



# SLURM Project Report

SLURM User Group Meeting  
October 9-10, 2012  
Barcelona, Spain

Morris Jette and Danny Auble  
[jette,da]@schedmd.com

# Agenda



- Status of version 2.5
- Plans for version 2.6
- Plans for later releases
- Other news

# Version 2.5 Contents



- Record for power consumption by job (Bull)
  - New *energy\_accounting* plugin infrastructure
  - Details in Wednesday presentation
- User control over CPU frequency (Bull)
  - New *srun --cpu-freq* option
- Added ability to reserve all nodes in a partition (Bull)
  - Reservation updated when nodes added to or removed from a partition
- Modified *sinfo* to report reservations (Bull)
  - New *sinfo --reservation* option

# Version 2.5 Contents



- Advanced reservation of cores, not only whole nodes (BSC)
  - With *select/cons\_res* plugin only
  - Not currently available for BlueGene systems
- Node *CPU\_Load* information available (SchedMD)

# Version 2.5 Contents



- Significant performance improvements (SchedMD)
  - Throughput up to 630 jobs per second
  - Details in Wednesday presentation
- Integration with IBM Parallel Environment (SchedMD)
  - New *launch* plugin infrastructure
    - New *launch/slurm*, *launch/poe*, *launch/runjob* plugins
  - New *switch/nrt* plugin
  - Details in Wednesday presentation

# Version 2.5 Contents

- Partition parameter *MaxCPUsPerNode* (NVIDIA/SchedMD)
  - Useful to reserve some CPUs for use with GPUs

```
# Excerpt from slurm.conf
#
JobSubmitPlugins=limit_gpu_use_by_partition # Site-specific script
#
NodeName=tux[1-128] CPUs=12 Gres=gpu:1
#
PartitionName=cpu Default=yes Nodes=tux[1-128] MaxCPUsPerNode=10
PartitionName=gpu Default=no Nodes=tux[1-128] MaxCPUsPerNode=12
```

# Version 2.5 Status



- Development largely complete
- Moving to test mode now
- Planning release in November

# Version 2.6 Plans




- Release 2nd quarter 2013
  - Continue with major release about every 6 months



# Version 2.6 Contents


## (Preliminary)



- Scheduling optimized for energy efficiency (Bull)
  - Based upon infrastructure/hardware power limits and job energy needs
  - Temperature aware
  - Details in Wednesday presentation
- License Management integration with FlexLM/Flexnet Publisher (Bull)

# Version 2.6 Contents

## (Preliminary)



- Integration with MapReduce (Greenplum/EMC/SchedMD)
  - Orders of magnitude performance improvement
  - Details in Wednesday presentation
- Support for Intel MIC (Many Integrated Core) processor (Bull/SchedMD)
- Finer-grained BlueGene resource management (SchedMD)
  - Partitions/queues and advanced reservations containing less than a whole midplane

# For Later Releases...

- Scalability and Throughput Improvements (Bull)
- Kerberos support for authentication (Bull)
- Multi-parameter scheduling (Bull)
  - More control over various limits and optimizations (power use, temperature, network topology, etc.)
- Virtualization Cloud Computing (Bull)
  - Jobs that deploy virtual machines
- Dynamic Job Integration with MPI (Bull)
  - Support job size changes and fault tolerance

# For Later Releases...



- Improved fault tolerance (SchedMD)
  - Hot spare resources
  - User API to get replacement resources and/or additional time
- Improved job step support (SchedMD)
  - Queuing (eliminate periodic retry)
  - Dependency support

# Other News



- Streamlined installation procedure on Cray systems with RPMs
  - <http://www.schedmd.com/slurmdocs/cray.html>
- Tutorials on YouTube
  - <http://www.schedmd.com/slurmdocs/tutorials.html>