CE 528 Cloud Computing

Lecture 1: Course Introduction
Fall 2026

Prof. Yigong Hu

About Me

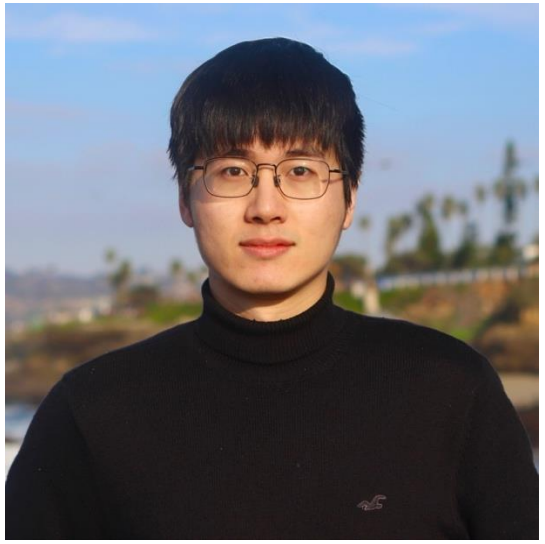
Prof. Yigong Hu

- Building reliable and fast systems
- From cloud platforms to ML systems
- https://yigonghu.github.io/

Office Hours

- Office: PHO335
- Mon/Wed: 4:00-5:00PM (or by appointment)

Course Staff



Wenbo Qian (TA)

- Office Hours: Tue 3:00-4:30 · Thu 4:30-6:00 · Fri 3:00-4:30
- Location: PHO 305/307

Zhakhangir Mamayev (grader)

- Office Hours: Mon 10:30-12:00 · Tue 1:00-2:30 · Wed 10:30-12:00
- Location: PHO 305/307

Course Website:

https://ec528.github.io/ec528/fall26/

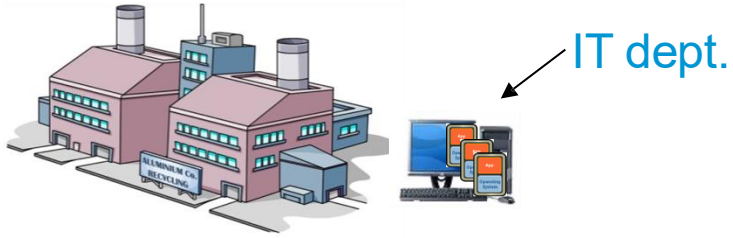
What is Cloud Computing?

- What do you think?

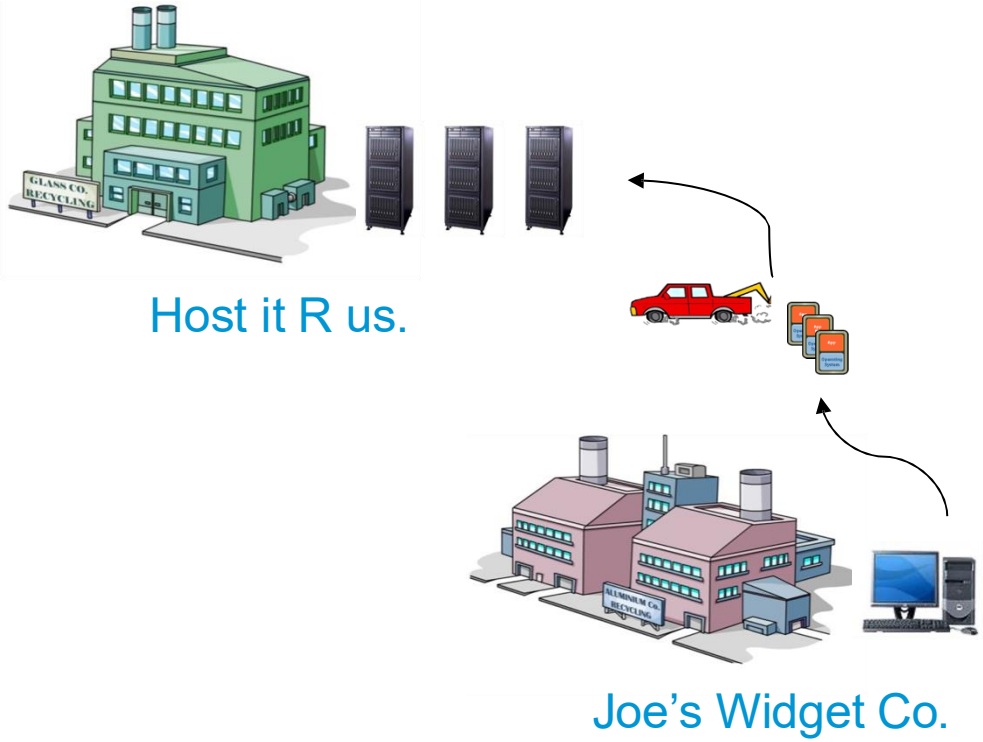
"Cloud computing is the ***delivery of computing services*** *–* ***servers, storage, databases, networking, software, analytics and more – over the Internet ("the cloud").*** *Companies offering these computing services are called* ***cloud providers*** *and typically charge for cloud computing services based on usage, similar to how you're billed for gas or electricity at home."*

  - https://azure.microsoft.com/en-gb/overview/what-is-cloud-computing/

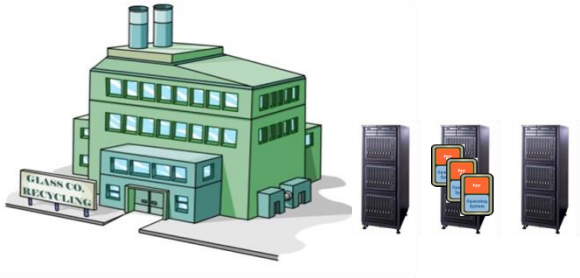
Cloud Computing: A Case



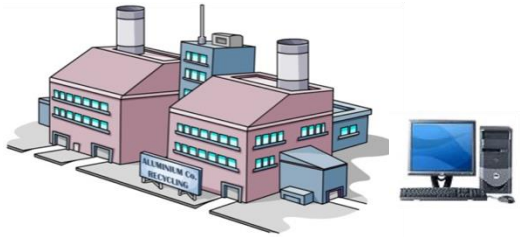
Cloud Computing: A Case



Cloud Computing: A Case

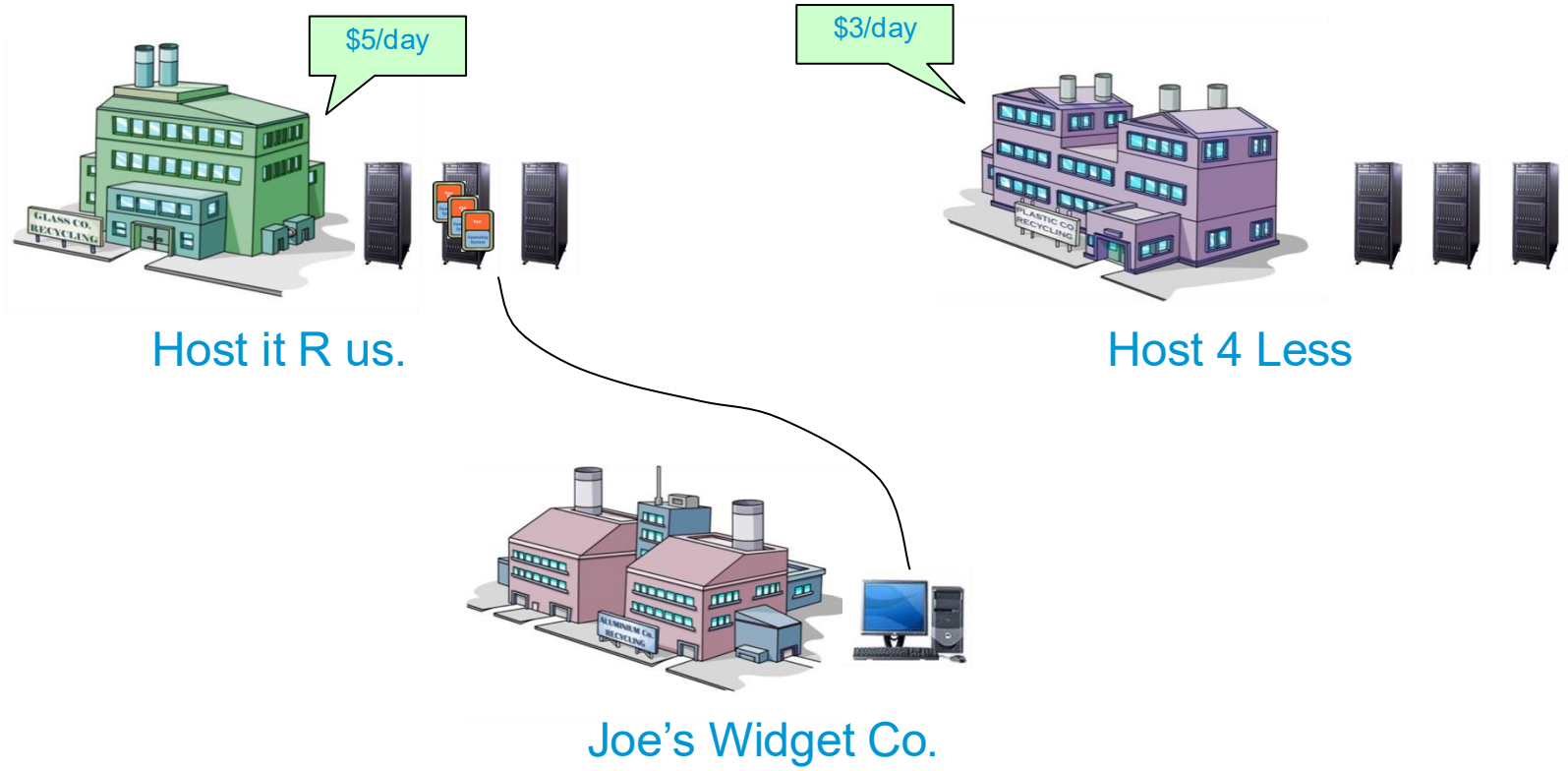


Host it R us.

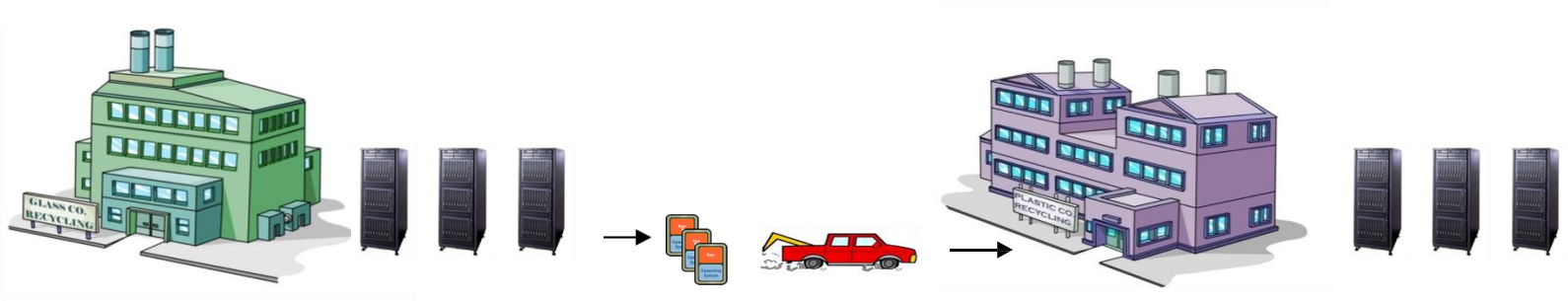


Joe's Widget Co.

Cloud Computing: A Case

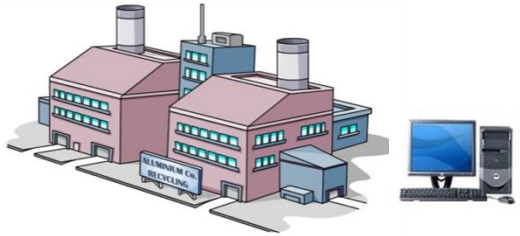


Cloud Computing: A Case



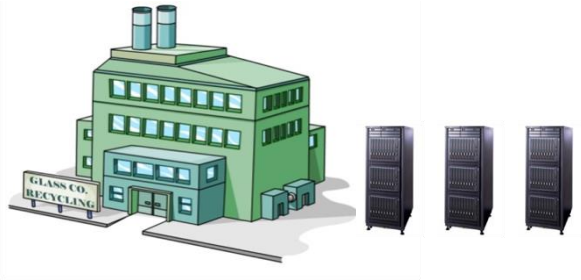
Host it R us.

Host 4 Less

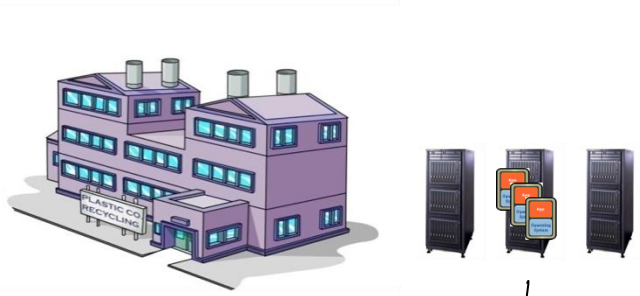


Joe’s Widget Co.

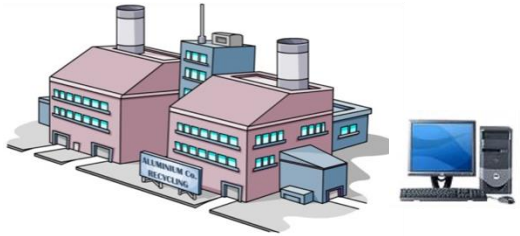
Cloud Computing: A Case



Host it R us.

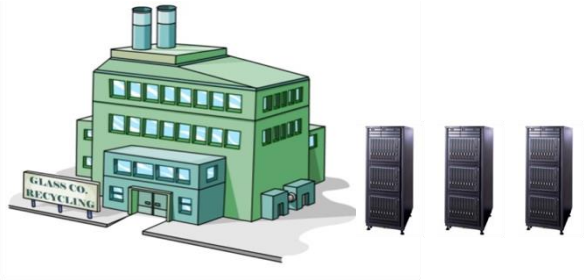


Host 4 Less

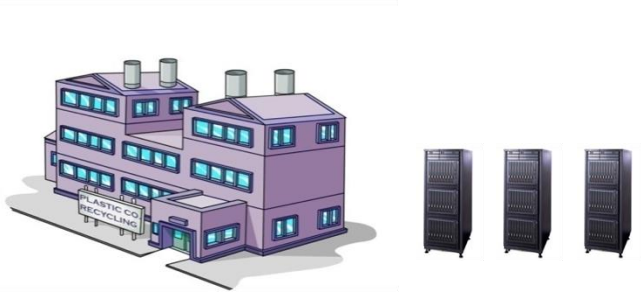


Joe's Widget Co.

Cloud Computing: A Case



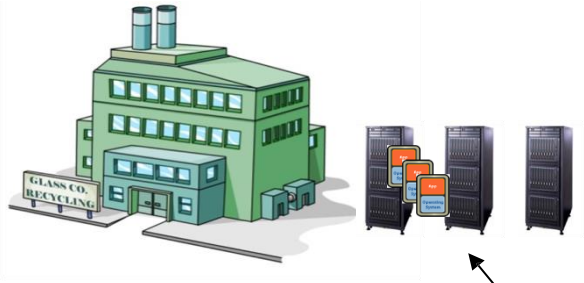
Host it R us.



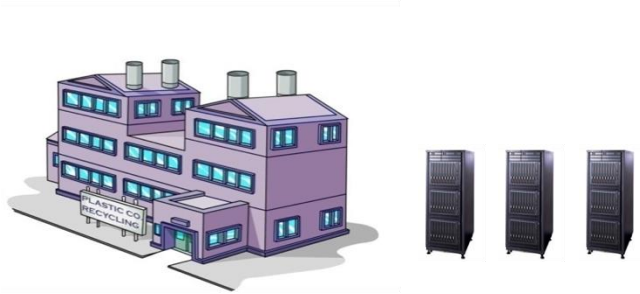
Host 4 Less

Snoopy's Startup

Cloud Computing: A Case



Host it R us.

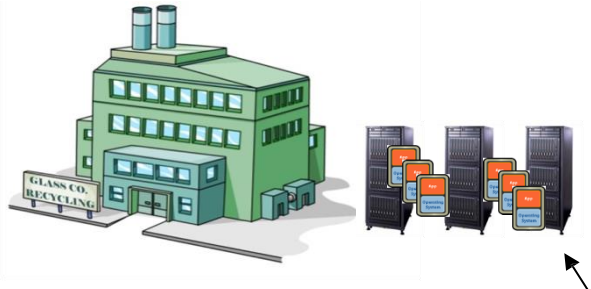


Host 4 Less

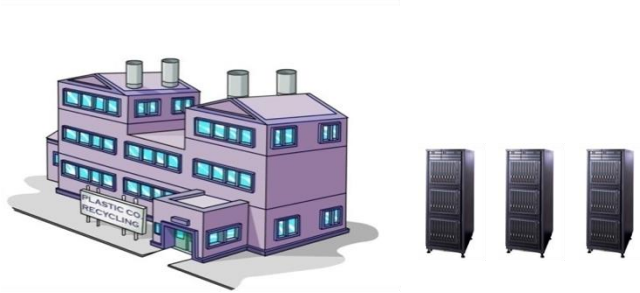
Dogspace

Snoopy's Startup

Cloud Computing: A Case



Host it R us.



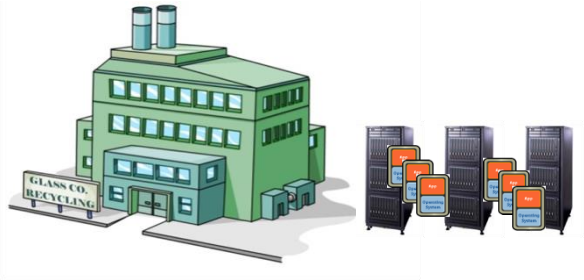
Host 4 Less

Dogspace

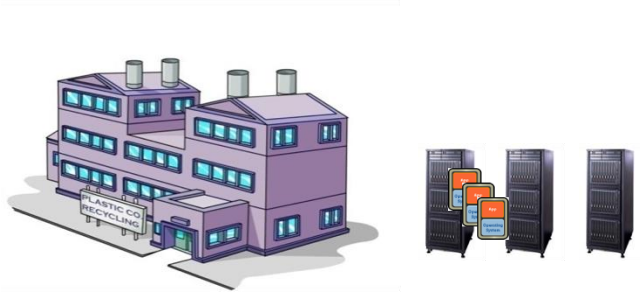
DogTube

Snoopy's Startup

Cloud Computing: A Case



Host it R us.



Host 4 Less

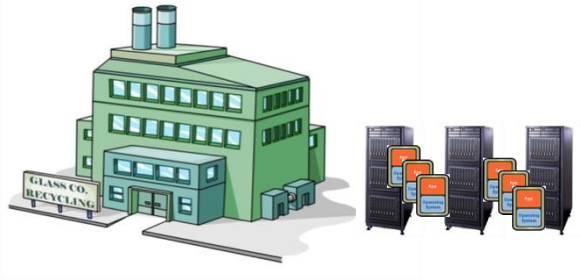
Dogspace

DogFlix

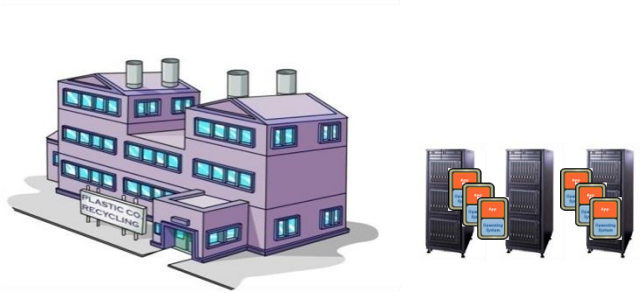
DogTube

Snoopy's Startup

Cloud Computing: A Case



Host it R us.



Host 4 Less

Dogspace

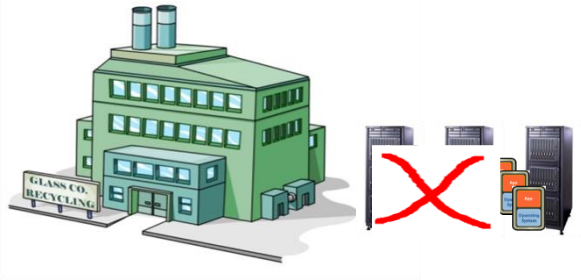
DogFlix

DogTube

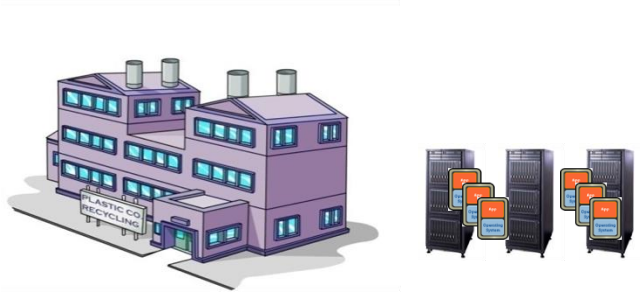
dBay

Snoopy's Startup

Cloud Computing: A Case



Host it R us.



Host 4 Less

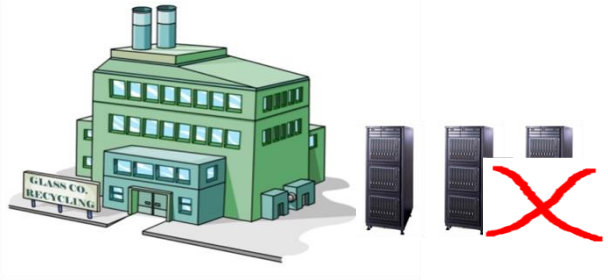
DogFlix

Snoopy's Startup

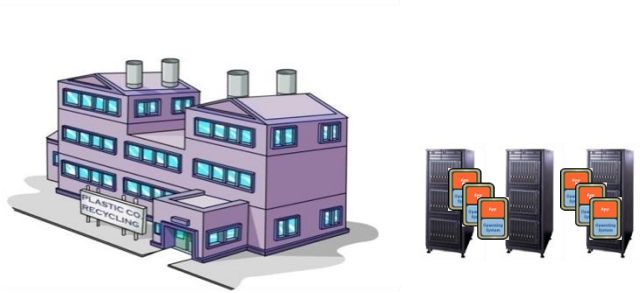
DogTube

dBay

Cloud Computing: A Case



Host it R us.



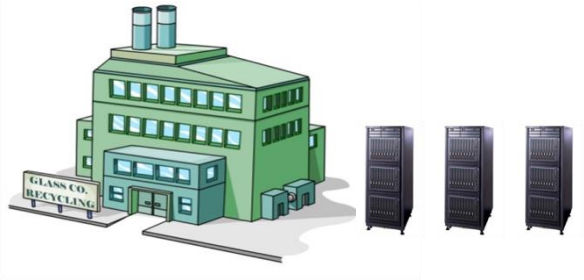
Host 4 Less

Snoopy's Startup

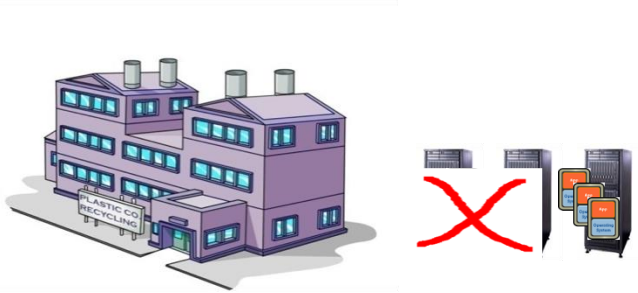
DogTube

dBay

Cloud Computing: A Case



Host it R us.

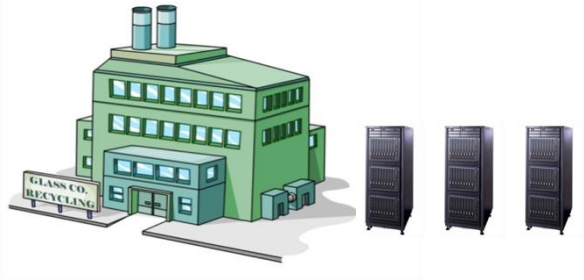


Host 4 Less

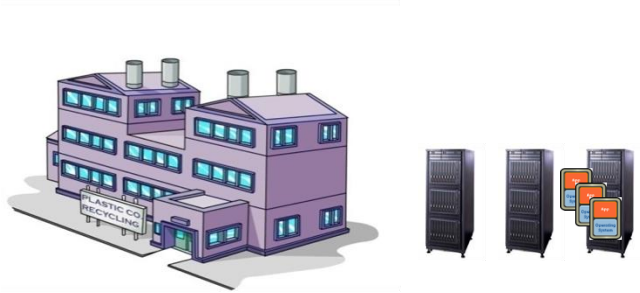
Snoopy's Startup

dBay

Cloud Computing: A Case



Host it R us.

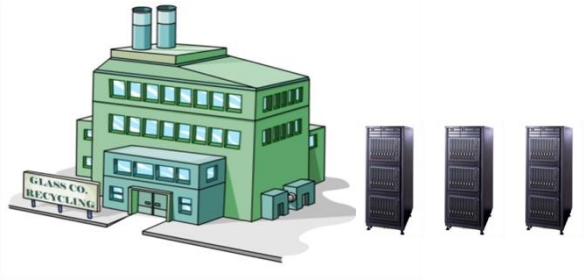


Host 4 Less

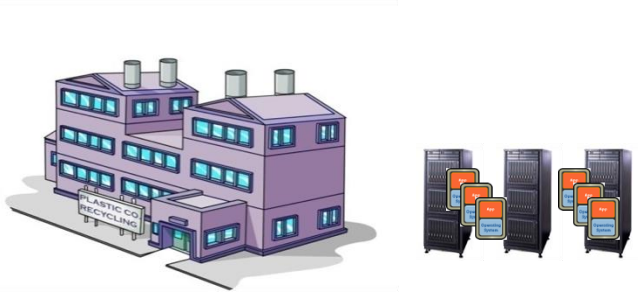
Snoopy's Startup

dBay

Cloud Computing: A Case



Host it R us.

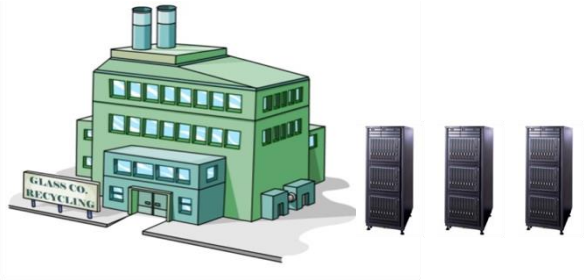


Host 4 Less

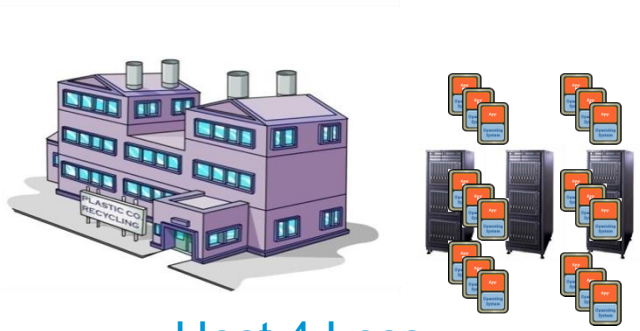
Snoopy's Startup

dBay

Cloud Computing: A Case



Host it R us.



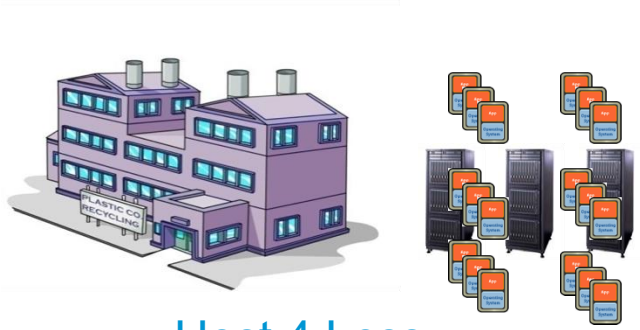
Host 4 Less

Snoopy's Startup

dBay

Cloud Computing: A Case

Host it R us.



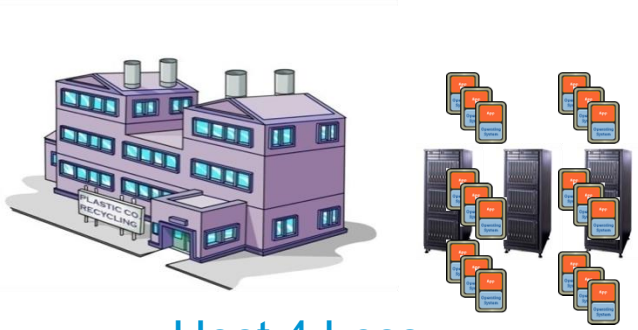
Host 4 Less

Snoopy's Startup

dBay

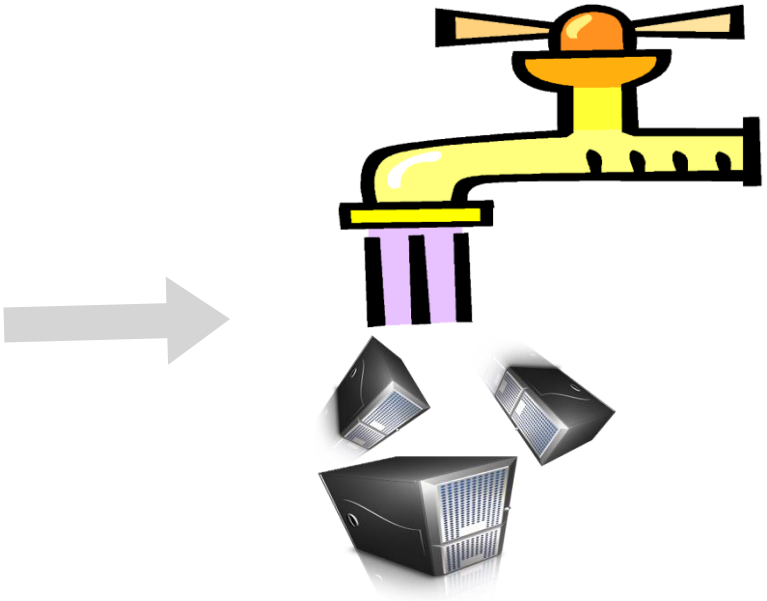
Cloud Computing: A Case

Host it R us.

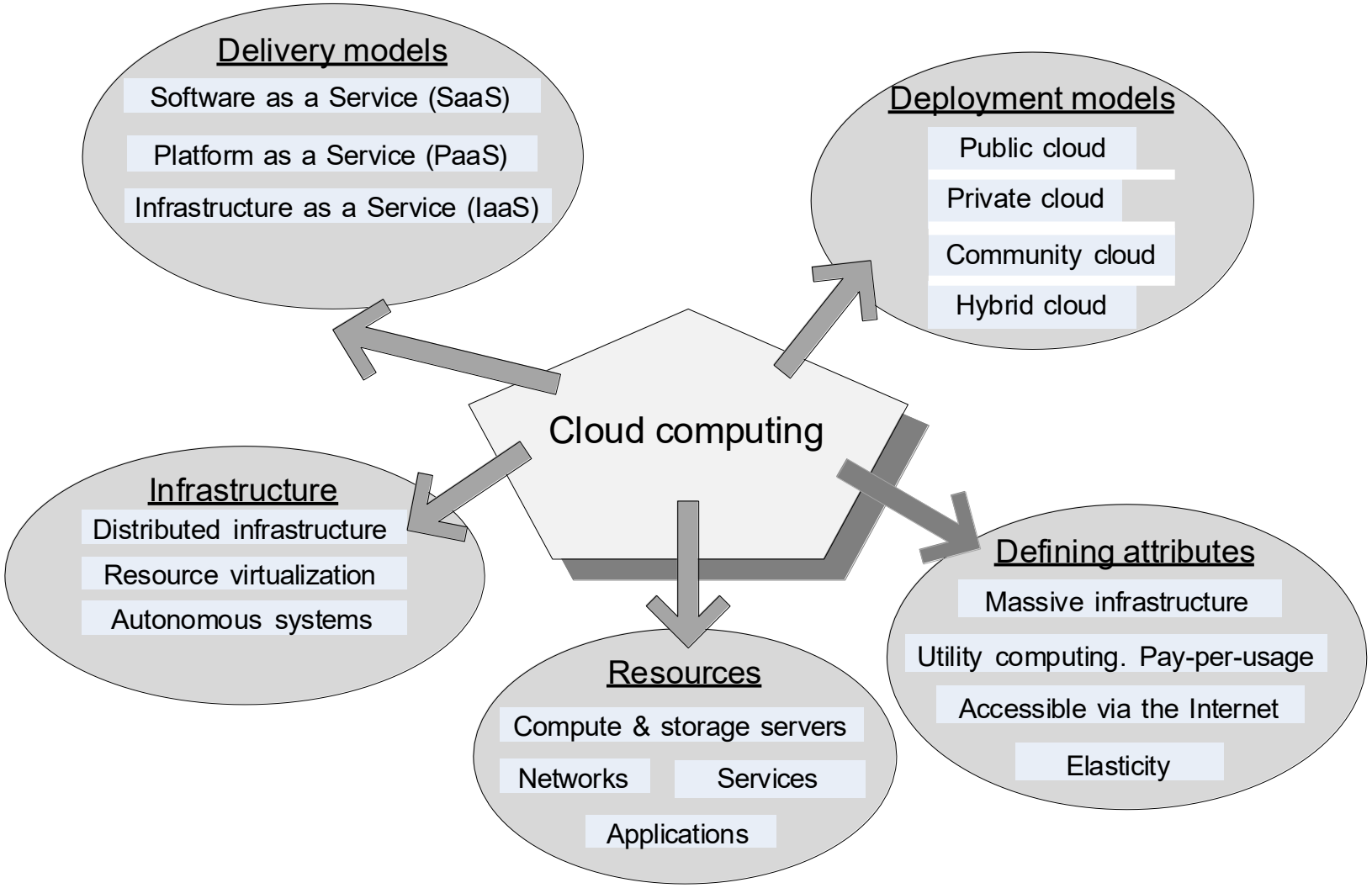


Host 4 Less

What We Want?



Cloud Computing Models, Resource, Attributes



Types of Clouds

1. **Public Cloud** - the infrastructure is made available to the general public or a large industry group and is owned by the organization selling cloud services.

2. **Private Cloud** – the infrastructure is operated solely for an organization.

3. **Hybrid Cloud** - composition of two or more Clouds (public, private, or community) as unique entities but bound by a standardized technology that enables data and application portability.

4. **Other types: e.g., Community/Federated Cloud** - the infrastructure is shared by several organizations and supports a community that has shared concerns.

Course Structure

- Distributed Systems Foundation
- Cloud Computing Foundation
- Advance topic on Cloud Computing
- Projects mentored by the industry

You can get to course website. This will be kept up to date, has all deadlines, demos…

Piazza is the source of truth.

Useful Link: http://nil.csail.mit.edu/6.824/2020/schedule.html
https://web.eecs.umich.edu/~manosk/eecs591-f21.html

Course Logistics

Good News: no exam in this class

Paper readings

- 1–2 papers per week.
- You do not need to write reviews, but you are expected to read them.

How You Are Evaluated

Project: 50% (team)

- Demo 1 - design proposal: 5%
- Demo 2 - working prototype: 10%
- Demo 3 - optimized system: 10%
- Final project result and report: 20%
- Presentation quizzes: 5%

Presentation: 10% (per presenter)

Individual contribution: 10% (peer + mentor)

Paper quizzes: 30%

- There are 18 quizzes in class:
  - MS/Ph.D. students: your highest 14 scores count.
  - Undergraduate students: your highest 12 scores count.

How You Are Evaluated

In-class Questions: Bonus

- A correct answer receives 0.5 bonus point, and an insightful answer receives 1 bonus point.
- A wrong answer does not incur any penalty
- Each student may receive at most one bonus per class
- Bonus points are capped at 3 points toward your final grade.
- If no one volunteers to answer, a student will be selected at random.

Course Project

There is a capstone project in the course.

Each team will be assigned a mentor from industry or a senior graduate student.

Project teams will consist of **4–5 students**.

Projects are proposed by mentors, and you may indicate your preferences among them.

Lets look at the projects

Peer Scoring

There will be an anonymous peer scoring process

- Individual contribution: 10% = 5% peer + 5% mentor
- Each team member rates every teammate on a scale of 1 to 10.

How it is scored

- 10 = carried their full share; 1 = contributed nothing
- Your score is the average of what teammates give you
- Peer points = 5 x min(average, 9) / 9
- An average of 9 or above earns full marks
- Your mentor also rates engagement and reliability

Ratings are anonymous - nobody sees who gave what

Project Presentation

3 in-class demos plus a final presentation; all groups present.

Before the demo

- Push slides and code to the demo branch by 12:00 noon
- Email me two multiple-choice questions by 5:30 PM
- You are teaching us - minimize acronyms
- Review your slides with your mentor first

Presenting

- 10 minutes to present, 5 minutes for questions
- At most two members per demo; everyone presents once

Paper Quizzes

18 quizzes, one per lecture

- One in every lecture except the first two
- 10 multiple-choice questions in 10 minutes
- Open book and open internet

Only your best scores count

- MS/Ph.D. students: your highest 14
- Undergraduates: your highest 12

There are no make-ups

- The extra quizzes are your absence buffer

Grading Scale

Grade thresholds

- ≥93 A, ≥88 A-, ≥83 B+, ≥78 B, ≥68 B-, ≥60 C

I may lower these. I will never raise them

- Based on course difficulty and the class distribution

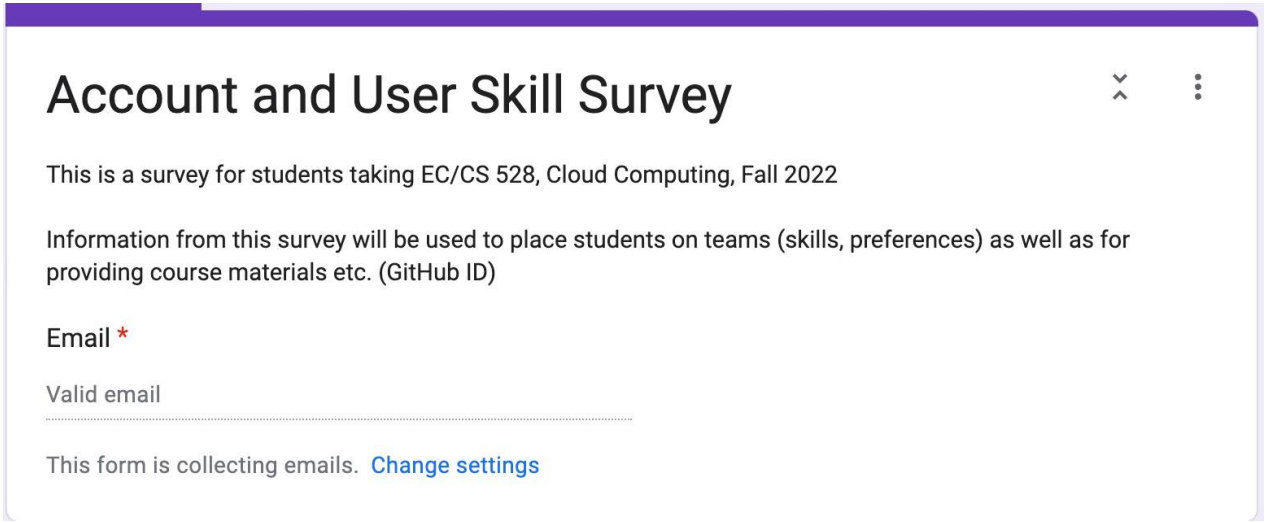
Do not spend the semester worrying

- In past offerings about half the class finished A or A-

Immediate Next Steps

Project matching

- We will post a project preference form on the course website.



- Fill out the form by 11:59 PM on 09/06.
- We will assign project teams and please reach out to your team member asap

Some Suggestions

Working as a team

Come to office hours for help with design problems

Be in touch with your mentors.

- Focus on the high-level questions
- Let the instructors know early if there are problems

But, Why Should You Stay in This Class, and This Field?

We Will Discuss

This is the future of all computing

It has profound impact on:

- how hardware is developed
- how software is developed
- the computer industry
- society as a whole

Course Contents: A Variety of Topics

- Overview of Cloud Computing
- Virtualization
- Distributed systems
- Datacenter architectures
- Storage systems
- Resource management
- Big data processing
- Datacenter networking
- Serverless Computing
- Performance Issue
- Reliability and AIOps
- LLM Infrastructure

So, Why is Cloud Computing Cool?

Shared Resources and Resource Management

- Cloud uses a shared pool of resources
- Uses Internet to offer **scalable** and **elastic** services.
- The term "**elastic computing**" refers to the ability of dynamically and on-demand acquiring computing resources and supporting a variable workload.
- Resources are metered and users are charged accordingly.
- It is more cost-effective due to resource-multiplexing. Lower costs for the cloud service provider are past to the cloud users.

Cloud Computing Benefits?

Resources, such as CPU cycles, storage, network bandwidth, are shared.

When multiple applications share a system, their peak demands for resources are not synchronized thus, **multiplexing** leads to a higher resource utilization.

Resources can be **aggregated** to support data-intensive applications.

Data sharing facilitates **collaborative** activities. Many applications require multiple types of analysis of shared data sets and multiple decisions carried out by groups scattered around the globe.

Cloud Computing Benefits(cont'd)

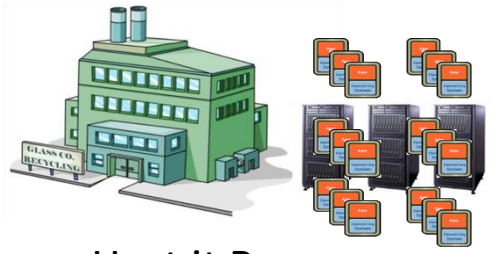
Eliminates the **initial investment costs** for a private computing infrastructure and the maintenance and operation costs.

Cost reduction: concentration of resources creates the opportunity to pay as you go for computing.

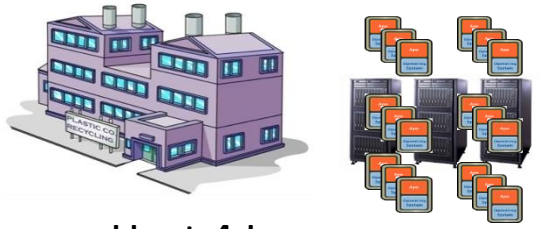
Elasticity: the ability to accommodate workloads with very large peak-to-average ratios.

User convenience: virtualization allows users to operate in familiar environments rather than in idiosyncratic ones.

Remember this?



Host it R us.

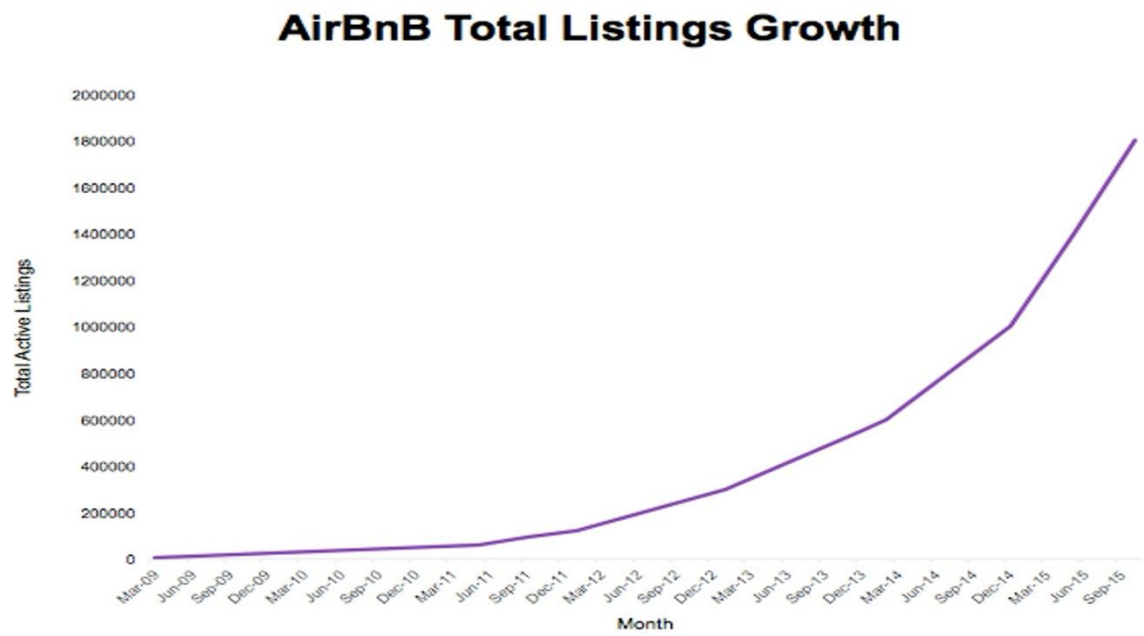


Host 4 Less

AirBnB Example

Success of market depends on network of renters and landlords

- starts really small



AirBnB

2010 – 24 EC2 instances, 300 GB of data

2015 – 1000 EC2 instances, 50 TBytes data

Grew up entirely on AWS, no data center, no capital purchases, no racking/stacking, no acquisition networking…

- 5-person operations team
- Piggyback on AWS for external network, availability zones

Rapid growth easily accommodated

Coursera

Massive on-line courses from Stanford, Duke…

Went from 0 to 3.2 million users in first year

Accessed from around the world

Spikes common, e.g., 75% increase in load in 5 minutes

Many Cloud Providers

Make computing resources available on demand

- through a growing set of simple APIs
- leverages economies of scale of large datacenters
- … for anyone with a credit card
- … at a large scale, if desired

Cloud Services Providers

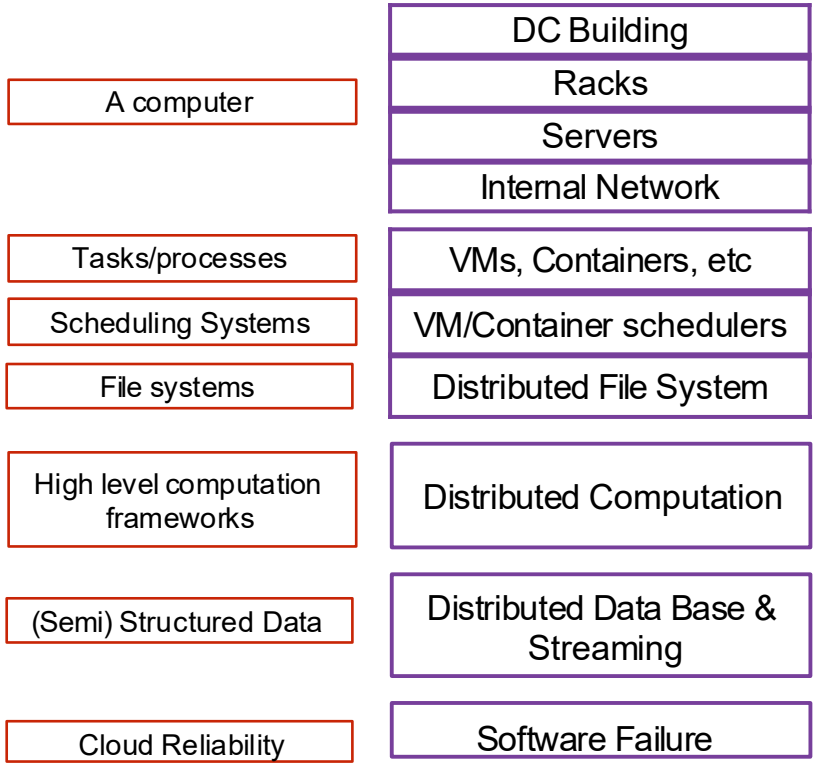
Amazon: Queue API in 2004, EC2 launched in 2006
Google: AppEngine in 2005, other services starting in 2008
Microsoft: Azure launched in 2008.

Millions of customers using these services

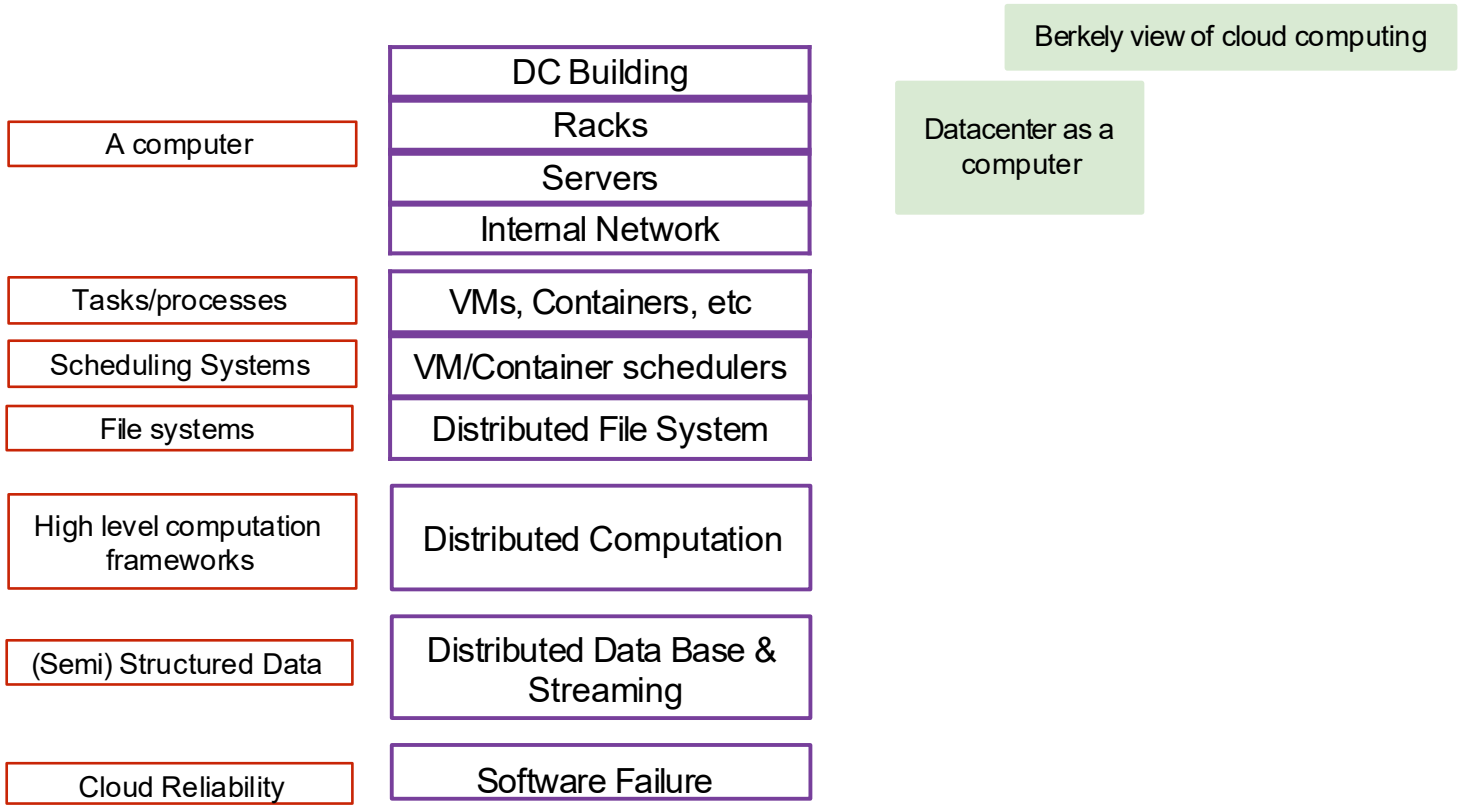
Shift towards these services is accelerating

Comprehensiveness of APIs increasing over time

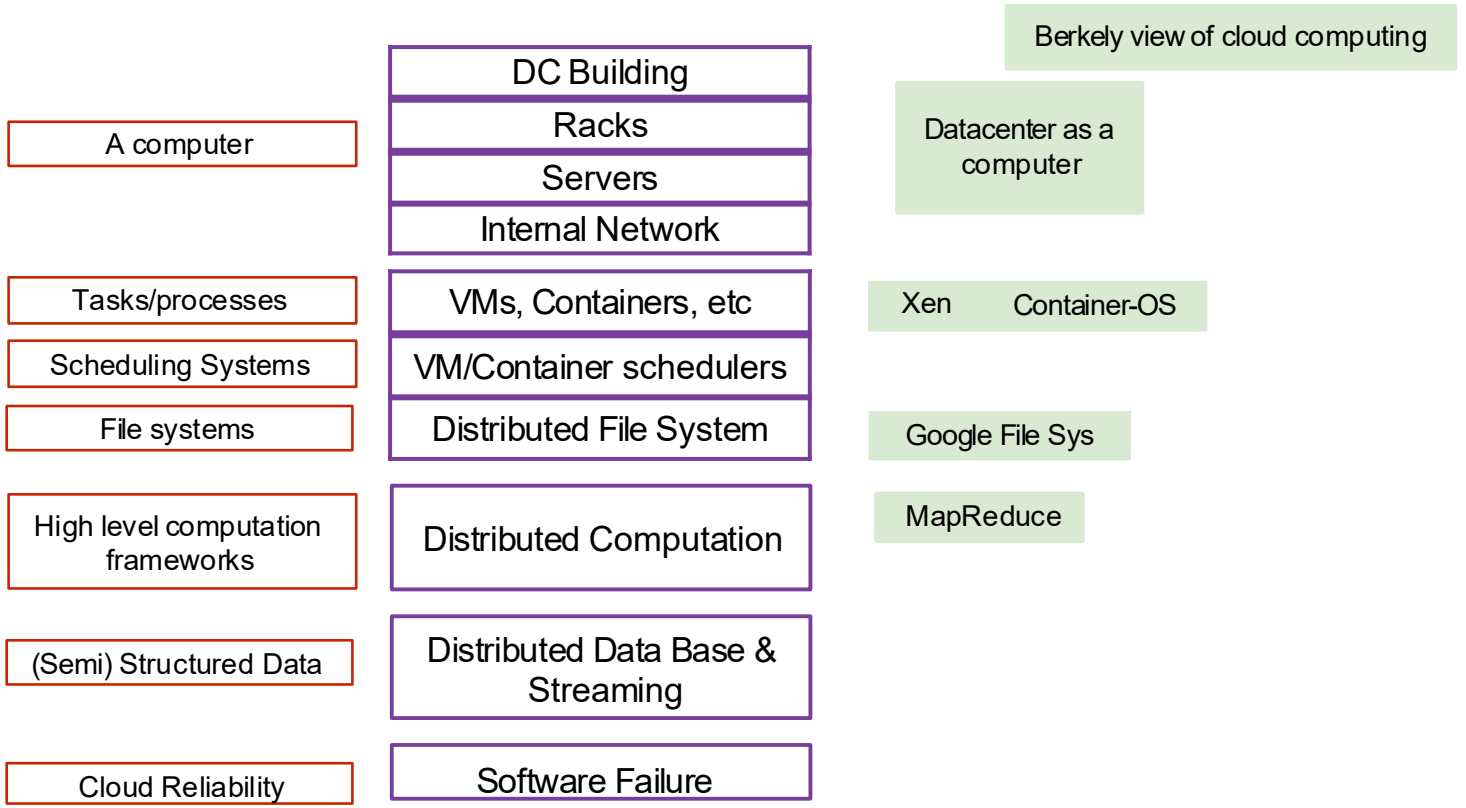
Top-Down View of the Course



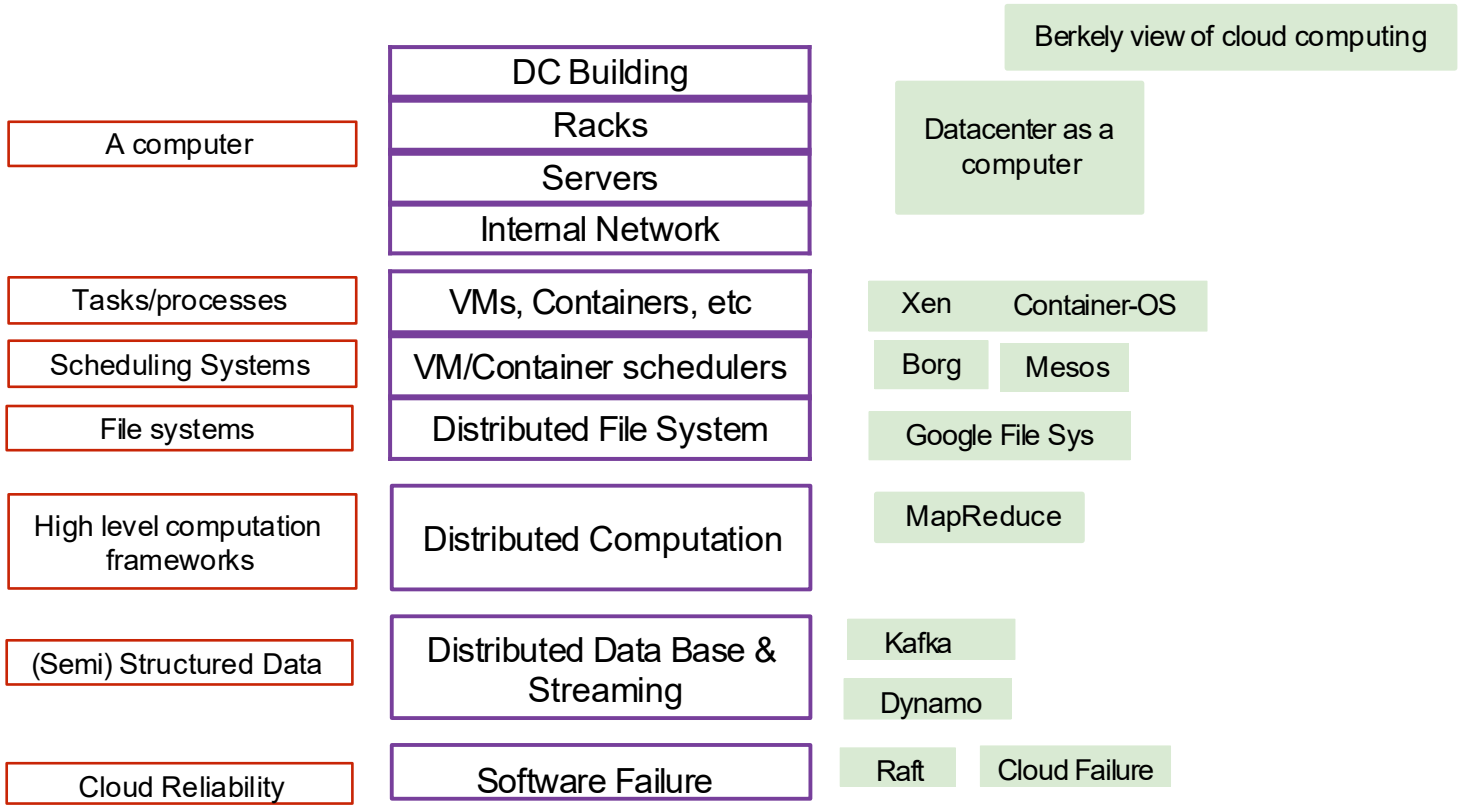
Top-Down View of the Course



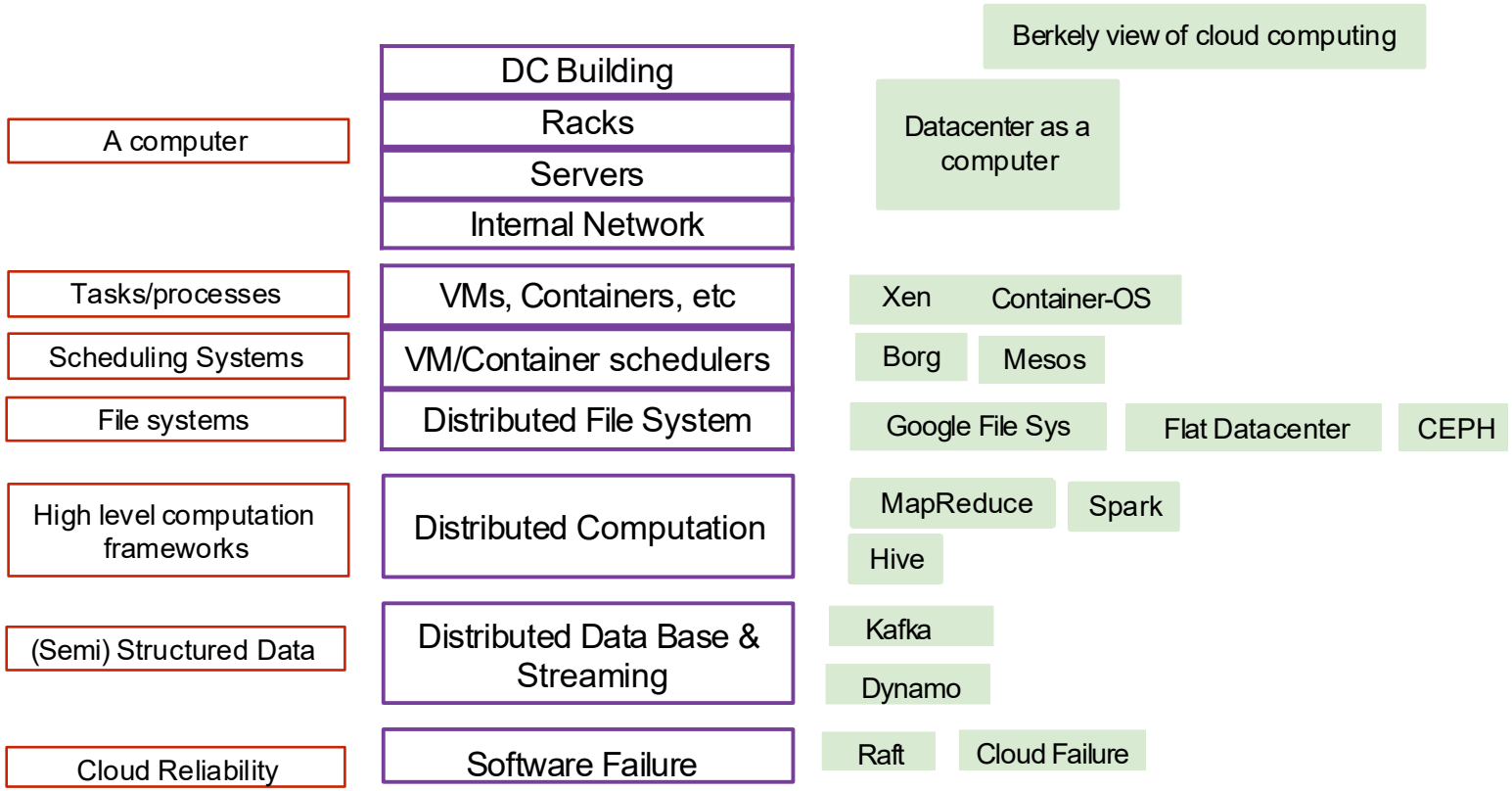
Top-Down View of the Course



Top-Down View of the Course



Top-Down View of the Course



How to Read

You May Think You Already Know How To READ, But…

Why Do We Need to Read Paper?

- Reading for classes
- Reach the edge of human knowledge
- Keeping up with your field
- Transitioning into a new areas
- Learning how to write better papers

Keshav's Three-Pass Approach: Step 1

A ten-minute scan to get the general idea

- Title, abstract, and introduction
- Section and subsection titles
- Conclusion and bibliography

What to learn: the five C's

- Category: What type of paper is it?
- Context: What body of work does it relate to?
- Correctness: Do the assumptions seem valid?
- Contributions: What are the main research contributions?
- Clarity: Is the paper well-written?

Keshav's Three-Pass Approach: Step 2

A more careful, one-hour reading

- Read with greater care, but ignore details like proofs
- Figures, diagrams, and illustrations
- Mark relevant references for later reading

Grasp the content of the paper

- Be able to summarize the main idea
- Identify whether you can (or should) fully understand

Keshav's Three-Pass Approach: Step 3

A several-hour virtual re-implementation

- Making the same assumptions, recreate the work
- Identify the paper's innovations and its failings
- Identify and challenge every assumption
- Think how you would present the ideas yourself
- Jot down ideas for future work

Next Time

- **The rise of cloud computing systems**