INFN Approach to Grid and Cloud Integration

**Valerio Venturi INFN-CNAF**

Workshop CCR e INFNGRID 2009

Palau 16 Maggio 2009

Outline

- Dynamic Virtual Environments

- Cloud computing

- Provisioning of resources from IaaS vendors

- Cloud computing vs. Grid computing

- Providing IaaS

Dynamic Virtual Environments

- In the previous presentation you saw a solution that allows Grid users to require that their jobs run on a specific virtual environment, i.e. on a virtual machines running a specified image

- Virtual machines are created on physical machine in the farm
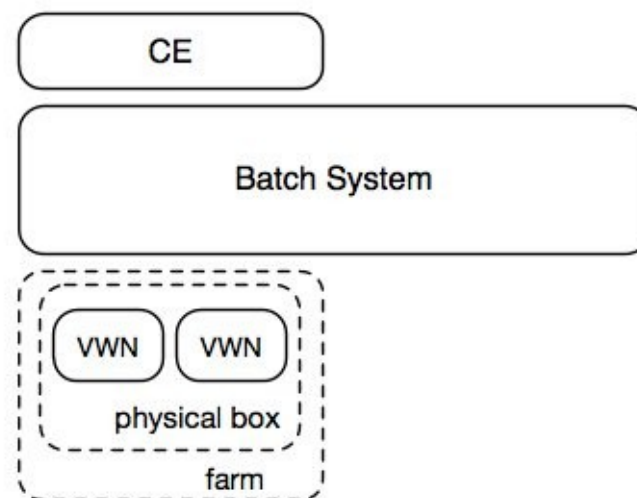
Dynamic Virtual Environments

- The user request to run on a virtual machine with a specific virtual image arrives to the batch system

- A virtual machine is created, and the job is sent there

- The details of how this currently works were presented in the virtualization session, and are not interesting here
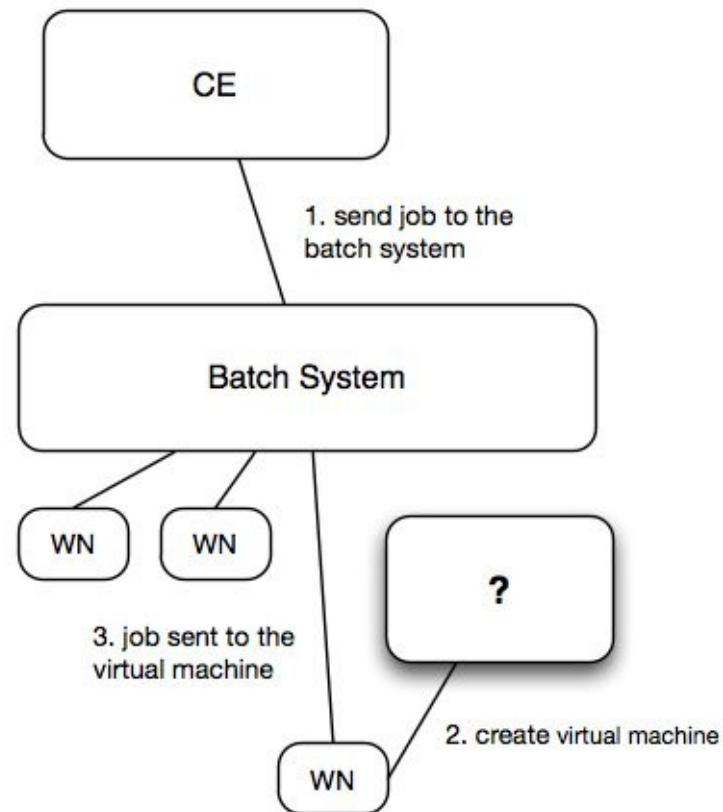
Dynamic Virtual Environments

- Creation of virtual machines running the request images is one of the central topic of cloud computing

- Natural to think about the possibility to create that virtual machine not from a phisycal machine at the farm but somewhere from a cloud vendor

CE

1. send job to the batch system

Batch System

WN   WN

3. job sent to the virtual machine

2. create virtual machine

WN
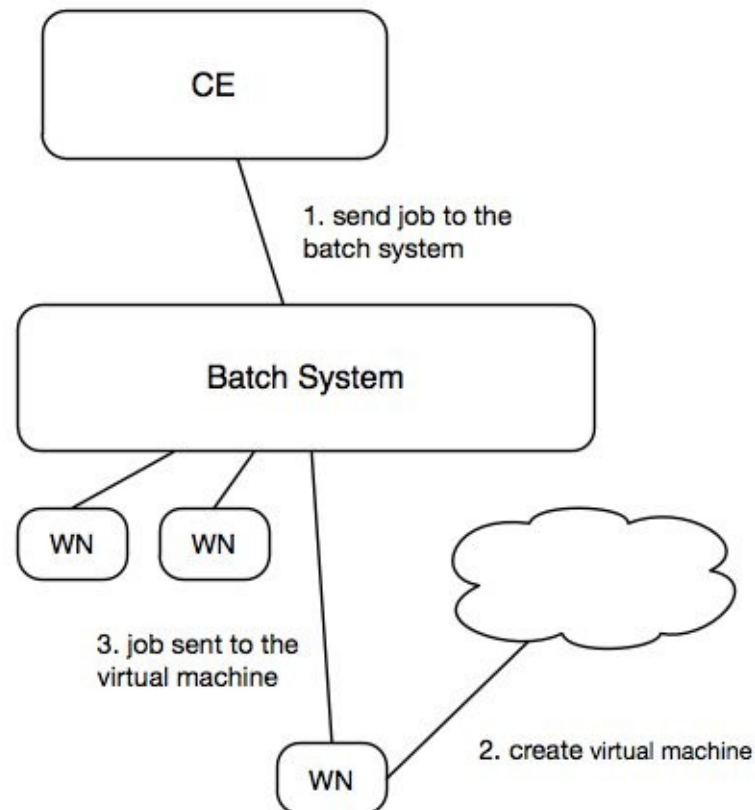
Cloud computing

- Cloud computing has emerged lately and is often presented as an alternative to Grid computing
  - Better alternative most of the time

- But what does that really mean?

  Computing paradigm that allows users to temporary utilize computing infrastructures over the network, supplied as a service at possibly one or more levels of abstraction
  Youseff, Butrico, Da Silva, Toward a Unified Cloud Computing Ontology, Grid Computing Environments Workshop, 2008

- Pretty general definition, let's see what's really out there

Cloud Computing

- Typically classified in three categories based on the level of abstraction of the services offered

- Software as a Service
  - Google Apps
  - Pretty far from our use cases, not even discussed here

- Platform as a Service
  - Google App Engine, Microsoft Azure

- Infrastructure as a Service
  - Amazon Web Services

Platform as a Service

- Provides programming language level environment for developers implementing their applications for and deploying them on the cloud

- Set of well defined APIs to interact with environment
    - Provides services like data storage, chaching, authentication, emails

- Interesting model? You may imagine a programming environment for scientific code, offering APIs for scientific computation

- The platform currently available are very different
    - Typically targeted to web appications development
    - Environment is very limiting: cannot fork processes, cannot use but few libraries
    - Limited portability, you develop for that platform

Infrastructure as a Service

- Provides basic infrastructure blocks to developers
    - Computational Amazon EC2
    - Data Amazon S3, Amazon SimpleDB

- In this presentation we will concentrate on
- The most important is Amazon EC2, but other are emerging GoGrid, Elastichosts

EC2

- The most famous example of computational IaaS

- Amazon Elastic Compute Cloud (EC2) is a Web service tha provides resizable compute capacity in the clouds

- Enables user to launch and manage virtual machine instances in Amazonàs data centers
    - Any number of instances, within minutes

- User has complete control over the instances
    - Firewall configuration, root access using RSA keys

Provisioning from IaaS

- Farm administrators may decide to buy computing power (virtual machines) from IaaS providers

- For temporary expanding the capacity of the farm in compute rush period

- If that will be in the future conveniente over traditional provisioning

- To differentiate the service offered
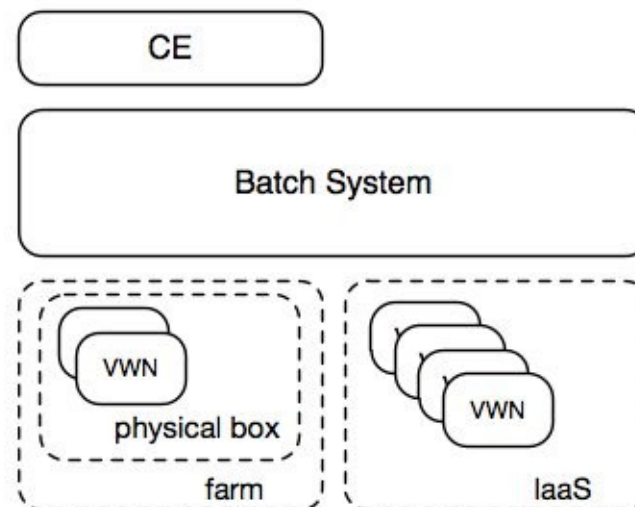
Provisioning from IaaS

- The alternative provisionig channel may be used for providing different better of services to users

- Jobs run only on machines in the farm

- Jobs run preferably on machines in the farm and on resources acquired from a IaaS provider when the desired level of service cannot be guaranteed

- Jobs run only on resources acquired from a IaaS provider

Some of the open issues

- **Economicity**. Cost analysis now not possible
  - Computing centers costs are difficult to estimate (housing, elettricity, machines..)
  - IaaS costs are going to change very quickly as only now other competitors than Amazon are entering the business
    - Amazon launched a new offer that decrease prices by 50%

# IaaS and Grid: another viewpoint

- So far we have seen how clouds and grids can coexist, and how cloud offers may be a chance for expanding the capacity of the Grid, by providing computing centers managers an agile provisioning of computing power for their farm

- This is not what we have heard in the last year

- What we have heard is that Grid was dead and Cloud would have solved all the problems

IaaS and Grid: another viewpoint

- Are IaaS and Grids competitors? According to the initial definition you would say so
  paradigm that allows users to temporary utilize computing infrastructures over the network

- But more practycally, see whatàs really available now

- Currently there's no cloud provider offering something similar to what grid computing is offering
  - Easy environment for execution of computational jobs

IaaS and Grid use cases

- IaaS and Grid use cases are different
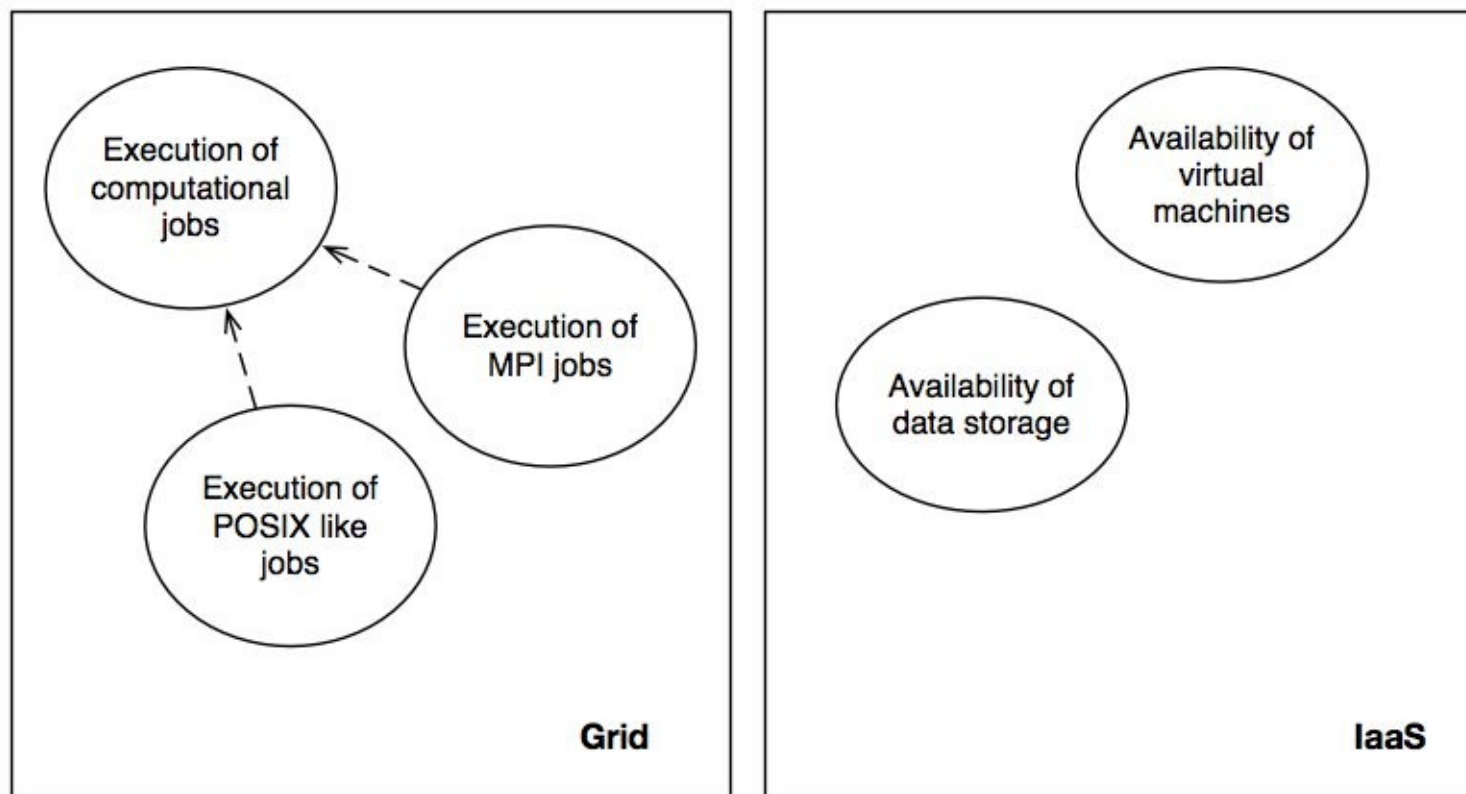
Providing IaaS services

- Is IaaS ever going to be INFN's business?

- Scientists who's use case is the execution of jobs will use Grid interfaces and services
  - We believe that our interfaces are a good answer to the scientists use cases, indeed they help shaping them

- There might be other use cases for INFN to provide IaaS services

Use cases

- **More efficient use of machines**. Research and development group at CNAF: 100 machines, used very seldom

- **Groups with special needs.** Is execution of jobs the use case for every scientists?

- **Easy deployment of service**. User won't have to ask administrators to run the VOMS server or add a CE, but would be able to instantly deploy thmeselves out of a prebuilt image

- **Provisioning of machines to externals**. Will ever an infrastructure built for scientist have idle power to resell and would that ever be convenient?

- Feedback on interest and practical use cases welcome.

IaaS provisioning: nice try

- The easy way to do that would be to allocate a set of machines for providing the service
  - Either developing a solution or using one of those available, like Eucalyptus or Globus Workspace

- Moving machines away from the farm is not an option

- At the end that we would like to go in the other direction, moving machines that are rarely used to the farm

IaaS provisioning: possible solution

- The batch system can assure the fair balance between the job execution use and the virtual machine provisioning use

- The batch system is used to schedule instances, thus maintaining the sharing

- No dedicated resources are kept away from job execution

- Consistent sharing, if you get a virtual machine running, you submit less jobs
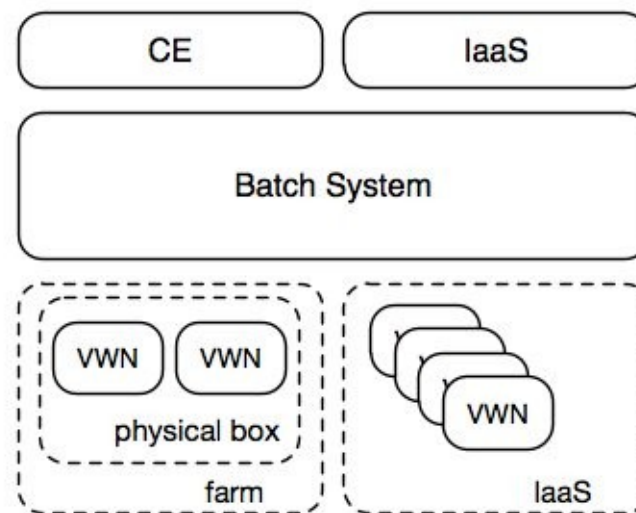
Open issues

- **Queuing**. Users donàt want to wait hours to get their virtual machines running. Currently not an issues given the current time jobs stay queued at T1.

Conclusions
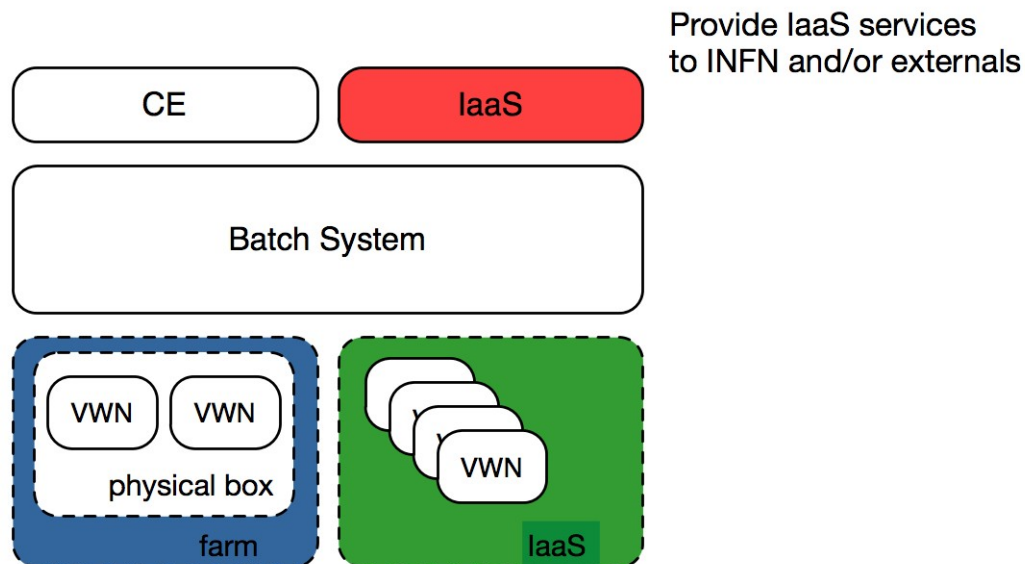
Provide IaaS services to INFN and/or externals

Users run jobs in virtual machines running a specified image

Expanding capacity of computing centers through provisioning from IaaS vendors

Other open issues

- Interfaces
  - EC2 the most known, and used also by other providers
  - OGF started the Open Cloud Computing Interface (OCCI) working group with the goal of delivering an API specification iimntially targeting specifically the lifecycle of virtual machines (http://www.occi-wg.org)

- Integration with Data IaaS

Akwnoledgements

- This talk is the result of a series of meetings done during the last months, that involved among the others but probably not only
  - CNAF: M.Cecchi, A.Ghiselli, A.Italiano, D.Salomoni, V.Venturi, E. Ronchieri, R. Zappi, S. Andreozzi, L. Magnoni, V.Ciaschini
  - Milano: D.Rebatto
  - Padova: M.Sgaravatto, L.Zangrando

- Further contributions are definitely welcome