Workload Management: NQE/LSF Status & Plans

Jack Thompson

Marketing Product Manager

SGI

jt@sgi.com

Brian MacDonald

Technical Relationship Manager

Platform Computing

brian@platform.com

41st Cray User Group Conference
Minneapolis, Minnesota

Agenda

- NQE Transition & Status
- Migration Program
- Status of LSF on SGI and Cray Systems
- LSF Plans
- Q&A

NQE Transition

NQE 3.3

- Final feature release

Next Steps

- ISV solutions prevalent
  - Core competency issue
  - Multi-vendor environment
- Partner solution best choice
- Platform Computing’s LSF

NQE Status

- Supported on SGI and Cray Systems
  - Support through year-end, 2004
  - Critical bugs fixed
  - Call center support
- Available for Cray SV1 systems
- Retired on non-SGI systems

LSF Migration Program

sgi

- Discounted pricing for systems licensed for NQE before February 1, 1999
  - Available through January 31, 2000
- Migration Guide
  - Developed jointly by Platform and SGI
- Professional services available
- Inclusion of key NQE features in LSF

PLATFORM

Strong relationship between SGI and Platform Computing engineering teams

LSF on SGI Systems

Current release is LSF 3.2

- Now available on IRIX, UNICOS, UNICOS/mk
  - Including Cray SV1
- Also on NT and Linux
- Available from SGI
  - LSF Standard Edition, LSF Parallel, LSF Client
- Available from Platform Computing
  - LSF Analyzer, LSF MultiCluster, LSF JobScheduler, LSF Make

Data Center Requirements

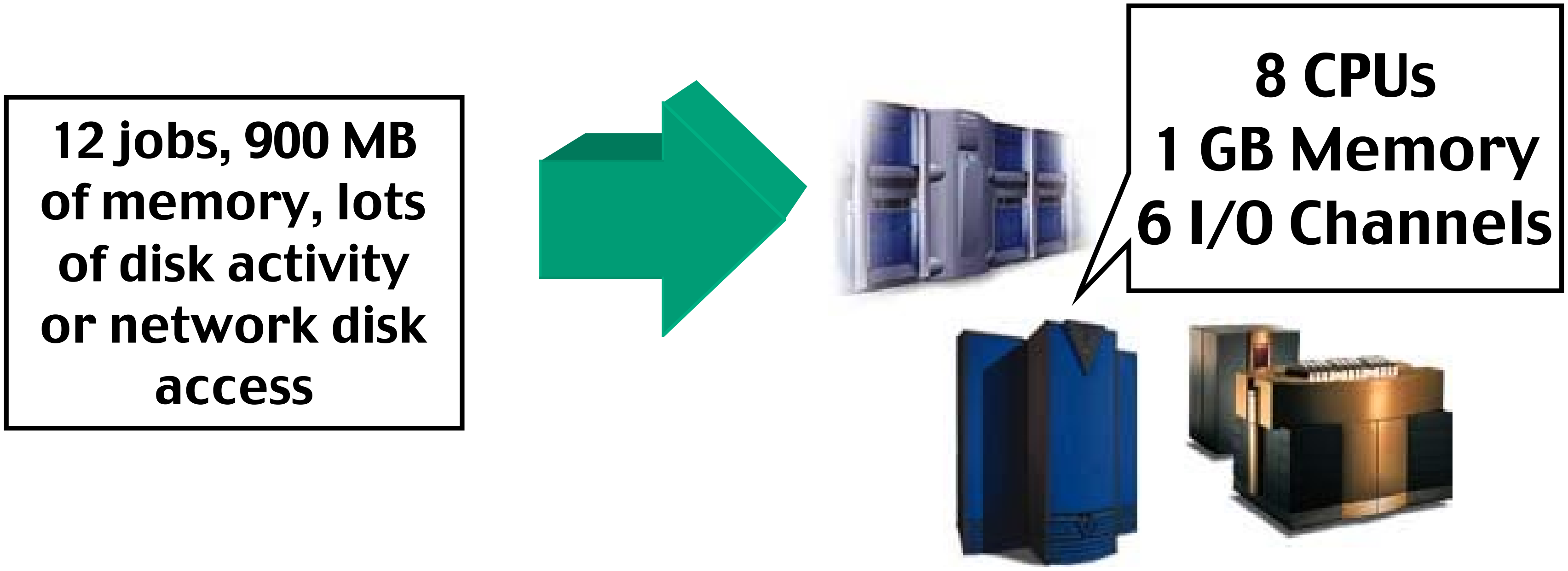
Environments for High Performance

- Single point of control and administration
- Logically present a single system image to users, applications and networks
- Application of policies across the consolidated platform – uniform across all machines
- Uniform policies to satisfy workload performance objectives in terms of throughput, turn around and response time
- Improved application availability – both for failures and planned outages

Defining Capacity Goals

LSF can be focused on throughput guarantees

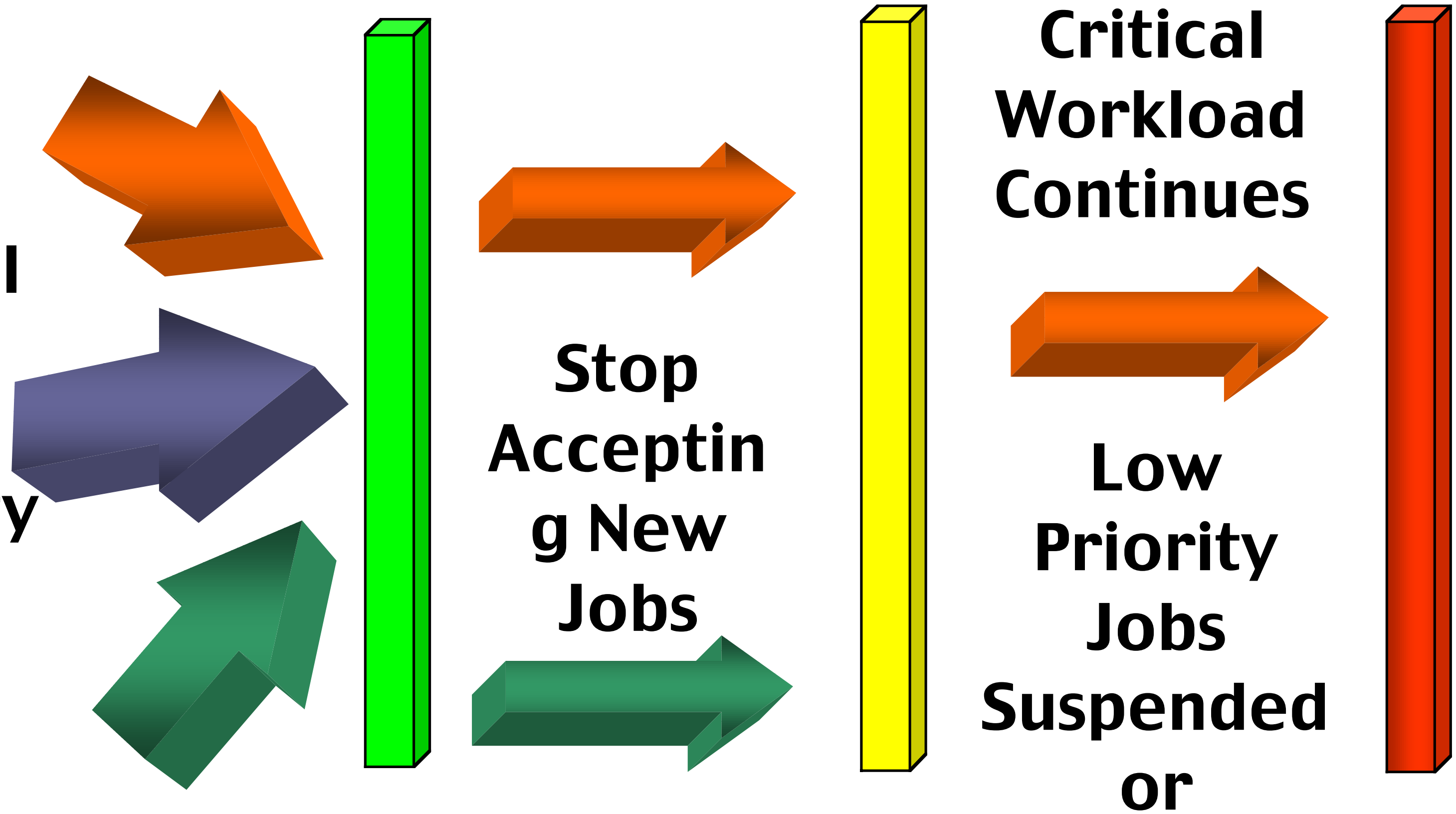
- **Run as much workload on the box, absolute performance not primary goal**



12 jobs, 900 MB of memory, lots of disk activity or network disk access

8 CPUs
1 GB Memory
6 I/O Channels

Thresholds for Execution



Critical and Lower Priority Jobs

Stop Acceptin g New Jobs

High Priority, Critical Workload Continues

Low Priority Jobs Suspended or Migrated

85 %

90 %

100 %

CPU Utilization

Defining Capability Computing

Clearly Stated Performance Goals

- Get my job done as quickly as possible using all necessary dedicated resources
- Avoid sharing and contention at all costs
- Problems can be tackled that otherwise could not be considered
- Mission critical applications gain the undivided attention of the computing infrastructure

Defining Capability Computing

Supporting the Exclusive Execution Model

- multi-box parallelism (Origin 2000)
- mixed operation large machines
- optimum support for Cray T3E
- committed product development in support of partitioning mechanisms
  - Miser (Q4 99)
  - Miser CPU sets (Q4 99)
  - OS service follow-on (XRS)

Resource Based Job Placement

Selection

- Match necessary conditions

Ordering

- Choose the best from eligible candidates

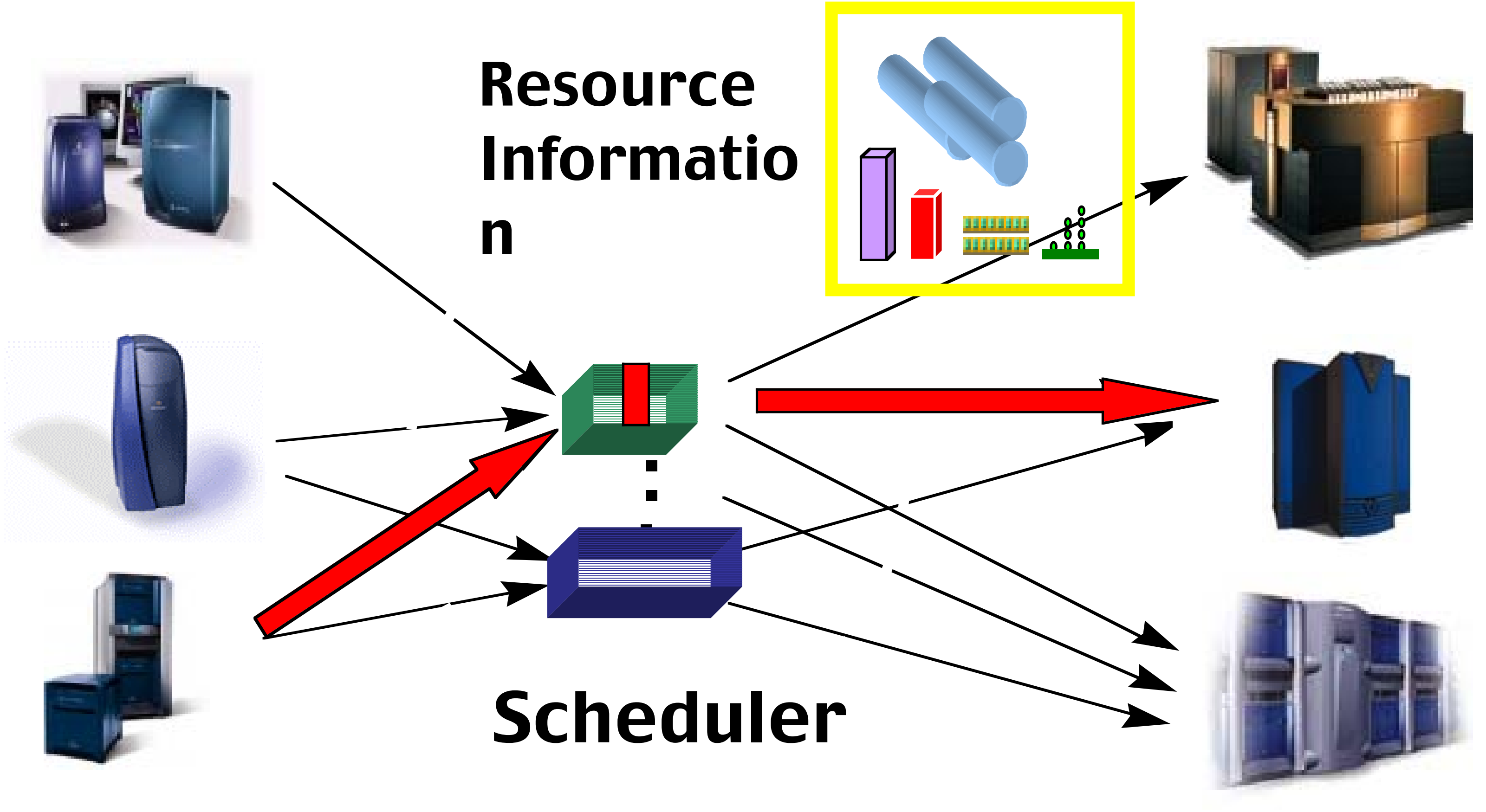
Reservation

- Adjust load values for selected hosts

Spanning

- Define locality of parallel jobs

Single Processing Image



Resource Information

Scheduler

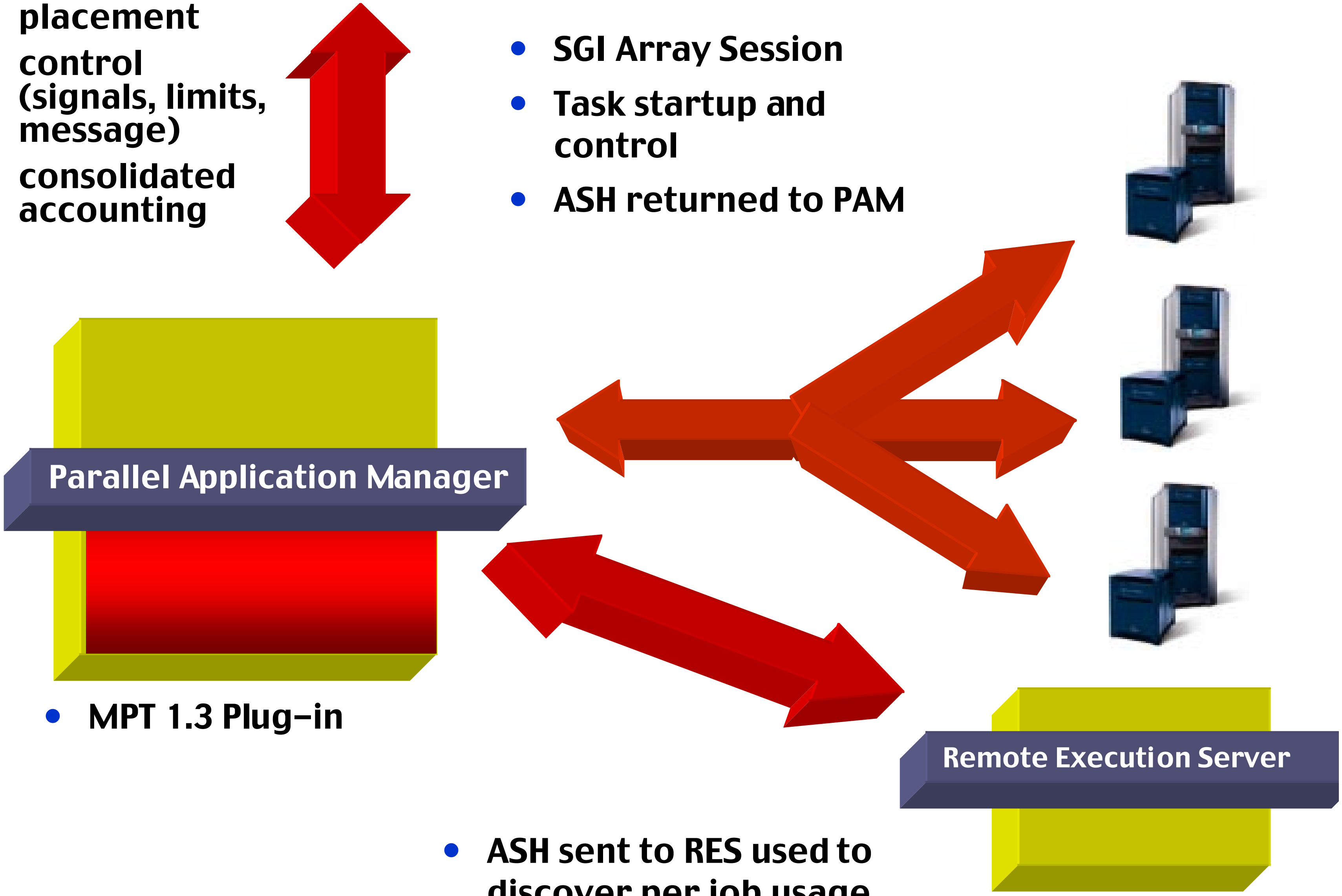
submission hosts batch queues server hosts

System Level Integration

sgi

- placement
- control (signals, limits, message)
- consolidated accounting

- SGI Array Session
- Task startup and control
- ASH returned to PAM



- MPT 1.3 Plug-in

- ASH sent to RES used to discover per job usage

Solutions Through Integration

sgi

ISVs, Custom Scientific and Commercial Applications transparently gain access to resource management services without changing their code

- **Application Checkpoint Restart**
- **Transparent host selection**
- **Accounting for ISV applications**

MARC ANSYS

FLUENT INCORPORATED

MSC The MacNeal-Schwendler Corporation

LSF Parallel 3.2

sgi

MPT 1.3

LSF 4.0 Enhancements

Scheduler

- Scalability improvements for all the bells and whistles turned on – Fair-share + Back-filling
  - 20,000 + jobs
- Dynamic re-configuration without re-start
  - lim and mbatchd
- Client query scalability
  - support for thousand's of clients
- Adaptive dispatch for high throughput, short running jobs
- Time dependent configuration for queues
  - different queue for night, same queue

LSF 4.0 Enhancements

Job Execution

- Improved Input/Output handling support
  - I/O Spooling
  - Admin defined spool directory
  - Job level CWD discovery enhancements
- Integrated FTA supported within LSF
- Job Flow
- Kill re-queue

Administrative Improvements

- Non-shared daemon configuration support
- Automatic host type and model detection