EU HPCs for ATLAS – updates since March SW week

Rod Walker, Andrej Filipcic

SuperMUC proposal

- Invited to submit application for MUC time
  - medium project <= 10M core hrs
- Application ready
  - already looks like a long shot
    - “But this is a serial application!”
    - at very least we need validated athenaMP(whole node)
- No outbound IP – need ARC CE
  - admins also resistant to special edge-node requirements

C2PAP

- Joint ATLAS-Astro cluster attached to SuperMUC
  - 2048 cores: same hw, IB, sw as SuperMUC
    - our nodes have local WN disk, more RAM(4GB/core)
    - can have serial queue, limited outbound IP, cvmfs on WN
    - ARC CE edge service agreed, but still pressure to use GT5
      - LRZ leads Globus Europe project!
  - Commissioning due end of May
  - Not much resources, but...
    - half way testing and potential way-in for SuperMUC usage
    - comes with 2FTEs for 4 years

C2PAP Manpower

- 1 admin plus 2 applications people each (HEP/Astro) for 4yrs – start Jul+Sep
- Role is rather vague, so flexible
  - both successful candidates have strong computing background, one is from ATLAS
    - parallelize ATLAS or G4 code, vectorization, event engine – i.e. contribute to Athena or G4
    - other suggestions and their interests taken into account

MPI Hydra

- MPI had to have their own new supercomputer
  - Hydra similar to SuperMUC
    - a little smaller and with GPUs too
- More helpful/keen than LRZ, so more progress
  - offer serial queue, but no outbound ip, no cvmfs
    - installing ARC CE (really underway)
    - plan to rsync cvmfs to gpfs
      - then link from /cvmfs on WNs (so no relocation problems)

Other

- CEA investigations suspended pending FR T1 tender
- Various inquiries and offers after the ATLAS weekly talk – not followed up
  - maybe wiki and a mailing list is the way forward
- MPI Hydra is current least awkward site
  - I (Rod) will work with the least awkward of the day
- CSCS – 36k core Cray XC30 (750TF)
  - Discussions, but not encouraging

Scandinavian HPCs

- Abel, Oslo, NO:
  - 11k (22k) core, 4GB mem/core, SLURM, part of NDGF-T1
  - Since 2004, big efforts to tune shared FS (GPFS, FhGFS)
  - Positive experience gained from ATLAS jobs to optimize for heavy I/O and heavy memory operations
  - 1k core pledged, up to 4k cores opportunistic (re-queue)
- Abisko, Umea, SE:
  - 16k core, 2GB/core, SLURM, whole socket scheduling
  - Purely opportunistic, few k cores
  - Available to ATLAS from summer
- In both cases, opportunistic usage is efficient with up to few h jobs.

Overall

- 10-20k (semi) opportunistic EU cores, a significant contribution to ATLAS resources, probably much more in the upcoming years
- Wildly varying site policies
  - From full external connectivity, cvmfs/nfs, large I/O jobs
  - To limited or no connectivity with synced/relocatable cvmfs copy
- In general:
  - WN input staging, output delivery not permitted – all HPCs rely on shared FS and little or no local disk
  - No grid software on WNs
  - ARC CE external batch filler and stager seems to better suite the site policies
  - Serial scheduling rarely an option – whole socket, whole node or even entire partition job allocations need to be dealt with

Needs on ATLAS side

- ARC CE more or less fits the site requirements
  - running as non-privileged user (tested)
  - Custom batch support – extending ARC backends
  - Custom OS support – difficulties mostly from globus, lfc, voms dependencies
- In some cases, a frontend server behind the firewall → on-site agent needed
  - Extending arcControlTower