Cray's Storage History and Outlook – Lustre+

Jason Goodman, Cray LUG 2015 - Denver





- Cray History from Supercomputers to Lustre
- Where we are Today
 - Cray Business
 - OpenSFS
- Flashback to the Future SSDs, DVS, and Beyond
- Questions



The future is seldom the same as the past

Seymour Cray June 4, 1995



Cray Timeline – for Lustre, Storage, and Leinies



COMPUTE | STORE | ANALYZE

Focused on Data-Driven Workflows and Industries

Compute



- Supercomputers
- Flexible Clusters
- Hybrid Architectures

Store



- High Performance Storage
- Tiered Storage & Archive

Analyze



- Graph Analytics
- Hadoop Solutions











Cray Investing in Lustre



OpenSFS – Original Founder and Board Member

- Cray, DDN, LLNL, ORNL
- Non-profit technical organization focused on high-end open-source file system technologies

Goals

- Collaboration among entities deploying leading edge HPC file systems
- Driving roadmap for future requirements into OpenSFS
- Supporting Lustre file system releases designed to meet these goals



Cray's Role



LUG 2015

8

OpenSFS – New 2015 Membership Offer

- Cray / OpenSFS offering 5K off first year membership
 - <u>http://opensfs.org/get-involved-with-lustre-for-free/</u>
- Use towards any membership level
 - Promoter, Adopter, or Supporter



Data-Driven Workflows

Managing Data from High Performance Storage to Deep Archives



Case Study: KAUST

• Industry partners

- Massachusetts Institute of Technology (MIT)
- London's Imperial College
- Hong Kong University of Science and Technology
- Woods Hole Oceanographic Institution
- Institut Français du Pétrole
- National University of Singapore
- The American University in Cairo
- Technische Universität München
- King Abdulaziz City for Science and Technology
- King Fahd University of Petroleum and Minerals
- Saudi Aramco
- Schlumberger



Plasma Turbulence

Results from a 4D drift kinetic plasma turbulence code. Simulations used upto 5 million core hours of Shaheen...

Geophysical Modeling A group of Geophysicists gained high resolution from Evanescent wave with Shaheen simulations ...



Simulation Study for (XFEL) The Electrical Engineering team settles down the theoretical grounds for X-Ray Free Electron Lasers (XFEL) ...



12

COMPUTE | STORE

KAUST Requirements

Petascale supercomputing system

- Replacing existing IBM[®] Blue Gene[®] (222 Tflops) with XC40
- "Burst buffer" caching tier

Parallel file system

- 500 GB/sec sustained performance
- Integration with tiered storage and burst buffer solution

• Tiered storage for archives

- 100 PB of capacity with two copies
- Tiered integration with parallel file system



KAUST Cray Solution

- Cray XC40[™]
 - 5.6 Pflops sustained performance with Cray Aries interconnect
 - 792 TB of memory
 - Cray DataWarp[™] with performance exceeding 1 TB/sec
- Cray Sonexion[®] 2000 Storage System
 - 17.2 PB of usable capacity with performance exceeding 500 GB/sec
 - Lustre[®] file system 2.5 with HSM extensions
- Cray Tiered Adaptive Storage (TAS)
 - Cray TAS Connector for Lustre and Versity Storage Manager
 - Spectra Logic T-Finity with 100 PB of capacity with IBM TS1150 drives

COMPUTE

ANALYZE









LUG 2015

STORE



Exascale Computing Memory Trends



COMPUTE | STORE | ANALYZE

Cray DataWarp I/O Accelerator – for Cray XC40

DataWarp overcomes the performance gap between compute and disk storage

- Pure Performance
 - Scale from 70 thousand to 40 million IOPS per system
- Breakthrough Efficiencies
 - 5x the performance of disk at the same cost
- Balanced and Cohesive architecture
 - No application changes needed
 - Quality of Service per-application

| STORE

DATAWARP[®]



DataWarp I/O Blades include SSD cards



ANALYZE

User-directed Checkpoint / Restart

- Explicit use case
- User asks for enough SSD to cover at least 2X memory
- High Bandwidth checkpoints are written to SSD
- Followed by an asynchronous trickle out to rotating storage between checkpoints

COMPUTE



LUG 2015

STORE

Los Alamos National Labs Early Collaboration on DataWarp



"LANL has been investigating burst buffer capabilities for years and the DataWarp technology in the Cray XC40 Trinity system will provide the first multi-petabyte multiterabyte/sec IO burst handling capability ever."

> Gary Grider, High Performance Computing Division Leader

Use case example: checkpoint/restart



COMPUTE

STORE

ANALYZE

Questions?

Resources

- cray.com/storage
- opensfs.org/get-involved-with-lustre-for-free/

• Cray people here at LUG

- Jason Goodman (<u>jasong@cray.com</u>)
- Cory Spitz
- Dave McMillen
- Patrick Farrell
- Chris Horn
- Charlie Carroll

Legal Disclaimer

Information in this document is provided in connection with Cray Inc. products. No license, express or implied, to any intellectual property rights is granted by this document.

Cray Inc. may make changes to specifications and product descriptions at any time, without notice.

All products, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

Cray hardware and software products may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Cray uses codenames internally to identify products that are in development and not yet publically announced for release. Customers and other third parties are not authorized by Cray Inc. to use codenames in advertising, promotion or marketing and any use of Cray Inc. internal codenames is at the sole risk of the user.

Performance tests and ratings are measured using specific systems and/or components and reflect the approximate performance of Cray Inc. products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

The following are trademarks of Cray Inc. and are registered in the United States and other countries: CRAY and design, SONEXION, URIKA and YARCDATA. The following are trademarks of Cray Inc.: ACE, APPRENTICE2, CHAPEL, CLUSTER CONNECT, CRAYPAT, CRAYPORT, ECOPHLEX, LIBSCI, NODEKARE, THREADSTORM. The following system family marks, and trademarks of Cray Inc.: CS, CX, XC, XE, XK, XMT and XT. The registered trademark LINUX is used pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Other names and brands may be claimed as the property of others. Other product and service names mentioned herein are the trademarks of their respective owners.

Copyright 2014 Cray Inc.