CLOUD COMPUTING

Chapter 9: Parallel and Distributed Software Systems

GHENT
UNIVERSITY

CLOUD COMPUTING

CLOUD COMPUTING

CLOUD COMPUTING

CLOUD COMPUTING

OUTLINE

1. Definition
   - Characteristics
   - Service models
   - Deployment models
   - Payment models
   - Advantages
   - Obstacles
2. Cloud Platforms
   - Amazon Web Services (AWS)
   - Microsoft Windows Azure
   - Google App Engine
3. Building blocks of an IaaS Cloud
   - Provisioning resources
   - Virtualization
   - Containers
   - Container management

GHENT
UNIVERSITY

DEFINITION

- A lot of different definitions are used !

- Some definitions are:
    - Allows users to request computing resources through web interfaces
    - Extreme automation through virtualization
        - Hide away physical resource details
        - Hide away tedious and error prone configuration issues from users
        - Provides an utility view to computing resources
    - Pay as you go
        - Resources on-demand
        - Services on-demand

GHENT
UNIVERSITY

DEFINITION

Cloud Provider	→ Utility Computing →	Application Provider	→ Web Applications →	End User

NIST definition deals with three attributes:
– Essential characteristics
– Service models
– Deployment models

NIST = National Institute of Standards and Technology

CHARACTERISTICS (1)

1. On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.

2. Broad network access: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).

3. Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
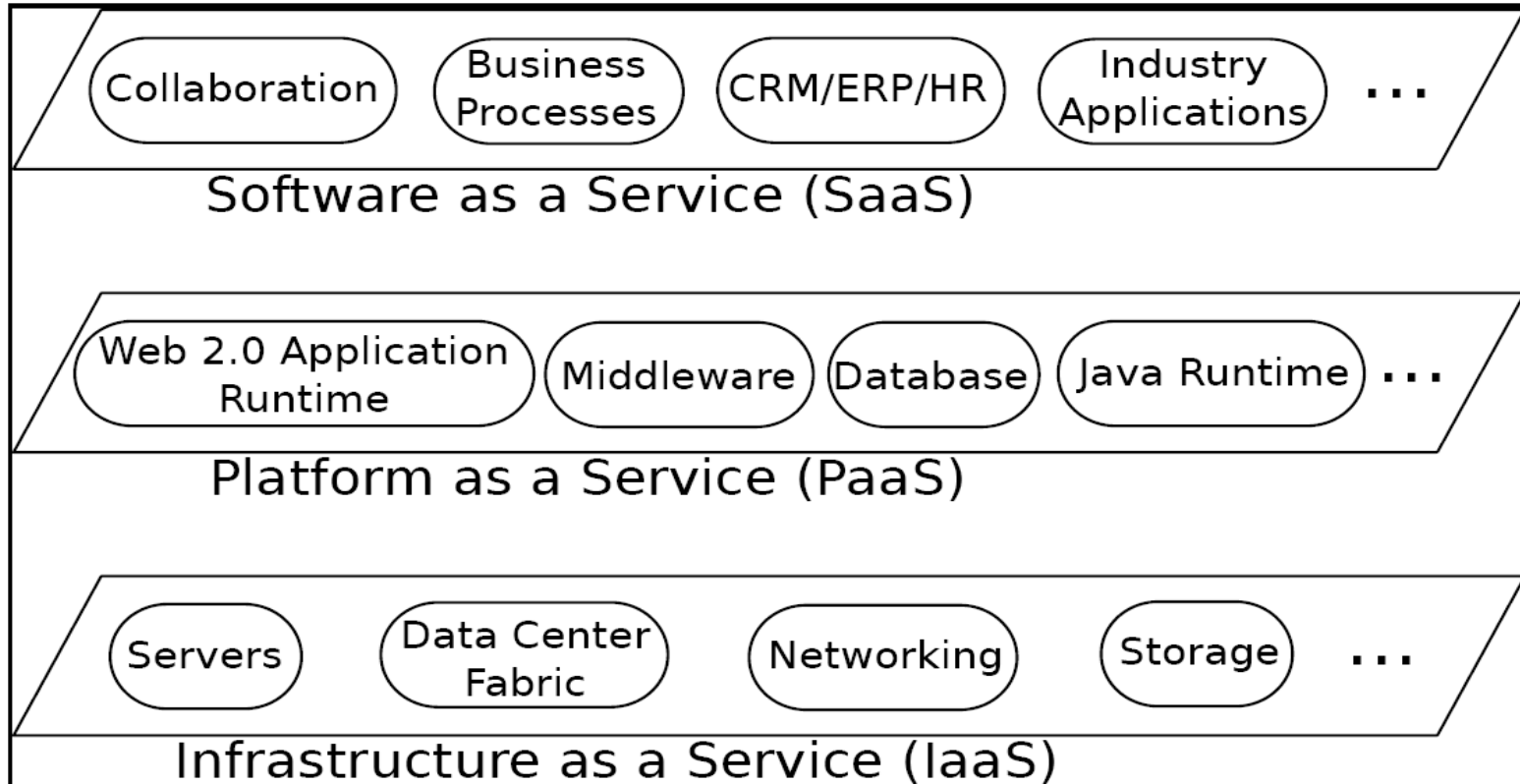
GHENT
UNIVERSITY

CHARACTERISTICS (2)

4. Rapid elasticity: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in.

5. Measured service: Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

GHENT
UNIVERSITY
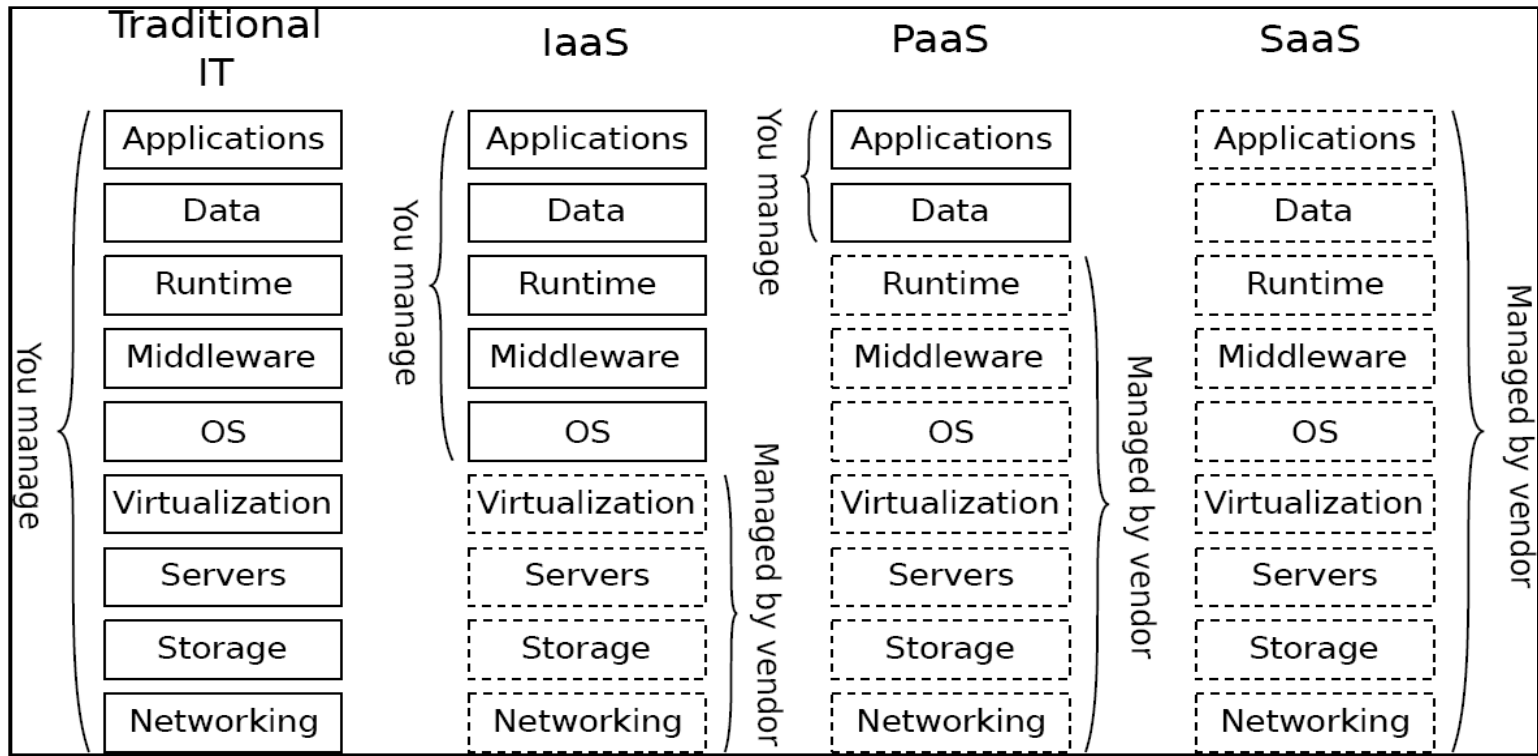
SERVICE MODELS

SERVICE MODELS

	Traditional IT	IaaS	PaaS	SaaS
	Applications	Applications	Applications	Applications
	Data	Data	Data	Data
	Runtime	Runtime	Runtime	Runtime
	Middleware	Middleware	Middleware	Middleware
	OS	OS	OS	OS
	Virtualization	Virtualization	Virtualization	Virtualization
	Servers	Servers	Servers	Servers
	Storage	Storage	Storage	Storage
	Networking	Networking	Networking	Networking

Traditional IT: You manage (all)

IaaS: You manage — Applications, Data, Runtime, Middleware, OS; Managed by vendor — Virtualization, Servers, Storage, Networking

PaaS: You manage — Applications, Data; Managed by vendor — Runtime, Middleware, OS, Virtualization, Servers, Storage, Networking

SaaS: Managed by vendor (all)
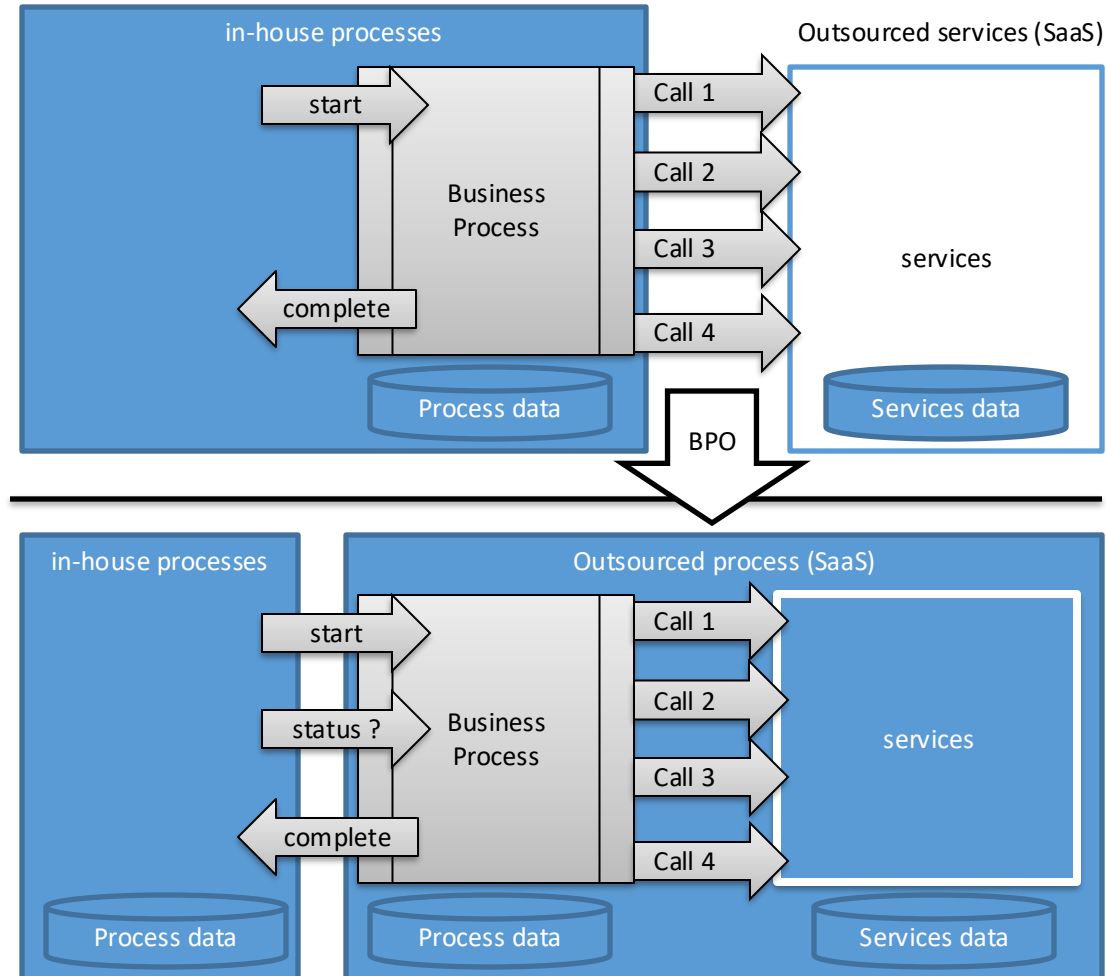
GHENT UNIVERSITY

SERVICE MODELS

- Cloud Software as a Service (SaaS)
  - Hotmail, Gmail, Google Docs, SalesForce
- Cloud Platform as a Service (PaaS)
  - Windows Azure, Google App Engine and force.com
- Cloud Infrastructure as a Service (IaaS)
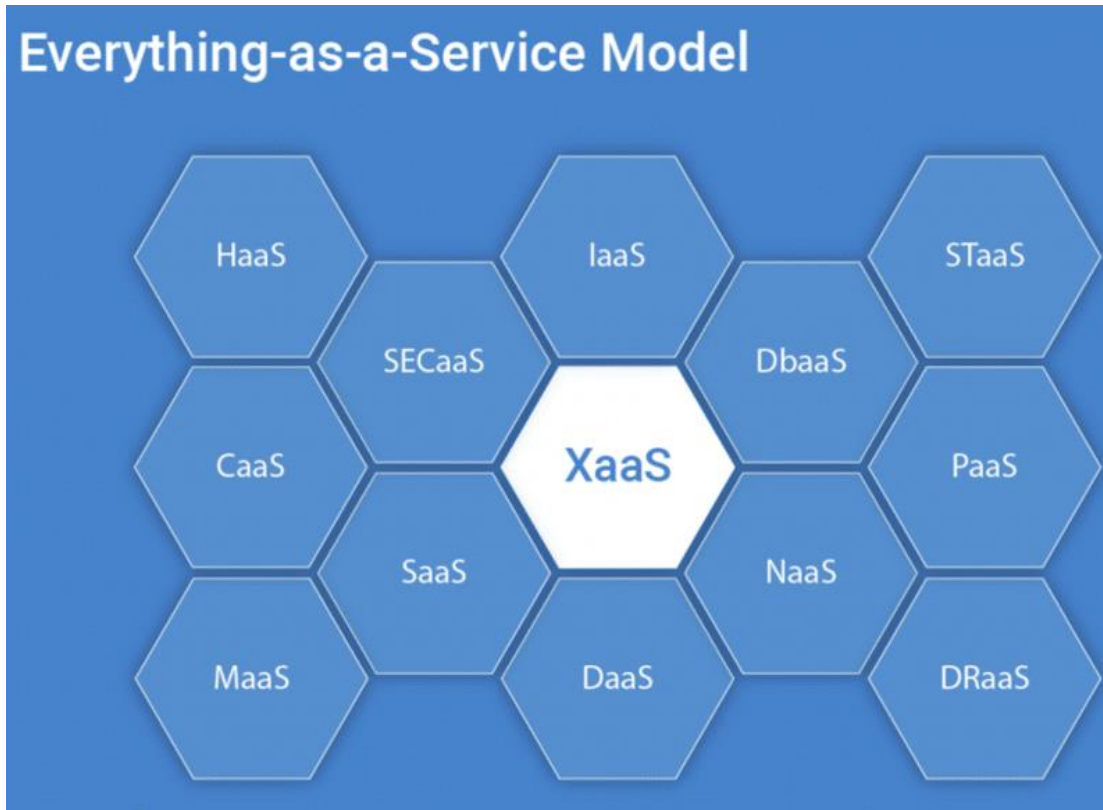 - Amazon, Rackspace

GHENT
UNIVERSITY

SERVICE MODELS

- BPaaS:

SERVICE MODELS

Everything-as-a-Service Model

DEPLOYMENT MODELS

1. Private cloud: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

2. Community cloud: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

3. Public cloud: The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

4. Hybrid cloud: The cloud infrastructure is a composition of two or more clouds (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).

GHENT
UNIVERSITY

DEPLOYMENT MODELS

1. Private cloud:

> datacenter of an organization

2. Community cloud:

> group of hospitals build community cloud for their requirements

3. Public cloud:

> Amazon, Google, Microsoft, etc.

4. Hybrid cloud:

> combination of the above, specific Azure focus

GHENT
UNIVERSITY

PAYMENT MODELS

1. Per-Instance billing: A common approach is to pay for every hour a VM or instance is used. This implies the instances need to be paid for, even if they are idle.

2. Reserved usage: In some cases, clients know they will be needing resources for longer periods, such as months or years. They can then make an up-front payment and reserve the instances for this time period, during which the instance will always be available, with lower hourly rates.

3. Bidding: In this approach, a maximum instance price is provided by the customer. The price of instances varies while based on the load of the cloud, so instances are more expensive when the utilization degree of the cloud is higher.

4. Actual usage: This approach is used in some PaaS clouds, and determines the cost based on the actual CPU cyles used by the applications. Only resources that are actually used need to be paid for in this model.

GHENT
UNIVERSITY

PAYMENT MODELS

https://aws.amazon.com/ec2/pricing/

Pricing Details

For region

US East (Virginia)

Hourly Fees

Total hourly fees will vary by instance type and EC2 region.

Software Pricing: Hourly ⬤ Annual

EC2 Infrastructure		Software	
Instance Type	Usage	Price	Savings
m1.medium	$0.087/hr	$3,750/yr	11%
m1.large	$0.175/hr	$8,500/yr	10%
m1.xlarge	$0.35/hr	$17,000/yr	10%
m2.xlarge	$0.245/hr	$14,000/yr	10%
m2.2xlarge	$0.49/hr	$29,000/yr	10%
m2.4xlarge	$0.98/hr	$58,000/yr	10%
c1.xlarge	$0.52/hr	$21,000/yr	10%
m3.medium	$0.07/hr	$3,750/yr	11%
m3.large	$0.14/hr	$7,800/yr	10%
m3.xlarge	$0.28/hr	$16,000/yr	10%
m3.2xlarge	$0.56/hr	$32,000/yr	10%

EBS Storage Fees ❓

$0.05 / GB / Month for EBS Magnetic Storage

Assumes On-Demand EC2 pricing
Learn about instance types 🔗

Data Transfer Fees not included
Learn more about Data Transfer Fees 🔗

Region: Asia Pacific (Mumbai)

|	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage
General Purpose - Current Generation					
t2.nano	1	Variable	0.5 GiB	EBS Only	$0.0062 per Hour
t2.micro	1	Variable	1 GiB	EBS Only	$0.0124 per Hour
t2.small	1	Variable	2 GiB	EBS Only	$0.0248 per Hour
t2.medium	2	Variable	4 GiB	EBS Only	$0.0496 per Hour
t2.large	2	Variable	8 GiB	EBS Only	$0.0992 per Hour
t2.xlarge	4	Variable	16 GiB	EBS Only	$0.1984 per Hour
t2.2xlarge	8	Variable	32 GiB	EBS Only	$0.3968 per Hour
m5.large	2	8	8 GiB	EBS Only	$0.101 per Hour
m5.xlarge	4	16	16 GiB	EBS Only	$0.202 per Hour
m5.2xlarge	8	31	32 GiB	EBS Only	$0.404 per Hour

**GHENT
UNIVERSITY**

ADVANTAGES

1. User does not need to own and configure machines
– Management of infrastructure left to cloud providers
– User needs to only worry about what to do with the machine/resource – not how to prepare it for that purpose
2. Request resources when needed
3. Simple web based interface for
- request resource
- monitor and manage resource
4. Extreme scaling
- Can scale the footprint from 1 server to 1000+ servers in a matter of few minutes or less
5. Economic model
- Rent vs lease
- Management cost is often higher than resource cost

GHENT
UNIVERSITY

ADVANTAGES

Three useful usage scenarios:

– Load varying with time

– Demand unknown in advance

– Batch analytics that can benefit from huge number of

resources for a short time duration
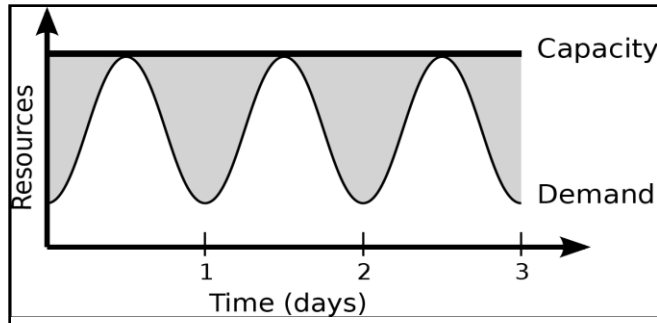
overprovisioning
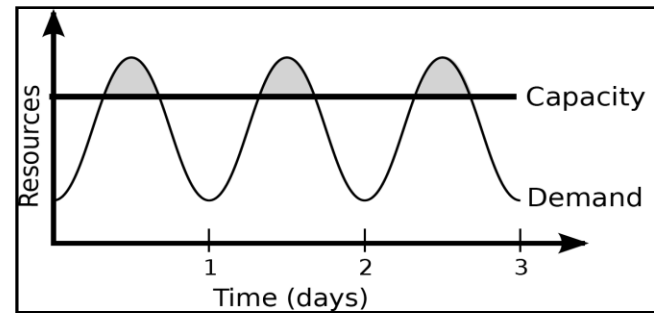
underprovisioning

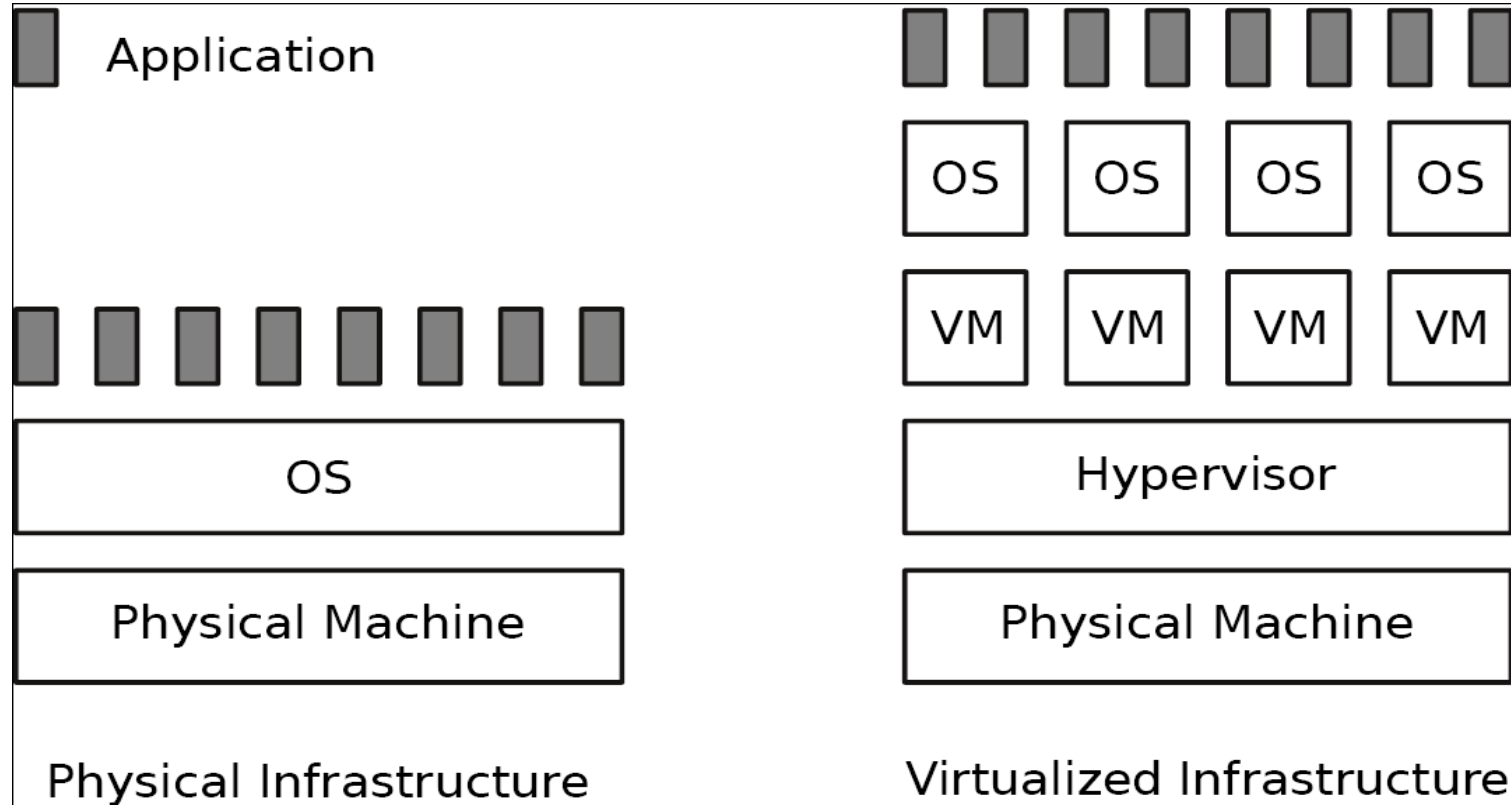# BUILDING BLOCKS OF AN IAAS CLOUD

PROVISIONING RESOURCES
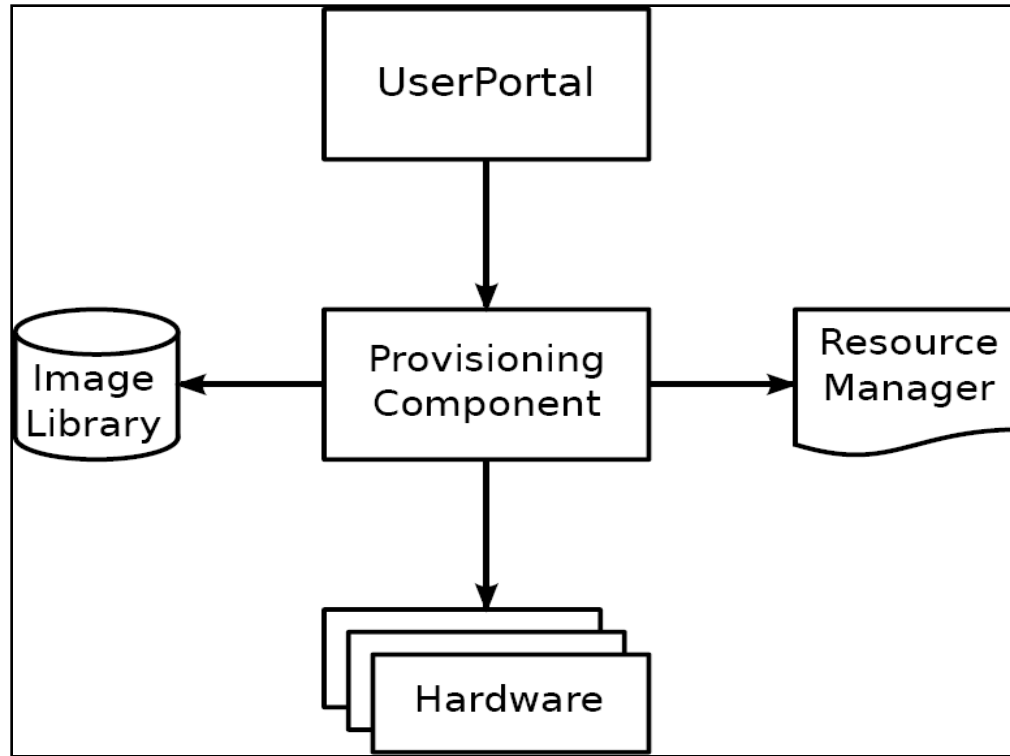
Current research challenges are the efficient packing of VMs and lowering storage requirements. Research into efficient allocation strategies, lowering the amount of hardware needed and thus the power consumption of clouds is also conducted.
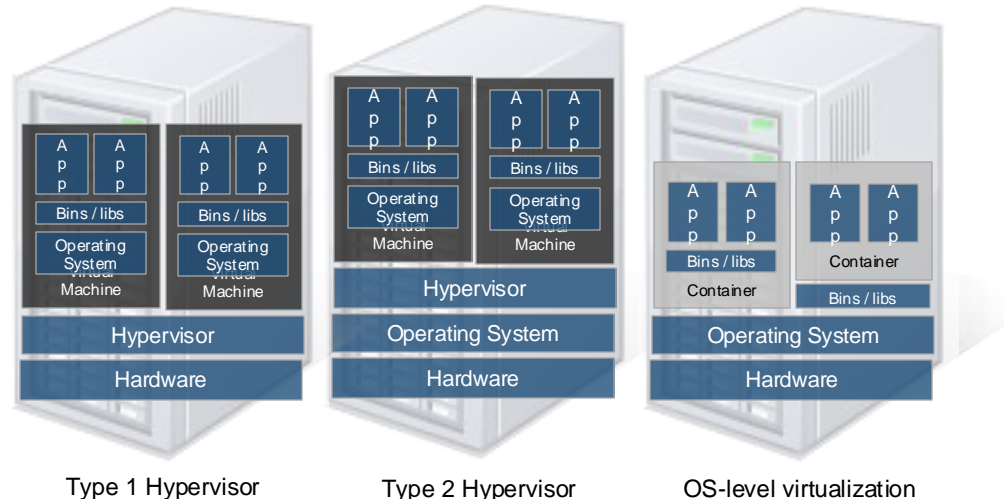
VIRTUALIZATION

CONTAINERS

VMs

- Virtualization of hardware
- Flexible, robust and safe, but fairly big performance hit

Containers

- Lightweight
- Better use of resources (sharing host OS and potentially binaries/ libraries)


Type 1 Hypervisor   Type 2 Hypervisor   OS-level virtualization

LIVE MIGRATION

— Needed for: maintenance, vertical scalability, energy efficiency.

— Live migration within the Local Area Network (LAN) of a single cloud is well supported by most commercial cloud platforms.

— Storage does not have to be migrated as this is typically located on a shared Storage Area Network (SAN) volume, accessible on the local network, as shown below:
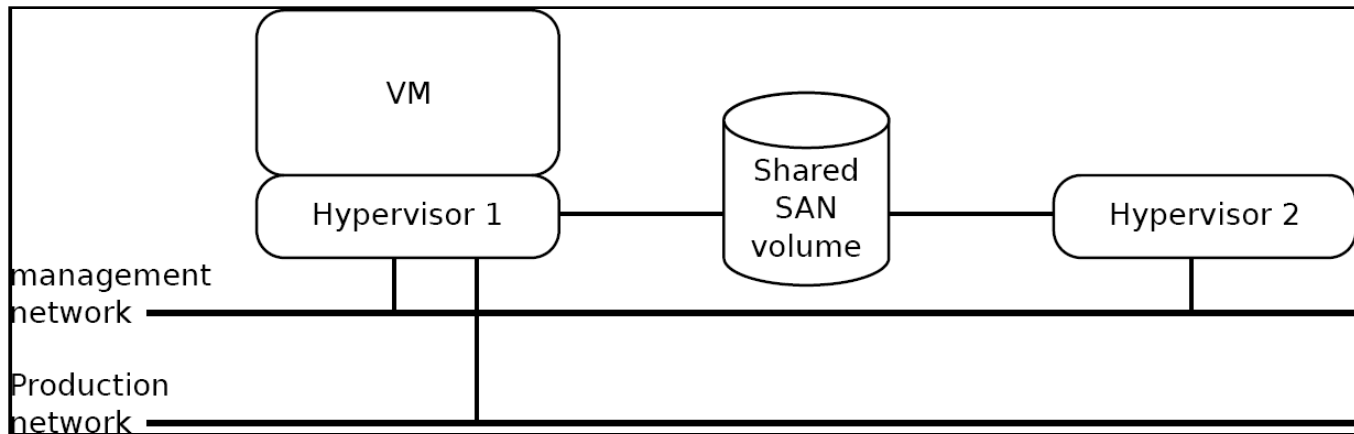
LIVE MIGRATION – NETWORK CONNECTIVITY (1)

– Option 1 : retain the old IP address
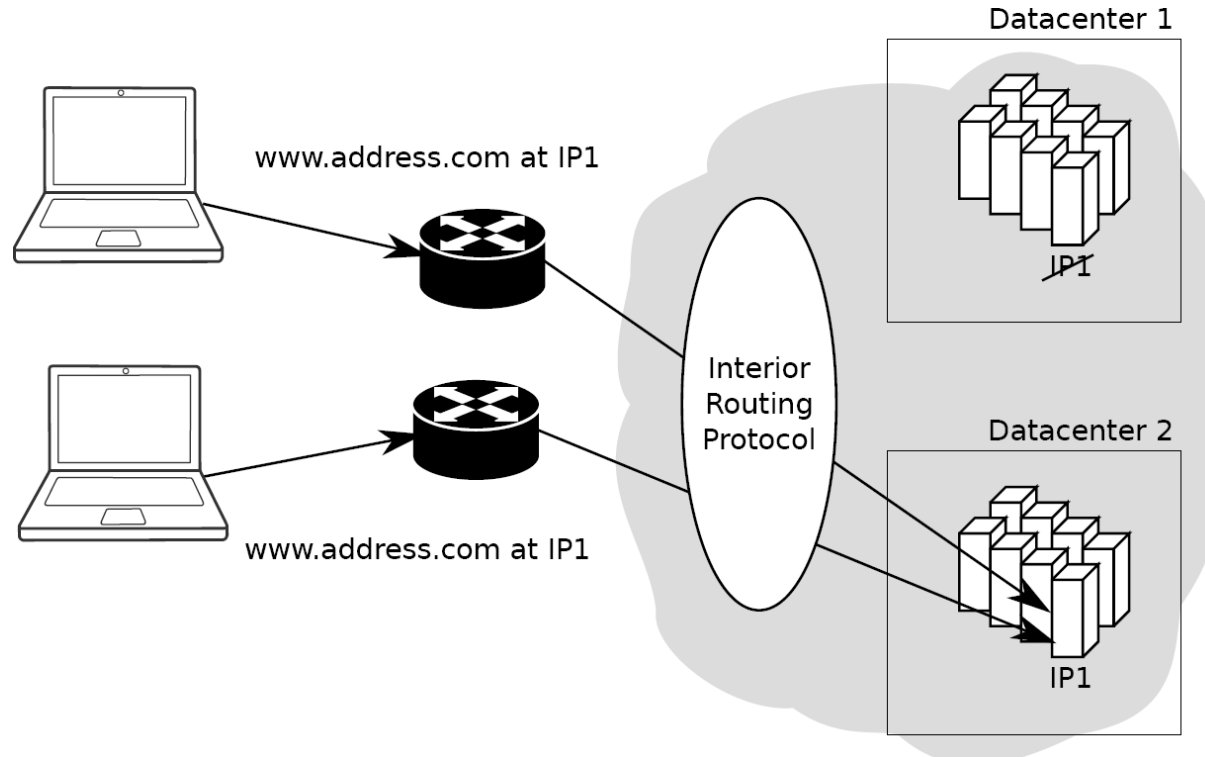
LIVE MIGRATION – NETWORK CONNECTIVITY (2)

– Option 2 : updating the IP address
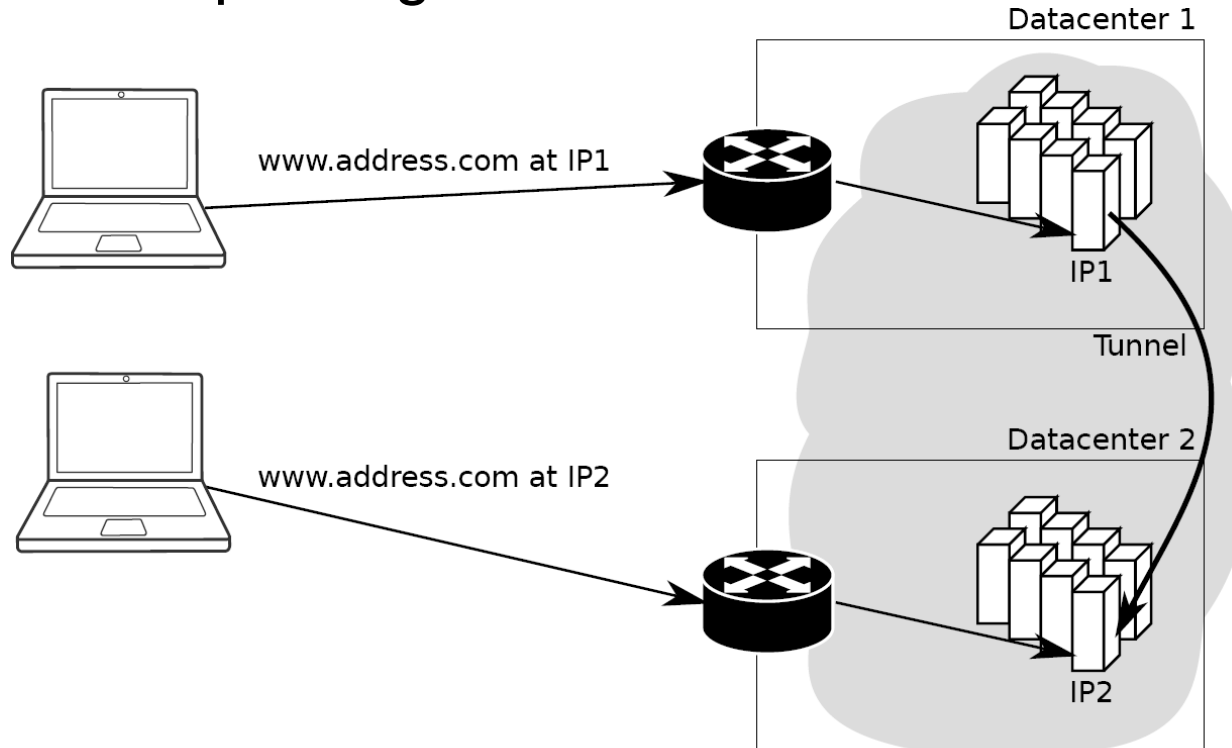
CONTAINER MANAGEMENT

read it

— Cloud, Edge and Fog Computing

— Resource Management

— Resource Elasticity

— Application Demand Profiling

— Monitoring, infrastructure demand profiling and resource utilization estimation

— Resource Pricing

— Dynamic resource allocation for containerized applications

— Dynamic Pricing

— Cloud management systems for bare-metal containers

— Management of a hybrid edge/fog/cloud environment

— Experimental validation of resource management strategies

— Towards serverless cloud computing

explain them using

See survey paper on Ufora:
"Resource Management in a Containerized Cloud: Status and Challenges"

GHENT UNIVERSITY

34