Resource Management for Cloud Computing

Supervisor : Dr. Pham Tran Vu

Presenters : Nguyen Viet Hung - 11070451

Tran Le Vinh - 11070487

Date : April 16, 2012

Contents

Introduction to Cloud Computing

Introduction to Cloud Computing

- Definitions
- History
- Service Layers
- Types
- Advantage
- Disadvantage

What is Cloud Computing ?

- **Well … It depends**
  - Each Cloud vendor seems to have their own definition
- **SYS-CON Media Inc [2].**
  - A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet.

What is Cloud Computing ?

- **NIST [6]**
  - A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.
    - *on-demand self-service*: a consumer can autonomously provision computing capabilities that is without requiring human interaction with the respective provider.
    - *rapid elasticity*: the above capabilities may be dynamically resized in order to quickly scale up (to potentially unlimited size) or down in according to the specific needs of the consumer.

What is Cloud Computing ?

- **CISCO [4].**
  - IT resources and services that are abstracted from the underlying infrastructure and provided *on-demand* and *at scale* in a *multitenant environment*.
    - *on-demand*: resources can be provisioned immediately when needed, released when no longer required, and billed only when used.
    - *at scale*: the service provides the illusion of infinite resource availability in order to meet whatever demands are made of it.
    - *multitenant environment*: the resources are provided to many consumers from a single implementation, saving the provider significant costs.

History of Cloud Computing

- **Related paradigms**
  - Grid and utility computing
  - Software as a Service (SaaS)
- **Earlier antecedents**
  - 1961, John McCarthy
    - Computation delivered as public utility
  - 1969, J.C.R. Licklider, ARPANET: Idea of an intergalactic computer network:
    - Access programs and data at any site, from anywhere

History of Cloud Computing

- **1999, Salesforce.com**
  - Delivering enterprise applications via a website
- **2002, Amazon web services**
  - Suite of cloud-based services including storage and computation
  - 2006, Amazon provided EC2 (Elastic Computing Cloud)

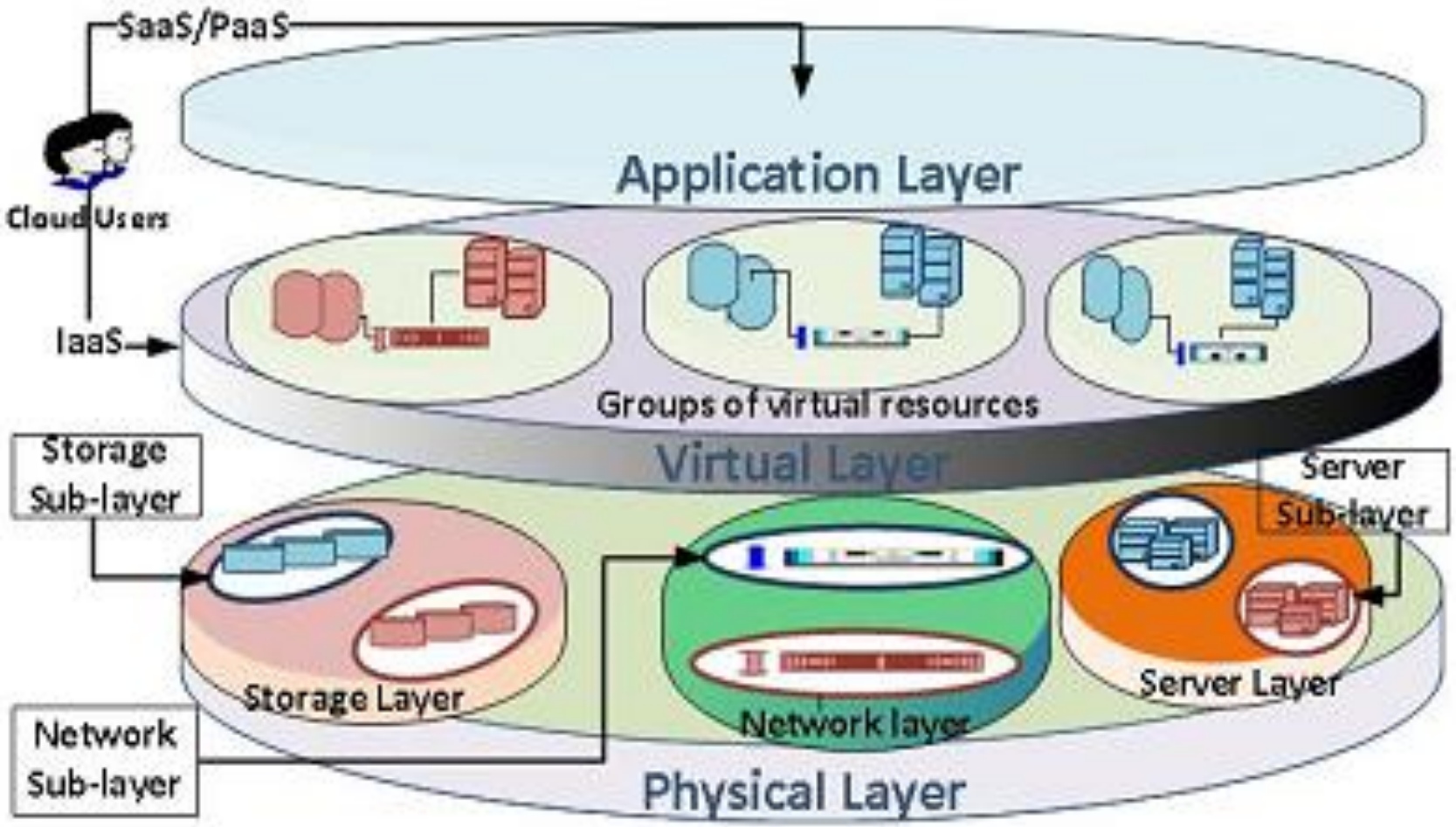
Service Layers of Cloud Computing

- **The four major layers in the cloud computing value chain**
  - Software as a Service (SaaS)
    - delivered over the network on a subscription and on-demand basis.
  - Platform as a Service (PaaS)
    - consists of run-time environments and software development frameworks and components delivered over the network on a pay-as-you-go basis.
    - presented as API to consumers
    - Ex: Google Apps Engine, Amazon Web Services

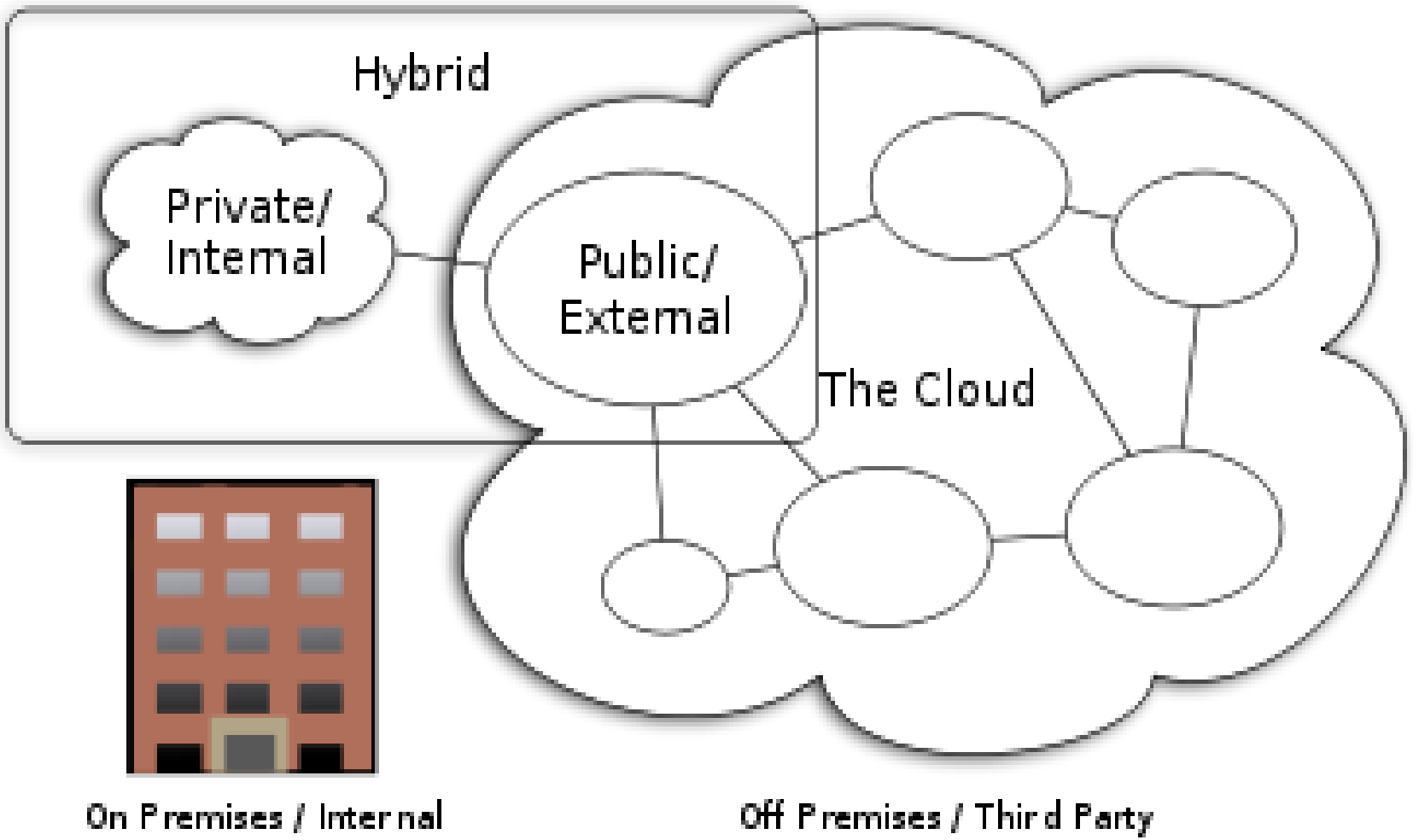
Service Layers of Cloud Computing

- **The four major layers in the cloud computing value chain**
  - Infrastructure as a Service (IaaS)
    - Compute, network, and storage are delivered over the network on a pay-as-you-go basis
  - IT foundation
    - The basis of the above value chain layers
    - Provide basic building blocks to architect and enable the above layers.

Cloud in 3D view



Types of Cloud Computing



Types of Cloud Computing

- **Public Cloud**
  - Computing infrastructure is hosted at the vendor's premises and shared between organizations.
  - The customer has no visibility over the location of the cloud computing infrastructure.
  - Example: Amazon, Google Apps, Windows Azure.

Types of Cloud Computing

- **Private Cloud**
  - Computing architecture is dedicated to the customer and is not shared with other organizations.
  - Expensive and more secure than public cloud.
  - Externally hosted ones as well as in premise hosted clouds.
  - Example: eBay

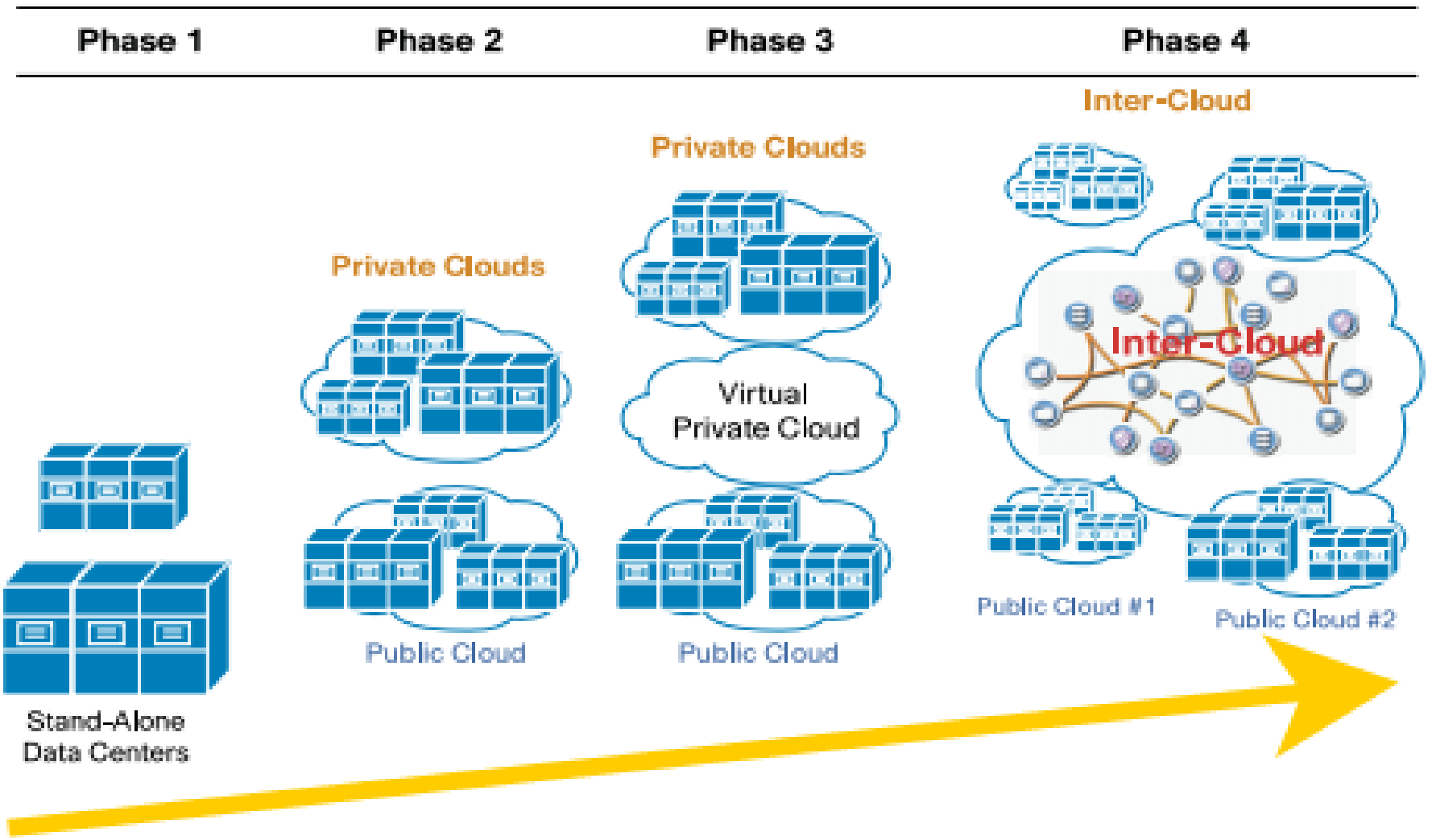
Types of Cloud Computing

- **Community Cloud**
  - Shared between the organizations of the same community with common concerns (security, compliance, jurisdiction).
  - Managed internally or by a third-party.
  - Hosted internally or externally.
  - The costs are spread over fewer users than a public cloud but more than a private cloud.

Types of Cloud Computing

- **Hybrid Cloud**
  - a composition of two or more clouds (private, community or public).
  - Hosted in the public cloud.
  - Secured Apps in the private cloud.

Phased Evolution of Cloud



Advantages of Cloud Computing

- **Cost Effective**
  - both the operating expenses and the capital expenses sides of the equation.
  - Only pay for what you use.
- **Flexible**
  - Derive from rapid provisioning of new capacity
  - Rapid relocation or migration of workload
- **Improved Automation**
  - Provisioned and de-provisioned in services.

Advantages of Cloud Computing

- **Focus on Core Competency**
  - Government agencies the benefits in order to focus on its core mission and core objectives and leverage IT resources as a means to provide services to citizens.
- **Sustainability**
  - Through leveraging economies of scale and the capacity to manage assets more efficiently, cloud computing consumes far less energy and other resources than a traditional IT data center.

Disadvantages of Cloud Computing

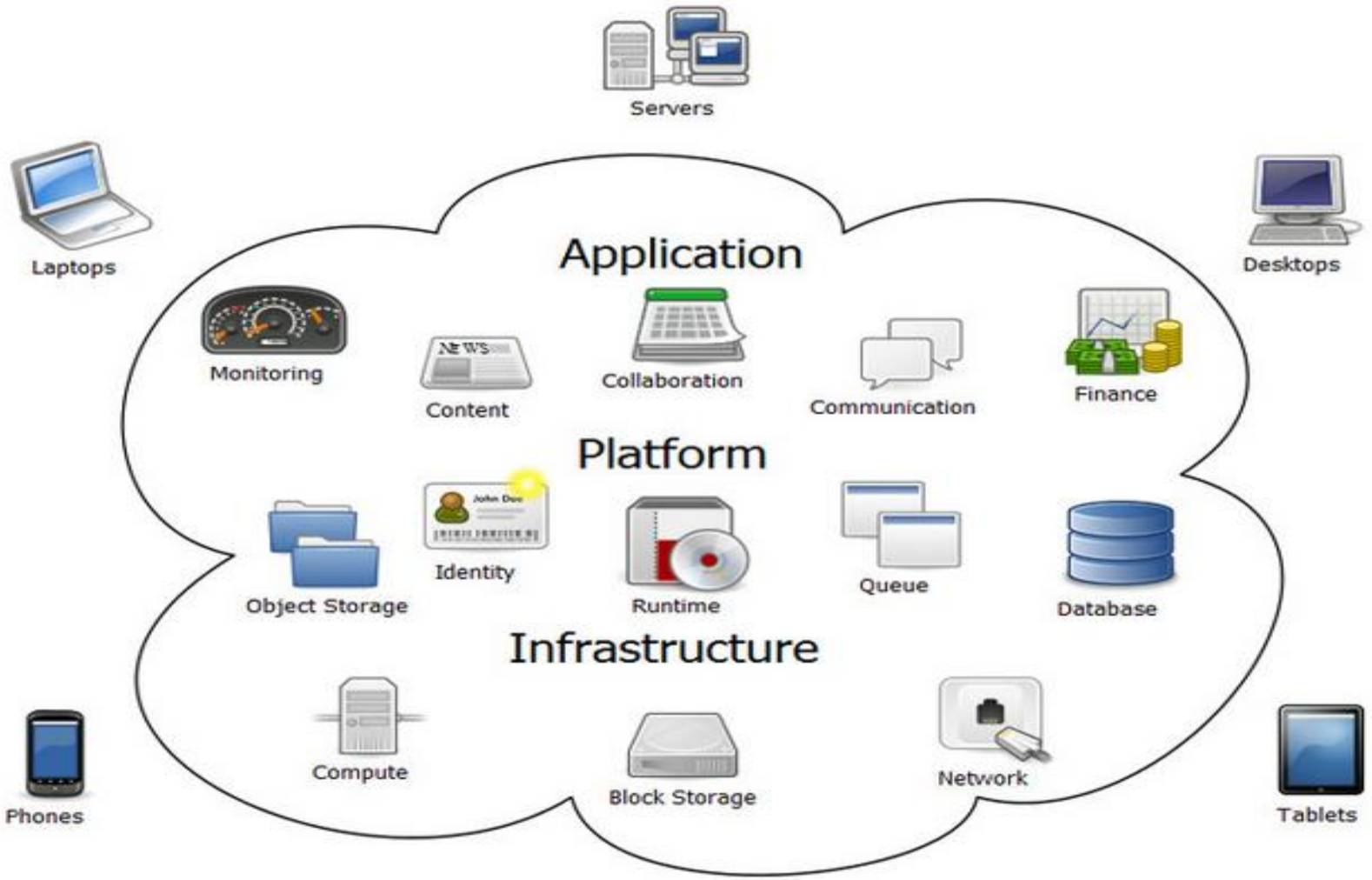
- Requires a high-speed internet connection
- Security and configurability of data
- Interoperability between cloud based systems

Introduction to Resource Management

Introduction to Resource Management

- What is resource ?
- Resource types
- Resource Management

What is resource ?



Resource types

- **Physical resource**
  - Computer, disk, database, network, scientific instruments…
- **Logical resource**
  - Execution, monitoring, communicate application ..

Resource Management

- The term resource management refers to the operations used to control how capabilities provided by Cloud resources and services are made available to other entities, whether users, applications, services.

RM in Cloud Computing Environment

RM in Cloud Computing Environment

- **On Cloud provider's view**
  - Provision resources on both HPC batch-job request and advanced reservation request on the same system
  - Energy efficient resource management in Data Centers
- **On Cloud service provider's view**
  - Renting cheapest resources on performance-constraints.
  - QoS to their cloud users.

RM in Cloud Computing Environment

- **On Cloud user's view**
  - Renting cheapest resources on performance-constraints
  - Cloud provider guarantees Service Level Agreement

RM in Cloud Computing Environment

- Compute Model
- Data Model
- Virtualization
- Monitoring
- Programming Model
- Security Model

Compute Model

- Resource in cloud being shared by all user at the same time.
- Allow latency sensitive application to operate natively on clouds with good enough level of QoS is being delivered to the end user.
- Major for Cloud Computing as Clouds grow in scale and number of users.

Compute Model

- **In Grid**
  - Use a batch-schedule compute model
  - Local resource manager such as PBS, Condor, SGE manages the compute resource for Grid site
  - Identify user to run job
- **Example:**
  - Run application for 60 minutes on 100 processor
  - Use queue

Data Model

- **Relate to mapping, partitioning, querying, movement, caching, replication..**
- **Data is stored at un-trusted host**
  - May not seem to make business sense for a cloud computing host to violate of its customers and access data without permission
  - Moving data off premises increases the number of potential security risks

Data Model

- **Data is replicated, often across large geographic distances**
  - Data availability and durability is paramount for cloud storage providers..
  - Data availability and durability are typically achieved through under-the-covers replication (i.e., data is automatically replicated without customer interference or requests).

Data Model

- **Data management problem**
  - Transactional data management
    - It is hard to maintain ACID guarantees in the face of data replication over large geographic distances.
    - There are enormous risks in storing transactional data on an untrusted host.

Virtualization

- **Has become an indispensable ingredient for almost every cloud**
  - Abstraction
  - Encapsulation
- **Cloud need to run multiple user applications and they appear to users as if running simultaneously and use all available resources.**

Virtualization

- **Abstraction**
  - Provide the necessary abstraction such that the underlying fabric (raw compute , storage, network) can be unified as a pool of resources and resource overlay (data storage services, web hosting environment).
- **Encapsulation**
  - Configured, deployed, started, migrated, suspended, resumed and stopped in each application.
  - Provided better security, manageability isolation

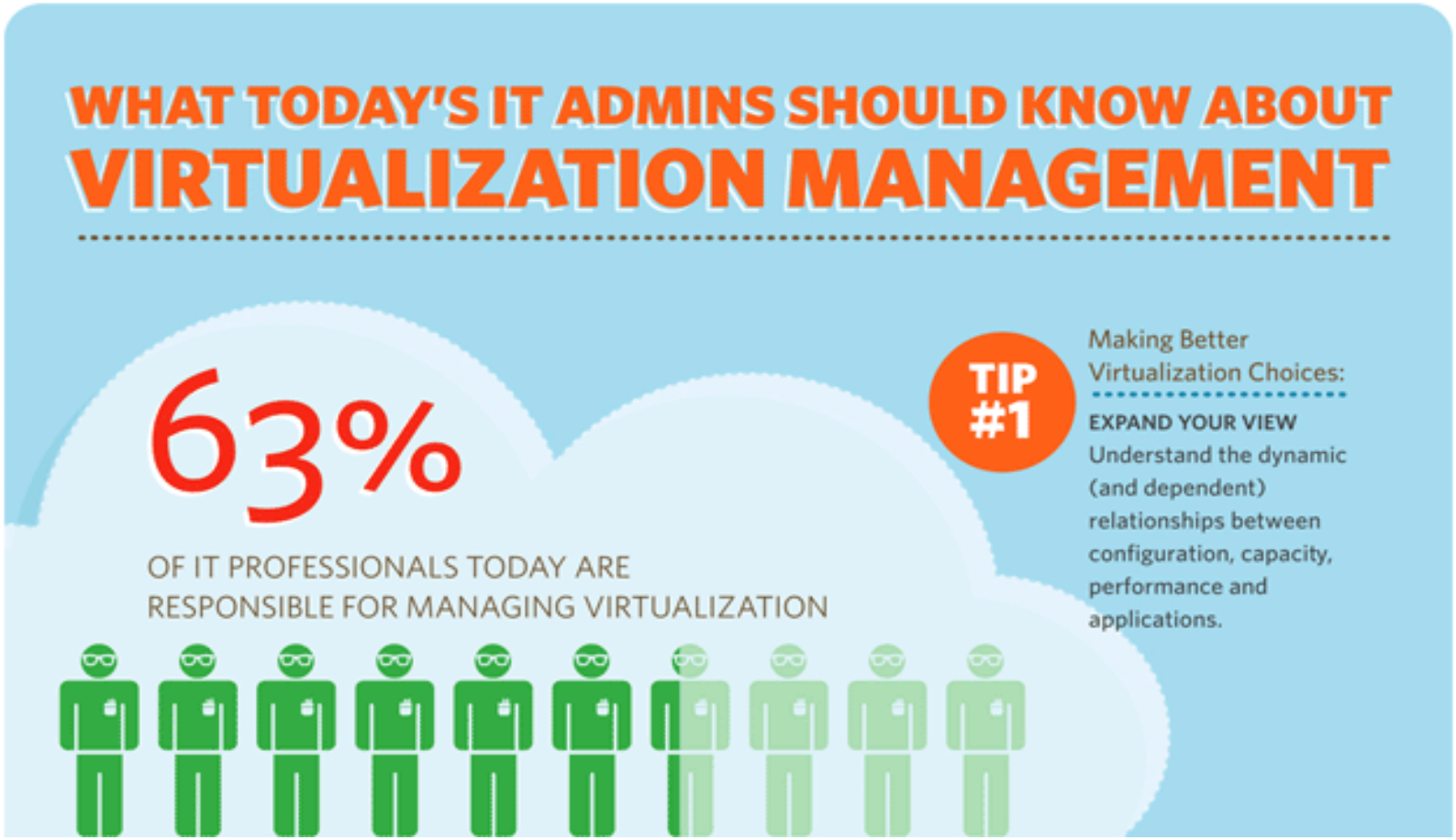
Virtualization

- **Cloud tend to adopt virtualization for some reason**
  - Server and application consolidation
    - Multiple application run on the same server
    - Resource can be utilized more efficiently..
  - Configurability
    - Resource requirement could differ for various app.
    - Large storage, some compute need to dynamically configure and bundle

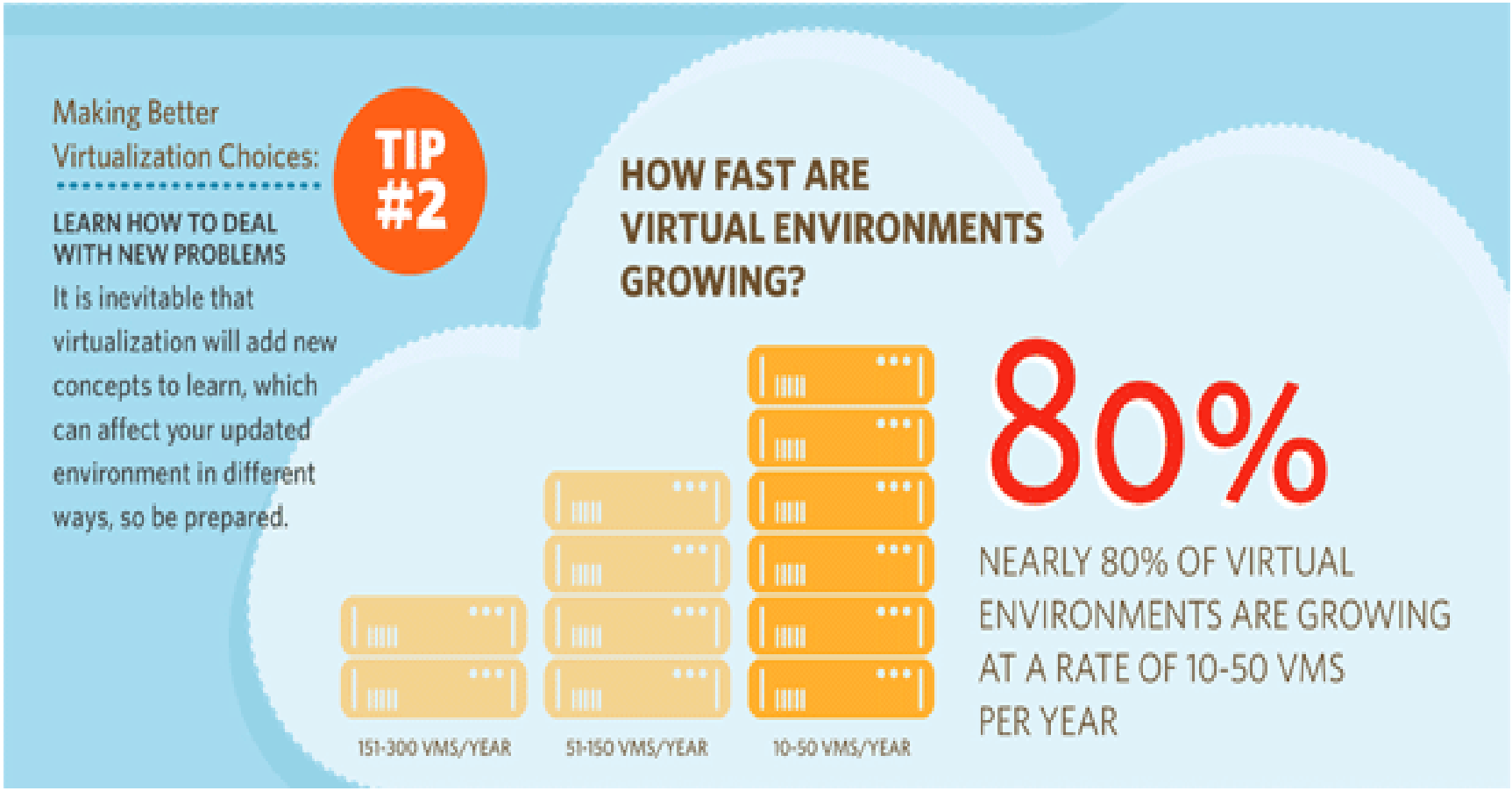
Virtualization

- **Cloud tend to adopt virtualization for some reason (cont.)**
  - Increased application availability
    - Quick recovery from unplanned outages
    - Virtual environment can be backed up and migrated with no interruption in service.
  - Improved responsiveness
    - Resource provisioning, monitoring and maintenance can be automated.
    - Common resource can be cached and re-used.

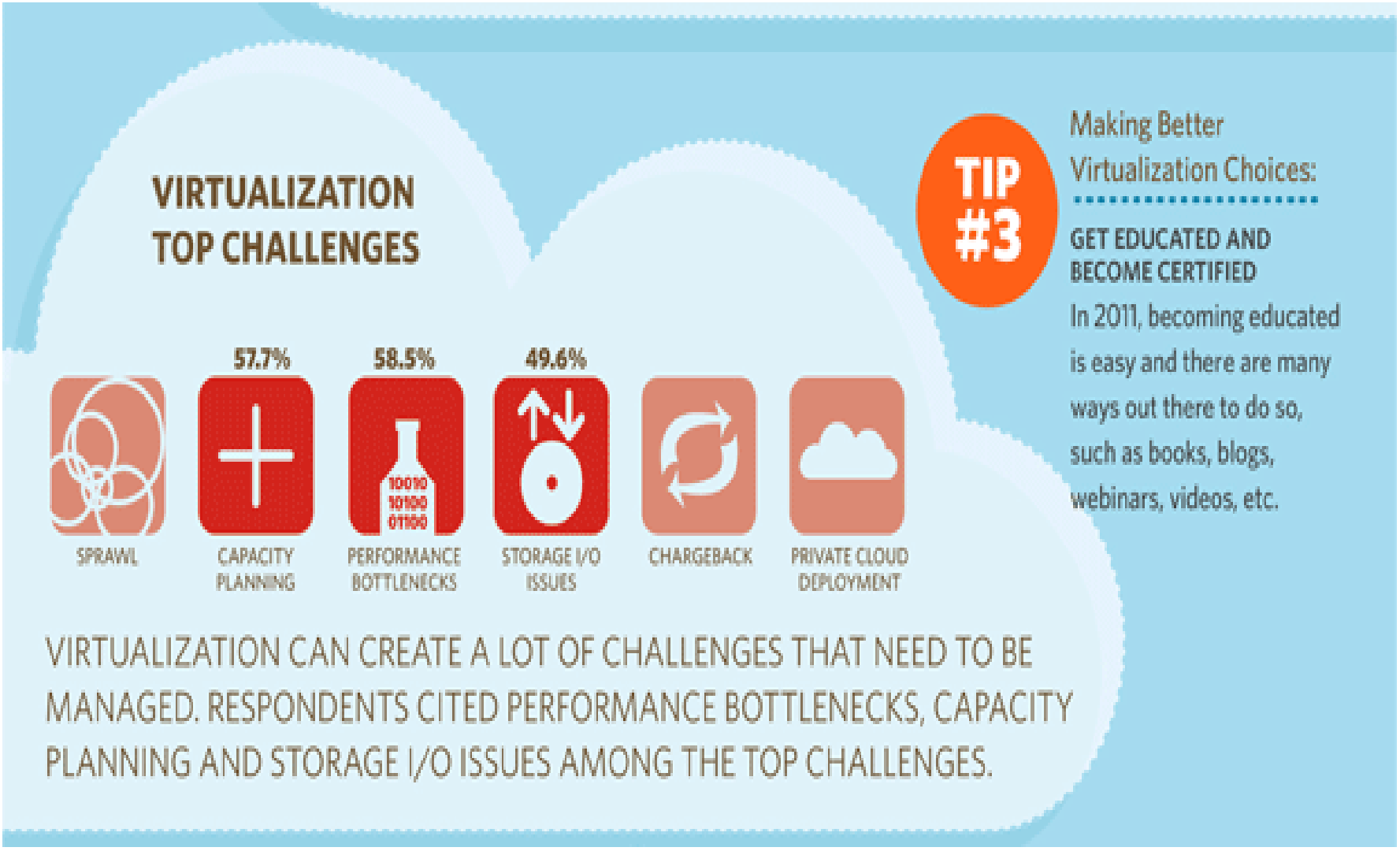
Virtualization



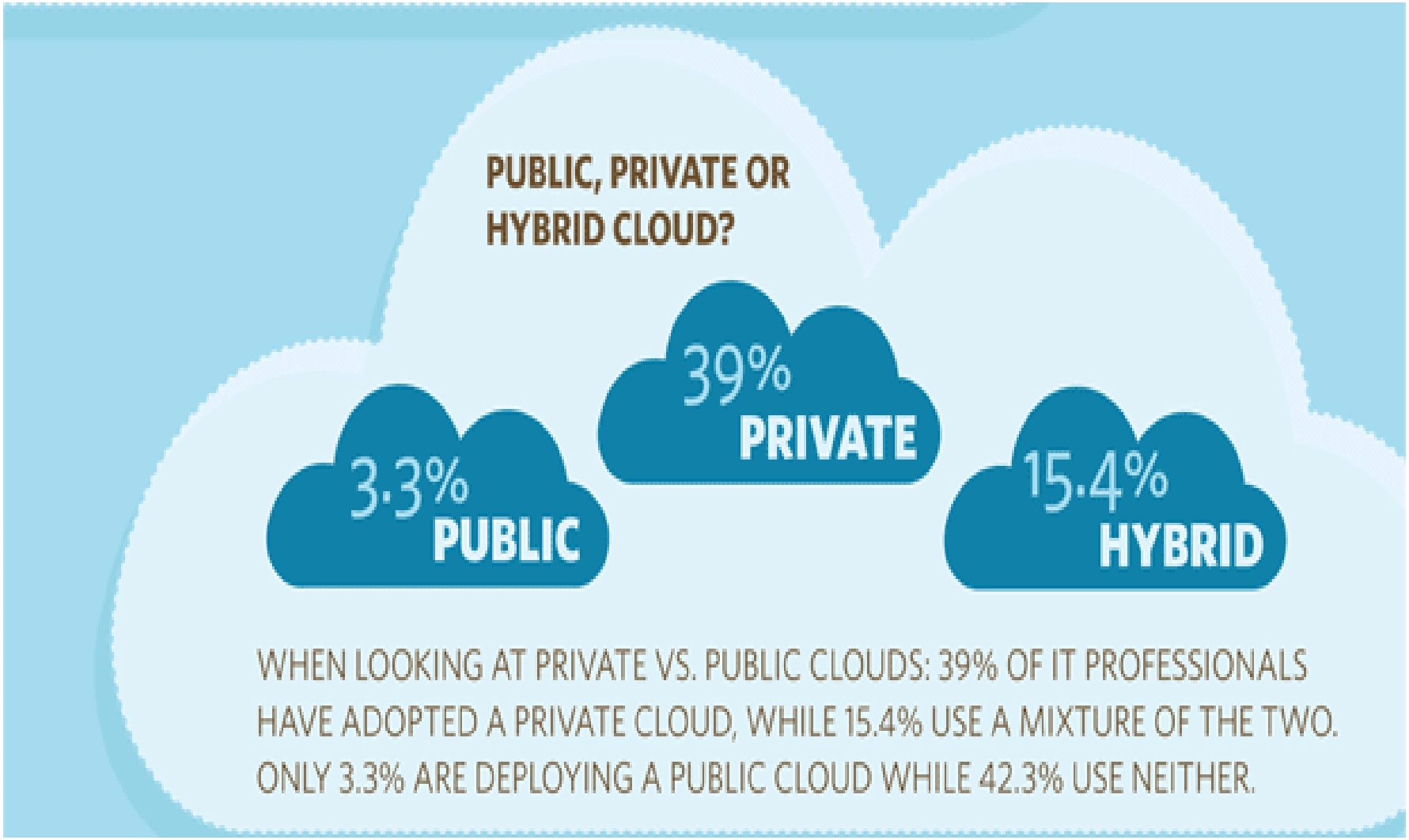
Virtualization



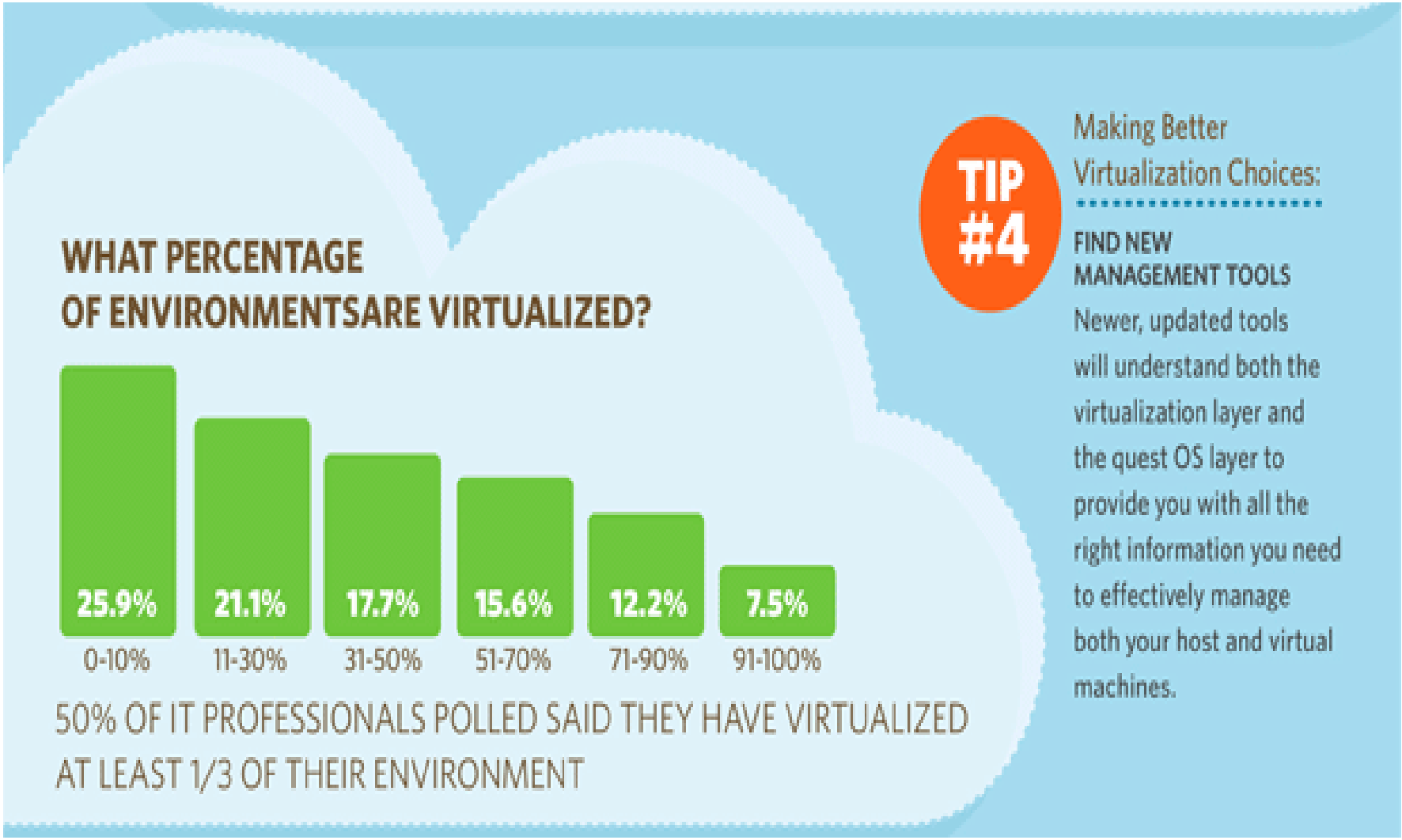
Virtualization



Virtualization



Virtualization



Virtualization

- **In grid**
  - Do not rely on virtualization
  - More due to policy and having each individual organization maintain full control resources
  - Virtual workspace
- **Performance is not problem**

Monitoring

- **The challenge that virtualization brings to Cloud is the potential difficulty in fine-control over the monitoring resource.**
- **In cloud**
  - Different levels of services can be offer to end user.
  - User is only exposed to a pre-defined API.
  - And lower level resources are opaque to user (PaaS, SaaS level some provider may choose to expose monitoring information at these levels).
  - User does not have liberty to deploy own monitoring infrastructure.

Monitoring

❖ In Cloud

- Limited information returned to user may not provide the necessary level detail for them to figure out what the resource status is.
- Require a fine balance of business application monitoring, enterprise server management, virtual machine monitoring and hardware maintenance.
- Other hand:
  - Users interacting with a more abstract layer that is potentially more sophisticated.
  - Abstract layer could respond to failures and QoS automatically.

Monitoring

- **In Cloud**
  - In future, cloud more sophisticated, self-maintenance and self-healing.
- **In Grid**
  - Have a different trust model in which users via their identity delegation can access and browse resources at different Grid sites and Grid resources are not highly abstracted and virtualized as in Clouds.

Programming Model

❖ In Cloud

- Have generally adopted Web Services APIs where users access, configure and program Cloud services using pre-defined as Web Services via HTTP and SOAP protocol.
- The integration and interoperability of all the services and application remain the big challenge.

Programming Model

- **In Grid**
  - Traditional parallel and distributed environment
  - Some issues:
    - Multiple administrative domain
    - Large variations in resource heterogeneity stability and performance.
    - Exception handling in highly dynamic environment

Security Model

- Rely on Web forms(SSL) to create and manage account information for end user
- Allow users to reset their passwords and receive new via Emails in an unsafe and unencrypted communication.
- User could user Clouds easily and instantly with credit card or Email.

Security Model

- **Seven risks:**
  - Privileged use access
  - Regulatory compliance
  - Data location
  - Data segregation
  - Recovery
  - Investigate support
  - Long-term viability

RM Solutions for Cloud Computing

RM Solutions for Cloud Computing

1. **Energy-Efficient Resource Management for Cloud Computing Infrastructures**
2. **Cloud Computing Resource Management through a Grid Middleware**

RM Solutions
Energy-Efficient Resource Management

Energy-Efficient

- Cloud computing is growing in popularity among computing paradigms for its appealing property of considering “Everything as a Service”.
- The goal of a Cloud infrastructure provider is to maximize its profit by minimizing the amount of violations of Quality-of-Service (QoS) levels agreed with service providers, and, at the same time, by lowering infrastructure cost.
- The energy consumption induced by the Cloud infrastructure, for running Cloud services, plays a primary role.

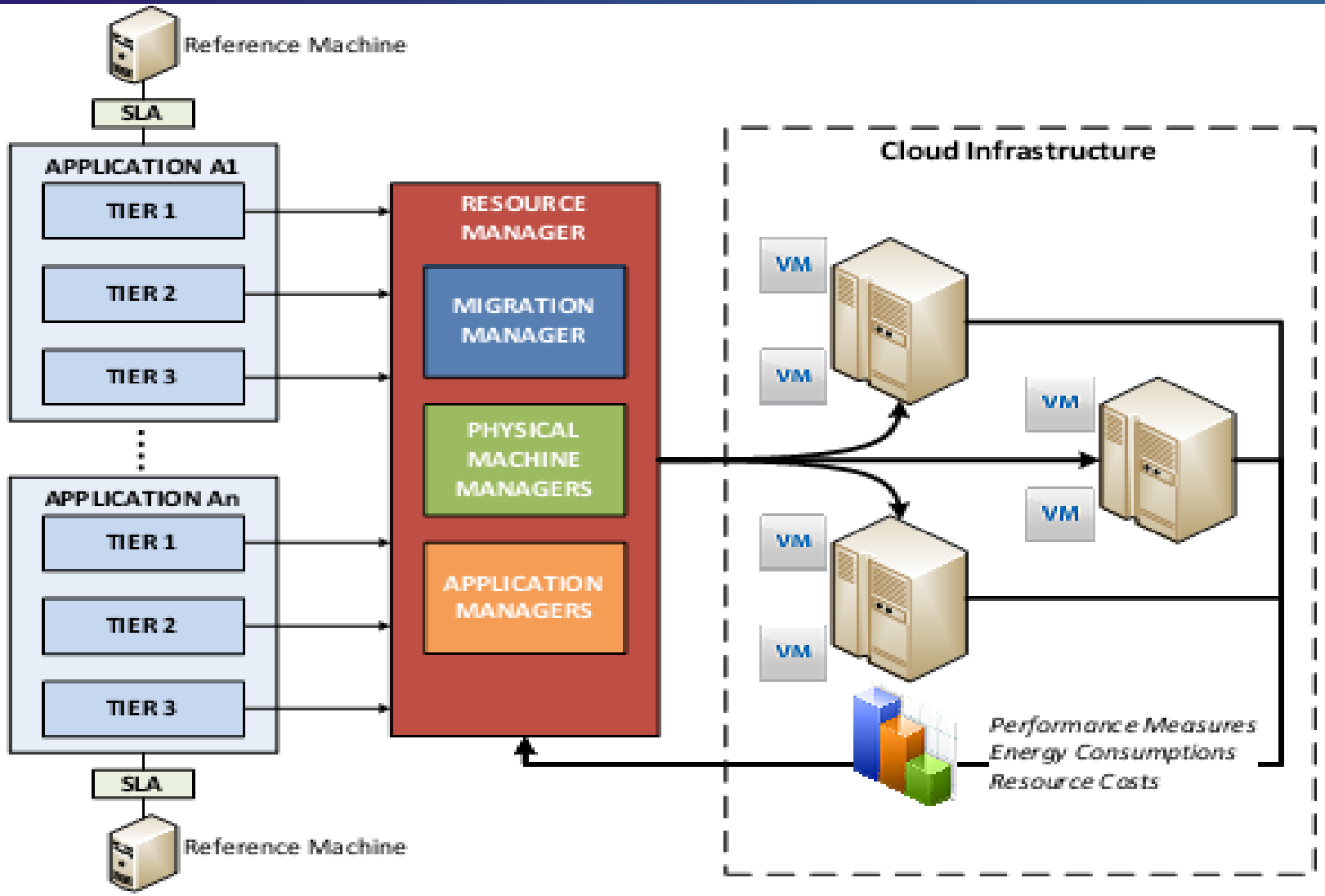
Energy-Efficient

- The minimization of QoS violations and, at the same time, the reduction of energy consumption is a conflicting and challenging problem
- We propose a framework to automatically manage computing resources of Cloud infrastructures in order to simultaneously achieve suitable QoS levels and to reduce as much as possible the amount of energy used for providing services.

The goal of our framework

- To provide automated resource management mechanisms and policies
- Monitor and maintain application performance targets
- Reduce energy consumption in Cloud computing systems.

The resource management framework



Resource Manager

- **Resource Manager accomplishes its goal by dynamically adjusting the fraction of the capacity of physical resources assigned to each VM, and, if needed, by migrating one or more VMs into other and more appropriated physical machine.**
- **Resource Manager:**
  - Application Manager
  - Physical Machine Manager
  - Migration Manager

Application manager

- **Provide the controlled application with the needed amount of resource capacity in order to satisfy its constraints**
  - Monitors the interested performance metrics.
  - Compares them with the related counterparts.
  - Computes the amount of resource capacity each tier should obtain to meet its constraints.
  - Forwards these resource capacity demands to Physical Machine Managers associated to physical machines.

Physical Machine Manager

- Satisfy CPU share demands coming from those Application Managers

Migration Manager

- Monitor, at long-time scale, performance targets and energy consumption.
- Decide which VMs need to be migrated to suitable physical machines (possibly by turning on additional physical machines, for achieving better performance) and which physical machines can be turned off (to save energy).

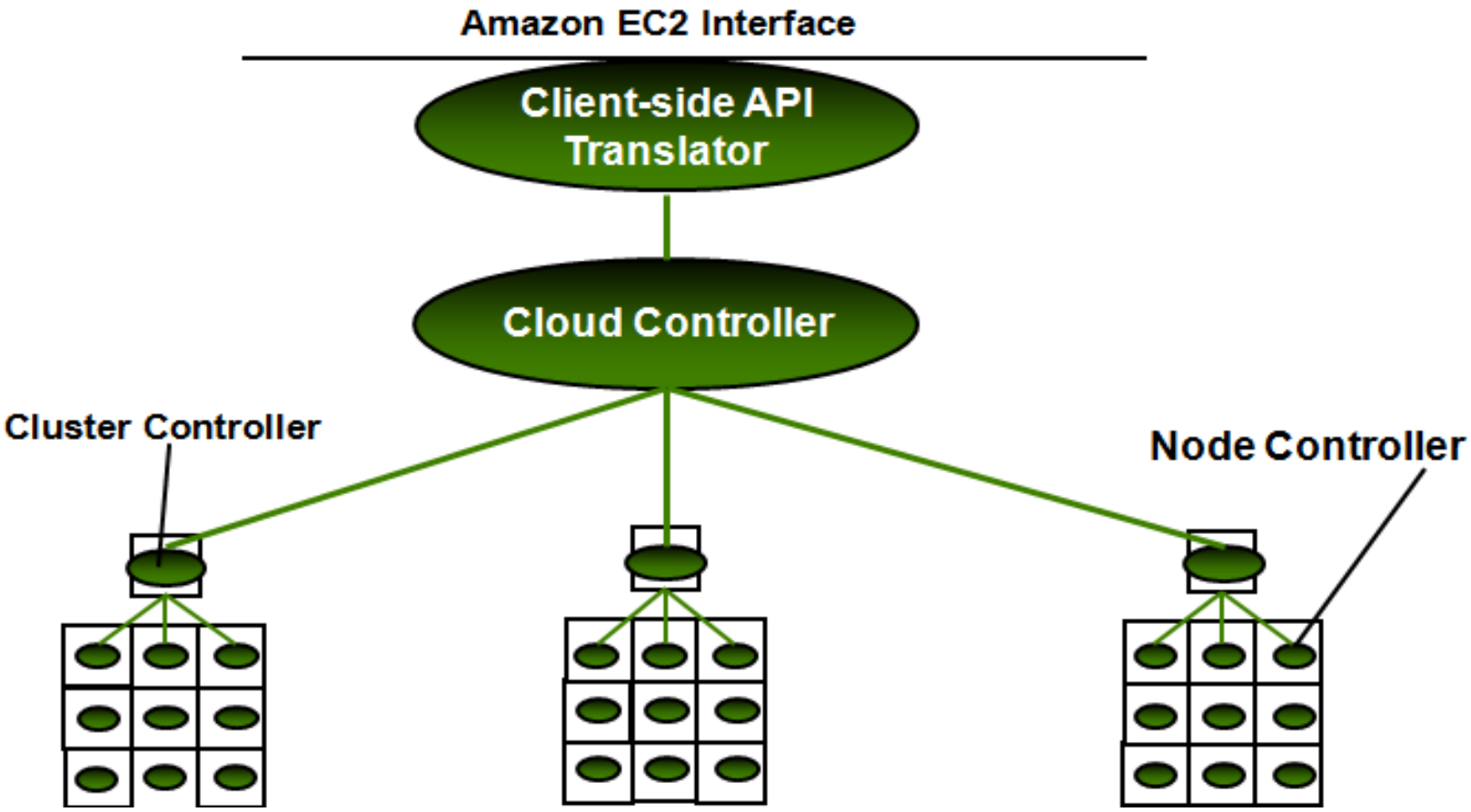
RM Solutions

Resource Management through Grid Middleware

Through Grid Middleware

- **Idea:**
  - Cloud
    - Aim at being dynamically scalable and offer virtualized resources as service over the Internet
    - Balanced when it comes to computational power and storage
  - Grid
    - High computational power
    - Low scalability
  - To get the best of both worlds, a Grid middleware can be used to manage and harness raw Cloud computational resources

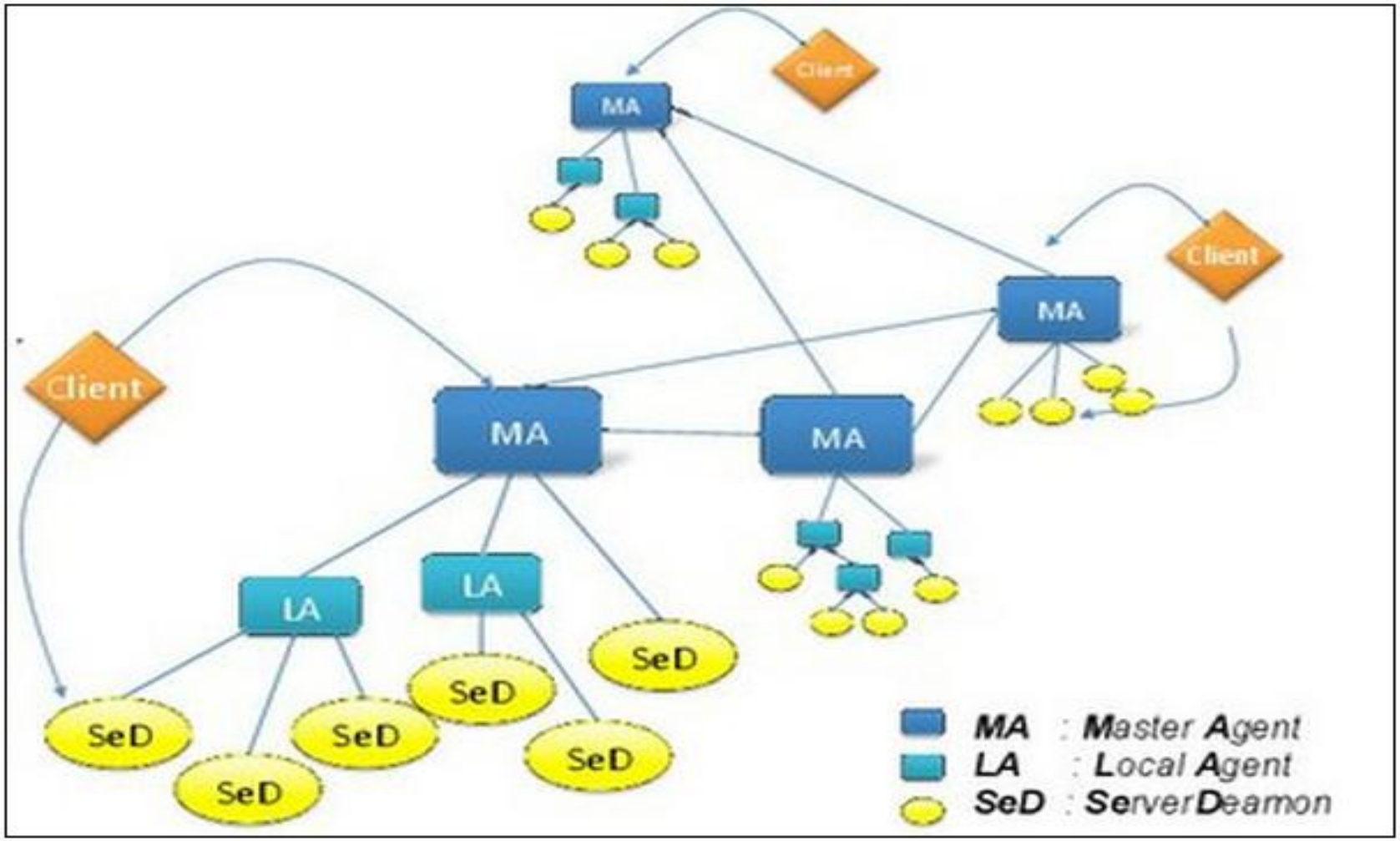
Eucalyptus


Amazon EC2 Interface
Client-side API Translator
Cloud Controller
Cluster Controller
Node Controller

Eucalyptus

- **Cloud Controller :**
  - handle client Cloud management requests
  - only node that is responsible for decision-making
- **Cluster Controller :**
  - keeping track of resource usage in its Cluster
- **Node Controller :**
  - Each physical machine that is to be a computing machine needs to have the NC service running
  - Two main responsibilities:
    - monitoring resource usage.
    - managing virtual resources.

DIET: The grid & cloud middleware



http://graal.ens-lyon.fr/DIET/diet-1/architecture

DIET component

- **Client**
  - an application which uses DIET to solve problems
- **Master Agent (MA)**
  - An MA receives computation requests from clients. Then the MA collects computation abilities from the servers and chooses the best one. The reference of the chosen server is returned to the client. A client can be connected to an MA by a specific name server or a web page which stores the various MA locations.

DIET component

- **Local Agent (LA)**
  - An LA aims at transmitting requests and information between MAs and servers. The information stored on an LA is the list of requests and the number of servers that can solve a given problem and information about the data.

DIET component

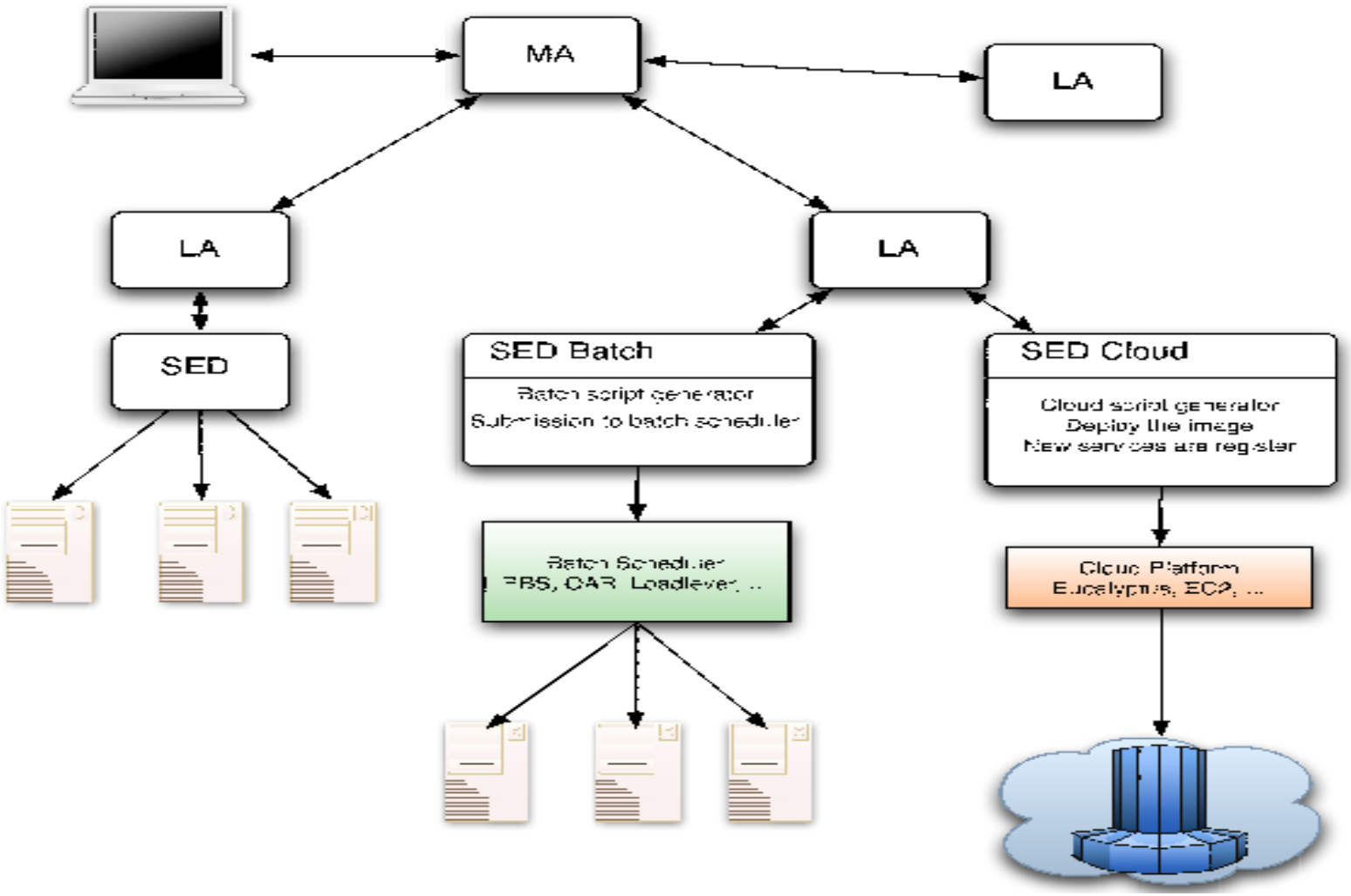
- **Server Daemon (SeD)**
  - A SeD is the point of entry of a computational server. It manages a processor or a cluster. The information stored on a SeD are the list of the data available on a server the list of problems than can be solved on it, and every information concerning its load (CPU capacity, available memory, etc.).

DIET-Solve is completely outside of EUCALYPTUS

❖ **DIET-Solve is completely outside of EUCALYPTUS**

- DIET-Solve and EUCALYPTUS architectures do not overlap at all in the sense.
- DIET-Solve agents or SEDs run separately with respect to the EUCALYPTUS controllers
- The DIET-Solve SED requests resources (compute nodes) to EUCA-LYPTUS when needed and uses the resources directly once they are ready.
- In this scenario scalability is limited because of the fixed DIET-Solve architecture that cannot scale easily, but the number of compute nodes takes full advantage of EUCALYPTUS's scalability.

Two scenarios (cont)



Cloud Computing Frameworks and Resource Management

Cloud Computing Frameworks

1. **Amazon** http://aws.amazon.com/ec2/
2. **Google** http://www.google.com/apps/intl/en/business/index.html
3. **Hewlett Packard** http://www.hpl.hp.com/research/cloud.html
4. **IBM** http://www.ibm.com/grid/
5. **Microsoft** http://www.microsoft.com/azure/default.mspx
6. **Salesforce.com** http://www.salesforce.com/cloudcomputing/

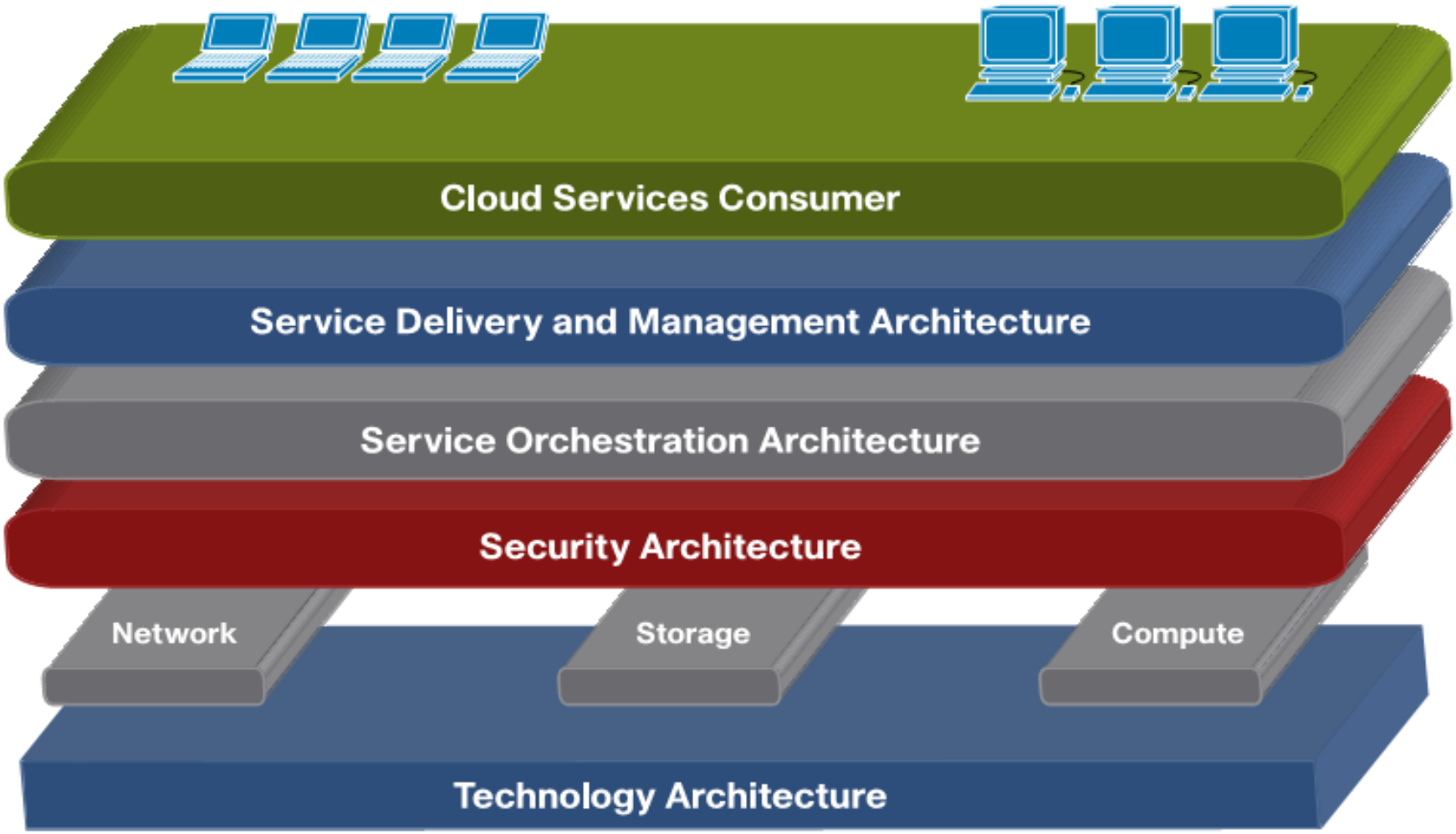
....

Cloud Computing Frameworks

1. Cisco Framework
2. Green cloud Framework

Cisco Framework

Cisco cloud


Cloud Services Consumer
Service Delivery and Management Architecture
Service Orchestration Architecture
Security Architecture
Network
Storage
Compute
Technology Architecture

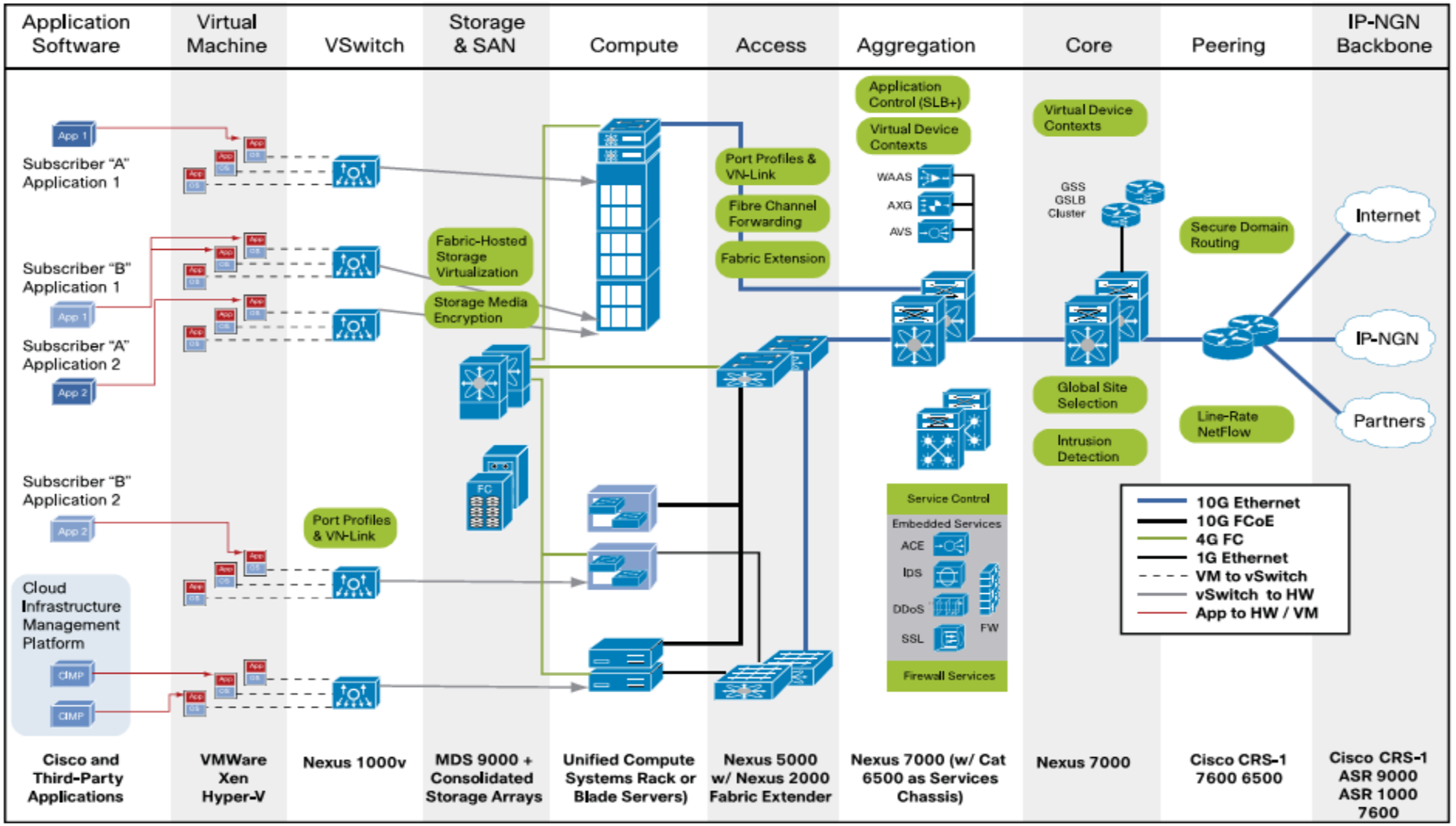
Cisco cloud reference architecture framework

- The foundation of this framework is the data center technology architecture, which consists of three salient blocks of network, compute, and storage.
- Security is blanketed as an end-to-end architecture across all aspects of the framework.
- Service Orchestration layer is implemented with configuration repository enablers. The configuration repository stores key information such as service catalogue, asset inventory, and resource-to-service mappings.

Cisco cloud reference architecture framework

- Service delivery management architecture is where infrastructure and service management function take place.
- Cloud service consumer: the consumer-facing layer, usually exposed via a portal-like solution and is defined, requested, and managed by the consumer.

Cisco Cloud Data Center



The Cisco UCS management framework

- Provides robust API for managing all system configuration and operation.
- Helps increase cloud data center staff productivity, enabling better management of storage, networking, computing, and applications to collaborate on defining service profiles for applications.
- Service profiles help automate provisioning, allowing cloud data center to provision applications in minutes instead of days.

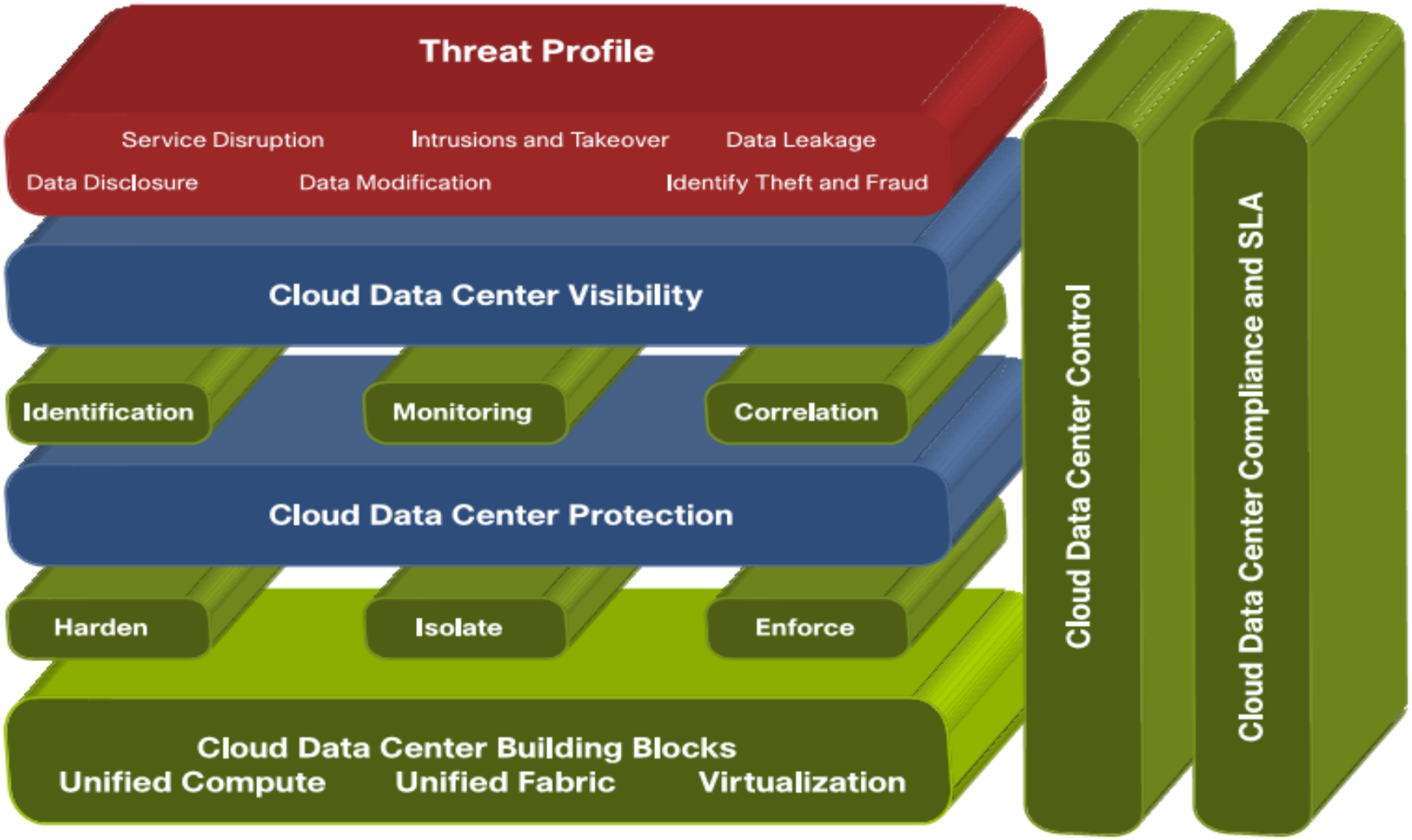
The Cisco UCS management framework

- Stateless computing: separate the access to the application from the execution of the application.
- Provides support for a unified fabric over a low-latency, lossless, 10-Gbps Ethernet foundation
- Extends virtualized data centers and creates a foundation for private clouds that federate with compatible virtual private clouds.

Cisco Secure Cloud Data Center Framework

- **Security: Traditional issues around data and resource access control, encryption, and incident detection are factors here**
- **Control: The ability of the enterprise to directly manage how and where data and applications are deployed and used**
- **Compliance and service-level management (SLA):refers to contracting and enforcement of service-level agreements between varieties of parties, and conformance with regulatory, legal, and general industry requirements.**

Cisco Secure Cloud Data Center

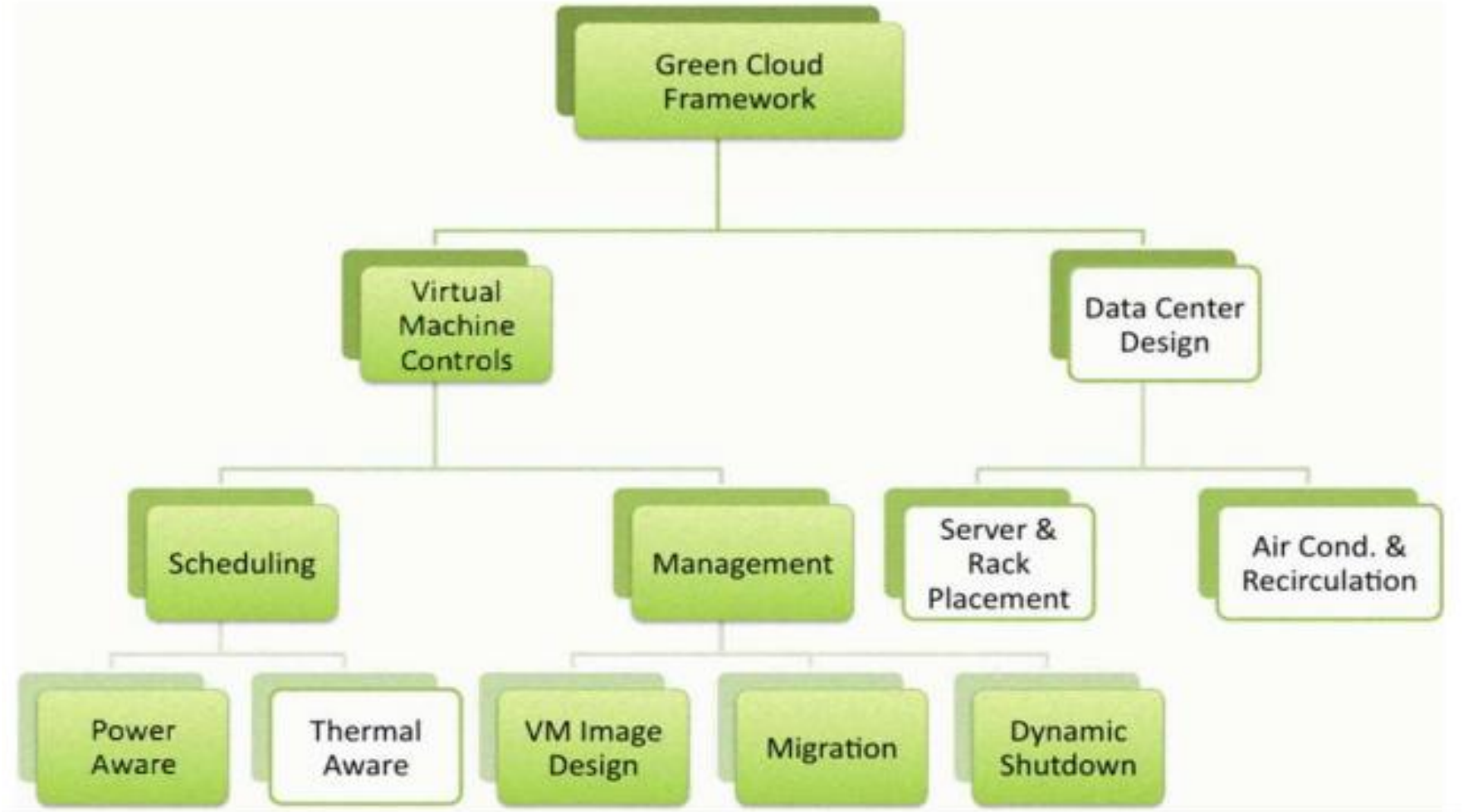


Green Cloud Framework

Green Cloud Framework

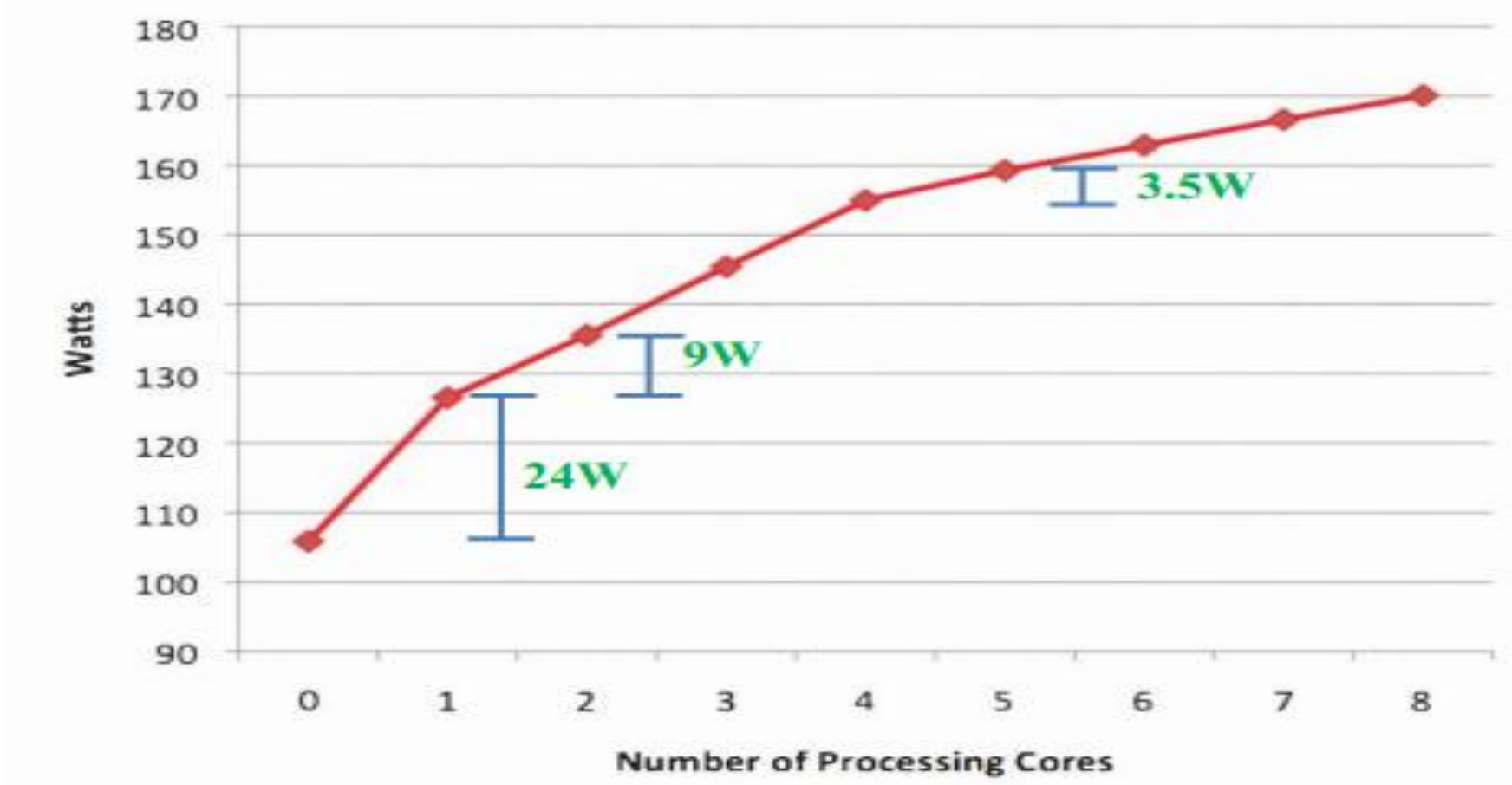
- **Applied to the Cloud in order to meet the goal of reducing power consumption.**
  - VM scheduling
  - VM image management
  - Advanced data center design

Green Cloud Framework



VM Scheduling & Management

Power consumption curve of Intel i7 920CPU



VM Scheduling & Management

❖ Scheduling Algorithm

Algorithm 1 Power based scheduling of VMs

```
FOR i = 1 TO i ≤ |pool| DO
  pe_i = num cores in pool_i
END FOR

WHILE (true)
  FOR i = 1 TO i ≤ |queue| DO
    vm = queue_i
      FOR j = 1 TO j ≤ |pool| DO
        IF pe_j ≥ 1 THEN
        IF check capacity vm on pe_j THEN
          schedule vm on pe_j
          pe_j − 1
        END IF
      END IF
    END FOR
  END FOR
  wait for interval t
END WHILE
```

VM Scheduling & Management

❖ VM Management

- Idle physical machines in a Cloud can be dynamically shutdown and restarted to conserve energy during low load situations.
- Live migration can be applied to Green computing in order to shift VMs from low load to medium load machines when needed.
- Low load servers can be shutdown when all VMs have migrated away.

VM Scheduling & Management

- **Service oriented virtual machine image**
  - Full OS VMs are scheduled often to carry out specific tasks. The images contain much more than they need.
    - Support a wide variety of hardware & software.
    - But each VM is typically designed for a specific task
  - A hypervisor provides the same virtualized hardware to each VM.
    - We want the OS within the VM to act only as a light wrapper which supports a few specific but refined tasks or services, not an entire desktop suite.

VM Scheduling & Management

- **Service oriented virtual machine image**
  - We concentrate on two areas: VM images and boot time
  - Customizing the VM environment
    - Remove the kernel modules which are not needed.
    - Remove the daemons and applications which are not needed.

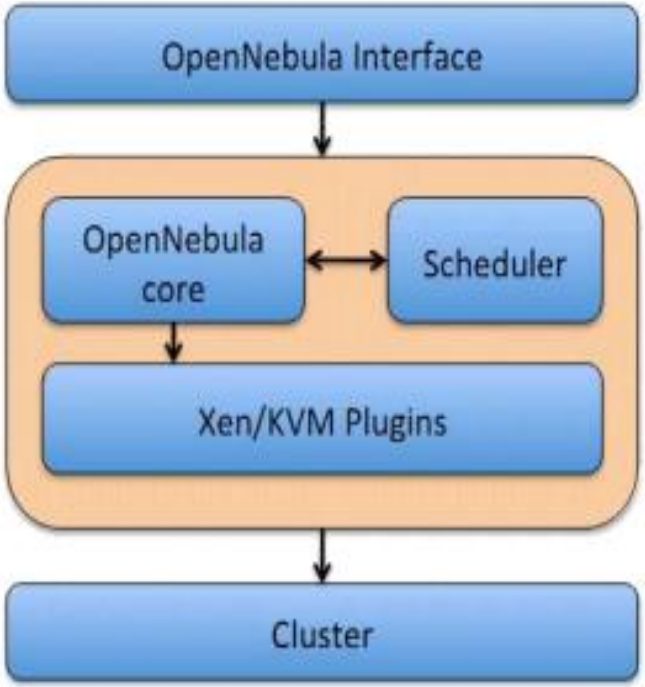
VM Scheduling & Management

- **Service oriented virtual machine image**
  - Reducing the boot time
    - *bootchart*: Profiles where boot-up system inefficiencies occur and to allow for optimization of the boot sequence.
    - *readahead*: Profiles the system startup sequence and use prefetching to load files into memory before they are requested.

Power Consumption Analysis

❖ Scheduling Analysis

- OpenNebula
  - Customized scheduler component
- 4 machines
  - CPU: Intel Core i7 920 2.6GHz
  - RAM: 12GB RAM
- Schedule 8 VMs



Power Consumption Analysis

Fig. 7. Illustration of Scheduling power savings

Power Consumption Analysis

- **VM image Analysis**
  - Top-down approach: Based on Ubuntu 9.04 Jaunty, removing all unnecessary packages.
  - Number of kernel modules were removed from the 2.6.28-11 kernel to speed up the init and modprobe processes.
- **Test boot time**
  - Basic Ubuntu 9.04 vs. Custom VM image
  - VMware server
    - CPU: Intel Core2 Duo 2.5GHz
    - RAM: 4GB

Power Consumption Analysis

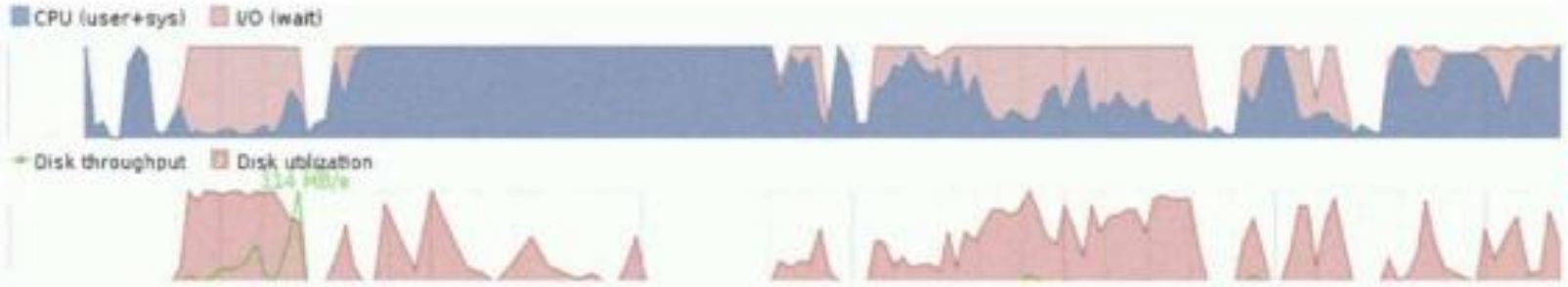


Fig. 8. Bootup chart of the default Ubuntu Linux VM image

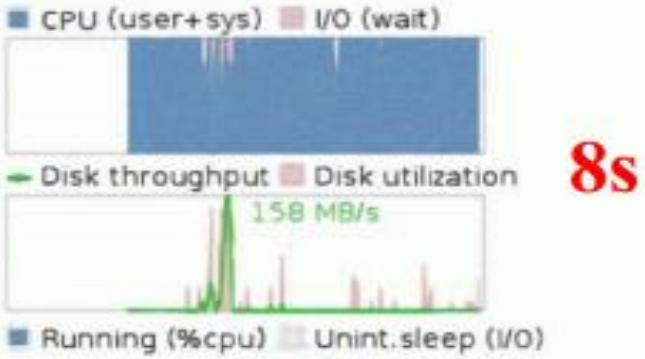


Fig. 9. Bootup chart of Minimal Linux VM image

Conclusions

Conclusions

- Cloud Computing
- Resource Management
- Some solutions & frameworks

References

1. **David F. Soll (2009).** *Cloud Computing - A General State of the Union.*
2. **Ian Foster et al (2009).** *Cloud Computing and Grid Computing 360-Degree Compared.* **Grid Computing Environments Workshop.**
3. **Andrew J. Younge et al (2010).** *Efficient Resource Management for Cloud Computing Environments*
4. **CISCO (2009).** *Cisco Cloud Computing – Data Center Strategy, Architecture, and Solutions***.**
5. **María S. Pérez (2008).** *Grid and Cloud Computing.*
6. **Marco Guazzone et al(2011).** *Energy-Efficient Resource Management for Cloud Computing Infrastructures.*

Thank You !