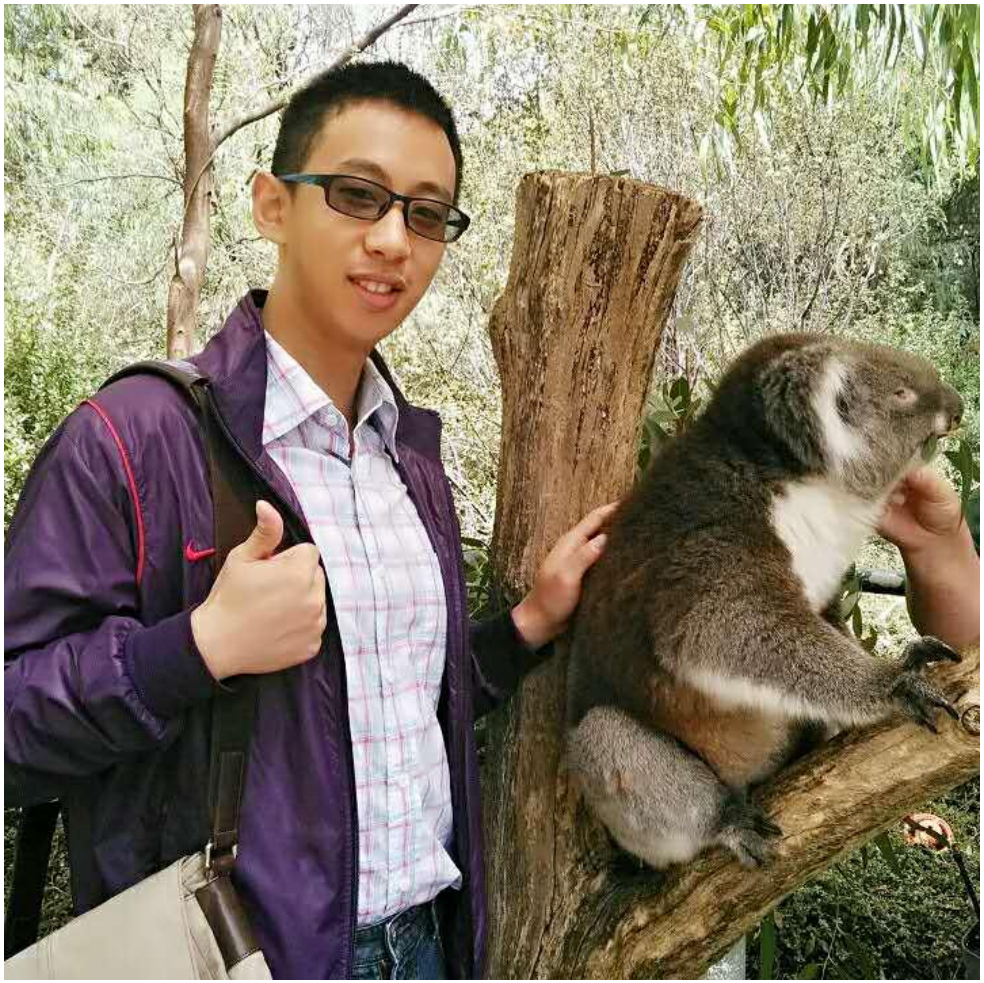
- Walid Baruni

Pittsburgh: Teaching Assistants

- Wei Luo

Pittsburgh: Teaching Assistants

- Yiming Zang

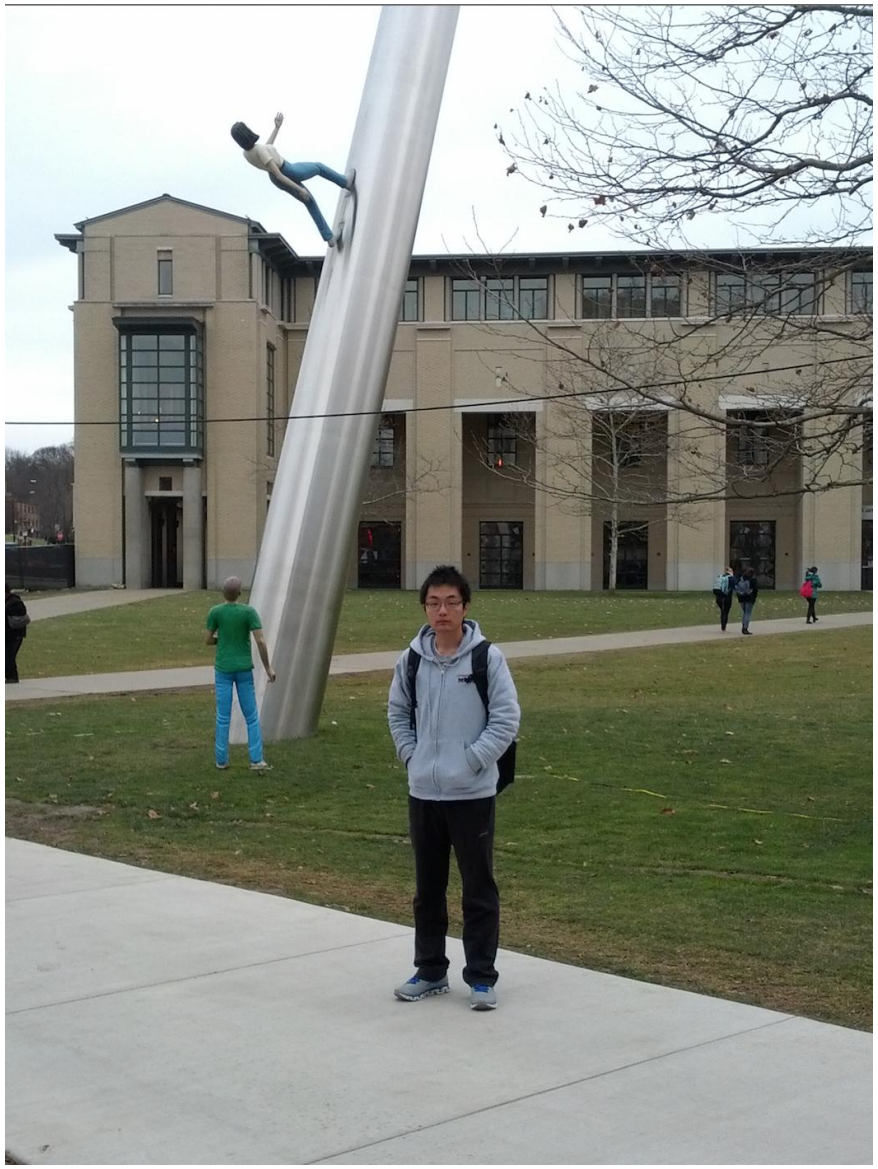


Pittsburgh: Teaching Assistants

- Yao Zhou

Pittsburgh: Teaching Assistants

- Zichang Feng



Silicon Valley: Teaching Assistant

- Abhishek Shivanna

Silicon Valley: Teaching Assistants

- Anshima Gupta

Silicon Valley: Teaching Assistant

- Chrysanthi Vandera

Silicon Valley: Teaching Assistants

- Mayank Singh Shishodia

Silicon Valley: Teaching Assistant

- Ozan Okumusog

Silicon Valley: Teaching Assistant

- Simba Tien

Silicon Valley: Teaching Assistant

- Yang Pan

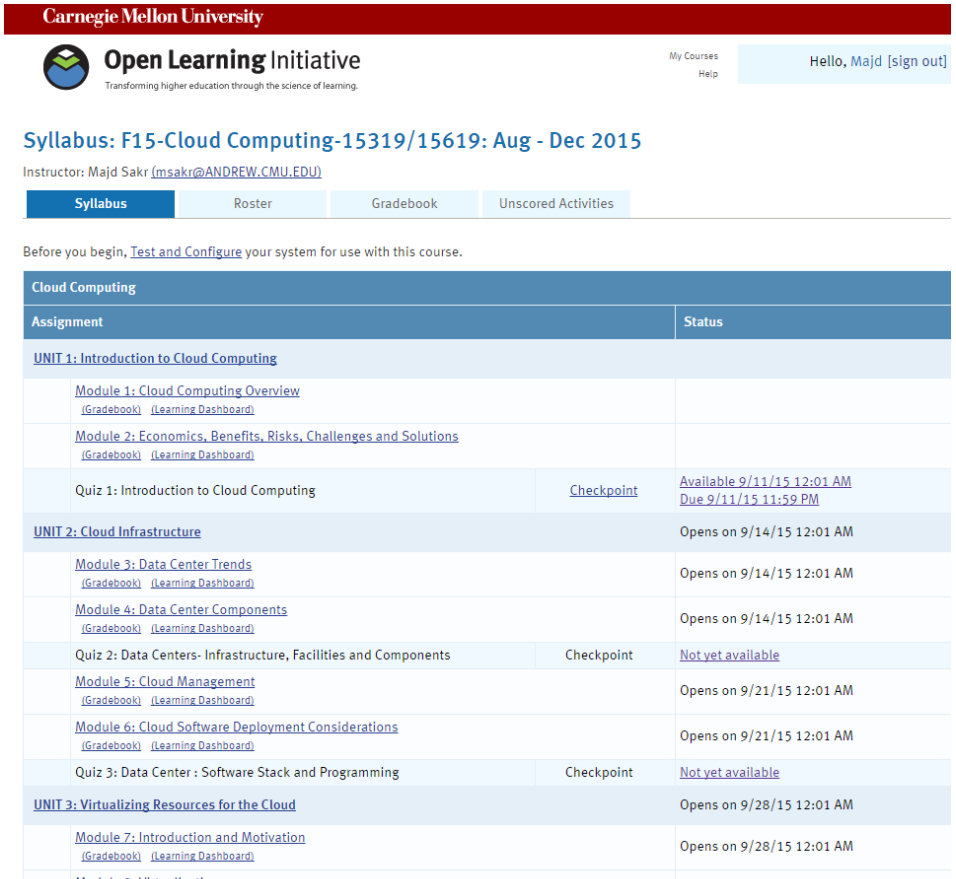
Adelaide: Teaching Assistant

- Lewis William Daly

Online Course Content - OLI

Course content is on the Open Learning Initiative:

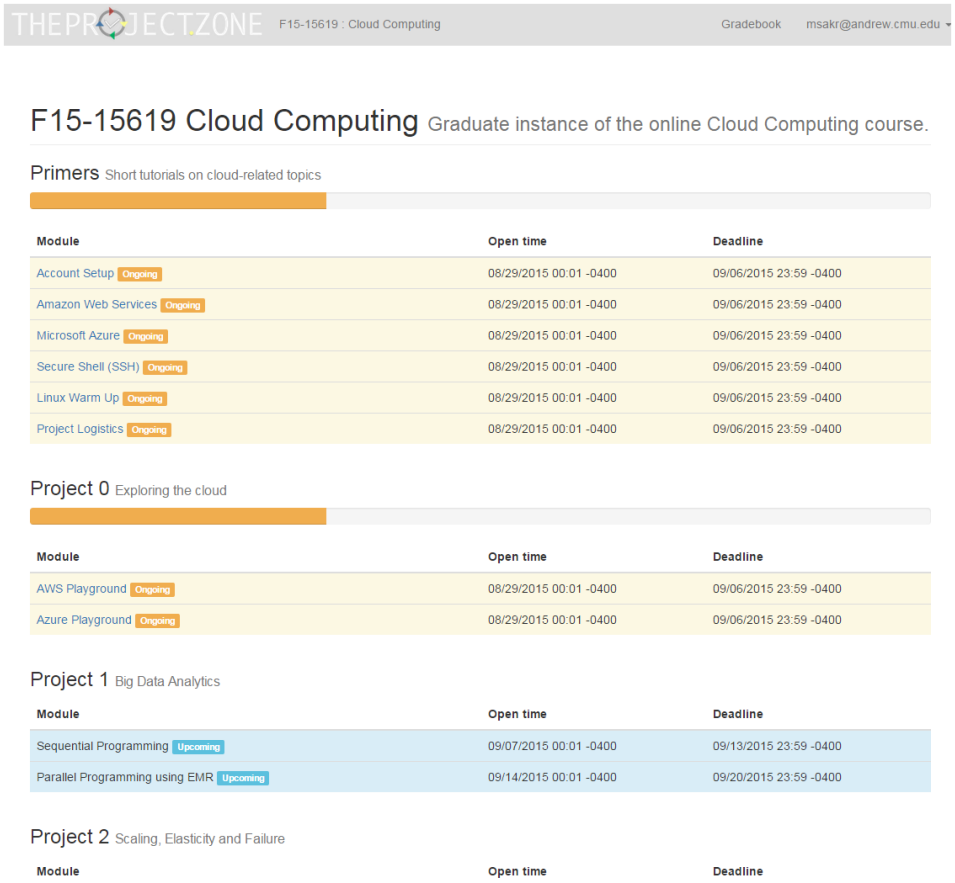
- Students are automatically registered
- Access to OLI is through Blackboard
  - blackboard.andrew.cmu.edu
- Check if Flash is installed
- Provide feedback on OLI
  - Bottom of each page
  - End of each module
- Do not copy or share content



TheProject.Zone

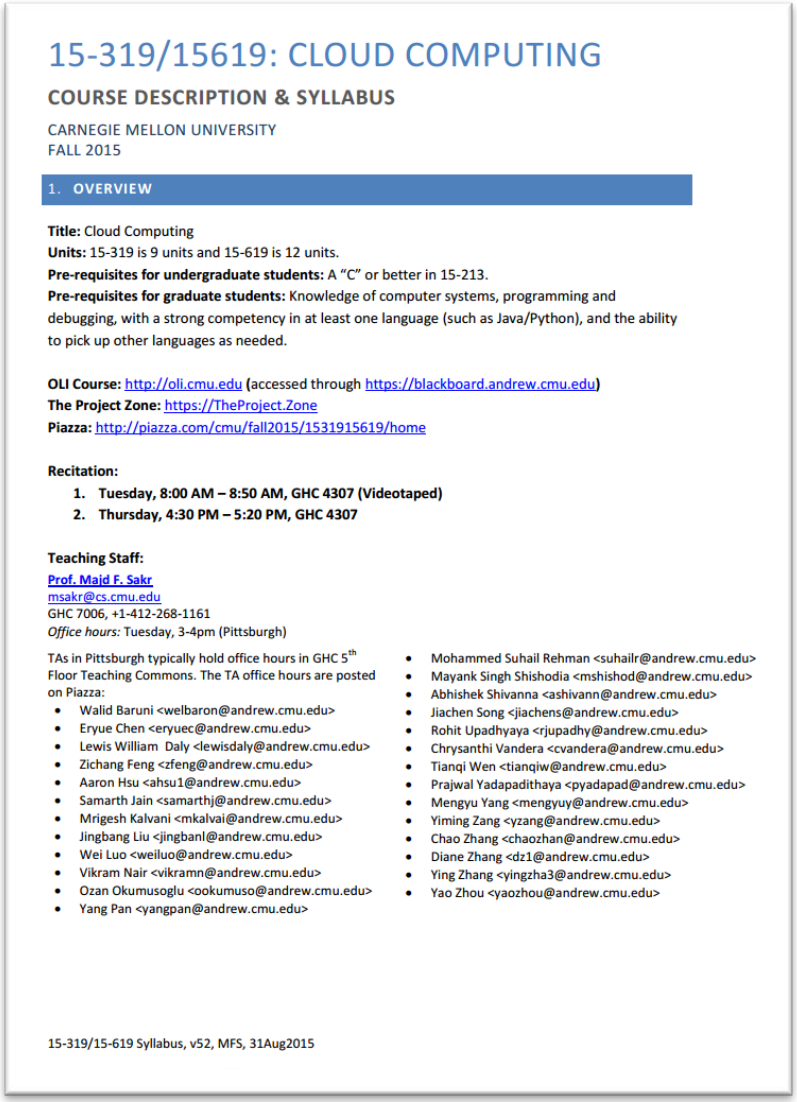
Course projects are on https://TheProject.Zone:

- 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
FALL 2015

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://blackboard.andrew.cmu.edu)
The Project Zone: https://TheProject.Zone
Piazza: http://piazza.com/cmu/fall2015/1531915619/home

Recitation:

1. **Tuesday, 8:00 AM – 8:50 AM, GHC 4307 (Videotaped)**
2. **Thursday, 4:30 PM – 5:20 PM, GHC 4307**

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:

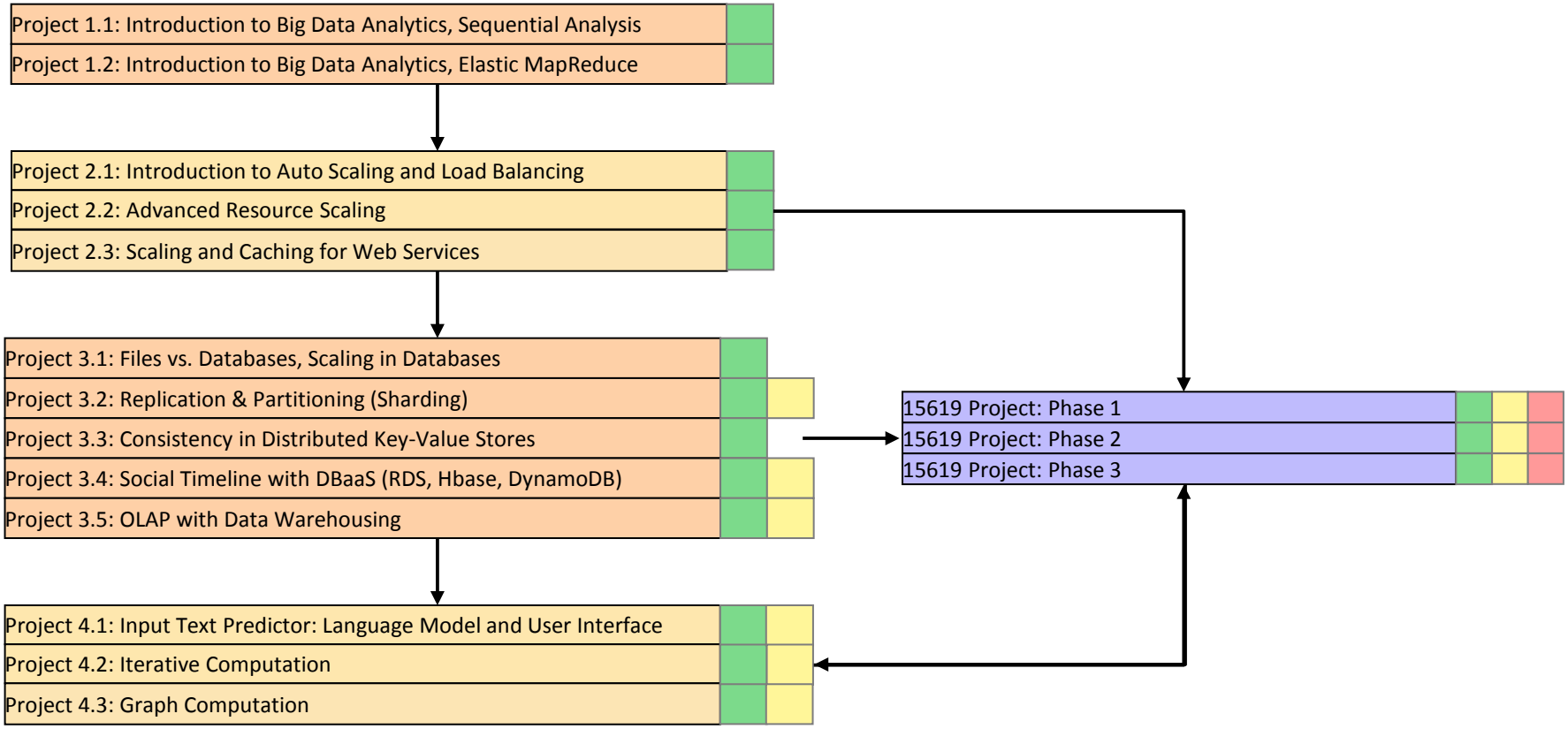
- Walid Baruni <welbaron@andrew.cmu.edu>
- Eryue Chen <eryuec@andrew.cmu.edu>
- Lewis William Daly <lewisdaly@andrew.cmu.edu>
- Zichang Feng <zfeng@andrew.cmu.edu>
- Aaron Hsu <ahsu1@andrew.cmu.edu>
- Samarth Jain <samarthj@andrew.cmu.edu>
- Mrigesh Kalvani <mkalvai@andrew.cmu.edu>
- Jingbang Liu <jingbanl@andrew.cmu.edu>
- Wei Luo <weiluo@andrew.cmu.edu>
- Vikram Nair <vikramn@andrew.cmu.edu>
- Ozan Okumusoglu <ookumuso@andrew.cmu.edu>
- Yang Pan <yangpan@andrew.cmu.edu>
- Mohammed Suhail Rehman <suhailr@andrew.cmu.edu>
- Mayank Singh Shishodia <mshishod@andrew.cmu.edu>
- Abhishek Shivanna <ashivann@andrew.cmu.edu>
- Jiachen Song <jiachens@andrew.cmu.edu>
- Rohit Upadhyaya <rjupadhy@andrew.cmu.edu>
- Chrysanthi Vandera <cvandera@andrew.cmu.edu>
- Tianqi Wen <tianqiw@andrew.cmu.edu>
- Prajwal Yadapadithaya <pyadapad@andrew.cmu.edu>
- Mengyu Yang <mengyuy@andrew.cmu.edu>
- Yiming Zang <yzang@andrew.cmu.edu>
- Chao Zhang <chaozhan@andrew.cmu.edu>
- Diane Zhang <dz1@andrew.cmu.edu>
- Ying Zhang <yingzha3@andrew.cmu.edu>
- Yao Zhou <yaozhou@andrew.cmu.edu>

Tentative Schedule

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

Week	Monday	OLI Content	Projects	15-619 Project	Quizzes
1	8/31/2015	Unit 1, Module 1	Primers/P0 (Sep 6)		
2	9/7/2015	Unit 1, Module 2	P1.1 (Sep 13)		Q1 (Sep 11)
3	9/14/2015	Unit 2, Module 3, 4	P1.2 (Sep 20)		Q2 (Sep 18)
4	9/21/2015	Unit 2, Module 5, 6	P2.1 (Sep 27)		Q3 (Sep 25)
5	9/28/2015	Unit 3, Module 7, 8, 9	P2.2 (Oct 4)		Q4 (Oct 2)
6	10/5/2015	Unit 3, Module 10, 11, 12	P2.3 (Oct 11)		Q5 (Oct 9)
7	10/12/2015	Unit 3, Module 13	P3.1 (Oct 18)	Project Out (Oct 12)	Q6 (Oct 16)
8	10/19/2015	Unit 4, Module 14	P3.2 (Oct 25)		Q7 (Oct 22)
9	10/26/2015	Unit 4, Module 15	P3.3 (Nov 1)	Phase 1 Due (Oct 28)	Q8 (Oct 30)
10	11/2/2015	Unit 4, Module 16, 17	P3.4 (Nov 8)		Q9 (Nov 6)
11	11/9/2015	Unit 5, Module 18	P3.5 (Nov 15)	Phase 2 Due (Nov 11)	Q10 (Nov 13)
12	11/16/2015	Unit 5, Module 19	P4.1 (Nov 22)		Q11 (Nov 20)
13	11/23/2015	Thanksgiving			
14	11/30/2015	Unit 5, Module 20, 21	P4.2 (Dec 6)	Phase 3 Due (Dec 2)	Q12 (Dec 4)
15	12/7/2015		P4.3 (Dec 11)		

Projects: Timeline and Dependencies



Grading

Course Elements	#	Weight
Projects	4 or 5	75%
OLI Unit Checkpoint Quizzes	12	25%

- All projects are equal weight
  - 18.75% for 15-319
  - 15% for 15-619
- Weekly quizzes (12 in total) are equal weight
  - ~2% for each quiz

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.

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 one	Cheating on one
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%

Course Administration

- Students are automatically registered on OLI through blackboard.andrew.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

Special Note on Amazon EC2

- 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

This Week

- Become familiar with content on OLI
  - Start reading Unit 1, Module 1
  - Quiz 1: Unit 1, Module 1 &2, Friday, September 11th, 2015
- Projects on TheProject.Zone
  - Primer and P0, due Sunday, September 6th, 2015
- Check that you were enrolled on Piazza
- Create an account on AWS and Azure (ASAP)
  - Submit your AWS account info using the link provided in the primers on TheProject.Zone

Questions?