Nimbus: Introduction

by various contributors
(see last slide)

Request for contributors

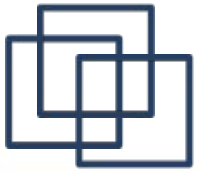
Hi and thanks for your contribution!

If you have modified something, I'm kindly asking you to
- put your name in the last slide
- write a quick note the next slide concerning what you have modified/added/deleted (since googledocs does not provide this feature, it only says who has modified the presentation and how many slides have been modified)

Massimo Canonico (mex@di.unipmn.it)

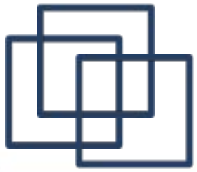
Revision history

01/12/11 - M. Canonico - sent the draft to cloud ML/forum to get feedback
18/11/11 - M. Canonico - first draft
17/11/11 - M. Canonico - uploaded on GD



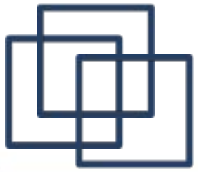
ToDO list

- Improve/Fix the layout
- Are there more information to put in this introduction?



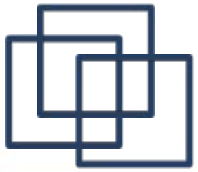
Outline

- What is Nimbus?
- Architecture
- Components
- Science Cloud: goals
- Projects and Cloud Computing
- Cloud Computing and projects
- Nimbus on FutureGrid



What is Nimbus?

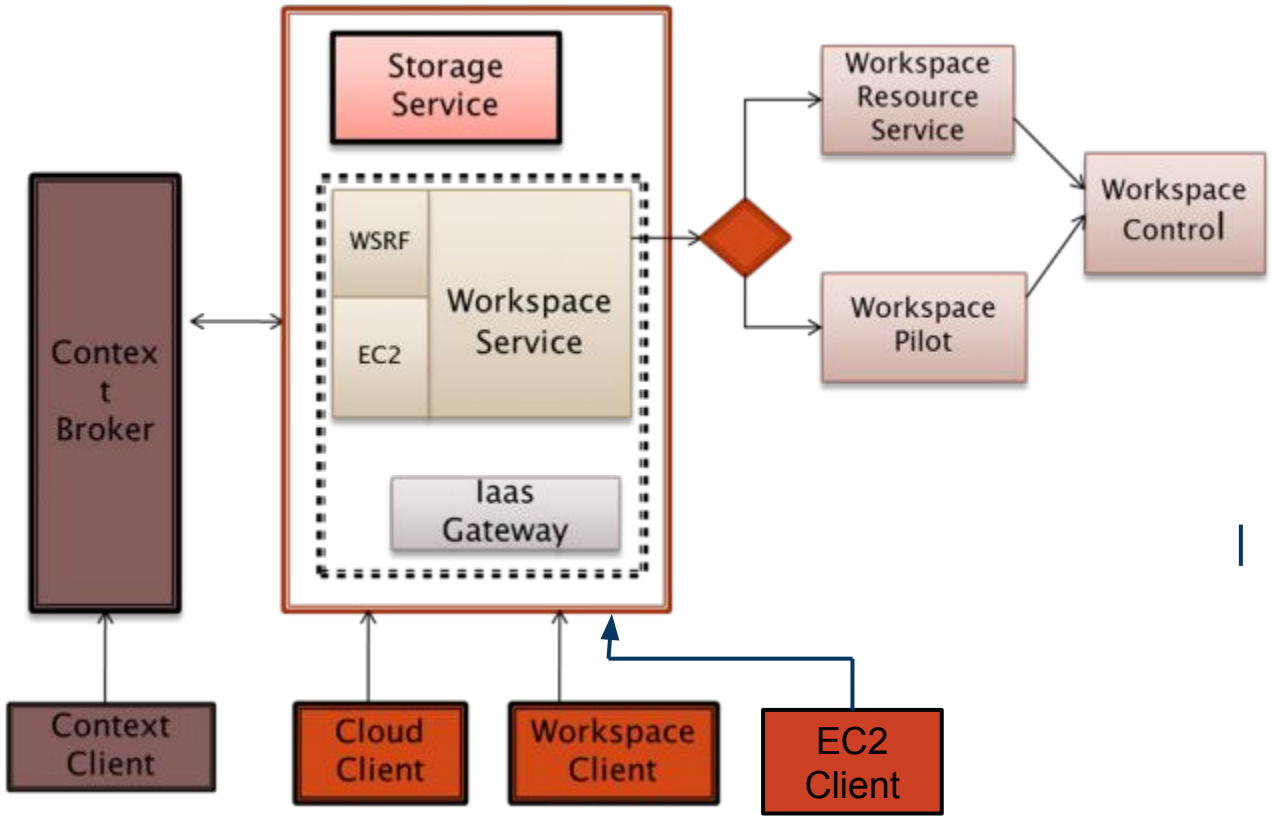
- Set of open source tools that together provide an "Infrastructure-as-a-Service" (IaaS)
- The programming framework used is Java and Python. The virtualization technologies supported are XEN and KVM.

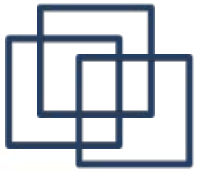


What is Nimbus (II)?

- An extensible open source Infrastructure-as-a-Service implementation
  - Turns your cluster into a cloud
- Why open source IaaS?
  - **Experiment and use:** make your own cloud or configure a private cloud
  - **Customize:** try new things, make the IaaS paradigm work for your application domain
- Particular interest in customization: scientific computing

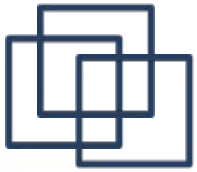
Architecture





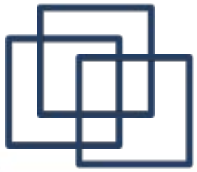
Components

- The **Workspace Service** is a standalone site VM manager that different remote protocol frontends can invoke.
  - A **Web Services Resource Framework (WSRF)** based remote protocol implementation
  - An **EC2 based remote protocol** implementation of their SOAP and Query APIs (partial) that supports EC2 Clients
- **Cumulus** is an open source implementation of the Amazon S3 REST API. It is used as the Nimbus repository solution and can also be installed standalone.
- The **cloudclient** aims to get users up and running in minutes with instance launches and one-click clusters.



Components (II)

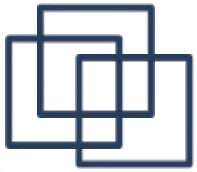
- The **WorkspacePilot** allows you to integrate VMs with resources already configured to manage jobs (i.e., already using a batch scheduler like PBS).
- **Resource manager** is used for a pool of physical nodes. It deploys and manages Workspaces on the nodes
- The **workspace-control** agent implements VMM and network specific tasks on each hypervisor.
- The **ContextBroker** allows clients to coordinate large virtual cluster launches automatically and repeatably.
- The **ContextAgent** lives on VMs and interacts with the Context Broker at VM boot.



Science Clouds: Goals

- Make it easy for scientific projects to experiment with cloud computing
  - Can cloud computing be used for science?

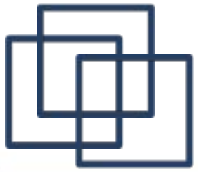
- Customize software in response to the needs of scientific projects
  - Start with EC2-like functionality and evolve to serve scientific projects: virtual clusters, diverse resource leases
  - Federating clouds: moving between cloud resources in academic and commercial space



Projects and Cloud Computing

- CloudBLAST: bioinformatics applications
  - CS research: investigate latency-sensitive apps, e.g. hadoop
  - Need access to distributed resources, and **high level of privilege** to run a ViNE router
  - Virtual workspace: ViNE router + application VMs

- STAR: a high-energy physics experiment
  - Needs resources with the **right configuration**
    - Complex environments: **correct versions** of operating systems, libraries, tools, etc all have to be installed.
    - **Consistent environments**: require validation



Nimbus on FutureGrid

- Hotel (University of Chicago) -- Xen
  - 41 nodes, 328 cores
- Foxtrot (University of Florida) -- Xen
  - 26 nodes, 208 cores
- Sierra (SDSC) -- Xen
  - 18 nodes, 144 cores
- Alamo (TACC) -- KVM
  - 15 nodes, 120 cores