

# Cloud Computing for science

Mariusz Witek



Institute of Nuclear Physics PAN, Kraków

# Plan

- Introduction
- Main features of cloud computing
- CC1 system – private cloud solution.
- Summary

# GRID – the global computer

WWW – revolution in access to information (invented at CERN).

- Uniform access to information around the world

GRID – uniform access to computing resources.

- GRID is seen as one big computer with enormous power
- Analogy to electric grid



- First successful GRID like project - SETI (*Search for Extra-Terrestrial Intelligence*)
  - Only small fraction of computing power of PCs are used.

After years of development, the computational GRID reached stability. WLCG (Worldwide LHC Computing Grid) is intensively used to store and process large samples of data - tens of PetaBytes from LHC experiments. But:

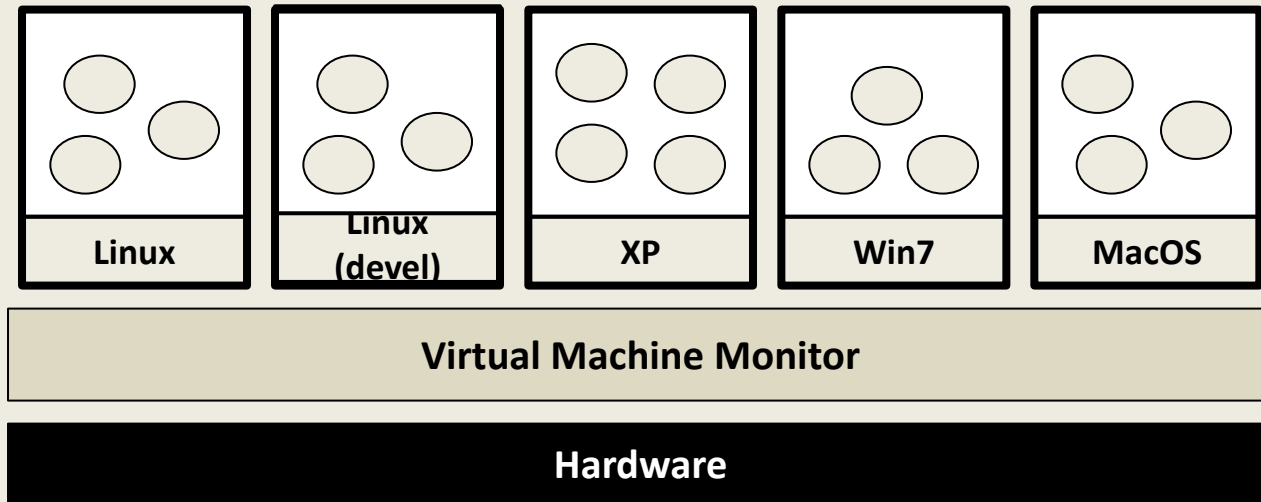
- Universality is limited – coexistence of various operating system is difficult. Dominated by Scientific Linux (RedHat based),
- Relatively heavy procedure to get the access to computing resources

The original idea of the computational grid came true when the technology of cloud computing appeared.

# Utility computing

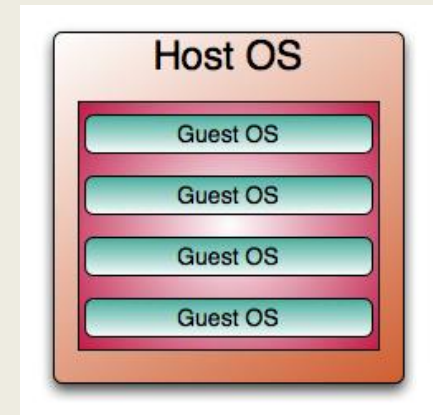
- Thanks to Cloud Computing, the long pending idea of Utility Computing has been realized.
  - Utility Computing - packaging of computing resources, such as computation and storage, as a metered service being sold.
- Example - start of small enterprise
  - No need to buy, install and maintain hardware. No financial risk if the initiative fails.
  - Computing resources available in a few minutes (servers running).
  - Easy scalable in case of success – buy additional resources (usually scalability is built into cloud services)

# Virtualization



- Virtual environment looks like real hardware to the guest operating system.
- Hardware virtualization is the key underlying technology. Virtualization of x86 architectures (Intel VT, AMD-V) was introduced in 2005.
- Popular VMMs: KVM, Xen, VMWare, VirtualBox

... an abstraction layer that allows multiple virtual computers, with heterogeneous operating systems to run in isolation, side-by-side on the same physical machine - VMWare

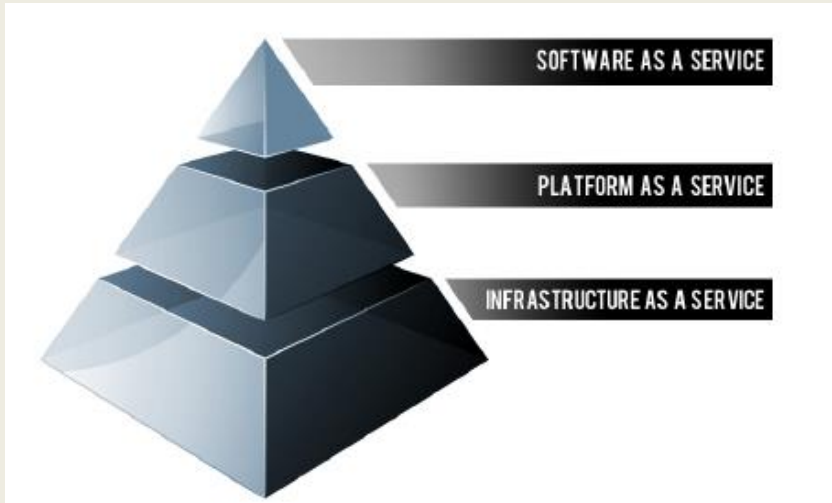


Important in multicore era

# Features of Cloud Computing

- Hardware point of view
  - The elimination of an up-front commitment by Cloud users
  - The ability to pay for use of computing resources on a short-term basis as needed
  - The illusion of infinite computing resources available on demand
- Software point of view
  - Access via self-service internet tools
  - Elasticity, scalability
  - Coexistence of various operating systems on the same hardware (efficiency gain)

# Types of Cloud Computing services



- **SaaS** - applications designed for end-users, delivered over the web
- **PaaS** - set of tools and services designed to make coding and deploying those applications quick and efficient
- **IaaS** is the hardware and software that powers it all – servers, storage, networks, operating systems

# Cloud Computing models

- **Public Cloud** – provided off-site over the Internet, available for everybody.
- **Private Cloud** – access restricted to a company, usually on private network
- **Hybrid Cloud** – a composition of private and public clouds. Typical example is private cloud with extension to public cloud (more resources needed or running not critical services from the security point of view).

	Public	Private
Start	No initial investment.	Purchase or use of existing infrastructure
Size	The illusion of having unlimited resources without the need to book in advance.	Limited resources but higher efficiency than in case of standard computing center. Stable operation requires the introduction of a quota system (in private cloud resources are available free of charge).
Economy	Cost optimization (pay per use). Scalability.	Scalability within available resources. Significant increase in efficiency.



# Commercial and OpenSource

- Commercial
  - Amazon Elastic Compute Cloud EC2 – 2006 (IaaS),
  - Google Apps (PaaS), Google Docs (SaaS) – 2009
  - Windows Azure
- Open source packages
  - Eucalyptus, Nimbus, OpenNebula, OpenStack, CC1
- OpenStack project (IaaS)
  - foundation consisting of > 200 companies.
  - Starts to be dominant OpenSource solution.
  - CERN based its Cloud Computing infrastructure on OpenStack.
    - Aim is 90% CERN IT capacity in the private cloud by 2015
    - Around 15,000 hypervisors, 150,000 virtual machines

# CC1 Project- Cracow Cloud One

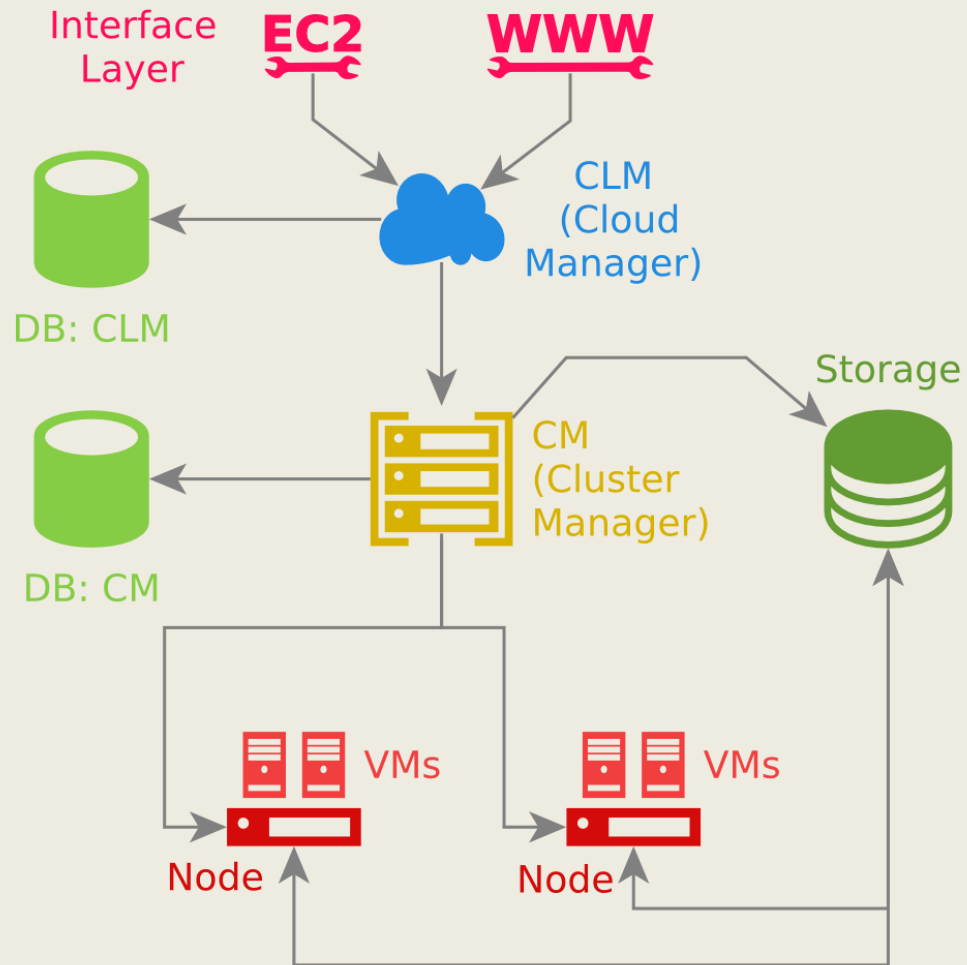
- **Project, nr POIG 02.03.03-00-033/09-04**
  - The “Cloud Computing for Science and Economy” project (acronym CC1) was started in 2009 at IFJ PAN. The project is financed from the structural resources within the grant POIG 02.03.03-00-033/09-04 funded by the European Commission and the Polish Ministry of Science and Education (Innovative Economy, National Cohesion Strategy).
  - The overall objective of the project is to reduce barriers to access the computer resources, for both scientific and commercial applications through the use of Cloud Computing technology.
- **The three phases of the project are**
  - development of a fully functional computing system in the form of Private Cloud and making it available to all users of the IFJ PAN. Stable IaaS system (Infrastructure as a Service) was put into operation at the beginning of 2012 and has been extensively used for the calculations by various research teams of the Institute.
  - implementation of a distributed architecture with a central management and sharing of computational resources of many distributed computer clusters.
  - dissemination of the benefits of cloud computing technology in the scientific community and among small and medium-sized businesses in the region. The installation procedure will be prepared and the source code repository will be made publicly available.



EUROPEAN UNION  
EUROPEAN REGIONAL  
DEVELOPMENT FUND



# CC1 structure



# CC1 system in IFJ PAN



## Blade IBM HS22

Xeon L5640 @ 2.27 GHz  
1000 cores,  
100 TB disk storage.  
Ethernet 10 Gb/s



## Network :

10 Gb/s external link  
10 Gb/s backbone of  
internal network

- Prototype working since 2011.
- Production quality system since beginning of 2012.
  - Intensively used by research groups.
  - Average CPU usage > 80%.
  - Useful feature- virtual farms

# CC1 features



- **Distributed structure** – the system consists of the central controller Cloud Manager and a number of Cluster Managers to control individual computing cluster.



- **Web interface** – provides intuitive, self-service user access to the system resources. The administration module contains a rich set of tools for user management and system configuration.



- **Groups of users** – self-organization mechanism allows users to create groups that share resources without the assistance of a system administrator.

# CC1 features



- **Virtual clusters** - the ability to create clusters of virtual machines based on a common image. Dedicated images allow the user to create a common environment with shared storage space and a ready-to-use batch system.



- **Virtual disks** – creation of permanent disks that can be dynamically connected to a virtual machine.



- **Elastic IP** - dynamical assignment of public IP addresses without the need to change the configuration of the VM operating system. IP addresses can be dynamically switched between running virtual machines.

# CC1 features



- **Sophisticated contextualization** - allows direct interaction with the virtual machine from the web interface. Encrypted communication protocol ensures security.



- **Quota system** – enforced limits on the number of processor cores, the total size of the disk space and the number of public IP addresses provide the stability of private cloud environment with limited resources.



- **Accounting** – makes it possible to measure the use of resources by individual users.

# CC1 features



- **Monitoring** - to view the load of virtual machines and to track the history of their use.



- **Administration module** – system management and configuration from the Web.



- **EC2 interface** - access to system resources with the popular Cloud Computing interface.



# CC1 installation – private cloud in one afternoon

- **Starting point:**
  - **Management server** – standard Debian.
    - All services inside: WI, CLM, CM, RM, DB (option to run on separate servers)
  - **N worker nodes** – standard Debian
  - **Disk array** – NFS exported
  - **Standard network** – switch, router.
- **Simple installation**
  - Very short installation instruction
    - Automatic procedure using standard Debian packages (.deb)

# Summary

- Utility computing became a reality with the advent of Cloud Computing technology (2006).
- Intuitive access to computing resources and self-service are two important features to be demonstrated on the example system – the CC1.