



IBM Research

Parallel Tools Platform (PTP) on Blue Gene/Q

Simon Wail

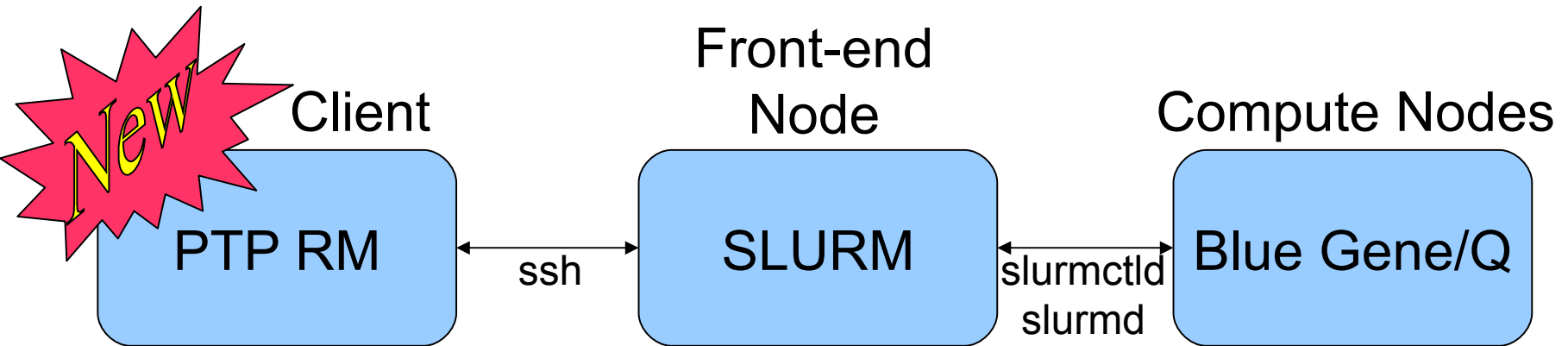
PTP and Blue Gene/Q

- **No support for Blue Gene in PTP until NOW**
- **Will demonstrate resource manager for PTP that interfaces with SLURM scheduler managing Blue Gene/Q**
- **SLURM is used on many HPC systems including Blue Genes**
- **SLURM will be the scheduler of choice for LLNL Sequoia Blue Gene/Q system**
- **SLURM (V2.3) already supports Blue Gene/Q**

System Monitoring with PTP

- **PTP has configurable resource managers to interface with remote HPC systems**
- **Resource Managers perform:**
 - Automatic job script generation
 - Job submission
 - Job and queue monitoring
 - Viewing job output
 - Overall system view
- **Uses “LLView” for graphical view of system load**
 - Developed by Juelich team for Blue Gene/P

PTP – SLURM – Blue Gene/Q



Run Configurations for SLURM – Blue Gene/Q

Run Configurations

Create, manage, and run configurations

[Application]: Project TamboMPIHelloWorld is closed

Name: RemoteMPIHelloWorld

Resources Application Arguments Environment Synchronize Common

Resource Manager: Slurm-BGQ-Batch

Basic SLURM Settings Advanced SLURM Settings Import SLURM Script

Name	Value	Description
Job Name:	<input type="text" value="ptp_sbatch"/>	Specify a name for the job allocation.
Queue:	<input type="text" value="main"/>	Request a specific partition for the resource allocation.
Number of nodes:	<input type="text" value="8192"/>	Request that a minimum of "minnodes" nodes be allocated to this job.
Wallclock Time:	<input type="text" value="1:00:00"/>	Set a limit on the total run time of the job allocation.
MPI Command:	<input type="text" value="srun"/>	Which srun command to use.
MPI Number of Tasks:	<input type="text" value="4"/>	Specify the number of tasks to run.
MPI Ranks per Node:	<input type="text" value="1"/>	Request the maximum ntasks be invoked on each node.

View Script View Configuration Restore Defaults

Apply Revert

Close Run

Filter matched 26 of 26 items

System Monitoring of 96 Rack Blue Gene/Q

System Monitoring - resourceManagers/Slurm-BGQ-Batch.xml - Eclipse

File Edit Navigate Search Project Run Design Window Help

Resource Managers

- Local-OpenMPI (Open MPI)
- Merri-Batch (LML_JAXB)
- Slurm-BGQ-Batch (LML_JAXB)
- Slurm-BGQ-Srun (LML_JAXB)

Active Jobs

step	owner	queue	wall	queuedat	dispatchd	totalcores	status
165	swail(586)	main	3600	2011-10-2	2011-10-2	512	RUNNING
166	swail(586)	main	3600	2011-10-2	2011-10-2	1024	RUNNING
167	swail(586)	main	3600	2011-10-2	2011-10-2	4K	RUNNING
168	swail(586)	main	3600	2011-10-2	2011-10-2	8K	RUNNING
169	swail(586)	main	3600	2011-10-2	2011-10-2	16K	RUNNING
170	swail(586)	main	3600	2011-10-2	2011-10-2	32K	RUNNING
171	swail(586)	main	3600	2011-10-2	2011-10-2	64K	COMPLETED
172	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
173	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
174	swail(586)	main	3600	2011-10-2	2011-10-2	512K	RUNNING
175	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
176	swail(586)	main	3600	2011-10-2	2011-10-2	128K	COMPLETED
177	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING

system: slurm Progress

R00

R000-M1	R001-M1	R002-M1	R003-M1	R004-M1	R005-M1	R006-M1	R007-M1
R000-M0	R001-M0	R002-M0	R003-M0	R004-M0	R005-M0	R006-M0	R007-M0

R01

R010-M1	R011-M1	R012-M1	R013-M1	R014-M1	R015-M1	R016-M1	R017-M1
R010-M0	R011-M0	R012-M0	R013-M0	R014-M0	R015-M0	R016-M0	R017-M0

R02

R020-M1	R021-M1	R022-M1	R023-M1	R024-M1	R025-M1	R026-M1	R027-M1
R020-M0	R021-M0	R022-M0	R023-M0	R024-M0	R025-M0	R026-M0	R027-M0

R03

R030-M1	R031-M1	R032-M1	R033-M1	R034-M1	R035-M1	R036-M1	R037-M1
R030-M0	R031-M0	R032-M0	R033-M0	R034-M0	R035-M0	R036-M0	R037-M0

R04

R040-M1	R041-M1	R042-M1	R043-M1	R044-M1	R045-M1	R046-M1	R047-M1
R040-M0	R041-M0	R042-M0	R043-M0	R044-M0	R045-M0	R046-M0	R047-M0

R05

R050-M1	R051-M1	R052-M1	R053-M1	R054-M1	R055-M1	R056-M1	R057-M1
R050-M0	R051-M0	R052-M0	R053-M0	R054-M0	R055-M0	R056-M0	R057-M0

R06

R060-M1	R061-M1	R062-M1	R063-M1	R064-M1	R065-M1	R066-M1	R067-M1
R060-M0	R061-M0	R062-M0	R063-M0	R064-M0	R065-M0	R066-M0	R067-M0

R07

R070-M1	R071-M1	R072-M1	R073-M1	R074-M1	R075-M1	R076-M1	R077-M1
R070-M0	R071-M0	R072-M0	R073-M0	R074-M0	R075-M0	R076-M0	R077-M0

Console

```
<terminated> RemoteMPIHelloWorld [Parallel Application] Slurm-BGQ-Batch: job 177
```

System Monitoring of 96 Rack BG/Q – Job Highlight

System Monitoring - resourceManagers/Slurm-BGQ-Batch.xml - Eclipse

File Edit Navigate Search Project Run Design Window Help

Resource Managers

- Local-OpenMPI (Open MPI)
- Merri-Batch (LML_JAXB)
- Slurm-BGP-Batch (LML_JAXB)
- Slurm-BGQ-Batch (LML_JAXB)**
- Slurm-BGQ-Srun (LML_JAXB)

Active Jobs

step	owner	queue	wall	queuedat	dispatchd	totalcores	status
165	swail(586)	main	3600	2011-10-2	2011-10-2	512	RUNNING
166	swail(586)	main	3600	2011-10-2	2011-10-2	1024	RUNNING
167	swail(586)	main	3600	2011-10-2	2011-10-2	4K	RUNNING
168	swail(586)	main	3600	2011-10-2	2011-10-2	8K	RUNNING
169	swail(586)	main	3600	2011-10-2	2011-10-2	16K	RUNNING
170	swail(586)	main	3600	2011-10-2	2011-10-2	32K	RUNNING
171	swail(586)	main	3600	2011-10-2	2011-10-2	64K	RUNNING
172	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
173	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
174	swail(586)	main	3600	2011-10-2	2011-10-2	512K	RUNNING
175	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
176	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
177	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING

system: slurm Progress

Console

```
<terminated> RemoteMPIHelloWorld [Parallel Application] Slurm-BGQ-Batch: job 177
```

LML DA Driver (slur...edu.au): (90%)

System Monitoring of 96 Rack BG/Q – Job Highlight

System Monitoring - resourceManagers/Slurm-BGQ-Batch.xml - Eclipse

File Edit Navigate Search Project Run Design Window Help

Resource Managers

- Local-OpenMPI (Open MPI)
- Merri-Batch (LML_JAXB)
- Slurm-BGP-Batch (LML_JAXB)
- Slurm-BGQ-Batch (LML_JAXB)**
- Slurm-BGQ-Srun (LML_JAXB)

Active Jobs

step	owner	queue	wall	queuedat	dispatchd	totalcores	status
165	swail(586)	main	3600	2011-10-2	2011-10-2	512	RUNNING
166	swail(586)	main	3600	2011-10-2	2011-10-2	1024	RUNNING
167	swail(586)	main	3600	2011-10-2	2011-10-2	4K	RUNNING
168	swail(586)	main	3600	2011-10-2	2011-10-2	8K	RUNNING
169	swail(586)	main	3600	2011-10-2	2011-10-2	16K	RUNNING
170	swail(586)	main	3600	2011-10-2	2011-10-2	32K	RUNNING
171	swail(586)	main	3600	2011-10-2	2011-10-2	64K	RUNNING
172	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
173	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
174	swail(586)	main	3600	2011-10-2	2011-10-2	512K	RUNNING
175	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
176	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
177	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING

system: slurm Progress

R00

R000-M1	R001-M1	R002-M1	R003-M1	R004-M1	R005-M1	R006-M1	R007-M1
R000-M0	R001-M0	R002-M0	R003-M0	R004-M0	R005-M0	R006-M0	R007-M0

R01

R010-M1	R011-M1	R012-M1	R013-M1	R014-M1	R015-M1	R016-M1	R017-M1
R010-M0	R011-M0	R012-M0	R013-M0	R014-M0	R015-M0	R016-M0	R017-M0

R02

R020-M1	R021-M1	R022-M1	R023-M1	R024-M1	R025-M1	R026-M1	R027-M1
R020-M0	R021-M0	R022-M0	R023-M0	R024-M0	R025-M0	R026-M0	R027-M0

R03

R030-M1	R031-M1	R032-M1	R033-M1	R034-M1	R035-M1	R036-M1	R037-M1
R030-M0	R031-M0	R032-M0	R033-M0	R034-M0	R035-M0	R036-M0	R037-M0

R04

R040-M1	R041-M1	R042-M1	R043-M1	R044-M1	R045-M1	R046-M1	R047-M1
R040-M0	R041-M0	R042-M0	R043-M0	R044-M0	R045-M0	R046-M0	R047-M0

R05

R050-M1	R051-M1	R052-M1	R053-M1	R054-M1	R055-M1	R056-M1	R057-M1
R050-M0	R051-M0	R052-M0	R053-M0	R054-M0	R055-M0	R056-M0	R057-M0

R06

R060-M1	R061-M1	R062-M1	R063-M1	R064-M1	R065-M1	R066-M1	R067-M1
R060-M0	R061-M0	R062-M0	R063-M0	R064-M0	R065-M0	R066-M0	R067-M0

R07

R070-M1	R071-M1	R072-M1	R073-M1	R074-M1	R075-M1	R076-M1	R077-M1
R070-M0	R071-M0	R072-M0	R073-M0	R074-M0	R075-M0	R076-M0	R077-M0

Console

```
<terminated> RemoteMPIHelloWorld [Parallel Application] Slurm-BGQ-Batch: job 177
```


System Monitoring of 96 Rack BG/Q – Row Zoom

System Monitoring - resourceManagers/Slurm-BGQ-Batch.xml - Eclipse

File Edit Navigate Search Project Run Design Window Help

Resource Managers

- Local-OpenMPI (Open MPI)
- Merri-Batch (LML_JAXB)
- Slurm-BGP-Batch (LML_JAXB)
- Slurm-BGQ-Batch (LML_JAXB)**
- Slurm-BGQ-Srun (LML_JAXB)

Active Jobs

step	owner	queue	wall	queuedat	dispatchd	totalcores	status
165	swail(586)	main	3600	2011-10-2	2011-10-2	512	COMPLETED
166	swail(586)	main	3600	2011-10-2	2011-10-2	1024	RUNNING
167	swail(586)	main	3600	2011-10-2	2011-10-2	4K	RUNNING
168	swail(586)	main	3600	2011-10-2	2011-10-2	8K	RUNNING
169	swail(586)	main	3600	2011-10-2	2011-10-2	16K	RUNNING
170	swail(586)	main	3600	2011-10-2	2011-10-2	32K	RUNNING
171	swail(586)	main	3600	2011-10-2	2011-10-2	64K	RUNNING
172	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
173	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
174	swail(586)	main	3600	2011-10-2	2011-10-2	512K	RUNNING
175	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
176	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
177	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING

system: slurm Progress

R01

R010	R011	R012	R013	R014	R015	R016	R017
R010-M1	R011-M1	R012-M1	R013-M1	R014-M1	R015-M1	R016-M1	R017-M1
R010-M0	R011-M0	R012-M0	R013-M0	R014-M0	R015-M0	R016-M0	R017-M0

Console

```
<terminated> RemoteMPIHelloWorld [Parallel Application] Slurm-BGQ-Batch: job_177
```

System Monitoring of 96 Rack BG/Q – Rack Zoom

System Monitoring - resourceManagers/Slurm-BGQ-Batch.xml - Eclipse

File Edit Navigate Search Project Run Design Window Help

Resource Managers

- Local-OpenMPI (Open MPI)
- Merri-Batch (LML_JAXB)
- Slurm-BGP-Batch (LML_JAXB)
- Slurm-BGQ-Batch (LML_JAXB)**
- Slurm-BGQ-Srun (LML_JAXB)

Active Jobs

step	owner	queue	wall	queuedat	dispatchd	totalcores	status
165	swail(586)	main	3600	2011-10-2	2011-10-2	512	RUNNING
166	swail(586)	main	3600	2011-10-2	2011-10-2	1024	RUNNING
167	swail(586)	main	3600	2011-10-2	2011-10-2	4K	RUNNING
168	swail(586)	main	3600	2011-10-2	2011-10-2	8K	RUNNING
169	swail(586)	main	3600	2011-10-2	2011-10-2	16K	RUNNING
170	swail(586)	main	3600	2011-10-2	2011-10-2	32K	RUNNING
171	swail(586)	main	3600	2011-10-2	2011-10-2	64K	RUNNING
172	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
173	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
174	swail(586)	main	3600	2011-10-2	2011-10-2	512K	RUNNING
175	swail(586)	main	3600	2011-10-2	2011-10-2	256K	RUNNING
176	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING
177	swail(586)	main	3600	2011-10-2	2011-10-2	128K	RUNNING

system: slurm Progress

R000

R000-M1

R000-M0

Console

```
<terminated> RemoteMPIHelloWorld [Parallel Application] Slurm-BGQ-Batch: job_177
```