



Pre-GDB on HPC Resource Utilization

Jakob Blomer

February 02, 2018

Welcome and Introduction

Supercomputers

- Large clusters with accelerators, cluster filesystem, fast interconnect
- Restricted area (most often): limited TCP/IP and Internet connectivity, user-land only, limited Linux environment, accelerator heavy
- Majority of machines is Intel based
- Surprisingly different from the HTC world, not genuinely made for HEP software
- Differences from machine to machine, from porting all code (Mira) to “virtual grid site” (CSCS, Comet)

Why we are interested in supercomputers

- Massive (theoretical) compute power, largest machines $>$ WLCG
- We can backfill and increase the machine's utilization
- We might be encouraged to use supercomputers by funding policy

- We come from the CernVM-FS perspective
 - Several issues identified during the last couple of years
 - Some technical, e.g. scale, platform issues (“can’t”)
 - Some policy (“won’t”)
 - We assume that is similar for other middleware building blocks
 - Data ingres and egress
 - Job management: preemption, unit of scheduling
 - Identity management
 - From our perspective: many successful but singular projects
- Create a forum for experiment experts and HPC site experts
e.g. mailing list, wiki/faq, list of contacts of other sites
- Provide input on HPC resources to WLCG

1. Which resource share comprises HPC now and in the future?
2. Are HPC systems part of the pledge (will they become part of the pledge)?
3. What sort of experiment tasks can/should we run on supercomputers?
4. Are there uncommon infrastructure pieces/features in the middleware that HPC systems critically rely on?
5. Do the HEP software stacks need to assimilate, e. g. MPI based job scheduling?