Lonestar 5: Customizing the Cray XC40 Software Environment

Cyrus Proctor and David Gignac

Robert McLay, Si Liu, Doug James, Tommy Minyard, and Dan Stanzione
Overview

• Introduction to TACC and LS5
• Motivation for XC40 software customizations
• Live demonstration
• Discussion of customizations made
Who We Are

• Texas Advanced Computing Center (TACC)
• Based at University of Texas in Austin
• 160 full-time technology & support staff
• 15K sq. ft. data center with 10 MW capacity
• >10 production and experimental systems
• >10K users in 2100 unique research projects
Our Focus Today: Lonestar 5

- Cray XC40
  - 5th in series; Production started in January 2016
  - 1252 Intel HSW/IVB nodes w/ Aries interconnect
  - 30K cores @ 1.25 PFLOPs
  - 5 PB & 20 PB Scratch & Work Lustre filesystems
Our Challenge:

• New system rollout replacing Lonestar 4
• How do we:
  – Transition 3500 LS4 users?
  – Support existing and future workflows?
  – Minimize disruption and need for retraining?
We Need Lonestar 5 to Provide

• Familiarity:
  – Recognizable to those who use other TACC resources

• Consistency:
  – Same Full Linux environment everywhere

• Flexibility:
  – Adaptive platforms that respond and evolve to meet user needs

…All while maintaining performance at scale
LS5 Modification Summary

- Customized network configuration
- Tailored workload management
- Shell startup behavior
- Hierarchical software stack
- Adaptive virtual esLogin nodes
- MPI distribution wrappers
- RPM development environment
Live Demonstration of Lonestar 5 XC40 Environment
Customized Network Configuration

- Direct Slurm traffic from esLogins to Slurm database node
- No need for internal login node or esWrapper module
- Ability to access compute nodes via ssh
Tailored Workload Management

- Run Slurm in “native” mode
- Custom Slurm plugins to qualify user jobs
- PAM Slurm module
- Extended Prolog and Epilog features
Virtual esLogin Nodes

• Solution for specialized workflows
• Quick deployment of a reproducible, agile configuration
• Capable of supporting International Traffic in Arms Regulations (ITAR) data
MPI Distribution Wrappers

- Access to Cray MPICH distribution
- Provides same expected performance found in Cray environment
- Includes scripts for mpicc, mpicxx, and mpif90 for Intel and GCC compilers
- Provide standardized job launcher (ibrun)
Shell Startup Behavior

• Provides consistent, full Linux environment on esLogin and compute nodes
• Allows two independent Cray and TACC startup environments
Hierarchical Software Stack

• Lmod: A new Environment Module system
• Location-based dependencies
• Automatic dependency resolution
• Multiple enhancements (e.g. reset, swap, collections)
RPM Build Environment

• Sandboxed change-rooted environment
• Uses a copy of compute node image
• Develop RPMs without impacting production status
• Capable of multiple, reproducible, and consistent instances
Conclusion

• We’re coming up on 5 months in production (tomorrow)
• We judge that the transition was smooth for our users
• Lonestar 5 offers a variation of the Cray environment and a highly customized vendor-independent, open science TACC environment
• More to be done: we are constantly learning, adapting and improving
Thank you for your time and attention!

{cproctor,dgignac}@tacc.utexas.edu