Virtual Machines on BiG Grid

BiG Grid
the dutch e-science grid

INFN Annual Meeting

May 2010

Sander Klous, Nikhef

Contents

- Introduction
  - Who are we? What do we do?
- Motivation
  - Why, What and How with VMs?
- Status
- Challenges
- Conclusions

Introduction

- Collaboration between
  - NCF: national computing facilities
  - Nikhef: national institute for subatomic physics
  - NBIC: national bioinformatics center
- Participation from Philips, SARA, etc.

Goal:

"Enable access to grid infrastructures for scientific research in the Netherlands"

https://wiki.nbic.nl/index.php/BigGrid_virtualisatie

BiG Grid VM Working Group

- Representatives from:
  - User Communities/Support, Operations and Security, Management, “Independent Chair”
  - Recipe for disaster
- Charge for phase 1:
  - ***“Provide a design and Proof-of-Concept for virtualized worker nodes that fulfill requirements of different user communities.”***
- Preliminary result:
  - Extensive report on all aspects that need to be addressed before BiG Grid accepts VMs.

Motivation: Why Virtual Machines?

- Site perspective
  - Resource flexibility (e.g. SL4 / SL5)
  - Resource management
    - Scheduling / multi-core / sandboxing
- User perspective
  - Isolation from environment
    - Identical environment on multiple sites
    - Identical environment on local machine

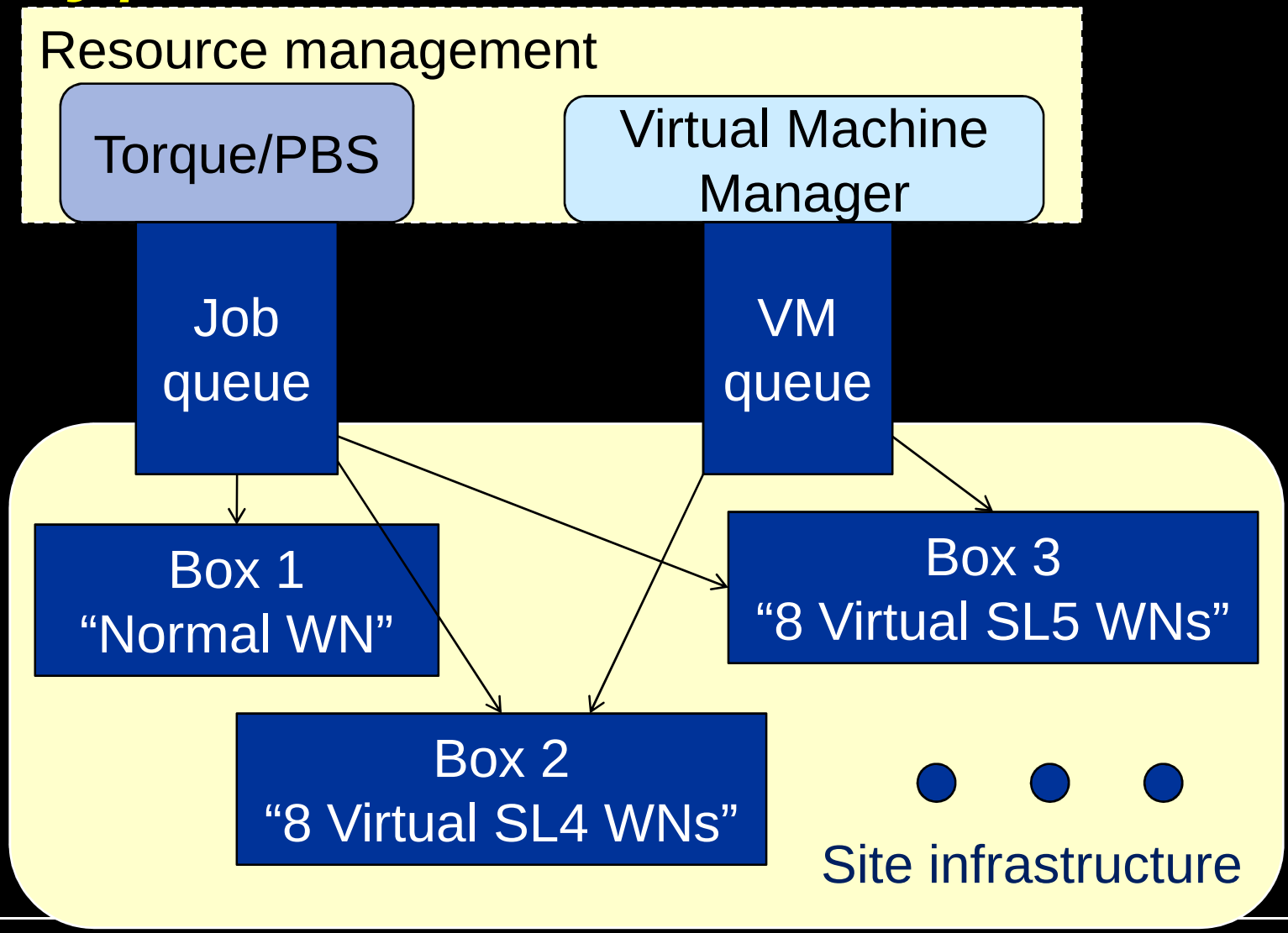
Different VM classes

- Class 1: Site generated Virtual Machines
  - No additional trust issues
  - Benefits for system administration
- Class 2: Certified Virtual Machines
  - Inspection and certification to establish trust
  - Requirements for monitoring / integration
- Class 3: User generated Virtual Machines
  - No trust relation
  - Requires appropriate security measures

Different VM Classes

STATUS

Typical use case Class 1 VM

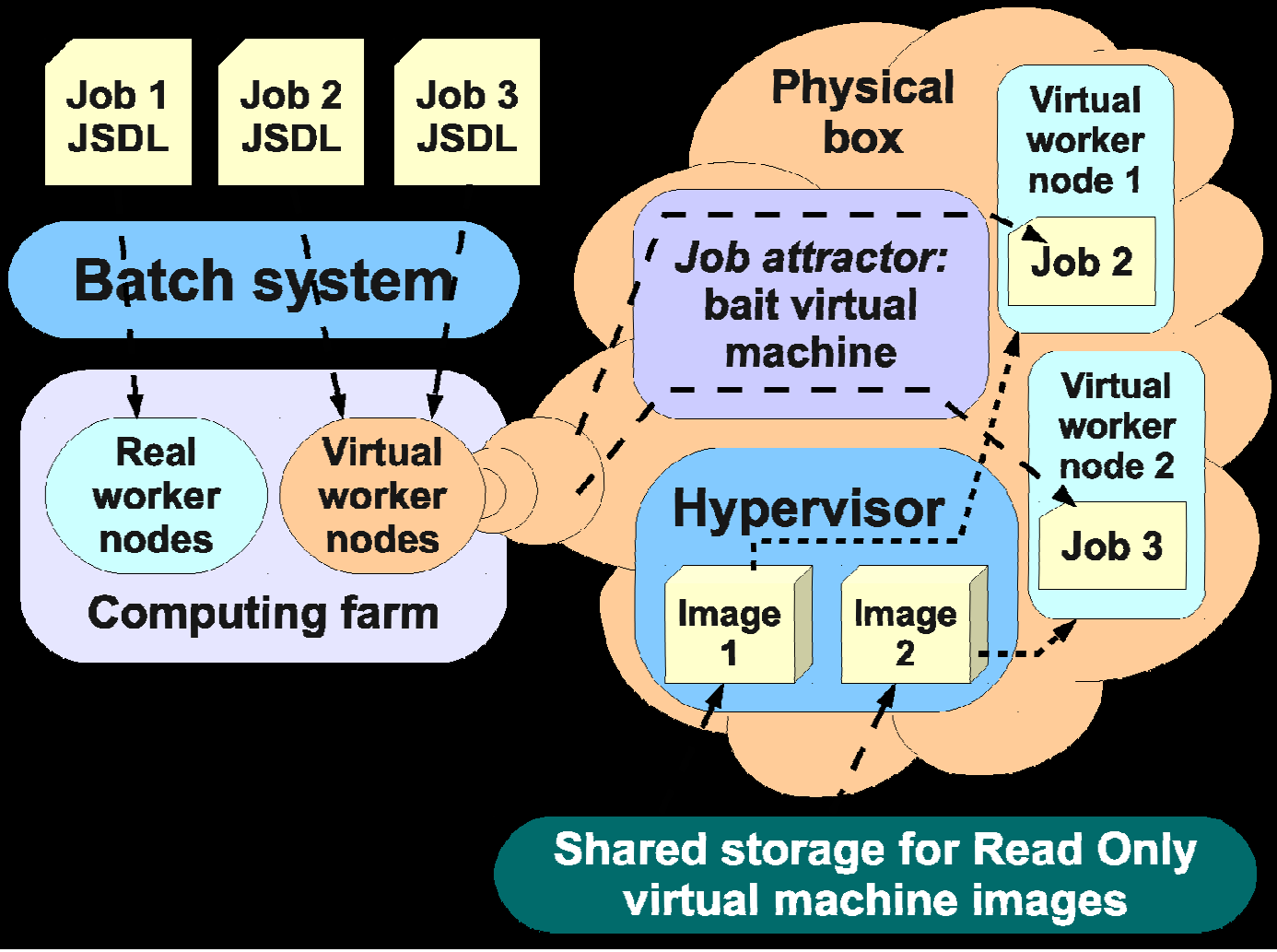


Typical use case Class 2 VM

Analysis on Virtual Machines

- Run minimal analysis on desktop/laptop
  - Access to grid services
- Run full analysis on the grid
  - Identical environment
  - Identical access to grid services
- No interest to become system administrator
  - Standard experiment software is sufficient

Between Class 1 and 2/3: WNoDeS



HEPiX VM working group

- Chaired by Tony Cass (CERN)
- Representatives from site operations
- Charge:

  "Enable virtual machine images created at one site to be used at other HEPiX (and WLGC) sites."
- Preliminary results:
  - Draft security policy
  - VM Image Catalog

    (straw man model and interface definition)

Draft Security Policy

http://www.jspg.org/wiki/Policy_Trusted_Virtual_Machines

"The aim is to enable Grid Sites to trust and instantiate endorsed VM images that have been generated elsewhere."

- Main points:
  - VM combined image split in base image (root install) , VO environment (user install)
  - Complete image is endorsed by an individual to confirm that it respects the policy requirements
  - Extendible with a site local policy

Policy Trusted Virtual Machines

Draft policy document - not yet approved or adopted.

This is Draft Version 1.4 (11 May 2010). Includes issues discussed at the working group meeting of 11 May.

Policy on the Endorsement of Virtual Machine Images

Table of contents [hide]

Introduction

This document describes the security-related policy requirements for the generation and endorsement of trusted virtual machine (VM) use on the Grid.

The aim is to enable Grid Sites to trust and instantiate endorsed VM images that have been generated elsewhere.

The virtualisation model addressed here is the use of virtual Grid worker nodes that act in a similar way to real worker nodes. Virtualis provides an efficient way of managing different configurations of worker node, e.g. the operating system used, and importantly different configured application environments for the VOs. The model addressed here, therefore, simply provides a different way of running auth work, transparent to the end user, exactly the same as if the user payload was running on a real worker node. There should be no nee more restrictions on virtual worker nodes running endorsed images as defined by this policy, than on real worker nodes in terms of acc trusted local services at the site.

This policy does not compel Sites to instantiate images endorsed in accordance with this policy nor limit the rights of a Site to decide instantiate a VM image generated by any other non-compliant procedures, should they so desire. The Site is still bound by all applica security policies and is required to consider the security implications of such an action on other Grid participants.

Definitions

The following terms are defined.

- **VM base image**: A VM image, including a complete operating system and all general middleware, libraries, compilers, programm utilities. All kernel and root-level configurations, including any that may be VO-specific, are included here.
- **VO environment**: The VO-specific middleware, application software, libraries, utilities, data and configuration which may be neces provide the appropriate environment for use by members of a VO. No kernel modifications or root-level configurations are included h
- **VM complete image**: The VM image resulting from the combination of the VM base image and the VO environment (if any).

BiG Grid Policy Addendum

- VO environment endorsed by VO person
- Complete image endorsed by non-VO person
  - Nikhef endorsers will defer responsibilities and liabilities for VO environment to VO person.
  - Written and signed deferral of all responsibilities and liabilities of VO environment, including:
    - Consequences of possible security leaks
    - Consequences of possible license violations

Other applicable policies

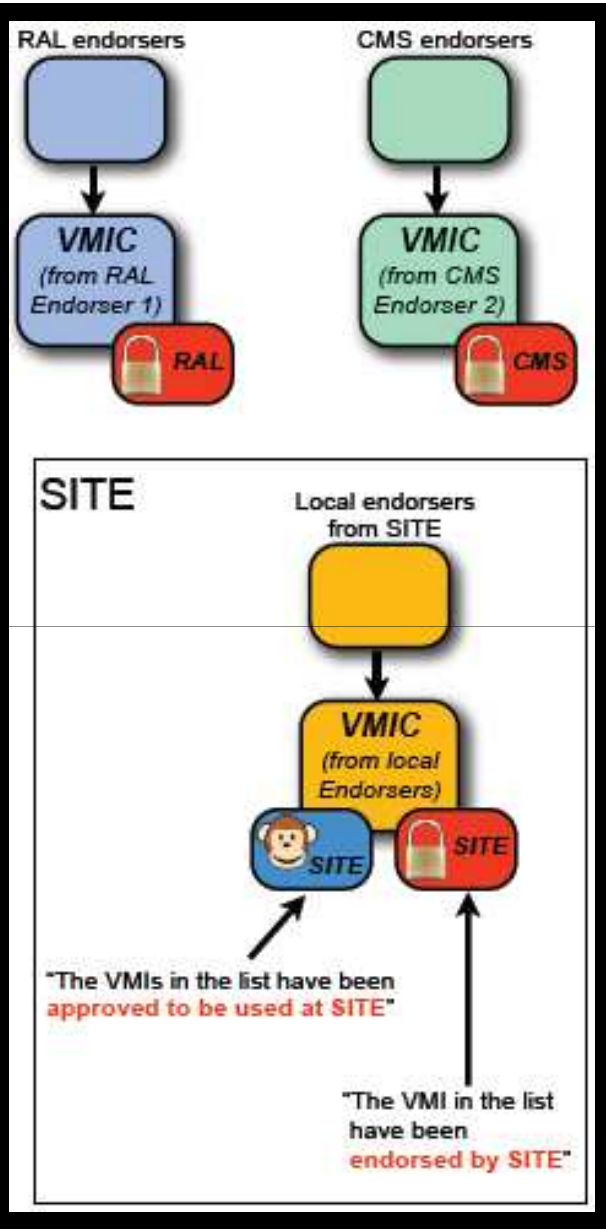
- *Grid Security Policy*, version 5.7a
- *VO Portal Policy*, version 1.0 (draft)
- *Big Grid Security Policy*, version 2009-025
  - *Grid Acceptable Use Policy*, version 3.1
  - *Grid Site Operations Policy*, version 1.4a
  - *LCG/EGEE Incident Handling and Response Guide*, version 2.1
  - *Grid Security Traceability and Logging Policy*, version 2.0
- *VO-Box Security Recommendations and Questionnaire*, version 0.6 (draft, not ratified)

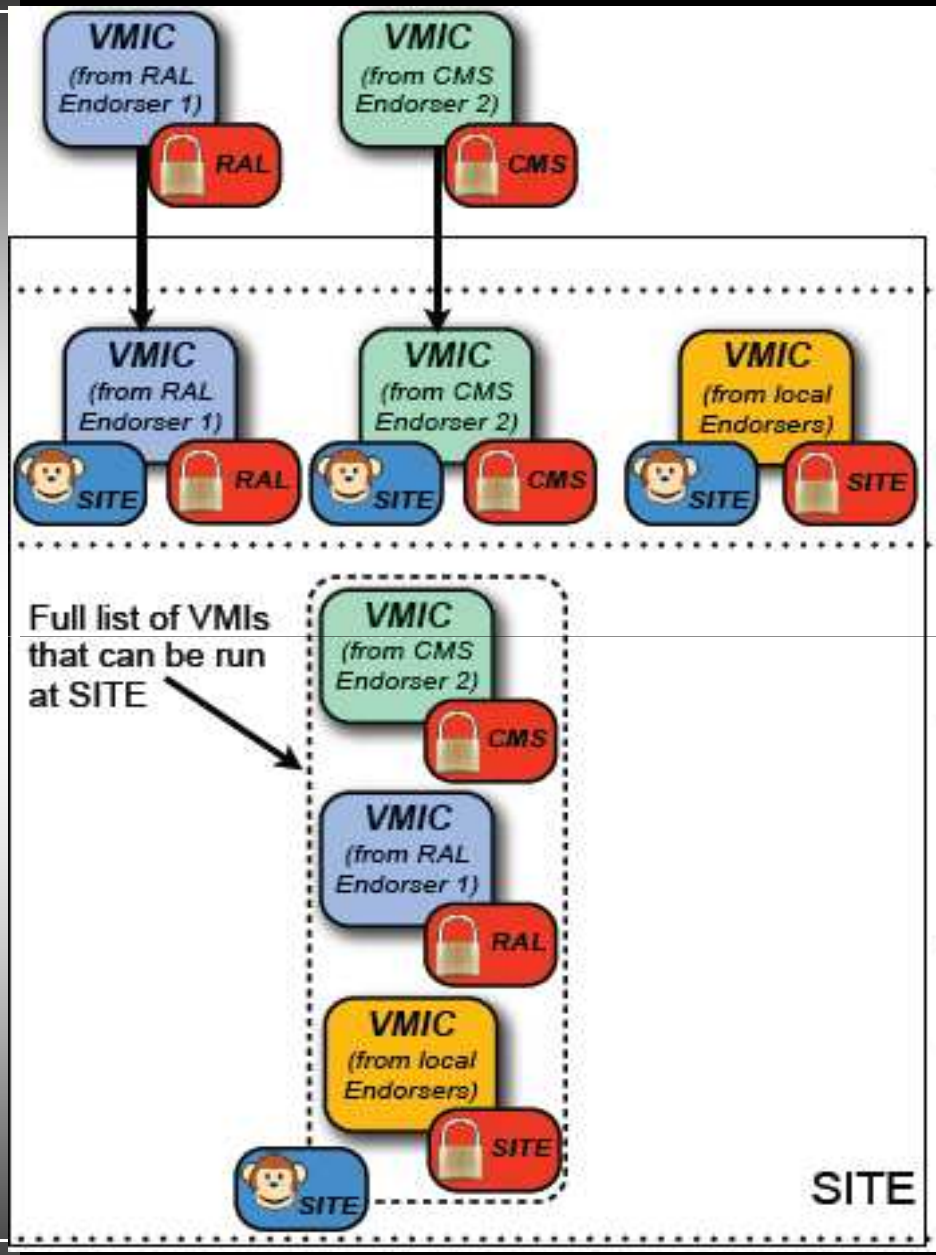
Relevant policy statements

- Network security is covered by site local security policies and practices
- A VO Box is part of the trusted network fabric. Privileged access is limited to resource administrators
- Software deployed in the grid must include sufficient and relevant site central logging.

VM Image Catalog

- Endorsed means:
  - Image created according to HEPiX policy requirements
  - Endorser responsible/liable
- Approved means:
  - Site will run this image
    - Acceptable endorser
    - Acceptable Meta Data





1. SITE decides to approve VMIs endorsed by (XYZ) RAL and CMS

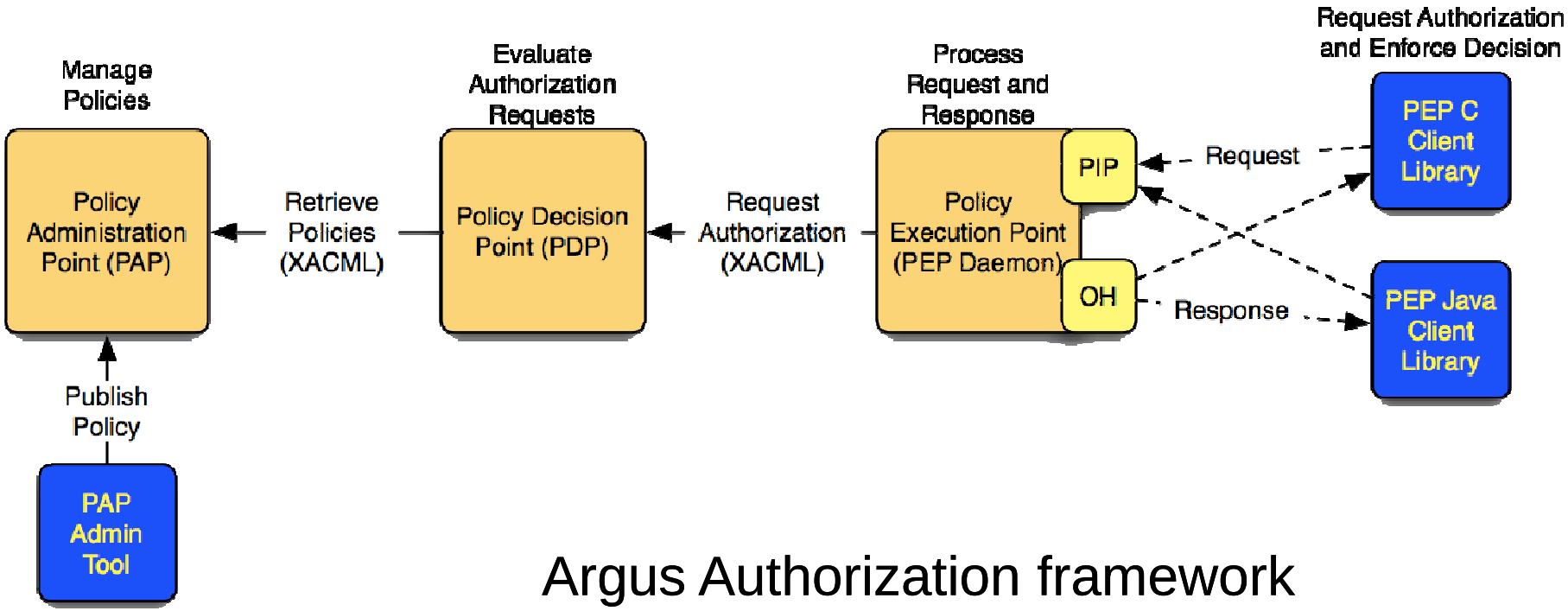
2. VMIs are approved(SITE)
Sites has fine-grained control over VMIs being approved (but can also approve them all)

3. The RAL and CMS VMIs are added to complement the VMIs produced locally

4. The resulting list of VMIs (endorsed by different entities) is approved by the local site

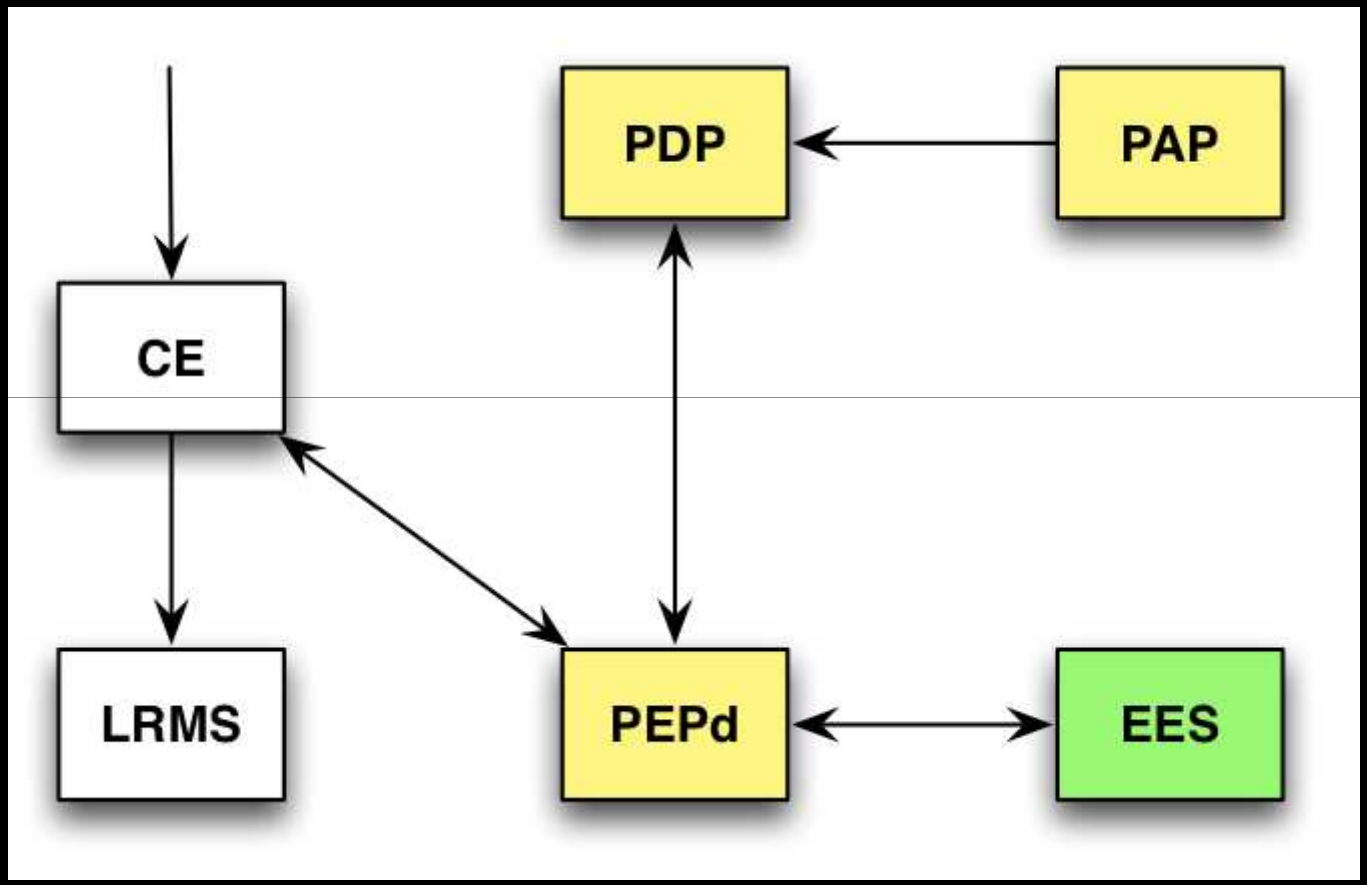
XYZ + SITE + SITE

How does this fit in the gLite middleware?



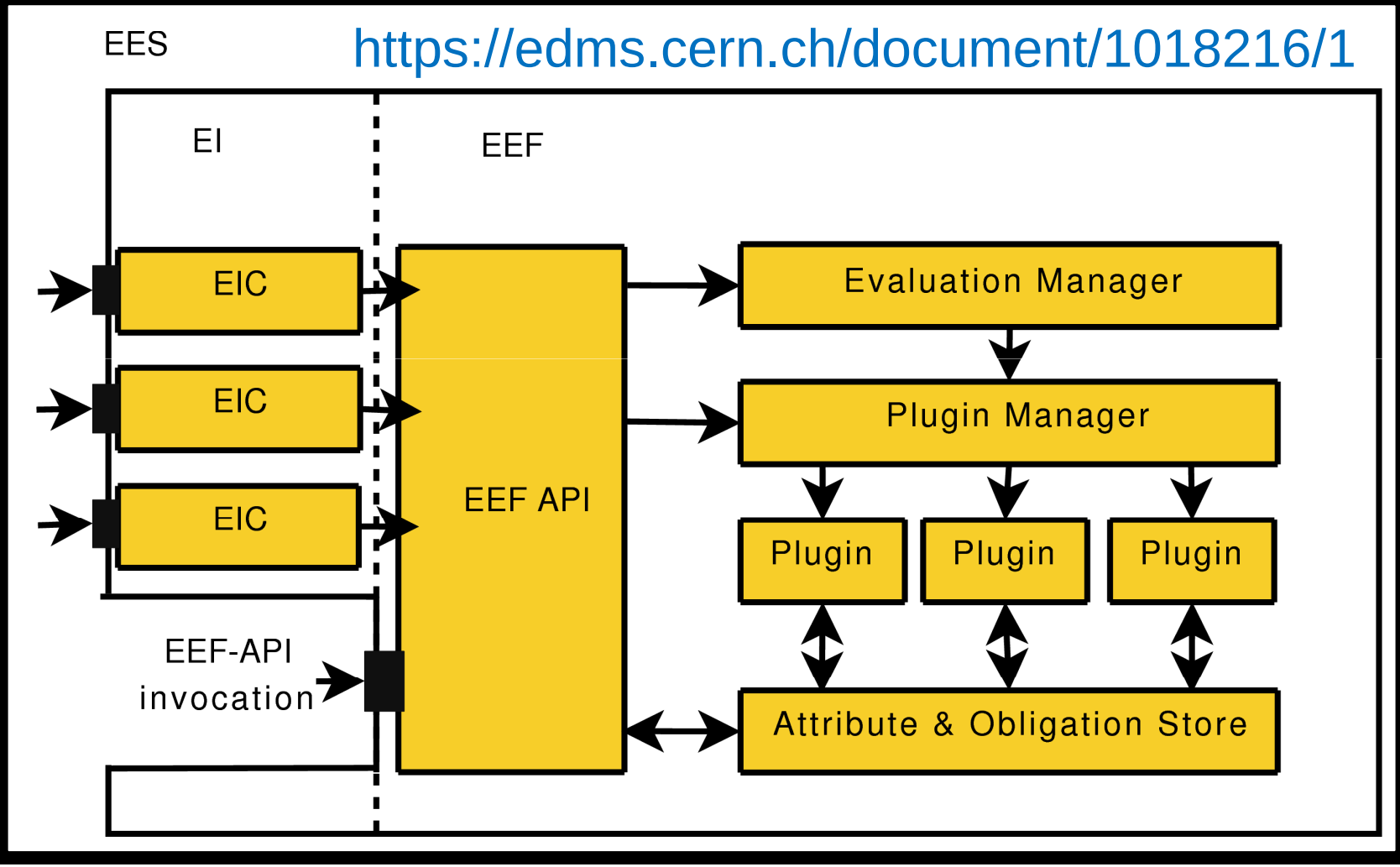
Argus Authorization framework

Local Policy Enforcement

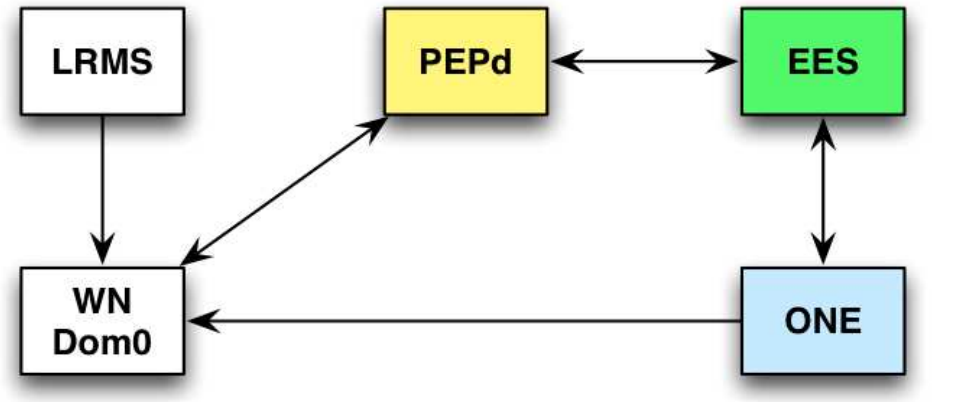


= Argus framework

Execution Environment Service



Deployment

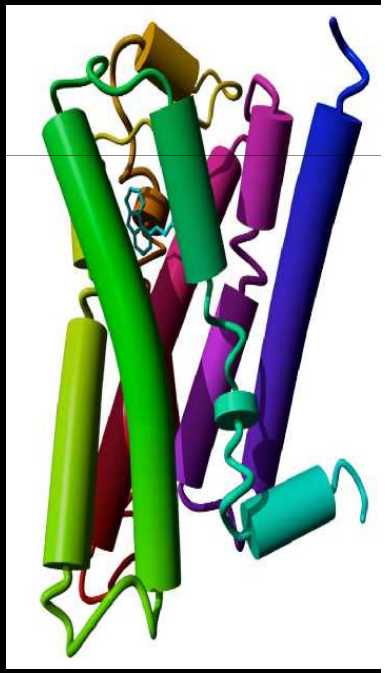


- A lot of similarities with WNoDeS, Nimbus
- Additional benefits compared to WNoDeS
  - Policy enforcement with standard middleware
  - Modular: EES middleware sets environment
  - OpenNebula is a standard VM Manager
- Disadvantages
  - Not operational yet, scalability unproven

Typical use case Class 3 VM

Identification and classification of GPCRs

- Requires very specific software set
  - Blast 2.2.16
  - HMMER 2.3.2
  - BioPython1.50
- Even non-x86 (binary) applications!
- Specific software for a single user
- No common experiment software



Cloudia project, SARA

- Make separate DMZ for Class 3 VMs
  - SSH accessible (Amazon like, OpenNebula based)
- Comparable to “Guest networks”
  - Only outbound connectivity
- Detection of compromised guests
  - Extended security monitoring
    - Packet inspection, netflows, hypervisor monitoring
- Small, but upgrade expected (500 kEuro)
- Floris Sluiter: floris@sara.nl

Email not displaying correctly? View it in your browser

HPC In the Cloud Weekly Update

Dedicated to Covering Enterprise & Scientific Large Scale Cloud Computing

May 11, 2010

Behind the Cloud

CRM in the Life Sciences: Why the Veeva Announcement Matters

With the recent FDA rulings governing how sales can interact with doctors, the importance of a CRM system to track the complex customer interactions has never been more critical. Read More...

Cloudview: An Interview with Dr. Ignacio M. Llorente

Today's entry contains feedback on the present and future of HPC in the cloud from from Dr. Ignacio M. Llorente, Professor in Computer Architecture and Technology, and the Head of the Distributed Systems Architecture Research Group at Complutense University of Madrid. Read More...

One Small Step for Man, One Giant Leap for HPC and Cloud

Gaia's infrequent but immense demands for mission-critical data processing created the need for one of the most convincing proof of concept measures tackled by cloud infrastructure and development giants, The Server Labs and RightScale. Read More...

- **SARA Opens Gate for HPC Cloud Researchers**

More Feature Articles...

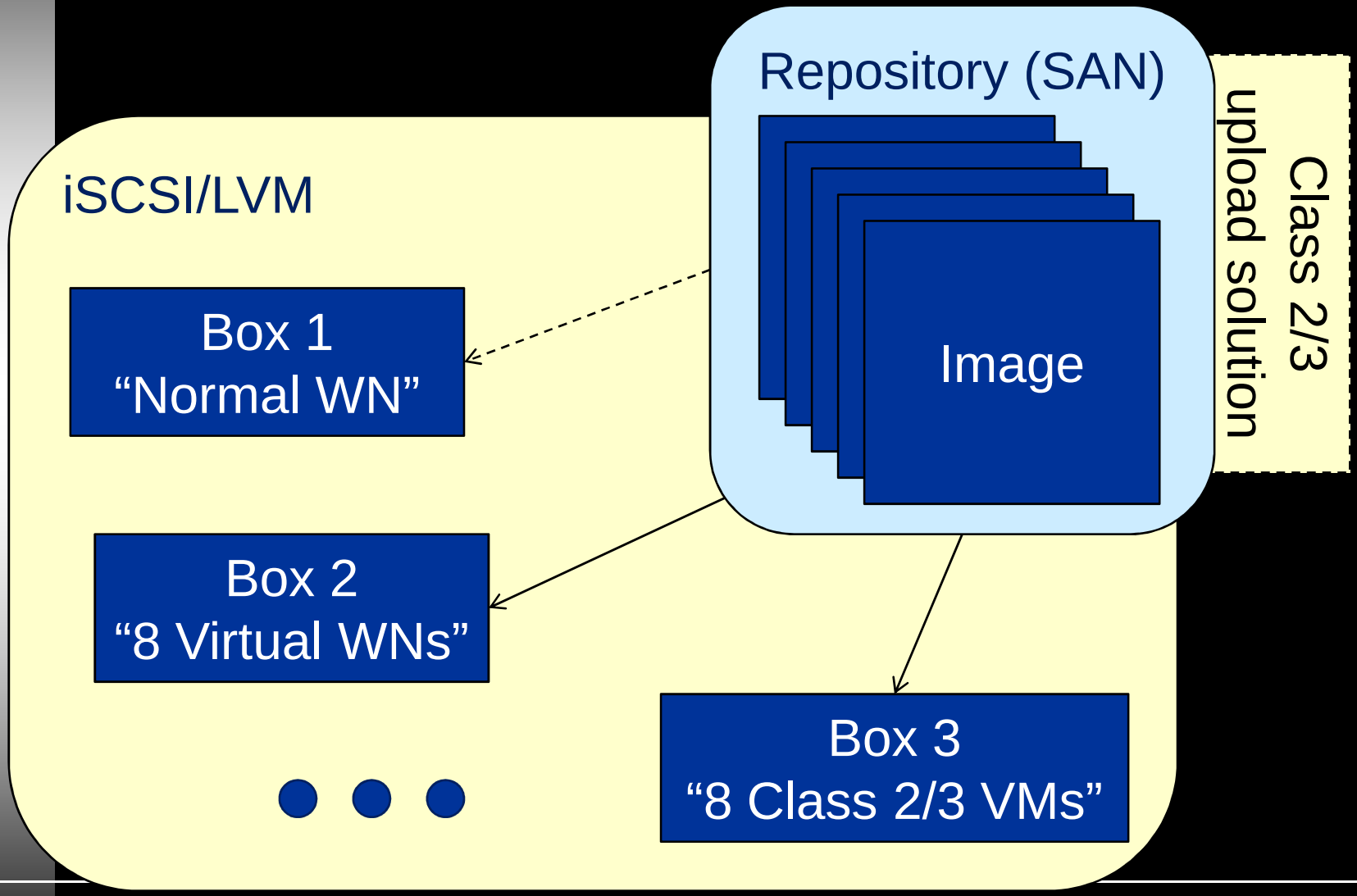
Top Headlines

- **Will Cloud Bring Supercomputing to the Masses?**

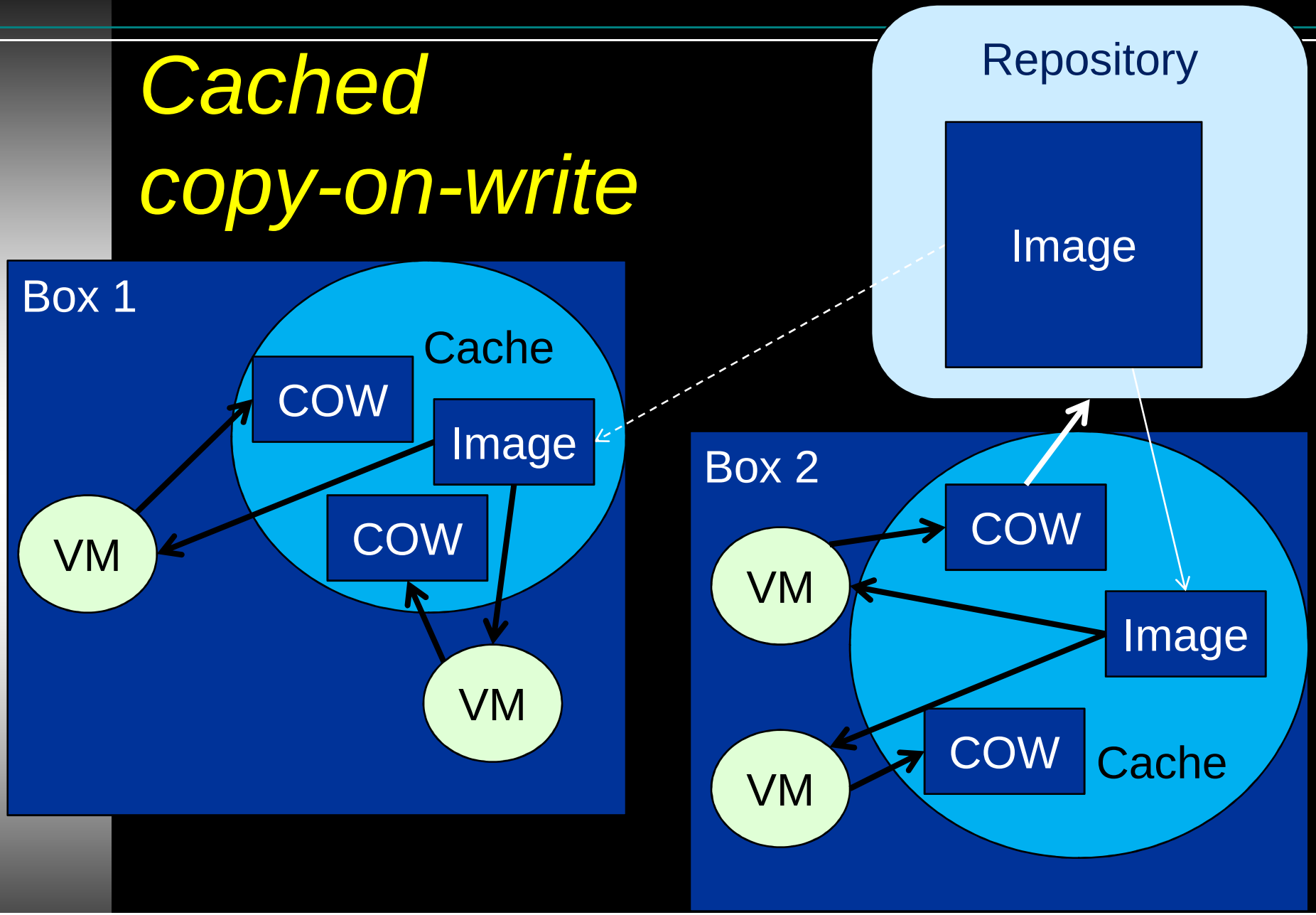
Scalability, Reliability, Maintainability, Usability, *bilities

CHALLENGES

Distributing VM images

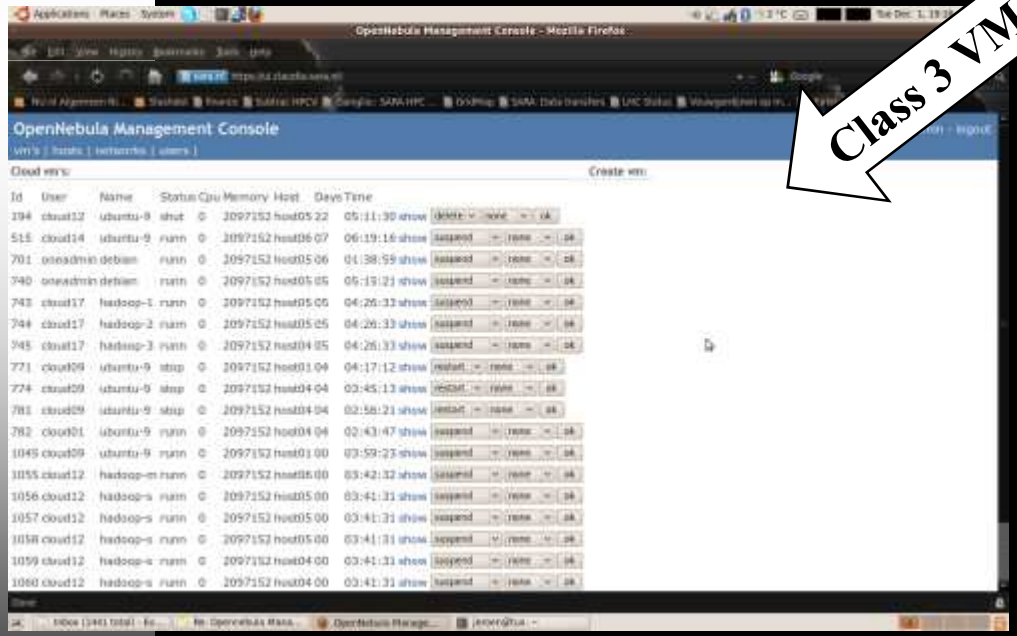


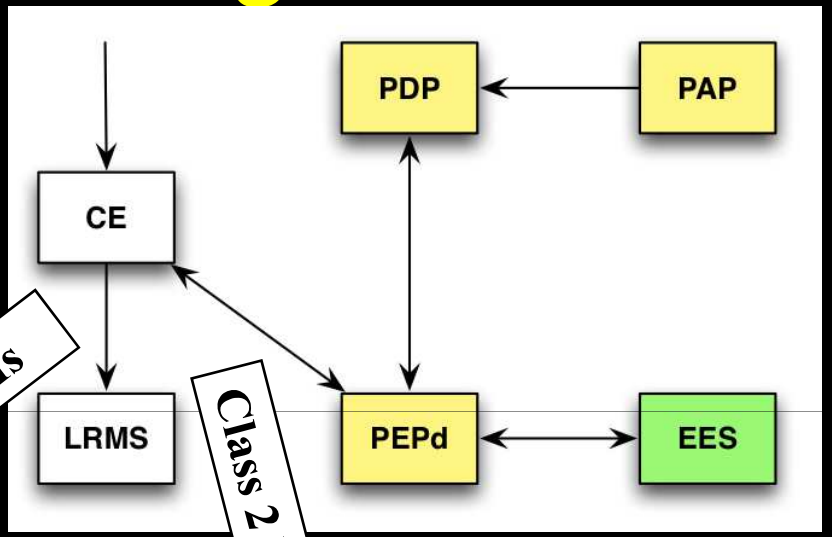
Cached copy-on-write



Hybrid solution: integration

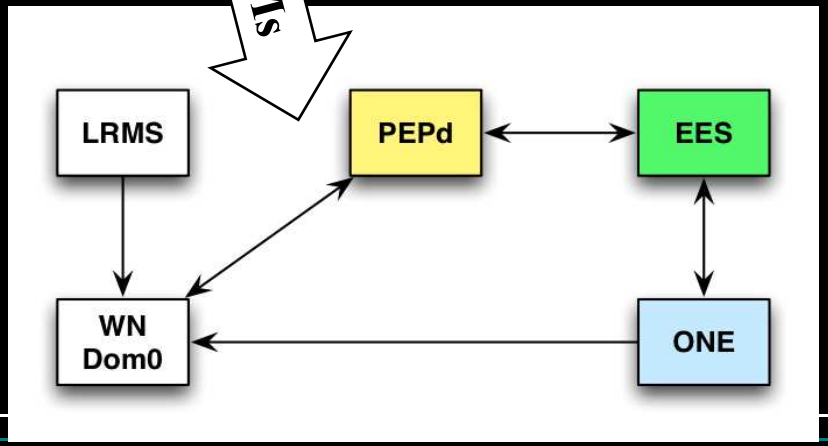
Cloudia Interface





Class 3 VMs

Class 2 VMs



Reliability / Usability

- Experiences from WNoDeS and Cloudia
- EGI focus point
  - See position paper on CloudScape II, 2010

http://www.ogfeurope.eu/Repository/FileScaricati/Cloudscape_II_Position_Papers_and_Professional_Profiles.pdf

- Industry developments
  - Grid was so far a purely scientific endeavor
  - Keep a close eye on industry for all the *bilities

Conclusions

- Integration of Grids and VMs is a hot topic
  - WNoDeS, HEPiX, BiG Grid, OpenNebula, etc.
- Many different approaches
  - Trusted/Untrusted, Grid like/Cloud like
- Scalability, reliability, maintainability, usability remain issues of primary concern.
  - Cloud like solutions: do they scale?
  - Grid like solutions: are they user friendly?