VM contextualization and image cataloguing in ATLAS

Fernando H. Barreiro Megino

CERN IT-ES-VOS

On behalf of the ATLAS ADC Cloud Computing R&D

ATLAS Cloud Computing R&D

- ATLAS Cloud Computing R&D effort since one year
  - Bits and pieces of ~10 persons
- Running MC production on PanDA queues with different cloud infrastructures
  - LxCloud, HelixNebula(CloudSigma), Stratuslab, FutureGrid, Canadian Clouds
- Tools to automatically scale up/down size of the cloud clusters
  - CloudScheduler and CloudFactory
- Getting experience with virtualized analysis clusters (batch/PROOF/PanDA)

- Until now: Mostly development and Proof-of-Concept phase
- Plans for the next months
  - Graduate our services to production
  - Improve automation and monitoring needed for scaling
  - Push the usage of as many "free" cloud resources as we are offered
  - Promote contact with and between sites
    - What are sites interested in?
    - Pioneering sites should advise others

- Rely on CVMFS for applications
- CernVM is convenient as our default VM image
  - Manual interaction to prepare the image and save a "golden image" for a site/context
    - (Needed since some commercial providers (in Helix Nebula) do not support contextualization at all)
    - Install additional packages (e.g. Condor, ganglia) & config
    - Need to rebuild golden image each time there is a change
  - HEPIX-compliant CDROM contextualization
    - Used in some PanDA analysis tests at LxCloud but not for our MC production

- Starting to test SL & CentOS using Puppet-based contextualization
  - Security: will sites allow us to change the image at boot time? How much?

Image management

- Currently the ATLAS image management is a purely manual operation:
  - No cataloguing, sharing or signing of images done so far
  - Each sub-project (~3 groups) builds and uses its own images
  - Starting to document/share available images between us

- Moving forward:
  - Image signing is needed for VO/site trust
    - Not just site trusting the image – also ATLAS trusting what it gets
  - Cataloging (StratusLab-Marketplace style) will be useful
    - but will still require a lot of manual interactions to get the images running on our clouds (with hugely varied Cloud APIs)

- Long term goal: Automatic image factory for making images, cataloging them, store site parameters, publishing and monitoring