Cloud Computing (ECS781P)

Week 1: Basics

Dr Arman Khouzani, Dr Felix Cuadrado

Electronic Engineering and Computer Science
Queen Mary University of London

January 12, 2018

Learning Outcomes

- Background and Motivation
  - Cloud Computing Origins
  - Motivation (Business Drivers)
  - Technology Innovations that inspired Cloud Computing
- Basic Concepts
  - Definition
  - The Five Essential Characteristics
  - Main Service Delivery Models
    - IaaS, PaaS, SaaS
  - Deployment Models
    - Private, Community, Public, Hybrid

Overview

Cloud Computing: Origins

- Computer Scientist *John McCarthy* in 1961:

"If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is ... The computer utility could become the basis of a new and important industry."

- **1963**: J.C.R. Licklider, a director at *Advanced Research Projects Agency (ARPA)*, introduced the concept of *Intergalactic Computer Network*.

- **1960** Leonard Kleinrock (@ARPANET):

"As of now, computer networks are still in their infancy. But as they [...] become more sophisticated, we will probably see the spread of computer utilities."

Cloud Computing: Origins

(*Fast Forward 30 years*):

- **Late 1990s**: `Salesforce.com` pioneered remotely provisioned services.

- **2002** `Amazon.com` launched *Amazon Web Services (AWS)* platform.

- **2006**: The term Cloud Computing emerged commercially. `Amazon` launched *Elastic Compute Cloud (EC2)*.

- **2009**: `Google` offered *Google App Engine*.

Cloud Computing Motivation (Business Drivers)

- From the point of view of the *consumer* of IT resources/services:
  - Relief from "Capacity" Planning
  - Cost Cutting
  - Business "Agility"
- From the point of view of the *providers* of IT resources/services:
  - Resource pooling, Multiplexing, and Multitenancy
  - Standardized Solutions and Services

Motivation (Business Drivers)

Relief from "Capacity" Planning:

- **Capacity**: The maximum amount of "work" that an IT resource can deliver per unit of time (communication, processing, memory, storage, etc.)
- Capacity planning strategies: *Lead*, *Lag*, and *Match*.
- The problem with Capacity Planning: There is a discrepancy between peak usage requirement and average usage. Peak usage itself fluctuates.
  - Accommodating for worst case peak usage will lead to **over-provisioning** and impose a huge investment cost.
  - Moderating expenditure can lead to not meeting the business needs at peak usages **under-provisioning** and problems like transaction losses, customer dissatisfaction, etc.

Motivation (Business Drivers)

Cost Cutting

- Two sources of IT-related costs:
  - **CAPEX**: Capital Expenditure (buying the Servers, Network Subsc., Internal Net with Firewalls, Disk Storage, System Software like OS, device drivers, database, middleware, etc, Application Software)
  - **OPEX**: Operational Expenditure (paying for technical staff, utility bills, upgrades and patches, license management, etc.)
- Instead of investing up-front for computing resources that may only be used sparingly and at less than peak capacity $\rightarrow$ signing-up with a service provider and only pay for resources that it uses.
- Hence, moving from CAPEX to OPEX model
- But also potentially reducing on OPEX (no datacenter cooling, patching, license mgmt, etc)

Motivation (Business Drivers)

Organizational Agility

- Agility: A measure of how responsive an organization is to *change*.
- This *change* can be:
  - "externally" driven, e.g., unanticipated shift (up or down) in customer demand, or requirements (like more reliability of a service); or
  - "internally" driven, e.g., when a business wants to try a new solution/product, or scrap an old one.
- Traditional IT provisioning has low agility Computing as a utility paradigm can address these problems.

Motivation (Business Drivers)

Resource pooling, Multiplexing, and Multitenancy

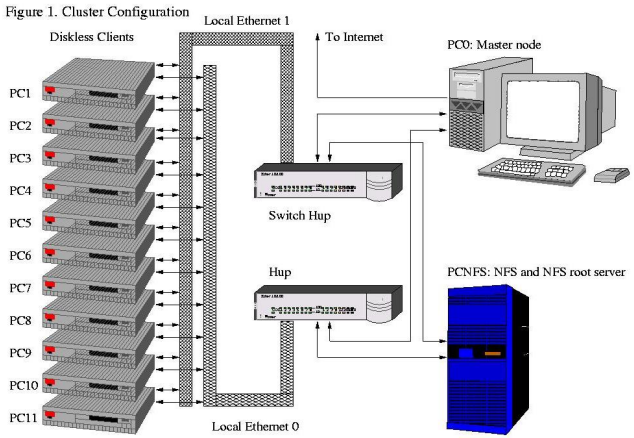
- Average computing resources by typical user at *peak*:
  - ▷ Proc.: 10% ▷ Mem.: 60% ▷ Net. BW: 20%
- At non-peak, it is even significantly less.
- Similar phenomenon in a data-centre:
  - ▷ e.g. Web / Application / Data-base Servers
  - Resource Pooling: Consolidating many individual resources to create a virtualized large resource.
  - Multiplexing: Pulling together statistically varying demands & serving them commonly by dynamically allocates resources as per a need basis.
  - ▷ Multiplexing Gain: *not every resource is required by all users/services at the same time!*
  - Multitenancy: Takes this idea even further; the same instance of a software can be used simultaneously by multiple user groups (*tenants*), hence less overhead.

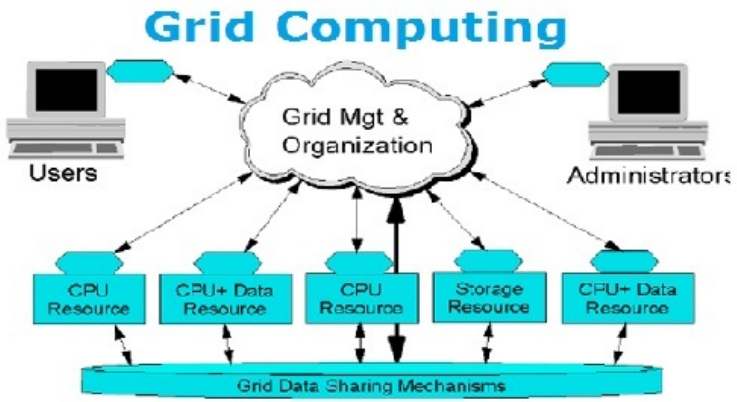
Motivation (Business Drivers)

Standardized Solutions and Services

- Instead of providing distinct services to different organizations with operationally unimportant but technically cumbersome idiosyncrasies, the provider offers a range of optimized, standard services to all.
- Hence, e.g., the provider does not have to re-invent all the security and reliability mechanisms for each new customer.
- Maintaining these standardized solutions is also much more smooth, e.g.:
  - *Patch Management* and *Updates* can be done for one core image, and all of its instances will be fixed.
- These two mean the Cloud Provider can save costs compared to traditional IT service providers.

Technology Innovations that inspired Cloud Computing

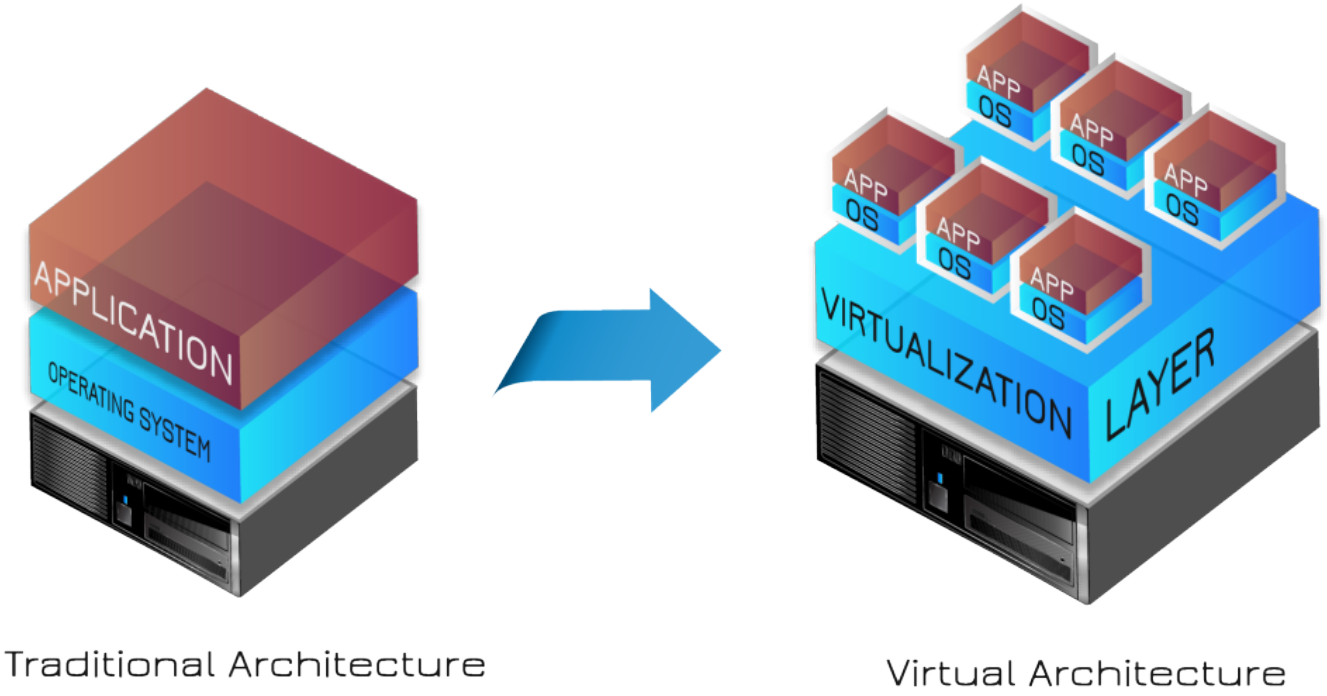




Clustering: almost identical hardware and OS for **failover** or failure recovery to increase availability (increasing the uptime), reliability (fault-tolerance – behaving as expected)

Grid Computing or a **super virtual computer**: pooling together heterogeneous computing resources (potentially geographically dispersed) and organizing them to emulate a single well-resourced machine.

Technology Innovations that inspired Cloud Computing



Traditional Architecture

Virtual Architecture

Virtualization: enabled to decouple applications from the underlying hardware, drivers, OS. Enabled sharing of the underlying processing capabilities by multiple users that need their own platform/hardware requirement. Enabled maintaining images for a standardized service and spawning as many copies of them as demanded, . . .

Basic Concepts

Basic Concepts: Definition

According to the National Institute of Standards and Technology (NIST) (in 2011):

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."

5 Essential Characteristics

5 Essential Characteristics of Cloud Computing

NIST lists 5 **essential chatacteristics** (defining properties) for cloud computing:

1. Ubiquitous (Broad) Network Access
2. On-demand Self-Service
3. Rapid Elasticity or Expansion
4. Resource Pooling
5. Measured Service

5 Essential Characteristics

Ubiquitous Network Access:

- The property that the cloud service (the IT resource that it provides) be widely accessible.
  - This requires the cloud to support a range of transport protocols, interfaces, and security technologies.

5 Essential Characteristics

On-demand Self-Service:

► The property that the cloud consumer can access and scale the offered cloud-based resources independently (through a graphical or command-line interface) and be able to control them at will, with no need for any decision-input/approval by the cloud provider or any disruptive impact on the cloud's operation. The cloud should also have the option to automate the process of resource provisioning whenever needed (at the runtime).

5 Essential Characteristics

Rapid Elasticity:

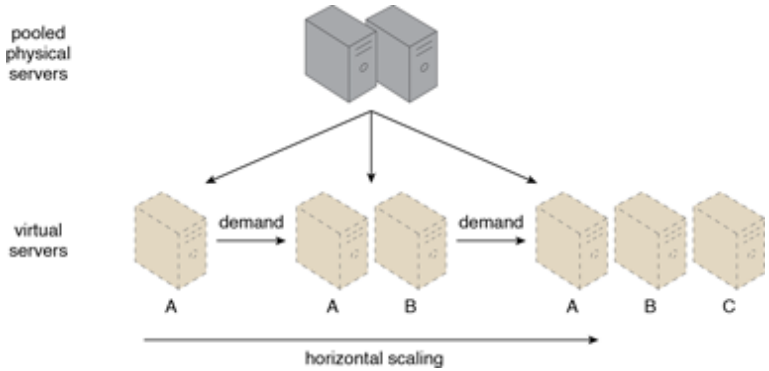
- **Elasticity (scalability)**:

  the degree to which a system is able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible.[1]

- There are two types of scaling:
  - Horizontal: Allocation/Releasing of resources *of the same type*, e.g. nodes, servers, instances, etc.
  - Vertical: Adding/removing capacity to/from (or replacing) an *existing* IT resource.

[1] Herbst, Nikolas Roman; Samuel Kounev; Ralf Reussner. *Elasticity in Cloud Computing: What It Is, & What It Is Not* ICAC 2013.

5 Essential Characteristics



Horizontal Scaling: Scaling Out / Scaling In

Vertical Scaling Scaling Up / Scaling Down

5 Essential Characteristics

Resource Pooling:

► Aggregation of computing resources such as storage, processing, memory, and network bandwidth across different physical machines so that they can be used by multiple users and dynamically assigned and released per demand.

5 Essential Characteristics

Measured Service:

► Cloud systems transparently monitor, control and report the usage of resources (e.g., storage, processing, bandwidth, and active user accounts) according to the abstraction level of the service, providing transparency to both the provider and consumer of the service.

Useful terminologies

- **Q: What are IT resources?**

▷ **A:** Any asset that provides an IT service. This includes physical/virtualized resources like Servers (computers where applications reside), Databases (computers where data resides in a structured way), Workstations or Clients (where the applications to communicate with and use the servers reside), network devices (access points, switches, routers, bridges, backbone, backhaul), or software (firmware, Drivers, OS, middleware, Applications), or human resources (IT/mgmt expertise)

- **Q:What's the opposite of a cloud-based resource?**

▷ **A: On-Premise**: in contrast to the "cloud"–based resources are remotely provisioned, and are scalable and measured, the conventional non-cloud based resources are called *on-premise* (in that they are located on the "premises" of the IT enterprise).

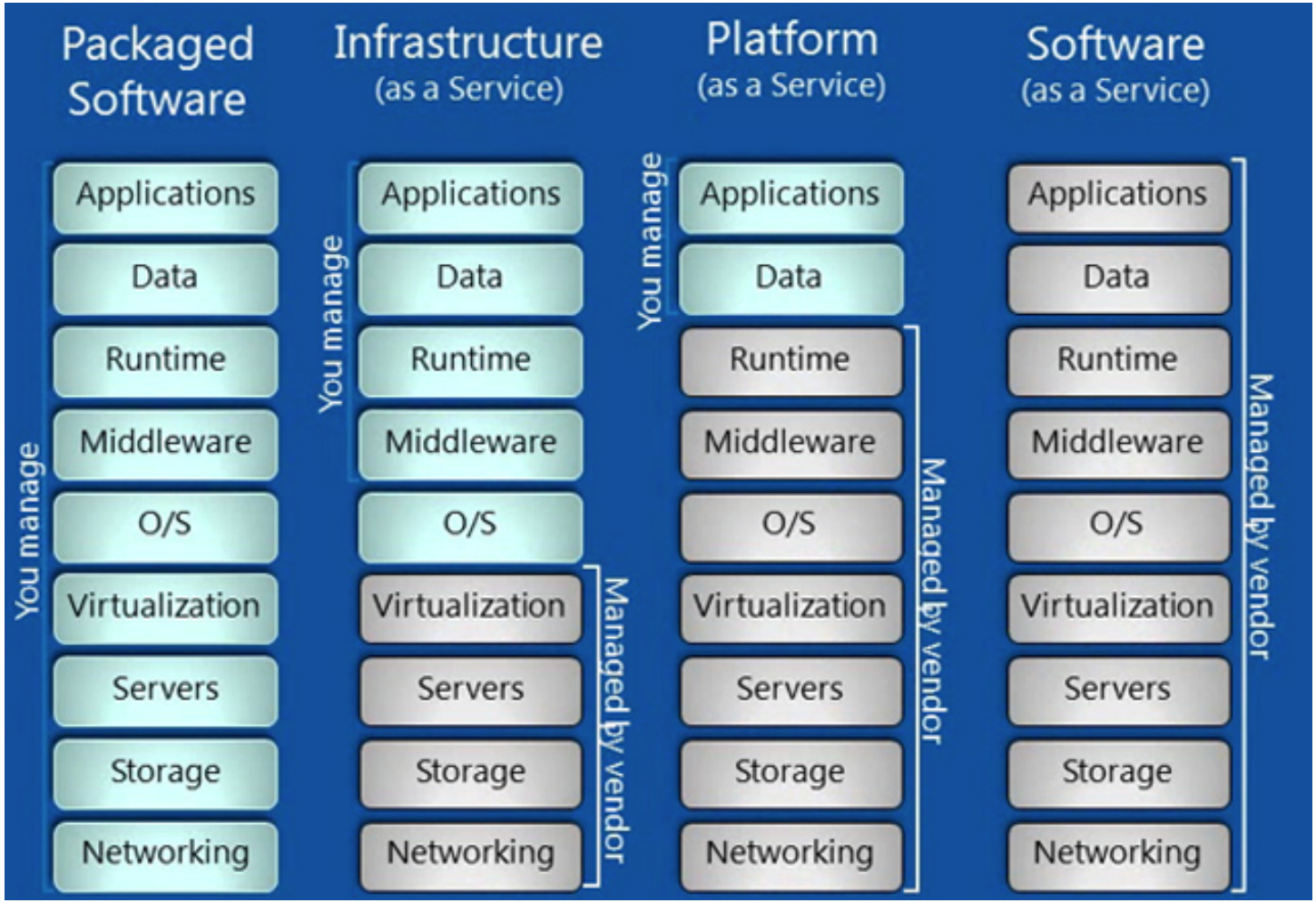
Cloud Service Models

NIST lists out three service (delivery) models based on the abstraction level:

- Infrastructure as a Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Service Models

3 Main Service Models



NIST lists 3 distinct service models: IaaS, PaaS, SaaS.

3 Main Service Models: IaaS

The NIST definition of Infrastructure as a Service (IaaS):[2]

> *"The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls)."*

[2] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011): `http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf`

3 Main Service Models: IaaS

- So the main resources being provided:
  ▷ Servers (with CPU, Cache, RAM, etc),
  ▷ Disk Storage ▷ Networking.
- In particular, the consumer has to install the OS, the application server (database server, the web server), job scheduling, resource usage governance, scaling, auto-scaling, caching, middle-ware (messaging between applications), distributed computing, compilers/builders, deployment and integration tools, all the necessary runtime libraries, add-ons and plug-inns, and of course, their applications (including their application security/API/admin/etc).
- What they get is a virtually "infinite" pool of computing resources that they can use-per-need and pay-per-use.

3 Main Service Models: IaaS

- `Instagram`: a success story of IaaS:
  - Three programmers on a bootstrap budget launch a photo-sharing application, 25,000 sign up the first day, 40 million in less than 2 year, when facebook acquired it for $\sim$ 1$ Billion.[3]
- Another good-use example:
  - When a consumer wants to build and deploy different server configurations, "test" and decide among them (even as an on-premise solution). So instead of hundreds of thousands of pounds this can be done in couple of tens.

[3] As of Sep 2017 it was about 600 million active users monthly: `https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/`

3 Main Service Models: IaaS

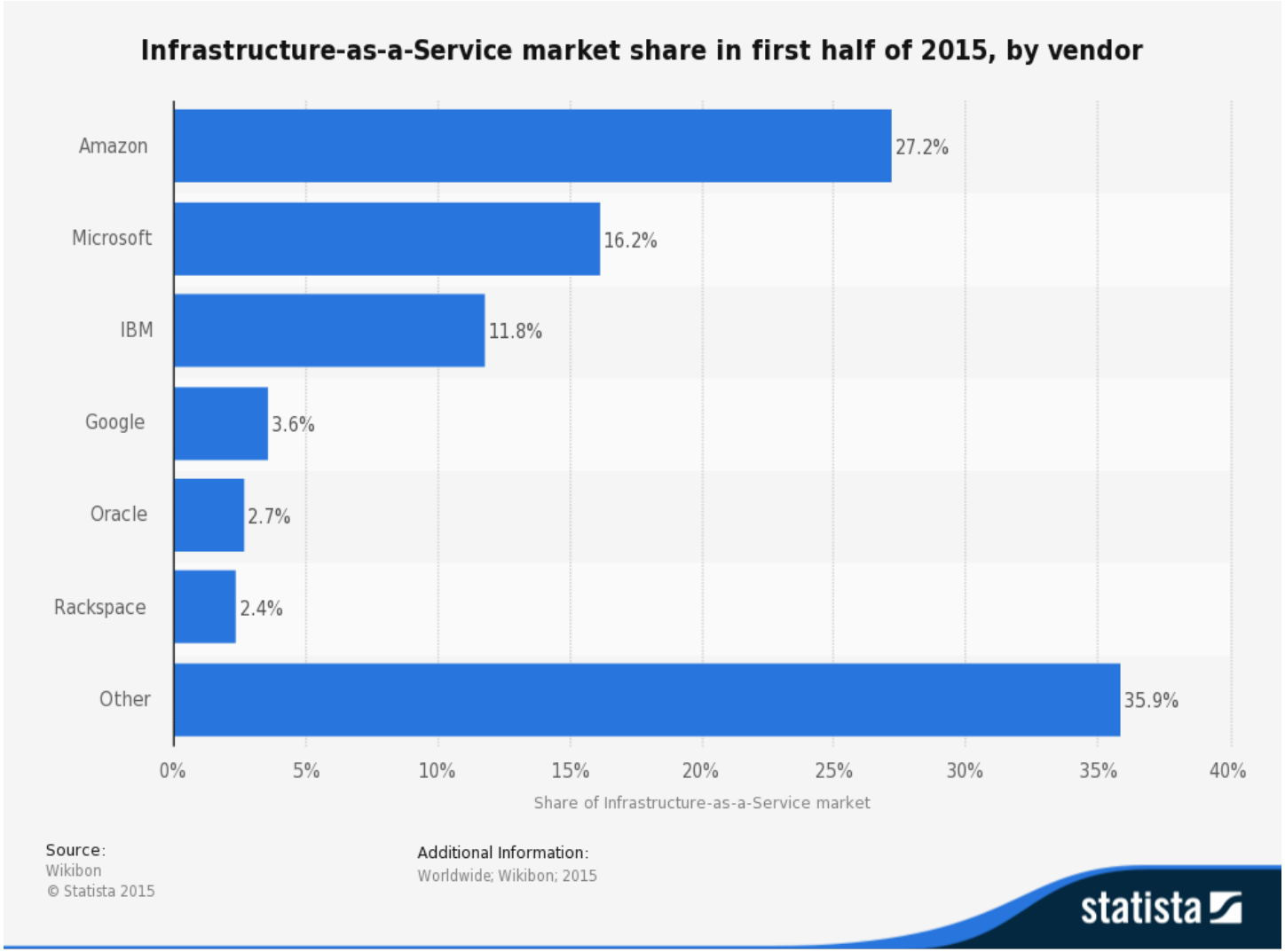
Some currently well-known IaaS vendors:

- Amazon Elastic Compute (EC2), Microsoft Azure, IBM Softlayer, Google Compute Engine, RackSpace Cloud Servers, vmware, CloudSigma, GoGrid, Joyent

Google Compute Engine

- But consumers can also build their own IaaS capabilities in-house (i.e., a private cloud) using **OpenStack** (say, to avoid vendor lock-in)

3 Main Service Models: IaaS



Source: https://www.statista.com/statistics/478143/iaas-vendor-market-share-ranking-worldwide/

3 Main Service Models: PaaS

NIST's definition of PaaS:[4]

> *"The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created* ***using*** *programming languages, libraries, services, and tools* ***supported by the provider****. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment."*

[4] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011): `http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf`

3 Main Service Models: PaaS

- PaaS provides a platform to *build*, run, deploy and *manage* applications on the cloud.
- Unlike in IaaS, the PaaS vendors take care of tasks like: automatic scaling based on traffic/usage load (scaling up/down and/or scaling out/in), which involves mechanisms like caching, asynchronous messaging, database scaling, etc., as well as providing the OS, the compilers/builders, the web/database servers, runtime libraries, testers, backups, deployment tools, etc., so that the consumers (developers) focus on the business logic and not re-inventing the wheel in implementing these IT-"plumbing" requirements.

3 Main Service Models: PaaS

Some famous PaaS providers (vendors):

- Amazon Web Services (AWS), Redhat OpenShift, Microsoft Azure, , Google App Engine, Heroku Enterprise, Forece.com and AppExchange by salesforce, SAP Hana, CloudBees, Engine Yard, Centurylink AppFog, IBM Bluemix...

- Besides these, tools like Dokku and OpenShift allow enterprises to build their own private PaaS.

3 Main Service Models: PaaS

- In return for agility and speed to market, consumers are constrained by the tools and software stacks that the vendor offers.
- They also give up a degree of control: e.g., they cannot manage memory allocation, stack configuration, number of threads, amount of cache, patch levels, etc.
- This means, e.g., that scalability is not as flexible as an IaaS solution (e.g. in the case of Instagram, they went with AWS, although the famous platform for Python was Google Apps Engine).
- Mature PaaS vendors also offer third-party add-ons (plug-ins, or extensions). These frees the developers from re-invent the wheel or obtaining and renewing licenses, install, patch, etc. and only pay-per-use:
  - Logging, Monitoring, Security, Caching, Search, E-mail, Analytics, Payments, etc.

3 Main Service Models: SaaS

NIST's definition of SaaS:[5]

> *"The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface (API). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings."*

[5] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011): `http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf`

3 Main Service Models: SaaS

Some very common SaaS applications are:

- Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), payroll, accounting, video-conferencing, emails, etc.

- There are thousands of SaaS providers, but the big money-makers serve enterprises.

3 Main Service Models: Quizz

- **Q1:** What is common among all consumers of an IaaS vendor?

3 Main Service Models: Quizz

- **Q1:** What is common among all consumers of an IaaS vendor?
- **A:** The infrastructure (each will have their own set of virtual machines)

3 Main Service Models: Quizz

- **Q1:** What is common among all consumers of an IaaS vendor?
- **A:** The infrastructure (each will have their own set of virtual machines)
- **Q2:** What is common among all consumers of a PaaS vendor?

3 Main Service Models: Quizz

- **Q1:** What is common among all consumers of an IaaS vendor?
- **A:** The infrastructure (each will have their own set of virtual machines)
- **Q2:** What is common among all consumers of a PaaS vendor?
- **A:** The platform. Each can have their own applications deployed on it.

3 Main Service Models: Quizz

- **Q1:** What is common among all consumers of an IaaS vendor?
- **A:** The infrastructure (each will have their own set of virtual machines)
- **Q2:** What is common among all consumers of a PaaS vendor?
- **A:** The platform. Each can have their own applications deployed on it.
- **Q3:** What is common among all consumers of a SaaS vendor?

3 Main Service Models: Quizz

- **Q1:** What is common among all consumers of an IaaS vendor?
- **A:** The infrastructure (each will have their own set of virtual machines)
- **Q2:** What is common among all consumers of a PaaS vendor?
- **A:** The platform. Each can have their own applications deployed on it.
- **Q3:** What is common among all consumers of a SaaS vendor?
- **A:** The application. Each can have their own customers using them.

3 Main Service Models: Quizz

- **Q4:** A cloud service is allowing me to create virtual machines, what service model (delivery model) is it?

3 Main Service Models: Quizz

- **Q4:** A cloud service is allowing me to create virtual machines, what service model (delivery model) is it?
- **A:** IaaS.

3 Main Service Models: Quizz

- **Q4:** A cloud service is allowing me to create virtual machines, what service model (delivery model) is it?
- **A:** IaaS.
- **Q5:** Which cloud service model allows me to run my own application?

3 Main Service Models: Quizz

- **Q4:** A cloud service is allowing me to create virtual machines, what service model (delivery model) is it?
- **A:** IaaS.
- **Q5:** Which cloud service model allows me to run my own application?
- **A:** Both IaaS and PaaS.

3 Main Service Models: Quizz

- **Q4:** A cloud service is allowing me to create virtual machines, what service model (delivery model) is it?
- **A:** IaaS.
- **Q5:** Which cloud service model allows me to run my own application?
- **A:** Both IaaS and PaaS.
- **Q6:** A cloud provider does not allow me to change the operating system (OS), which cloud service model is this?

3 Main Service Models: Quizz

- **Q4:** A cloud service is allowing me to create virtual machines, what service model (delivery model) is it?
- **A:** IaaS.
- **Q5:** Which cloud service model allows me to run my own application?
- **A:** Both IaaS and PaaS.
- **Q6:** A cloud provider does not allow me to change the operating system (OS), which cloud service model is this?
- **A:** Could be either PaaS/SaaS.

3 Main Service Models: Quizz

- **Q7:** My enterprise wants to run our applications "as is" on the cloud (migrate to cloud) in the current month. Which cloud service model is the most appropriate?

3 Main Service Models: Quizz

- **Q7:** My enterprise wants to run our applications "as is" on the cloud (migrate to cloud) in the current month. Which cloud service model is the most appropriate?
- **A:** IaaS.

3 Main Service Models: Quizz

- **Q7:** My enterprise wants to run our applications "as is" on the cloud (migrate to cloud) in the current month. Which cloud service model is the most appropriate?
- **A:** IaaS.
- **Q8:** Between the "network architect", "application developers (programmers)", and "end-users" who is the targeted consumer of PaaS cloud model?

3 Main Service Models: Quizz

- **Q7:** My enterprise wants to run our applications "as is" on the cloud (migrate to cloud) in the current month. Which cloud service model is the most appropriate?
- **A:** IaaS.
- **Q8:** Between the "network architect", "application developers (programmers)", and "end-users" who is the targeted consumer of PaaS cloud model?
- **A:** application developers.

3 Main Service Models: Quizz

- **Q7:** My enterprise wants to run our applications "as is" on the cloud (migrate to cloud) in the current month. Which cloud service model is the most appropriate?
- **A:** IaaS.
- **Q8:** Between the "network architect", "application developers (programmers)", and "end-users" who is the targeted consumer of PaaS cloud model?
- **A:** application developers.
- **Q9:** On the SLA (service level agreement) of which cloud service model you are most likely to find information about the available database servers?

3 Main Service Models: Quizz

- **Q7:** My enterprise wants to run our applications "as is" on the cloud (migrate to cloud) in the current month. Which cloud service model is the most appropriate?
- **A:** IaaS.
- **Q8:** Between the "network architect", "application developers (programmers)", and "end-users" who is the targeted consumer of PaaS cloud model?
- **A:** application developers.
- **Q9:** On the SLA (service level agreement) of which cloud service model you are most likely to find information about the available database servers?
- **A:** PaaS.

Other Cloud Service Models

These main service models need to be updated, as new models are invented:

- Container-as-a-Service (CaaS)
- Business Process as a Service (BPaaS)
- Information-as-a-Service (InaaS)

Deployment Models

4 Deployment Models

NIST lists the following 4 deployment models:[6]

Private: The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

- **Q:** Which enterprises are likely to use a private cloud? And why?

[6] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011)

4 Deployment Models

NIST lists the following 4 deployment models:[6]

Private: The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

- **Q:** Which enterprises are likely to use a private cloud? And why?
- **A:** Those with deep pockets and concern for security/privacy or strict regulations (government, finance, healthcare), e.g. Data Protection Act (DPA), Service Organization Control (SOC), Payment Card Industry Data Security Standard (PCI DSS), etc.

[6] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011)

4 Deployment Models

Community: The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

- *Examples of such communities*: all banks of a financial zone, all hospitals of a city, all universities of a country, etc.

4 Deployment Models

Public: The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

4 Deployment Models

Hybrid: The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).