Science & Technology Facilities Council

Tier 1
A Forward Look

Ian Collier

GridPP 31

September 25th 2013

A Forward Look

“A happy man is too satisfied with the present to dwell too much on the future.”

Science & Technology Facilities Council

Setting the Scene

A number of developments in Tier 1 over last 4 years position us to be able to adapt & respond better

- Comprehensive configuration management & provisioning system
- Virtualisation for (nearly) all grid & core services
- Better internal information systems
- Cloud testbed

Grids: what did we achieve?

And fail to achieve?

- Solved our problem of making effective use of distributed resources
- Made it work at huge scale
- Effective to ensure all collaborators have access to the data
- Networks are a significant resource
- Federation of trust and policies – important for future

- Cluster computing/ grids not suitable/ needed for many sciences
- Operational cost is high
- Very complex middleware was not (all) necessary
- Many tools were too HEP-specific

What I've been reading

Update of the Computing Models of the WLCG and the LHC Experiments

https://indico.cern.ch/materialDisplay.py?materialId=0&confId=212501

I'll talk about some of what I think this might mean for the Tier 1

Science & Technology Facilities Council

Less domain specific tools

“Another theme that is coming to the fore is the desire to use solutions that are not specific or peculiar to HEP, where this is realistic. Use of standard software, protocols, and tools where possible again helps to make the WLCG infrastructure and services easier to maintain, operate and evolve, as long as such standard solutions can satisfy our requirements.

On the other hand there are clearly areas where LHC is unique – such as the need for a globally federated data distribution and access service.

However, even in such an area LHC may be unique today, but there are many other sciences that will soon have similar problems to solve, and so HEP should ensure that tools and services that we develop might reasonably be eventually made available to those other communities.”

Less domain specific tools

- Desire for standard tools where possible.
  - Should make maintenance easier
- Remain some areas of uniqueness
- Even those may be common to emerging scientific applications

VMs & Cloud interfaces

“Several sites are also supporting virtual machines through cloud management software. Several experiments are testing direct submission of work to such cloud interfaces. This is useful in order to be able to use opportunistic cloud resources (several such have been offered recently), and for those sites that deploy such software, it may make sense in the future to use those cloud interfaces, rather than through the existing CE’s. Thus, we may start to see a gradual evolution from the existing grid mechanisms to private clouds.

“An interesting aspect of this is that there are large support communities behind such cloud software, while support for the CE software essentially depends on a much smaller community, and has relied on specific additional funding for that support, which may not be there in the longer term.”

VMs & Cloud interfaces

- Much work on opportunistic clouds by VOs & experimenting with IaaS infrastructure at sites
- Can imagine evolution to private (federated) clouds - and preference for cloud interfaces
- Larger support community (esp vs CEs)

Multi-core, whole node, VMs

"Virtual machines may also help to improve the efficiency of CPU use, by appropriate provisioning of multi-core job slots, or provisioning the "bare-metal" machine. In either case it becomes the responsibility of the application to optimise the use of the cores available. Various strategies could be available for that, …"

ie VO

Science & Technology Facilities Council

Multi-core, whole node, VMs

- *May* well be easier to schedule in virtualised IaaS cloud

Outlook for WMS

“ Since the demands on a site are now relatively straightforward with the widespread (and almost ubiquitous) use of pilot jobs, it is clear that we may envisage a simplification of the services needed to support submission of tasks to a computing site.

- There will be no requirement for the gLite WMS in future. The remaining use cases will be migrated as soon as possible.
- A CE at a site could be simplified to the key functions: job submission to the local batch system; connection to the grid accounting; logging and “control” of users. If common pilots jobs were used, a site “CE” could be a simple pilot submitter.
- Introduction of cloud interfaces, could eventually replace a grid CE. WLCG should agree a common (subset) of possible interfaces, such as “EC2-like” , or agree to use a common abstraction library, that interfaces to common cloud management software.
    - Experiments need to be able to use such interfaces for any potential opportunistic resources – so since many sites are introducing cloud software, why not just use them directly?
    - Use of VM’s will allow whole-node (or multi-core) scheduling in a simple way.
- Reduced required for a traditional batch system (for LHC) with VMs and pilots jobs. Such uses could certainly be simplified, and avoid scaling limitations. ”

Outlook for WMS

- The writing is on the wall
- No longer needed by main LHC Vos
- Remaining question is what is best replacement for small/non-LHC VOs

Which means…

- New batch system > easier support for multi-core & whole node
- Also makes dynamic changes to resources easier
- IaaS tools for infrastructure management are interesting to us separately
    - Cloud platform is useful for internal testing & development
    - Active interest in GridPP & WLCG (& EGI) cloud work
    - If/when VOs demand can move pledged resources to cloud (or have both interfaces to same resources)
    - When we are happy with performance cost we could just run batch system on IaaS cloud. (NOT saying when that might be.)

Which means…

- ARC CE already appears easier to run - but requirements may become simpler
- Cloud interfaces with VOs supplying images leave us with simpler hypervisors to maintain
    - The parts we maintain become even more interchangable
    - Smaller VOs may need support from someone to develop & maintain images
- We have to do some work to support those who currently depend on WMS

Storage

- Until now it has been unavoidably complicated.
- Disk servers are not interchangeable.
- We have to respond promptly when there are problems.
- We do this pretty well, but it is expensive
- Next generation storage research project told us that where we are (Castor) is good - for now.

Storage

- Disk & tape separation opens up options for disk
- Will naturally evolve in coming years
- My hunch - migration to non-domain specific solution in 2-5 years
    - Not possible to say what solution
    - Actively investing HDFS & ceph for other use cases
- One aim would be more interchangeable storage units making OOH easier - unlikely to be cost free though

And so…

- Cannot know how things (VO requirements, available technology) will develop
- Maintain & improve our ability to adapt & respond
- We need to continue to test/prove & develop new ways of running things.
    - New batch system (more or less) complete
    - Less & simpler Grid interfaces
    - IaaS cloud & exposed cloud interfaces
    - New storage system
        - More interchangeable components
    - ...who knows exactly...