Commodity Technology Systems 2 (CTS-2) Update

September 2022

Matt Leininger
CTS-2 Technical Lead
Livermore Computing
& Tri-Lab CTS-2 Team
ASC Platform Timeline

Fiscal Year

ATS-1: Trinity (LANL/SNL)
ATS-2: Sierra (LLNL)
ATS-3: Crossroads (LANL/SNL)
ATS-4: El Capitan (LLNL)
ATS-5 (LANL)
ATS-6 (LLNL)
CTS-1
CTS-2
CTS-3
Astra (SNL)
Vanguard-2 (SNL)
Vanguard-3 (SNL)

Acceptance
Plan
NRE
Use
Retire

March 2022
A Tri-lab (LLNL, LANL, Sandia) joint procurement to provide a common hardware platform for robust capacity computing platforms. CTS provides common simulation & operating environment when coupled to Tri-lab software stack (TOSS) & Tri-Lab Common Environment (TCE).

**What is CTS?**

**CTS Advantages**

- Reduce total cost of ownership associated with system procurements, integrations, and support.
- Rapid deployment of systems into production.
- Helps enable TCO reductions realized from TOSS.
- Supports ASC, DP & other computing programs.
- Strategy adaptive to new technology & workloads.
- Supports computing demands for HPC, data sciences, ML/AI, and visualization.
A Brief History of NNSA Commodity Systems

Commodity Platforms have successfully delivered to simulation programs at LLNL, Sandia, and LANL for 15 years and growing.
CTS-2 and Tri-Lab software teams (TOSS & TCE) work together throughout the CTS-2 procurement, deployment, and platform lifetimes.
CTS-2 Contract Awarded to Dell Technologies

- Compute Node Servers: Dell C6620 (4 nodes in 2U)
- Login/Mgmt/Gateway Servers: Dell 760 2U
- Intel Sapphire Rapids Xeon CPUs
- HPC Network: Cornelis Networks Omni-Path or Mellanox InfiniBand
- Liquid Cooling: CoolIT direct-to-chip liquid cooling
- Power Density: Up to ~66kW/compute rack
- Power: 480V 3 phase
- Options for GPUs: NVIDIA, AMD, & Intel
First Dell CTS-2 Systems Arriving in Late 2022

- System deliveries begin in August 2022
- Intel Sapphire Rapids CPUs (4\textsuperscript{th} Generation Xeon-SP)
  - 56 cores/socket or 112 cores/node
  - \textbf{Default}: 256 GB DDR5
  - Alternative: 512 GB DDR5
  - Alternative: 128 GB HBM2e
- Cornelis 200 Gb/s High Speed Network
- Software Environment
  - TOSS4 based on RHEL 8.x
  - Tri-Lab Common Environment (TCE2)

### CTS-2 Scalable Unit (SU)

<table>
<thead>
<tr>
<th>Nodes</th>
<th>CPU Cores</th>
<th>Memory Capacity</th>
<th>Theoretical Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>192</td>
<td>~21.5K</td>
<td>DDR5: 49 TB</td>
<td>1.5 PF/s (FP64)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDR5: 98 TB</td>
<td>3.0 PF/s (FP32)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HBM2e: 25 TB</td>
<td>6.1 PF/s (FP16/BF16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>49 PF/s (BF16/new TMUL op)</td>
</tr>
</tbody>
</table>
CTS-2 Base Node Architecture & Performance Estimates

**CTS-2 Base Architecture**

CPU + DDR5
- 2 socket 56 core 2.0+ GHz CPU
- 16x16 GB DDR5 DIMM no chipkill
- Total memory 256 GB DDR5
- ~2.2-2.4 GB/CPU core
- ~615 GB/s memory BW (theoretical)
- ~460 GB/s memory BW (actual)

<table>
<thead>
<tr>
<th>Sapphire Rapids Processor</th>
<th>Overall Node FOM</th>
<th>HPCG</th>
<th>LAGHOS</th>
<th>Quicksilver</th>
<th>SNAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDR (relative to CTS-1)</td>
<td>4.9x</td>
<td>3.2x</td>
<td>6.2x</td>
<td>5.3x</td>
<td>6.5x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sapphire Rapids Processor</th>
<th>Mini-app Node Performance</th>
<th>Power efficiency</th>
<th>Perf per $</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDR (relative to CTS-1)</td>
<td>4.9x</td>
<td>1.7x</td>
<td>~3x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sapphire Rapids Processor</th>
<th>DGEMM (Est.)</th>
<th>STREAM (per Socket)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDR</td>
<td>3.1 – 3.3 TF</td>
<td>230 GB/s</td>
</tr>
</tbody>
</table>

CTS-2 DDR will be available summer 2022
Maximize small/medium job throughput with 2:1 Tapered Network and more compute nodes!
Scalable Units are the “Legos” that scale to Multi-SU platforms

<table>
<thead>
<tr>
<th>#SU’s</th>
<th># Nodes</th>
<th>PFlops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>192 - 200</td>
<td>1.45 – 1.5</td>
</tr>
<tr>
<td>2</td>
<td>384 - 400</td>
<td>2.9 – 3.0</td>
</tr>
<tr>
<td>4</td>
<td>768 - 800</td>
<td>5.8 - 6</td>
</tr>
<tr>
<td>6</td>
<td>1,152 – 1,200</td>
<td>8.7 – 9.0</td>
</tr>
<tr>
<td>8</td>
<td>1,536 – 1,600</td>
<td>11.6 – 12.0</td>
</tr>
<tr>
<td>12</td>
<td>2,304 – 2,400</td>
<td>17.3 – 18.0</td>
</tr>
</tbody>
</table>

One or more of each cluster size may be deployed at the Tri-labs
CTS-2 CoolIT Direct-to-Chip Liquid Cooling

CoolIT CHx750 CDU

CoolIT CHx80 CDU

Dell C6620 Compute Node
Evolution of Power & Cooling Requirements for Commodity Systems

Technology power density has reached a threshold where liquid cooling is a requirement for HPC platforms!

- TLCC1
  - 26 kW per rack
  - 26 kW air cooled
  - 208V power

- TLCC2
  - 28 kW per rack
  - 28 kW air cooled
  - Fielded 1 liquid cooled solution
  - 208V power

- CTS-1
  - 24 kW per rack
  - 8 kW air cooled
  - 16 kW liquid cooled
  - 480V & 208V power

- CTS-2
  - Next generation
  - 66 kW per rack
  - All solutions liquid cooled
  - 21 kW air cooled
  - 45 kW liquid cooled
  - 480V power

Power and Cooling Trends for Commodity Systems

<table>
<thead>
<tr>
<th></th>
<th>TLCC1</th>
<th>TLCC2</th>
<th>CTS-1</th>
<th>CTS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilowatts (kW) / rack</td>
<td>75</td>
<td>75</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>Air</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liquid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- TLCC1
  - 26 kW per rack
  - 26 kW air cooled
  - 208V power

- TLCC2
  - 28 kW per rack
  - 28 kW air cooled
  - Fielded 1 liquid cooled solution
  - 208V power

- CTS-1
  - 24 kW per rack
  - 8 kW air cooled
  - 16 kW liquid cooled
  - 480V & 208V power

- CTS-2
  - Next generation
  - 66 kW per rack
  - All solutions liquid cooled
  - 21 kW air cooled
  - 45 kW liquid cooled
  - 480V power
CTS-2 Architecture:
4th Generation Intel Sapphire Rapids Xeon CPUs

- Direct evolution of TLCC2 and CTS-1 CPUs
- Mature software solutions and ecosystem
- 2-socket servers with 56 CPU cores/socket
- 256 GB DDR5 per node / ~2.3 GB/core
- 8 channels DDR5/socket
- Same CPU as NNSA ASC Crossroads platform
- Smooth application transition from CTS-1 → CTS-2 → ATS-3 Crossroads
- BFloat16 and other new CPU features to support accelerated ML
Phase 1 Deployment Plan for CTS-2 Platforms

- First wave of LLNL Systems: 256 GB DDR5 Base Configuration

<table>
<thead>
<tr>
<th>Lab</th>
<th>Name</th>
<th># SU’s</th>
<th># Nodes</th>
<th>FP64 PetaFlops</th>
<th>Total Memory Capacity</th>
<th>Delivered to Lab</th>
<th>Est. General Availability</th>
<th>Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLNL</td>
<td>Mutt</td>
<td>1 SU</td>
<td>192</td>
<td>1.5 PF/s</td>
<td>49 TB</td>
<td>8/16/22</td>
<td>9/22</td>
<td>ASC (CZ)</td>
</tr>
<tr>
<td>LLNL</td>
<td>Whippet</td>
<td>&lt; 1 SU</td>
<td>32</td>
<td>0.25 PF/s</td>
<td>8 TB</td>
<td>8/16/22</td>
<td>9/22</td>
<td>ASC (RZ)</td>
</tr>
<tr>
<td>LLNL</td>
<td>Poodle</td>
<td>&lt; 1 SU</td>
<td>32</td>
<td>0.25 PF/s</td>
<td>8 TB</td>
<td>8/16/22</td>
<td>9/22</td>
<td>ASC (CZ)</td>
</tr>
<tr>
<td>LLNL</td>
<td>Hound</td>
<td>2 SU</td>
<td>384</td>
<td>3.0 PF/s</td>
<td>93 TB</td>
<td>9/27/2022</td>
<td>1QCY2023</td>
<td>ASC (RZ)</td>
</tr>
<tr>
<td>LLNL</td>
<td>Bengal</td>
<td>6 SU</td>
<td>1,152</td>
<td>9.1 PF/s</td>
<td>295 TB</td>
<td>9/27/2022</td>
<td>1QCY2023</td>
<td>ASC (SCF)</td>
</tr>
<tr>
<td>SNL</td>
<td>Amber</td>
<td>8 SU</td>
<td>1,536</td>
<td>12.1 PF/s</td>
<td>393 TB</td>
<td>11/8/2022</td>
<td>1QCY2023</td>
<td>ASC</td>
</tr>
<tr>
<td>LLNL</td>
<td>Dane</td>
<td>8 SU</td>
<td>1,536</td>
<td>12.1 PF/s</td>
<td>393 TB</td>
<td>12/20/2022</td>
<td>1QCY2023</td>
<td>ASC, Inst., PSAAP (CZ)</td>
</tr>
<tr>
<td>SNL</td>
<td>Stout</td>
<td>8 SU</td>
<td>1,536</td>
<td>12.1 PF/s</td>
<td>393 TB</td>
<td>1/31/2023</td>
<td>1QCY2023</td>
<td>Inst.</td>
</tr>
</tbody>
</table>

Additional CTS-2 systems deployment will depend on funding, power, cooling, and data center space permit.
CTS-2: Mutt, Whippet, & Poodle
Supply Chain Issues, Lease to Own, & Deployment Timeline

- Dell and CTS-2 components suppliers all report supply chain issues
  - Some components have 6-9 months lead time
  - These issues will likely continue through at least all of 2022, if not longer.

- Dell working with suppliers to mitigate risks as much as possible
  - NNSA supplying accurate forecasts for system orders will help.
  - Dell prioritizing CTS-2

- NNSA received a Defense Priority (DPAS) rating for CTS-2
  - Places CTS-2 systems on high priority list with system integrator and component suppliers.
  - CTS-2 will have a DX rating (perhaps DO for some systems)
  - El Capitan has a “DO” DPAS rating
  - DPAS will help, but the application of rating is on specific components & quantities – not entire systems
Ordering a new CTS-2 system takes time
- 0-2 months for modifying build of materials for specific system and data center
- 1-2 months for setting up lease to own (LTO)
- Formally order system
- 1-2 months for Dell to acquire parts (best case)
- 1 month to build, test, and deliver to lab
- At least 1 month to integrate, test, and stabilize system at Lab

4–8 months needed from time program decides to order system until system production
CTS-2 Market Inflection Point:
CPU+DDR vs. CPU+HBM Price Parity

**CPU + DDR5**
- 2S ~64 core 2.0+ GHz CPU
- 16x32 GB DDR5 DIMM no chipkill
- Total memory 512 GB DDR5
- ~5+ GB/CPU core
- ~615 GB/s memory BW (theoretical)
- ~492 GB/s memory BW (actual?)

**CPU + HBM**
- 2S ~64 core 2.0+ GHz CPU
- 8x16 GB HBM2e no chipkill
- Total memory 128 GB HBM2e
- ~1.3 GB/CPU core
- ~3,200 GB/s memory BW (theoretical)
- ~2,200 GB/s memory BW (actual?)

Is 3-4X the memory BW but ¼ the memory capacity a good trade-off? Are ASC codes and capacity workloads ready and able to utilize the extra bandwidth? Would at least a 30% performance improvement be worth ¼ of the memory capacity?
CTS-2 Next Steps

- CTS-2 platforms arriving in early to mid-2022
- More detailed information on CTS-2 will be available late summer 2021
- Early deployment experiences and performance in late summer 2022
- Questions?
Thanks to Tri-Lab CTS-2 Team!

- CTS-2 Tri-Lab Technical Leads
  - Matt Leininger (LLNL)
  - Trent D’Hooge (LLNL)
  - Randy Scott (SNL)
  - Steve Monk (SNL)
  - Amanda Bonnie (LANL)
  - Jim Lujan (LANL)
  - Ian Karlin (LLNL)
  - Jim Foraker (LLNL)
  - Jim Silva (LLNL)
  - Brian Ryjuin (LLNL)
  - Riyaz Haque (LLNL)
  - Clay England (LLNL)
  - Jeff Ogden (SNL)
  - Mike Glass (SNL)
  - Si Hammond (now NNSA HQ)
  - Doug Pase (SNL)
  - Phil Regier (SNL)
  - Brad Settlemyer (LANL)
  - Brett Holman (LANL)
  - Jesse Martinez (LANL)
  - Timothy Randles (LANL)
  - Galen Shipman (LANL)
  - Trent D’Hooge (LLNL)
  - Chris DePrater (LLNL)
  - Brandon Hong (LLNL)
  - Rigo Delgado (LLNL)
  - Lance Weems (LLNL)
  - Larry Baca (SNL)
  - Anthony Agelastos (SNL)
  - Jay Livesay (SNL)
  - Dave Martinez (SNL)
  - David Smith (SNL)
  - Eloy Romero Jr. (LANL)
  - Nathan Debardeleben (LANL)
  - Hai Ah Nam (LANL)
  - Howard Pritchard (LANL)
  - Mike Lang (LANL)
  - Gary Ward (LLNL)
  - Sofia Delgadillo-Marruco (SNL)
  - Patrick Padilla (LANL)

- Former members
  - Josip Loncaric (LANL)
  - Ian Karlin (LLNL)
  - Robin Goldstone (LLNL)

Many additional subject matter experts provided reviews and feedback.
Questions?

CTS-2 Lead: Matt Leininger (matt@llnl.gov)

Cluster Integration Lead: Trent D’Hooge (dhooge1@llnl.gov)

CTS-2 Integration Lead: Jim Silva (silva50@llnl.gov)