Optimize HPC with PBS Professional & OpenHPC

Scott Suchyta

November 2017
ALT AIR AT A GLANCE

Founded 1985
Headquartered in Troy, MI US

67 offices
in 23 countries

$313M
2016 Billings

50+
ISV partners under our unique, patented licensing model

2500+
Engineers, scientists and creative thinkers

5000+
Customers globally

60,000+
Users
Altair is the Only Company That…

...makes HPC middleware:

Altair | PBS Works™

...develops HPC applications:

Altair | HyperWorks®

...and uses these to do HPC every day:

Altair | ProductDesign
PBS Professional: Optimize HPC

Workload Management & Job Scheduling

EFFICIENT AND EFFECTIVE UTILIZATION OF EXPENSIVE HPC RESOURCES

- **Ubiquity**: Open Source and commercial licensing
- **Performant**: faster time-to-results, better throughput and utilization
- **Secure**: EAL3+ security certification and MLS support
- **Intelligent**: policy-driven and topology-aware scheduling
- **Scalable**: proven to run millions of jobs per day
- **Green Provisioning™**: power management and control
- **Robust**: known in the industry for stability and support
- **Open Architecture**: implement virtually any new policy
The HPC World

Public Sector
- Risk takers
- Early adopters
- Natural Collaborators
- Open Source

Private Sector
- Risk averse
- Later adopters
- Natural Competitors
- Commercial

PBS Pro®
Dual-licensing
Aggressively Open & Community Oriented

• Opening the full core of PBS Professional
  - not just a weak subset or older version

• Reorganizing to behave as one of the
  (hopefully many) contributors

• Using community-accepted practices
  - OSI-approved license
  - GitHub, JIRA, open roadmaps, …

• Focusing on longevity
  - For a viable, sustainable community

www.pbspro.org
Industry-leading workload manager and job scheduler for high-performance computing

Downloads

Support Options
Communities and Commercial

Source Code
Fork Me on GitHub
pbspro.org – Community Forum

Join the conversation by connecting via the community forum:

**Announcements**
Important updates relevant to the entire PBS Pro community

**Users**
General questions and discussions among end users

**Developers**
Technical discussions among developers
To dive in deeper and learn more about the project and what the community is up to, visit:

**Contributor's portal**
Includes roadmaps, processes, how-to articles, coding standards, release notes, etc.

**Source code**
Includes full source and test framework — fork and submit pull requests to contribute

**Issue tracking system**
Includes bugs and feature requests and status

pbspro.org – Contributor Portal
Is this “OpenPBS”? 

No, it’s full PBS Professional…
(Please call it “PBS Professional” or “PBS Pro”.)

OpenPBS was from 1999.
Nurturing a community “standard” HPC stack

Founding member with Intel, HPE, others
  + Technical Steering Committee (Scott Suchyta)

PBS Professional included since v1.2
  (released at SC16)

“Altair’s decision to open source PBS Professional, an established, production hardened HPC middleware technology, is a key milestone in moving the OpenHPC community closer to achieving exascale computing,” says Charles Wuischpard, vice president and general manager of HPC Platform Group at Intel Corporation.
PBS Pro: Speed, Scale, Resilience & The “Whole” HPC World

• 5x more scalable
  • Tested to 50,000 nodes
  • Tested to 500 concurrent portal users

• 10x faster
  • Both scheduling & analytics

• 10x smaller footprint
  • Analytics memory & disk

• 100% health check framework

• Dual licensing: Commercial & Open Source

PBSPro.org
# A Few Examples of Innovation Projects . . .

<table>
<thead>
<tr>
<th>Feature</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultiSched (parallel schedulers)</td>
<td></td>
</tr>
<tr>
<td>Green Provisioning™ (power mgmt)</td>
<td></td>
</tr>
<tr>
<td>Conditional resource requests</td>
<td></td>
</tr>
<tr>
<td>Filtering resource requests</td>
<td></td>
</tr>
<tr>
<td>Release limited resources on suspend</td>
<td></td>
</tr>
<tr>
<td>ARM64 (coming soon)</td>
<td></td>
</tr>
<tr>
<td>Release nodes from running jobs</td>
<td></td>
</tr>
<tr>
<td>Add fairshare to job sort formula</td>
<td></td>
</tr>
<tr>
<td>Move scheduler config to qmgr</td>
<td></td>
</tr>
<tr>
<td>Fast job launch (PMIx)</td>
<td></td>
</tr>
<tr>
<td>IPv6 support</td>
<td></td>
</tr>
<tr>
<td>DRMAA v2 support</td>
<td></td>
</tr>
<tr>
<td>PBS Plugins</td>
<td></td>
</tr>
<tr>
<td>Docker &amp; nvidia-docker</td>
<td></td>
</tr>
<tr>
<td>Backfill using estimated walltime</td>
<td></td>
</tr>
<tr>
<td>10k jobs/min throughput</td>
<td></td>
</tr>
<tr>
<td>MUNGE</td>
<td></td>
</tr>
<tr>
<td>MUNGE</td>
<td></td>
</tr>
<tr>
<td>Hadoop integration</td>
<td></td>
</tr>
<tr>
<td>Native Package Manager Installs</td>
<td></td>
</tr>
<tr>
<td>CYLC Workflow Engine</td>
<td></td>
</tr>
<tr>
<td>Modify reservation start/end</td>
<td></td>
</tr>
<tr>
<td>Xeon Phi</td>
<td></td>
</tr>
<tr>
<td>Active-active resilience</td>
<td></td>
</tr>
<tr>
<td>Burst buffer (NVRAM) scheduling</td>
<td></td>
</tr>
<tr>
<td>Job equivalence classes</td>
<td></td>
</tr>
<tr>
<td>MLS with SELinux (multi-level security)</td>
<td></td>
</tr>
<tr>
<td>Job &amp; Fairshare Formulas</td>
<td></td>
</tr>
<tr>
<td>Advanced Allocation Management</td>
<td></td>
</tr>
</tbody>
</table>
PBS Architecture & Communication

Inter-daemon Communication Using TPP

- PBS Scheduler
- PBS Server
- pbs_comm
- PBS Commands
- PBS MoM
- Job Task
- MPI
- TPP
- TCP

© 2017 Altair Engineering, Inc. All rights reserved.
Tunable Scheduling Formula

Define Any Policy – Including On-the-Fly “Exceptions”

Simple formulas are very simple (big jobs go first)

\[ \text{ncpus} \times \left( \text{walltime}/3600.0 \right) \]

Complex formulas are pretty simple too… (adds priority accrual for smaller jobs, high-priority queue, deferred queue, “run this job next”)

\[ (\text{ncpus} \times \left( \text{walltime}/3600.0 \right)) \times \text{Wsize} + \\
(\text{eligible_time}/3600.0) \times \text{Wwait} + \\
\text{special_p} \]
Topology-aware Scheduling

Faster, Predictable Performance and Better Utilization

Inter-node & intra-node placement
Switches, clusters, and NUMA

All networks
Infiniband, Ethernet, custom

Dynamic (runtime changeable)

All topologies

Average runtimes
~ 45% Faster

** actual US Customer Reported Results
PBS Plugins ("Hooks")

Framework for Agility and Innovation

- Meet unique enterprise requirements
- Deliver site-specific and 3rd party integrations
- Support platform-specific features on day 1
- Build novel extensions
- Crowdsource solutions
- Share (even sell) via PBS ecosystem

Say “Yes!” to unique requests
Learn More @SC17

Join Altair Tiki Booth at SC17
Booth #1117

User technology talks, PBS Works demos, free hands-on tutorials… and much more