Workflow-Driven Storage Solutions for Data Tiering and Archiving

Jason Goodman
PASIG 2015
San Diego
Agenda

● Cray Overview
  ● Storage and Data Management Business
  ● Storage and Data Management Products

● Workflow-Driven Storage
  ● “Big Data” - Lifecycles and Workflows

● Data Tiering and Open Archiving
  ● Cray Tiered Adaptive Storage (TAS) Overview

● KAUST Case Study

● Summary
The future is seldom the same as the past

Seymour Cray
June 4, 1995
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

- Earth Sciences
- Manufacturing
- Energy
- Life Sciences
- Higher Education
- Financial Services
- Government and Defense
Cray Storage and Data Management Business

- **End-to-end solutions for HPC storage and data management**
  - From ingest to deep archives

- **Global organization**
  - 24/7 global support team
  - Single point of support

- **Proven track record and leadership**
  - Over 200PBs of primary disk installed or under contract
  - Nearly an Exabyte of Cray-integrated solutions for archives

Proven Expertise & Best Practices

Best-in-Class Storage Systems

Holistic Storage Architectures
Cray Storage Systems

Scale-Out Lustre® Storage
- Integrated, Modular and Compact
- Simplified Deployment and Management
- Precision Performance

Open Archiving and Tiered Storage
- Complete and Open Archiving
- Fluid Data Management and Migration
- Lustre Integration
Workflow Ready Solutions - Span Entire Stack

- Qualified End-to-End
- Cray System Architectures
- Aries, InfiniBand, Ethernet
- Cray Sonexion
- Cray TAS
- Application I/O Optimization
- File Systems Leadership & Expertise
- Scalable Networking Experts
- Best in Class Storage Systems

Performance Optimization
Data-Driven Workflows

Managing Data from High Performance Storage to Deep Archives
“Big Data” (an overused marketing term)

- **Data Types**
  - Structured & Transactional
    - Databases and Analysis
  - Unstructured (our primary focus)
    - Persistent file data
    - Growing file sizes—and data sets

- **3 V’s: Volume, Variety, Velocity**
  - **Volume**
    - Lot’s of data – and large file systems
  - **Variety**
    - Mainly unstructured
    - Mix of file types and sizes
  - **Velocity**
    - Data is “fast” early in the lifecycle
      - IOPS and throughput
    - Archived data still needs to be accessed, less frequently
Early in the “Big Data” Lifecycle – All Performance

- Throughput and I/O
- Parallel Access
- Performance Scaling

Throughput (MB/sec)

Day 1 | 30 Days | 60 Days | 90 Days | 180 Days | 360 Days | 2 Years

Capacity (PB)

0 | 30
0 | 15
0 | 10
0 | 5
And Later, when Data Infrequently (if ever) Accessed
Common Workflow

Acquire
- Data Collection and Ingest
- High-Speed Data Movers
- Media

Store and Process
- CRAY DataWARP
- CRAY Linux x86 Linuxers

Archive
- Users and Applications

Transparent Data Movement and User Access

Storage and Data Management
Cray Tiered Adaptive Storage

Deployment-ready tiered storage system – engineered on open systems and open formats
Cray Tiered Adaptive Storage (TAS)
Open Archiving and Tiered Data Management

Simplify
Automated and transparent storage tiering

Preserve
Open data management and migration

Reduce
Right data on the right storage technology

Flexible Tiered Storage Management
Simplify with Cray Tiered Adaptive Storage

- **Transparent tiering for users**
  - Data always accessible regardless of tier
  - File system appears infinitely large
  - Files always visible from the file system

- **Automated data management**
  - Policy-based data management
  - 24x7 data management
  - Multiple copies and disaster recovery

- **Integrate with Cray storage**
  - Cray TAS Connector for Lustre HSM
Familiar HSM Actions

- Transparently **Archive** from disk cache to archive media
- Manage disk space or **Release** archived files from disk
- Automatically **Stage** released files back when accessed
Preserve with Cray Tiered Adaptive Storage

● **Open data format**
  - Based on POSIX TAR
  - Data is accessible without TAS
  - No vendor lock-in

● **Data protected at scale**
  - Support for 100’s of PB of managed data
  - Integration with Lustre HSM

● **Technology migration**
  - Support for multigenerational data management
  - Migrate with technology

the default setting is **Open**
Reduce with Cray Tiered Adaptive Storage

- Manage time-value of data
  - Keep only the most active data on the most expensive tier
  - Separate policies for demanding applications or users
  - Right data on right media

- Built with growth in mind
  - TAS scales with your needs without capacity walls

- Cray TAS Connector for Lustre
  - Extend data management to Lustre parallel file systems

- Single point of support
  - World-class Cray support for the entire solution
KAUST Supercomputing Center and Digital Archives

Solution Overview
Supercomputing Research at 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
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
Cray TAS for KAUST

● Cray TAS Connector for Lustre File System
  ● Seamless and scalable tiered storage for Lustre file systems
  ● Policy engine to automatically manage user data 24x7
    ● Eliminates backup windows that become impractical at these scales
  ● Data accessible to users and applications throughout the data lifecycle
    ● Available regardless of storage tier the data resides on

● Cray TAS
  ● Back-end storage management for Lustre file system
  ● Versity Storage Manager (VSM) to manage archived data
  ● Two copies of data for long term preservation on IBM TS1150 media
Questions?

- **Resources**
  - cray.com/storage
  - cray.com/tas

- **Cray people (Crayons) here at PASIG**
  - Jason Goodman (jasong@cray.com)
  - Mike Selway (mselway@cray.com)
  - Scott Pearson (spearson@cray.com)
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.'s 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.