OpenSFS, Lustre, and HSM: an Update for LUG 2014

Cory Spitz and Jason Goodman

Lustre User Group 2014
Miami, FL

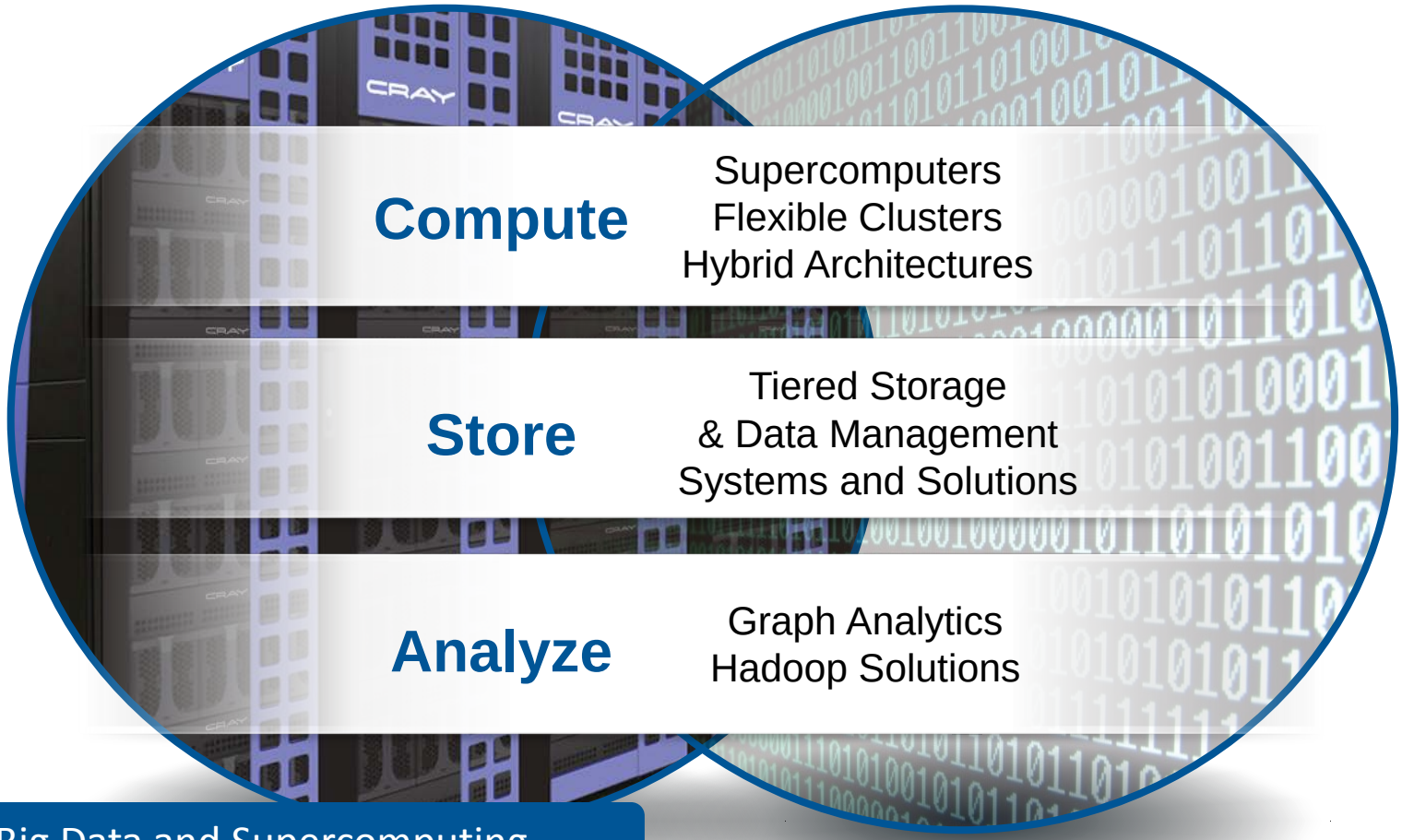
Safe Harbor Statement

This presentation may contain forward-looking statements that are based on our current expectations. Forward looking statements may include statements about our financial guidance and expected operating results, our opportunities and future potential, our product development and new product introduction plans, our ability to expand and penetrate our addressable markets and other statements that are not historical facts. These statements are only predictions and actual results may materially vary from those projected. Please refer to Cray's documents filed with the SEC from time to time concerning factors that could affect the Company and these forward-looking statements.

Agenda

- **Cray Storage and Data Management**
- **Cray and the Community**
  - OpenSFS – our role
  - TWG, CWG, BWG, MWG
  - What we offer the community
- **Lustre – and Cray's role**
- **HSM**
- **Summary**

We Build Computational Tools That Help Change The World



Merging Big Data and Supercomputing

Cray Storage & Data Management - Pillars

Experts in workflow-driven storage, optimized for scale and results

Your Trusted Expert

- Proven experts in parallel systems & storage
- 150 Lustre deployments
- 120 petabytes primary storage installed
- Exascale leadership in storage performance and scalability

Scale Optimally

- Scale-as-you-go performance from GB/s to 1TB/s in a file system
- Fluid capacity scalability from terabytes to exascale-capable archives
- Quality assurance and stress testing for the largest production environments

Results Faster

- Simplify and reduce time to deployment
- Fastest in-production Lustre file system
- Reduced time to results by 24x at NCSA
- Reduce storage footprint by 50% for petascale systems

Massively Scalable Storage Solutions for Big Data & Supercomputing

Cray Customers

PDC Center for
High Performance Computing

COMPUTE | STORE |

What Our Storage Customers are Saying

We immediately saw success from the perspective of stability and performance. Our bandwidth numbers were higher than the previous vendor's, using the exact same hardware. We went from the file system being our biggest issue to the least of our issues, with Cray.

– Jim Lujan, HPC Project Leader, LANL

"Some of the science teams have been able to do 3 years worth of work in 3 months."

– Michelle Butler, Head of Storage & Networking, NCSA Blue Waters project

Cray was chosen at Pawsey because Cray is the most credible and reliable partner and best understood the requirements. Knowing we have Cray onsite is very important. If Cray can't do it, nobody can.

– Dr. George Beckett, Deputy Director & Head of Supercomputing Team

Cray's Storage Portfolio - Overview

CRAY SONEXION®

CRAY® TIERED ADAPTIVE STORAGE

Powered By Versity

Scalable building blocks
- Best-of-breed storage technologies
- Open systems and software

Scale optimally – small to large systems
- Gigabytes to terabytes of performance
- Terabytes to exabytes of capacity

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

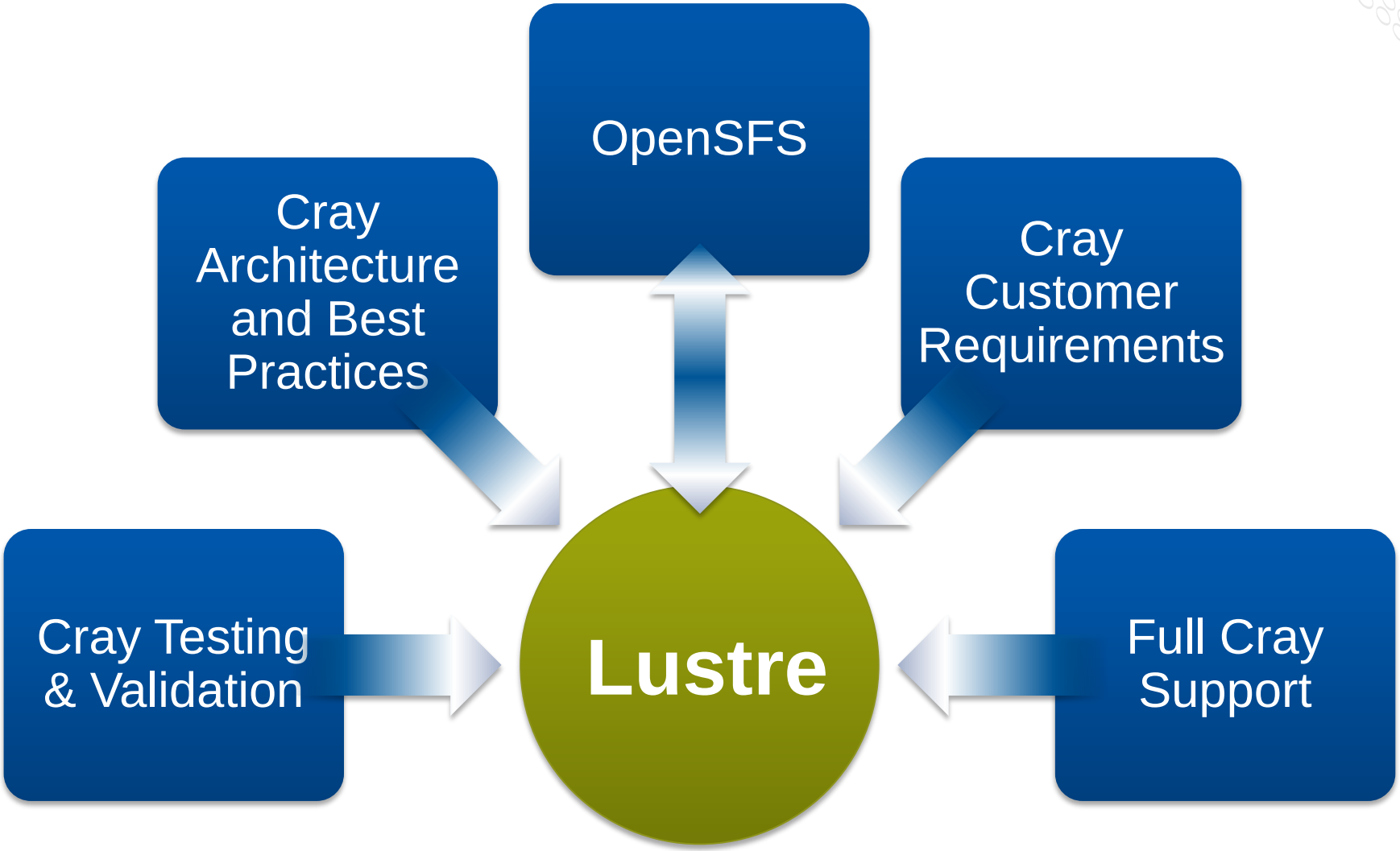
- Collaboration among entities deploying leading edge HPC file systems
- Driving roadmap for future requirements into OpenSFS
- Supporting Lustre file system releases designed to meet these goals

Lustre development process reestablished

OpenSFS partnership created

Multi-stage roadmap in place

Cray's Role


OpenSFS
Cray Architecture and Best Practices
Cray Customer Requirements
Cray Testing & Validation
Lustre
Full Cray Support

Cray's release strategy for Lustre

Three Goals

- Value – build on our OpenSFS investment
- Efficiency - leverage common "Lustre" version across products & releases
- Excellence - maintain performance & Cray-level quality at scale

Tactics and strategy

- Work with community at head of development (master)
- Provide Cray Test feedback of master and release candidates
- Leverage both feature and community maintenance branches

Plan added enhancements independently

- Lustre development is moving rapidly
- Watch for regressions; new features don't destabilize core functionality

How Cray benefits Lustre and the community

- **Testing! – what and how we test**
  - Cray tests all of the stack, save socklnd
  - Scale testing
  - Regression testing
  - Performance testing
  - Failure injection
  - Interop testing (supporting more interop than canonical release scope)
  - Upgrade and migration testing
  - We constantly test master and release branches with automated test suites
- **We get lots of real-world exposure**
  - Cray model: feature releases plus patches or maintenance release plus patches
  - We regularly update our releases and we plan to release each feature release
- **We give back, tracking bugs and patches**
  - We ensure that we carry minimal amount of patches
  - Our process: we don’t close tickets until fix is landed to master
- **Support**
  - Ensure customers have path forward to new versions of Lustre

Addressing Lustre quality

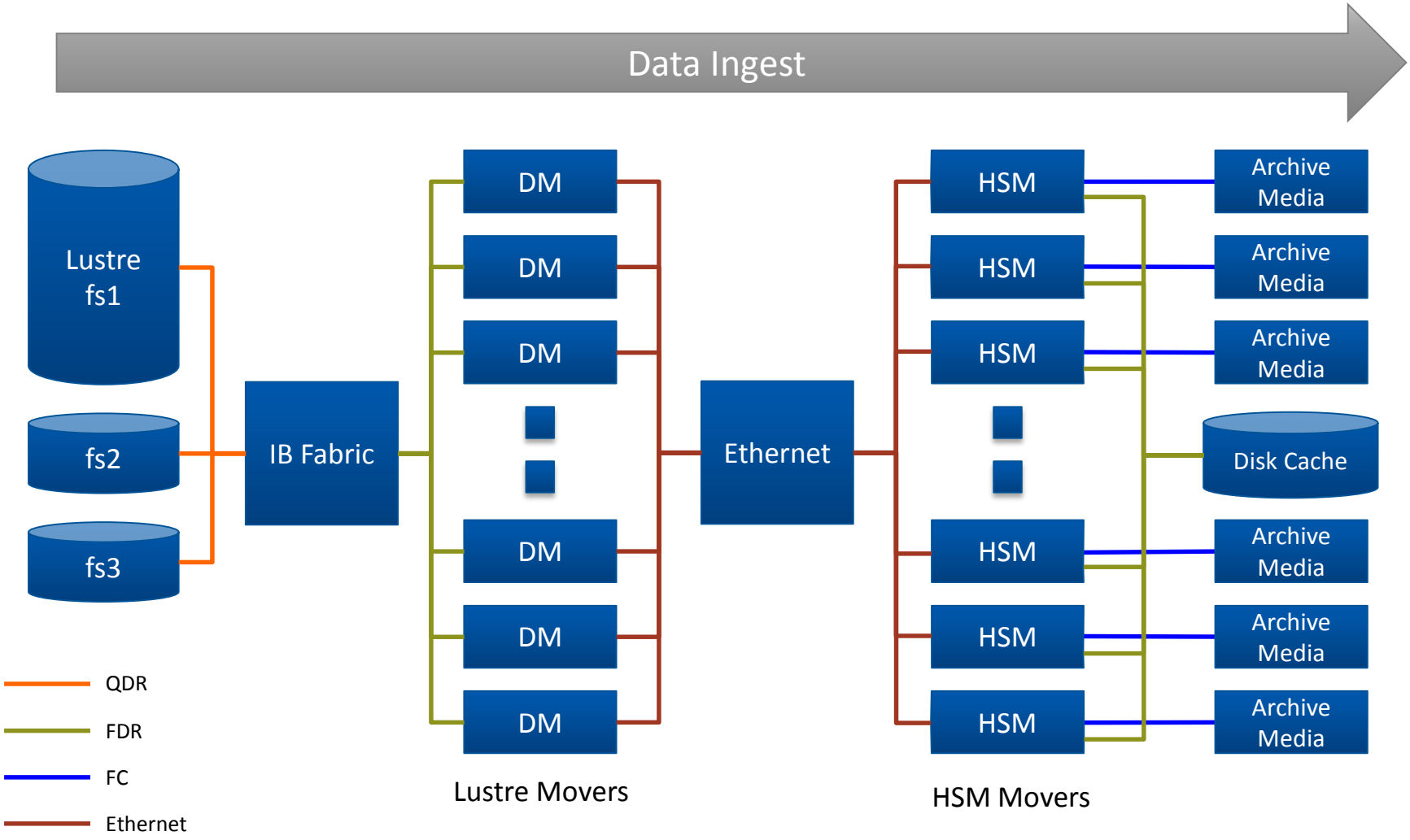
- Collaboration essential
- Goal: improve both feature testing and release testing
- Test improvements, methodologies, and tools
- Address technical debt
- Address design complexity
- Internals documentation
- Resources at scale
- Work with the TWG & CDWG!

Examples of Work & Focus Areas

- **LNET**
  - gnilnd
  - RAS & re-routing
- **RAS**
- **APIs and Development**
  - Engaging Lustre community for Open Fabrics Alliance
  - MPI-I/O
- **Scaling**
  - DNE scale testing
  - Pingless clients with imperative recovery and client eviction
- **Testing**
- **HSM deployment**

Lustre HSM – Cray's Approach

Traditional HSM Implementation – Complex



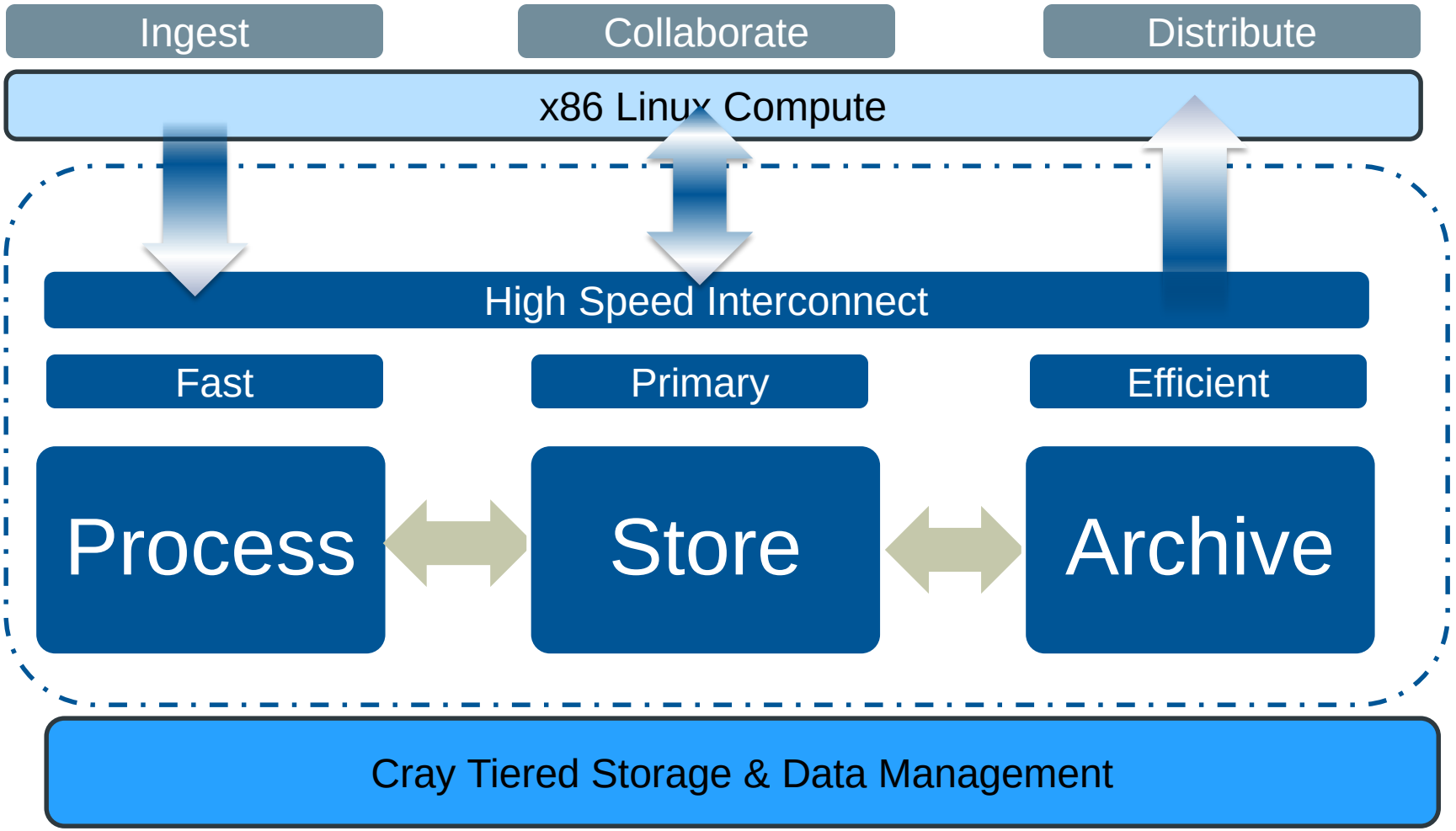
Cray Goals for HSM and Archiving

Data Management and Access across Storage Tiers

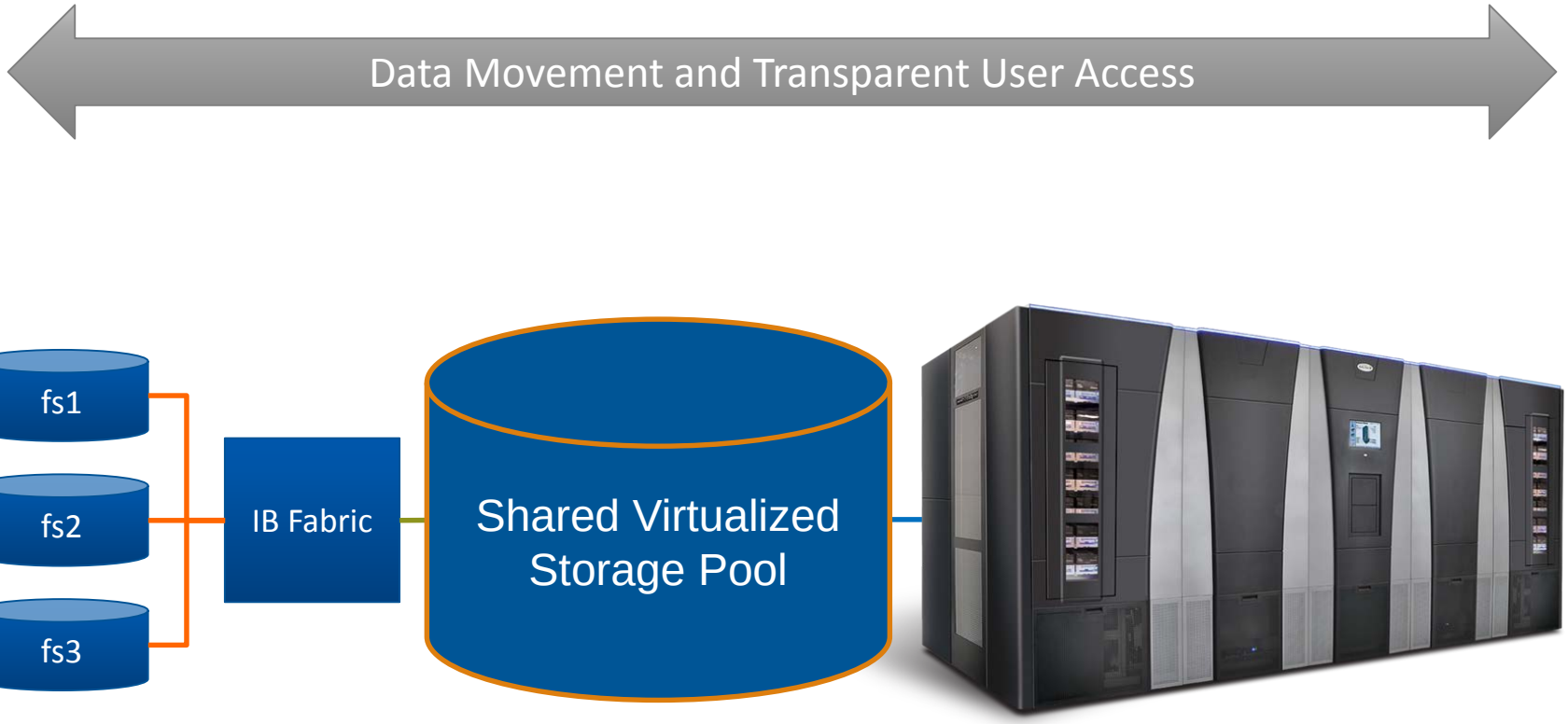
- **Simplicity**
  - Use familiar, policy-based data management best practices
  - System management – planning, deploying, operating, and modifying the system should be easy
  - Lifecycle management of all storage hardware and software
  - In place data migration through open format technologies / standards
- **Fluid expandability and scalability**
  - Performance scalability using best-of-breed SSD and SAS
  - Capacity expansion should be media agnostic and exascale-capable
- **Open, vendor-independent architecture**
  - Open format Hierarchical Storage Management (HSM)
  - Open source Linux OS and tools
  - Flexibility in choice of media technologies – i.e., best of breed storage
- **Data continuously accessible and protected**
  - Driven by available requirements of data set and users
- **Quality and dependability at scale**
  - Solutions should work as advertised
  - Single point of support for entire solution, if possible

Sample HSM Workflow

Managing Lustre data across tiers

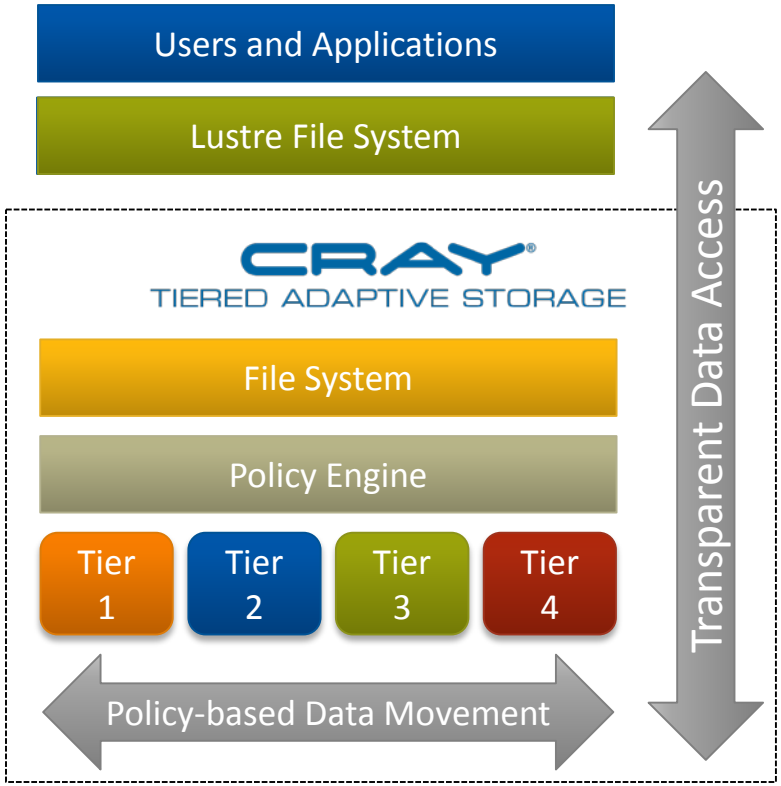


Cray TAS – Simplifying HSM



Cray Tiered Adaptive Storage for Big Data

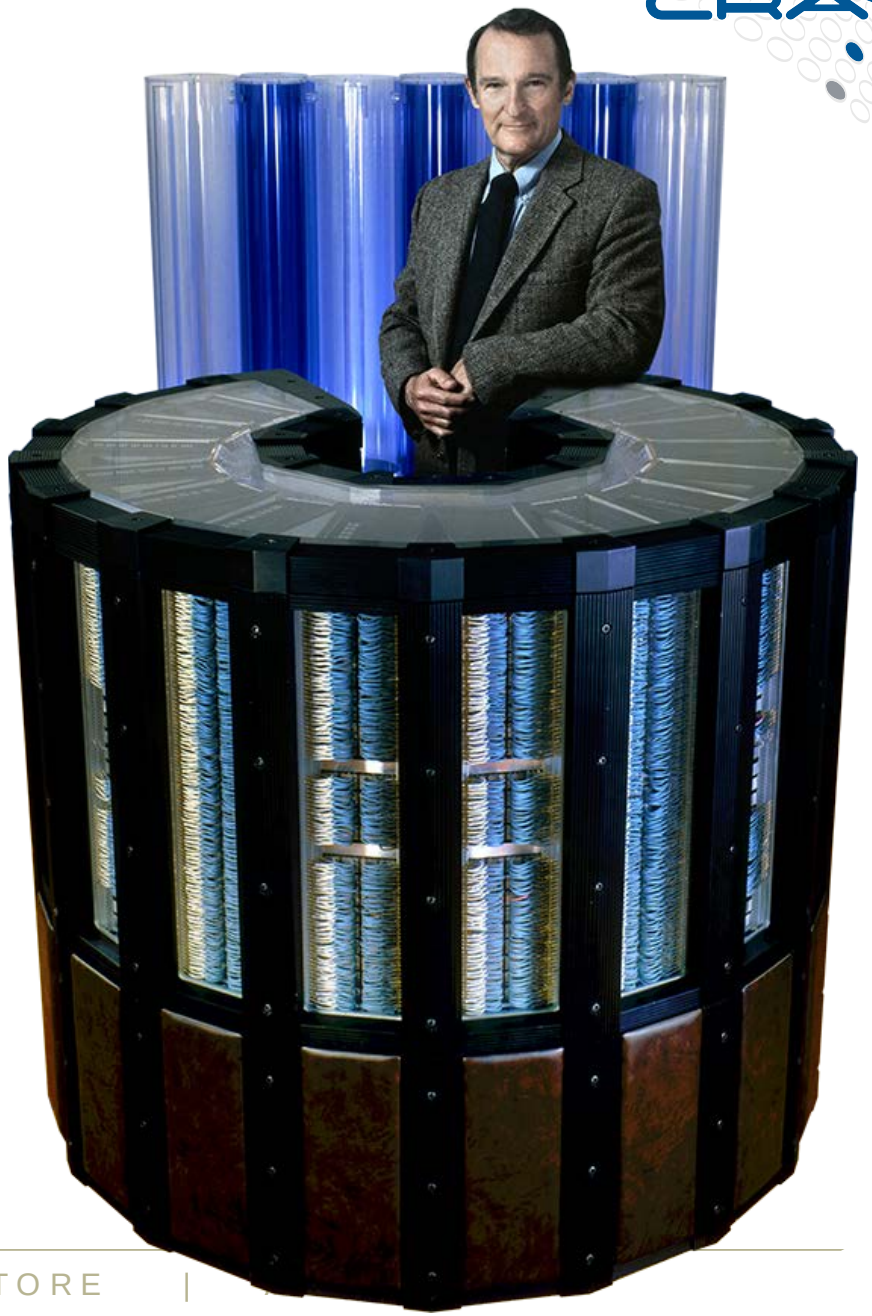
- **Virtualize storage**
  - Single interface to multiple tiers
  - File systems appear infinitely large
  - No user interaction required
- **Protect data at scale**
  - Multiple copies of files
  - Disaster recovery capabilities
- **Flexible storage tiers**
  - Scale the correct tiers to your needs
  - Support for both disk and tape
- **Transparent for users and apps**
  - Maintain ease of use for your customers
- **Extensible to Lustre file system**
  - Lustre file system integration
  - Maintain transparency throughout



Summary and Call to Action

- **Storage Leadership**
  - Founding member and current board member of OpenSFS
  - High performance storage solutions at all scales
  - Exascale vision
- **Testing at Scale**
- **Joint Collaborations**
  - NCSA, ORNL, et al
- **Let's Talk!**

The future is seldom the same as the past



Seymour Cray
June 4, 1995

CRAY®

THE SUPERCOMPUTER COMPANY

COMPUTE | STORE | ANALYZE

Cray Tiered Adaptive Storage

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 associated model number marks, are 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 trademarks used in this document are the property of their respective owners.