Grid and Cloud Computing

María S. Pérez
Facultad de Informática
Universidad Politécnica de Madrid
mperez@fi.upm.es

Outline

- Challenges not yet solved in computing
- Grid computing
- Cloud computing
- References

Challenges not yet solved in computing

Grand Challenge Applications

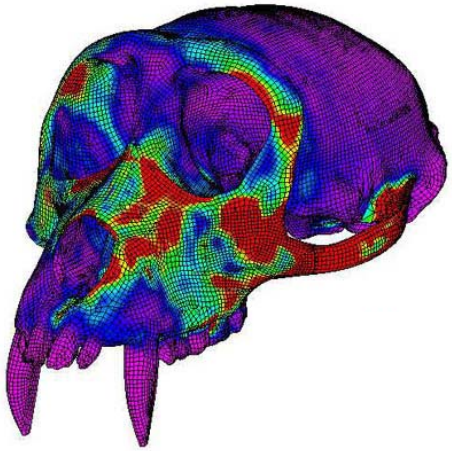
Aerospace:

Life sciences:

E-commerce:

Earth sciences:

Biology:



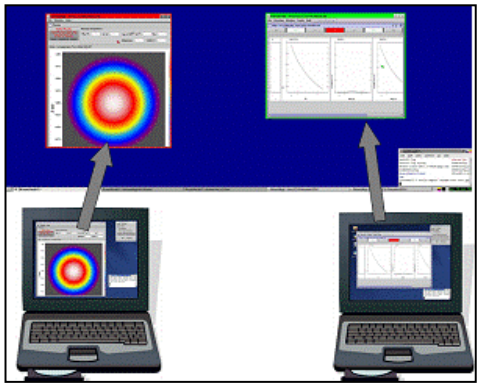
Possible Scenarios

- A biochemist exploits 10,000 computers to screen 100,000 compounds in an hour
- 1,000 physicists worldwide pool resources for petaop analyses of petabytes of data
- Civil engineers collaborate to design, execute, & analyze shake table experiments
- Climate scientists visualize, annotate, & analyze terabyte simulation datasets
- An emergency response team couples real time data, weather model, population data
  - Source: *Slides "The Challenges of Grid Computing" Ian Foster*

Means of Solving the Problem

Cluster Computing

Intranet Computing



Grid Computing

Cloud Computing

Grid Computing

Grid Computing

- Grid Computing is based on the philosophy of information and electricity sharing, allowing us to access to another kind of heterogeneous and geographically separated resources
- Grid provides the sharing of:
  - Computational resources
  - Storage elements
  - Specific applications
  - Equipment
  - Other
- Thus, Grid is based on:
  - Internet protocols
  - Ideas of parallel and distributed computing

A Three Point Checklist

- *"A Grid is a system that...*
  - *1) ...coordinates resources that are not subject to a centralized control...*
  - *2) ...using standard, open, general-purpose protocols and interfaces...*
  - *3) ...to deliver nontrivial qualities of services."*

Ian Foster

What is the Grid? A Three Point Checklist (2002)

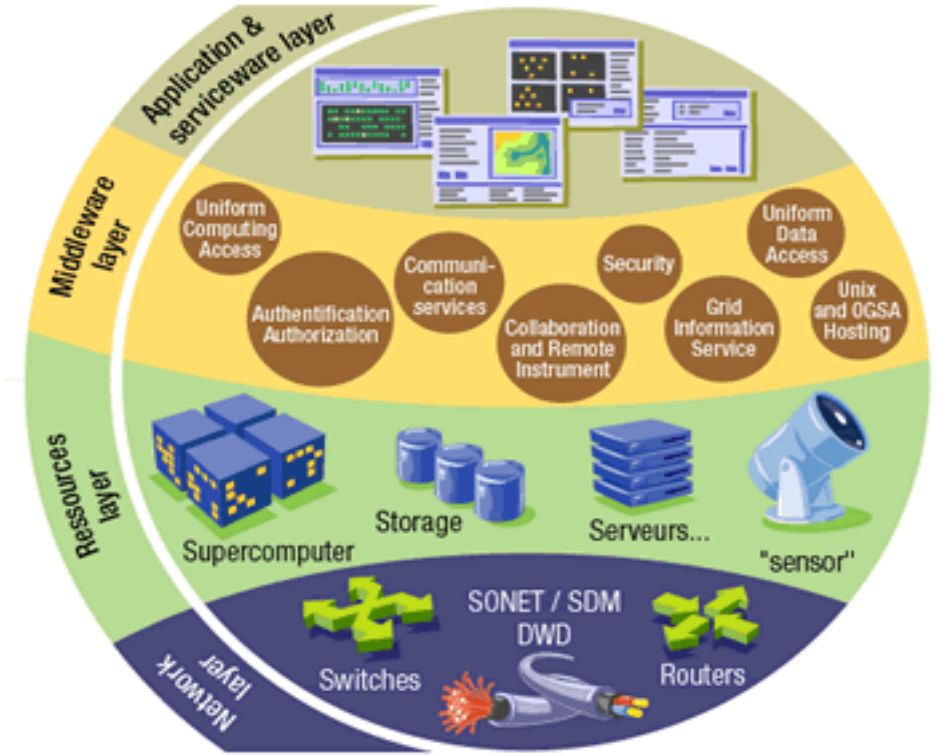
The "Grid" Scenario

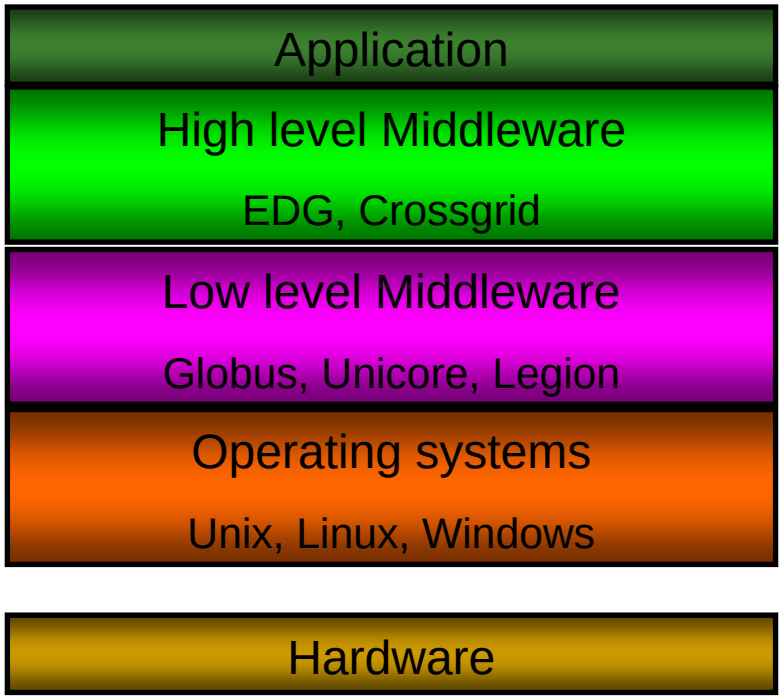
- Flexible, secure, coordinated resource sharing among individuals and institutions
- Enable communities (*virtual organizations*) to share geographically distributed resources in order to achieve a common goal
  - In applications which cannot be solved by resources of an only institution
  - Or the results can be achieved faster and/or cheaper

Idiosyncrasy of the scenario

- Dynamic virtual organizations
  - A set of individual and/or institutions which share rules
  - Large or small
  - Static or dynamic
- Kind of resources
  - Heterogeneous resources
  - Computers, storage, sensors, networks, etc.
- Coordinated problem solving
  - Distribution
  - Collaboration
  - Trust, policies, negotiation, payment
- Challenges
  - Security
    - Authentication
    - Authorization
  - Resource access
  - Resource discovery
  - Scheduling
  - Data management…

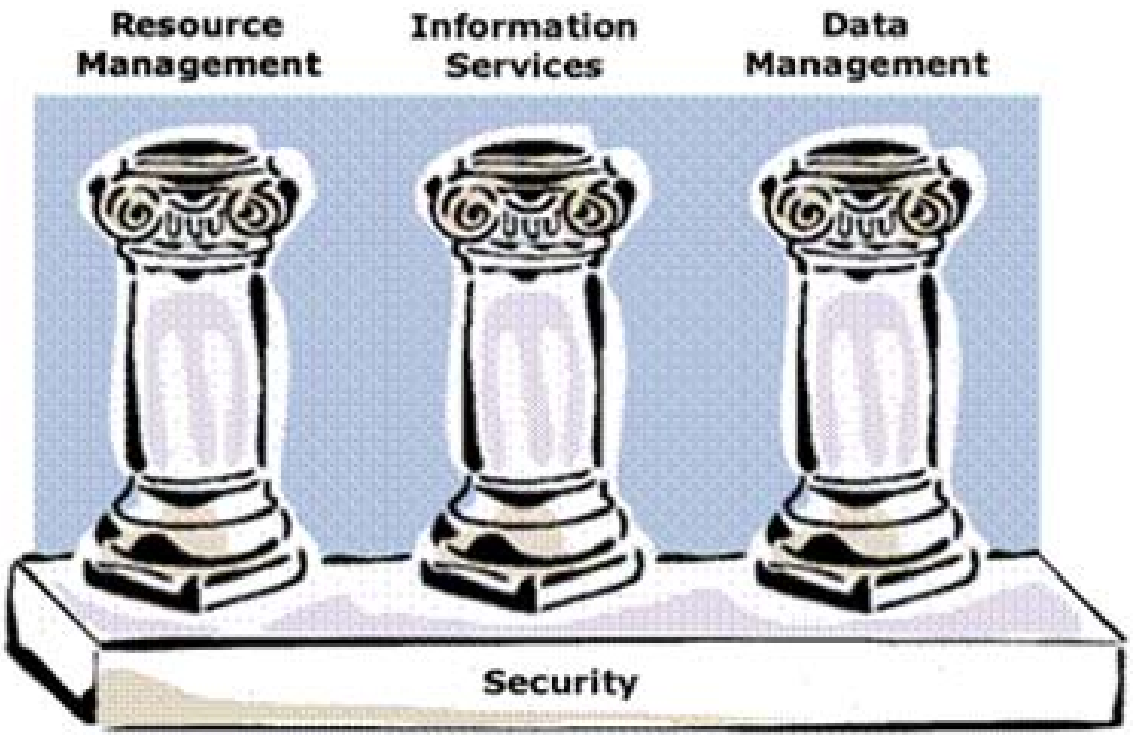
Grid Architecture





Source: Grid Café: What is the Grid?
http://gridcafe.web.cern.ch

Basic pillars



Need of security

- Distributed resources
- No centralized control
- Different resource providers
- Each resource provider uses different security policies

Security in Grid

- Generic Security Services (GSS)
  - Authentication, delegation, integrity and confidentiality
  - Public Key Infrastructure (PKI) with X.509 certificates
  - Kerberos
  - Secure Socket Layer (SSL)
- Grid Security Infrastructure (GSI)
  - Delegation
  - Single Sign-On ⟶ Proxy certificates

Certificate request

- A user asks for a certificate to a Certification Authority (CA)
- The CA checks the user identity
- Then, the CA signs the request, creating a certificate, and returning it to the user
  - Certificates can be cancelled
    - Certificate Revocation List (CRL)
- The aim of the certificates is described in the certificate policy (CP)

Information Systems

- Provide information on:
  - The Grid itself
    - The user may query about the status and performance of the Grid
  - Grid applications
- Register and monitor resources
- Standardization is required to interoperate among different grids projects
  - Globus: MDS (Monitoring and Discovery Service)
  - European Data Grid: R-GMA (Relational Grid Monitoring Architecture)
  - UNICORE: Incarnation Database (IDB)

Data Grid

- Set of storage resources and data retrieval components which allows applications to access data by means of special software mechanisms
- Data grid problems:
  - Data location
  - Replication
  - I/O performance

Data Transfer

- GridFTP: Protocol to data transfer in a secure way in a grid environment
  - Extends FTP protocol
  - Use Grid Security Infrastructure (GSI)
  - Several storage systems provide GridFTP interfaces:
    - Castor
    - EDG's SRM
- Reliable File Transfer (RFT): Grid Service which provides interfaces to manage and monitor file transfers by using GridFTP servers

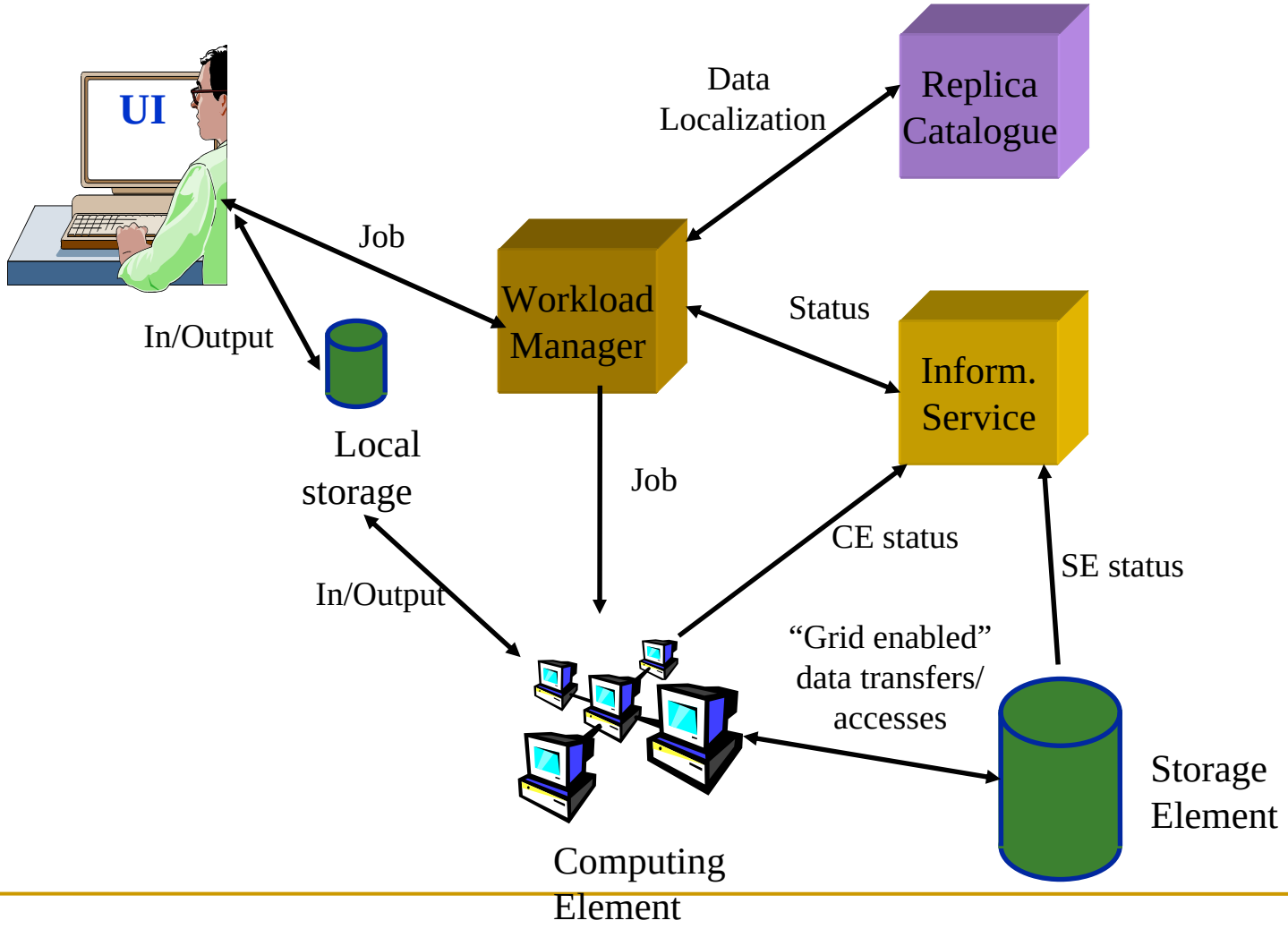
Data replication

- Due to the complexity of a grid environment, the existence of file replicas could be advisable
- Need of identifying and locating replicas
- Replica Location Service (RLS): a Grid Service for registering data replicas and later discovering
  - Mappings between logic and fisical identifiers
  - Database for metadata

Resource management system

- Resource Management includes the efficient use of computing and storage resources
  - Processor time
  - Memory
  - Storage
  - Network
- User-transparent
- Interacts with the rest of Grid components

Job submission



Job queue managers

Condor-G: Condor High-Throughput Computing Project

http://www.cs.wisc.edu

Portable Batch System (PBS)

Sun Grid Engine (SGE)

Grid Computing approaches

mix-and-match

Object-oriented

Legion

Internet-WWW

Web-based technologies

Problem Solving Approach

Market/Computational Economy

Nimrod-G/CPM

Grid Projects

- Globus: http://www.globus.org
- EGEE: http://www.eu-egee.org/
- TeraGrid: http://www.teragrid.org
- CrossGrid: ***http://www.crossgrid.org***
- EU-DataGrid: *http://www.eu-datagrid.org*
- *IrisGrid: http://www.rediris.es/irisgrid*
- myGrid: http://www.mygrid.org.uk/
- …

Cloud Computing

Cloud Computing

- Future scenario:
  - No computing on local computers
  - Third-party compute and storage facilities
- Cloud Computing:
  - *"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"**
- Is it just a "fashion" name for grid computing?

Cloud computing history

- 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

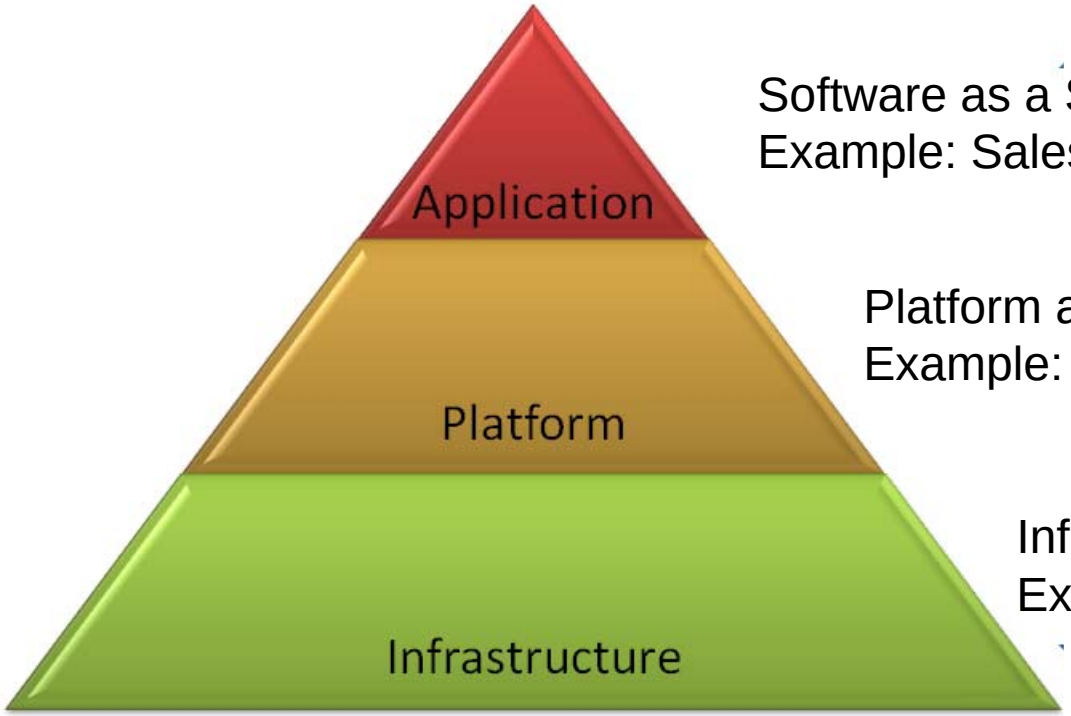
Source: "A history of cloud computing", Arif Mohamed, March 2009

Cloud computing history

- 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)

Source: "A history of cloud computing", Arif Mohamed, March 2009

Cloud computing layers



Software as a Service (SaaS).
Example: Salesforce.com

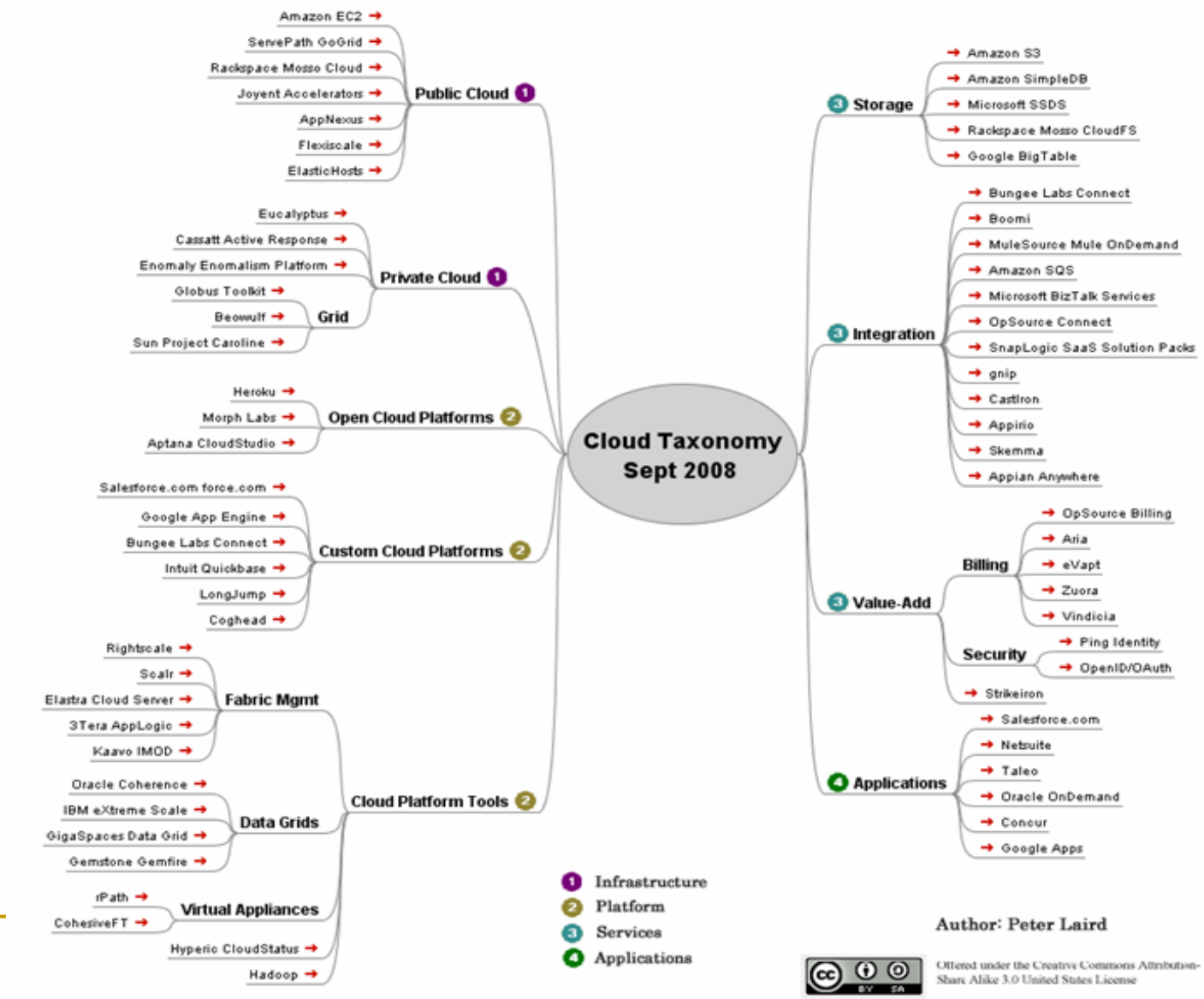
Platform as a Service (PaaS)
Example: Microsoft Azure

Infrastructure as a Service (IaaS)
Example: Amazon Web Services

Cloud Deployment Models

- Private cloud
  - Managed by an organization
- Community cloud
  - Shared by several organizations
  - Intended to one community
- Public cloud
  - General public
  - Owned by an organization selling cloud services
- Hybrid cloud
  - Composed by two or more clouds

Cloud Taxonomy



Cloud computing characteristics

- Elasticity: Resource allocation can be increased or decreased according to the demand
- Scalability: the cloud scales according to the demand
- Self-service provisioning: Cloud customers accesing cloud services
- Standardized interfaces: Standard APIs
- Billing service: A pay-as-you-go model

Cloud computing characteristics

- Large scale
  - Economies of scale
- Service oriented architecture
  - Stateless
  - Low coupled
  - Modular
  - Semantically interoperable
- Virtualization
  - It enables elasticity
  - Cost savings
- Autonomic computing
  - Self-configuring
  - Self-healing
  - Self-optimizing
  - Self-protecting

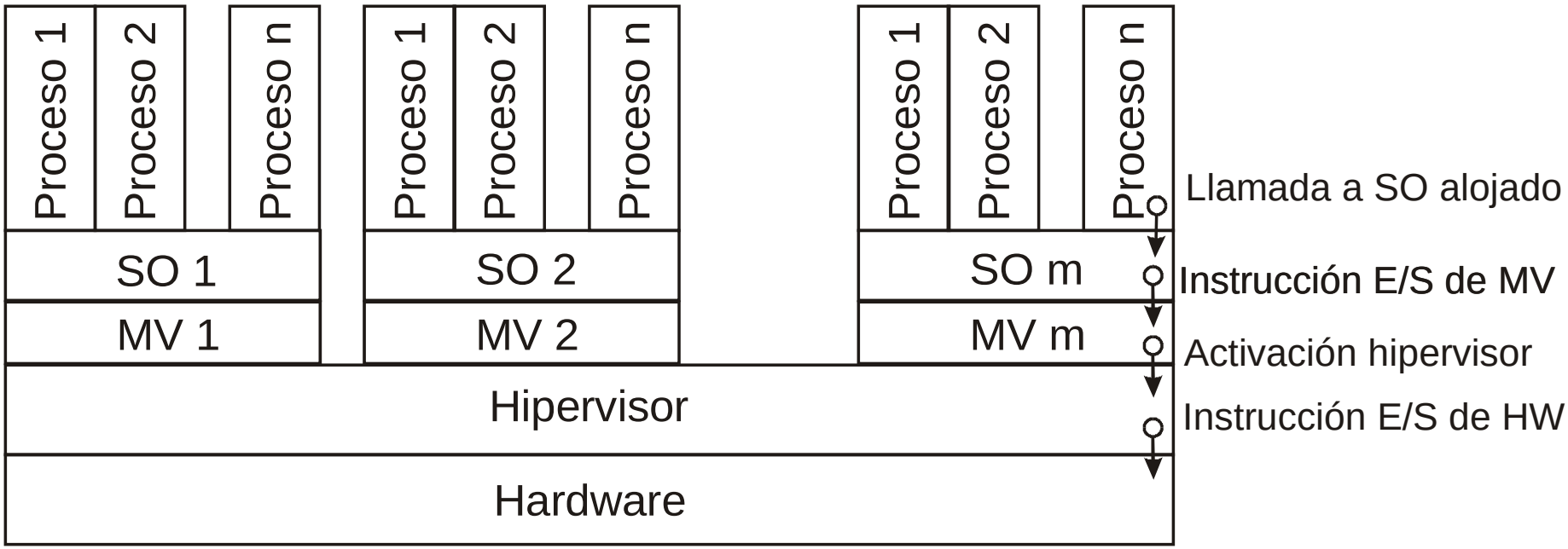
Virtualization

- Virtual machine: “An efficient isolated duplicate of a real machine”
  - Gerald J. Popek and Robert P. Goldberg (1974). *"Formal Requirements for Virtualizable Third Generation Architectures"*. Communications of the ACM 17 (7): 412 –421
  - Properties:
    - Equivalence: identical behavior
    - Resource control: complete control of the virtualized resources
    - Efficiency: A dominant fraction of machine instructions must be executed without VM intervention
- Origins: CP-40 (IBM 1967)
  - Current processing capacity relieves the inefficiency of VM
  - Some processors support virtualization
- Some important definitions:
  - Host machine: hardware that runs the virtual machine software
  - Host operating system: operating system that runs the virtual machine software
  - Hypervisor: software layer that provides the virtualization
  - Guest system: operating system

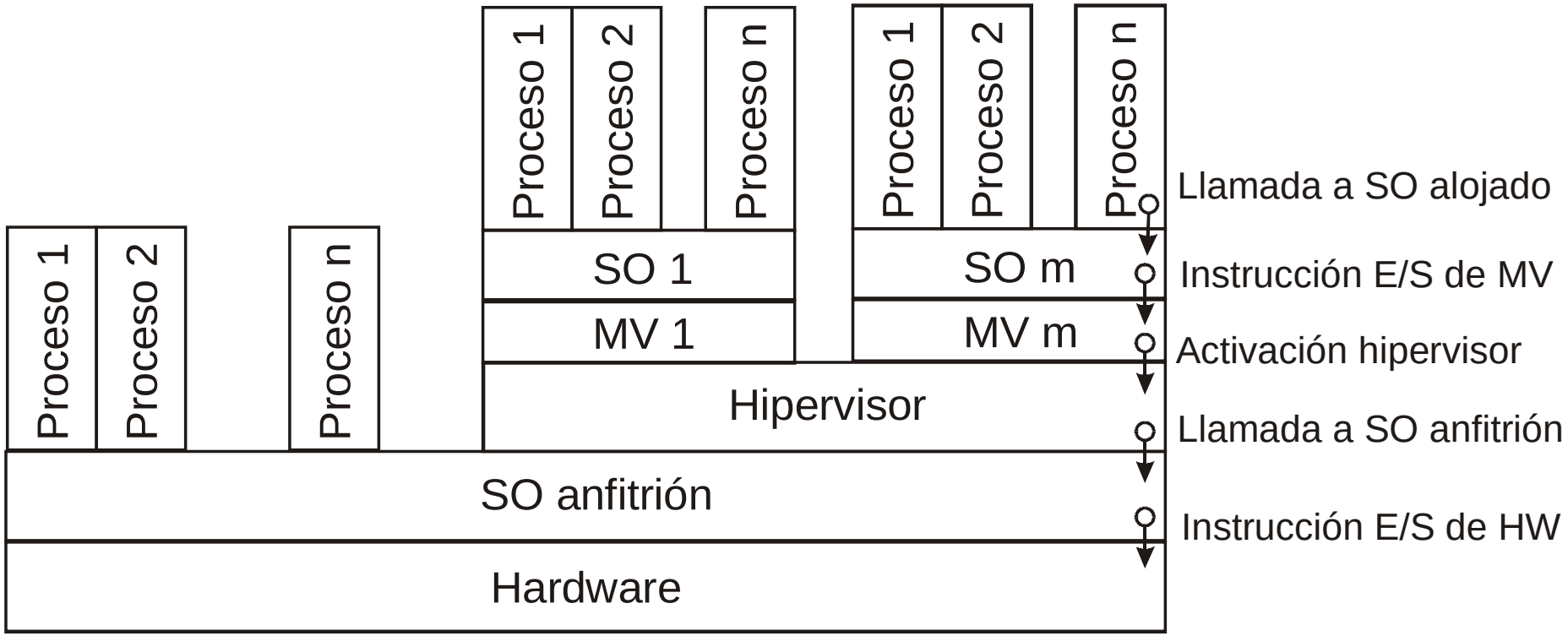
Virtualization

- Kinds of VM
  - Process VM: It supports a single process. It provides a platform-independent environment abstracting underlying hardware and/or operating systems
  - System VM: It supports the running of a OS.
    - Hypervisor enables the sharing of resources among several VMs
    - 2 kinds:
      - I (Native VM): Hypervisor running on HW
      - II (Non native VM): Hypervisor running on host OS

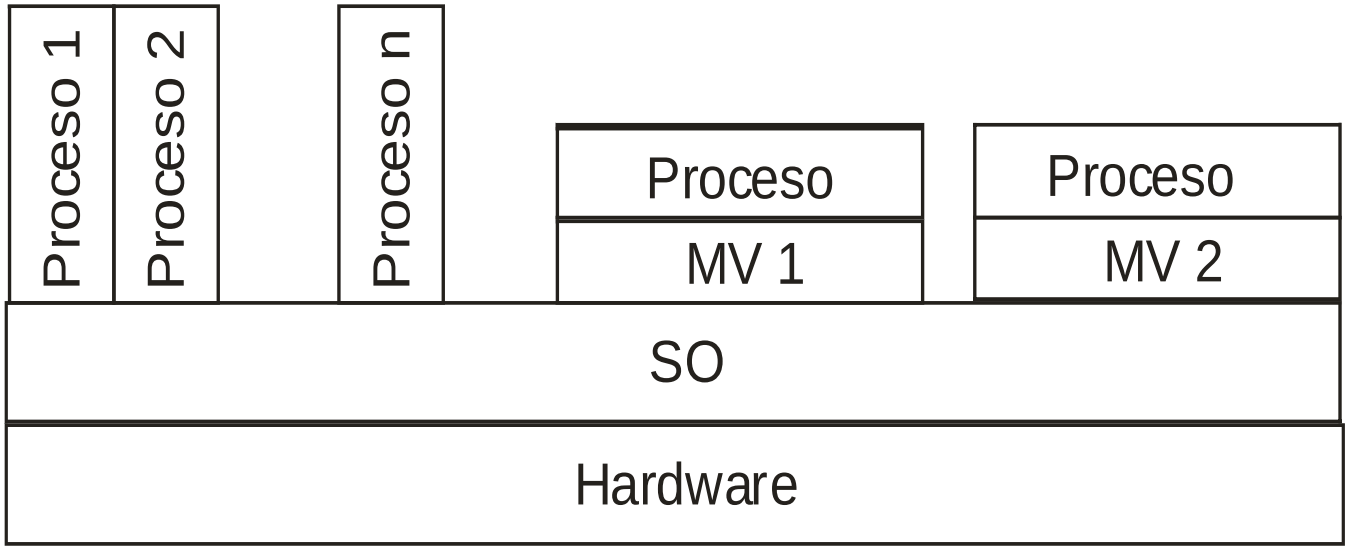
Native VM (Hypervisor architecture)



Non Native VM (Hosted architecture)



Process VM


Proceso 1
Proceso 2
Proceso n
Proceso
MV 1
Proceso
MV 2
SO
Hardware

Virtualization approaches

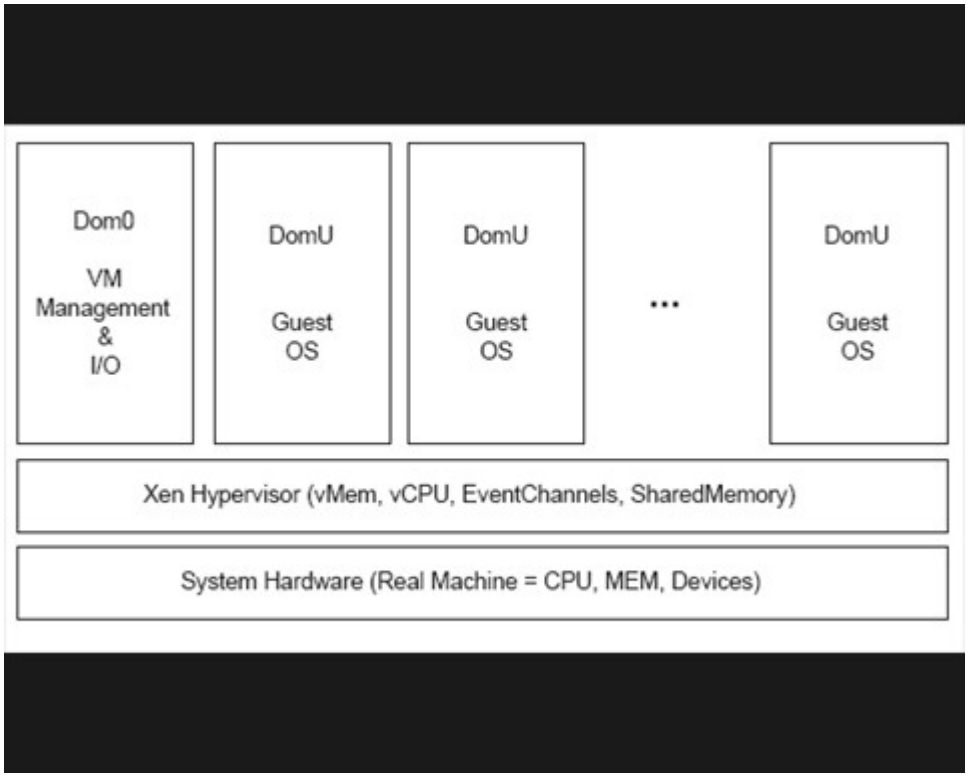
- Full virtualization
  - Unmodified guest operating system
  - Nondisruptive migration to virtualized environments
  - Example: Vmware, a combination of direct execution and binary translation techniques to achieve full virtualization of an x86 system
- Paravirtualization
  - Modified guest operating system
  - Advantages:
    - No need for binary translation
    - Potential performance advantages for specific workloads requiring modified operating system kernels
  - Problem: It is impossible to modify “closed” source operating systems (e.g., Microsoft Windows)
  - Example: Xen (open source)
- Hardware virtualization support
  - Virtualization extensions to the x86 architecture by Intel (Intel VT) and AMD (AMD V)
    - New processor instructions to assist virtualization software
    - First-generation hardware: CPU virtualization only
    - Later generations are expected to include memory and I/O virtualization as well
  - Multicore processors also promote the adoption of virtualization
  - This approach reduces the need to paravirtualize guest operating systems

Virtualization advantages

- Cost savings
- Operational efficiency
- Flexibility
- Coexistence of several OS
- OS debugging
- Run legacy systems
- Backup machines

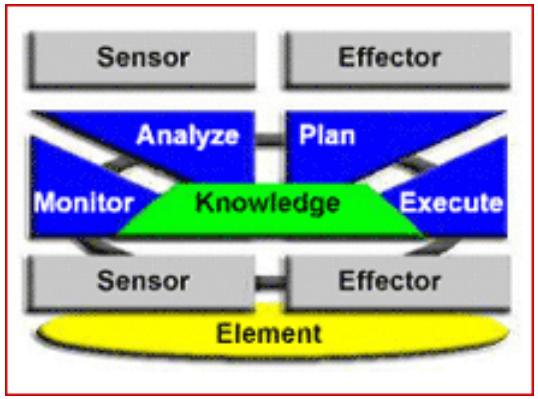
Virtualization

- Examples:
  - VMWare
  - Xen
  - Sun xVM
  - Microsoft Virtual PC
  - Microsoft Virtual Server
  - VM from IBM



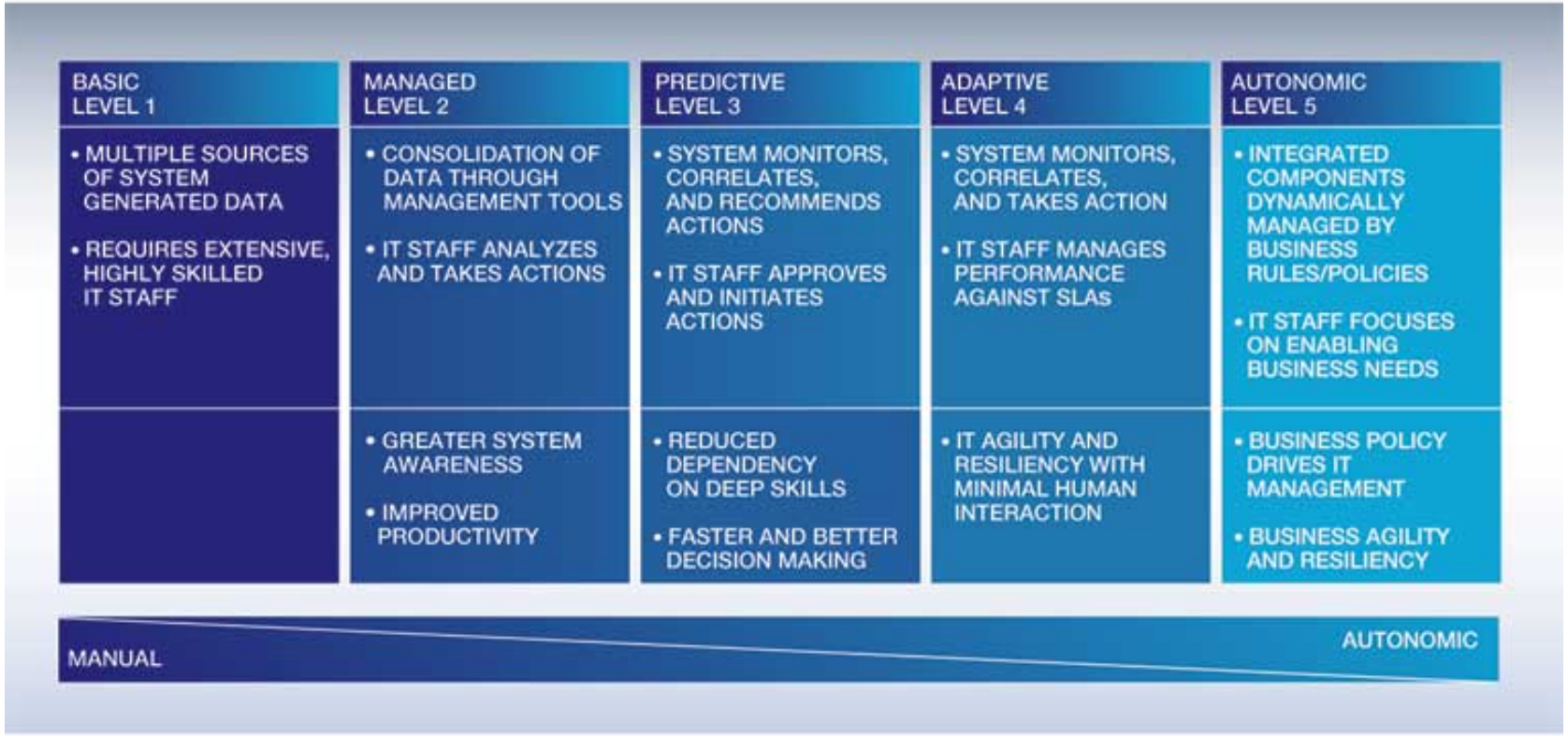
Autonomic Computing

- Decreasing the complexity of the environment in order to enhance its performance
- Based on biological systems, more specifically on the nervous system. Multiple unconscious tasks:
  - Check blood pressure
  - Adjust body temperature



Autonomic levels

Figure 2 Evolving to autonomic operations

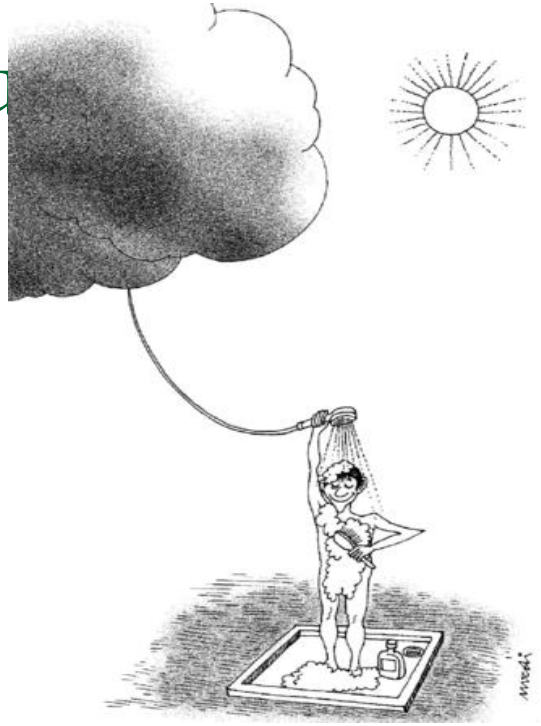


From *IBM Global Services and Autonomic Computing*, IBM White Paper, October 2002; see http://www-3.ibm.com/autonomic/pdfs/wp-igs-autonomic.pdf.

Autonomic features

- Self-Configuring: Automatic adaptation to dynamic environments
- Self-Healing: Discovering, diagnosing and reacting to failures according to specific policies
- Self-Optimizing: Monitoring resources and making decisions according to monitored data
- Self-Protecting: Detecting and identifying attacks against the system and acting in these situations

Advantages of cloud compu



- Lower computer and software costs
- Enhanced software updates
- Unlimited storage capacity

Disadvantages of cloud computing

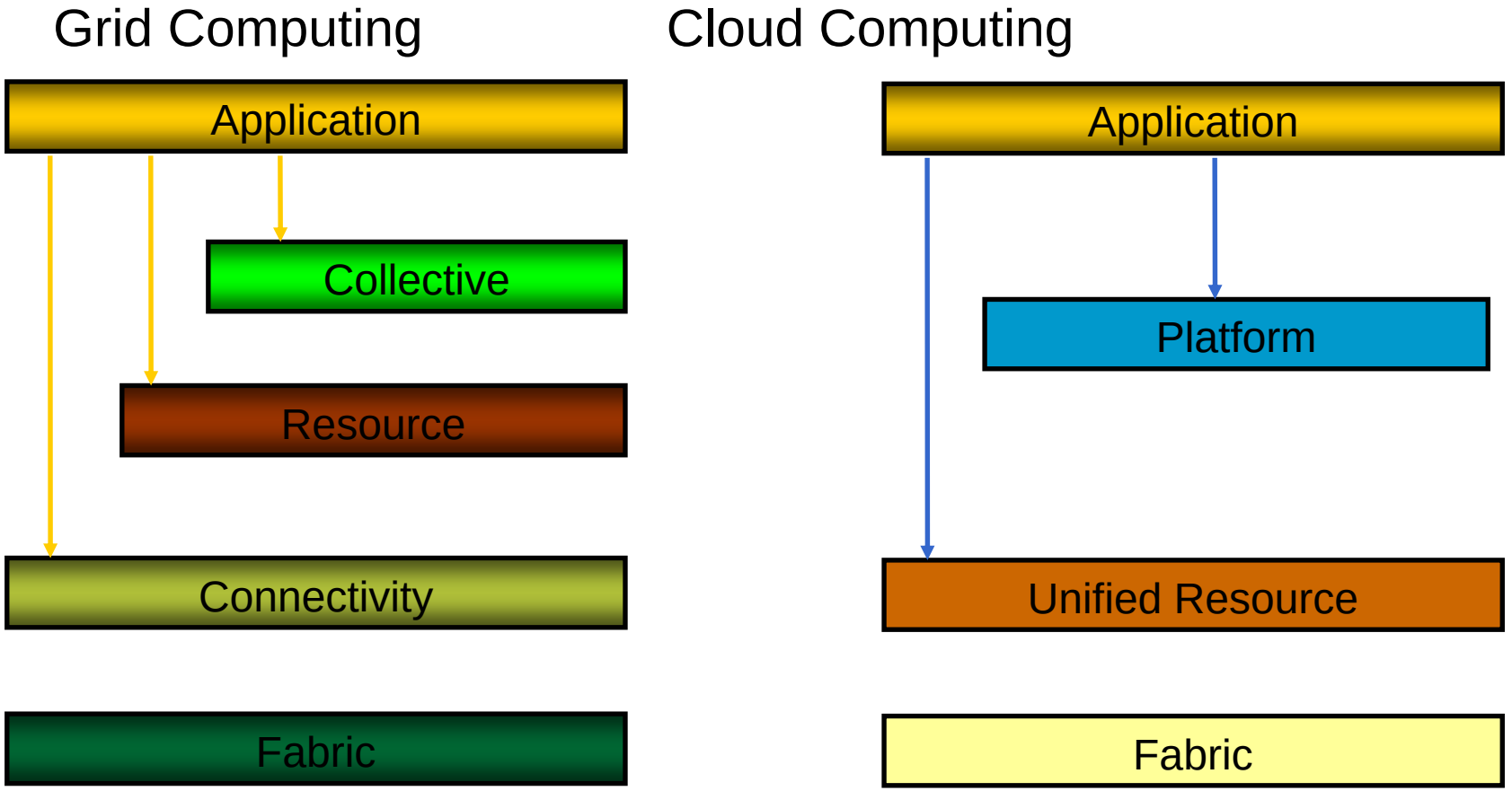
- Requires a high-speed internet connection
- Security and confiability of data
- Not solved yet the execution of HPC apps in cloud computing
- Interoperability between cloud based systems

Grid Computing vs Cloud Computing

- Business model:
  - Grid computing: project-oriented, in which it is possible to spend an amount of service units, generally CPU hours
    - Example: TeraGrid, proposals for the increasement of computational power
  - Cloud computing: customers pay providers on a consumption basis (such as electricity)
    - Example: EC2 from Amazon (instance-hour consumed), S3 from Amazon (GB-Month of storage)

Grid Computing vs Cloud Computing

- Architecture:



Grid Computing vs Cloud Computing

- Resource management:
  - Grid computing:
    - Batch-based computing model: Use of LRM (local resource managers), such as Condor, PGS or Sun Grid Engine.
    - Data model: location transparency, use of a distributed metadata catalog. Data storage usually depends on a shared file system (PVFS, Lustre).
    - Virtualization is not so important, although there are some initiatives
    - Widely use of Ganglia as monitoring system
  - Cloud computing:
    - Computing model: Resources in the cloud shared by all users. More number of users.
    - Data model: Google's MapReduce system running on top of the Google File system (Replicated chunks of data)
    - Virtualization is key in Cloud Computing
    - Difficult to obtain a high level of detail in monitoring. In the future, clouds will be self-maintained.

Grid Computing vs Cloud Computing

- Programming model:
  - Grid computing:
    - MPICH-G2
    - GridRPC
    - Workflow systems
    - WSRF
  - Cloud computing:
    - MapReduce model:
      - “Map”: Applying a specific operation to a set of items, obtaining a new set of items
      - “Reduce”: Aggregation on a set of items
      - Hadoop: Open source implementation of the MapReduce model
    - Scripting (Java Script, PHP, Python)

Grid Computing vs Cloud Computing

- Security model:
  - Grid computing:
    - GSI
  - Cloud computing:
    - Simpler model and less secure
    - Use of SSL and Web forms
    - A challenge not solved in clouds

Grid Computing vs Cloud Computing

- They share many goals
- They are different in many aspects
- But, they are complementary
  - *"…[Cloud computing] is indeed evolved out of Grid Computing and relies on Grid Computing as its backbone and infrastructure support."**

References

References

- “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”. I. Foster, C. Kesselman, S. Tuecke, International J. Supercomputer Applications, 15(3), 2001.
- “The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration”. I. Foster, C. Kesselman, J. Nick, S. Tuecke, Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002
- The Globus Project, http://www.globus.org

References

- Modeling Stateful Resources with Web Services, http://www.globus.org/wsrf
- The WS-Resource Framework, http://www.globus.org/wsrf
- [*] "Cloud Computing and Grid Computing 360-Degree Compared," I. Foster, Y. Zhao, I. Raicu, S. Lu, Grid Computing Environments Workshop, 2008. GCE '08 , vol., no., pp.1-10, 12-16 Nov. 2008
- “A history of cloud computing”, Arif Mohamed, March 2009

References

- "Understanding Full Virtualization, Paravirtualization, and Hardware Assist" Technical Paper of Vmware
  - http://www.vmware.com/files/pdf/VMware_paravirtualization.pdf
- Nimbus project:
  - http://www.nimbusproject.org
- OpenNebula.org:
  - http://ww.opennebula.org

Grid and Cloud Computing

María S. Pérez
Facultad de Informática
Universidad Politécnica de Madrid
mperez@fi.upm.es