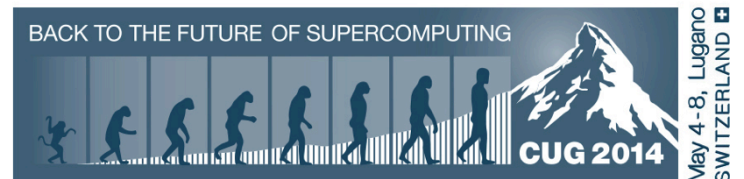# Enhanced Job Accounting with PBS Works and Cray RUR:
Better Access to Better Data

Scott Suchyta
May 2014

Thanks to Carl Albing, Andrew Barry, & Jason Coverston

BACK TO THE FUTURE OF SUPERCOMPUTING

CUG 2014

May 4 - 8, Lugano
SWITZERLAND

Agenda

- **Brief Introduction**

- **Overview**

  - Problem Statement

  - PBS Analytics

  - PBS Plugins

- **How it Works**

- **Hiccups?**

- **Closing**

Altair Overview

Founded ...

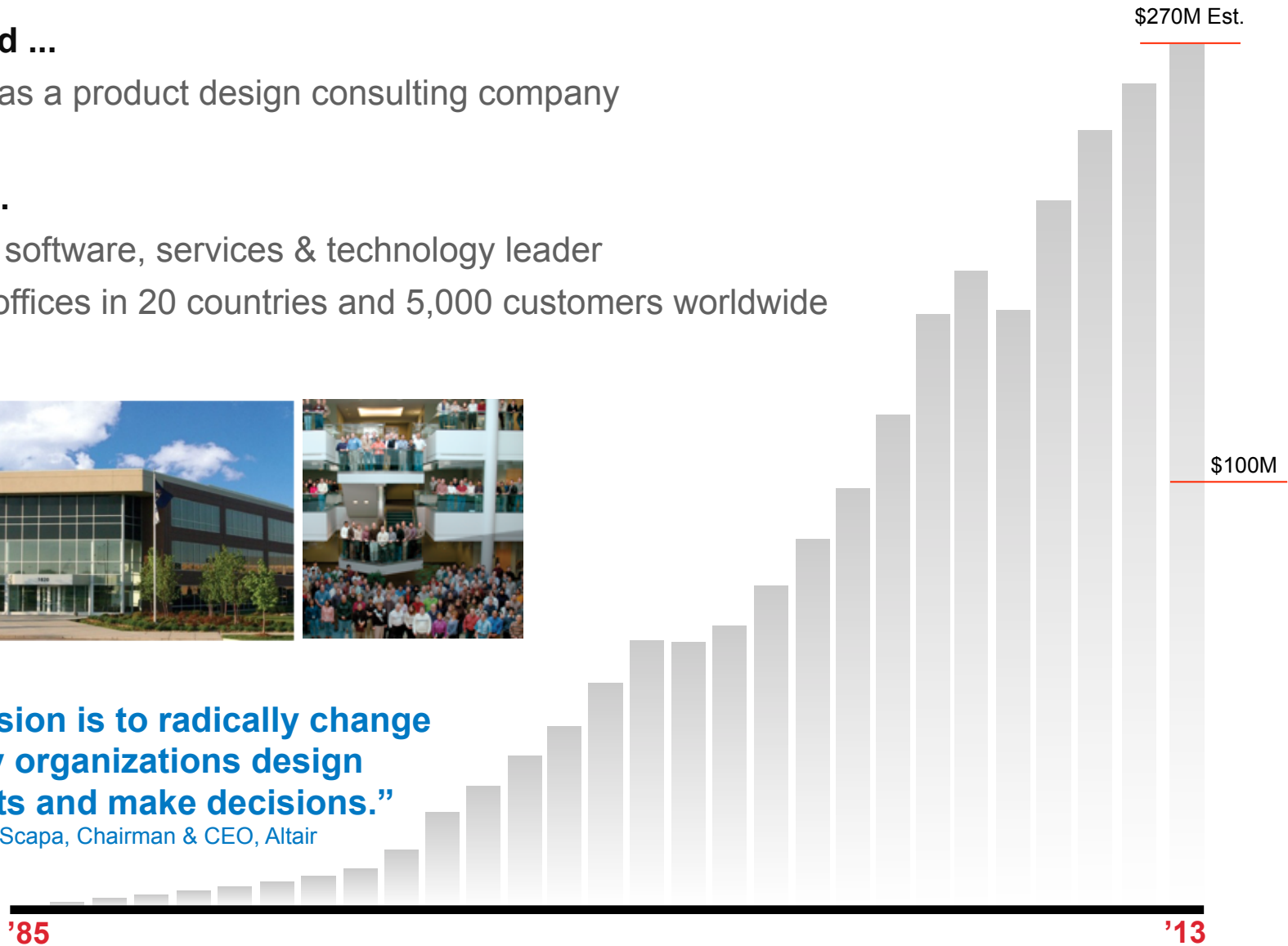In 1985 as a product design consulting company

Today ...

A global software, services & technology leader

with 48 offices in 20 countries and 5,000 customers worldwide

"Our vision is to radically change the way organizations design products and make decisions."
– James R. Scapa, Chairman & CEO, Altair

$270M Est.

$100M

'85   '13

3

Innovation Intelligence®

28
Years of Innovation

48
Offices in 20 Countries

2050
Employees Worldwide

Global Presence

Seattle, USA	Montreal, Canada	Lund, Sweden	Delhi, India	Beijing, China
Salt Lake City, USA	Toronto, Canada	Gothenburg, Sweden	Pune, India	Shanghai, China
Mountain View, USA		Leamington Spa, UK	Chennai, India	
Los Angeles, USA		Bristol, UK	Hyderabad, India	Tokyo, Japan
Austin, USA	Detroit, USA	Manchester, UK	Bangalore, India	Osaka, Japan
Houston, USA	Milwaukee, USA	Stuttgart, Germany		Nagoya, Japan
	Charlotte, USA	Cologne, Germany	KL, Malaysia	
	Huntsville, USA	Hamburg, Germany		Seoul, Korea
Mexico City, Mexico		Hanover, Germany		
		Munich, Germany	Tel Aviv, Israel	Taipei, Taiwan
		Graz, Austria		
		Paris, France		
	Sao Paulo, Brazil	Lyon, France		
		Sophia Antipolis, France		Melbourne, Australia
		Toulouse, France		
		Torino, Italy		
		Madrid, Spain		
		Thessaloniki, Greece		

48 offices across 20 countries

Customers

Automotive	Aerospace	Heavy Equipment	Government	Life/Earth Sciences	Consumer Goods	Energy

5,000 customers worldwide

Altair's Divisions and Companies

△ Altair

HyperWorks®

Engineering Simulation
and Optimization Software

ProductDesign Solutions

Product Innovation and
Development Consulting

solidThinking®

Simulation-driven Industrial
and Concept Design Software

Enterprise Solutions

Cloud-based Business and Engineering
Analytics Software and Consulting

PBS Works™

High Performance Computing
Software and Consulting

Staffing Solutions

Technical Staffing and
On-premise Consulting

Project Overview

What we did, in a nutshell

- **Project**: Build an integration between Cray RUR and PBS Analytics

- **Goal**: Enhanced Job Accounting -- Better Access to Better Data

- **How we did it:** Used a PBS plugin (execjob_epilogue) to extract Cray RUR metrics and updated the job accounting information with the PBS Server. Finally, PBS Analytics digests the PBS Professional accounting logs.

- **Results:** Useful metrics, from Cray RUR, are incorporated in the PBS Professional accounting logs.

Problem Statement & Use Case

Business Drivers: Why is job accounting needed?

- Are projects being delayed due to lack of the **right IT resources**?

- How can we **plan for software and hardware growth**?

- How can we **share and rationalize** expensive software and hardware?

- Are we **meeting our service level agreements** with our customers?

- How can we **track usage** from geographically separate groups?

- How can we **optimize license use** and minimize unused licenses?

- How can we **reduce waste to zero**?

Maximize the Value of Your IT Assets

- **Allocate costs** (chargeback) to projects, business units and regions

- **Accurately plan capacity growth** of software licenses and hardware by forecasting use based on real data

- **Reduce spending** by sharing expensive licenses and raising utilization

- **Get more work done** (via PBS Professional) without impacting interactive license use during valleys in real-time license usage

- **Meet project deadlines** by minimizing IT bottlenecks

- **Improve contract terms** when negotiating the purchase or renewal of licenses by understanding actual usage and cost data
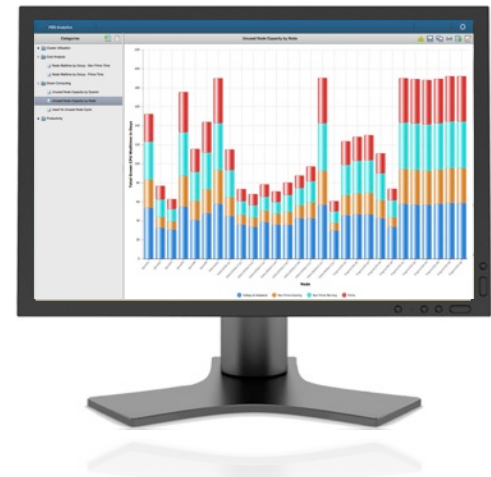
PBS Analytics

PBS Analytics

HPC Accounting & Analytics Portal
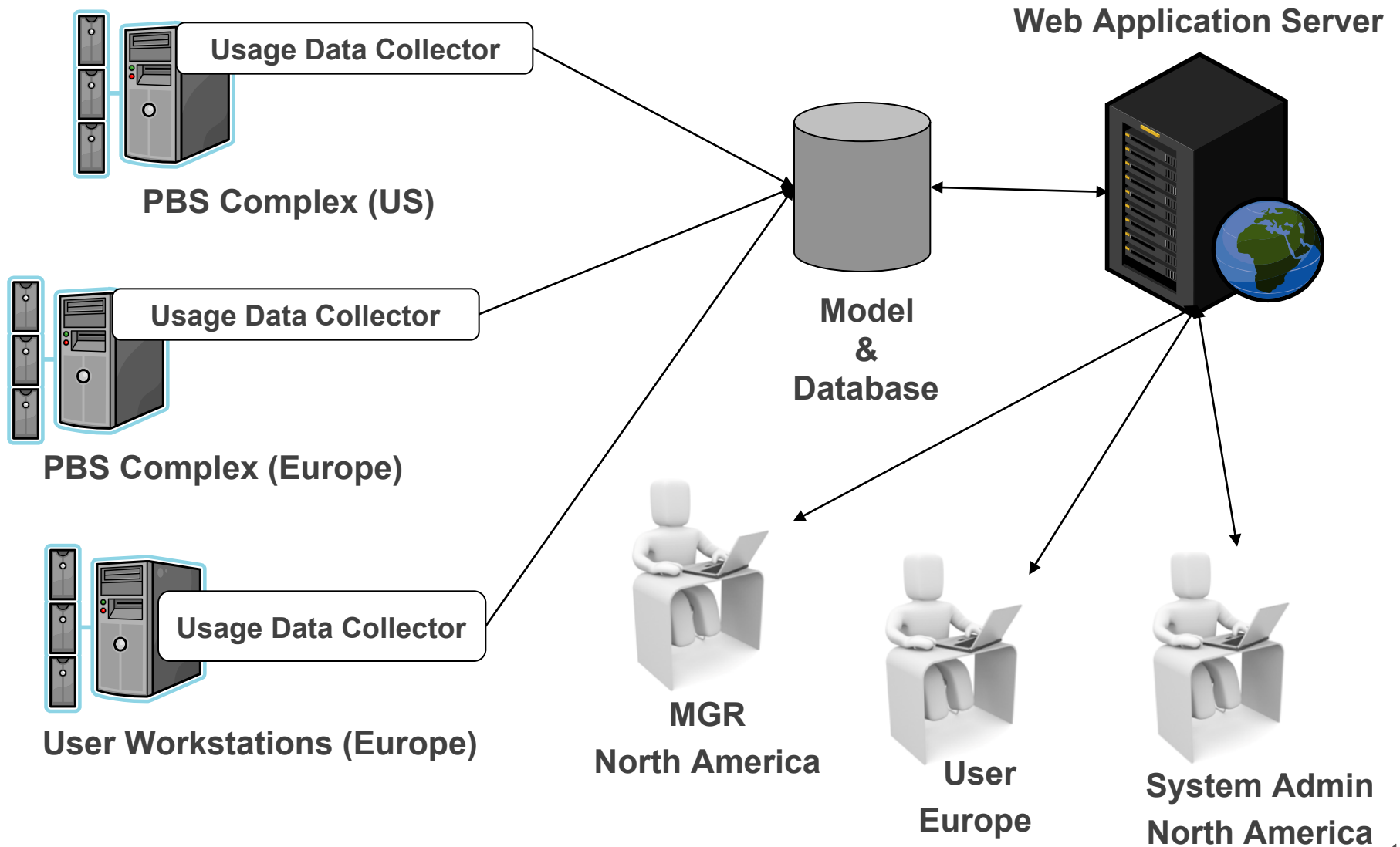## Visualize Historical Usage for Optimized Returns on HPC Investments

- Allocate Costs & Plan for Future Capacity

- Visualize Workload & Historical Usage

- Drill-down to Underlying Data

- Filter by Project, App, User, Group, Queue, Host, …

- Canned reports out-of-the-box, and customize your own

- Aggregate data from multiple PBS Professional servers

- Slideshow Mode for Continuous Display (in Lobby)

Maximizing our license utilization means we don't have to buy a new license, set up another workstation, and hire another engineer to keep up with demand.
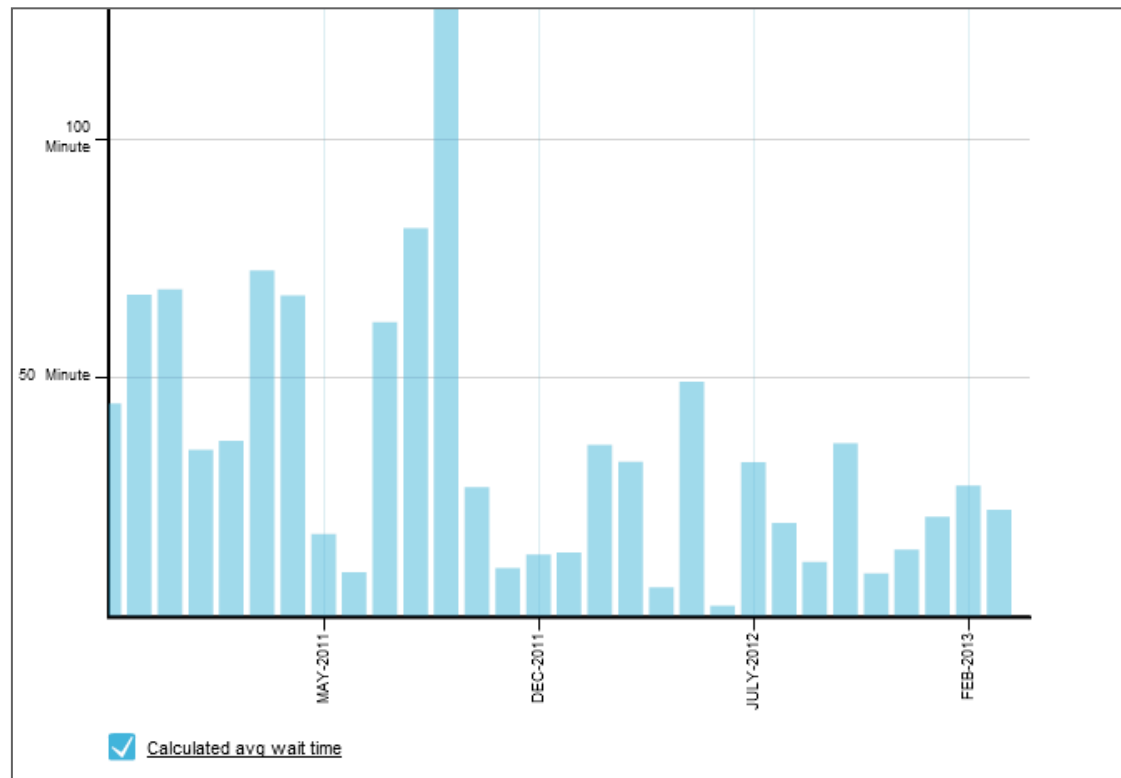
– Trelleborg

View | Optimize | Forecast

PBS Analytics Architecture

△ Altair

Usage Data Collector

PBS Complex (US)

Usage Data Collector

PBS Complex (Europe)

Usage Data Collector

User Workstations (Europe)

**Model
&
Database**

Web Application Server

**MGR
North America**

**User
Europe**

**System Admin
North America**

15

Example: Analyze Wait Times to Improve Service…

Do we need more hardware? More licenses?

Example: System is lightly loaded...

Users mostly got their jobs running straight away.
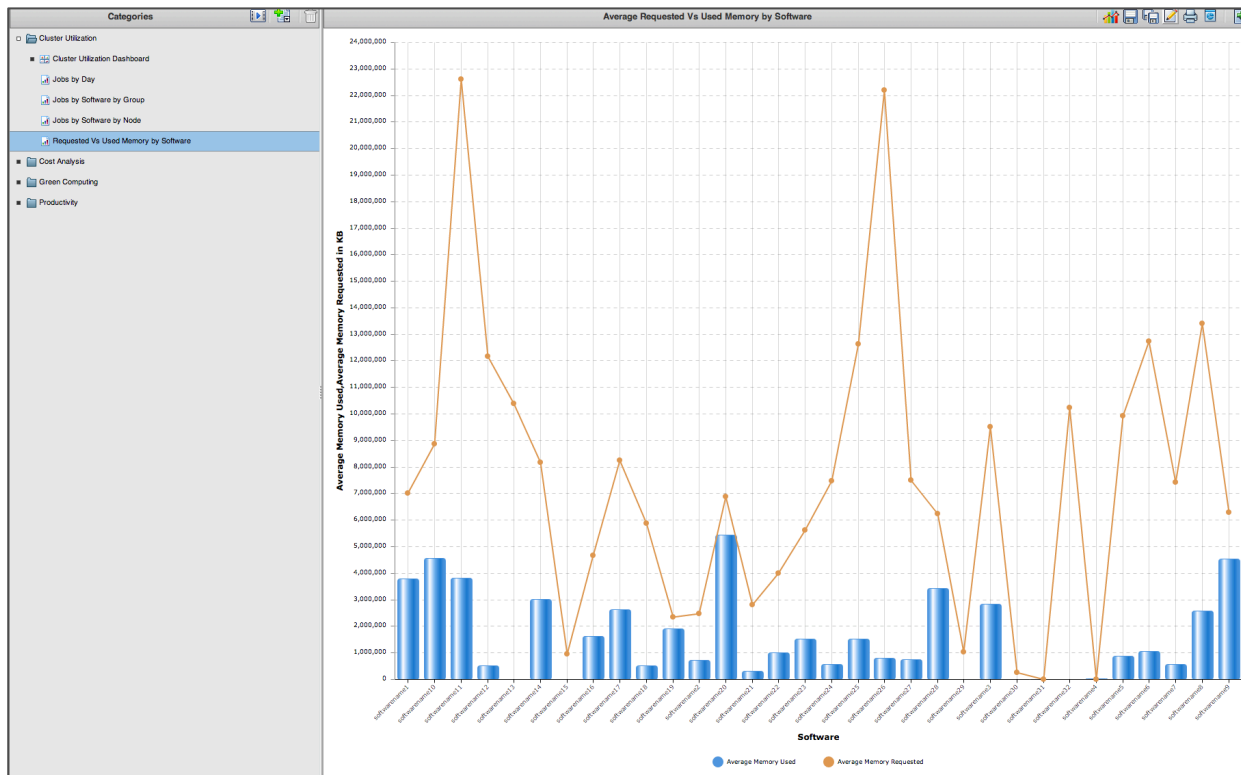
There was very little wait time.

Hmmm, in Oct. '12 users had to wait longer… well that's due to the new project in Oct '12!
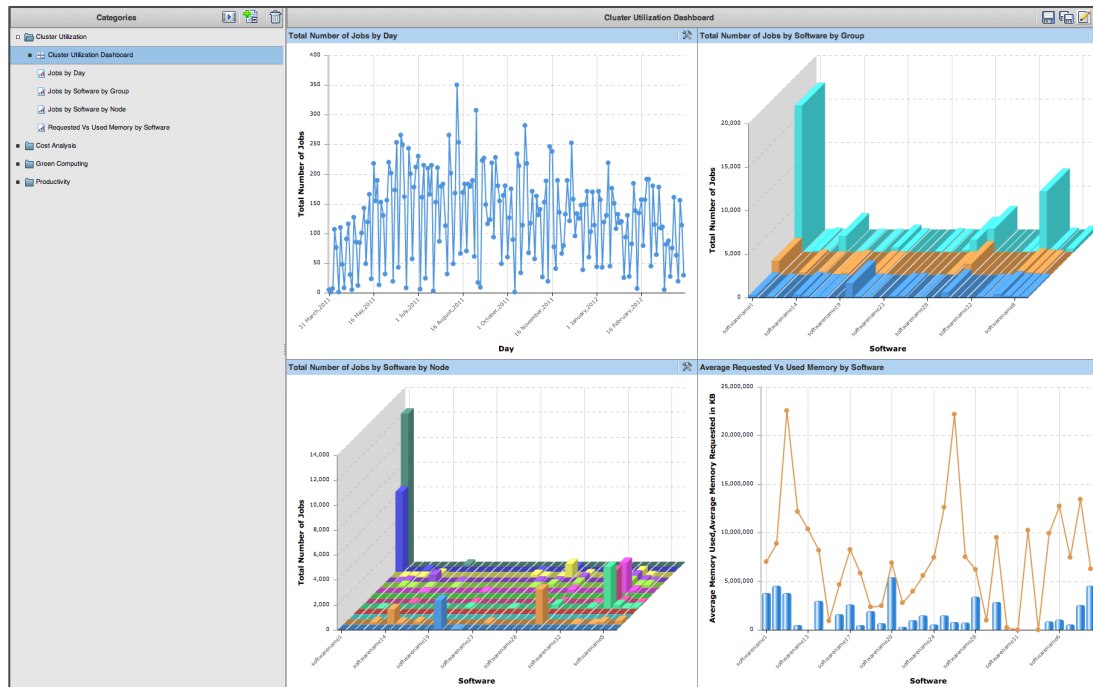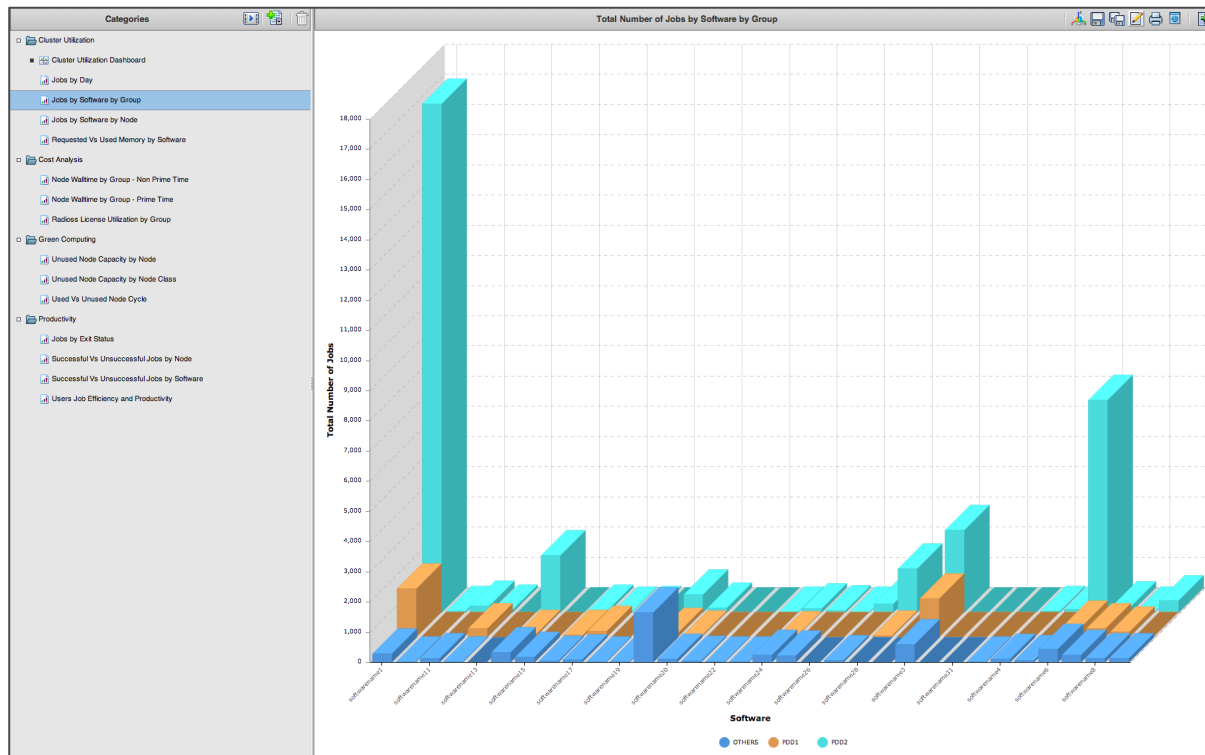
Example: Users Request Too Much Memory…

Users are asking for more memory than needed.

**This could have caused some jobs to wait unnecessarily…**

Example: Engr is the Most Active Group

There were hardly any jobs from contractors!
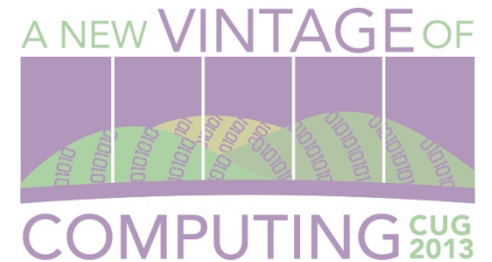
Do we get any value out of them?

Example: Most Used Software…

Lets buy more licenses of OptiStruct as it is our most used software

How can we increase the usages of RADIOSS – lets conduct training

Why are we still running Nastran?

Is it worth maintaining LS Dyna?
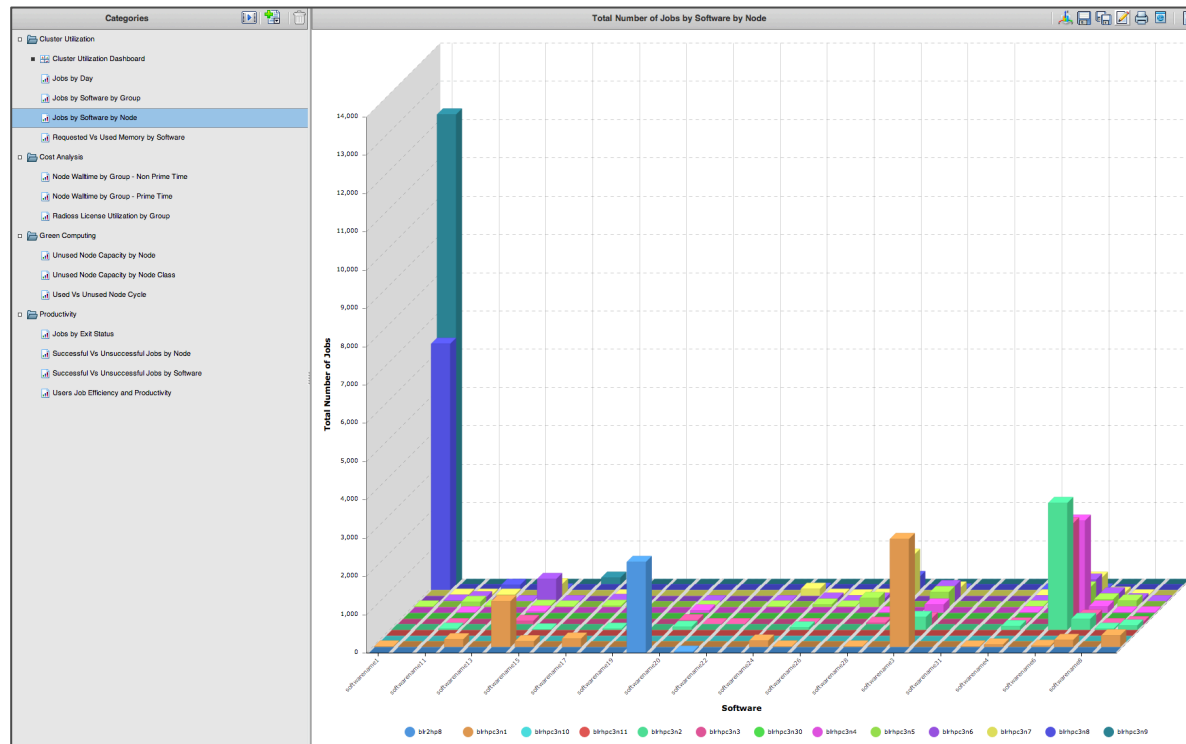
Let's Recap

PBS Plugins
"Hooks"

A NEW VINTAGE OF COMPUTING CUG 2013

Napa Valley, California • May 6–9

PBS Plugins ("Hooks")

Change / augment capabilities in the field, on-the-fly, without source

Unified data model based on industry-standard Python

Admission control events

- Validation, allocations, on-the-fly tuning, novel limits, logging, …

Job execution events

- Parallel node setup / cleanup
- Periodic monitoring
- Job termination
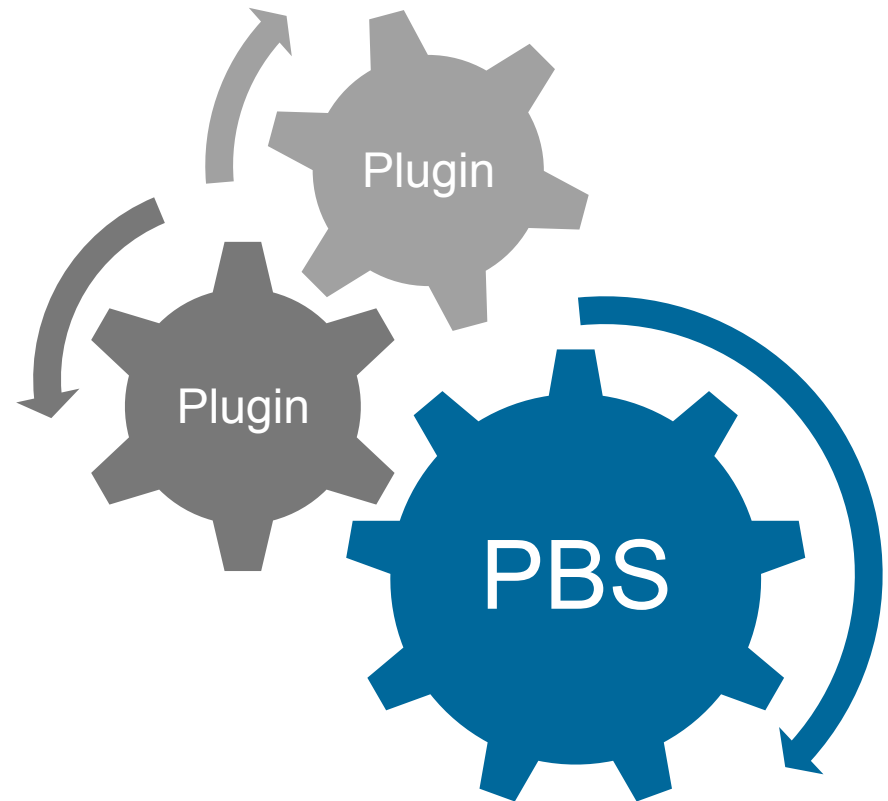
Plugin Examples

Plugins Deliver Real Capabilities

Fault detection & mitigation

- Mitigate "black hole" syndrome

Customized access control

Customized runtime environments

Allocation management

The pbs Module

- **Natural mappings to PBS objects**

  ```
  q = s.queue("workq")
  q.total_jobs  ← returns the # of jobs on "workq"
  q.job("22.fest") ← returns a job in the queue
  ```

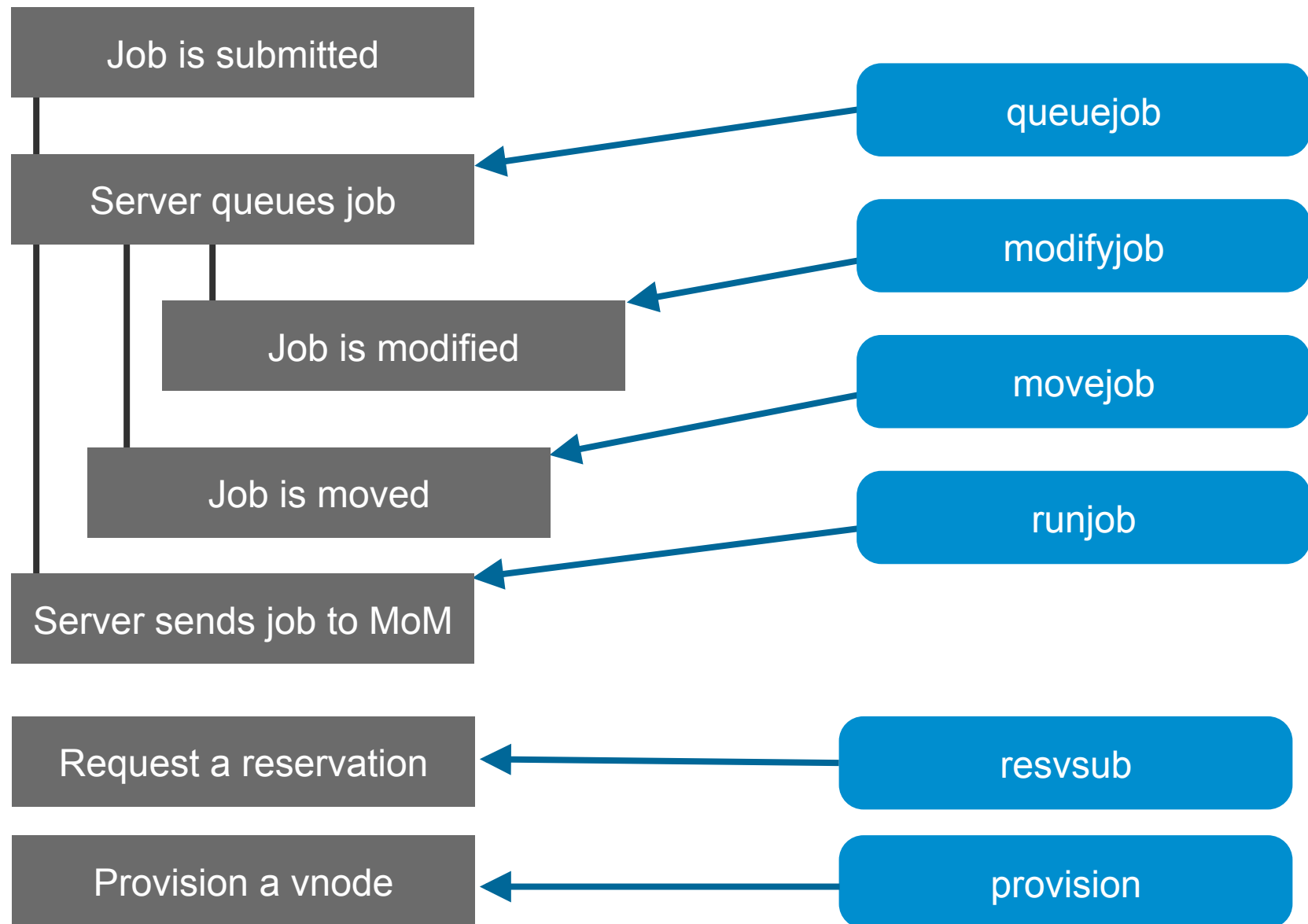- **Write log/debug info directly to PBS logs**

  ```
  pbs.logmsg(pbs.LOG_DEBUG, "Hooks are awesome!")
  ```

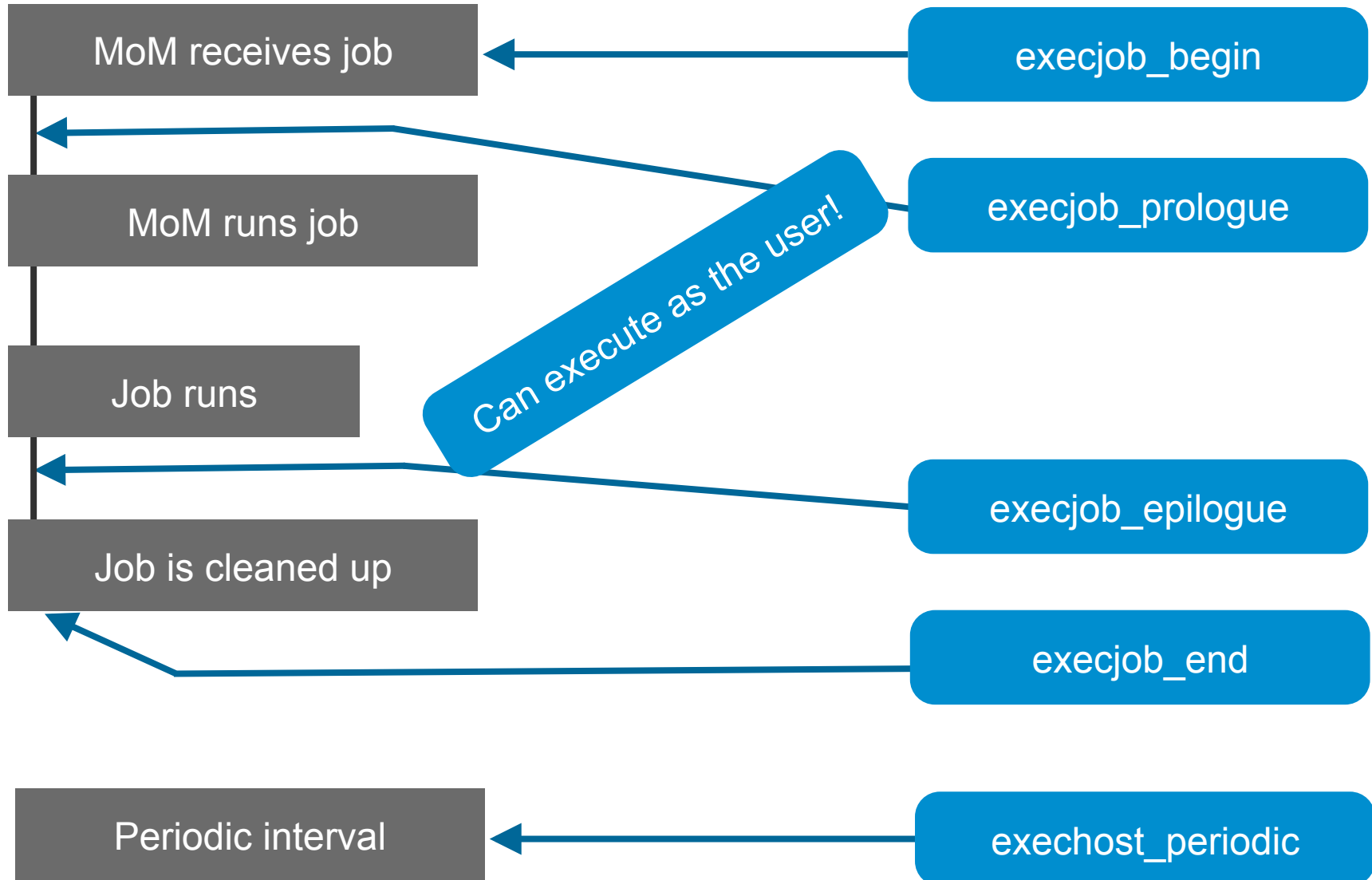- **Standard Python exception handling**

- **Multiple hooks per event (including ordering execution)**

24

Plugins: Admission Control and Management

△ Altair

Job is submitted	
	queuejob
Server queues job	
	modifyjob
Job is modified	
	movejob
Job is moved	
	runjob
Server sends job to MoM	
Request a reservation	resvsub
Provision a vnode	provision

25

Plugins: Job Execution

△ Altair

| MoM receives job | ← | execjob_begin |

| MoM runs job | | execjob_prologue |

Can execute as the user!

| Job runs | | |

| Job is cleaned up | | execjob_epilogue |

| | | execjob_end |

| Periodic interval | ← | exechost_periodic |

Job Execution Events (MOM Hooks)

Job Lifecycle

1. Set up

2. Stage-in file(s)

3. Prologue

4. Launch job

5. Epilogue

6. Stage-out file(s)

7. Clean up

PBS Professional v12

1. Adds Setup Hook (before environment…)

2. …

3. Adds Prologue Hook (replaces Prologue)

4. …

5. Adds Epilogue Hook (replaces Epilogue)

6. …

7. Adds Cleanup Hook (after obit…)

Executed by all pbs_mom associated to the job

Plus…

- Periodic hooks
- Can execute as the user
- Configurable via qmgr
- Debugging facilities

Cray Use Cases for MoM Hooks

- **Pre-job Health Checks**
  - Is ALPS really running?
  - Call node health checker before the job goes further
  - Set resource value(s) on PBS node

- **Amend User Environment**
  - Set environment variables to control job/application functions
  - Run script in same user environment as job will run in
  - Start user based accounting

- **Post-job Metrics & Health Checks**
  - Requeue checkpointed jobs
  - Reset the resources_used on a job that will be requeued
  - Set a flag for power usage & update a custom resource with values to be captured in accounting logs

The "How"

A little customization

- **Custom Cray's Resource Utilization Reporting (RUR) output plugin**

- **PBS MOM Hook – execjob_epilogue**

- **Customize the PBS Analytics charts**

Custom RUR outputplugin

- **Output unique RUR file for each job**
  - PBS_HOME/spool/rur.PBS_JOBID

```
<python_code>

outputfile = os.environ['PBS_HOME'] + "/spool/rur."+str(jobid)

<python_code>
```

31

PBS MOM Hook – execjob_epilogue

- **Define custom resources in PBS_HOME/server_priv/resourcedef**

- **Define & import execjob_epilogue hook**
  - Extract the RUR metrics from PBS_HOME/spool/rur.PBS_JOBID
  - PBS MOM records new job metrics with PBS Server

```
<python_forloop>

e.job.resources_used[taskstats_metric] = rur_data[jobid]
['taskstats'][taskstats_metric]


<python_forloop>

e.job.resources_used[energy_metric] = rur_data[jobid]['energy']
[energy_metric]
```

Customize PBS Analytics
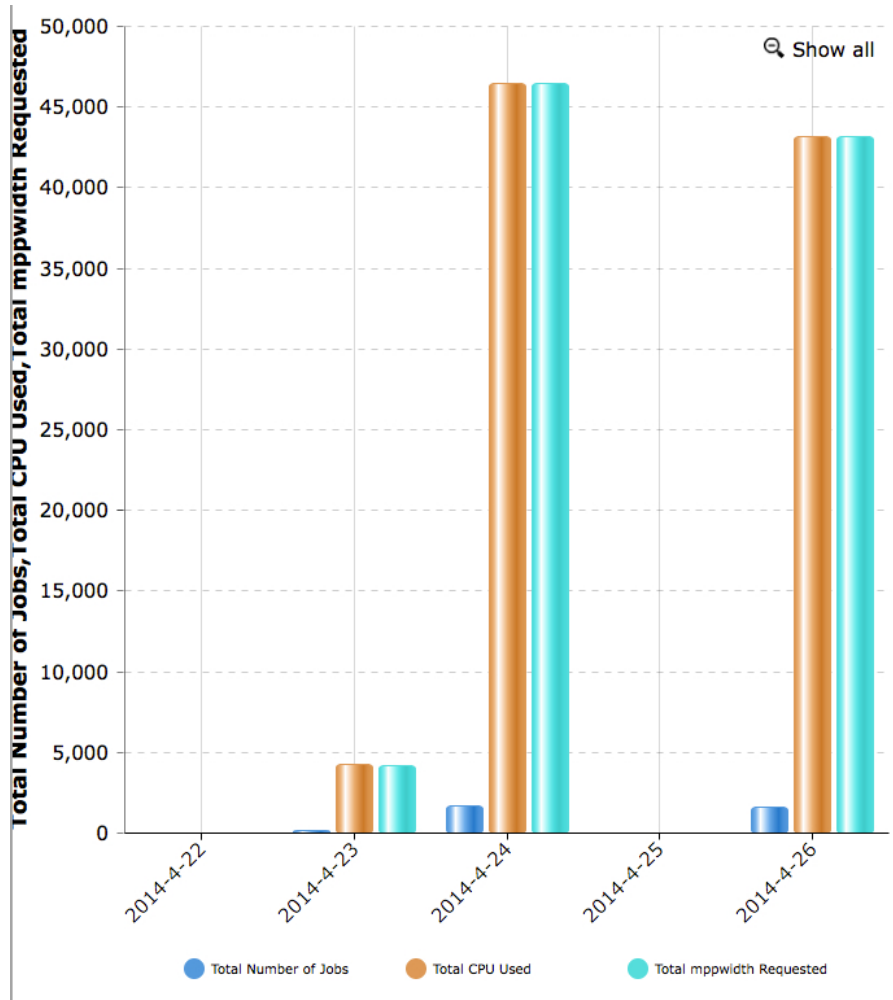
- **Define custom resources in parser config**
  - mpp*
  - RUR metrics

- **Create custom reports**

PBS Professional Accounting Logs

04/26/2014 22:18:36;E;1993.opal-p1;user=crayadm group=crayadm project=_pbs_project_default
accounting_id="0x500000669" jobname=memcheck queue=workq ctime=1398568647 qtime=1398568647
etime=1398568647 start=1398568693 exec_host=opal-p1/27+opal-p1/81+opal-p1/82+opal-p1/83+opal-
p1/84+opal-p1/85+opal-p1/86+opal-p1/87 **exec_vnode=(opal-p1_63_0:ncpus=1)+(opal-p1_63_0:ncpus=1)+**
(opal-p1_63_0:ncpus=1)+(opal-p1_63_0:ncpus=1)+(opal-p1_63_0:ncpus=1)+(opal-p1_63_0:ncpus=1)+
(opal-p1_63_0:ncpus=1)+(opal-p1_63_0:ncpus=1) Resource_List.arch=XT Resource_List.mpiprocs=1
Resource_List.mppnodes=28-31,36-37,60-63 Resource_List.mppwidth=8 Resource_List.ncpus=8
Resource_List.nodect=8 Resource_List.place=free Resource_List.select=8:vntype=cray_compute
session=17741 alt_id=200334 end=1398568716 Exit_status=0 resources_used.cpupercent=0
resources_used.cput=00:00:00 resources_used.mem=6536kb resources_used.ncpus=8
resources_used.vmem=151236kb resources_used.walltime=00:00:20
resources_used.min_accel_power_cap=245 resources_used.max_power_cap_count=1
resources_used.max_power_cap=425 resources_used.wcalls=200 resources_used.max_accel_power_cap=245
resources_used.max_vm=11992424 resources_used.utime=89040000 resources_used.rcalls=1068
resources_used.min_accel_power_cap_count=1 resources_used.wchar=4486
resources_used.minfault=2930297 resources_used.accel_energy_used=1472 resources_used.pgswapcnt=0
resources_used.min_power_cap_count=1 resources_used.max_rss=4684 resources_used.bkiowait=1277753
resources_used.majfault=0 resources_used.etime=320838348 resources_used.btime=1398568696
resources_used.min_power_cap=425 resources_used.stime=8368000
resources_used.max_accel_power_cap_count=1 resources_used.coremem=347273426
resources_used.energy_used=2851 run_count=1

PBS Analytics – Visualize!

PBS Analytics – Visualize!

PBS Analytics – Visualize!

37

Any Hiccups?

Of course, but they were minor

- **PBS Analytics**

  - Limit on the number of custom resources a site can define – to be increased

  - mpp* resources are 'foreign' – to be recognized as native resources, easy workaround

  - "Non-Additive" custom resources support – ENGR provided custom JAR, to be GA next release (12.4)

- **PBS Professional**

  - Memory reported on login node is wrong – to be fixed, easy workaround

- **Resource Utilization Reporting**

  - RUR output plugins lack consistency in output format – to be changing with options to output in json-list and json-dict

In Closing...

Altair Knows HPC

Altair is the only company that:

makes HPC tools…

develops HPC applications…

…and **uses** these to solve real HPC problems

500 Altair engineers worldwide

use HPC every day *for*

real-world modeling

& simulation

41

Hot off the Press: DWD Picks PBS Professional

- **Customer**: Germany's national meteorological service Deutscher Wetterdienst (DWD)

- **Solution**: PBS Professional for workload management on two Cray XC30 systems

- *"PBS Professional provides the **bulletproof reliability and flexibility** we need to manage our critical numerical weather prediction workloads and ensure **faster, more reliable computing performance**."*
  - *-- Dr. Dieter Schröder, CIO at DWD*

- *"PBS Professional is a proven workload management product and an **excellent choice** for organizations and businesses with **large, complex systems** where uptime and reliability are top priority.'*
  - *--Peg Williams, SVP HPC at Cray*

- Read more at http://www.pbsworks.com/

Top Global Weather Service Selects Altair for Petascale Workload Management
Germany's DWD Picks PBS Professional® to Manage HPC Workload on Cray XC30 Systems

TROY, Mich., May 6, 2014 – Altair today announced that one of the world's premier numerical weather prediction centers, Germany's national meteorological service Deutscher Wetterdienst (DWD), has selected Altair's PBS Professional for workload management on its two Cray XC30 supercomputing systems.

Researchers and scientists at DWD will utilize the petascale systems to produce higher resolution and more accurate global and regional weather predictions to help fulfill the organization's wide range of weather responsibilities, including the meteorological safeguarding of aviation and shipping and the issuance of official warnings of weather occurrences that could become dangerous.

"At DWD it is our mission to ensure the highest levels of accuracy in weather and climate services," said Dr. Dieter Schröder, CIO at DWD. "Our researchers and scientists rely on their supercomputing systems for high availability, performance and scalability and need a workload manager proven to operate in large, complex production environments. PBS Professional provides the bulletproof reliability and flexibility we need to manage our critical numerical weather prediction workloads and ensure faster, more reliable computing performance."

Announced Tuesday May 6

"Everyone relies on weather prediction and the forecasting systems," said Big Mooney, chief technology officer at Altair. "As computing becomes more accurate and timely, forecasts require a level of reliability that is above that of other HPC areas. Reliability is a key strength of PBS Professional, and we're honored that DWD chose Altair for these compute-intensive, mission-critical weather-computing services."

The PBS Professional workload management product has been used for more than thousands of customer sites to manage distributed computing resources across geographic boundaries. PBS Professional is the flagship product for PBS Works™, a comprehensive suite for optimizing HPC and cloud computing environments including portal products for job management, analytics and remote visualization.

Numerical weather prediction uses many different types of simulations, modeling techniques and components at different spatial and temporal scales. The meteorological models used are extremely data-intensive, and the computational and data requirements to simulate these models are massive. In addition, with increases in available data and better understanding of the requirements for improving forecast accuracy, demands for handling complexity and big data movement grow even higher. Thus, weather centers require not only scalable architectures that support highly compute-intensive applications and efficient data movement – they also require a well-integrated solution built by experts who understand the experience of architecting and delivering systems at scale and complexity.

Leaders in the field of HPC solutions for weather and climate centers, Cray and Altair have collaborated to develop an integrated solution offering a breakthrough in system utilization strategy and high-performance workload scheduling and management. Designed with the demands of numerical weather prediction in mind, the Cray-Altair solution enables weather facilities to exploit peak performance capabilities while improving service quality and utilization rates. The solution is described in a white paper published by Altair and Cray: www.pbsworks.com/cray-weather-wp.

"Cray and Altair have a longstanding partnership, and we are pleased that the combination of the Cray XC30 supercomputer and PBS Professional has proven to be a compelling solution for DWD's meteorological workloads," said Peg Williams, senior vice president of high performance computing systems at Cray. "PBS Professional is a proven workload management product and an excellent choice

42

For More Information

Altair

- **Stop by Altair's table at CUG**

- **Contact us**

  Scott Suchyta

  scott@altair.com

- **Visit us online**

  www.altair.com

  www.pbsworks.com

Thank You!

" PBS is the **best choice** for complex HPC workload management.
–Scuola Normale

" It's pretty **easy to see the value**.   –QIMR

" Altair's expertise and dedication to success is **unbeatable**.
–Weizmann Institute

Why Altair for HPC Workload Management?

✓ Market leader for over 20 years

✓ Single comprehensive suite with integrated portals for job management and monitoring

✓ Proven at thousands of global sites

✓ Unparalleled security (EAL3+, SELinux)

✓ Powerful, easy to use plug-in framework for extensive customization capabilities

✓ Global support with experts in 20 countries

No other vendor offers the breadth and robustness of Altair's suite of HPC and engineering products and services