HEPiX Virtualisation Working Group

Status, June 22nd 2010

Tony.Cass@cern.ch

Agenda

- Objective / Background
- Status
- Comments/Thoughts
- Summary

Agenda

- Objective / Background
- Status
- Comments/Thoughts
- Summary

Objective

- Enable virtual machine images created at one site to be used at other HEPiX (and WLCG) sites.
  - As worker nodes: end user does not have root access and virtual image does not contain an execution "payload".
    - No requirement that images use only a single processor.
- Background
  - Clear resistance by many sites during 2008/9 to use of virtual images created elsewhere.
- Approach
  - Establish trust
    - guarantee to sites that images are secure and can be configured to ensure that sites can fulfill obligations --- e.g. for logging and accounting.
    - guarantee to experiments that images won't be changed; users can be 100% sure of software environment.

Agenda

- Objective / Background
- Status
- Comments/Thoughts
- Summary

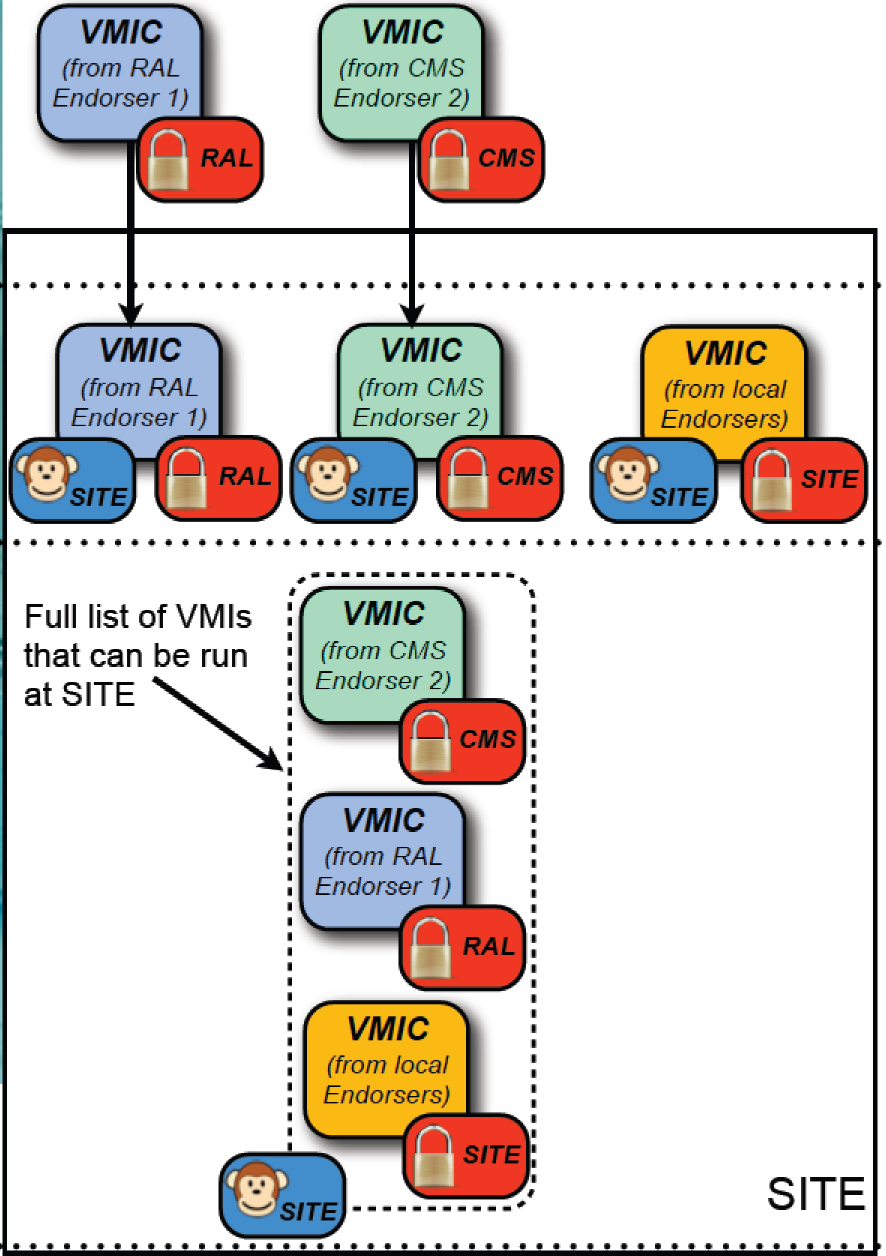
Four work areas

- Policy for Image Generation
- Image Cataloguing & Distribution
- Image Contextualisation
- Support for Multiple Hypervisors

Four work areas

- Policy for Image Generation
  - Document prepared establishing obligations for people preparing images for use at remote sites (“endorsers”). See https://edms.cern.ch/document/1080777.
    - Basic point: “[endorsers] are held responsible by the Grid and by the Sites for checking and confirming that a VM complete image has been produced according to the requirements of this policy and that there is no known reason, security-related or otherwise, why it should not be trusted.”
    - Comments welcome; document has been distributed to WLCG security group & GDB.
- Image Cataloguing & Distribution
- Image Contextualisation
- Support for Multiple Hypervisors

Heterogenous example



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

Four work areas

- Policy for Image Generation
- Image Cataloguing & Distribution
- Image Contextualisation
  - i.e. customising images to meet site policies
    - Amongst other things, required to enable sites to meet Grid security policies, e.g. for traceability of user activity.
  - How
    - ISO CD filesystem attached to VM at instantiation. VM invokes two scripts from the filesystem, one prior to network configuration and one at the very end of the start up process.
  - What is not contextualised?
    - Image & software changes (python version…) forbidden.
      - No patching: in case of security concerns, site must refuse to instantiate image and inform endorser.
- Support for Multiple Hypervisors

Four work areas

- Policy for Image Generation
- Image Cataloguing & Distribution
- Image Contextualisation
- Support for Multiple Hypervisors
  - Sites interested in KVM and Xen (full and para-virtualised)
  - A procedure has been documented for a generation process that allows images to be run with both KVM and Xen.

Agenda

- Objective / Background
- Status
- Comments/Thoughts
- Summary

Comments/Thoughts

- VM image generation provides VOs with absolute control over installed software.
  - Criticism: software changes too often.
    - Is fixed OS image + CernVM FS for experiment software an interesting combination?
  - Why not just use CernVM?
    - Can sites contextualise image? If so, no reason per se why not if image is endorsed.
      - But this a personal view, not representing the HEPiX working group.
- Instantiation of VM images allows direct connection to pilot job framework, bypassing local batch system.
  - Some sites are keen to move in this direction now, although others are wary.
    - But this is a personal view, not representing the CERN batch team!
  - Dynamic instantiation of images to reflect instantaneous demand is an attractive scenario.
    - But this is a personal view, not representing the CERN batch team!

Agenda

- Objective / Background
- Status
- Comments/Thoughts
- Summary

Summary

- A year ago, sites were rejecting any possibility of running remotely generated virtual machine images.
- Today, we have the skeleton of a scheme that will enable sites to treat trusted VM images exactly as normal worker nodes.
  - This enables
    - VOs to be 100% sure of the worker node environment
    - (potentially) inclusion in the VM image of the pilot job framework enabling “cloud like” submission of work to sites.
- **Active involvement of VOs is now highly desirable** as we move towards delivering a proof-of-concept system.
- Nothing in what is being done
  - prevents sites that wish to do so from implementing Amazon EC2-style instantiation of user generated images, or
  - precludes use of CernVM.