The Abdus Salam
International Centre
for Theoretical Physics

2494-25

Workshop on High Performance Computing (HPC) Architecture and Applications in the ICTP

14 - 25 October 2013

Which e-infrastructure for my computation?

Stefano Cozzini
CNR/IOM Democritos & eXact lab, Trieste

Which e-infrastructure for my computation ?

Stefano Cozzini

CNR/IOM Democritos and eXact lab - Trieste

Agenda

- Introduction:
  - Computing infrastructures and e-infrastructure
- What is available on the market ?
  - HPC/GRID/CLOUD
- Cloud computing
- Conclusion: Which infrastructure do you need ?

What is a computing infrastructure?

What is e-infrastructure ?

- E-infrastructure includes:
  - Networks (internet, light paths…)
  - Computers (workstations, servers, HPC…)
  - Access controls (security, AAA…)
  - Middleware (metadata…)
  - Finding tools (portals, search engines…)
  - Digital libraries (bibliographic, text, images, sound…)
  - Research data (national and scientific databases, individual data…)

infrastructure for computational science

- powerful and modern clusters of multicores.
  - hardware
  - software

High Performance Computing

- pooling of resources geographically distributed
- distribute collaborations

GRID COMPUTING

- Infrastructure as a Service

CLOUD COMPUTING

Elements of the computational environment

- powerful and modern parallel hardware (HPC)
- pooling of resources geographically distributed (GRIDs)
- Infrastructure as as service (CLOUD)

HARDWARE

- Scientific software: Models &analysis
- Software for distribute collaboration and data sharing

SOFTWARE

- IT- skilled computational scientists
- Strong partnership between IT people and scientists

BRAINWARE

HPC stands for:

- High Performance Computing
- The term is most commonly associated with computing used for scientific research.

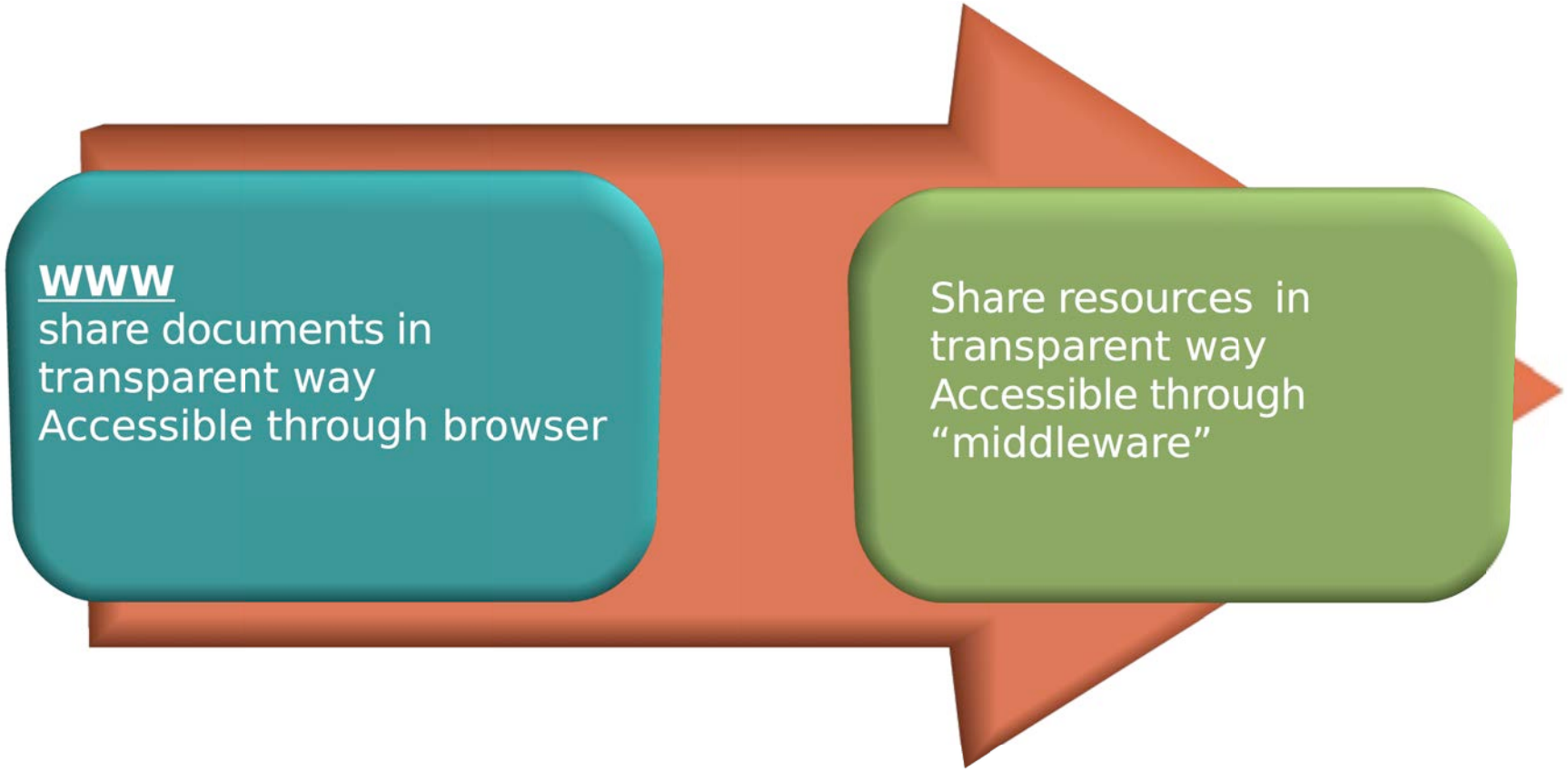
[wikipedia]

- it is not only on hardware but involves software and people as well
- Performance is not always what is matter..

To reflect a greater focus on the productivity, rather than just the performance, of large-scale computing systems, many believe that HPC should now stand for High Productivity Computing.

[wikipedia]

The GRID paradigm (2000)



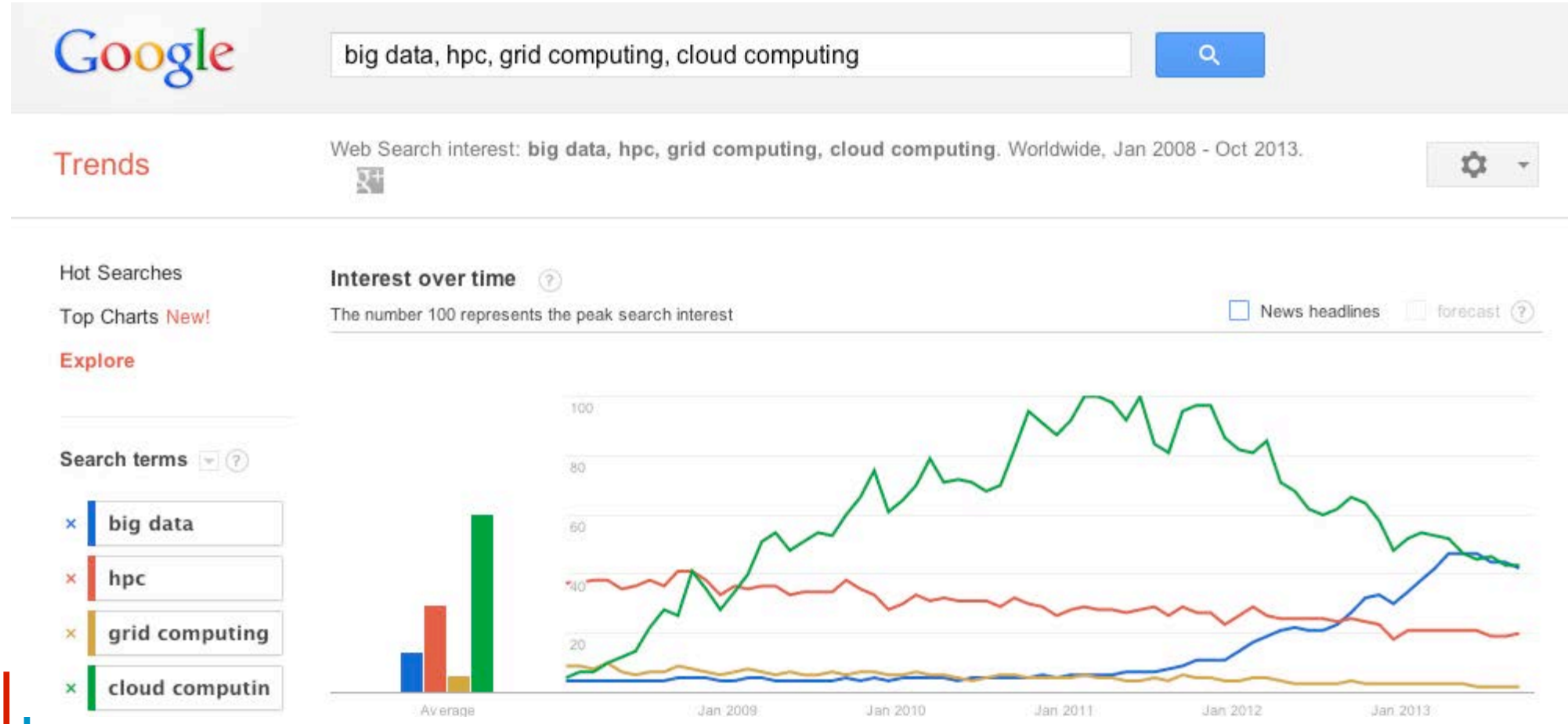
Why the GRID?

- Motivation: When communication is close to free we should not be restricted to local resources when solving problems.
- A Grid Infrastructure built on the Internet and the Web to enable and exploit large scale sharing of resources
- It should provides Scalable Secure Reliable mechanisms for discovery and for remote access of resources.

Grid Resource

- Storage systems
- Computer clusters
- HPC clusters
- Supercomputers (IBM SP, etc)
- Databases
- Keyword: heterogeneous as regards hardware and software

CLOUD: a buzzword..



Cloud Computing..

TECNOLOGIA

Ora l' Europa punta sul "clound computing"

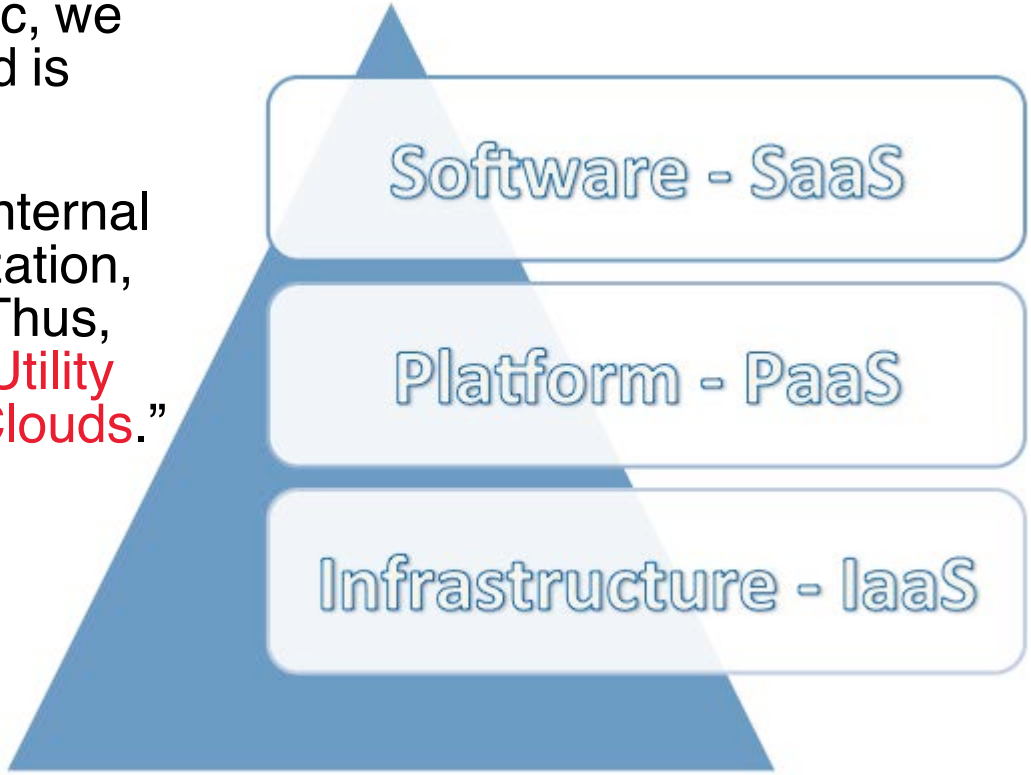
Ecco l'agenda digitale Ue:
con standard comuni, obblighi
chiari e compatibilità provider
sono previsti 2,5 milioni
di posti di lavoro entro il 2020

What is CC?

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services.

The services themselves have long been referred to as Software as a Service (SaaS). The data center hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing.

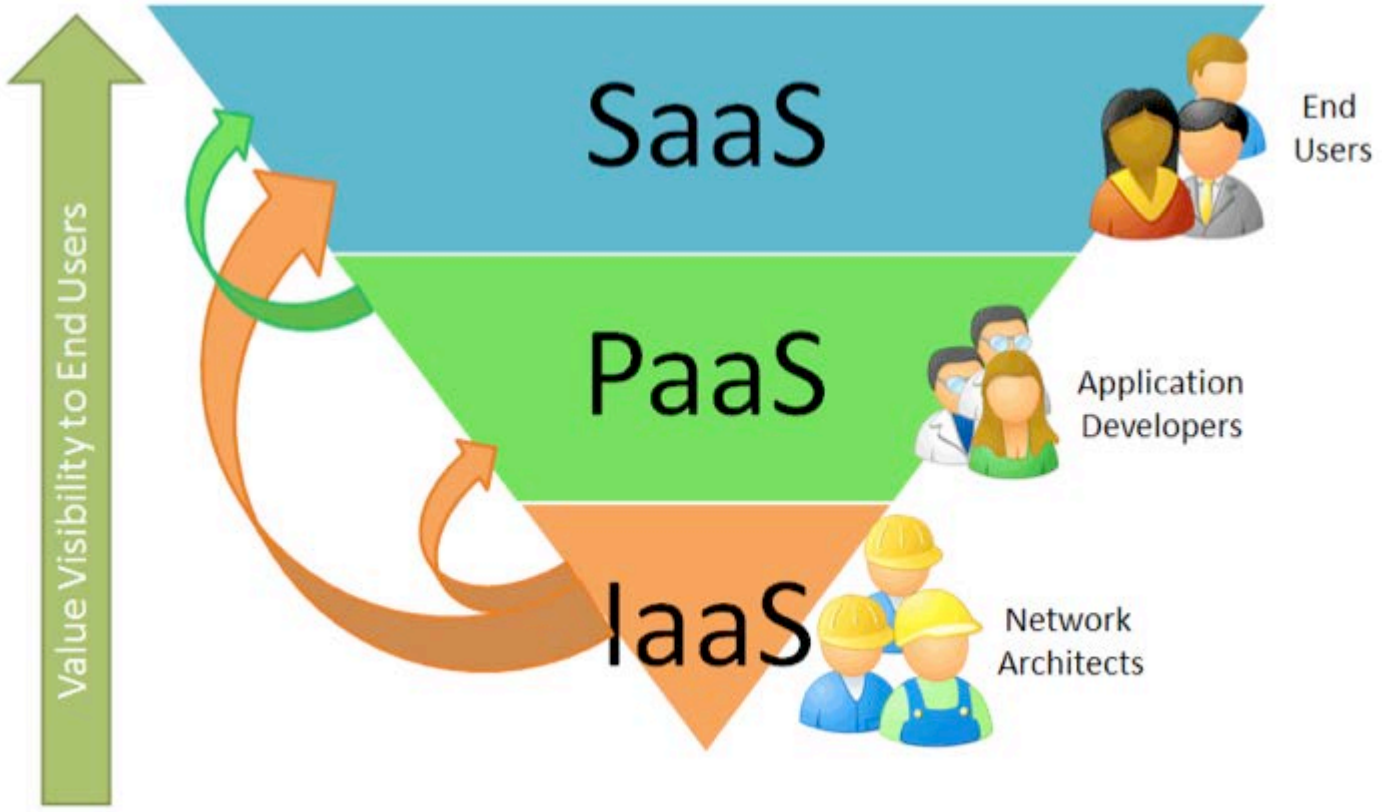
We use the term Private Cloud to refer to internal data centers of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds." (Berkeley RC)



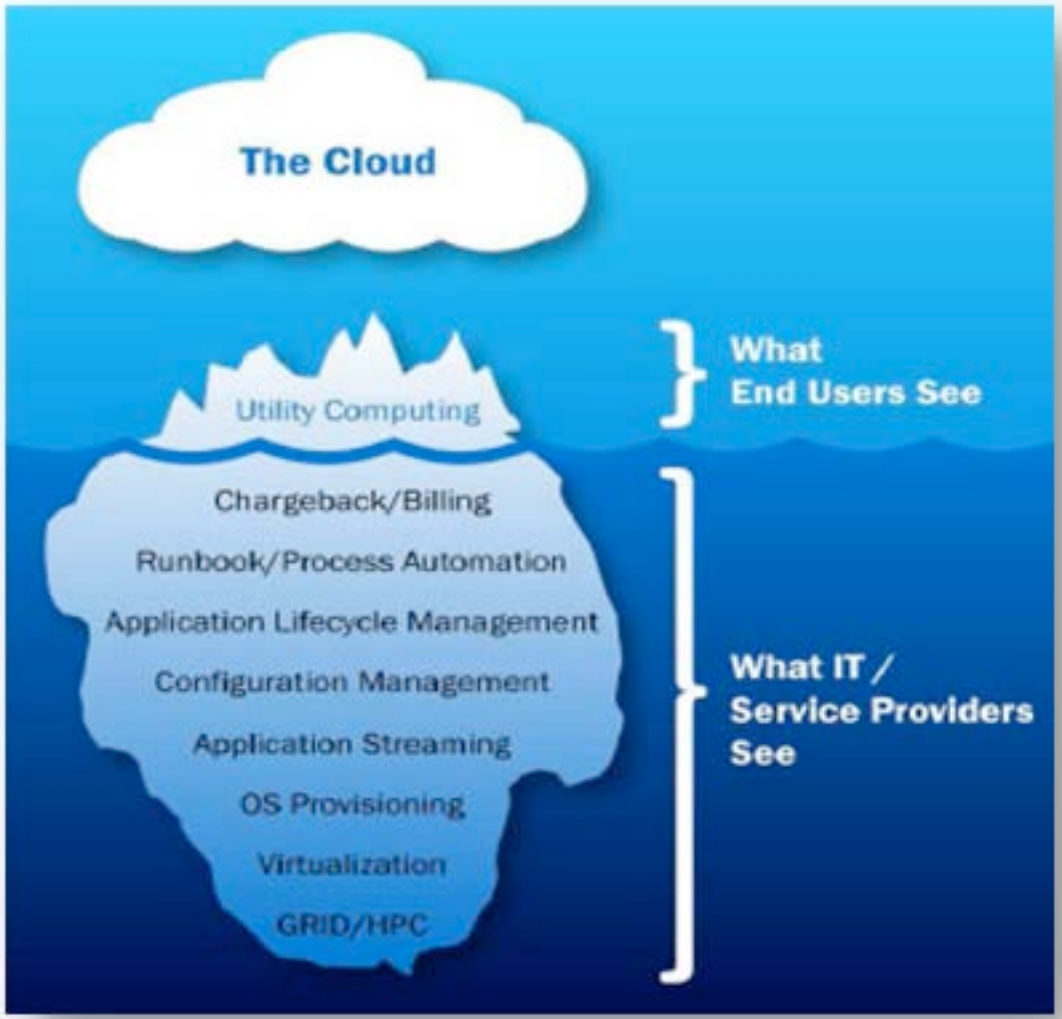
Basic Definition

- A model of computation and data storage based on “pay as you go” access to “unlimited” remote data center capabilities
- A cloud infrastructure provides a framework to manage scalable, reliable, on-demand access to applications
- Cloud services provide the “invisible” backend to many of our mobile applications
- High level of elasticity in consumption
- Historical roots in today’s Internet apps
  - Search, email, social networks
  - File storage (Live Mesh, Mobile Me, Flicker, …)

How far are from end users?



Why the CLOUD ?



How can I use a Cloud

- Amazon, Google, etc.
- Make your own one:
  - Eucalyptus (http://www.eucalyptus.com/)
  - OpenNebula (GLOBUS)
  - OpenStack (see later)
  - WS API, compatible with Amazon EC2 and S3

Building your own computational infrastructure

- Open source software + commodity off the shelf hardware provides now tools to build low cost HPC infrastructure
  - based on clusters
- GRID infrastructures are just outside..
  - they can provide a looot of resources
- CLOUD are outside: to give it a try..

Which computational infrastructure do you want ?

Elements of a computational infrastructure

- Hardware
  - The basic bricks
- Software
  - To make hardware usable
- People
  - installers/sys adm. /planners/ users etc..
- Problems to be solved
  - Any action in building such an infrastructure should be motivated by real needs

Lesson learned

- Role of human interaction is fundamental and sometimes underestimated
- CLOUD/GRID/HPC users speak different language from CLOUD/GRID/HPC providers
- E-infrastructure providers need to adopt their language NOT viceversa
- Technical tools are important BUT not sufficient to overcome the user inertia..

Our final message:

- Understand your computational problems before
  - buying/building a cluster !
  - accessing a grid !
  - Accessing a Cloud !