

# Cloud pre-GDB Summary

Michel Jouvin  
LAL, Orsay  
[jouvin@lal.in2p3.fr](mailto:jouvin@lal.in2p3.fr)

September 2014

# Meeting Facts

- Agenda: <https://indico.cern.ch/event/272791/>
- Summary (in progress):  
<https://twiki.cern.ch/twiki/bin/view/LCG/20140910PreGDB>
- Well attended
  - > ~30 local
  - > ~10 remote
  - > ATLAS and LHCb well represented
  - > CMS: many people travelling to a meeting in the US
  - > ALICE: absent?

# WG Goals

- ◉ WG is about exploring the possibility to use private/community clouds as a replacement for grid CEs
  - > No intention to cover all the aspects of cloud usage by experiments
  - > Focus on shared clouds rather than dedicated resources to one experiment
    - But many lessons can be learnt from private clouds...
  - > This perspective is part of the WLCG future directions presented/discussed at last WLCG workshops
    - Copenhagen and Barcelona
- ◉ Build on existing work in experiments: do not start a new huge R&D project
  - > No manpower available to do it
  - > Tackle all foreseen operational issues: scheduling, accounting, security...

# Since January Meeting

- Meeting summaries:  
<https://twiki.cern.ch/twiki/bin/view/LCG/CloudDiscussions>
- Mailing list very quiet but work progressing on several topics
  - > Graceful termination of VMs: Machine/job features infrastructure being deployed into production
    - Followed by WLCG OpsCoord, dedicated TF (S. Roiser)
  - > Cloud accounting published to APEL
    - In particular EGI Federated Cloud
  - > Target share implementation
    - A. McNab started an OpenStack implementation of Vac concept (vcycle)
    - CREAM CE team started the implementation of a “fairshare scheduler” for OpenStack with request queuing

# WLCG Context Evolution

- Reduction of funding for operations and MW development is continuing, in particular in Europe: urgency to reduce the operational/development cost
  - EGI-Inspire ending: follow-up project will have less support to operations
  - Budget crisis in many countries leading to personnel reduction
- Computing provisioning is a key area for savings: ideally require nothing specific from sites to benefit from their resources
  - Setting up/managing a grid CE is “complex” and **specific to our community**
  - Cloud seen as a more pervasive technology for sites, no MW development required by the community

# Current Meeting Goals

- Foster the current work on open issues for replacing a grid CE by a cloud
  - > Make possible for sites interested to do it
  - > No intent to make the move compulsory...
  - > Does not prevent other solutions, like a cloud behind a CE or a batch system
    - Already working in fact... See Uvic presentation today
- A comprehensive summary of where we are presented in Barcelona by L. Field, based on CERN experience
  - > This summary was used as the starting point of this meeting
- Expected outcome of the meeting: realistic milestones toward direct usage of shared clouds

# Target Shares...

- Ability to do a dynamic partitioning of a shared clouds
  - > Goal: achieve something similar to fairshare in batch systems
- 2 implementations started in parallel in the last months
  - > Vcycle (A. McNab): a service that is in charge of instantiating VMs according to VO target shares
    - Based on the Vac ideas, implemented as a cloud client
    - For pilot-based VOs: no attempt to contact the pilot framework, the VM will do it
    - Possibly cloud MW agnostic: will support EC2 and OCCl
  - > FairShareScheduler (Padova): a drop-in replacement for nova-scheduler, queuing of requests
    - OpenStack specific, based on SLURM scheduler
    - Currently in test in Bari
    - Trying to get it into mainstream OpenStack (BLAZAR)

# ... Target Shares

- ◎ Both approaches seen as possibly complementary
  - > Vcycle-like component is probably a requirement: addresses the capacity management needs, which is more than scheduling
  - > A fair share scheduler can help in situation when some VOs don't rely on pilot jobs
- ◎ Milestone for the next 6 months: increase deployment of both solutions
  - > Vcycle: already several deployment in the UK, will try to find other sites
  - > FairShareScheduler: a few OpenStack sites interested to evaluate it

# Accounting

- ◎ CERN report on the work done to benchmark VMs
  - > Every VM classified in “HW types”
  - > HW types defined from live information exposed in the VM
    - CPU type, number of cores, mem speed...
  - > HS06 benchmark run on a large number of instance for each type
    - As a job prologue
    - HW type performance: “pessimistic mean” of the result
  - > A table built with the performance of each HW type
  - > When starting a VM (can) retrieve its perf from the table
    - Could be exposed to payload through machine/job features
    - Accuracy assessment at CERN: better than 10%
- ◎ A few shortcomings identified in APEL for clouds
  - > Number of cores and HS06 perf of VM not collected
  - > Work needed in collection agent and portal: will be done

# Security...

- ⦿ Cloud is a paradigm change compared to grid with respect to responsibilities sharing between sites and VOs
  - > VOs will have sysadmins responsibilities, e.g. VM image maintenance
  - > Change in liability: sites will not be liable for VM image problems
- ⦿ Cloud is not necessarily a major change in incident handling compared to today (pre-glexec) grid operations
  - > Will need a cooperation between sites and VOs to get all the information needed for incident handling
    - Proved to work today in grid

# ... Security

- Need to ensure that both VOs and sites collect the information required for traceability in clouds
  - > Nothing really new...
  - > But the risk of loosing transient information in images
    - Probably no alternative to central syslogging at sites: are all the sites doing it
  - > VOs already uploading centrally a lot of logs/information to be able to debug applications
    - Can we take advantage of it for security incident handling
- Proposal: create a short-lived TF to do the traceability gap analysis
  - > Are experiments and (some) sites ready to participate?
    - Contact me: I'll send a more formal request...
  - > Outcome/proposals must be compatible with available manpower...

# Conclusions

- Despite being best effort, a significant work is being done and progresses on all main issues
- Target shares was the real showstopper and solutions foreseen
  - Already tested at some scale... in particular the Vac/Vcycle approach
- Accounting: CERN work allowed to identify the missing pieces
  - Mainly a problem of manpower to implement things
  - No major development, improvements to what exists (APEL)
- Security: risk that as usual the necessary work is postponed...
  - Will make difficult to solve issues that would be reasonably easy to address before going to large-scale production