Using & Managing HPC Systems

Presenter: William Lu, Ph.D., Marketing Director
Date: October 7, 2008
Challenges on HPC systems

Using and managing HPC systems, organizations are facing some challenges:

**Complex hardware**
- Performance relies on CPU, memory, storage, interconnects,…

**Reliability**
- Commodity hardware has good performance/cost, but…

**Users competing resources**
- Why my jobs are waiting?
What is HPC Management Software?

The software between the application and the operating system...

...that enables users to develop, run and manage compute or data intensive applications and manage HPC systems.
4,000,000 Managed CPUs
2,000 Customers worldwide
500 Employees in 15 offices
16 Years of profitable growth
1 Leader in HPC
Industries Served by Platform

- CERN
- DoD, US
- DoE, US
- ENEA
- Georgia Tech
- Harvard Medical School
- Japan Atomic Energy Inst.
- MaxPlanck Inst.
- MIT
- SSC, China
- Stanford Medical TACC
- U. Of Georgia
- U. Tokyo
- Washington U.

Financial Services
- BNP
- Citigroup
- Fortis
- HSBC
- KBC Financial
- JPMC
- Lehman Brothers
- LBBW
- Mass Mutual
- MUFG
- Nomura
- Prudential
- Sal. Oppenheim
- Société Générale
- Airbus
- BAE Systems
- Boeing
- Bombardier
- Deere & Company
- Ericsson
- Honda
- General Electric
- General Motors
- Goodrich
- Lockheed Martin
- Nissan
- Northrop Grumman
- Pratt & Whitney
- Toyota
- Volkswagen
- Agip
- BP
- British Gas
- China Petroleum
- ConocoPhillips
- EMGS
- Gaz de France
- Hess
- Kuwait Oil
- PetroBras
- Petro Canada
- PetroChina
- Shell
- StatoilHydro
- Total
- Woodside

Oil & Gas

Electronics
- AMD
- ARM
- Broadcom
- Cadence
- Cisco
- Infineon
- MediaTek
- Motorola
- NVidia
- Qualcomm
- Samsung
- Sony
- ST Micro
- Synopsys
- TI
- Toshiba

Life Sciences
- Abott Labs
- AstraZeneca
- Celera
- DuPont
- Eli Lilly
- Johnson & Johnson
- Merck
- National Institutes of Health
- Novartis
- Partners Health Network
- Pharsight
- Pfizer
- Sanger Institute

Other Industries
- AT&T
- Bell Canada
- Cingular
- DreamWorks Animation
- SKG
- GE
- IRI
- Telecom Italia
- Telefonica
- Walt Disney Co.
Solutions with Partners

Platform OCS 5 and Platform Manager integrated in Dell cluster systems

Platform LSF, Platform Manager form key parts of Unified Cluster Portfolio

Platform enterprise solutions support a wide range of IBM HPC systems

Platform delivers first certified Intel® Cluster Ready solution, Platform OCS 5

Integrates Platform LSF and Platform Symphony in grid solutions

Platform OCS 5 powers the Red Hat® HPC Solution

OEMs Platform’s core technology in SAS® applications
Platform HPC Management Solutions

Platform MPI
Platform LSF
Platform LSF MultiCluster
Platform LSF License Scheduler
Platform LSF Session Scheduler
Platform Process Manager
Platform EGO
EnginFrame

Platform Manager
Platform Analytics
Platform RTM
Platform VM Orchestrator

Platform Open Cluster Stack 5: Accelerate & manage single Linux cluster
• Platform MPI contains *forward looking* MPI features:
  – Parallel Checkpoint/Restart
  – Dynamic affinity (load monitoring)
  – Windows CCS support (OS agnostic clusters)
  – Dynamic process creation
  – Network Failover
  – Replace run time (.so file) without recompile/relink
Platform MPI Benchmarks

FL5L2 Performance

Platform MPI outperforms by ~30%
Powerful workload scheduler – Platform LSF

- Robust, fault tolerance, and high performance
- Powering mission critical environment
- Special price for education

By scheduling workloads intelligently according to policy, Platform LSF reduces application run-time and optimizes resource use.

VIRTUALIZED VIEW OF COMPUTE, NETWORK AND STORAGE RESOURCES
Platform Computing Simplifies

• Why we use Platform products: simplicity and flexibility
  ◦ User base: hundreds of heterogeneous users
  ◦ Mixed modes: from EP to MPI
  ◦ Flexibility: open API
  ◦ Monitoring: reporting and analytics
  ◦ Ease of use: simple and clearly documented
Centralized monitoring/reporting – Platform RTM

- Tool for LSF Administrators and Operators
- Near Real-Time Monitoring for LSF and extendable to other physical devices and application software
- Drill-down to Individual job statistics
- It allows you to graphically see your clusters job history without the use of `b-commands`
- Based on Open Source Cacti
- *Platform RTM enables system administrators to make timely decisions for proactively managing compute resources against business demand...*
RTM Shows Detailed Job Statistics

### General Information
- **Status**: DONE
- **Job ID**: 046227
- **Job Host**: symph4
- **Username**: ddunlap
- **Start Time**: 2006-11-29 12:27:00
- **Exit Status**: 0

### Submit Detail
- **Project**: Dyna
- **Submit Host**: symph4
- **Submit Time**: 2006-11-29 12:24:00
- **Application Name**: airbag.deploy.dyna
- **User Name**: ddunlap
- **Resource Requirements**: select(mem=256M, swp=256M)
- **Submit Command**: 
  ```bash
  #/bin/sh #SBATCH -P 'Dyna' #SBATCH -J 'airbag.deploy.dyna' #SBATCH -q normal #SBATCH -n 1 #SBATCH -R 'select(mem=256M, swp=256M, max_scratch=2...'

### Execution Environment
- **Current/Last Status**: 43.27 Minutes
- **System Time**: 11 Seconds
- **User Time**: 43.08 Minutes
- **Mem Used**: 40.957M
- **Swap Used**: 106.891M
- **PGID(s)**: 10464

### Graphical History
#### Memory Consumption for Job
![Memory Consumption Graph]
- **Maximum Memory**: 35.67 M
- **Average Memory**: 106.82 M
- **Minimum Memory**: 40.96 M
- **Peak Memory**: 106.83 M

#### Memory Variation for Job
![Memory Variation Graph]

#### CPU Time for Job
![CPU Time Graph]
- **Maximum CPU Time**: 104.53 M
- **Average CPU Time**: 2.64 M
- **Minimum CPU Time**: 0.60 M
- **Peak CPU Time**: 9.60 M

#### CPU Time Variation for Job
![CPU Time Variation Graph]
Single Cluster All-in-One Solution

- Simplify HPC cluster

HPC Management Software – for Single Linux Cluster

Platform Open Cluster Stack (OCS)

- Intel Software Tools Kit
  - MPI & HPX Kit
  - OFED tools Kit
  - Intel Cluster Ready Kit
- Platform Lava PVFS2
- Nagios, Cacti NTOP
  - Core Kusu Management Tools

Develop | Schedule & Run | Manage

Red Hat HPC Solution = RHEL + Platform OCS 5
Creating a OCS 5 Installer Node

Boot from CD → Choose Language → Configure Networks → DNS, & Gateway

Root Password → Partitioning & LVM → Adding Kits → Install Summary

Installs Packages → Installer Node Boot

Installation complete. Installer node is ready to use.
### Platform OCS 5 Features

**Operating Systems:**
- RHEL 4, 5*
- Fedora 6
- Fedora 7
- CentOS 5

**Message Passing Libraries:**
- MPICH 1,2
- MVAPICH 1
- OpenMPI
- Intel MPI*

**Workload Management:**
- Platform Lava
- Platform LSF*

**Resource Management:**
- Platform EGO*

**Application Portals:**
- NICE Engine
- Frame*

**Certification Tools:**
- Intel Cluster Checker+

**Applications & Benchmarks:**
- Atlas
- blacs
- bonnie++
- fftw
- hdf5
- iozone
- iperf
- linpack
- modules
- netcdf
- scalapack
- stream

**Networking Hardware:**
- Gigabit Ethernet
- Cisco Infiniband
- Qlogic Infiniband
- Mellanox Infiniband

**Cluster Provisioning Methods:**
- Package
- Diskless
- Imaged

**Repository Management:**
- Multiple repositories
- Repository Snapshots
- Full Dependency Checking
- Package management

**Cluster Update:**
- Update from RHN
- Update from Yum repository
- Update individual rpms

**Configuration Management:**
- Node Templates/Classes
- Cluster File Synchronization
- Driver Management
- Cluster Network Configuration

**Cluster Administration**
- **Cacti Configured to monitor**
  - Memory on a node
  - CPU Load on a node
  - Logged In users
  - Processes on a node
- logwatch configured to monitor activity on the cluster
- Pdsh for administering nodes on the cluster

**Nagios configured to manage & Notify**
- Nodes in the cluster
- Web server
- Load
- logged in users
- DNS Server status
- MySQL Server status
- NTP daemon status
- Node 'PING' status
- Root Partition disk
- SSH Daemon
- SWAP Space
- Total Processes on node
- Email Notifications to Administrator

**Plone CMS containing:**
- Manual pages for OCS
- Documentation for Platform OCS 5
- List of installed Kits on the cluster

---

*Optional Commercial products that must be purchased separately
+Available only from Intel Cluster Ready Partners
• **Workload based dynamic power management**
  – Turn on/off machine
  – Avoid hot spot
  – Slow down CPU speed when MPI task is waiting.

• **Workload driven dynamic provisioning**
  – Switch machines between Windows and Linux automatically based on workload

• **“Cloud” Toolkit**
  – User web portal
  – Dynamic provisioning
  – Workload scheduling
  – Accounting and billing
“Platform has been proactive, involved and very, very friendly in providing support.”

Henry Neeman  
Director, Oklahoma University  
Supercomputing Centre

“Platform’s standard of support has been excellent.”

Tim Cutts  
Platform LSF Administrator  
Sanger Institute
Summary

• Platform’s Value proposition
  – Total solution for HPC management
  – 16 years of HPC experiences and broader customer base
  – Top quality technical support and services
Platform

www.platform.com
info@platform.com
1-877-528-3676 (1-87-PLATFORM)