Slurm Roadmap

Versions 16.05 and beyond

Morris Jette, Danny Auble (SchedMD)
Yiannis Georgiou (Bull)

Slurm User Group 2015

Version 16.05

- Version 16.05 to be released in May 2016
  - Maintaining 9 month release cycle
- Some key features already determined
- Other possible features still under discussion
  - Optimize agility

Federated Cluster Support

- Provides enterprise-wide workload management
  - Manage work-flow through various phases on different hardware: pre-processing, computation, post-processing, visualization
  - Optimize system utilization and responsiveness
  - Can be configured to support ultra-high throughput
- Separate presentation with more details

Intel Knights Landing Support

- Improved controls over task layout
  - 60+ cores with 2-D mesh interconnect
- User control over NUMA memory and MCDRAM configuration
  - Node re-boot required for configuration changes

PMIx Version 2.0 Support

- Improved MPI scalability and performance
- https://www.open-mpi.org/projects/pmix/

Support for heterogeneous resources

- These developments have as goal to extend the job specification language of SLURM to support MPMD (Multiple Program Multiple Data)
- With the support of heterogeneous resources, the idea is to introduce a new type of jobs named job packs which will be described by a set of pack groups, each pack group having the same resources requirements.
- Examples of executions illustrating the targeted capability :
  - srun -n 2 -c2 ./app1 : -n 4 --mem-per-core 256 --gres=gpu:2 ./app2
  - Or
  - sbatch -n 2 -c 2 : -n 4 --mem-per-core 256 –gres=gpu:2 ./script.sh
  - cat script.sh
    - srun --pack-group 0 ./app1 : --pack-group 1 ./app2
    - srun –pack-group=[0,1] ./app

Energy fairsharing

- Provide incentives to users to be more energy efficient
  - Based upon the energy accounting mechanisms and normal fairsharing
    - Accumulate past jobs energy consumption and align that with the shares of each account
    - Implemented as a new multi-factor plugin parameter in SLURM
    - Energy efficient users will be favored with lower stretch and waiting times in the scheduling queue

Yiannis Georgiou, David Glesser, Krzysztof Rzadca, Denis Trystram
A Scheduler-Level Incentive Mechanism for Energy Eciency in HPC
(In proceedings of CCGRID 2015)

Beyond 16.05

- Scheduling
  - Scheduling within Energy Constraints
  - Multi-objective resource selection
  - Machine Learning Optimizations
- Hybrid environments
  - Deploy Big Data workflows and Cloud environments upon HPC clusters
- Enable SLURM simulation in very large scales
  - Proposal for flexible solution enabling comparisons with new generation RJMS

Questions?