CERN

HPC Service status and plans

Carolina Lindqvist, Philippe Ganz, Nils Høimyr

IT/CM

20/10/2016

High Performance Computing (HPC)

- **Applications and use cases that do not fit the standard batch HTC model. Typically parallel MPI applications**
  - Theory Lattice QCD studies (TH)
  - Accelerator physics, beam simulation, plasma simulations... (BE, TE)
  - Computation Fluid Dynamics, CFD (EN, EP, HSE)
    - Additionally structural analysis, field calculations (EN,PH,TE), currently mainly on Windows fat boxes (run by colleagues in IT/CDA)
- **Job duration often very long, (e.g. several weeks for CFD and QCD)**
  - Stability of OS and environment critical
  - MPI application performance require fast interconnects with low latency between nodes in a cluster. Some applications also require fast access to shared storage

Current “HPC” facilities in IT

- lxbatch resources (SLC6) running MPI over LSF:
  - eng, cfd (Xeon 12 core low latency 10Gb)
  - Spacecharge Large batch nodes running MPI in one box
- New Theory QCD Infiniband cluster
  - 72 Quanta E4 16 core / 64Gb
  - Infiniband FDR interconnects
  - Puppet HG ithpc/linux/qcd, soon bi/hpc/qcd
- Dedicated (recent) clusters for TE plasma simulations and HSE CFD
  - Quanta 16 core / 128Gb with low latency 10Gb ethernet
  - Puppet HG ithpc/linux, nodes running CC7
- Windows Engineering HPC service (Ansys, Comsol etc, IT-CDA)

HPC service - challenges

- Special hardware (Infiniband interconnects, storage requirements)
- Sensitive environment stack with respect to system updates
- Limited MPI and HPC competence (need to draw on expertise in the group and department)
- Application competence often missing in the user community
  - HPC application level support very expensive
  - Go as far as possible with documentation and examples
- Central batch HPC resources vs special use cases (e.g. TH)
  - **Funding** for a general HPC facility only next year
  - Batch service currently in transition (LSF to Condor)

HPC service - roadmap

- TH cluster – released to users
  - Further integration with batch in progress
- Investigate options for HPC with HTCondor (Tech student)
  - Condor parallel universe? Other batch scheduler?
    - MPI job submission with HTCondor?
    - Backfill with HTCondor on HPC resources
- Build up initial batch HPC with existing hardware and extend with new Infiniband cluster in 2017
  - JIRA: project BBC, component: HPC
  - Project outline: http://cern.ch/batchprojects

HPC project - people

- Nils (lead, service management)
- Technical student Philippe Ganz
  - Focus on MPI performance, job submission, batch scheduler and HTCondor evaluation and application launch templates
- Fellow Carolina Lindqvist
  - Focus on service development, notably integration with the batch service and other IT infrastructure, monitoring and scalability

HPC – related stuff

- Windows HPC for Engineering applications
  - Windows HPC cluster operated by CDA
  - Possible use of Linux HPC back-end in the future
- Applications running requiring fat boxes
  - Memory requirements, e.g. 512 Gb to 1 Tb
    - Engineering applications (Ansys HFSS, CST...)
    - Microelectronics
    - Other use cases that cannot scale with MPI across clusters for application or licencing reasons
  - Requests for such resources will come our way – how to provision them?

Current installation – TH Cluster

- Puppet-managed installation
  - CephFS-based home directories for users.
  - Modules for OFED/Mellanox driver and MPI installations.
  - Basic monitoring using Lemon sensors and Kibana dashboards.
- Current work
  - Preparing the configuration to be applied on the whole new cluster.
  - Using OSU benchmarks to compare performance of MPI on old cluster vs. new configuration.

Current installation – Batch cluster

- Also a puppet-managed installation
  - Users' home directories on AFS.
  - Reusing module for OFED driver and MPI installations.
  - Reusing monitoring components.
- Current work
  - Preparing the configuration to be applied on the whole new cluster, and using OSU benchmarks to check MPI performance.
  - Enabling the parallel universe in HTCondor and configuring the nodes as HPC worker nodes which accept HPC jobs.

Questions?

CERN

www.cern.ch