NCAR’s Next Procurement: Meeting Users’ Reliability and Storage Demands

DAVID L HART
NCAR User Services Manager
iCAS 2019 — 12 SEPTEMBER 2019
Where we are: NCAR’s Cheyenne system

HPE ICE XA Cluster with 4,032 dual-socket Intel Broadwell nodes
- No GPGPU nodes
- Heterogeneity limited to 64/128 GB nodes

“Conventional” 5.34-PFLOPS cluster aimed at conventional HPC modeling capabilities and practices
- What the users wanted at the time

Times have changed.

https://doi.org/10.5065/D6RX99HX
Preparing for NWSC-3: NCAR’s third petascale system

• A lot has happened since NCAR began procuring (ca. 2015) and deployed (2017) Cheyenne
  – Machine learning
  – Cloud maturity in HPC
  – Dynamic technology landscape
  – Containers
  – Pangeo, Jupyter Notebooks & Hubs
  – Workflow engines (Cylc, Rocoto) and continuous integration in model development
  – Storage and data management requirements

• While many of these existed earlier, most fully entered mainstream HPC and/or Earth systems science only in the past few years.
NWSC-3 procurement schedule

- NCAR modified its procurement process to address uncertainties in the technology space.
- Notably, we issued a “Request for Information” followed by daylong co-design meetings with vendors.
  - Opportunities to explore alternatives, clarify misconceptions, and set expectations
- We kept roughly the same process for gathering science requirements and analyzing our workload.
  - But we gleaned new insights

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late 2018 – Mid-2019</td>
<td>Benchmark design Technology briefings and co-design meetings Science requirements &amp; workload analysis</td>
</tr>
<tr>
<td>Summer 2019</td>
<td>Preparation &amp; review of Technical Specifications</td>
</tr>
<tr>
<td>Early 2020</td>
<td>RFP release</td>
</tr>
<tr>
<td>Mid-2020</td>
<td>Vendor selection and approval</td>
</tr>
<tr>
<td>Mid-2020 – Early 2021</td>
<td>Facility preparation</td>
</tr>
<tr>
<td>Mid-2021</td>
<td>Phase 1: Delivery, installation and acceptance</td>
</tr>
<tr>
<td>Early 2022</td>
<td>Phase 2: Delivery, installation and acceptance</td>
</tr>
<tr>
<td>Late 2022</td>
<td>Decommission Cheyenne</td>
</tr>
</tbody>
</table>
The initial context for the NWSC-3 procurement

- We approached users in terms of four key questions
  - Make the complexity a bit more tractable
  - Encapsulate the major hardware choices anticipated by CISL
- **Question A:** How much to spend on compute versus storage?
  - A = 80% has been our typical investment
- **Question B:** How much to spend on HPC versus high-throughput computing?
  - B = 99% in the past
- **Question C:** How much to spend on CPU-based nodes versus GPU-accelerated nodes?
  - C = 100% for Cheyenne
- **Question D:** How much to spend on SSD disk?
The NWSC-3 Science Requirements Advisory Panel (SRAP)

- Group of 44 modelers, software engineers, and computational scientists
  - NCAR and University participants
  - Covering NCAR’s primary research domains, model development groups, and experts in data assimilation & machine learning
- SRAP discussed several input sources over three meetings
  - White papers of their 5-year science objectives
  - Cheyenne workload analysis
  - Community survey
- Final set of SRAP recommendations agreed to by “ballot,” and letter of consensus prepared.

NCAR’s Next Procurement — D. Hart — iCAS 2019
What we learned from the workload analysis – part 1

• Extreme scalability not demonstrated by user activity
  – Job scale on Cheyenne only slightly larger than Yellowstone patterns

• Need for large node-level memory not demonstrated by user jobs
  – More than 95% of Cheyenne jobs fit within the usable 45-GB on regular nodes
  – 21% of Cheyenne nodes have 128-GB memory
What we learned from the workload analysis – part 2

• 78% of all jobs scheduled on Cheyenne to date have been single-node, short-duration
  – But account for only 2% of core-hours delivered (40M core-hours)
  – PBS getting a non-HPC workout!

• Storage usage patterns do not show user need for substantial I/O bandwidth
  – No apparent need for I/O bandwidth any greater than the 300 GB/s available from Cheyenne to its file system
What we learned from the community survey – part 1

- Top Cheyenne aspects to improve
  - Reliability/availability/stability
  - Storage capacity, retention periods, data management tools
  - High-throughput job support

- Top Cheyenne aspects to keep
  - Flexible software environment
  - HPC capability and performance
  - Help Desk / Support team
  - Integrated storage and analysis environment

“If you could improve one thing about Cheyenne…”

“More storage”
“Reliability/stability”
“High Throughput jobs”
“Availability”
“Longer storage retention”
“Data management tools”
“User environment/software”
“Documentation/FAQ”
“Scheduler (various aspects)”
“Data science/analysis support”
“DevOps environment”
“Memory per node”
“Cloud computing integration”
“CPU mode (7)”
“GPUs”
“Green power sources”
“Larger allocation”
“Longer run time”
“Training, user education”
What we learned from the community survey – part 2

• Respondents would support greater investment in storage capacity
  – As well as more investment in development and analysis systems

• Even split on a non-trivial (~20%) investment in GPGPU

• Traditional batch access likely to remain preferred access method
  – But growing interest in containers, Jupyter, cloud storage integration, and ML/DL

How would you split the NWSC-3 budget between compute & storage?

<table>
<thead>
<tr>
<th>Percentage of Compute</th>
<th>Number of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than Cheyenne</td>
<td>5</td>
</tr>
<tr>
<td>80% Compute (like Cheyenne)</td>
<td>25</td>
</tr>
<tr>
<td>75%-70% Compute</td>
<td>30</td>
</tr>
<tr>
<td>65%-60% Compute</td>
<td>40</td>
</tr>
<tr>
<td>Less than 60% Compute</td>
<td>50</td>
</tr>
</tbody>
</table>
What we learned from the SRAP white papers

- SRAP white papers & meeting discussions echoed the workload study and survey responses
- Cheyenne’s compute capability was rarely a topic of in-person discussions
  - Plans for large-scale science covered in the white papers
- Top user issues were
  - Availability and reliability (not compute capability)
  - Storage capacity and policies (not SSDs, I/O bandwidth)
- Emerging system needs
  - Much more data assimilation
  - GPU-based modeling
  - Machine learning
  - Automated testing for model development
Five final SRAP recommendations

• Worth waiting for high-bandwidth memory—to a point
  – SRAP was briefed on general findings from the vendor co-design meetings
• No need to acquire user-accessible SSD-based file system
• Phased deployment for storage to allow for flexibility over the production period
• A substantial GPU partition needed for GPU-based applications and machine learning
• Enhanced reliability and availability features, where cost effective and feasible
Findings incorporated into our RFP technical specifications

• Reliability & availability
  – Changes to Cheyenne environment to allow more non-HPC components to be usable when HPC system is down
  – Explored notion of “cluster of clusters”

• Storage capacity
  – Reviewing compute-storage balance
  – Working with NCAR labs to quantify the trade-offs
  – No SSD-based user file system

• Capacity workload
  – Plan to deploy a larger, dedicated development environment
  – And expand the analysis environment

• GPU-based modeling
  – Plan to acquire non-trivial GPU-based partition
Storage challenges going into NWSC-3 era

- Challenges are not technical
  - I/O bandwidth is abundant
  - Short-term storage is plentiful
- The challenge is storage over time
  - Users want a year or more to analyze model runs
  - Users want data sets available for sharing for 1-5 years after initial publication
  - Users want (some) key results preserved for more than 5 years
- Accrued data becomes a greater challenge than managing access to compute resources
  - Opportunity costs of “data in residence”
- Furthermore, analyzing petabytes of data output is qualitatively different than analyzing tens or even hundreds of terabytes
CMIP6 effort highlights what other modelers are *not* doing

• At NCAR, the CMIP6 campaign involved extensive planning (beyond central CMIP6 planning)
  – Selecting and prioritizing runs
  – Calculating the aggregate compute cost
  – Estimating the total storage output
• Coordinated effort across NCAR labs
  – Manage HPC and storage access
  – Coordinating compute runs
  – Supporting post-processing and data management workflows
• Optimized post-processing scripts using lossless data compression
  – Selective variable output
• Well-defined procedures for metadata, output format, and data to be stored

Most other modeling experiments do not follow the same procedures.
Certainly lack the high-level coordination NCAR-wide across large-scale runs, and the consistency of practice.
“Right-sizing” the storage balance

- Understanding trade-offs between node capacity vs. storage capacity
  - Quantifying compute decrease per unit storage increase as budget shifted
- Calculating the storage needed for longer scratch retention
  - NCAR seeing roughly 1 PB scratch per week of retention
- Moving project space to extensible, slower bandwidth portion of the infrastructure
  - Retaining flexibility to acquire additional disk to meet user needs
Continuing storage evolution

Current systems and plans

- High-bandwidth POSIX disk
- **Slower-bandwidth POSIX disk**
  - Expanding use for medium-term workflow and analysis needs
  - Capacity over performance
- **Object store**
  - Exploring use cases for data sharing and preservation
- **Tape**
  - Current hardware end of life
  - Reduced scope and intended uses
- **Cloud**
  - Increased interest for both data sharing and cold archive

Plans from two years ago

- **SSD-based file system for NWSC-3**
  - No user-based needs identified
- **High-bandwidth POSIX disk**
  - Still in place, but constrained
  - For sharing curated collections (??)
- **Warm archive on disk**
  - In place and expanding
- **Replacement tape archive**
  - Largely similar scope and use
NCAR expanding use of and integration with the Cloud

• **Science @ Scale plan**
  – As described by Jeff de la Beaujardière earlier today
  – Increased analysis and data sharing via on- and off-premises cloud

• **Computing in the cloud**
  – HPC still cheaper on premises
  – Unique analysis, interactive needs may be best deployed as cloud-hosted resources
  – High-availability capacity for critical use cases

• **Storage in the cloud**
  – Especially disaster recovery copies

The cloud has changed user expectations for research HPC.
A new context for NCAR’s HPC procurement

• Our three original questions still relevant
  – Address the key needs foreseen by users
• As we have progressed through our process, new questions have emerged
  – How should NCAR adjust its investments in storage to support users at new scales of data?
  – How much and for what purposes should NCAR invest in cloud services and integration?
  – How do we meet head-in-the-clouds expectations?
  – How can NCAR support users in adapting their practices, behaviors, and workflows?
• Even as HPC remains the core, the solution shows that HPC is no longer an island
Questions, complaints, criticisms: David Hart, dhart@ucar.edu

For more details on NCAR’s NWSC-3 procurement, see https://www2.cisl.ucar.edu/resources/nwsc-3

Thanks to many persons at CISL, including Rich Loft and Irfan Elahi.