August 12, 2015

Amanda Katona,
Customer Advocacy

Purdue University Case Study

How Flash-based Collaboration Drives Faster Time to
Insight, Discovery & Results

Why Storage Matters

"Data underpins almost all research today, so having a scalable storage platform that can meet the varying demands of many diverse communities is vital to any modern university."

Mike Shuey
Research Infrastructure Architect

Topics

- ► **About Purdue University**
  - Research areas
  - Requirements
  - Challenges
- ► **About the Solution**
  - Environment
  - SFX
  - Results
- ► **About DDN**

About Purdue University

► Located: West Lafayette, Indiana
► #22 of US Research Universities
  • 400+ Laboratories
  • 140 Research Centers & Institutes
► #1 U.S. Campus Supercomputing Computational Capacity
  • Conte #73
  • Rice #164
  • Carter #426

About Purdue's Research Areas

Through computer simulation, theories can be explored and experiments can be conducted that were unthinkable just a few years ago.
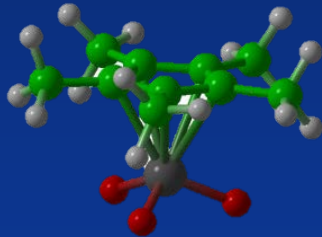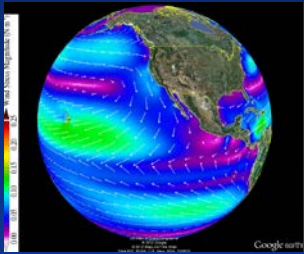
High performance computing (HPC) is the engine that realizes such simulations:

▶ Climate change predictions

▶ Molecular simulations for drug design

▶ New theories of particle physics

Purdue's Requirements

1,000+ Researchers across 300+ projects are using HPC to store and analyze massive amounts of data

Machines get old, storage fills up, and it's a challenge that has to be addressed constantly

Purdue's infrastructure MUST . . .
- Support tons of concurrent users across 100's of concurrent projects
- Provide uninterrupted access to centralized compute, network and storage resources
- Be maintained and upgraded regularly

Purdue's Requirements
Enabling Collaborative Research

► Massively parallel data access

► Millions of independent file requests

► All without imposing undue performance penalties on any research group
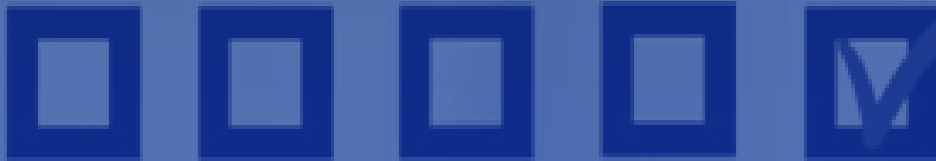
Purdue's Storage Purchasing Challenges

► Determining how much storage is needed

► Gathering diverse requirements of several of the top research areas

- Computational Nanotechnologies
- Aeronautical & Astronomical Engineering
- Mechanical Engineering
- Genomics
- Structural Biology

One of the largest areas of data growth is . . .

▶ Liberal Arts, Sociology, Ethnographic Research
- Across multiple continents
- Lots of large video files
- Tons of data collected on culture & behavior

Purdue University's Storage Upgrade
What they were looking for

▶ A single, site-wide file system

▶ Accessible from multiple internal HPC systems

▶ Fast, unrestricted access to data

▶ Highly scalable to support fast growing data volumes

- Performance
- Scalability
- Reliability
- Price/performance
- Ability to handle the diverse workloads
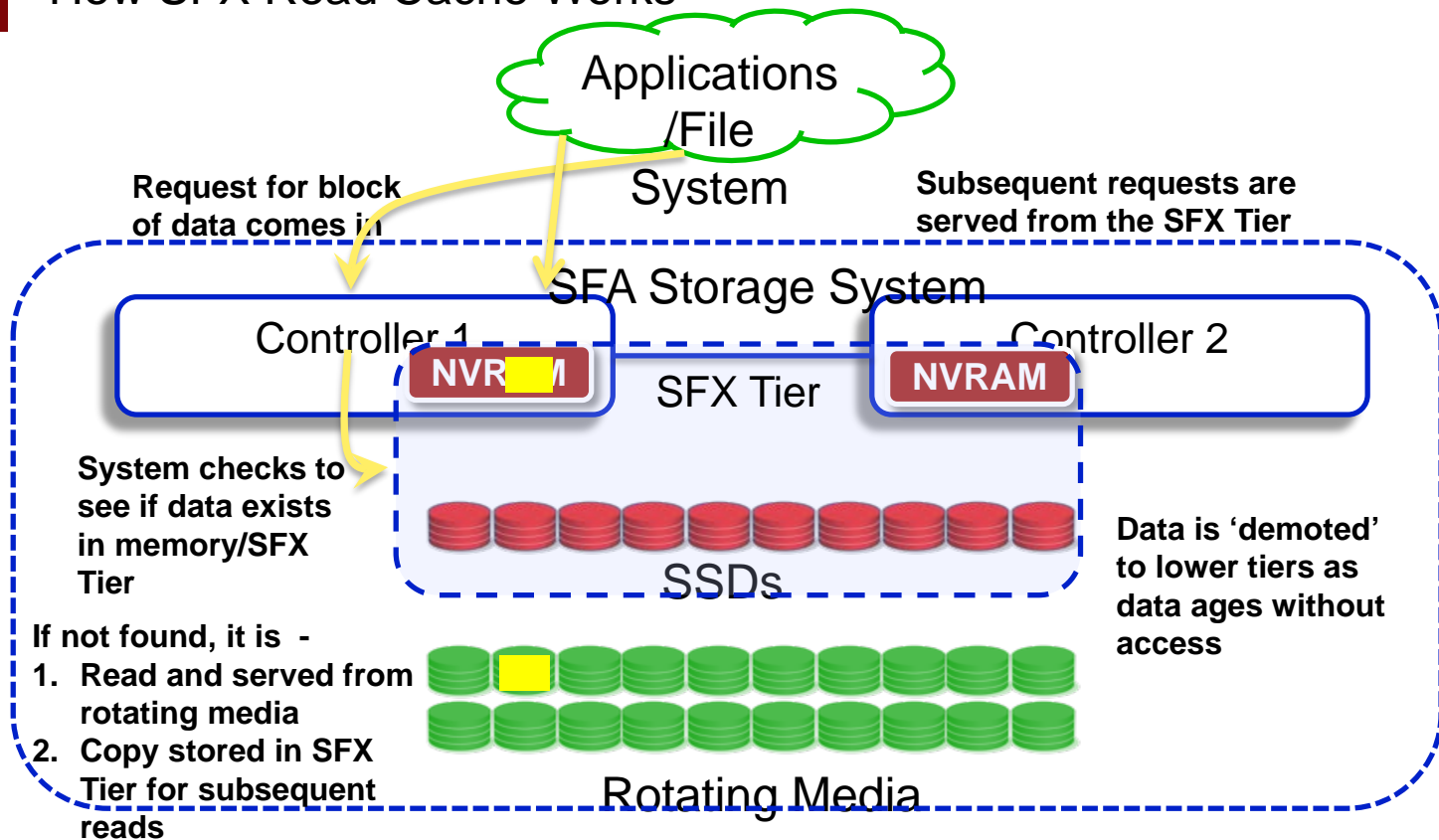
Leveraging one of the nation's largest campus-wide cyber infrastructures for research

► *Data Depot* – a fast, reliable, high-capacity data storage service for researchers across all fields of study

- Facilitates file sharing across research groups
- Transfers of large amounts of data between local systems, national labs, other 3rd parties

Solution- Introducing: *SFX*

How SFX Read Cache Works

Applications /File System

Request for block of data comes in

Subsequent requests are served from the SFX Tier

SFA Storage System

Controller 1

Controller 2

NVRAM

NVRAM

SFX Tier

System checks to see if data exists in memory/SFX Tier

SSDs

Data is 'demoted' to lower tiers as data ages without access

If not found, it is -
1. **Read and served from rotating media**
2. **Copy stored in SFX Tier for subsequent reads**

Rotating Media

ddn.com

DDN STORAGE

BUSINESS APPLICATION
Flash Memory Summit
2014 BEST OF SHOW

End-to-End Solution Delivered by DDN
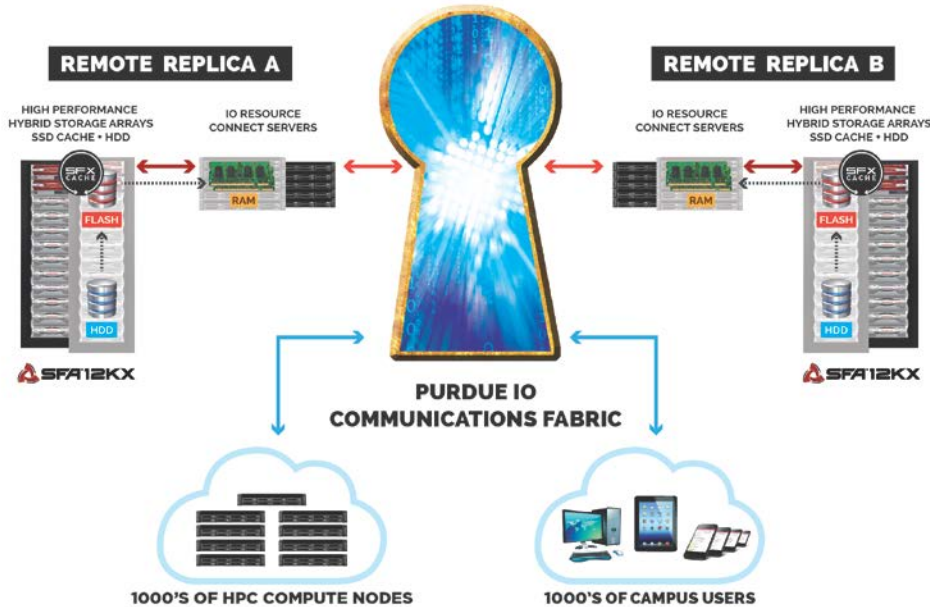
- ▶ 6.4 petabytes of block storage capacity for the parallel file system
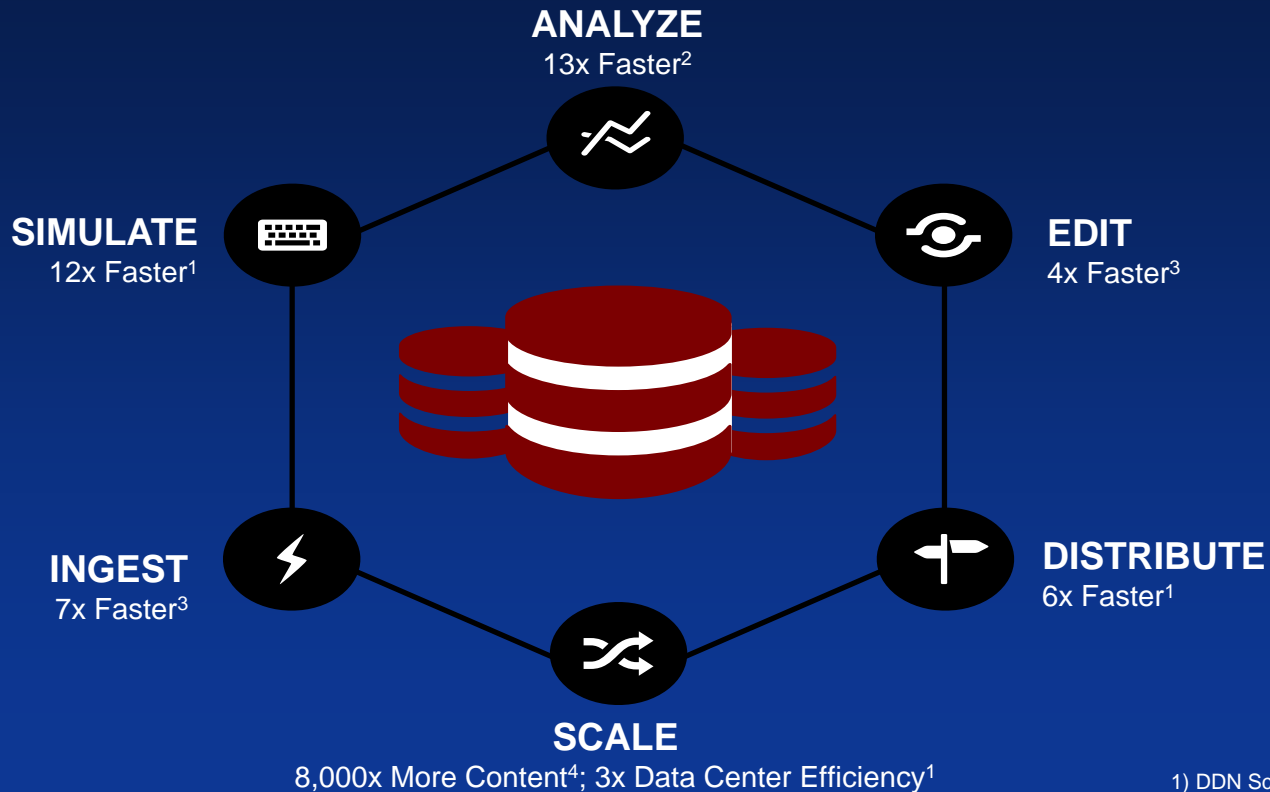- ▶ Extremely cost-effective
- ▶ Included SFX® data acceleration technology
- ▶ A single system for high performance research and *Data Depot*
  - Recently accessed data can be cached on a large pool of solid-state storage, built into the same array as our primary data storage

Results Delivered

▶ What Took 2 minutes . . .
Now Takes Just 2 seconds!
- Projects can access millions of small files held on dedicated solid-state modules while continuing to stream very large data files at the same time

▶ 10X Improvement in reads
- 9/10 random reads are served from SFX cache!

▶ Faster & Lower Cost
- Than buying an all flash array

DDN Technology to Accelerate & Scale

ANALYZE
13x Faster[2]

SIMULATE
12x Faster[1]

EDIT
4x Faster[3]

INGEST
7x Faster[3]

DISTRIBUTE
6x Faster[1]

SCALE
8,000x More Content[4]; 3x Data Center Efficiency[1]

1) DDN Scalers vs. EMC Isilon
2) SFA12K vs. EMC VNX (SAP HANA)
3) SFA12K vs. NTAP e5500
4) WOS vs. Quantum Stornext

DDN Technology Behind The World's Leading Data-Driven Organizations

HPC & Enterprise Big Data Analysis
- Wellcome Trust Sanger Institute
- AIRBUS
- SUZUKI
- TOTAL
- ceol
- SAS
- STANFORD UNIVERSITY
- OAK RIDGE National Laboratory

Cloud, Web, & Telco Infrastructure
- Level (3)
- bezeq
- PayPal
- rgb NETWORKS
- FAMILYSEARCH WHERE GENERATIONS MEET
- netstream That's IT.
- gettyimages
- 115.com

Media & Entertainment
- UNIVERSAL
- CBS SPORTS NETWORK
- WB
- SONY
- SPEED CHANNEL
- MEDIASET
- starz
- deluxe

Government, Academic Research, Surveillance
- Public Transport Authority
- REDWIND CASINO
- Raytheon
- HARRIS
- NORTHROP GRUMMAN
- Maricopa County
- BOEING

DDN I/O & Performance Leadership

1998

DDN Founded

2000-2005

1st Customer (NASA)

First Real-Time Appliance for High-Scale Big Data

DDN's First Parallel File System Offering ft. Lustre®

Achieved 100+GB/s File System Target

EXAScaler™

2006-2010

First in Data Center Density

First in Bandwidth + IOPS

First In-Storage Processing™

Largest Private Storage Company

First Hyperscale Object Storage

STORAGE FUSION ARCHITECTURE

WOS®

2011-2014

Achieved 1TB/s File System Target

First Application-Aware Hybrid Caching Revolutionizing Big Data

Exceeded 5PB/rack

SFX™

IME™

2015

Complete End-to-End Big Data Lifecycle Integrated Solutions

Thank You!

Keep in touch with us

sales@ddn.com

@ddn_limitless

company/datadirect-networks

2929 Patrick Henry Drive
Santa Clara, CA 95054

1.800.837.2298
1.818.700.4000

DDN STORAGE

ddn.com