BG/L System Management

Susan Coghlan
HPC Manager, MCS
BG/L Consortium System Software and
Applications Workshop
April 19-20, 2006

Key BlueGene System Management Areas

- BG/L hardware
  - Replacement
  - RAS events
- Driver and boot images
  - Installation/Upgrades
  - Management
- Service Node
  - bglmaster and discovery processes
  - Network configuration
- Diagnostics
  - Preventative Maintenance (PM)
  - Debugging problems
- Storage
  - Mount management
  - Debugging

Key BlueGene Management Areas continued…

- **DB2**
  - Startup
  - License mgt
  - Standard DB maintenance - backups, etc.
- **Resource Manager**
  - Partition management - definitions, dependencies, state
  - Scheduling policies, reservations, queue mgt
  - Job cleanup
- **MPI**
  - mpirun wrappers
  - mpicc/mpif90/etc. wrappers
- **Compilers and Libraries**
  - Installation/Upgrades

Stability of the system - it's great

- Hardware
  - Ours - 31.5 days MTBI
  - Quick and easy replacement of all nodes, fans and power modules
  - Burst of failures this winter
    - *6 nodes within 40 days (starting 1/5/2006)*
- Replacement of the actual hardware very quick, but, shutdown and restart of midplane very slow and requires that the admin remember to do certain tasks in certain orders
- Software
  - Somewhat stable - we are seeing failure to exit mpirun processes cleanly again
- We spend about 10 hrs a week managing our 1024 node system
  - Most of that is helping users and dealing with oversubscription, not fixing the system

Diagnostics

- Important to run on a regular basis
- Recent improvements very welcome
  - Shorter time to complete base test
  - Easier to run
- Still hard to run a single specific test
  - I'm not positive about that because it is still…
- Poorly documented
  - What are the tests?
  - What do they do?
  - Which ones should be run when?
  - Which failures are 'real' failures?
- I'd like to be able to…
  - Run tests based on the output of other tests
  - Incorporate my tests into the diagnostics framework

DB2 issues

- Useful for storing RAS events, information about the system and even job history
- Required in order to run a job - bad idea
    - If your license disappears, jobs will no longer run
- Problems with discovery of the hardware
    - keeps biting us
- Security!
    - Default password - you can change this, carefully
    - Separation of privileges

Common Problems

- Still no easy way to know when software updates and efixes are available
  - IBM is creating a new document that shows the most current versions
  - Very helpful, but, not easy to automate checks
- Compilers, libraries are still very confusing
- Regression testing after a driver update is not happening
- Everyone is writing their own command line version for searching the RAS event DB
- Too easy to miss a step in the somewhat convoluted driver updates
- Efix install script failures - isn't anyone walking thru the instructions prior to releasing the fix?
- Documentation is still lacking:
  - Steps to shutting down and bringing up the system
  - Changes with a new driver
  - RAS error message translation

Future System Improvements

- Remove the database requirement for the running of jobs
- More efficient and reliable launch and termination of user jobs
- CLI-based tools for monitoring system and user jobs
  - The web browser is nice, but it should be an extra to a solid set of CLI tools
- Better debugging tools
  - user
  - system manager
- Security, particularly on the database side
  - User id management within the mmcs world
- Better tools for the installation and upgrading of the drivers/efixes
  - Validate that everything is installed correctly
- Regression test suite for driver upgrades
- Ability to hook resource managers into the underlying system
- Documentation not an after thought!
- QA on the software

System Manager's Resources

- Mailing lists
  - admins-wg @ bgconsortium.org
  - discuss @ bgl.mcs.anl.gov
- Upcoming activities
  - "Everything you wanted to know, but were afraid to ask, about X" documents
  - Collections of top issues for communication to IBM on a regular basis