Lightweight sites: computing resources

CHEP 2016

Alessandra Forti, Maarten Litmaath
on behalf of WLCG Ops Coordination

v1.1

Rationale

One of the goals of WLCG Operations Coordination activities is to help simplify what the majority of the sites, i.e. the smaller ones, need to do to be able to contribute resources in a useful manner, i.e. with large benefits compared to efforts invested.

Classic grid sites may profit from simpler mechanisms to deploy and manage services. Moreover, we may be able to get rid of some service types in the end.

New and existing sites may rather want to go into one of the cloud directions that we will collect and document.

There may be different options also depending on the experiment(s) that the site supports.

There is no one-size-fits-all solution. We will rather have a matrix of possible approaches, allowing any site to check which ones could work in its situation, and then pick the best.

Introduction

- The May GDB afternoon had a session on lightweight sites
  - A summary is available here
- The introduction defined boundary conditions
- In particular, here we are not (directly) concerned with storage and data access
  - Those aspects fall under the Data Management Coordination initiative approved in the Sep MB
- Instead, we focus on what it takes to provide computing resources to the experiments

How to enable computing

- Services that currently are or may be needed to enable computing at a site:
  - Computing Element
  - Batch system
  - Cloud setups
  - Authorization system
  - Info system
  - Accounting
  - CVMFS Squid
  - Monitoring
- Storage service deployment may also profit from generic simplifications pursued here

OSG and EGI

- In OSG every WLCG site mainly supports just a single LHC experiment
- The sites are managed in close collaboration with the US project in each experiment
  - US-ATLAS, US-CMS, US-ALICE
- Both US-ATLAS and US-CMS are already working on lighter ways to provision their resources
  - Ubiquitous Cyberinfrastructure, Virtual Clusters
  - Tier-3 in a box, Pacific Research Platform
- In EGI the situation is a lot more complex
  - Multi-experiment sites, many countries/cultures/projects/…, more MW diversity, experiments have less influence, …
- Here we should focus on the EGI sites then
  - While learning from the OSG sites

T2 vs. T3 sites

- T3 sites have not signed the WLCG MoU
  - Typically dedicated to a single experiment → can take advantage of shortcuts
- T2 sites have rules that apply
  - Accounting into EGI / OSG / WLCG repository
  - EGI: presence in the info system, at least for Ops VO
  - Security regulations
    - Mandatory OS and MW updates and upgrades
    - Isolation
    - Traceability
    - Security tests and challenges

T2 simplifications

- Reduce the catalog of required services, where possible
- Replace classic, complex services with new, simpler portfolio
- Simplify deployment, maintenance and operation of services

- Some sites could offer a partial portfolio, e.g. just “worker nodes”
  - Need to get their contributions properly recognized

Cloud systems

- Tap directly into local cloud deployments at sites
  - But OpenStack etc. may not be so easy for them either
- Paradigm shift: batch slots → VM instances
  - Need to ensure proper accounting
- Experiments can handle this today, but:
  - would like to see less variety, fewer interfaces to deal with
  - do not want to become sysadmins of the acquired resources
- Other supported VOs may be unable to use such resources
  - Sites may still need to run their classic setups instead
    - Or even in parallel

Configuration

- Slow but steady move towards Puppet
  - YAIM still being used for many services at many sites
  - Some prefer Ansible, CFEngine, Chef, Quattor, Salt, ...
    - But they presumably know what they are doing!
- Small sites also need help there
- Shared modules?
  - Not a lot of evidence yet?
- DPM comes with self-contained mini Puppet
  - An idea for other services?

Simplified deployment

- Can we provide VM images that require little local configuration?
  - The easiest test case may well be the Squid for CVMFS
  - Complex services would benefit more
  - We can at least get their package contents right !
- Would containers be better?
  - Could even work at sites without a (compatible) cloud infrastructure
- Ready-made (HW +) SW solutions?
  - Deployed in a DMZ like perfSONAR
  - Remotely operated by experts

Next

- From the GDB session summary we can get ideas on where to put effort
  - Mind that nothing comes free!
- To identify what might help sites the most, we have prepared a questionnaire
- Any question will come with some context explaining the implications of a particular technology choice or operations model
  - We cannot merely shift the work from T2 sites to the experiments, CERN-IT or T1 sites
  - Instead we seek an overall reduction of complexity and operational costs, benefiting the sites in particular

Technical questions (1)

1. **Does your site support a single LHC experiment?**
2. **Must your site support other VOs on the same resources?**
3. **Does your site require classic grid services?**
4. **Consider reduction of the required services or phase space?**
5. **Might your site benefit from a repository of OpenStack images?**
6. **Might your site benefit from a repository of Docker containers?**
7. **Might your site benefit from a repository of Puppet modules?**
8. **Might your site benefit from install / configuration wizards?**

Technical questions (2)

9. **Could your site supply WNs dedicated to the experiment(s)?**
   - If the WN resources used by the experiment(s) do not have to be shared with other VOs, they could be made to join a central HTCondor pool at CERN and directly receive jobs from there.
   - In this scenario there is no need for a local batch system, nor a CE or other auxiliary grid service, except for a local CVMFS Squid.
   - The WNs just pop up under control of the site. They can be in the form of physical machines, VMs or containers.
   - Proper accounting and experiment shares per site would be implemented in the central pool.

Technical questions (3)

10. **Could your site run VM images supplied by WLCG or the experiments?**
    - Joining the central HTCondor pool at CERN (cf. Instant Glidein) or receiving work directly from experiment task queues.
    - Started e.g. through Vac or Vcycle.

11. **Could your site use Vac-in-a-Box with self-contained hypervisors?**
    - The first Vac-in-a-Box hypervisor is installed with a pre-configured image and then subsequent machines install from it without manual intervention.
    - Each hypervisor machine runs all required services including DHCP, TFTP, Vac and even a Squid for CVMFS.
    - Configuration is obtained from a central website, allowing site admins to control experiment shares etc.

Technical questions (4)

12. **Allow remote access to your local cloud infrastructure?**
    - Remote instantiation of VMs has operational scalability drawbacks compared to the previous options.

13. **Allow remote access to a DMZ for the experiment(s)?**
    - Also this option has a scalability concern.

The questionnaire has been launched on Oct 4th and its results will help to select where to put our efforts.

Conclusions and outlook

- The WLCG ecosystem is complex
- Many small sites currently need to invest efforts that are not commensurate with their size nor available funding
- Multiple areas are being investigated to allow small sites to become more lightweight
- Sites are envisaged to be able to pick the best choice from a matrix of solutions
- WLCG should thus evolve toward increased flexibility and sustainability