14-848 Cloud Infrastructure

SPRING 2024

MOHAMED FARAG

FARAG@CMU.EDU

Carnegie Mellon University

Agenda

- Welcome and Introductions

- Focus Areas of this Course

- Course Syllabus & Schedule

- Class Expectations

- Introduction to the Cloud

- Next Steps

Why is this course Important?

In recent years, the demand for Cloud Computing and Infrastructure jumped significantly.

- **"The global cloud computing market size is projected to grow from $677.95 billion in 2023 to $2,432.87 billion by 2030", Fortune Business Insights.**

- **The most in-demand hard skills of 2023 lists Cloud Computing among the top 10 hard skills.**

Why is this course Important? – Cont'd

The course touches on several topics on the Cloud Computing Hype Cycle, 2023 released by Gartner R&D
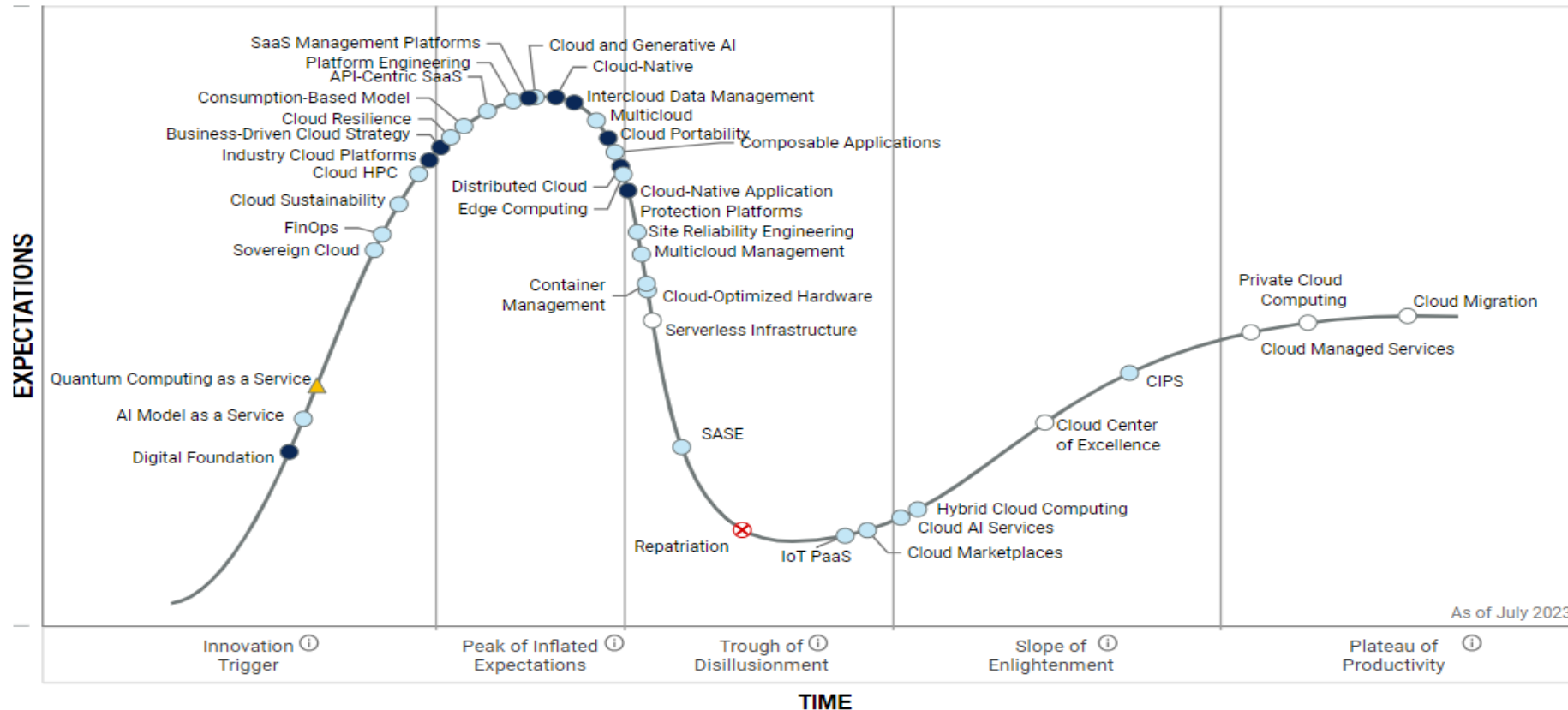
Focus Areas of the Course

Cloud Technologies

Cloud Providers:
- Amazon Web Services
- Google Cloud Platform
- Microsoft Azure

Big Data Processing Platforms:
- Apache Hadoop
- Apache Spark

NoSQL Database:
- DynamoDB and BigTable

Large Language Models:
- Vertex AI

DevOps

Deployment Scaling and Orchestration:
- Docker
- Kubernetes

Infrastructure-as-a-Code
- Terraform

Cloud Infrastructure
- Metaverse
- Data Centers
- Edge Computing & Fog Computing

Why Do We Care about Cloud Computing?

Expectations for Incoming Students

- ***You are expected to know Python or are willing to learn it.***

  - A Python recording will be released next week for members who need support with Python

- ***You are expected to have a basic understanding of Computer Virtualization.***

  - If you don't know about virtualization, check out this reading: https://www.vembu.com/blog/physical-server-vs-virtual-machine-choice-open/

Instructor and TA Introductions

Instructor:

- Mohamed Farag: farag@cmu.edu

TAs:

- Shuhuai Lin: shuhuail@andrew.cmu.edu
- Tian Xie: tianxie@andrew.cmu.edu
- Shubham Kachroo: skachroo@andrew.cmu.edu

Course Logistics

- Lectures are offered in-person only, but recordings will be made available after the lectures.

  - Please allow for some lead-time in the beginning of the semester for the recordings to be released.

- Lecture slides are delivered via TopHat during the lecture and will be posted on Canvas under Modules section. Sign up for a free TopHat account and join the course with the following code: **177225**

- Starting from next lecture, each lecture will have an in-class quiz.

- Students who have approved accommodation shall contact the course instructor to figure out how the instructor can meet their needs

Course Logistics – Office Hours (Zoom/In-Person)

Days/Timeframes	11:30am-12:30pm ET	12:30-1:30pm ET	2:30-3:30pm ET	4-5pm ET	5-6pm ET	9-10pm ET
Monday		Shuhuai				
Tuesday	Tian	Shubham				Shubham (only by appointment)
Wednesday	Mohamed	Shuhuai		Tian (In-person PIT) / Shuhuai (In-person SV)		Shubham
Thursday	Shubham				Shubham (In-person PIT) / Shuhuai (In-person SV)	Tian
Friday	Tian					

(orange)	Instructor Office Hours - Conducted remotely via Zoom - URL can be found on Canvas
(blue)	TA Office Hours - Conducted remotely via Zoom - URL can be found on Canvas
(gray)	In-Person OHs held at PIT in the INI Project Room 205 - 4616 Henry St., Student Floor
(gray)	In-Person OHs held at SV in room 123 on the first floor of building 23

- All Office Hours will use the same Zoom URL:
  - https://cmu.zoom.us/j/92004227250?pwd=YnpiTW1qSXdscWRvbWYzd2t0elptQT09
- If you have short questions and you don't want to wait in the Zoom room for extended time, please book a 15-min discussion via this URL (limited slots): https://calendly.com/skachroo/cmu-14848-ta-oh-slots

Carnegie Mellon University

Course Logistics – Piazza Hours

	12:30-1:30pm ET	1:30-2:30pm ET	4-5pm ET	5-6pm ET	7-8pm ET	9-10pm
Piazza OHs						
Monday			Tian		Tian	
Tuesday	Tian			Shuhuai		
Wednesday		Shuhuai			Tian	
Thursday	Shubham					Shubham
Friday	Shubham					Shubham
Saturday	Shuhuai					

- Please note that the team will be spending the allocated timeslot to answer as many questions as they can from Piazza.

Carnegie Mellon University

Course Logistics – Office Hours - Cont'd

- Use Course Piazza to ask asynchronous questions that require instructor and/or TA help

- Use the Student Space Slack Channel to find a teammate for your course project (No instructor or TA help is offered there)

Course Assessment –
Final Exam on April 25th during the lecture

Final Exam	Project	Assignments	Quizzes
15%	20%	40%	25%

Final Exam: is an open-note test.
- o Students will have access to all the **PDFs for lectures, readings and HW solutions**. Students can **bring any hard-copied materials with them**.
- o Students are required to follow the schedule of their registered section. **On the scheduled final lecture of each section, final exam will be released only to the registered students of the corresponding section**. Each section will have its final exam version(s).
- o Exam will be offered via **Lockdown Browser and no knowledge exchange is allowed among students during the exam.**
- o Students are expected to install and test Lockdown browser on their machines ahead of the exam. If students face an issue with Lockdown browser installation, students must reach out to the instructors **no later than 2 weeks** before the final exam date.
- o **Sharing hard-copied notes is prohibited during the exam.**

Course Assessment – Cont'd – Final Exam Waiver

Final Exam	Project	Assignments	Quizzes
15%	20%	40%	25%

You can get full score on the final exam if you obtain TWO of the following certifications two weeks before the final exam:

Google Cloud: Associate Cloud Engineer

AWS Certified Solutions Architect – Associate

Microsoft Certified: Azure Developer Associate

GCP Professional Data Engineer

AWS Certified Data Engineer - Associate

Microsoft Azure Data Engineer Associate

Certified Kubernetes Application Developer

Late acquisition of these certificates beyond the deadline won't be accepted. Students who would like to waive their final exam must upload a proof of obtaining two certificates two weeks prior to the exam.

Late submissions are not accepted. Please plan ahead as no excuses will be provided if the only available exam slots are beyond the deadline.

Obtaining only 1 certificate won't provide any partial benefit on the final exam score

Carnegie Mellon University

Course Assessment – Cont'd

Final Exam	Project	Assignments	Quizzes
15%	20%	40%	25%

- **Homework Assignments:** there will be 7 homework assignments provided throughout the semester covering the practical aspects of the class. There will be good learning curve that students will have to take on their own.

- Students will have 3 days to submit the assignment after the due date with a late penalty. Late penalties are applied based on the timestamp of the last code commit on GitHub and it will follow this equation (no matter whether the delay is in minutes or in hours):
  - 5 points for up to 24 hours delay
  - 15 points for the next 24 hours delay
  - 25 points for the next 24 hours delay
  - 100 points penalty (no grade) after this time.

After homework grades are released, a Canvas announcement will be made with a link to submit regrade requests. Regrade requests can be made for 24 hours via the URL that is provided on the Canvas announcement and CANNOT be submitted via email.

Course Assessment – Cont'd

Final Exam	Project	Assignments	Quizzes
15%	20%	40%	25%

- **Course Project:** Each student will have the option to peer with a team member for the project and you will choose one of two project options to submit. This project leverages most of the topics and practices that are covered throughout the semester. Course details are released in Week-3. Project submission deadline is April 4th 11:59PM ET.

- **Quizzes:** there will be 1 quiz published on Canvas after each lecture with a specific access code. The access code will be revealed during the lecture to the registered students of the corresponding section.
  - Quizzes will start next lecture.
  - You will receive two excused absences from Quizzes for emergencies, sickness, etc.
  - If you need more time, get an approval from your faculty advisor (your professor and not the administrative person)

Course Grade Scheme

+/- are used to provide granularity in equal intervals in B and C ranges

Grade	Percentage Interval
A/A-	[85-100%], A starts from 93
B	[70-85%)
C	[55-70%)
D	[40-55%)
R (F)	Below 40%

Carnegie Mellon University

Course Schedule

Date	Topic	Notes
Week-1	- Introduction & Syllabus - Virtualization Basics	- System Setup homework released
Week-2	- Containerization - Guest Speech	- System Setup homework deadline. - Docker homework released
Week-3	- Containerization Demo - Deployment Orchestration - Pokémon Go Case Study	- Course Project released
Week-4	- Deployment Orchestration Demo - Cloud Computing Concepts	- Docker homework deadline - Kubernetes homework is released
Week-5	- Cloud Computing Paradigms - Cloud Data Storage	
Week-6	- NoSQL Database: DynamoDB - Infrastructure-as-a-Code	- Kubernetes homework deadline - DynamoDB Homework released
Week-7	- Infrastructure-as-a-Code (Cont'd) Terraform	- DynamoDB Homework deadline - Course Project Checkpoint
Spring Break		

Course Schedule – Cont'd

Week-8	- Introduction to Hadoop - Hadoop HDFS	- Terraform Homework released
Week-9	- Hadoop MapReduce	- Terraform Homework deadline - Hadoop MapReduce homework is released.
Week-10	- Hadoop MapReduce Cont'd - Big Data Algorithms	- Hadoop MapReduce homework deadline
Week-11	- Introduction to Spark	- Course project submission
Week-12	- Spark (Cont'd)	- Apache Spark homework released
Week-13	- Cloud Security & Privacy - Edge Computing and Fog Computing	- Apache Spark homework deadline
Week-14	- Introduction to running LLMs on the Cloud using Vertex AI - Final Exam	

Expectations down the Road!

- For some cloud platforms like Azure and AWS, you will have to use the Free Trial version which includes a step for you to add your Credit Card information. You will not be billed if you used the cloud platform correctly.

- For the final exam, we will use the Lockdown browser and you will have access to the lecture PDFs via the browser. You will receive few training attempts on the final exam environment before the exam date.

Course Delivery and HW Notes

- Lecture materials will be released on Canvas prior to the lecture.

- Annotations will be added on the slides while playing them on TopHat but you won't be able to download the TopHat slides.

- All HW assignments (starting from HW-2) will be submitted via GitHub classroom.

- First HW assignment focuses on System Setup. It's released on Canvas and you can submit it until Thursday September 7th, 11:59PM ET.

Other Syllabus Information

- Please read the remaining sections of the Course Syllabus.
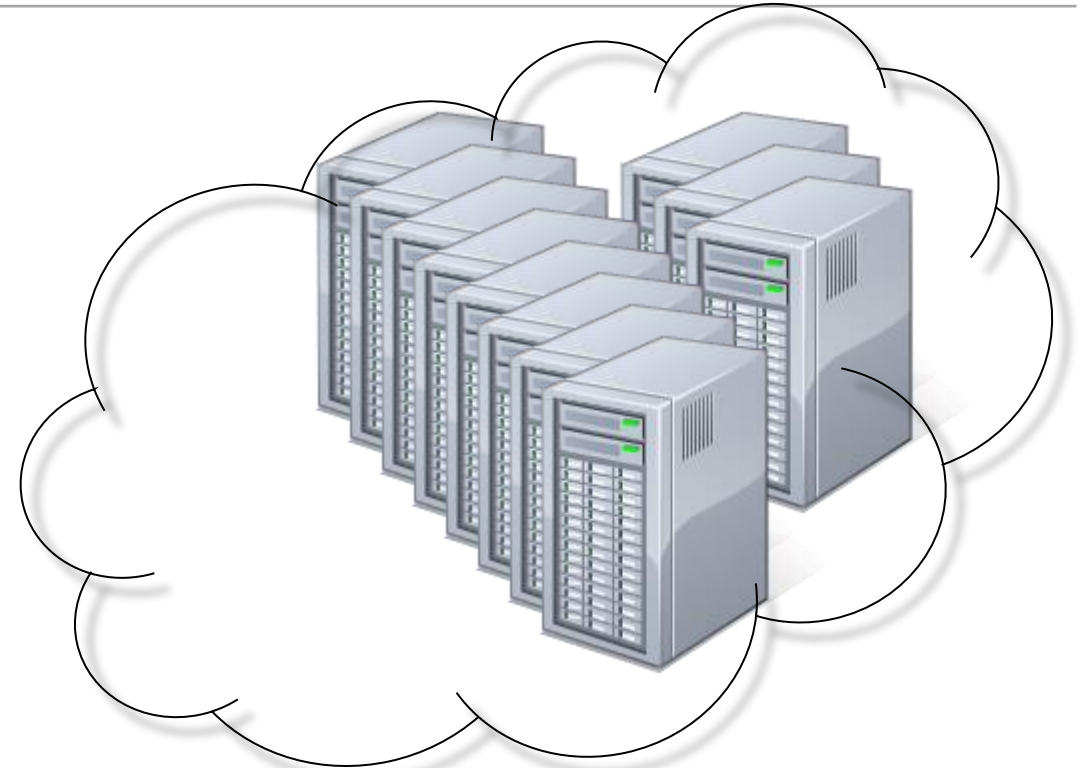- The Syllabus can be found on Canvas under the Modules section

Carnegie Mellon University

What is Cloud?

Carnegie Mellon University

What is the Cloud?

- It's a cluster!
- It's a supercomputer!
- It's a datastore!
- It's superman!

- None of the above?

- All of the above?

- In simple words:
  - Cloud = Lots of storage + compute cycles nearby + network bandwidth

Cloud Computing – NIST Definition

A *model* for enabling convenient, on-demand network access to a *shared* pool of *configurable computing* (e.g., servers, storage, applications, and services) and networking resources that can be *rapidly provisioned* and *released* with *minimal management* effort or service provider interaction

What is Cloud Computing in the REAL WORLD?

"Cloud" refers to large Internet services running on 10,000s of machines (Amazon, Google, Microsoft, etc)

"Cloud computing" refers to services by these companies that let external customers rent cycles and storage

- Amazon EC2: virtual machines at 8.5¢/hour
- Amazon S3: storage at 21¢/GB/month
- Google Cloud AppEngine
- Windows Azure
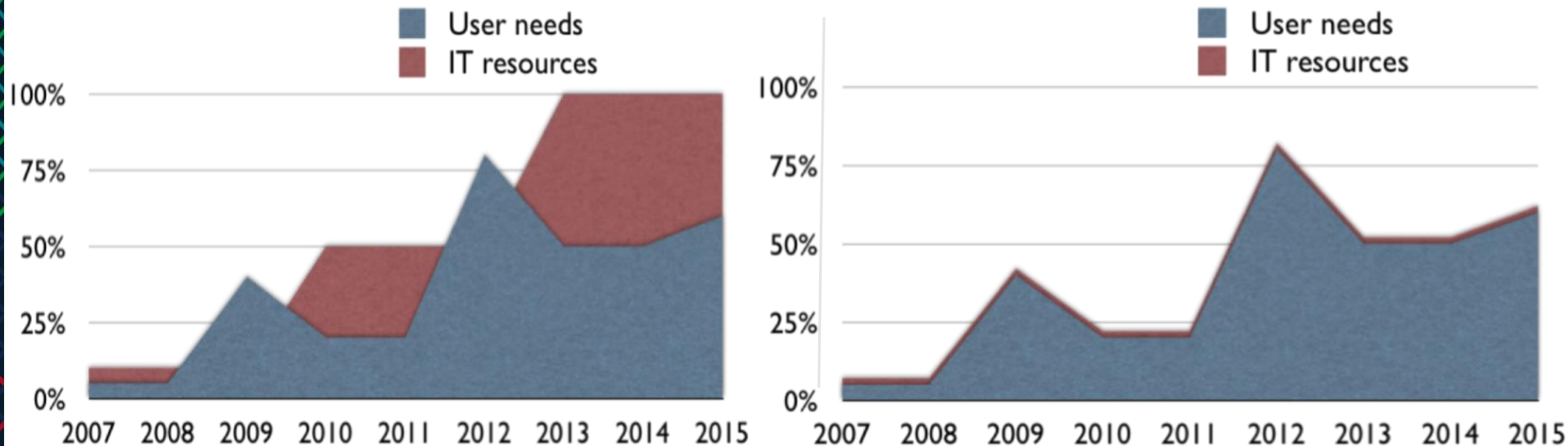
Examples of Cloud Services

- Dropbox
- Google Drive
- Microsoft OneDrive
- Apple iCloud

- Netflix - hosted on AWS
- Google search – Google Cloud
- Google Docs, Sheets, and Slides
- Facebook
- Skype
- Twitter

Company Infrastructure and User Needs
With vs. Without Cloud

Which diagram reflects the IT resources and User needs for Infrastructure using the Cloud? And Why?

Conventional Computing Infrastructure vs Cloud Computing Infrastructure

Conventional	Cloud
Manually Provisioned	Self-provisioned
Dedicated Hardware	Shared Hardware
Fixed Capacity	Elastic Capacity
Pay for Capacity	Pay per Use
Capital & Operational Expenses	Operational Expenses
Managed via System administrators	Managed via APIs

Cloud Benefits

Cloud Computing Service Model
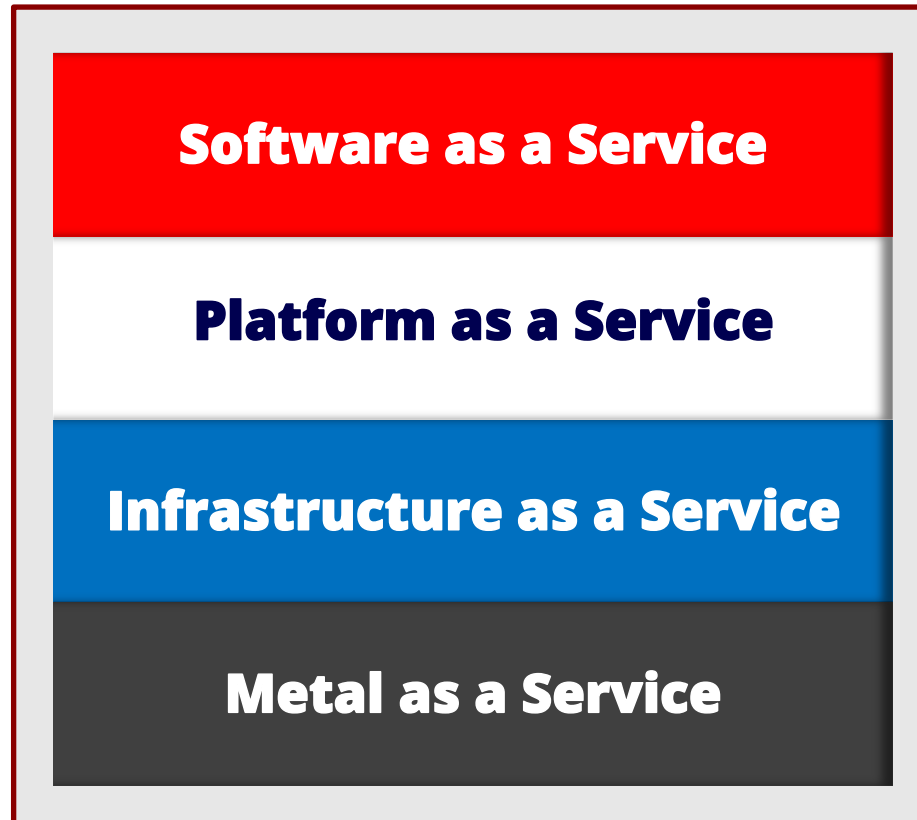
SaaS: Software as a Service

- Provided with access to application software in the cloud

    - On-demand software

- Most applications can be run directly from web browser

- Largest cloud market

- Examples: Google Apps, Microsoft Office 365, Oracle's Netsuite, SAP's Concur, Cisco WebEx, GoToMeeting

Software (as a Service)

- Applications
- Data
- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

Managed by vendor

PaaS: Platform as a Service

- Provides computing platforms which typically includes operating system, programming language, execution environment, database, web server, etc. to build cloud applications.

- Applications using PaaS inherit cloud characteristic such as scalability, high-availability, multi-tenancy, SaaS enablement, and more.

- Examples: Google App Engine, AWS Elastic Beanstalk, Salesforce.com, Amazon EMR, MS Azure HDInsight, GCP Dataproc

Platform
(as a Service)

You manage
- Applications
- Data

Managed by vendor
- Runtime
- Middleware
- O/S
- Virtualization
- Servers
- Storage
- Networking

IaaS: Infrastructure as a Service

- Offers storage and computing resources that developers and IT organization use to deliver custom business solutions

- Examples: Amazon EC2, VMWare vCloud, GCP Compute Engine
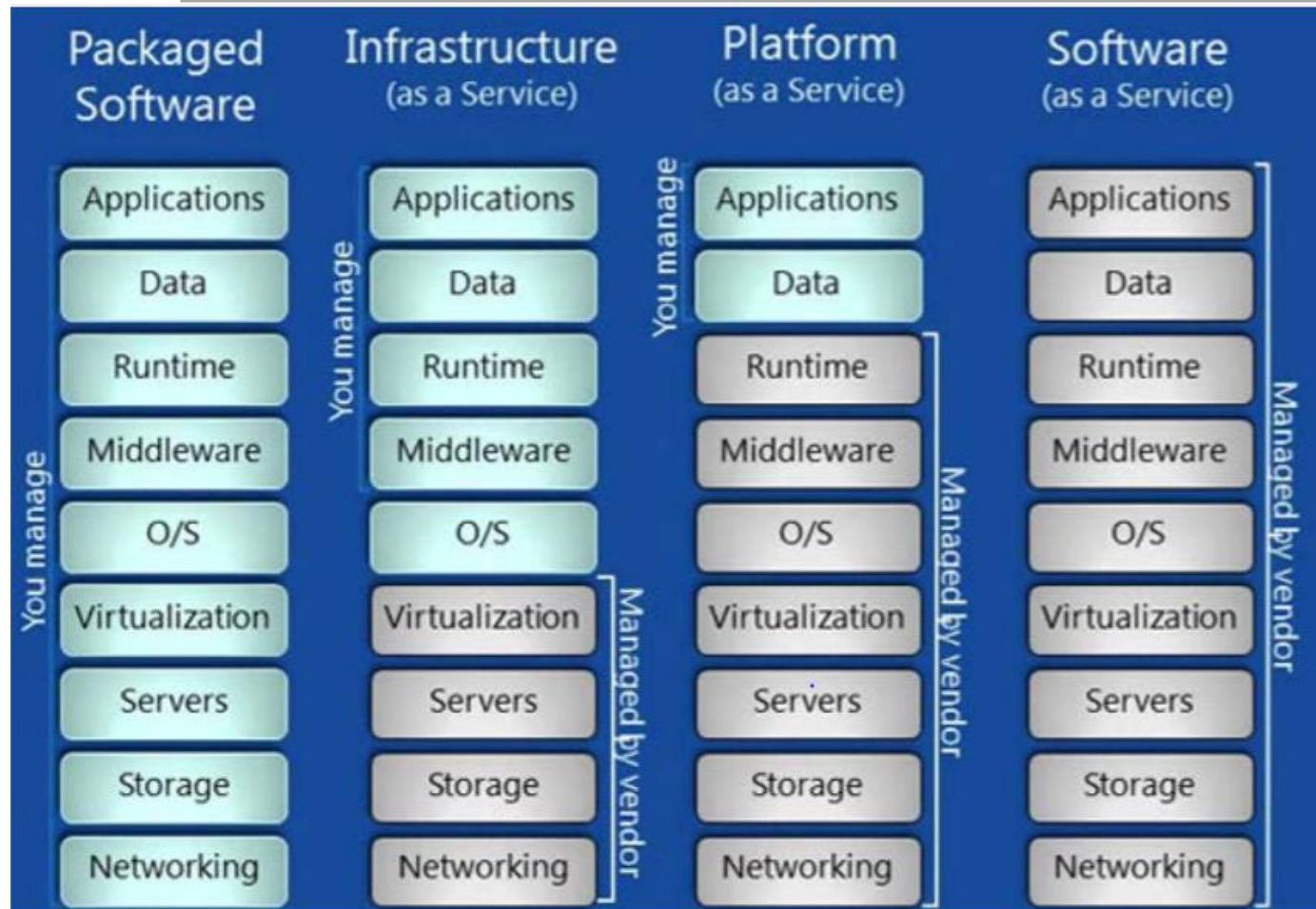

Carnegie Mellon University
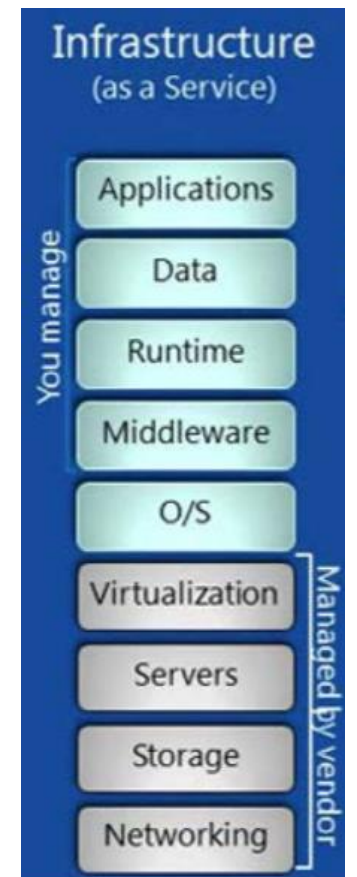
MaaS: Metal as a Service

- Combines the flexibility and scalability of the cloud with the ability to harness the power of physical servers.

- Examples: Juju

- For more information, watch this video (optional): https://www.youtube.com/watch?time_continue=280&v=FBCKCO45xIw
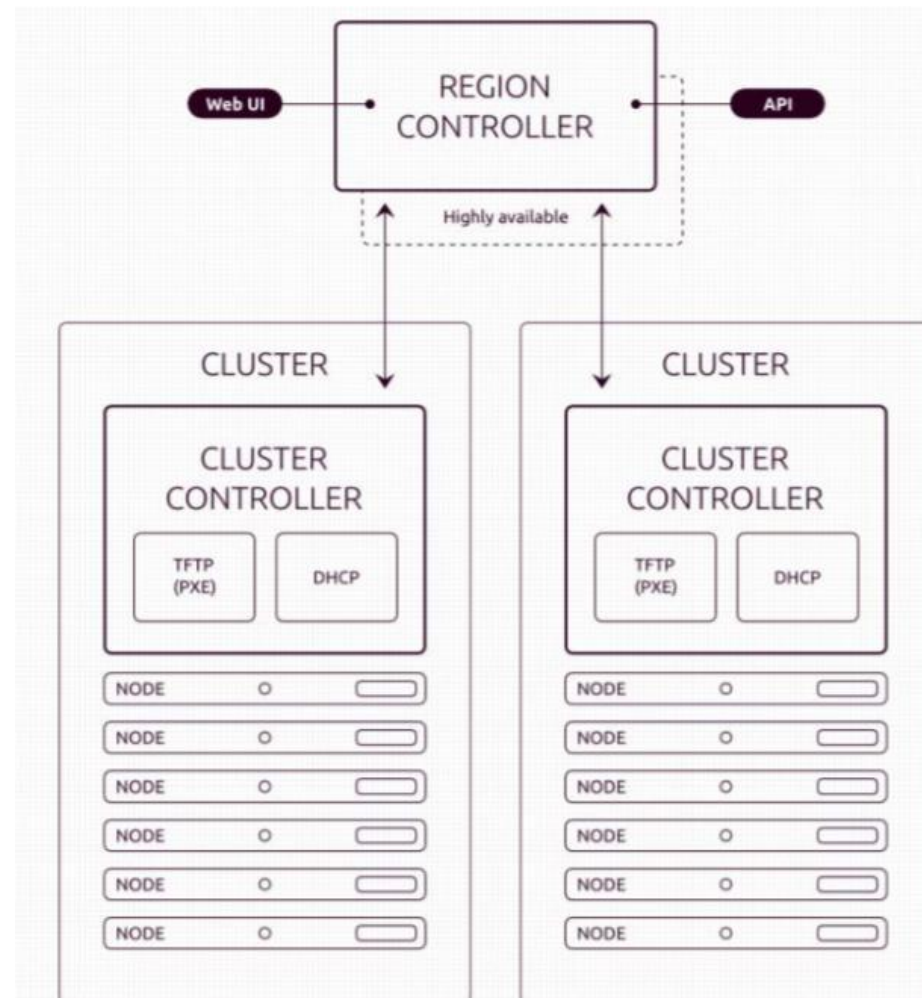
Cloud Computing Service Model

In MAAS

- You will have the option to control everything!

PaaS or IaaS?

Vendor Lock-in: the ability to use "what you manage" in cloud environment with different cloud provider.

- PaaS may lock-in applications by requiring users to develop apps based on their specific APIs.

- If you are using PaaS, it might be difficult to switch to different vendor.

Development Tools

- PaaS providers usually allow a set of development tools for their users to shorten development time.

- Another trick for vendor lock-in!

Cloud Computing Enablers

Data Center + Virtualization

Next

- Complete Course Entry Survey:

  - https://forms.gle/z2sE6W2yfBnv2ydM6

- Sign-up for the course on TopHat.

- Join the Course Piazza

- Join the Student Space on Slack

- Join the Student Slack Workspace

- Familiarize yourself with the location of the in-person OHs

- Check Homework-1 PDF

Carnegie Mellon University

Waitlisted Students

- All materials for first two weeks will be uploaded here