Cloud pre-GDB Summary

Michel Jouvin
LAL, Orsay
jouvin@lal.in2p3.fr

GDB, January 2014

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 at last WLCG workshop in Copenhagen
- 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…

Last Meeting Focus (July)

- Review progress/work about cloud usage in each experiment
  - Most of the work done with HLT farms… which are not shared clouds
  - An important milestone for integrating cloud backends in pilot factories
- Review concrete work about implementing graceful VM termination as discussed in March
  - Integrated into job/machine features framework
- Continue discussions on possible models for non static sharing of resources
  - Target share approach vs. economic models
  - Vac approach: resource provider decides the next VM to run rather than waiting for submission
    - Volunteer computing model rather than cloud approach

Pre-GDB Facts

- Well attended : ~40 people
  - Including local+remote North America participation
- All 4 WLCG experiments presents
- Indico and presentations
  - https://indico.cern.ch/conferenceDisplay.py?confId=272783
  - 3 topics: accounting, security/traceability, target shares
- Summary
  - https://twiki.cern.ch/twiki/bin/view/LCG/20140114PreGDB

Accounting…

- A working prototype based on APEL
  - Developed in the EGI Federated Cloud TF
  - Based on UR 1.1 from OGF + SSM for transport
  - Collection of data, push to central DB, publishing on portal
  - If needed, integrating Amazon is doable by parsing bills
- How to normalize a VM CPU power?
  - HS06 remains the best VO-independent metrics
    - Doesn't preclude a VO from collecting other metrics for internal use
  - Job/machine features is the way to publish CPU power to the job
    - This also needs to be pushed to the accounting
    - A benchmark job may be run on the VM as part of the machine features
  - Don't need a very precise measurement: +/- 20% would be ok

… Accounting

- WC time vs. CPU time
  - Stay with CPU time: homogeneity with grid
  - WC time is even more difficult to normalize
  - Allow a site to overcommit resources to work around potential VM CPU inefficiencies

Security/Traceability…

- Several security experts absent but meeting well prepared by Romain’s team and EGI SVG
  - See presentation
  - Triggered a lot of discussions
- Several pre-production and production resources: time to discuss issues and find appropriate solutions
- Current security model based on 2 roles
  - Endorser: responsible for producing/maintaining the image
  - VM operator: person/entity instantiating the VM (with full rights)
  - Both must be distinct from end user to be trustable
- Main discussion around achieving traceability
  - Traceability requirements remain the same as for grid
    - Need proper configuration of central logs
    - Some new logs may be needed (e.g. VM instantiated)
  - Absence of root access for end users remain critical

… Security/Traceability

- User compartmentalization is required on "multi-user VMs"
  - Allow to block one specific user, avoid interference between users
  - Avoid to bring complex part of grid SW to do the mapping…
  - Proposal: on-the-fly creation of a new userid for each user payload
    - No attempt to make a unique mapping for each DN
    - Check that the DN is not banned before: provide a simple tool to do it
    - Provide a glexec-like functionality using standard sudo
- Vulnerability handling requires the ability to terminate VMs in a reasonnable amount of time
  - Job/machine features will provide the mechanism
- Policy discussion: no time, missing experts
  - To be done at a future meeting

Target Shares…

- Follow-up discussion for target share implementation
  - Almost a consensus that implementing an economic model without real money is far too complicated…
  - Almost a consensus that without it it is impossible to replace a CE/batch system
    - Batch systems provide target share through the fair share mechanism
- 2 main difficulties
  - No queuing of requests: impossible to arbitrate the next request to start to rebalance shares
    - Reintroducing queuing is not desirable: better to use a batch system
  - When some resources have been freed, no guarantee that the next request will be from a VO under its quota
    - Temporarily refusing VOs over their quota may be difficult to manage (when? How long?)

… Target Shares

- Avoid asynchronous processing of requests
  - Queuing for requests only: how to manage credentials cache, risk of starting a resource too late…
- Avoid building a complex service over a cloud MW
  - Avoid hacking/forking the cloud MW to support such a service
- Proposal: start long lived VMs for a VO up to its target share, then start spot instances with a short minimum lifetime
  - Use job/machine features to let the job know the lifetime
  - Will ensure a minimal VM turnaround
    - A cloud is supposed to be a large resource (several Kcores)
  - Use temporary overcommitment to let a VM start when another one is reclaimed
  - Explore the ability to dynamically adjust quota based on log file analysis (request pressure per VO)

Conclusions

- Good meeting with lively discussions!
  - Progress toward common understanding and solutions
  - Concrete work going on in several places
- Agreement on a "minimalistic" approach
  - Do not develop complex services or require big changes in exp SW
  - Do not "fork" cloud MW
- Not yet clear that a cloud without a batch system can replace a CE in a context where queuing is managed by the VO
  - Larger agreement that it would bring advantages for VOs if possible
- Egroup for discussions: register for if you are interested
  - project-lcg-gdb-clouds-wg@cern.ch
- A future meeting probably in Spring