CS 524

Introduction to Cloud Computing

Module 2: NIST definitions, basic building blocks, IT industry transformation (from Public Cloud/Shadow IT to Private Cloud), Network Function Virtualization

Igor Faynberg

OUTLINE

- Evolution: Mainframes, Grid computing, Cloud Computing
- NIST Definitions
- IT industry transformation
  - in-house data centers
  - Shadow IT
  - Private IT
- Services
- Technology building blocks

What makes Cloud Computing Different from the Mainframe Environment

- In the Cloud, an *internet appliance* can actually be a PC

  In the mainframe environment, the user has no independent computing power or storage

- In the Cloud, one gets one's own *virtual machine* (and even the whole infrastructure) and is unaware of other users

  In the mainframe environment a user gets a share of a machine and is typically aware of other users

- In the Cloud, a user gets theoretically infinite computer power

  In the mainframe environment the computing power is bounded (and typically depends on what other users use)

What makes Cloud Computing Different from the Grid Computing Environment

- In Cloud computing, only Cloud is actually a grid

  In the Grid computing environment, each end-point participates in computing

- In Cloud computing, a user is a client

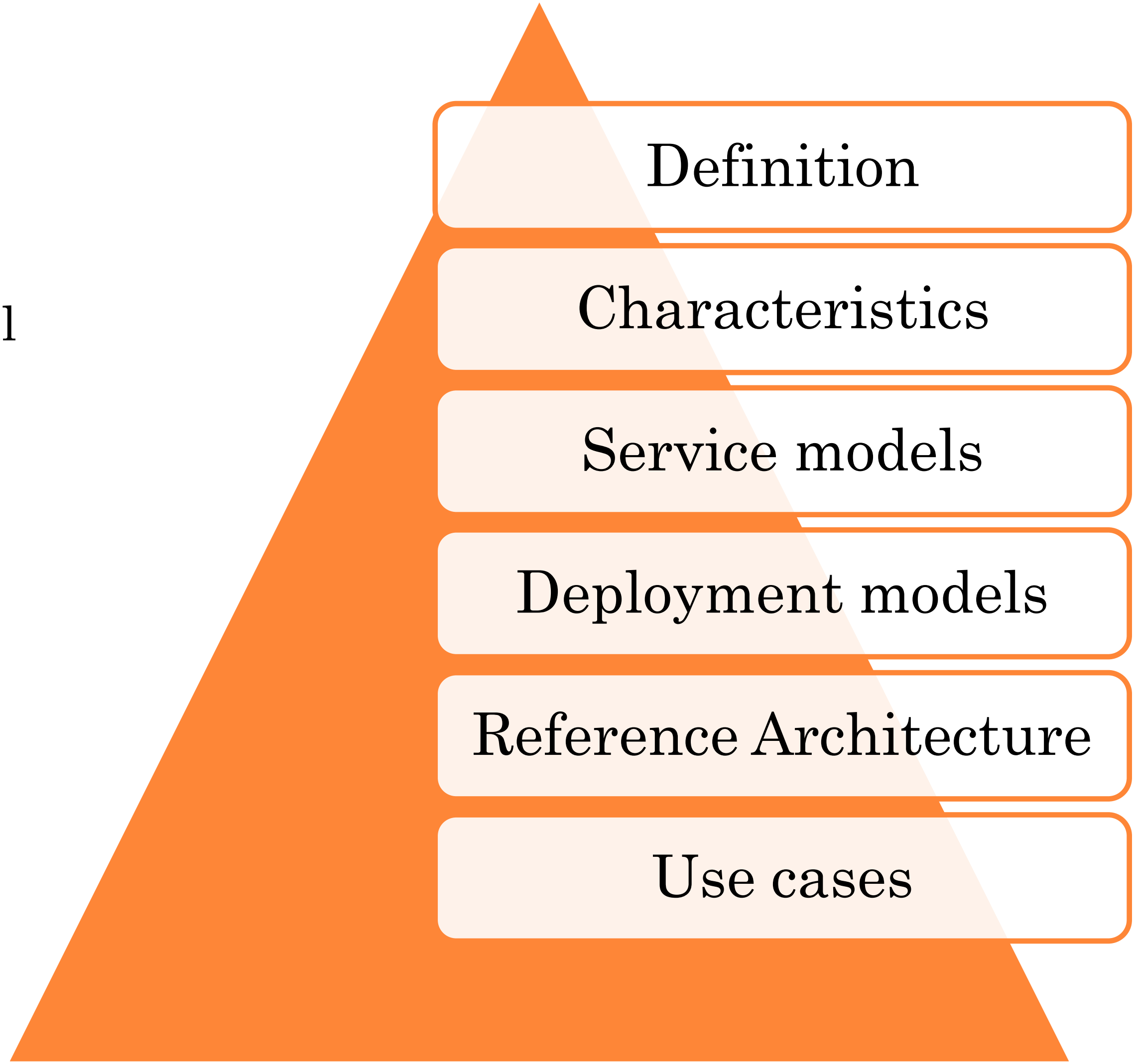
  In the Grid computing environment, a user is both a client **and** a server

- In Cloud computing, there is central control

  In the Grid computing environment there is no central control

Bringing in some structure…

Source: US National Institute of Standards and Technology (NIST)



Definition

Characteristics

Service models

Deployment models

Reference Architecture

Use cases

Cloud Computing: Definition

"… 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 of service provider interaction. This cloud model promotes availability and is composed of **five essential characteristics**, **three service models**, and **four deployment models**."

CHARACTERISTICS:

1. On-demand self-service

   = automatic provisioning of computing capabilities (CPU + storage)

2. Broad network access

   = *standard* mechanisms supporting multiple platforms (phones+PCs+…)

3. Resource pooling

   = one dynamic pool of physical and virtual resources (bandwith+CPU+storage+…) shared among multiple consumers using a ***multi-tenant*** model

4. Rapid *elasticity*

   = scalability + speed of acquisition and release

5. Measured Service

   = measurements to ensure automatic control and optimal resource use

Service Models

1. Cloud Software as a Service (SaaS)

   = specific application access

2. Cloud Platform as a Service (PaaS)

   = a platform (languages and other software tools), on which new software can be developed and deployed

3. Cloud Infrastructure as a Service (IaaS)

   = broad resource pool (i.e., multiple machines—in different locations, CPU, storage, network access)—without management responsibilities—in which new systems can be deployed)

Deployment Models

1. Private cloud

   = one organization (but may be managed by a third party)

2. Community cloud

   = several organizations united by a common set of requirements (security, policy, compliance, etc.)

3. Public cloud

   = available pretty much to all

4. Hybrid cloud

   = a combination of the above

Cloud Computing Reference Architecture

Actor	Definition
Cloud Consumer	uses services of *Cloud Provider*
Cloud Provider	makes services available to *Cloud Consumers*
Cloud Auditor	independently assesses the services of *Cloud Provider*
Cloud Broker	manages services of *Cloud Provider* and negotiates its relationship to *Cloud Consumers*
Cloud Carrier	provides connectivity and transport of services from *Cloud Providers* to *Cloud Consumers*

Service Level Agreement (SLA)

SLA documents the agreement between the Cloud Consumer and Cloud Provider

Deployment Use Cases

	A) Within Trust Boundary	B) Crossing Trust Boundary
1. Centralized (one administrative domain)	1A •"Simple" authentication •VM management •Storage management •Service discovery •Workflow management •Auditing • Virtual organizations	1B •SLAs in support of governance requirements •"Strong" authentication (PKI) •Certification of VM isolation (through hardware and hypervisor support) •Data encryption
2. Distributed (crossing domains)	2A •*Peer-to-peer* (P2P) service discovery •P2P SLA and performance monitoring •P2P workflow management •P2P auditing •P2P security •P2P virtual organization management	2B + P2P SLAs in support of governance requirements

An overarching set of NIST requirements

1. Creating, accessing, updating, and deleting data objects in clouds
2. Moving virtual machines and virtual appliances among clouds
3. Selecting the best *IaaS* vendor for private, externally-hosted cloud
4. Providing tools for monitoring and managing multiple clouds
5. *Orchestrating* services across clouds
6. Discovering cloud resources
7. Evaluating Service Level Agreements (SLAs) and penalties
8. Auditing

An Example: *Amazon Elastic Compute Cloud (EC2 or actually* EC^2)

A *web service* that provides resizable computing capacity in the cloud with these options:

- **On-Demand Instances** – pay as you go (no long-term commitments)
- **Reserved Instances** – make one-time payment for each instance you want to reserve and get a discount on the hourly charge for that instance
- **Spot Instances** – bid on unused Amazon EC2 capacity and run those instances for as long as their bid exceeds the current *spot price*, which changes periodically based on supply and demand

EC 2 Features

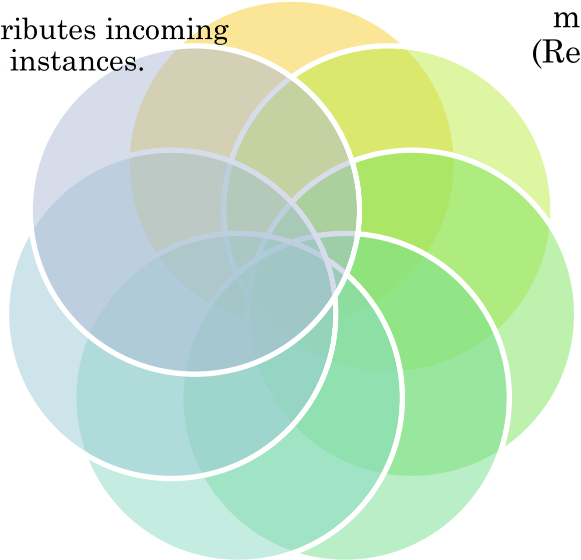
High I/O Instances
Provide customers very high, low latency, random I/O access to their data (no-SQL and relational databases)

Elastic Block Storage (EBS)-Optimized Instances enable EC2 instances to use fast I/O provisioned on an EBS volume (within 10% performance bracket 99.9% of the time.

Elastic Block Store offers persistent storage for Amazon EC2 instances.

Multiple Locations provides the ability to place instances in multiple geographic locations (Regions and Availability Zones).

Elastic Load Balancing automatically distributes incoming application traffic across multiple EC2 instances.



Auto Scaling allows to scale automatically the EC2 capacity up or down according to given conditions.

Elastic IP Addresses

CloudWatch provides monitoring for cloud resources and applications.

High Performance Computing (HPC) Clusters supports applications that demand parallel processing.

Amazon Virtual Private Cloud enables enterprises to connect isolated data centers.

AWS Marketplace
AWS Marketplace is an online store that helps to find, buy and quickly deploy software that runs on AWS.

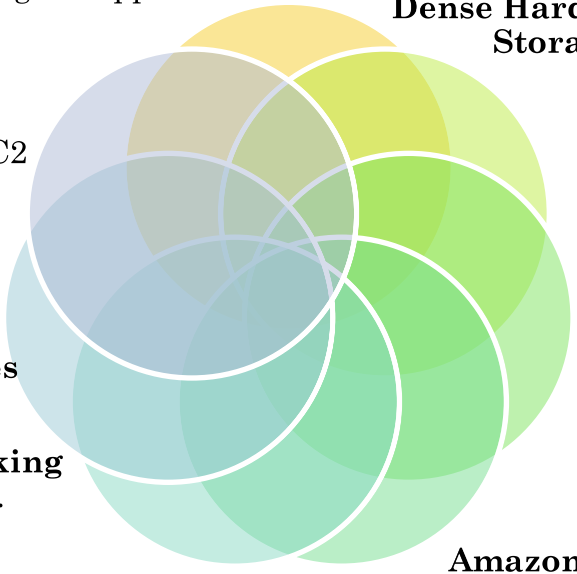
VM Import/export enables import of virtual machine images from a user's existing environment to EC2 instances as well as export back.

More EC2 Features

Flexible Storage Options

Bare Metal Instances
ideal for workloads that require access to hardware feature sets (such as Intel® VT-x), or for applications that need to run in non-virtualized environments for licensing or support requirements.

Optimize Compute Performance and Cost with Amazon EC2 Fleet
Provision compute capacity across EC2 instance types.



Dense Hard Disk Drive (HDD) Storage Instances

High Performance Computing (HPC) Clusters

Elastic Fabric Adapter
Fast interconnect for HPC clusters

Pause and Resume Your Instances

.

Enhanced Networking

.

Graphic Processing Unit (GPU) Compute Instances

.

Amazon Time Sync Service

.

IT Industry Transformation

Creating software-based services before the Cloud:

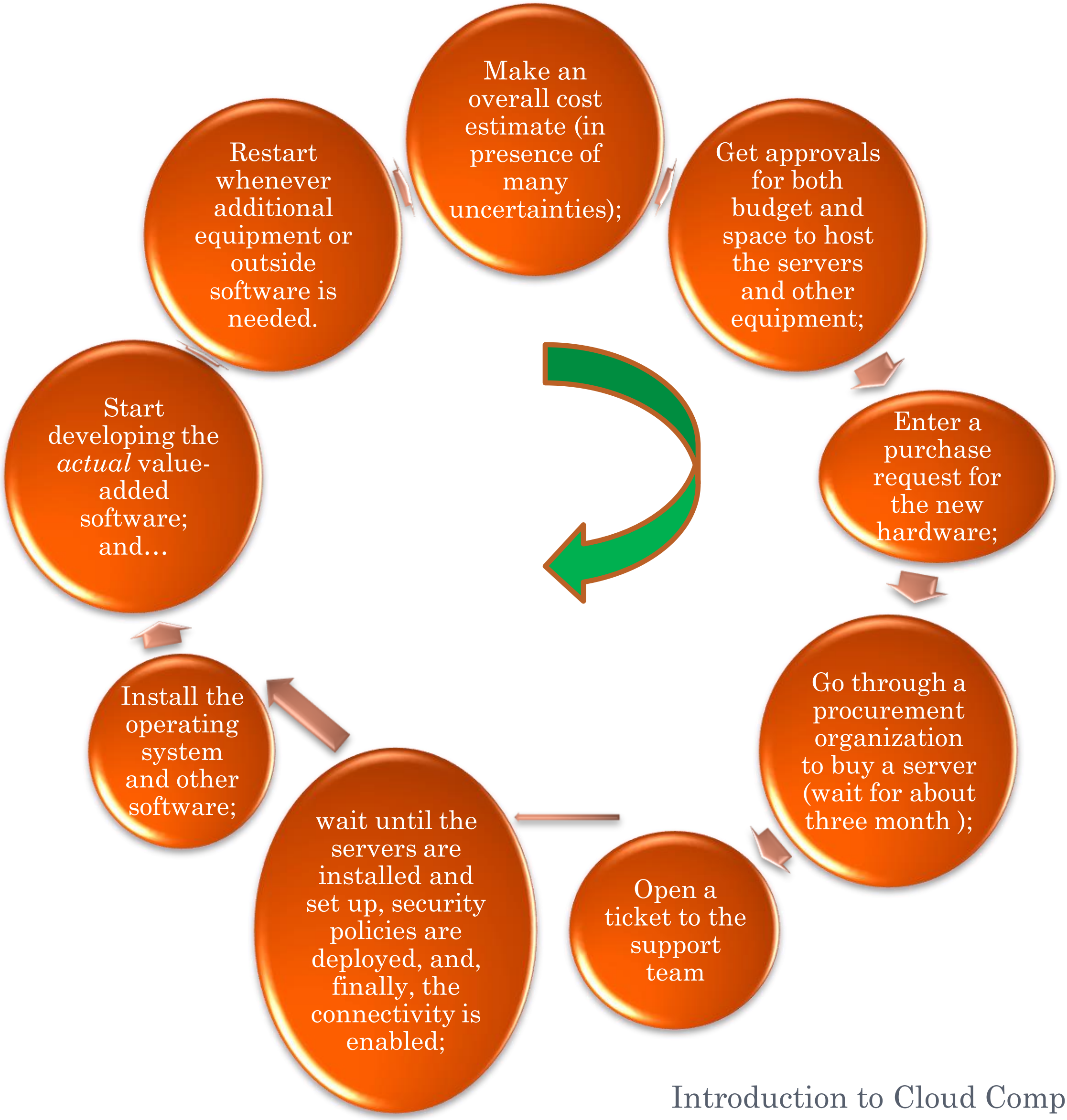
- High upfront investment, *capital expenditure (CAPex)*, in hardware and network connections based only on speculation regarding the demand
- Risk of losing this investment
- On-going *operational expenditure (OPex)*
- Slow *time-to-market*
- Required (and expensive) expertise in the physical infrastructure (servers, switching, storage) and software reliability infrastructure (clustering, monitoring)

Case Study (*Zing Interactive Media*), *2000*

- Rent space on a hosting site—the "cage" on an AT&T hosting facility
- Anticipate the peak use amount and develop a redundancy schema
- Specify the technical requirements for the servers to meet the capacity plans
- Negotiate vendor- and support contracts
- Lease T1 lines for the "cage" connectivity and pay for them regardless of the actual use (!)
- Purchase and install the computing and networking gear (load balancers, switches, firewalls, and other *networking appliances*)
- Purchase and install the **software platform** (operating systems, databases, software libraries, compilers)
- Hire an IT team of networking experts, a systems administrator, and a database administrator
- Deploy and maintain the unique software that actually delivered the service

And all of this was on top of the actual cost of building the product!

In a general case: an IT manager starting a new project has to



Make an overall cost estimate (in presence of many uncertainties);

Get approvals for both budget and space to host the servers and other equipment;

Enter a purchase request for the new hardware;

Go through a procurement organization to buy a server (wait for about three month);

Open a ticket to the support team

wait until the servers are installed and set up, security policies are deployed, and, finally, the connectivity is enabled;

Install the operating system and other software;

Start developing the *actual* value-added software; and…

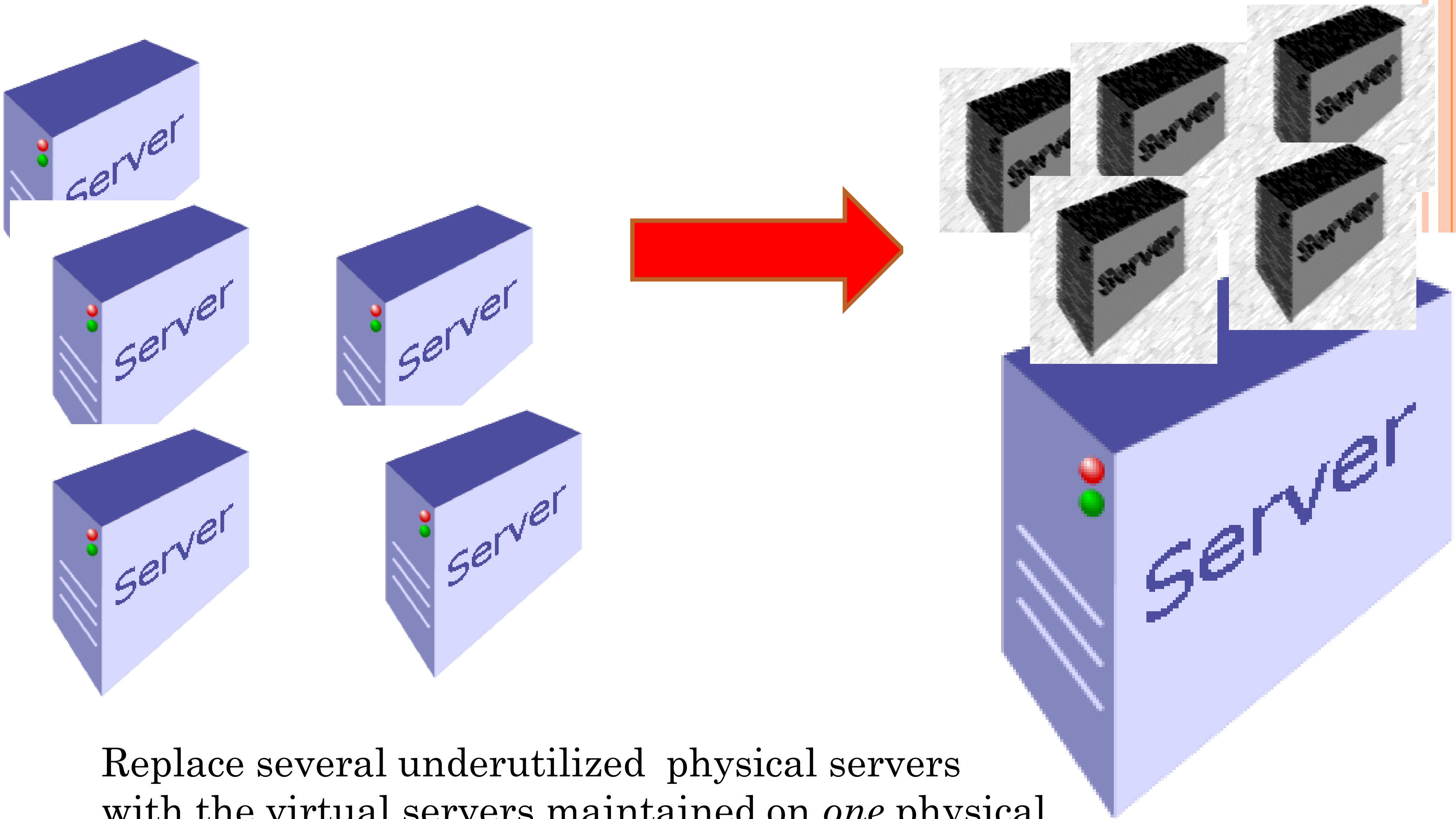
Restart whenever additional equipment or outside software is needed.

IT Transformation

- Took place to streamline the operations to keep both the developers and the service providers focused on the matter of *added value*
- Proceeded (roughly) in two phases:
  - Virtualization (rediscovered for the new purposes)
  - The full-blown Cloud

Cloud = Virtualization + Networking + Automation

VIRTUALIZATION (THE LATE CONTEXT)



Replace several underutilized physical servers with the virtual servers maintained on *one* physical

The achievement of virtualization

- Significant reduction of CapEx and OpEx (2/3 of the IT budget is devoted to the maintenance of hardware!)
- Improvement in availability (deploy a new instance in no time!)
- Improvement in reliability (easy to run several instances simultaneously)
- Improvement in higher CPU utilization

A new problem: Software licensing

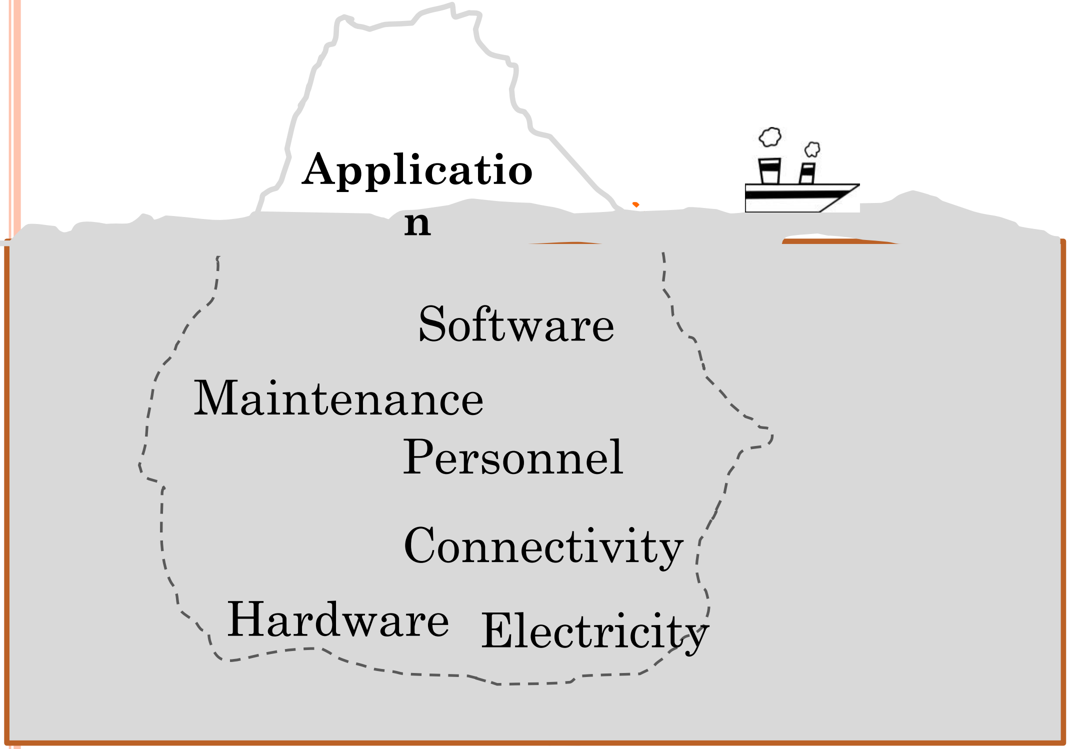
A remaining old problem: Static configuration

The Cloud as a major step forward

- "Pay-as-you-go" business model
- Full automation
- Low cost of ownership
- Availability of multiple platforms and services (databases, load balancing, Big Data/analytics, machine learning… the list is still growing!)
- Life-cycle management: auto-deployment, monitoring, auto-healing—all rule-based and defined by a user.

Consequence: The developer can concentrate on the application!

Before…



Applicatio
n

Software

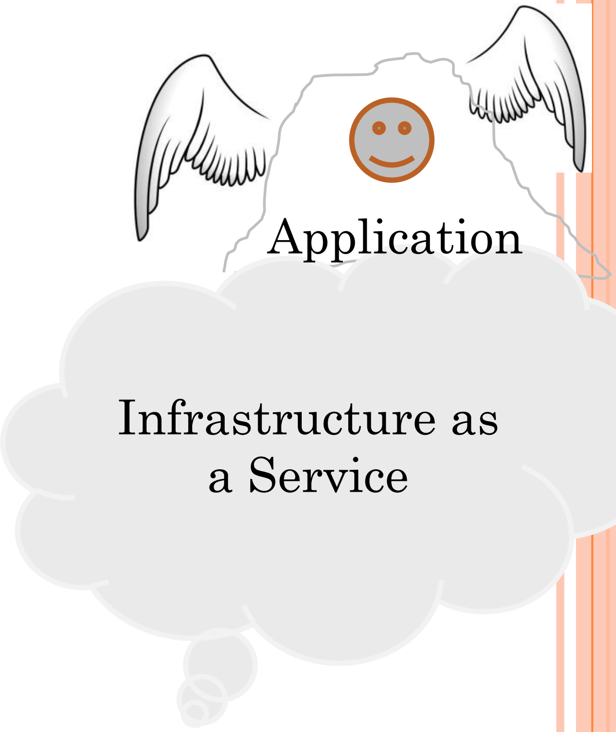
Maintenance

Personnel

Connectivity

Hardware Electricity

…and After



Application

Infrastructure as
a Service

AND SO THE ENTERPRISE IN-HOUSE DEVELOPERS

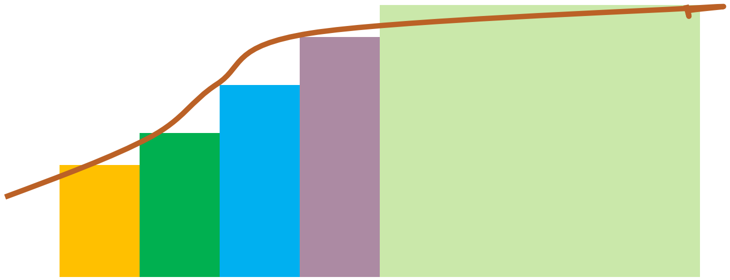
- Bypassed the IT department and

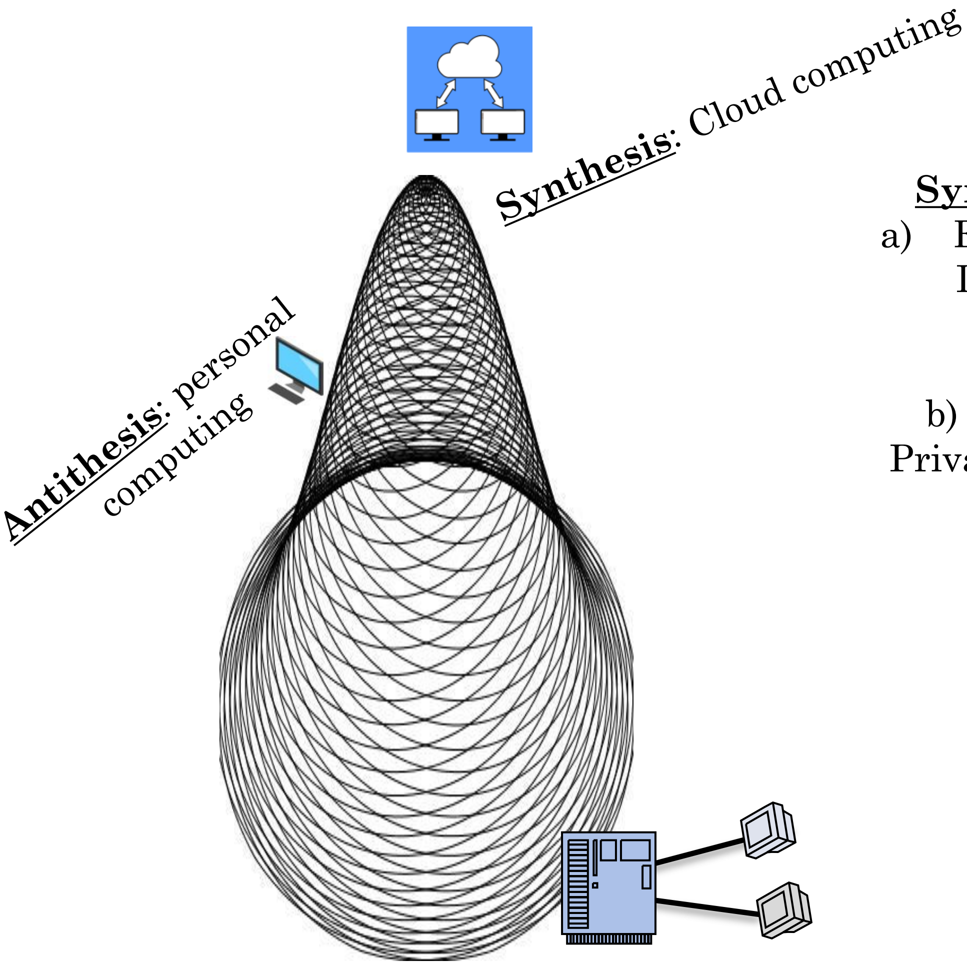
- Took the development to the Public Cloud (all one needed was a credit card!) and created the

Shadow IT

But the pure Public Cloud may be a security risk (and it is certainly a risk to job security)!

- CIOs observed this trend
- Thus two developments:
  - in-house *Private Cloud* (using *VMWare or OpenStack*)
  - *Virtual Private Cloud* as a service
- The business trend is to start in a public Cloud and then build the private Cloud after reaching a *scale-up plateau*





Synthesis: Cloud computing

Antithesis: personal computing

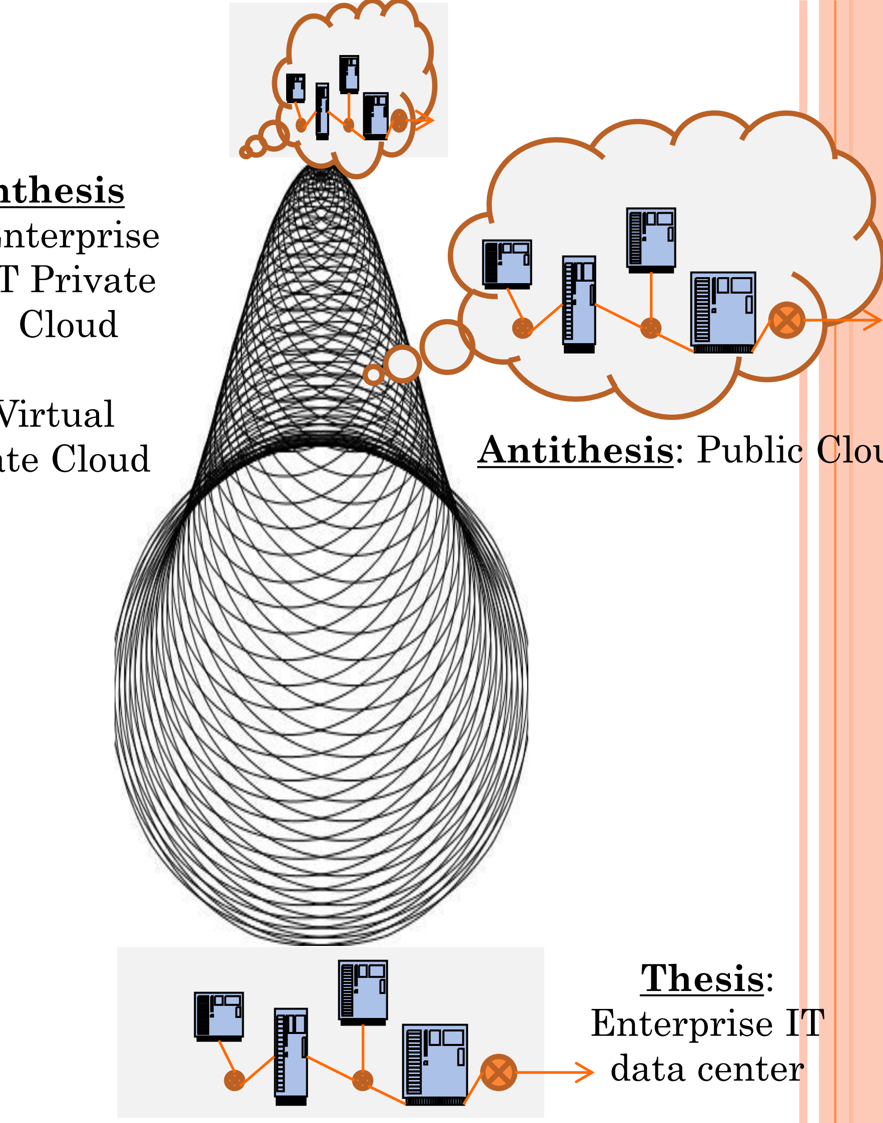
Thesis: mainframe computing

a) From mainframe to Cloud

Synthesis

a) Enterprise IT Private Cloud

b) Virtual Private Cloud



Antithesis: Public Clou

Thesis: Enterprise IT data center

b) From the IT data center to Private Cloud

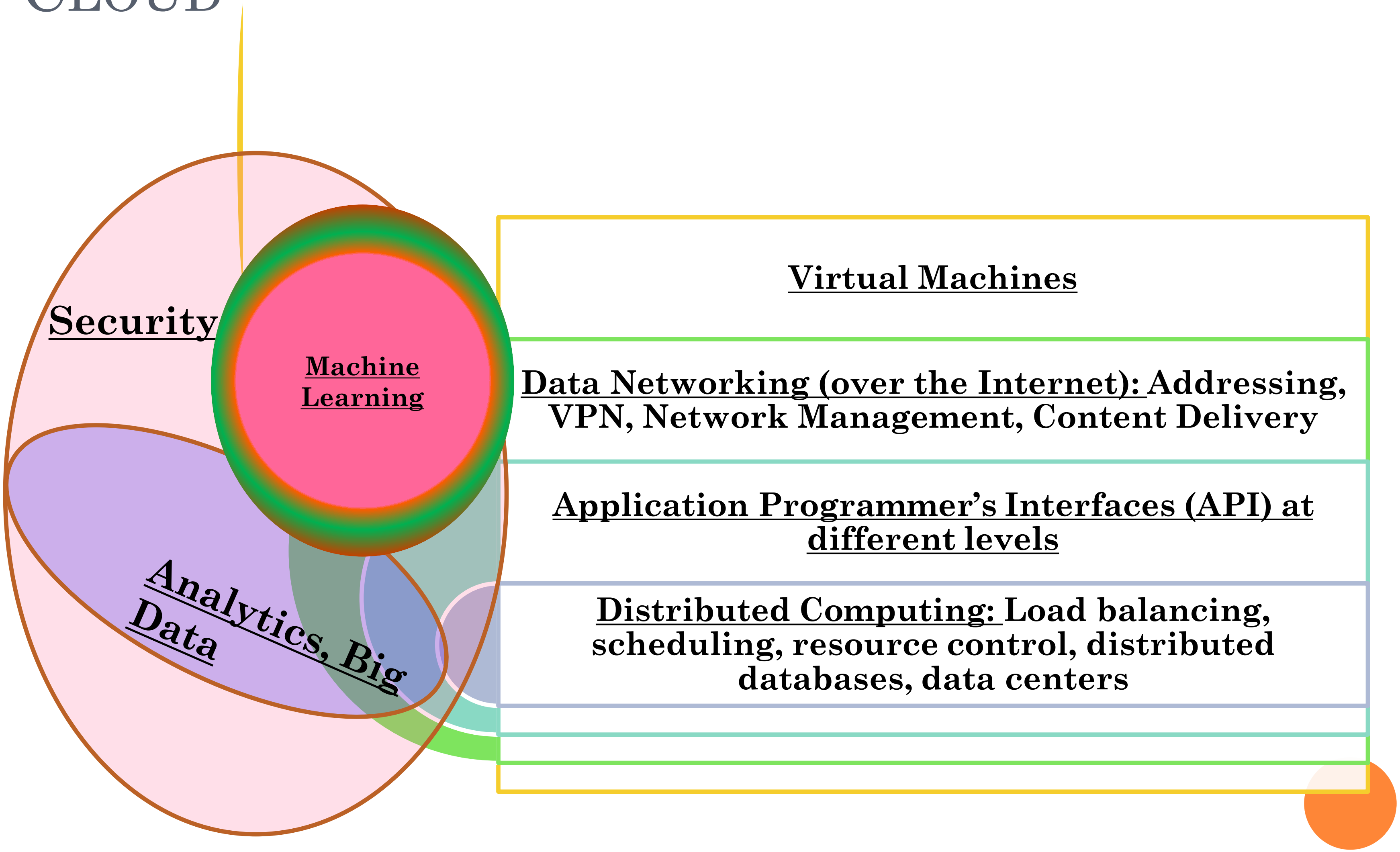
A Business Model Around Cloud

- Infrastructure examples
  - Google: more than 1,000,000 physical servers
  - Amazon: “space” for up to 2,000,000 virtual machines
  - Commodity hardware managed by small operational teams
- Cloud Providers’ model
  - An array of pricing options balancing end-users flexibility and budget
  - Higher prices for special services (Hardware Security Module, Virtual Private Cloud)
  - Introduction of *spot pricing*
- Software Vendors’ model
  - Proprietary software (*VMWare)*
  - Open-source software (*OpenStack, CloudStack)*

Network Operator's Cloud: Network Function Virtualization

- Designed to be "Telecom-grade"
- Aims at
  - Operational Improvements
  - Cost reduction
  - Streamlining high-touch processes
  - Reduction of the development time
  - Reduction of replacement costs
  - Reduction of the equipment costs
- Mandates replacing physical deployment with the virtual network functions deployed *dynamically, on-demand* across the network on *Common Off-The-Shelf (COTS)* hardware

Basic Building Blocks of the modern Cloud



Virtual Machines

Data Networking (over the Internet): Addressing, VPN, Network Management, Content Delivery

Application Programmer's Interfaces (API) at different levels

Distributed Computing: Load balancing, scheduling, resource control, distributed databases, data centers