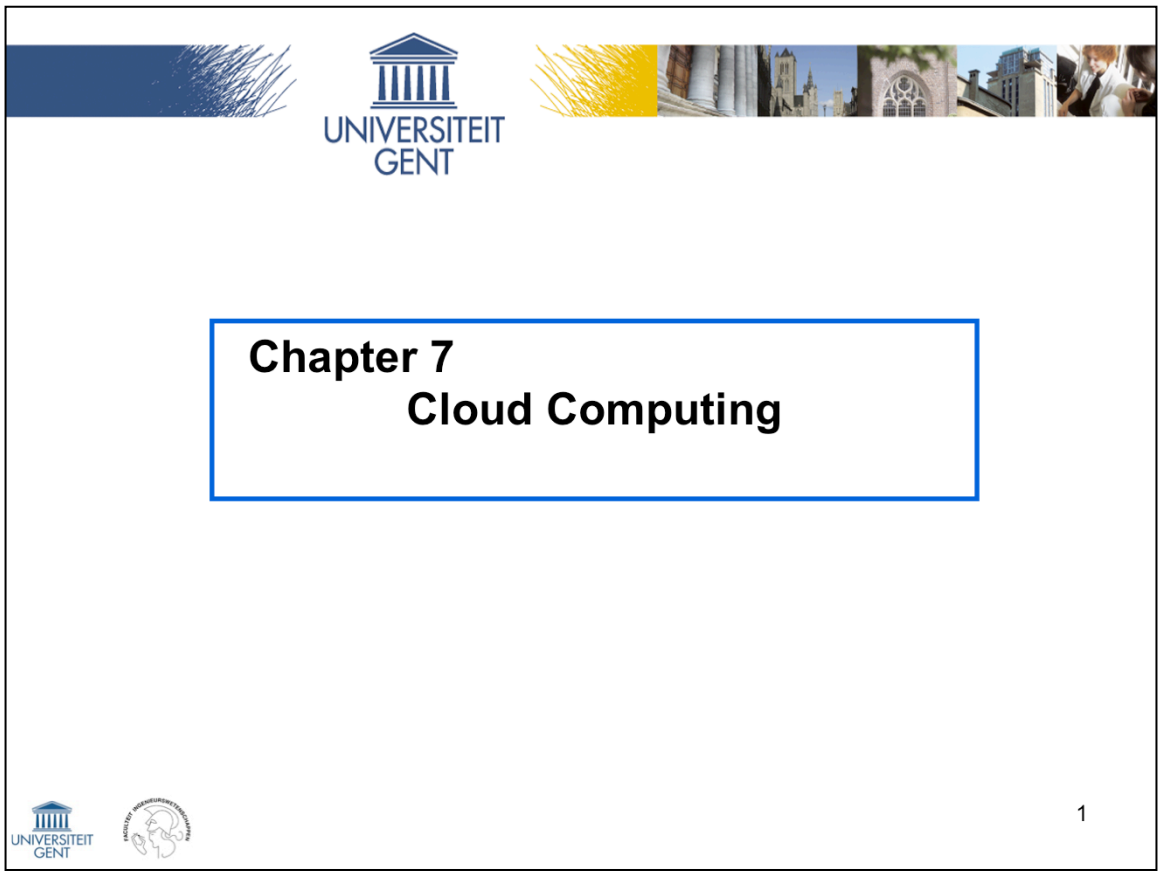

UNIVERSITEIT
GENT
Chapter 7
Cloud Computing
UNIVERSITEIT
GENT
1

Cloud Computing

Cloud Computing

Cloud Computing

Cloud Computing

UNIVERSITEIT GENT

5

Cloud Computing: conferences

CALL FOR PAPERS

Service Management

- Multimedia service management
- Managed service provisioning
- OTT service management
- Data service management
- IT service management
- Hosting
- Data center management
- Cloud computing
- Virtualized infrastructure management
- Infrastructure as a Service
- Platform as a Service
- Software as a Service
- Management as a Service

UNIVERSITEIT GENT

6

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
   - Virtual Images
   - Virtual Applicances
4. Enterprise applications
   - Cloud migration and planning
   - Live Migration


7

After an overview of the basic concepts (programming language generations and compilation versus interpretation), the following programming paradigms are detailed: procedural, functional, object oriented and declarative. Moreover, scripting languages will be explained. The next three sections focus on an overview of Java, C/C++ and C# and their important concepts.

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

Definition

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).

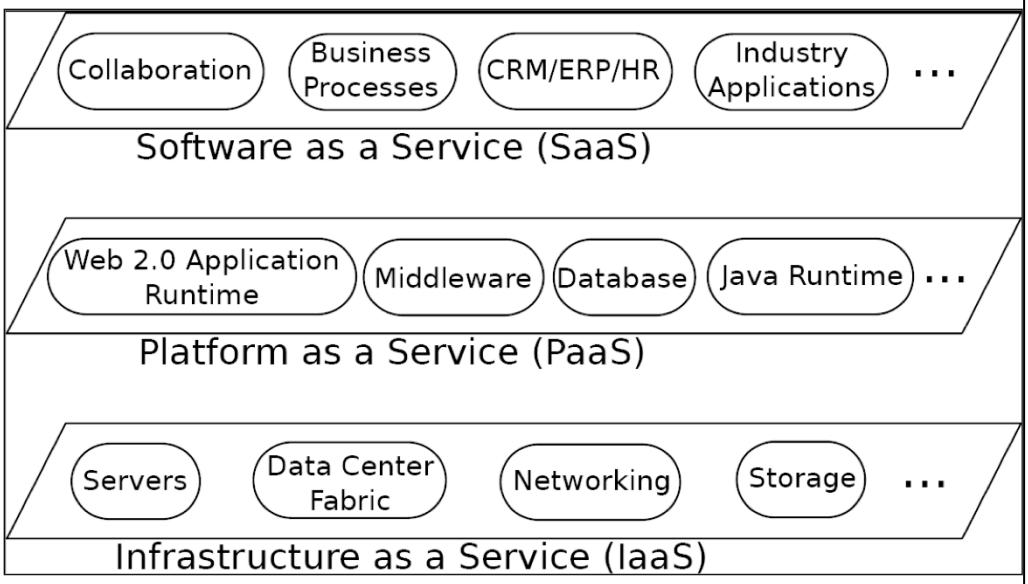
3. Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

Characteristics (2)

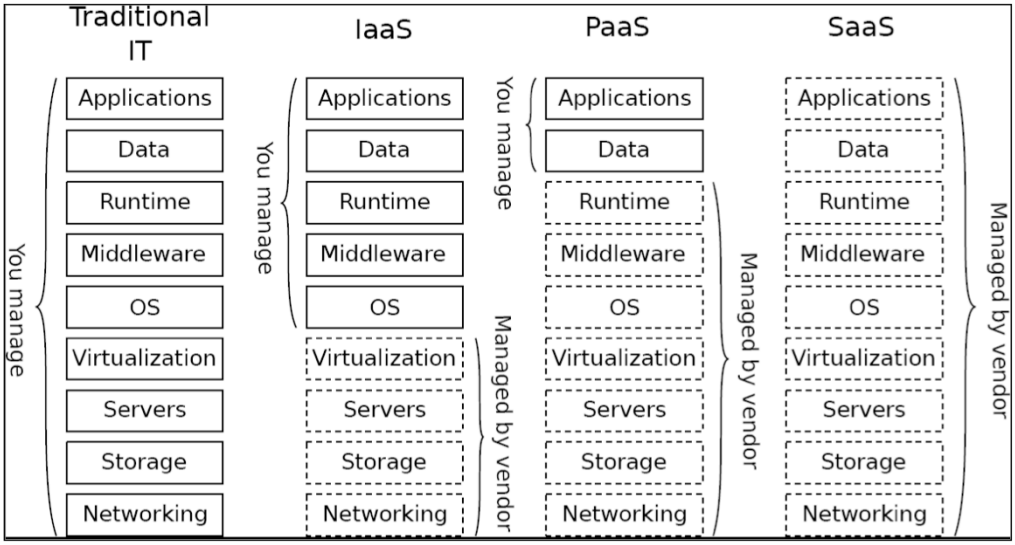
4. Rapid elasticity: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in.

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

Service models



Service models

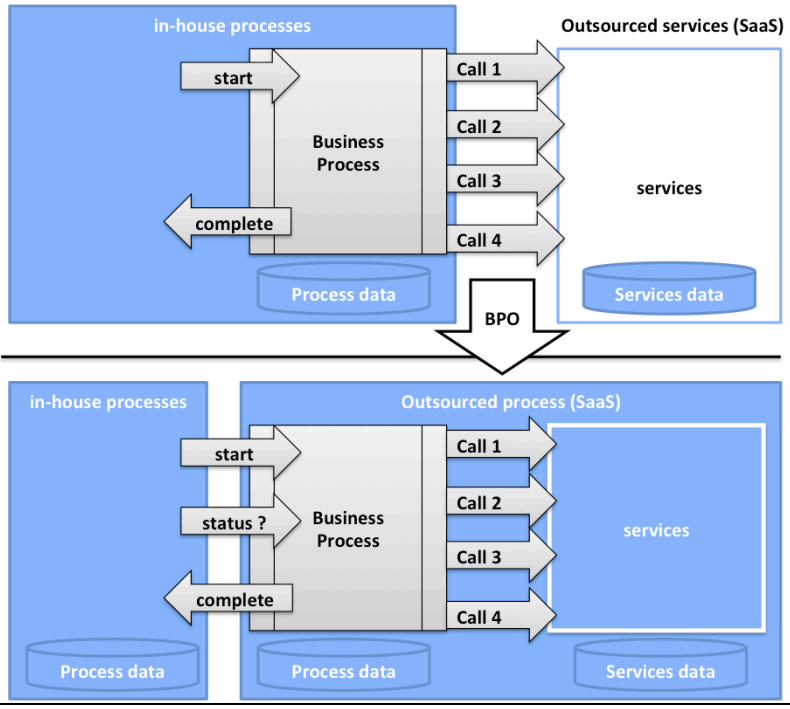


Service models

- 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

Service models

- BPaaS:



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).

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

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.

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

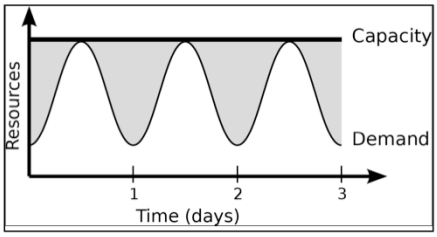
Advantages

Three useful usage scenarios:

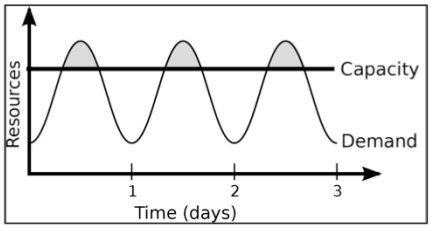
– Load varying with time

– Demand unknown in advance

– Batch analytics that can benefit from huge number of resources for a short time duration



overprovisioning



underprovisioning

Obstacles

Table 2. Top 10 obstacles to and opportunities for growth of cloud computing.

	Obstacle	Opportunity
1	Availability/Business Continuity	Use Multiple Cloud Providers
2	Data Lock-In	Standardize APIs; Compatible SW to enable Surge or Hybird Cloud Computing
3	Data Confidentiality and Auditability	Deploy Encryption, VLANs, Firewalls
4	Data Transfer Bottlenecks	FedExing Disks; Higher BW Switches
5	Performance Unpredictability	Improved VM Support; Flash Memory; Gang Schedule VMs
6	Scalable Storage	Invent Scalable Store
7	Bugs in Large Distributed Systems	Invent Debugger that relies on Distributed VMs
8	Scaling Quickly	Invent Auto-Scaler that relies on ML; Snapshots for Conservation
9	Reputation Fate Sharing	Offer reputation-guarding services like those for email
10	Software Licensing	Pay-for-use licenses

Source: Above the Clouds: A Berkeley View of Cloud Computing

Amazon Web Services (AWS)

IaaS Cloud Example: Amazon EC2

– Amazon EC2 provides public IaaS Cloud

– User uses a portal to request a machine with specific resource

  – CPU, memory, disk space

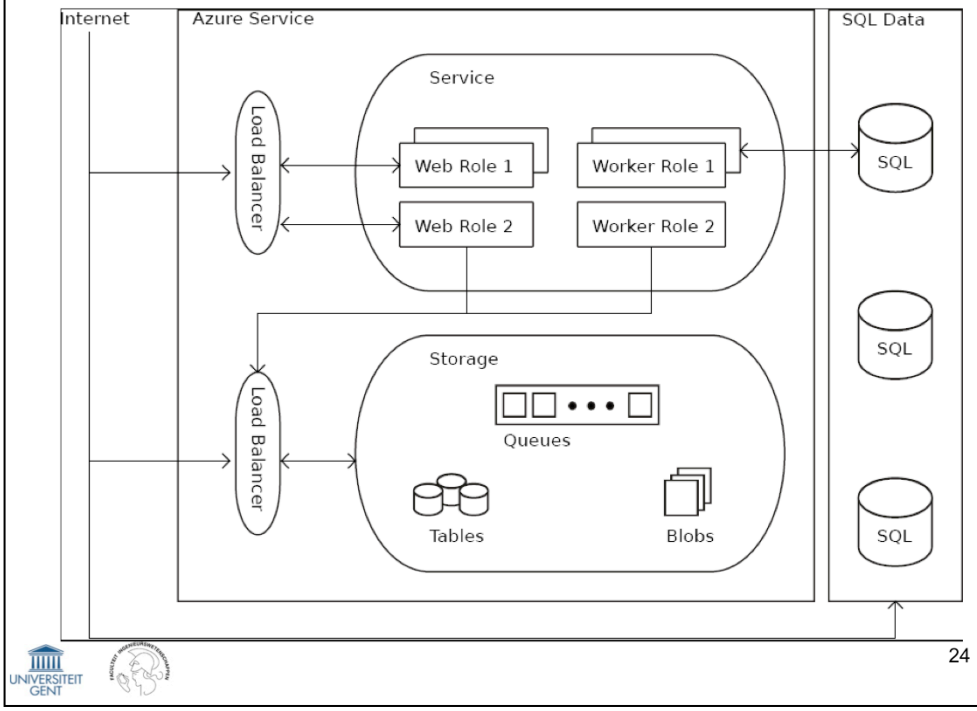
  – Pre-built OS and possibly middleware

– No default support for automatic scaling, failover: extra services are needed

Microsoft Windows Azure

– Applications need to be written using .NET libraries
– More flexible than Google AppEngine
– Able to provide automated scaling
– Between Application framework and hardware virtual machines

Microsoft Windows Azure



Google App Engine

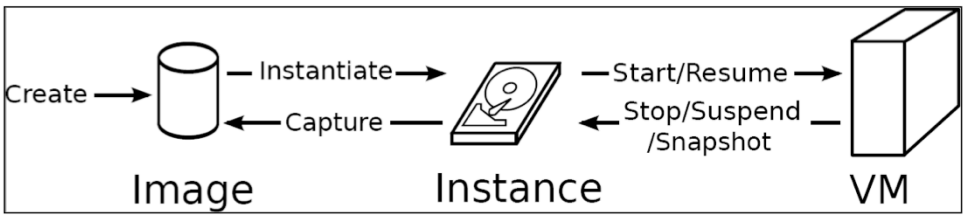
- PaaS model
- Provides a platform to host web applications
- App Engine SDK for programming (Python and Java support)
- A set of primitives (datastore, URL fetch, memcache, JavaMail, Images, authentication..)
- User focuses on developing the application in this framework
- Once deployed, scaling, availability etc. are handled by Google AppEnginePlatform

Building blocks of an IaaS Cloud

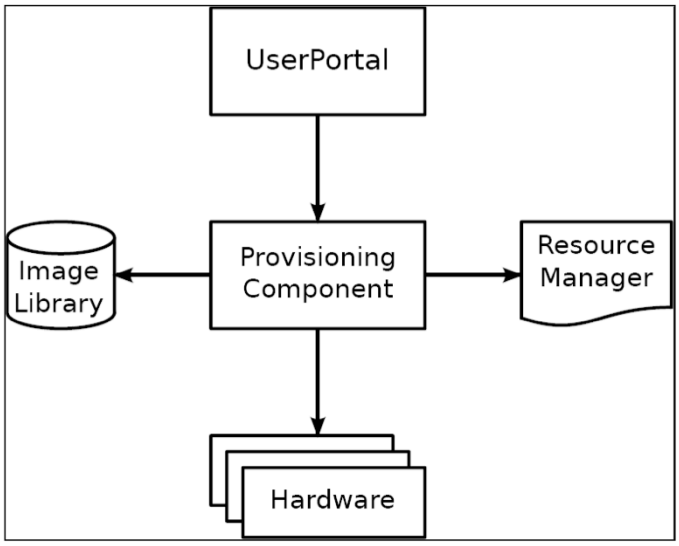
Point of view of cloud provider:

- Provisioning resources
- Virtualization
- Virtual Images
- Virtual Applicances

Provisioning resources

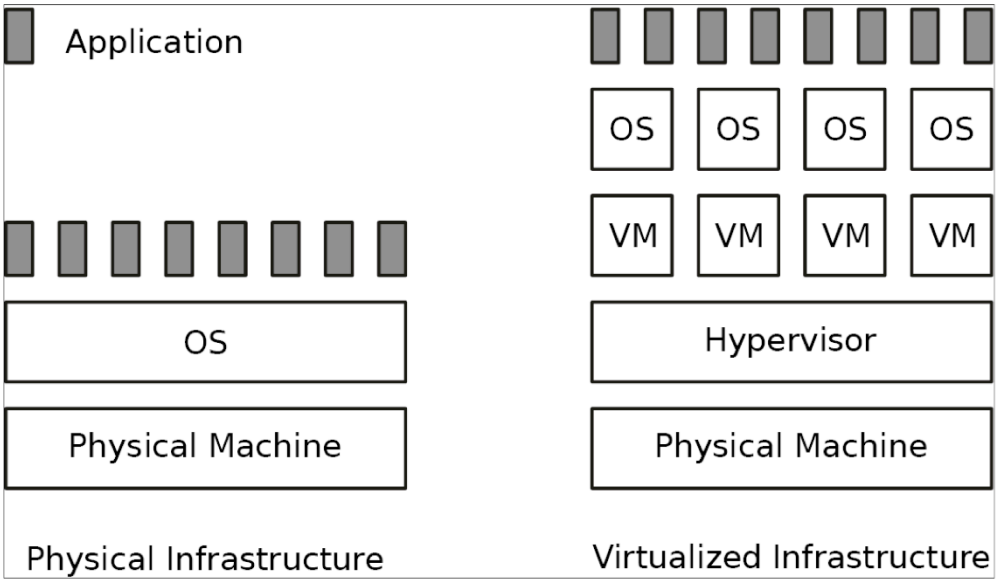


Provisioning resources



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



Virtual Images

To provide a pre-built software stack for a virtual machine, a virtual image is used. At VM creation time, it boots from this stack.

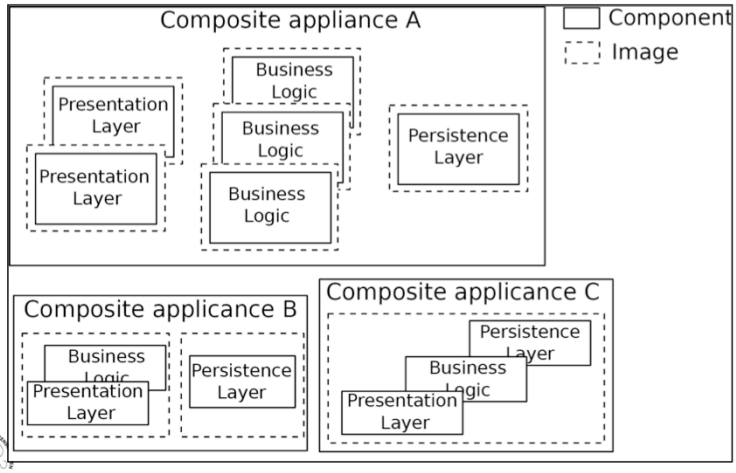
Table 10.1: The contents of a VMware image

Extension	Description
.vmdk	Stores the contents of the VM's hard disk drive
.nvram	Stores the state of the VM's BIOS
.vmx	Stores the setting of the virtual machine
.vmxf	Stores supplemental configuration information
.vswp	Swap file

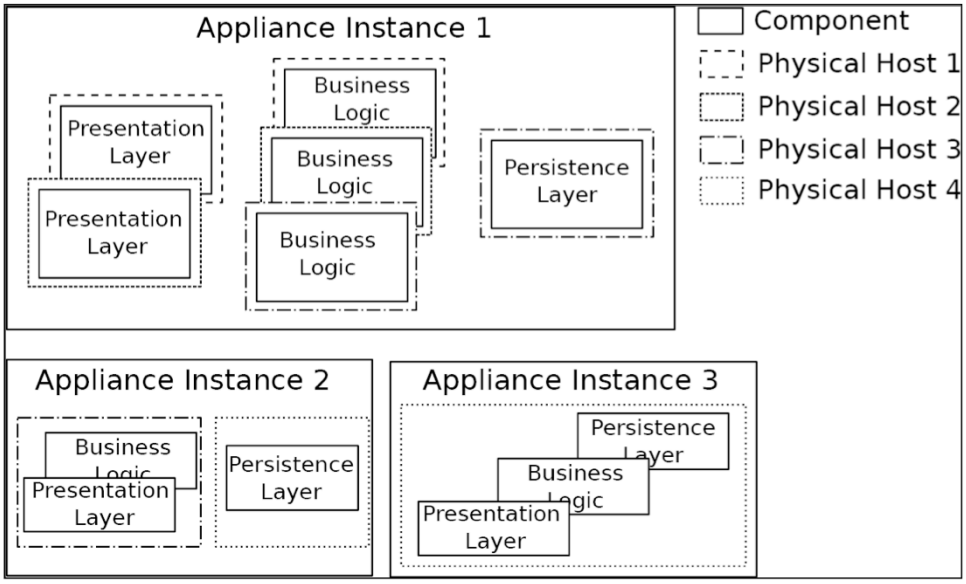
Active research areas concerning these repositories are the de-duplication of files over multiple images, ensuring files occurring in two different images are only stored once, further decreasing the required storage, and caching of images to improve VM construction times.

Virtual Applicances - definition

A virtual appliance is an image of a pre-built software stack that can be instantiated.
This software stack can be only an operating system, an operating system and application middleware such as a webserver, or a full software stack with a running application.
One of the largest advantages of these virtual appliances is that a large part of the configuration complexity is already taken care of.



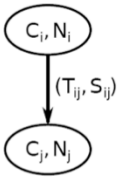
Virtual Applicances - instantiation



Enterprise applications on the Cloud

1. how applications can be migrated to cloud environments
   - planning required
2. how applications can be migrated within a cloud and between clouds
   - live migration solutions

Cloud migration and planning



Input:

N_i = number of servers in component C_i

T_{ij} = number of messages per second along (i, j)

S_{ij} = average size of messages along (i, j)

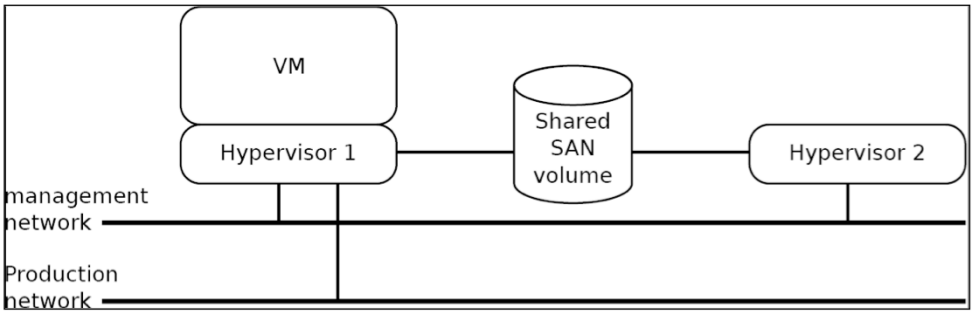
To determine:

m_i = number of servers of the component C_i to migrate to the cloud ($m_i \leq N_i$)

Live Migration

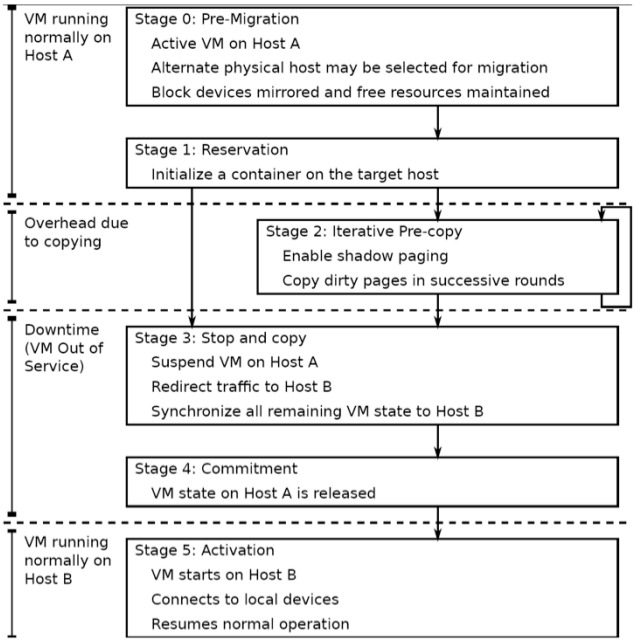
Needed for: maintenance, vertical scalability, energy efficiency.

- 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

Live migration of memory within data center:



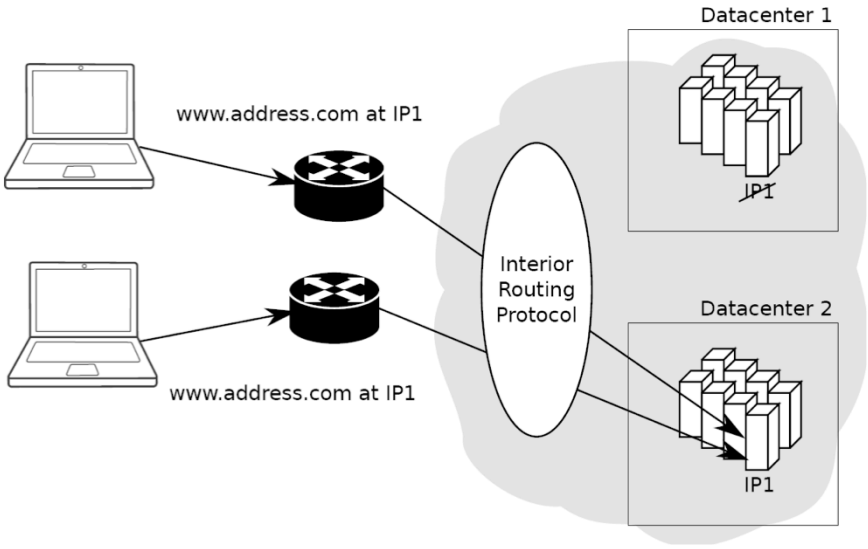
Live Migration – copy models

Different tactics for the migration of VMs within a datacenter:

<table>
<tr><td rowspan="2">Pre-copy</td><td colspan="2">Image File Transfer</td><td>Memory Migration</td></tr>
<tr><td colspan="3">Intercept, record and transfer written blocks</td></tr>
<tr><td rowspan="2">Post-copy</td><td rowspan="2">Memory Migration</td><td colspan="2">Background Image Copy</td></tr>
<tr><td colspan="2">On-demand Fetching</td></tr>
<tr><td rowspan="2">Pre+Post-copy</td><td rowspan="2">Image File Transfer</td><td rowspan="2">Memory Migration</td><td>Background dirty block copy</td></tr>
<tr><td>On-demand Fetching</td></tr>
</table>

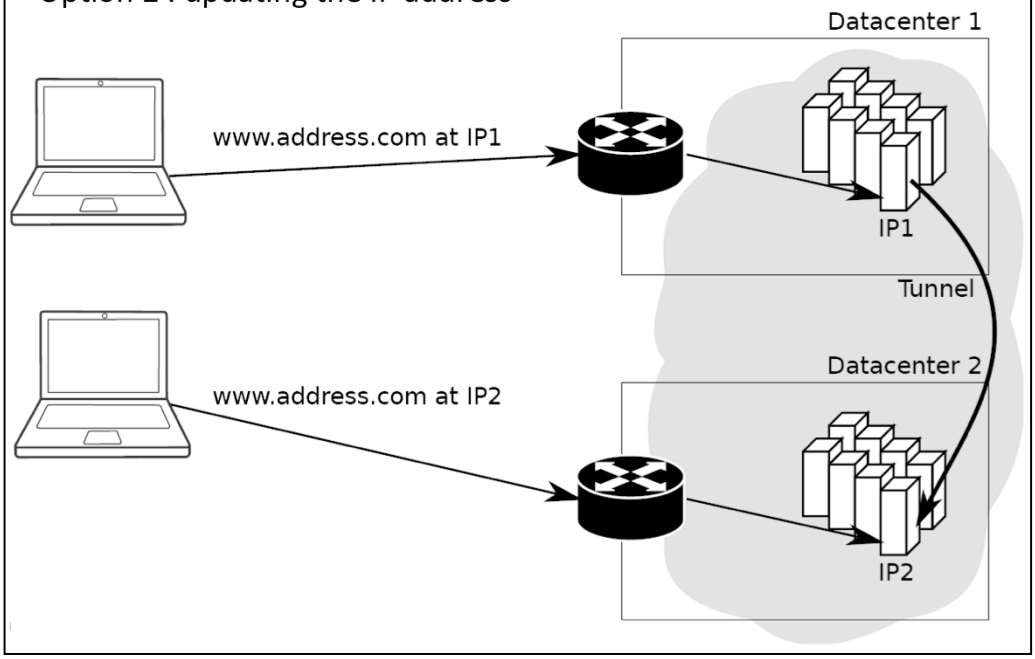
Live Migration – network connectivity (1)

Option 1 : retain the old IP address

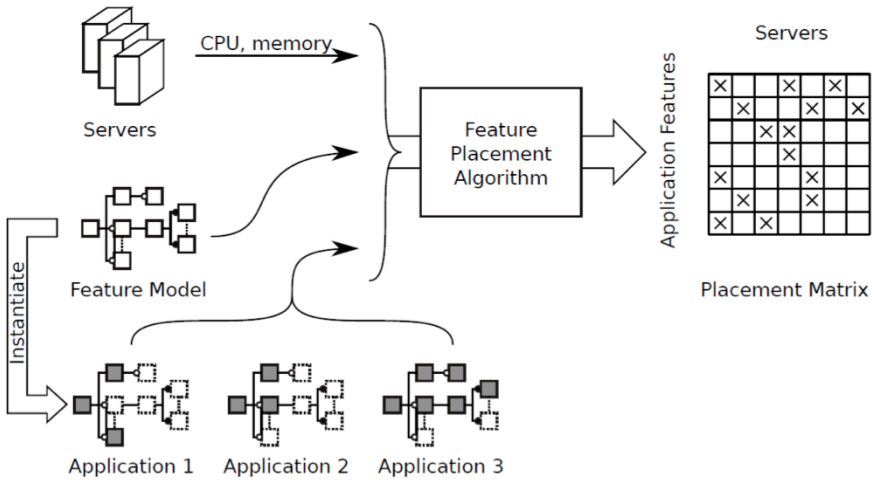


Live Migration – network connectivity (2)

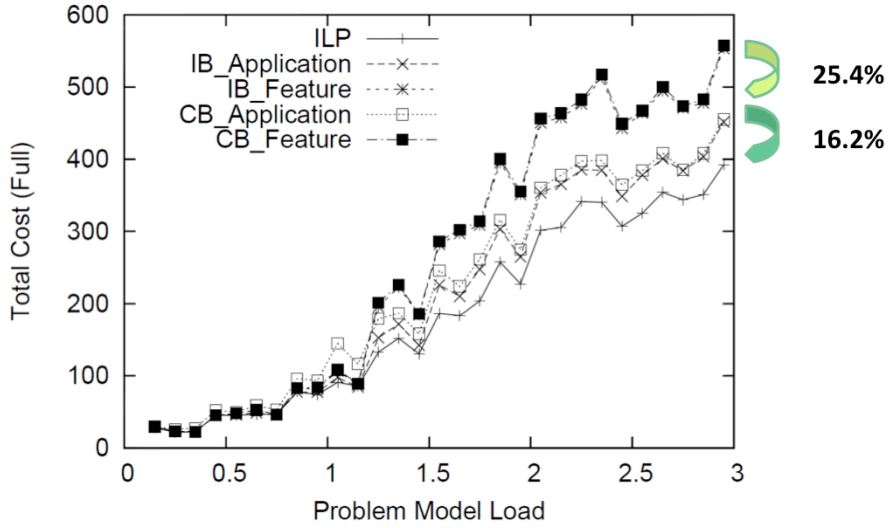
Option 2 : updating the IP address



Advanced SaaS Clouds



Feature placement of customizable multi-tenant applications



Cloud Computing status

1. Real-time analysis of management data in clouds
2. Designing and deploying resource management systems for large-scale interclouds
3. Scalability of cloud resource management systems
4. Dynamic pricing for cloud resources

Cloud Computing status

5. Resource allocation in clouds based on dynamic application performance metrics
6. Edge-computing – fog computing
7. Security management in SaaS applications
8. Network Function Virtualization (NFV)