Cloud Discussions

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

Pre-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

Today's 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
- One topic added: 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?
- One topic left outside of the today's discussion: accounting
  - Accounting and 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
  - Endorsed by EGI, WLCG…
  - Probably need to reopen the with cloud experience in mind
    - Not existing when the first version of the policy was defined
  - Root access and its consequences have to rediscussed
    - 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

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
- General agreement that a contextualization mechanism is needed
  - Generally scripts executed at startup
  - Must be common to all infrastructures to enable the use of a unique image on many different infrastructures
  - Several transmission method for contextualization data generally supported: "CD-Rom", http server…
- HEPiX proposed a mechanism based on amiconfig
  - Focus on site contextualization: attempt to prevent user contextn
  - Well integrated into CERNVM

… Image Contextualization

- Since then, CloudInit emerged as the new 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
  - Worries about the effort evolved in moving to it
    - StratusLab report: non-zero but minor
- 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

VM Instantiation

- Mainly a matter of interfaces…
- General agreement that interfaces are not really important
  - Only CMS insists on EC2 because of Condor
  - Others already have several backends (Atlas) or are using abstract API like libCloud (DIRAC) or CERNVM Cloud
  - 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

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
- Not a complete agreement that such a mechanism is needed
  - But probably required to implement a fairshare strategy by shutting down long-lived VM
  - Else experiments will have to accept arbitrary killing of tasks in execution… or it will not be difficult to implement long-lived VMs in shared resources
- Gavin's proposal
  - Based on SLAs, launch a VM with X minimum days of lifetime and Y minimum hours of shutdown notice

…VM Duration

- Open question: do we need a (common) mechanism to gracefully reclaim resources
  - Some says "no" to graceful reclaim as it is not a feature available in commercial clouds
  - But most think we need this mechanism not only for scheduling but for site management without static lifetimes
    - A site may need to shut down many VMs for various reasons: just kiilling them is not acceptable
  - A mechanism to communicate the information to the user has been proposed by HEPiX: a well know file
  - It is not clear how this file is updated after started and if it is a purely internal site matter or if it would help to have an agreed method
    - Can it be done by signaling VMs, even long in advance?
    - Does it requires something installed in the image?

VM Scheduling

- "Fairshare-like" scheduling in clouds: generated a lot of debate (flame?!)
  - Many people don't like the word "fairshare" as they worry we just want to import an inappropriate concept from grid…
  - But generally an agreement that it is good to be able to allocate resources to the actual needs of users without a static "partitioning" of the resources and guarantee a certain share of resources over time

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
  - Do we care about reporting public cloud usage into WLCG central accounting?
    - It is an experiment decision to use them, not part of WLCG infrastructure
- Accounting and VM benchmarking: what to report? How to ensure consistency between sites?
  - Easy to invent a very complex system…