

International Context and September Computing Model Workshop

Andrew McNab
University of Manchester
DUNE, GridPP





Overview

- International Context
 - WLCG, Tiers, HTC vs HPC
 - Funding models
 - Need for a Computing Model
- Computing Model Workshop, Sept 9-11(+12), FNAL
 - “Shopping List” of requirements to agree this week as input to September

WLCG, OSG, EGI

- EGI has evolved from EU DataGrid c.2001
 - In principle used by many disciplines, but bulk of the workload is still HEP
- LHC Computing Grid started to guarantee meeting the needs of the LHC experiments
 - “Worldwide” added shortly after
- In theory, WLCG = EGI + OSG + others
- Some things are shared, but often people talk of WLCG as equivalent to OSG
 - “CMS uses WLCG and OSG” etc



WLCG structure

- Management Board with representatives from the four LHC experiments and the national partners (eg STFC in the UK)
 - DUNE now invited to attend as an observer/associate
- Grid Deployment Board is a monthly “Town Hall” meeting open to everyone
 - Talks about topical issues; proposed new technologies
 - Next meeting is at FNAL during our September workshop
 - Mostly meets at CERN
 - Experiments and major sites well represented
- Many working groups and task forces to co-ordinate new developments or rollout of new technologies
 - eg Authorization Working Group



WLCG sites

- Sites are classified as Tier-0 (CERN), Tier-1, or Tier-2
 - “Tier-3” is used by some experiments for local resources or for opportunistic use of a site
- Two things make a Tier-1 site
 - 24/7 support (ie an alarm ticket causes someone on duty to get a text message at any time of day or night)
 - A tape archive
- Within a particular country, Tier-1s are (usually) the largest sites, but some Tier-2s are larger than smaller Tier-1s elsewhere
- WLCG partners make pledges of CPU/Disk/Tape to LHC experiments for their sites
 - Accounting and Availability/Reliability monitoring are then used to verify that pledges are delivered

HTC vs HPC vs Cloud

- Almost all sites are (or present as) conventional High Throughput Computing grid sites
 - Lots of independent batch jobs and lots of fast storage
- However, some countries are offering High Performance Computing sites, perhaps asking to include it in their pledge
 - The US for several years, now also Italy, Brazil, ...
 - Fast interconnects? No local disk? GPUs? Limited/no outbound networking? KNLs? No CernVM-FS?
- Some countries (eg IRIS in the UK) are offering cloud resources to user communities
 - eg an experiment gets access to the OpenStack API rather than the site just using OpenStack internally



Funding

- Different countries/projects have different funding models
- Some, like GridPP, have been very “economic”
 - Any site can run jobs for any approved experiment
 - Each round of funding handed out pro-rata to each site based on qualifying work done and storage provided
 - This creates a marketplace of sites trying to attract jobs and experiments trying to have reliable jobs
- Some, like IRIS also in the UK, are very top down
 - Experiments, telescopes etc apply for CPU/storage
 - The case needs to include some elements of a Computing Model to justify how the resources will be used
 - So we need some kind of evolving (proto)DUNE CM **already**



Computing Model Workshop, 9-11(+12) Sept

- If this Data Model workshop is “what we need to do”
 - Then the CM workshop is “how we will do it”
- We need to take the resource requirements produced this week ...
 - ... and map them onto the WLCG, HPC, Cloud etc etc resources that are/will be available
 - ... and we need to discuss what technologies (RUCIO etc) we need to adopt or adapt or create to do that
 - ... and we need to map the proposed workflows onto that
 - eg if we wanted to reprocess data every year, would that mean campaigns to stage raw data off tape? Is that feasible with the sites we will have?

Computing Model workshop week

Monday	Tuesday	Wednesday	Thursday	
<p>Session 1 9:00 - 12:30</p>	<p>Session 3 9:00 - 12:30</p>	<p>WLCG Grid Deployment Board 9:00 - 15:40</p>	<p>Resource Board? (structure/ role)</p>	<p>FIM4R</p>
<p>Session 2 13:30 - 17:30</p>	<p>Joint session with WLCG Authz WG 13:30 - 17:30</p>		<p>Operations? (including staff effort)</p>	
		<p>Session 4 16:00 - 17:30</p>		

CM agenda will appear at <https://indico.fnal.gov/event/21231/>



Registration for September workshops

- Computing model workshop
 - <https://indico.fnal.gov/event/21231/>
 - **Registration deadline Friday 16 August: see “Fermilab - General Info” on the Indico event’s minisite for FNAL site access requirements**
- Same for GDB: <https://indico.fnal.gov/event/21232/>
 - GDB agenda at <https://indico.cern.ch/event/739882/>
 - Authz preGDB at <https://indico.cern.ch/event/739896/>
- Same for Federated Identity Management 4 Research
 - Agenda/registration <https://indico.fnal.gov/event/21374/>



Requirements from the Data Model workshop

- Data volume/rate estimates
- Processing CPU requirements including Working Group productions
- Processing storage access requirements
 - eg how much data needs to be accessed + how often
- Monte Carlo CPU requirements
- User-initiated analysis CPU + storage requirements
- Quality of Service for sites (eg any Tier-1 QoS required?)
- Job geometries: hours x procs x memory x local disk
- CPU features (AVX etc); GPUs; ???
- Other things that come up during this week



Summary and next steps

- We're expanding (being welcomed!) into the wider WLCG environment
- WLCG itself is becoming more varied
 - And other projects like IRIS in the UK are appearing
- The Computing Model workshop in September gives us time to discuss how to implement the Data Model we produce this week
 - **Note registration deadline of Friday this week**
- During this week, please try to bear in mind the inputs we need for September
 - Don't use up time on the CM details though - we'll do that at FNAL