



Pre-GDB on HPC Resource Utilization

Summary

Jakob Blomer (CERN)

GDB, 14 February 2018

Scope and Goal

- HPC systems have been among our resources for 5+ years, with a growing share
 - From CernVM-FS experience, several issues were identified during the last couple of years
 - Some *technical*, e. g. scale, platform issues
 - Some *policy*, e. g. connectivity, availability of kernel features
- From our perspective: many successful but singular projects, unlike grid sites, HPC sites and experiments lack a common forum
- I Bird suggested a pre-GDB to gauge the interest

- Invited HPC sites *and* LHC experiments
- Half day [▶ agenda](#)
- ~20 participants on site plus ~20 remote participants

HPC Center/Machine	x86 cores	Phis	GPUs	Presentation
NCSA Blue Waters (US)	360k	-	-	▸ Slides
CSCS Piz Daint (CH)	110k	-	5k	▸ Slides
NERSC Cori (US)	86k	10k	-	▸ Slides
TACC Stampede 2 (US)	80k	4k	-	▸ Slides
SDSC Comet (US)	46k	-	0.3k	▸ Slides
IT4I Salomon (CZ)	24k	-	0.5k	▸ Slides
LNCC SDumont (BR)	18k	0.1k	0.4k	▸ Slides
UFreiburg NEMO (DE)	15k	-	-	▸ Slides

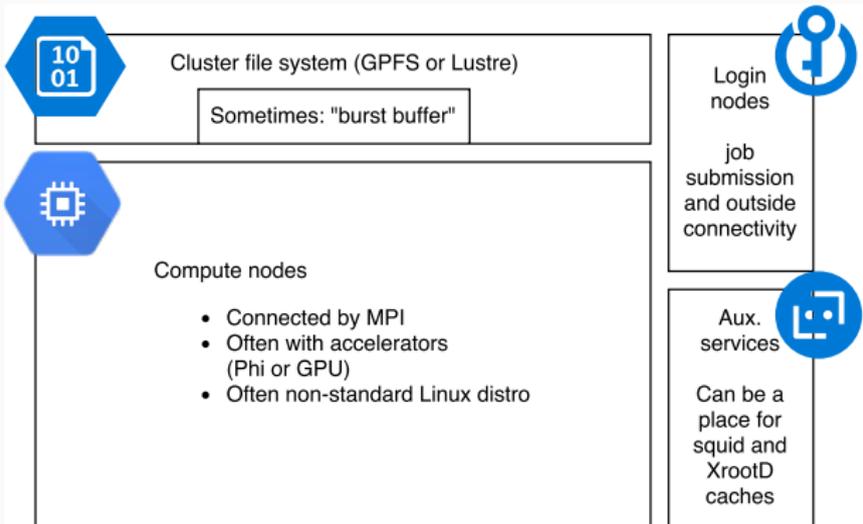
(Note: These are just the sites presented at the pre-GDB. There are more, and large facilities that are used by LHC experiments.

Notable sites are, for instance, **OLCF/Titan**, **ALCF/Theta**, **SuperMUC**.)

Experiment perspectives from [▸ ATLAS](#), [▸ CMS](#), and [▸ LHCb](#)

- Mostly Geant4 simulations, first steps towards reconstruction
- There are multiple execution modes in use:
core-hour grants, backfilling, *part of the grid pledge*
- Spikey resource availability

- ATLAS: 5 % to 10 % of simulations
- CMS: "use HPC for standard operations in Run3 and to rely on significant contribution from HPC for Run4. [...] Need to work with WLCG to properly acknowledge HPC resource use."
- Largest computing growth rate expected for supercomputers
Example: NERSC plans for $> \times 10$ increase in FLOPS by 2024
(Note: the HPC growth is not exclusively in the x86 cores)



- There is a wide spectrum: from ported code (e. g. Mira) to transparent grid site (e. g. Comet)
- Search for integration of HPCs into resource pool in a cost effective way both for HPC sites and experiments

1. Remote job submission via SSH (ARC CE, Bosco)
ATLAS is starting to deploy Harvester as a way for remote job submission.
2. Event-level granularity for job preemption
3. Data ingress and egress through SSH,
many HPC sites use gridftp servers for data transport in and out of the centers.
4. HPC compute node internet connectivity impedes remote I/O
5. Shifter container technology for code distribution
(efficient cvmfs fanout to container image needed)
6. Non-standard cvmfs provision techniques, e. g.
special cache layout, partial tarballs unpacked onto shared file system
7. Accelerators difficult to exploit

- HPC share expected to contribute significantly by run 4
- HEP has little influence on HPC policies and purchases
→ a common voice would help
- A minimum common denominator between experiments and HPC sites would help for *cost effective* exploitation of supercomputers

Suggestions

1. An expert forum for the exchange between WLCG and HPC sites, starting with a mailing list
2. Documentation of existing approaches, e. g. on HSF web site, on WLCG/HSF workshop
3. Nomination of coordinators for follow-up efforts in U.S. and Europe; interest in a working group has been expressed