Cloud pre-GDB Summary

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

GDB, March 2013, KIT

Last GDB Summary

- Try to converge on some **concrete steps** implementable in existing (private) clouds that could be tested with real world applications
- Egroup discussion has been good for bootstrapping the discussion and do some kind of brainstorming…
- Converging on a workplan will probably require a more formal meeting
  - Would be better if it was mostly F2F: Karlsruhe (Tues. afternoon) is probably the only possibility in a reasonable timeframe but at least 2 conflicting meetings (ATLAS, ROOT)
  - If not possible, fall back to a Vidyo meeting

Pre-GDB Agenda

- 3 initial topics identified based on January discussions a little bit reorganized after the initial discussion
  - Image contextualization
  - VM instantiation and duration
  - VM scheduling to achieve fairshare-like resource sharing
- Security model
  - In particular, is it still a goal/requirement to prevent root access to VMs
  - Impact on possible/acceptable contextualization strategies
  - Need for a JSPG policy update?
- Accounting
  - VM benchmarking: what to report? How to ensure consistency between sites?

Security…

- Trusted images: definition currently based on a JSPG policy proposed early in the HEPiX WG
  - Corner stone: no root access to the VM
  - Endorsed by EGI, WLCG a few years ago…
  - Probably need to reopen the discussion based on cloud experience
    - Not existing when the first version of the policy was defined
    - Root access is a key feature of every cloud… difficult to prevent it!
    - Role of a policy if root access is accepted?
- Liability and level of traceability currently available
  - Goal: have the same level of traceability back to the user as we have in the grid (with glexec)
  - If root access to VM accepted, how to enforce it

… Security

- Agreement: no root access needed/envisioned for the end users in WLCG VOs
  - Root access restricted to the user who instantiate the VM: the pilot factory user
    - May need to further refine what actions are allowed/disallowed
  - This specific user in the VO is liable for root account usage: it is its responsibility to ensure that no other user of the VM is enabled to use it
  - Identity must be switched to a non root user to execute any payload
    - Need to evaluate/discuss with experts if glexec may be used in the cloud context to trace identity switching
  - Passing user credential to a VM is better done on an encrypted connection
    - 1 possibility is to do it with SSH using root (the only accessible account)

Image Contextualization…

- Contextualization: way to pass data to the image at instantiation time
  - Only clean way to pass credential to an image
  - Site and/or contextualization
- User contextualization acceptance strongly related to root access debate
  - User contextualization is a way to bypass root access restrctions…
  - … but in the cloud world user root access to a VM is a basic feature
- HEPiX proposed a mechanism based on amiconfig
  - Focus on site contextualization
    - Controlled user contextualization also possible
  - Well integrated into CERNVM

… Image Contextualization

- Since then, CloudInit emerged as the new de-facto standard
  - Based on the same concepts as amiconfig
  - More data input mechanisms: backward compatible for the user
  - More user contextualization oriented: a lot of flexibility added
    - Including ability to execute arbitrary scripts
- Agreement: CloudInit is the way to go for the future but we can live for the time being with CloudInit and amiconfig
  - CMS already played with CloudInit and amiconfig but no attempt to convert one to the other
    - StratusLab report: non-zero but minor
  - No real impact on the user/VO if the input data syntax is the same
    - Unfortunately this is not generally the case
  - Need to wait more concrete plans from CERNVM

VM Instantiation

- Mainly a matter of interfaces…
- General agreement that interfaces are not really important
  - Most VO using abstract API like libCloud (DIRAC) or CERNVM Cloud
    - CMS may consider DeltaCloud: supported by Condor thus coming for free
  - One (non convincing) standardized interface recommended/used by EGI federated cloud TF : OCCI (OGF)
    - Interface not well designed
    - Implementations available for several cloud MW but not mainstream for any of them
    - Contextualization not supported
  - One emerging new standard: CIMI
    - Proposed by the same organization as CDMI (DTMF?)
    - Soon to be proposed as an ISO standard
    - May want to follow further developments with it…

VM Duration…

- Long-lived VMs are requested by several Vos
  - But require a way for a VO to shut down a no longer needed VM
- Main topic is the graceful stop of a VM
  - Overlap with VM scheduling discussion
  - Now recognized as a feature required as a counterpart to long-lived VOs
- Proposal from previous discussions
  - Based on SLAs, launch a VM with X minimum days of lifetime and Y minimum hours of shutdown notice
    - Probably X is not really needed and Y should be part of the SLA
- Mechanism to publish information to VM user should be independent from any cloud MW implementation
  - HEPiX well-known file proposal looks as a good starting point

…VM Duration

- How the file is updated is out of the scope of our discussions
  - Site decision: site should use contextualization to install what is necessary at VM instanciation time to ensure the proper update of the information
    - Eg.: cron job
    - A site can prefer to use a shared file system
- Be pragmatic: start something addressing the main needs but do not try to embrace all the possible use cases
  - First step: demonstrate ability to send an advance notification to the VM user, play with different SLAs for VMs in the normal share of a VO and those above it (sort of spot instances)
    - Termination date for a VM should be given in absolute time
  - Left outside short term plans: ability to reclaim the VO a certain number of VMs rather that specific VMs

VM Scheduling

- “Fairshare-like” resource sharing: agreement that we want to avoid static partitioning of resources
  - Graceful termination of VMs opens a way to implement this fair sharing still enabling one VO to take advantage of the underused resources by another VO
  - Difficulty: how the cloud scheduler can discover requests by other VOs that are under their quota
    - Batch sytems can do it because they have a queueing mechanism but there is no such feature in clouds. A reason to keep them?
    - Do we want to implement (see implemented!) a mechanism for a VO to let a site know they would like more resources: risk of reinventing a complex system
  - As an alternative, explore economic models where VOs are given credits and where the price of a VM increases with its duration an the number of VMs owned by a VO.

Accounting

- General agreement about using wall-clock time accounting for the cloud world
    - Concerns about funding agency reactions if they think we inefficiently use the infrastructure, even though the VO is responsible
- How to report doesn't seem to be problem for private clouds
    - APEL has demonstrated its ability to do the job
        - See work done by EGI federated cloud TF
    - This is not WLCG responsibility to report public cloud usage into WLCG central accounting
        - But an experiment is required to do such an accounting
- VM benchmarking: what to report? How to ensure consistency between sites?
    - Easy to invent a very complex system… Must be avoided!
    - Not specific to clouds but they may offer a possibility to improve the situation

Conclusions

- Good/better consensus on important issues to tackle for making possible to use a cloud as a CE replacement
  - Batch-less interaction with the compute resources
  - First priority: demonstrate a basic feature to do resource reclaim
    - Graceful termination of VMs
- Probably the end of a first phase of our work: reach enough consensus on issues to devise a work plan
  - Still some details to be discussed/done…
  - But the most important is now to try to implement ideas discussed and review them afterwards
- Another similar meeting foreseen next Spring
  - May or June GDB slots: please report known conflicts
  - Requires some practical work/testing to be done before…