Outline

• ASC Computing Strategy
• Project Drivers and Procurement Process
• Platform Architecture Overview
• Schedule and Status
• Questions, and maybe some answers
ASC computing strategy

• Approach: Two classes of systems
  – Advanced Technology: First of a kind systems that identify and foster technical capabilities and features that are beneficial to ASC applications
  – Commodity Technology: Robust, cost-effective systems to meet the day-to-day simulation workload needs of the program

• Investment Principles
  – Maintain continuity of production
  – Ensure that the needs of the current and future stockpile are met
  – Balance investments in system cost-performance types with computational requirements
  – Partner with industry to introduce new high-end technology constrained by life-cycle costs
  – Acquire right-sized platforms to meet the mission needs
Advanced Technology Systems

• Leadership-class platforms
• Pursue promising new technology paths with industry partners
• These systems are to meet unique mission needs and to help prepare the program for future system designs
• Includes Non-Recurring Engineering (NRE) funding to enable delivery of leading-edge platforms
• Trinity (ATS-1) will be deployed by ACES (New Mexico Alliance for Computing at Extreme Scale, i.e. Los Alamos & Sandia) and sited at Los Alamos
• ATS-2 will be led by LLNL, ATS-3 by ACES, etc
Trinity Project Drivers

• Satisfy the mission need for more capable platforms
  – Trinity is designed to support the largest, most demanding ASC applications
  – Increases in geometric and physics fidelities while satisfying analysts time to solution expectations
  – Foster a competitive environment and influence next generation architectures in the HPC industry

• Trinity is enabling new architecture features in a production computing environment (ATS Components)
  – Tightly coupled solid state storage serves as a “burst buffer” for checkpoint/restart file I/O & data analytics, enabling improved time to solution efficiencies
  – Advanced power management features enable measurement and control at the system, node and component levels, allowing exploration of application performance/watt and reducing total cost of ownership
  – Trinity’s architecture will introduce new challenges for code teams: transition from multi-core to many-core, high-speed on-chip memory subsystem, wider SIMD/vector units
Trinity/NERSC8 Procurement Process Timeline

- ACES (LANL/SNL) Project started November 2011
- Market Survey started January 2012
- Partnered with LBL/NERSC on RFP (NERSC 8) March 2012
- CD-0, Draft Technical Requirements and RFI issued December 2012
- Formal Design Review completed April 2013
- Independent Project Review (Lehman) completed May 2013
- CD-1, Trinity/NERSC8 RFP issued August 2013
- Technical Evaluation of the proposals completed September 2013
- Initial negotiations for both systems completed November 2013
- NNSA Independent Cost Review completed Jan 2014
- CD-2/3, NERSC8 awarded April 2014
- CD-2/3, Trinity awarded July 2014 after Best and Final Offer (BAFO)
Trinity Platform Solution

• Cray has been awarded the contract, July 2014
  – Based on mature Cray XC30 architecture

• with Trinity introducing new architectural features
  – Intel Knights Landing processor
  – Burst Buffer storage nodes
  – Advanced power management system software enhancements

• A single system that contains both Intel Haswell and Knights Landing (KNL) processors
  – Haswell partition satisfies FY15 mission needs (well suited to existing codes) and fits the FY15 budget profile.
  – KNL partition delivered in FY16 results in a system significantly more capable than current platforms, provides the application developers with an attractive next generation target, and fits the FY16 budget profile.

• Managed Risk
  – Cray XC30 architecture minimizes system software risk and provides a mature high-speed interconnect
  – Haswell partition is low risk as technology is available Fall CY14
  – KNL is higher risk due to new technology, but provides a good path for codes teams to transition to many-core architecture
Trinity High-Level Architecture

Cray Compute and Storage Infrastructure for “Trinity”

Cray XC Family Supercomputer

- Compute (Intel “Haswell”) >9500 nodes
- Compute (Intel Xeon Phi) >9500 nodes

42PF Total Performance and 2.1PiB of Total Memory

- Gateway Nodes
- Lustre Routers
- Burst Buffer

Cray Development & Login nodes

40 GigE Network

GigE Network

2x 648 Port IB Switches

41PB file system

41PB file system

82 PB Usable ~1.7 TB/sec - 2 Filesystems

Cray Sonexion® Storage System
# Trinity Architecture Details

<table>
<thead>
<tr>
<th>Metric</th>
<th>Trinity</th>
<th>Haswell Partition</th>
<th>KNL Partition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node Architecture</td>
<td>KNL + Haswell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Capacity</td>
<td>2.11 PB</td>
<td>&gt; 1 PB</td>
<td>&gt;1 PB</td>
</tr>
<tr>
<td>Memory BW</td>
<td>&gt;6 PB/sec</td>
<td>&gt; 1 PB/s</td>
<td>&gt;1PB/s + &gt;4PB/s</td>
</tr>
<tr>
<td>Peak FLOPS</td>
<td>42.2 PF</td>
<td>11.5 PF</td>
<td>30.7 PF</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>19,000+</td>
<td>&gt;9,500</td>
<td>&gt;9,500</td>
</tr>
<tr>
<td>Number of Cores</td>
<td>&gt;760,000</td>
<td>&gt;190,000</td>
<td>&gt;570,000</td>
</tr>
<tr>
<td>Number of Cabs (incl I/O &amp; BB)</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PFS Capacity (usable)</td>
<td>82 PB usable</td>
<td>&gt; 8x Cielo</td>
<td></td>
</tr>
<tr>
<td>PFS Bandwidth (sustained)</td>
<td>1.45 TB/s</td>
<td>&gt; 10x Cielo</td>
<td></td>
</tr>
<tr>
<td>BB Capacity (usable)</td>
<td>3.7 PB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB Bandwidth (sustained)</td>
<td>3.3 TB/s</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Compute Node Specifications

<table>
<thead>
<tr>
<th></th>
<th>Haswell</th>
<th>Knights Landing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Capacity (DDR)</td>
<td>2x64=128 GB</td>
<td>Comparable to Intel® Xeon® processor</td>
</tr>
<tr>
<td>Memory Bandwidth (DDR)</td>
<td>136.5 GB/s</td>
<td>Comparable to Intel® Xeon® processor</td>
</tr>
<tr>
<td># of sockets per node</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td># of cores</td>
<td>2x16=32</td>
<td>60+ cores</td>
</tr>
<tr>
<td>Core frequency (GHz)</td>
<td>2.3</td>
<td>N/A</td>
</tr>
<tr>
<td># of memory channels</td>
<td>2x4=8</td>
<td>N/A</td>
</tr>
<tr>
<td>Memory Technology</td>
<td>2133 MHz DDR4</td>
<td>MCDRAM &amp; DDR4</td>
</tr>
<tr>
<td>Threads per core</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Vector units &amp; width (per core)</td>
<td>1x256 AVX2</td>
<td>AVX-512</td>
</tr>
<tr>
<td>On-chip MCDRAM</td>
<td>N/A</td>
<td>Up to 16GB at launch, over 5x STREAM vs. DDR4</td>
</tr>
</tbody>
</table>
Trinity Capabilities

• Each partition will accommodate 1 to 2 large mission problems (2 to 4 total)

• Capability relative to Cielo
  – 8x to 12x improvement in fidelity, physics and performance
  – > 30x increase in peak FLOPS
  – > 2x increase in node-level parallelism
  – > 6x increase in cores
  – > 20x increase in threads
The Trinity Center of Excellence & Application Transition Challenges

- **Center of Excellence**
  - Work with select NW application code teams to ensure KNL Partition is used effectively upon initial deployment
  - Nominally one application per laboratory (SNL, LANL, LLNL)
  - Chosen such that they impact the NW program in FY17
  - Facilitate the transition to next-generation ATS code migration issues
  - This is NOT a benchmarking effort

- **Intel Knights Landing processor**
  - From multi-core to many-core
  - > 10x increase in thread level parallelism
  - A reduction in per core throughput (1/4 to 1/3 the performance of a x86-64 core)
  - MCDRAM: Fast but limited capacity (~5x the BW, ~1/5 the capacity of DDR4 memory)
  - Dual AVX-512 SIMD units

- **Burst Buffer**
  - Data analytics use cases need to be developed and/or deployed into production codes
  - Checkpoint/Restart should “just work”, although advanced features may require code changes
Trinity Project Team

Trinity Executive Committee
ASC Execs, LANL
ASC Execs, SNL
ACES Co-Directors
Project Manager
System Architect

ACES Co-Directors
Gary Grider, LANL
Bruce Hendrickson, SNL

Trinity Project Director
Manuel Vigil
Chief Architect
Doug Doerfler

NNSA OCIO
Advisors and Compliance

Federal Project Director
NNSA

System
Architecture
Doug Doerfler
Josip Loncaric

Center of Excellence
Rob Hoekstra
Tim Kelley
Shawn Dawson

Project Management, Security
Manuel Vigil
Jim Lujan
Alex Malin

Facilities and Trinity Installation
Ron Velarde

System Integration and Deployment
David Morton

Operations Planning
Jeff Johnson
Bob Ballance

Acquisition
Darren Knox

System Architecture
Doug Doerfler
Josip Loncaric

Center of Excellence
Rob Hoekstra
Tim Kelley
Shawn Dawson

Project Management, Security
Manuel Vigil
Jim Lujan
Alex Malin

Facilities and Trinity Installation
Ron Velarde

System Integration and Deployment
David Morton

Operations Planning
Jeff Johnson
Bob Ballance

Acquisition
Darren Knox

System Architecture
Doug Doerfler
Josip Loncaric

Center of Excellence
Rob Hoekstra
Tim Kelley
Shawn Dawson

Project Management, Security
Manuel Vigil
Jim Lujan
Alex Malin

Facilities and Trinity Installation
Ron Velarde

System Integration and Deployment
David Morton

Operations Planning
Jeff Johnson
Bob Ballance

Burst Buffer
Cornell Wright

Advanced Power Management
Jim Laros

Hardware Architecture
Scott Hemmert

Acceptance
Jim Lujan

External Networks and Archiving
Parks Fields, Kyle Lamb

System Software Stack
Daryl Grunau

File System
Brett Kettering

Application Readiness
Cornell Wright
Joel Stevenson

R&D

Software Architecture
Kevin Pedretti

Viz
Laura Monroe
Questions

Manuel Vigil
mbv@lanl.gov

Douglas Doerfler
dwdoerf@sandia.gov