MY472 - Week 3: Cloud Computing

Pablo Barbera & Akitaka Matsuo

October 16, 2018

About me: Akitaka Matsuo

- Born and raised in Japan
- Did post graduate study in Political Science at Rice University in Houston
- Research Fellow in Data Science at Methods@LSE:
  - Previously Research Officer at Nuffield Centre for Experimental Social Sciences, University of Oxford
  - Ph.D in Politics, Rice University
- My research and work:
  - Scaling methods (latent scaling, Bayesian methods)
  - Political text analysis (including parliamentary speeches, social media texts)
  - Author of R packages for text analysis
- Web presence
  - website: http://amatsuo.net
  - github: https://github.com/amatsuo

Your turn!

Sorry for asking you to do this again, but please introduce yourself

- Name
- Career goal
  - Research? Professional? Still weighing options?
- What do you want to get from this course?

Cloud Computing

Cloud Computing

Senario

- Suppose that you want to run a deep learning project
  - Need a lot of computer resource

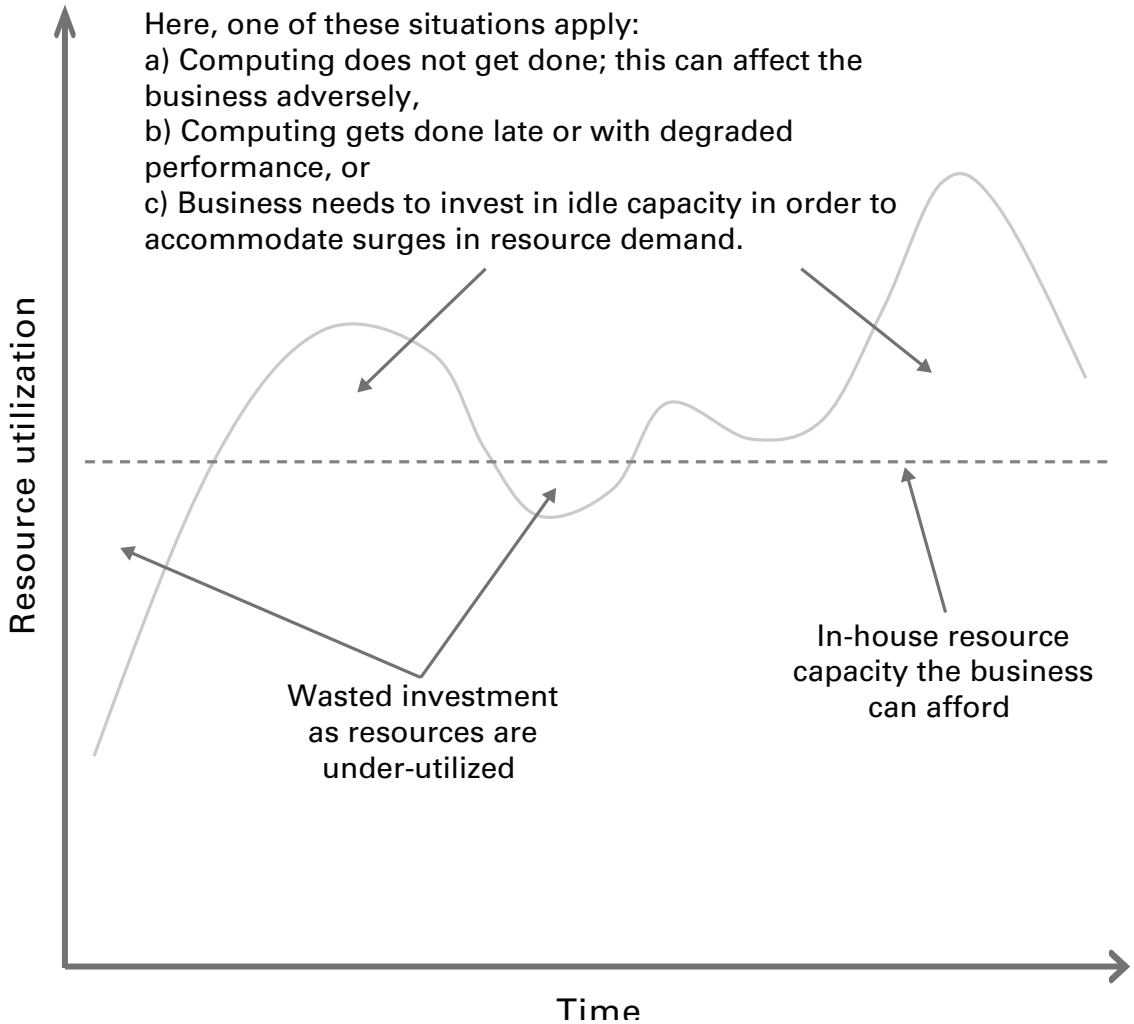
So. . .

Let's buy a powerful comptuer!

I can analyze anything!

But this is what happens. . .

- Your computation environment is fixed
- Your demand fluctuates



Also, all sorts of things can go wrong. . .

- Hardware breaks down
- Software needs patch
- Intenet connection is gone

So what if...

- You just have to pay what you use?
- You can deploy things in the matter of hours/minutes/seconds?

That's what cloud computing is for!

What is cloud computing?

Let me ask your experience

- Experience in Linux/Unix console?
- Experience use virtual machine softwares on your computer (e.g. VirtualBox, Parallels)?
- Having account on major cloud platforms (AWS, Microsoft Azure, Google Cloud)?
- Ever used HPC environments?

You are already using cloud application services

These are the examples of Software as a Service:

- email
  - gmail
  - exchange mail
- storage
  - Google drive
  - Dropbox
- software
  - Google docs
  - Adobe

But, beyond the use of cloud services, we will try to learn about and deploy the cloud computation resources.

Definition

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models." (From *NIST Definition of Cloud Computing*)

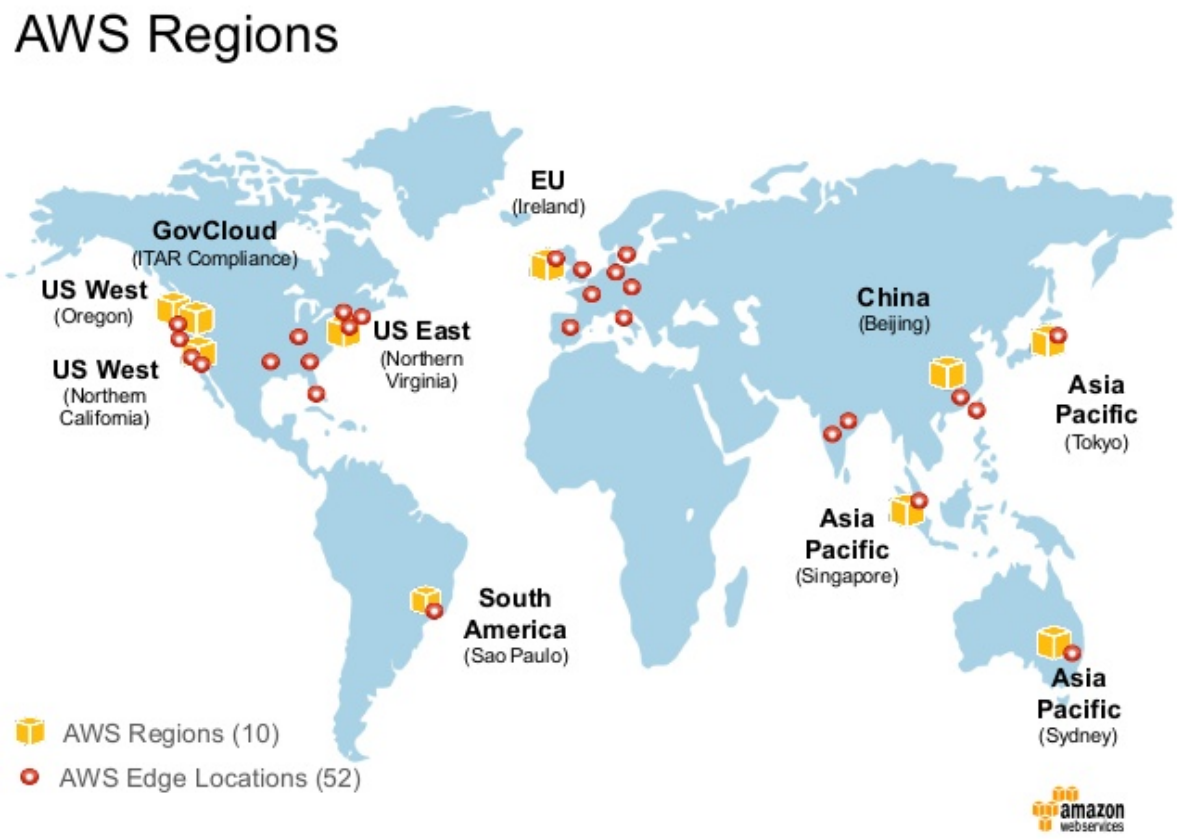
Technical perspective: Virtualisation

- In cloud computing, computers are virtualized
  - Similar idea to virtual machine on your computer
  - In data centers connected to the network, there are a number of hardwares which host a number of virtual machines

Data centers

- For major providers of cloud service, there are data centers everywhere and you can deploy resources in any location.

Example AWS



Data centers

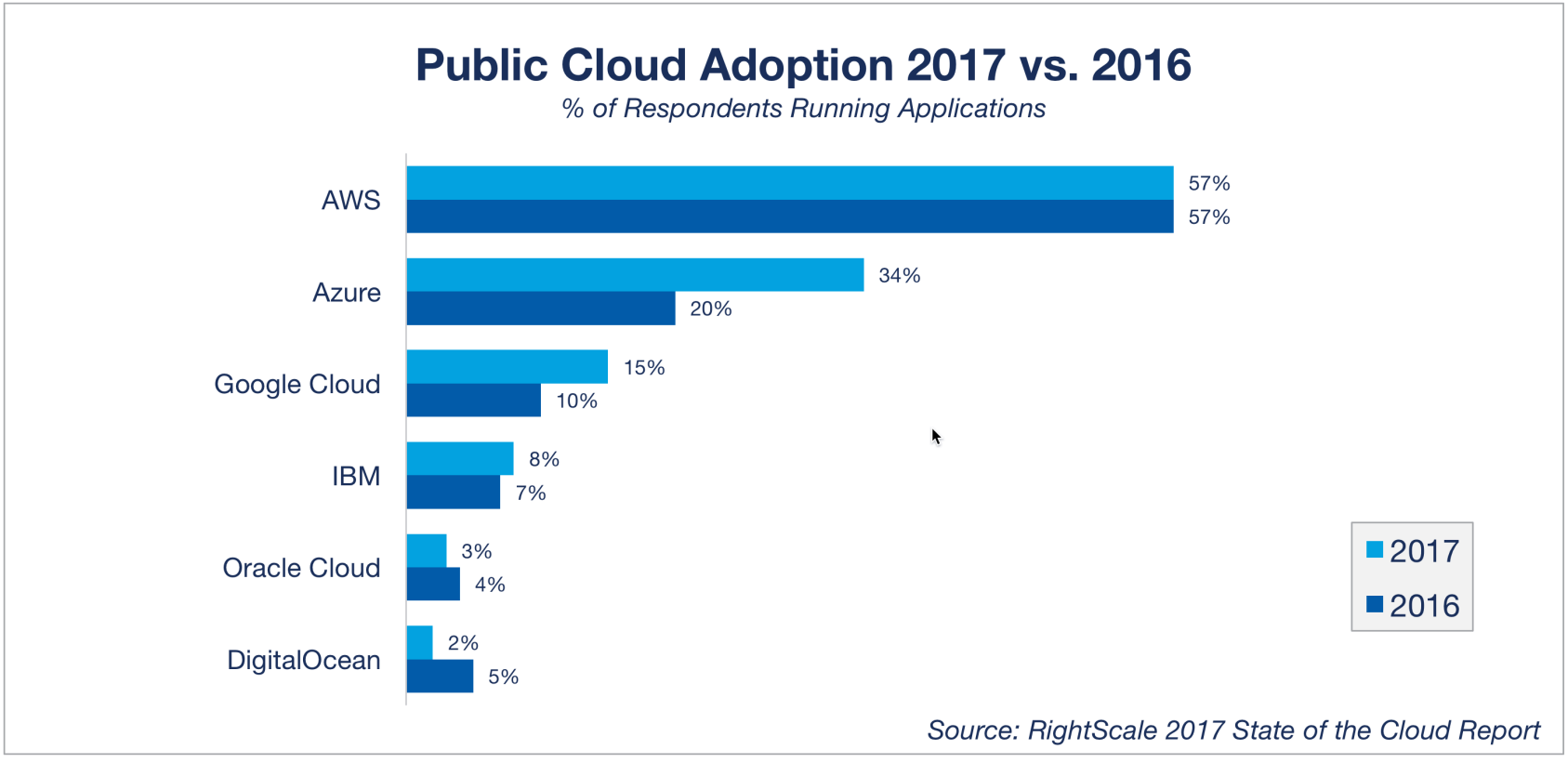
- Physial locations of data centers are usually undisclosed (for security reasons)

Technical Perspective: Virtualisation

Difference between virtualisation on your machine and cloud

- elasticity
  - readily **scalable** in short time
    - deploy the same app on a bigger instances
    - set up auto-scaling with load balancers
- measured service provision
  - billed based on the usage (e.g. CPU/memory/storage/network)

Managed cloud service providers



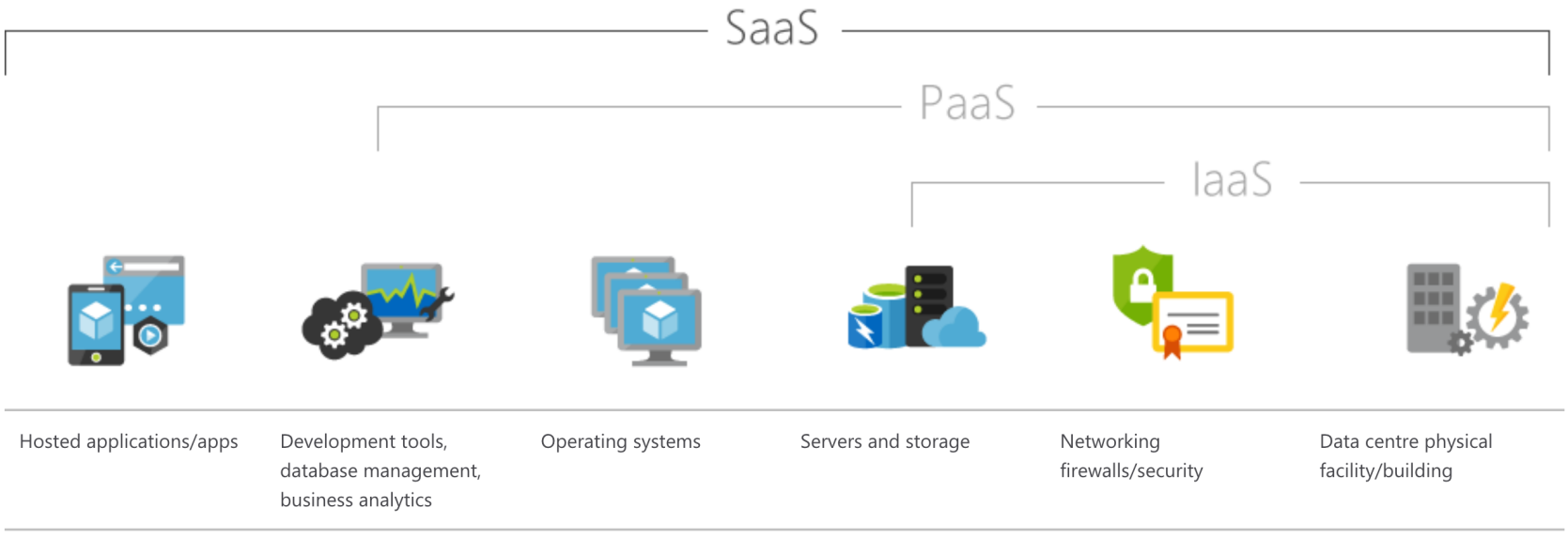
From *4mation*

Conceptual perspective: Level of abstraction

To what extent, you want a service provider to be responsible?

- Infrastructure? (IaaS)
- Platform (= OS)? (PaaS)
- Software (SaaS)

aaS = as a Service



We mostly work with PaaS in the lab

Some use cases of PaaS

- Data storage
  - Store important data reliably and securely
- Computation
  - Set up a virtual machine and run a computationally intensive tasks
- Data acquisition
  - Stream tweets (although it's sort of discontinued)
  - Create a crawler to regularly check news websites
- Data hosting
  - Set up databases online

Advantage and disadvantage of cloud computing

Cloud computing: Advantages

- Availability
  - In a given time, what's the chance that the service is available?
- Durability
  - What's the chance of losing the data?
- Scalability
- Cost efficiency
- Security

Advantages: Scalability

Definition: "The ability for something designed to operate at one measure of size to operate successfully at other sizes. The term is commonly used in relation to the development of shared computer applications that are intended to be used by large numbers of users. Of necessity, developments take place with a small number of test users. Unless the application is carefully designed to take account of the interactions that will arise when it is called on to service a large number of users, it may well fail to operate at all, or to operate only with an unacceptable level of service. An application that successfully expands its numbers of supported users is said to be scalable."

(From Oxford Dictionary of Computer Science)

Simply put, the question is whether the computational system you have created can seamlessly scale up?

Advantages: Cost efficiency

- In many cases, it is cheaper to use cloud computing services
  - You may need to use high performance computing for a few times
- Total cost of ownership (TCO) model
  https://aws.amazon.com/tco-calculator/

Advantages: Security

When you use cloud services, some of the security concerns (security "of" the cloud) are taken care by the service providers.

For example:

- Update computer systems
- Automated backup
- Network connection in the cloud
- Physical security of infrastructure

However, you also are responsible for other concerns

- Network connection between client and cloud
- Data encryption
- Application security

Cloud computing: Disadvantages

- Cost?
- Communication with the cloud
  - Unless you prepare dedicate line to connect to the cloud, your communication with the cloud is going through the Internet.
    - reliable?
    - secure?
- Service providers' problem
- Legal compliance
  - e.g. GDPR requirement of encryption, data security, data physical location

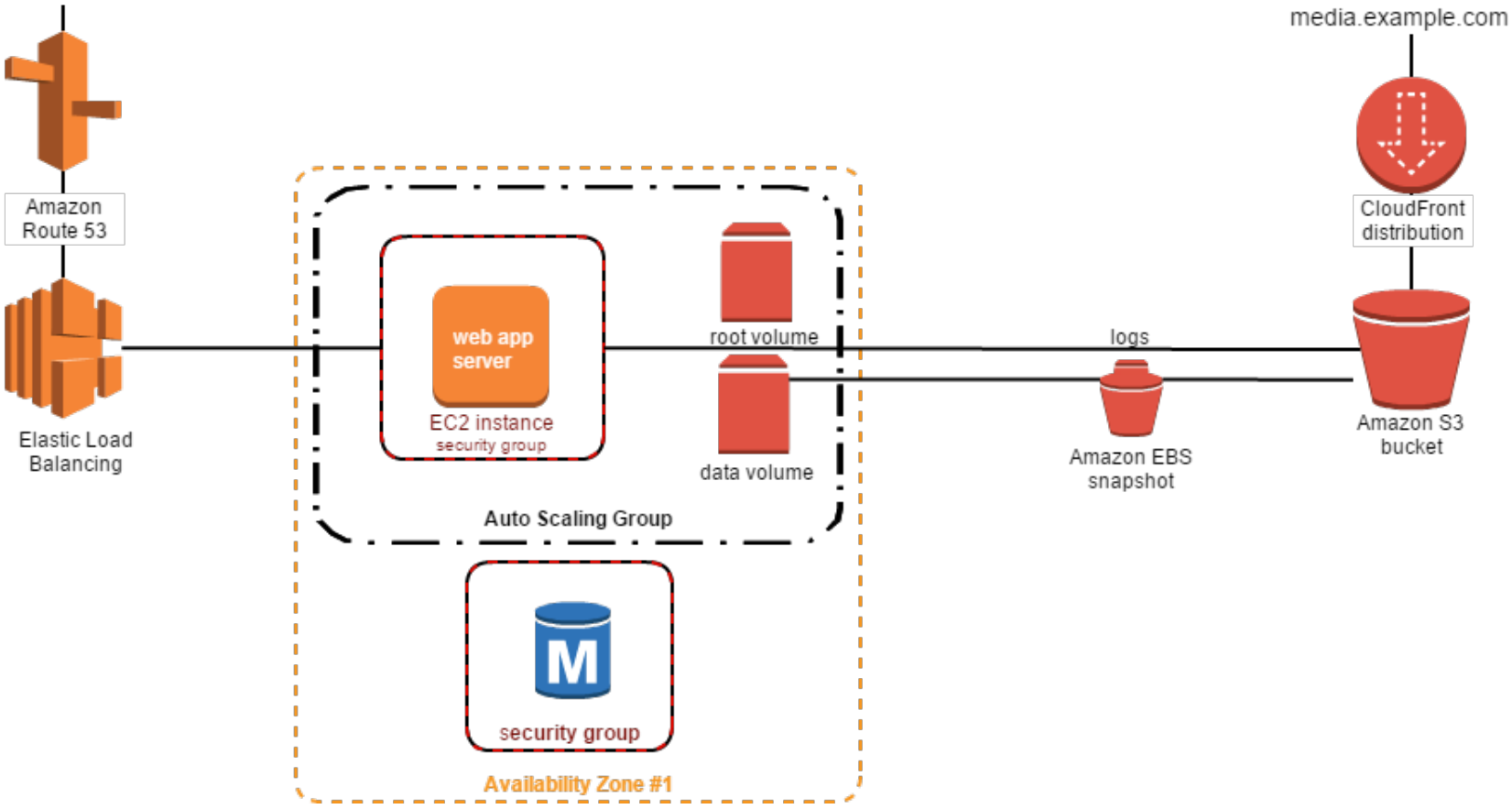
Cloud trend: Going serverless

Severless computing

- So far what we have seen is based on the idea of “replacing what you have done with cloud service”
- Serverless computing goes one step further

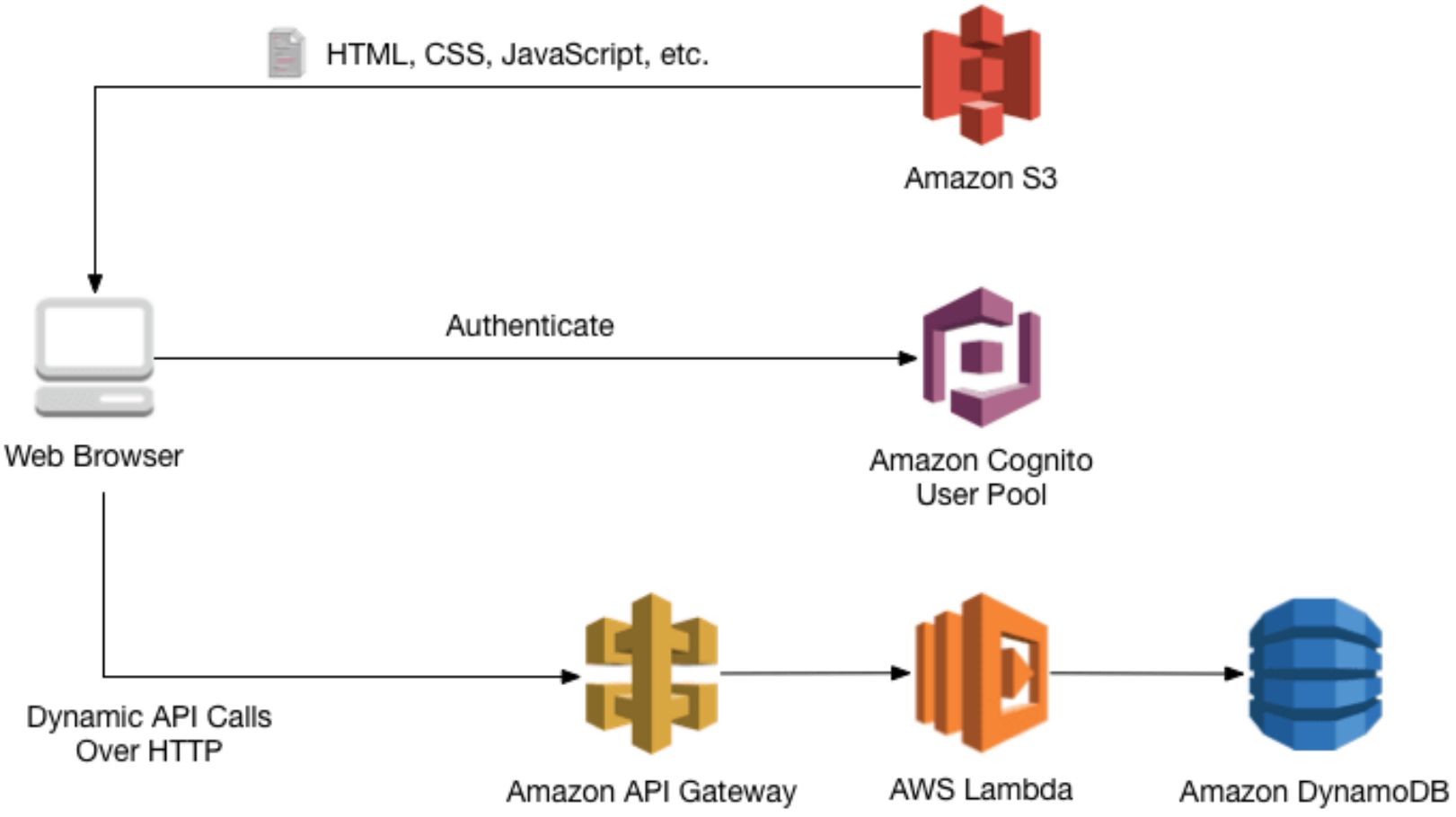
Web app with servers example

Link



Serverless example

Link



Advantages and disadvantages of serverless computing

Advantages

- Good for real time insights
- Scalable
- Usually cheaper

Disadvantages

- New/probably unfamiliar idea
- Not for memory intensive tasks (e.g. 5min limit for AWS Lambda)
- Not really for tasks that takes long time to process

Severless computing services

The core of serverless computing is a service to run codes without a server

- AWS Lambda
- Azure “Functions”
- Google Cloud “Functions”

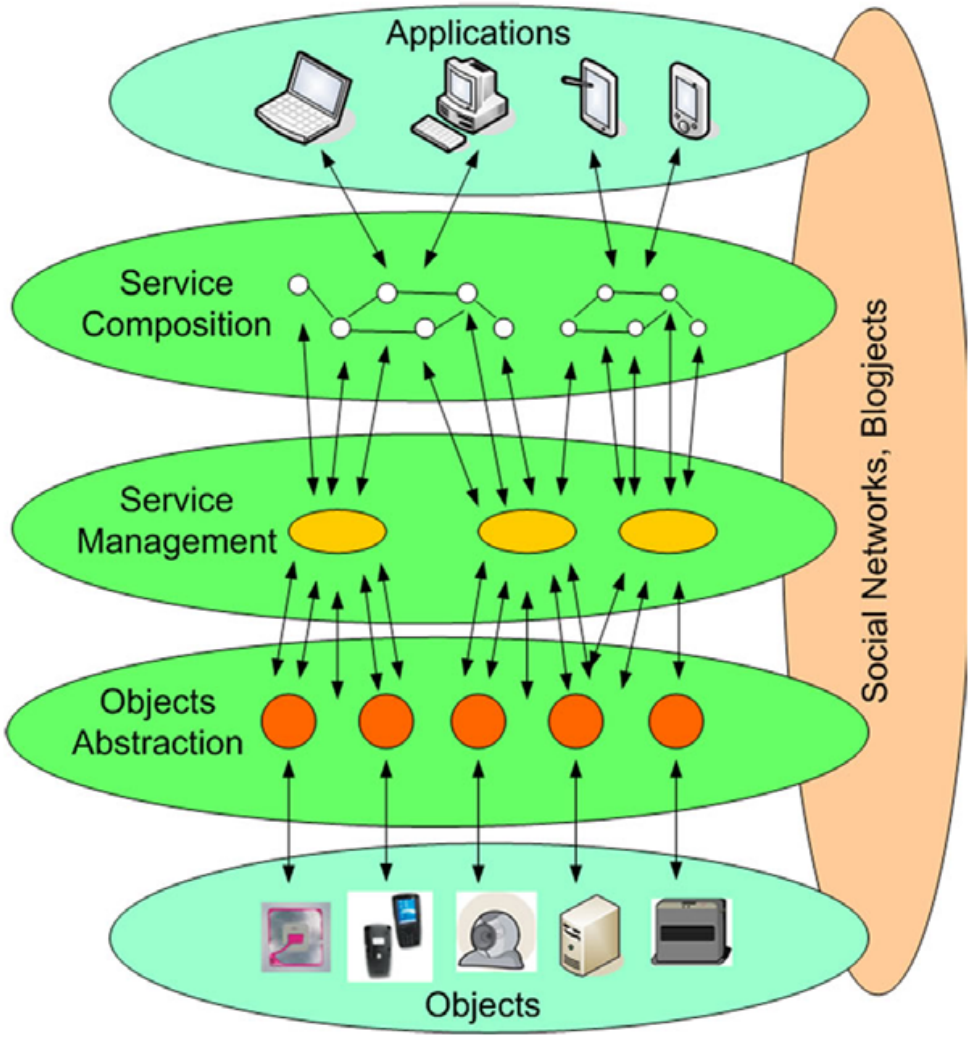
Future step: Internet of Things (IoT)

Definition

The internet of things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

From *Tech Target*

IoT: Illustration



Cloud and IoT

- IoT: A number of objects have sensors and send the data to cloud.
- Cloud: Aggregation of data from object, generate insights, send them to your devices (e.g. computer/smartphone) q
- Example: Healthcare (Botta et al, 2016)
  - "collecting patients' vital data via a network of sensors connected to medical devices, delivering the data to a medical center's Cloud for storage and processing, properly managing information provided by sensors, or guaranteeing ubiquitous access to, or sharing of, medical data as Electronic Healthcare Records (EHR)"

Week 3 Lab

From now, we will see the general overview of Amazon Web Service, and then in the lab, we will actually do some work with AWS.

Before your lab

- Accept the invitation to the AWS Educate classroom (create AWS Educate Starter Account)
  - This account will let you play with AWS without risk (i.e. account compromised)
- For Windows users, install git from https://git-scm.com/download/win
  - We need bash for git