IBM Spectrum Scale Update

UK User Group Meeting 2017
Manchester – May 9th+10th, 2017

Ulf Troppens

The world is changing …

IBM Spectrum Scale

2005

The world is changing …

IBM Spectrum Scale

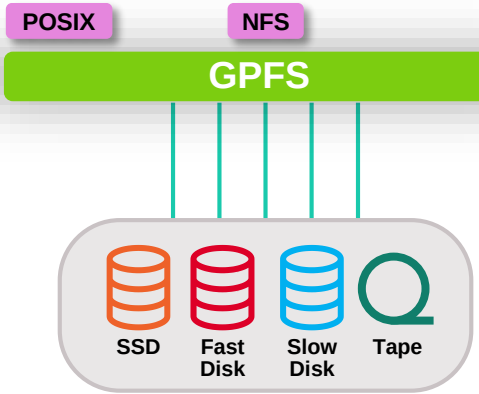
2005

2013

GPFS is changing …

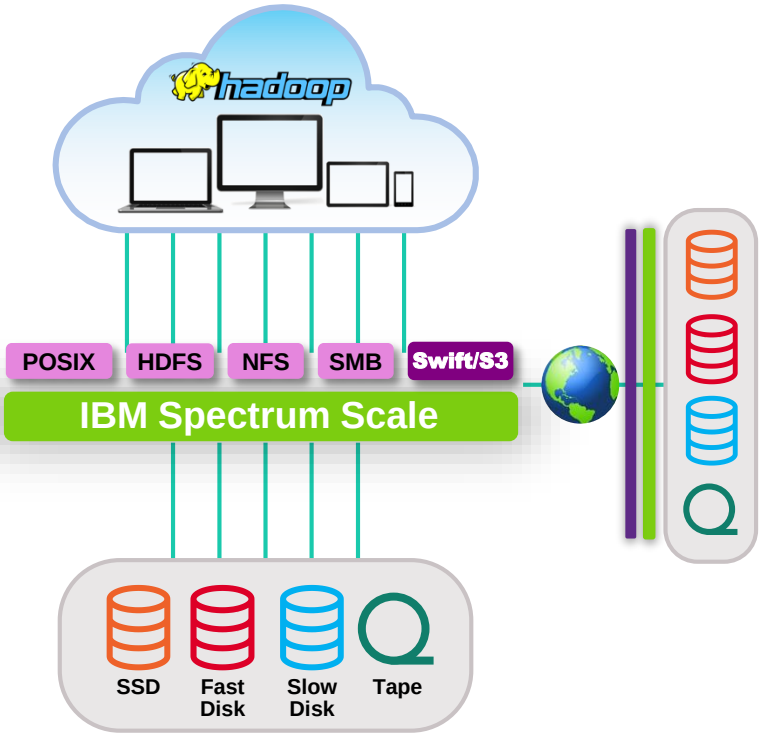
IBM Spectrum Scale

- 1993: Started as “Tiger Shark” research project at IBM Research Almaden as high performance filesystem for accessing and processing multimedia data
- Next 20 years: Grew up as General Parallel File System (GPFS) to power the world’s largest supercomputers
- Since 2014: Transforming to IBM Spectrum Scale to support new workloads which need to process huge amounts of unstructured data



IBM Spectrum Scale

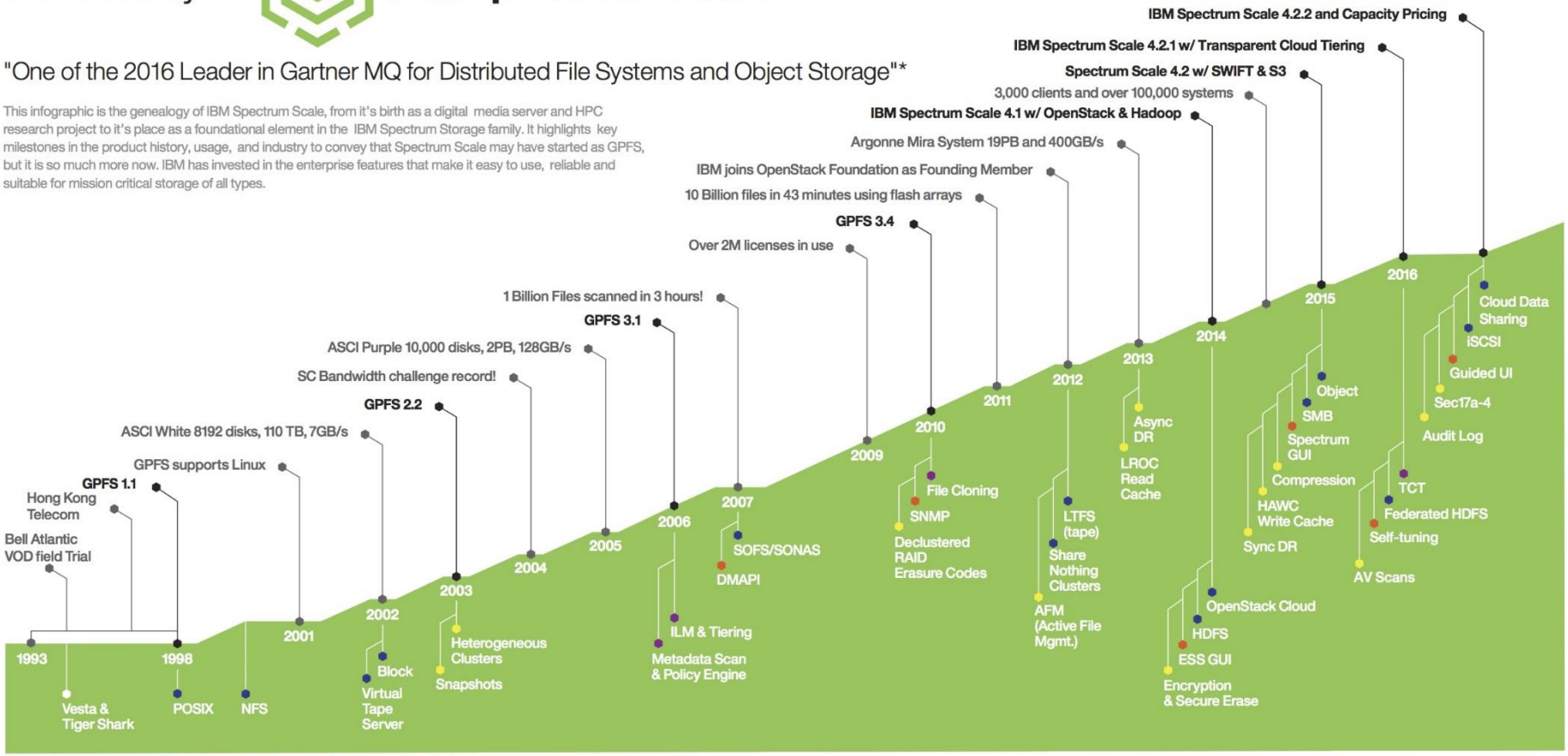
- Based on GPFS, a robust, fast and mature parallel file system
- BUT: If you still just think GPFS, you miss:
  - Support for workflows which for example inject data via object, analyze results via Hadoop/Spark and view results via POSIX
  - Storing and accessing large and small objects (S3 and Swift) with low latency
  - Automatic destaging of cold data to on premise or off premise object storage
  - Exchange of data between Spectrum Scale clusters via object storage in the cloud
  - Storing and starting OpenStack VMs without copying them from object storage to local file system
  - GUI, Grafana Bridge, REST API
  - iSCSI boot
  - And many, many more



The History of IBM Spectrum Scale

"One of the 2016 Leader in Gartner MQ for Distributed File Systems and Object Storage"*

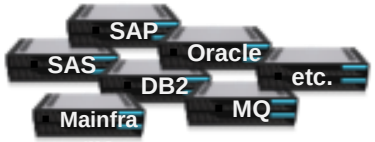
This infographic is the genealogy of IBM Spectrum Scale, from it's birth as a digital media server and HPC research project to it's place as a foundational element in the IBM Spectrum Storage family. It highlights key milestones in the product history, usage, and industry to convey that Spectrum Scale may have started as GPFS, but it is so much more now. IBM has invested in the enterprise features that make it easy to use, reliable and suitable for mission critical storage of all types.

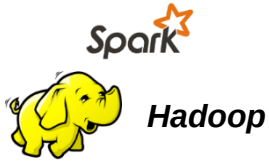


* Gartner, Magic Quadrant for Distributed File Systems and Object Storage, 20 October 2016, Document No. G00307798

Unleash New Storage Economics on a Global Scale

Client workstations





Users and applications

Single name space

POSIX
NFS
SMB/CIFS
HDFS Connector
Object
S3
Swift
OpenStack
Cinder
Glance
Manila
Docker
kubernetes

IBM Spectrum Scale

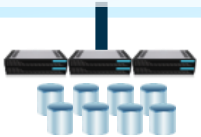
Automated data placement and data migration

Active = Hot+Warm = "Online"

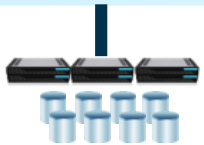
Archive = Cold = "Offline"

Flash

Disk



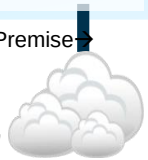
Storage Rich Servers



Object Storage

Tape

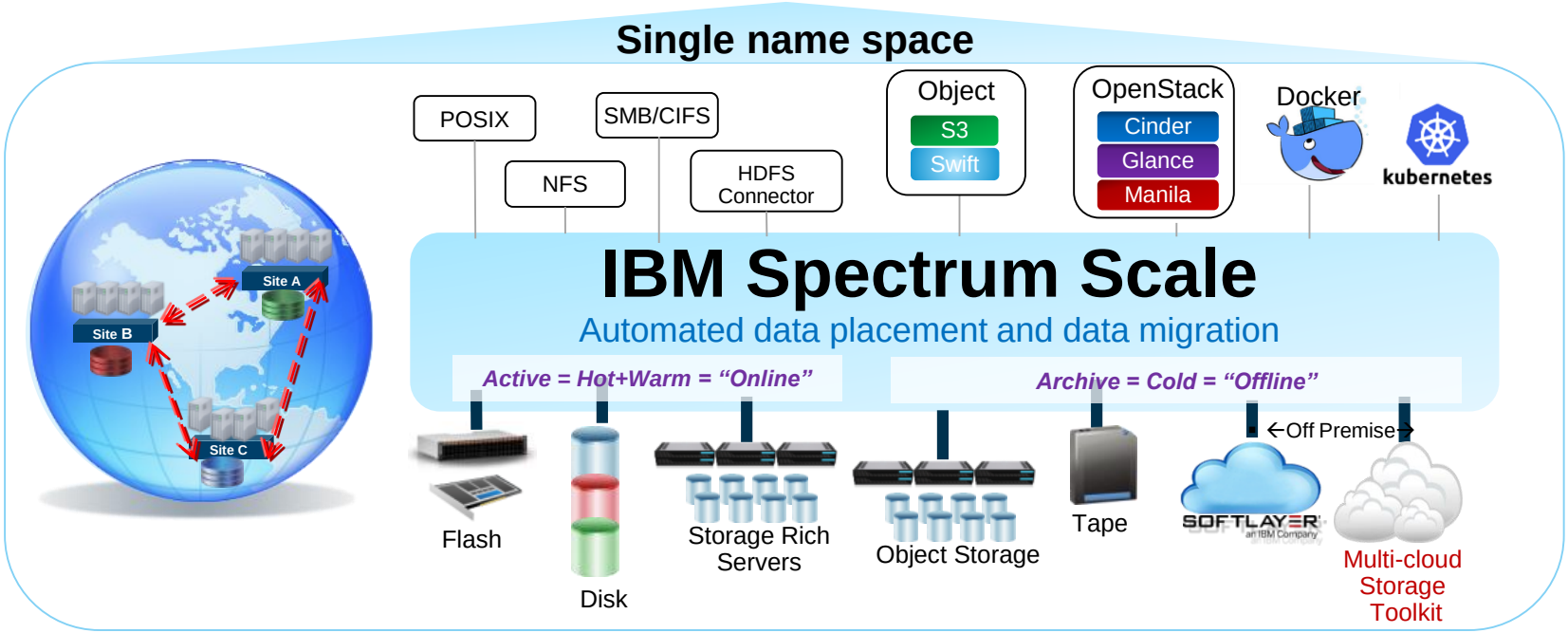
←Off Premise→



Multi-cloud Storage Toolkit

1) Understand Use Case

Core HPC | **File-based Workflows** | **Application Acceleration** | **Virtual Infrastructure**



2) Define the Spectrum Scale Configuration

Core HPC | **File-based Workflows** | **Application Acceleration** | **Virtual Infrastructure**

IBM Spectrum Scale

File Services / File Storage Virtualization / Application Integration
Software-defined Storage Freedom

Active = Hot+Warm = "Online"

Archive = Cold = "Offline"

Flash

Disk

Storage Rich Servers

Object Storage

Tape

←Off Premise→

SOFTLAYER an IBM Company

Multi-cloud Storage Toolkit

3) Select Best of Breed Storage

Core HPC | **File-based Workflows** | **Application Acceleration** | **Virtual Infrastructure**

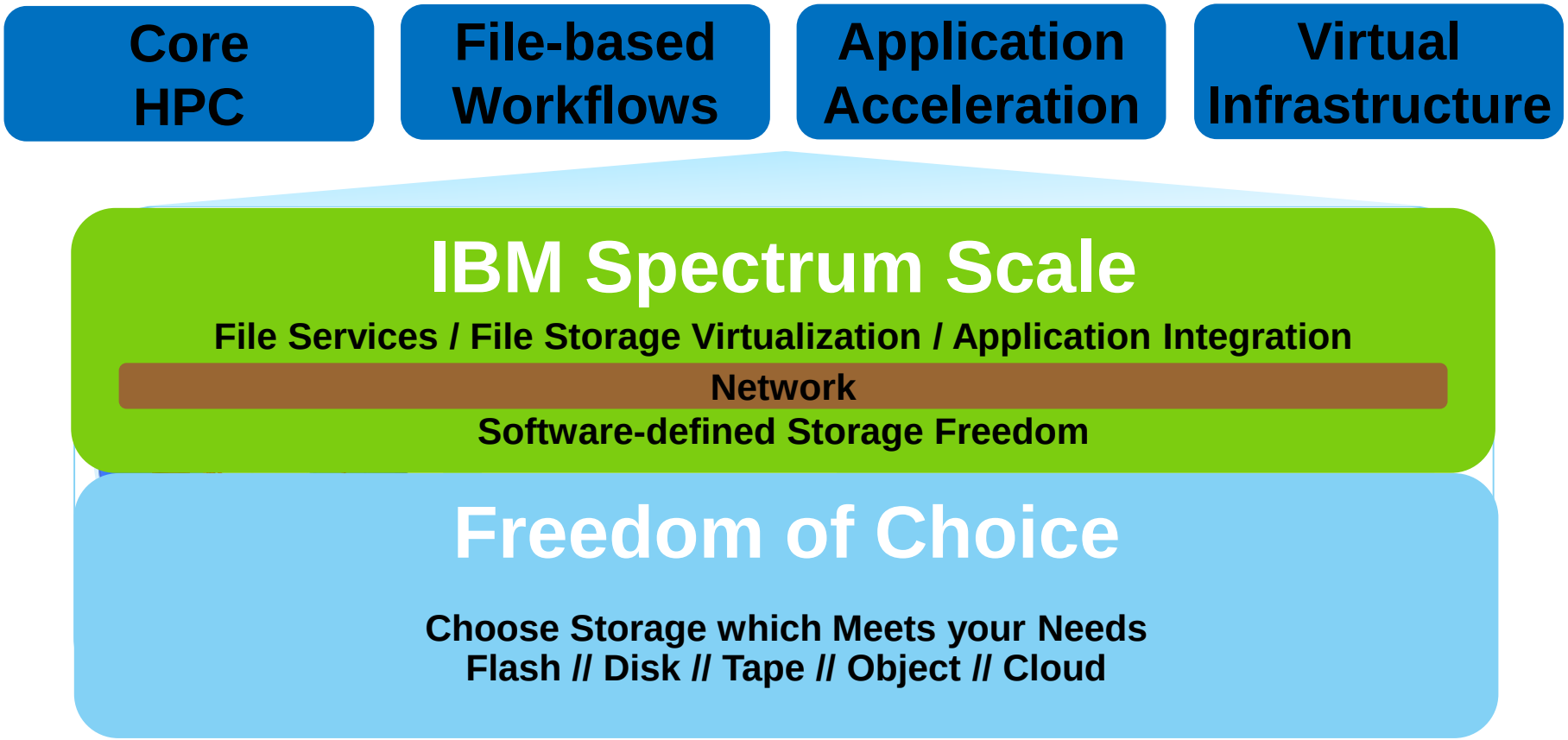
IBM Spectrum Scale

File Services / File Storage Virtualization / Application Integration
Software-defined Storage Freedom

Freedom of Choice

Choose Storage which Meets your Needs
Flash // Disk // Tape // Object // Cloud

4) Network is Integral Part of End-to-end Solution



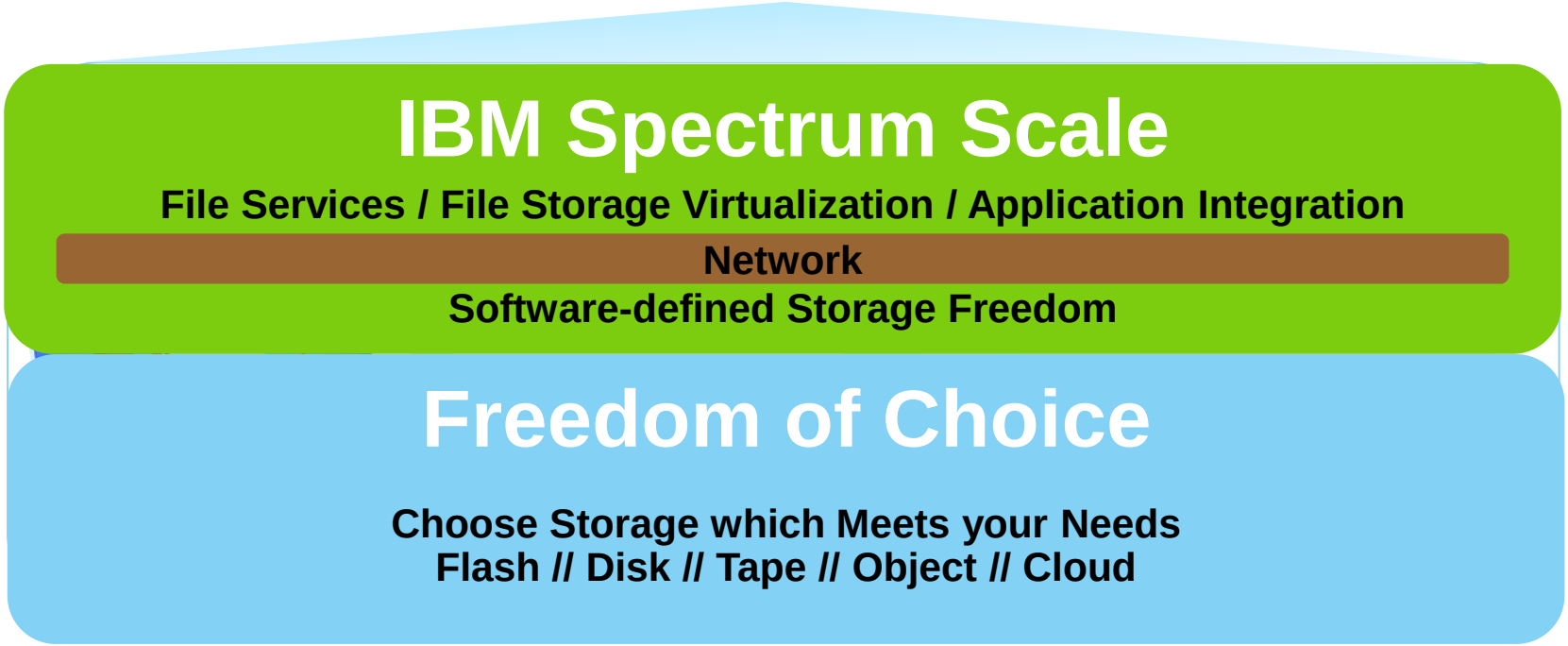
Examples – Core HPC

Core HPC

File-based Workflows

Application Acceleration

Virtual Infrastructure



What is CORAL

Collaboration of DOE Oak Ridge, Argonne, and Lawrence Livermore National Labs

- Established in early 2014 to leverage supercomputing investments, streamline procurement processes and reduce costs to develop supercomputers
  - "High-performance computing is an essential component of the science and technology portfolio required to maintain U.S. competitiveness and ensure our economic and national security"
    - U.S. Secretary of Energy Ernest Moniz
- Two new High Performance Computing (HPC) awards announced in November 2014
  - Both CORAL awards leverage the IBM Power Architecture, NVIDIA's Volta GPU and Mellanox's Interconnected technologies to advance key research initiatives for national nuclear deterrence, technology advancement and scientific discovery
  - Oak Ridge National Laboratory's (ORNL's) new system, Summit, is expected to provide at least five times the performance of ORNL's current leadership system, Titan
  - Lawrence Livermore National Laboratory's (LLNL's) new supercomputer, Sierra, is expected to be at least seven times more powerful than LLNL's current machine, Sequoia.

Source: http://energy.gov/articles/department-energy-awards-425-million-next-generation-supercomputing-technologies

CORAL Systems

- **LLNL's Sierra system**
  - ~4000 Power9 nodes with GPU acceleration
  - ~2.3 PB system memory (include DDR & HBM; does not include NVMe)
  - Dual-rail InfiniBand EDR fat tree network or better
  - ~120 PF
  - ~9 MW
- **ORNL's Summit system**
  - ~4500 Power9 nodes with GPU acceleration
  - ~ 2.7 PB system memory (include DDR & HBM; does not include NVMe)
  - Dual-rail InfiniBand EDR fat tree network or better
  - ~200 PF peak
  - ~13 MW

http://files.gpfsug.org/presentations/2016/SC16/11_Sarp_Oral_Gautam_Shah_Spectrum_Scale_Enhancements_for_CORAL_v2.pdf

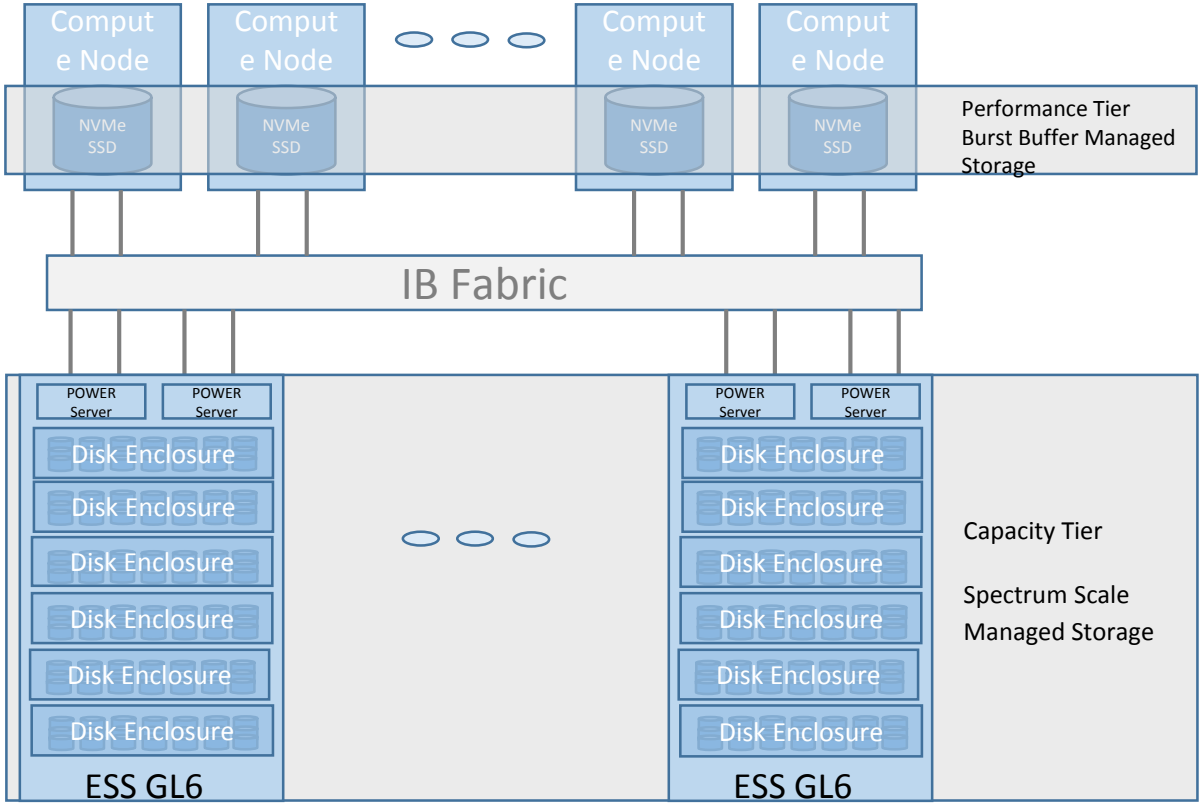
CORAL System Storage Overview

Storage architecture

- Need for a burst buffer/performance tier
  - Lowers traditional spinning disk & lower power consumption
  - Node local NVMe SSD managed by Burst Buffer Software
- Capacity requirement – ESS Storage
- Performance/Scaling requirements – Spectrum Scale Software

http://files.gpfsug.org/presentations/2016/SC16/11_Sarp_Oral_Gautam_Shah_Spectrum_Scale_Enhancements_for_CORAL_v2.pdf

CORAL Storage Overview



http://files.gpfsug.org/presentations/2016/SC16/11_Sarp_Oral_Gautam_Shah_Spectrum_Scale_Enhancements_for_CORAL_v2.pdf

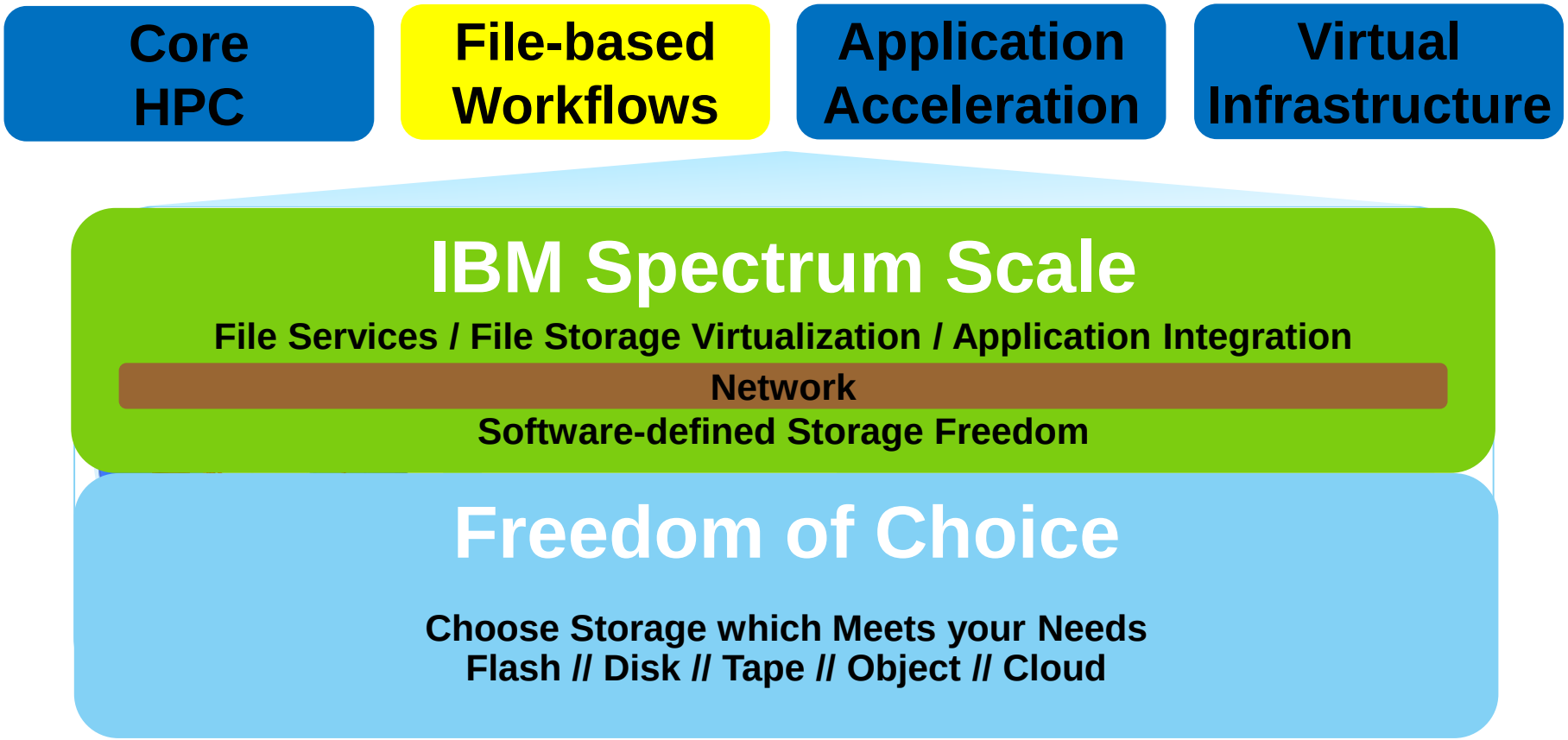
Spider 3 @ OLCF

Spider 3 is a center-wide single namespace POSIX file system to serve all OLCF resources eliminating data islands, and enabling seamless data sharing between resources

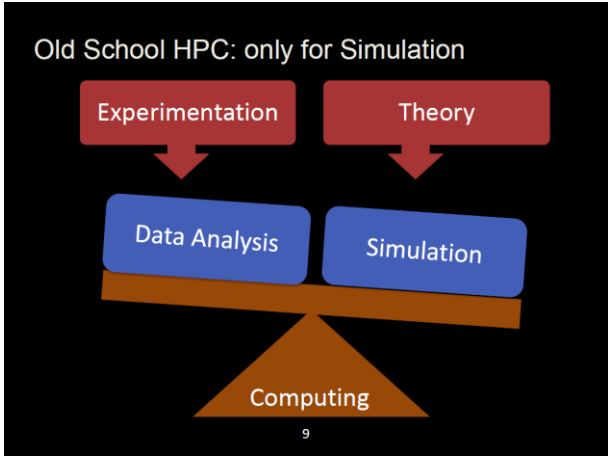
- Built on IBM's Elastic Storage Server based on Power 9 Processor and uses Spectrum Scale (formerly known as GPFS) parallel filesystem technology utilizing GPFS Native RAID with 8+2 redundancy
- Provides a usable capacity of 250 PB
- Performs at an aggregate sequential peak read/write bandwidth of 2.5 TB/s
- Performs at an aggregate random peak read/write bandwidth of 2.2 TB/s
- Provides rich metadata performance; single directory parallel create rate of 50,000/s
- Provides rich interactive performance; @32 KiB I/O 2.6 million IOPs
- Disk-based, with tens of thousands of disks
- Connected to OLCF's SION 3 SAN with IB EDR
- Will also serve as the Summit Burst Buffer sink and source on the end-to-end I/O path

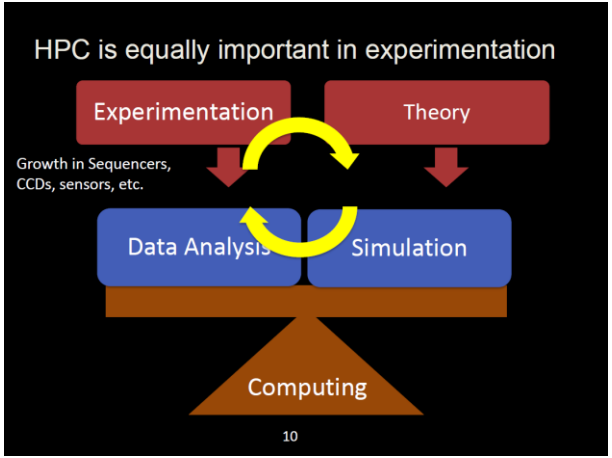
➔ See here for supporting Spectrum Scale Enhancements:
http://files.gpfsug.org/presentations/2016/SC16/11_Sarp_Oral_Gautam_Shah_Spectrum_Scale_Enhancements_for_CORAL_v2.pdf

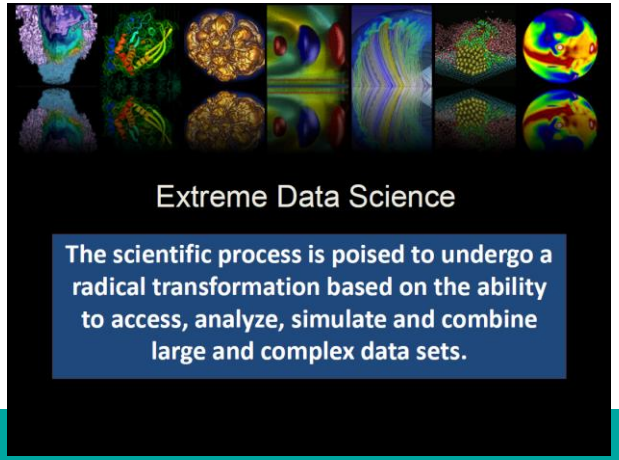
Examples – File-based Workflows



Changes in Science





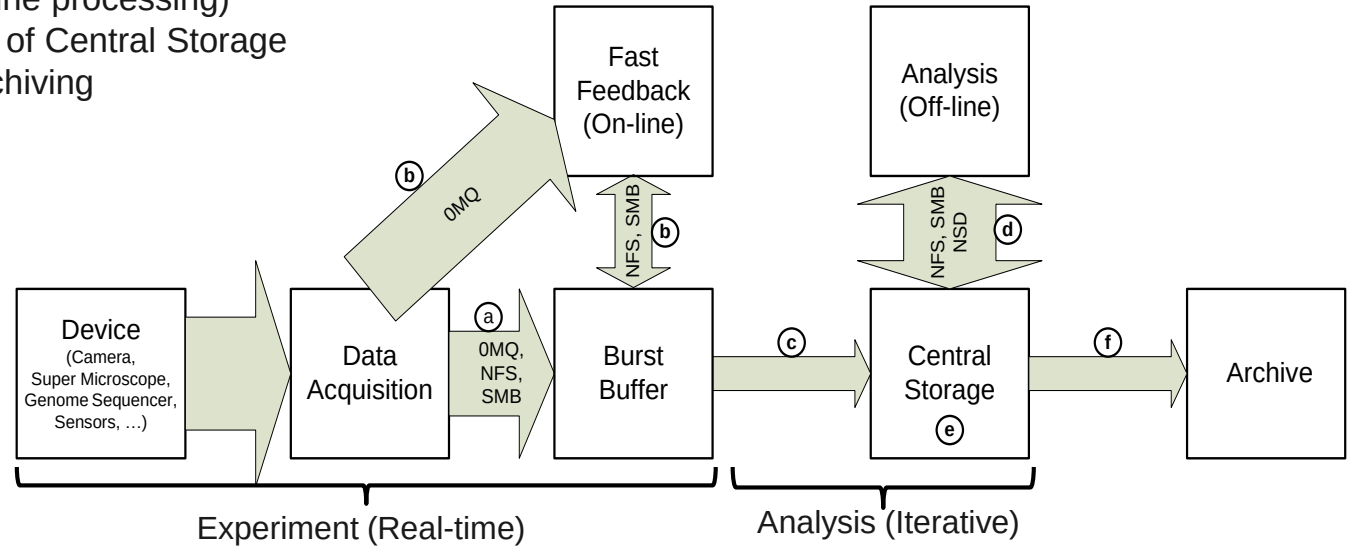


- Data driven science is getting the norm
- HPC gets integrated into the experiments to analyze huge amounts of measured data
- This shift is seen in scientific research and in industry (e.g. life-science, automotive)

https://science.energy.gov/~/media/ascr/ascac/pdf/meetings/201609/Yelick_Superfacility-ASCAC_2016.pdf

Typical Workflow for Data Intensive Science

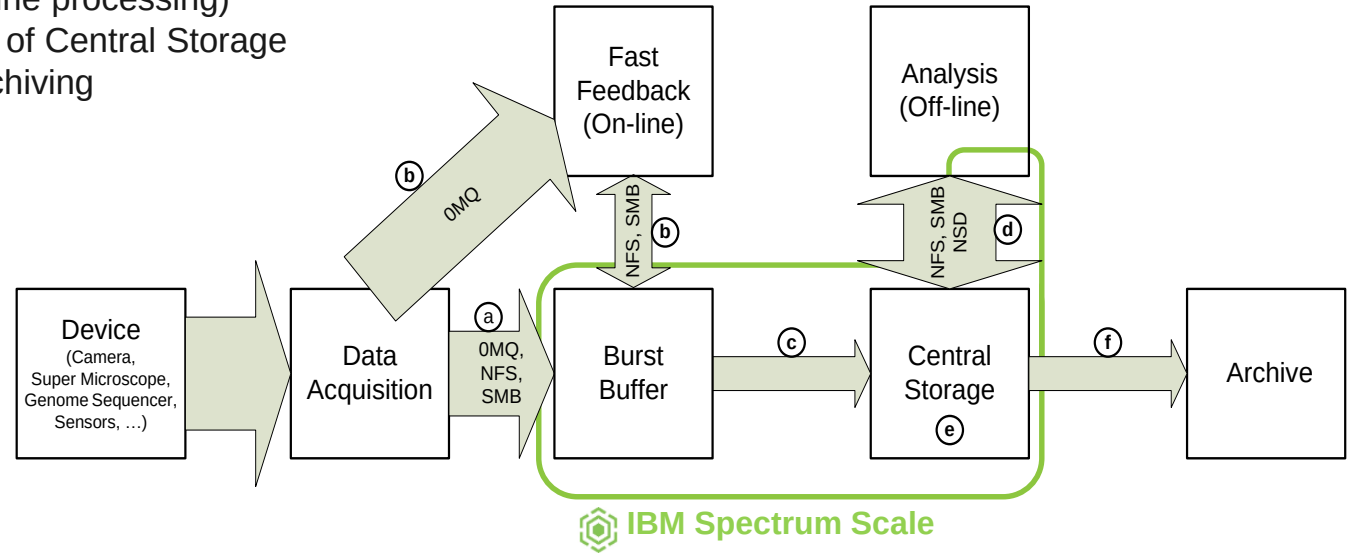
a) Real-time data ingest (data acquisition)
b) Visualization and near real-time analysis (online processing)
c) Data movement from Burst Buffer to Central Storage
d) Deep analysis (offline processing)
e) Data management of Central Storage
f) Long-term data archiving



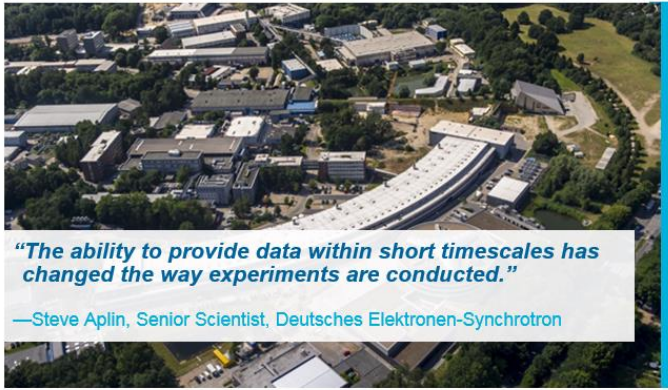
- Scientists need access to data during each stage of the workflow

Typical Workflow for Data Intensive Science (continued)

a) Real-time data ingest (data acquisition)
b) Visualization and near real-time analysis (online processing)
c) Data movement from Burst Buffer to Central Storage
d) Deep analysis (offline processing)
e) Data management of Central Storage
f) Long-term data archiving



- Scientists need access to data during each stage of the workflow
- IBM Spectrum Scale has proven to support this workflow



"The ability to provide data within short timescales has changed the way experiments are conducted."

—Steve Aplin, Senior Scientist, Deutsches Elektronen-Synchrotron

Business challenge

Research center Deutsches Elektronen-Synchrotron (DESY) found that increasingly resource-intensive experiments was affecting storage system performance, limiting research. How could the organization handle over five gigabytes of data streaming into its computing center every second?

Transformation

With a flexible, high-performance storage solution from IBM, DESY can meet growing demand cost-effectively. Scientists can now start analyzing the data in just a few minutes, instead of days, accelerating ground-breaking research.

Business benefits:

Ensures
DESY can easily maintain a multi-PB library of research data to meet growing demand and remain an attractive research destination

Rapid
access to millions of data points accelerates research and helps lead to breakthroughs

Increases
administration efficiency with automated data management, improving DESY's service offering

DESY

Making the next breakthrough in scientific research possible with the latest in storage innovation

DESY, Deutsches Elektronen-Synchrotron, is a national research center in Germany that operates particle accelerators and photon science facilities used to investigate the structure of matter. DESY is housed in Hamburg and Zeuthen, Germany, and attracts over 3,000 scientists from over 40 countries annually.

Solution components

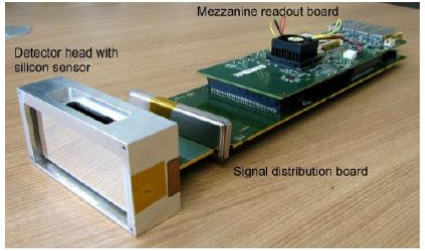
- IBM® Spectrum Scale™
- IBM Spectrum Scale RAID
- IBM Elastic Storage™ Server GS1
- IBM Elastic Storage Server GL4 and GL6
- IBM Power® S822L
- IBM Systems Lab Services

Share this

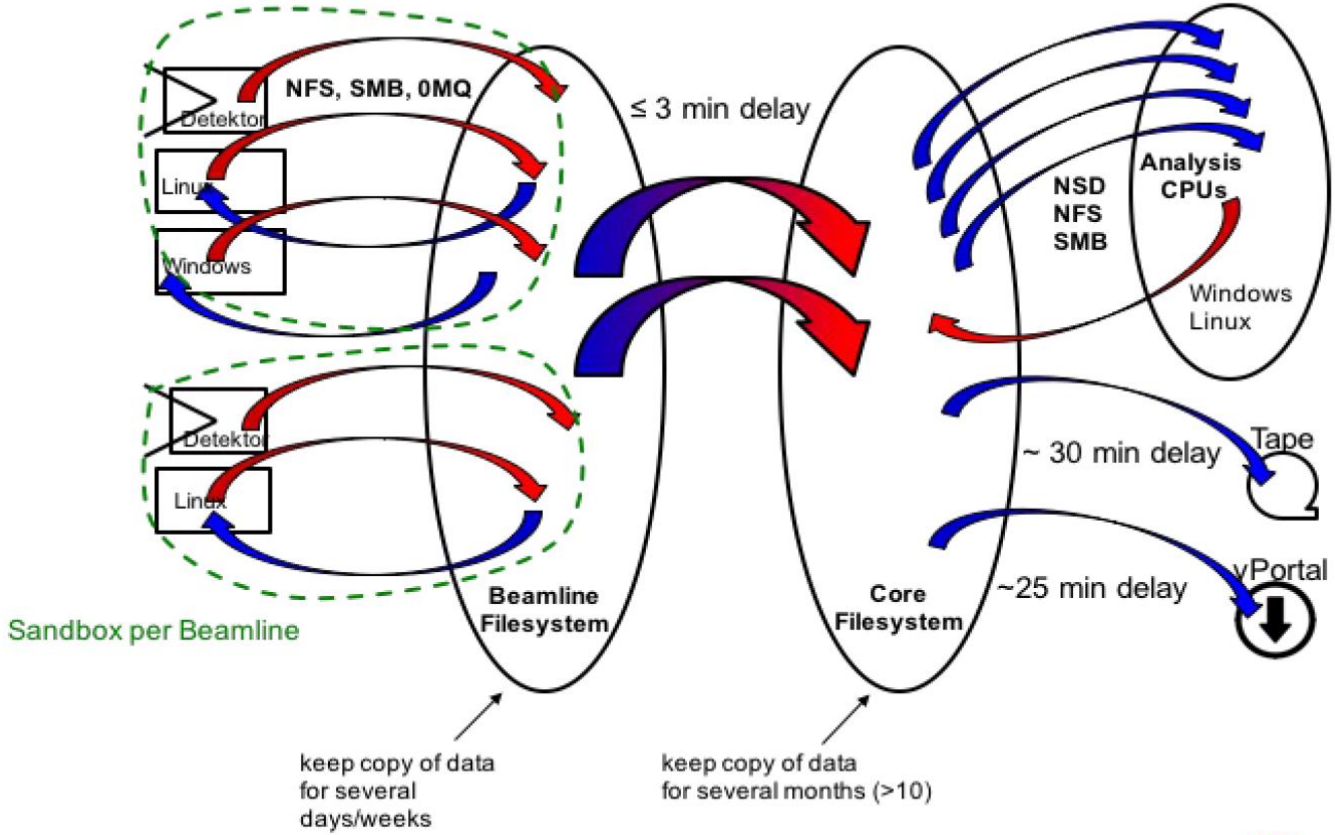
Current and Future Detector Rates

- Detectors exceeded capabilities of prev. system:
  - Pilatus 300k: 1,2 MB Files @ 200 Hz
  - Pilatus 6M: 25 MB files @ 25 Hz
    7 MB files @ 100 Hz
  - PCO Edge: 8 MB files @ 100Hz
  - PerkinElmer: 16 MB + 700 Byte files @ 15 Hz
  - Lambda: 60 Gb/s @ 2000 Hz (Future)
  - Eiger: 30 Gb/s @ 2000 Hz (Future)

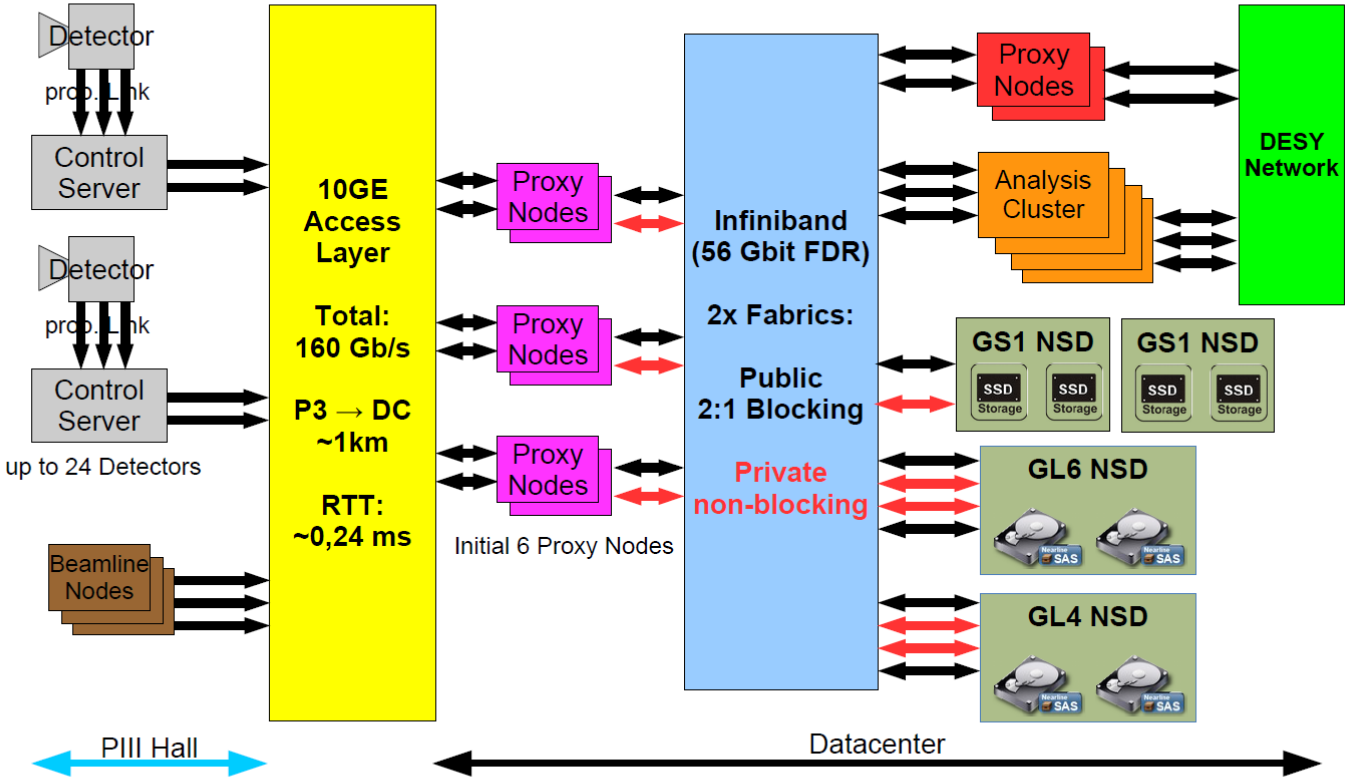
- GPFS is now used to handle those rates
  - SMB/NFS sufficient for current detectors
  - Future detectors need new methods



from the cradle to the grave



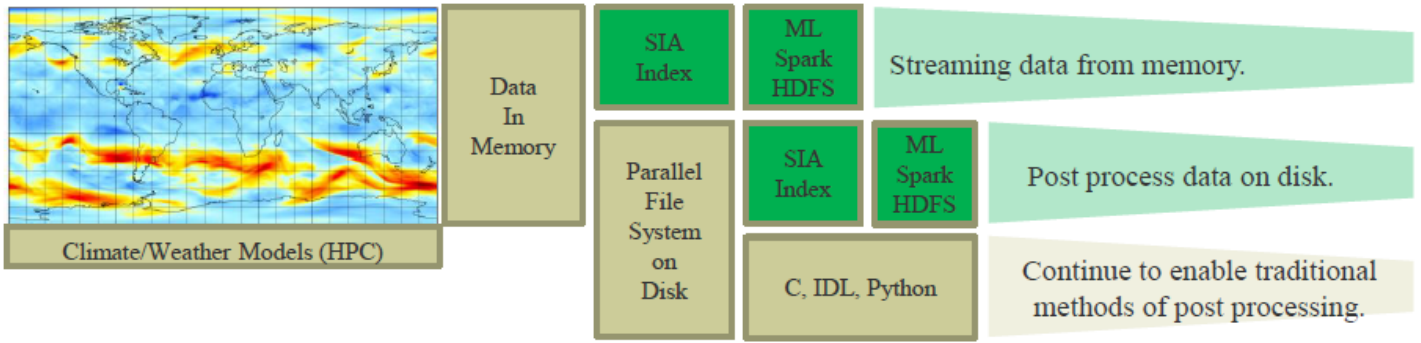
ASAP³ Architecture



Resources

- Detailed whitepaper published by DESY at CHEP2015 http://iopscience.iop.org/article/10.1088/1742-6596/664/4/042053
- DESY presentation at IBM Edge 2015: http://www.slideshare.net/UlfTroppens/desy-ibm-edge2015-technical-computing-for-photon-science-20150520v2
- DESY presentation at ALICE, ATLAS, CMS & LHCb Second Joint Workshop on DAQ@LHC: https://indico.cern.ch/event/471309/contributions/1981091/attachments/1257042/1856128/gpfs-for-p3xfel.pdf
- IBM customer story with video and four page pdf: http://ibm.co/1qCIAuL

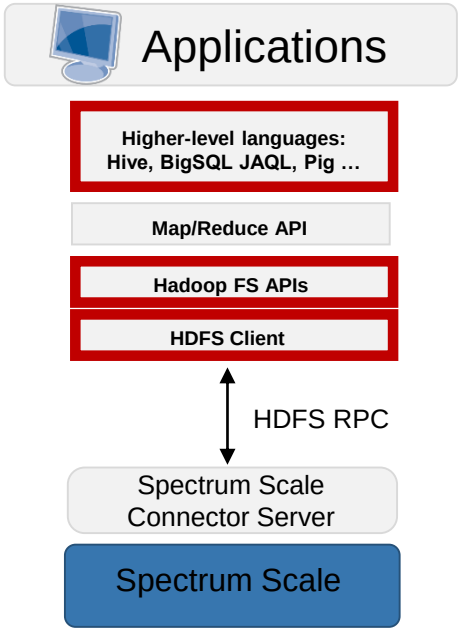
Future of Data Analytics



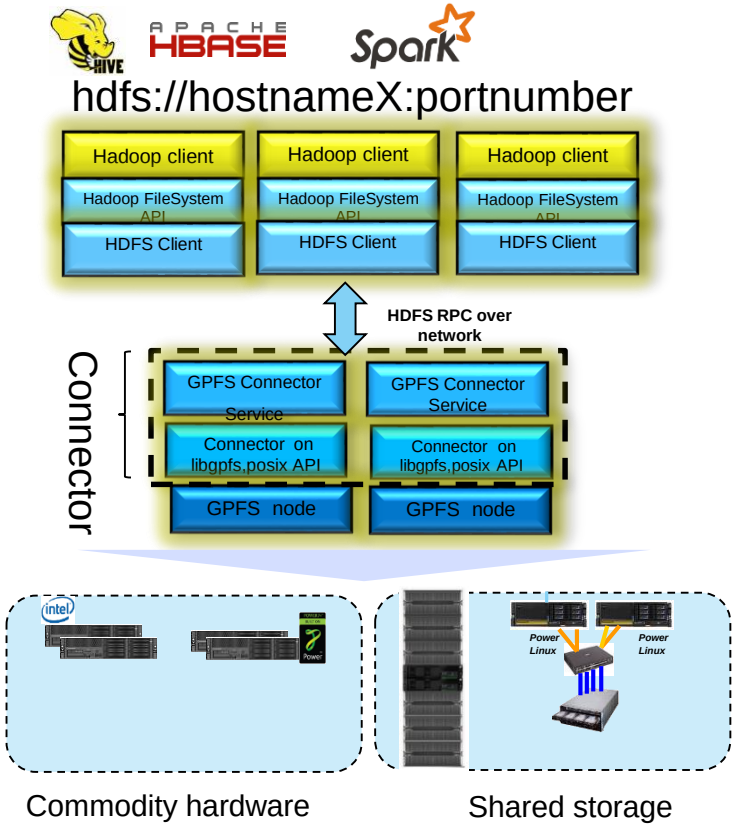
- Future HPC systems must be able to efficiently transform information into knowledge using both traditional analytics and emerging *machine learning* techniques.
- Requires the ability to be able to index data in memory and/or on disk and enable analytics to be performed on the data where it resides – even in memory
- All without having to modify the data

http://files.gpfsug.org/presentations/2016/SC16/06_-_Carrie_Spear_-_Spectrum_Sclale_and_HDFS.pdf

New Spectrum Scale HDFS Transparency Design



Supported Hadoop versions: 2.7.1



IBM Delivers New Platform to Help Clients Address Storage Challenges at Massive Scale

Las Vegas, NV (IBM PartnerWorld) – 14 Feb 2017: IBM (NYSE: IBM) and Hortonworks (NASDAQ: HDP) today announced the planned availability of Hortonworks Data Platform (HDP®) for IBM Elastic Storage Server (ESS) and IBM Spectrum Scale. The agreement with Hortonworks will lead to certification of Hortonworks HDP on Power with IBM Spectrum Scale and Hortonworks HDP on x86 with IBM Spectrum Scale.

https://hortonworks.com/press-releases/ibm-delivers-new-platform-help-clients-address-storage-challenges-massive-scale/

Store everywhere. Run anywhere.

Content Repositories

Challenge

Object storage for static data

- Seamless scaling
- RESTful data access
- Object metadata replaces hierarchy
- Storage efficiency

Spectrum Scale Swift & S3

- High-performance for object
- Native OpenStack Swift support w/ S3
- File or object in; Object or file out
- Enterprise data protection
- Spectrum Scale RAID (ESS)
- Transparent ILM
- Encryption of data at rest and Secure Erase

Store everywhere. Run anywhere.

Analytics without complexity

Challenge

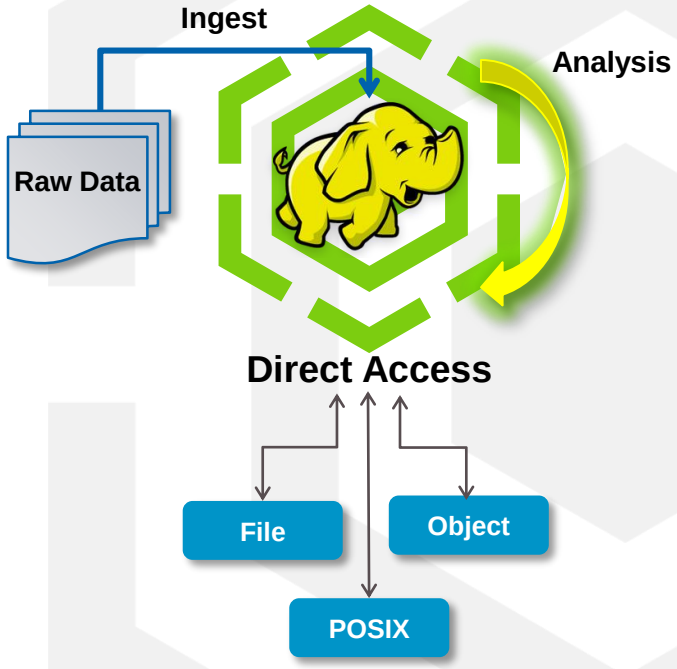
Separate storage systems for ingest, analysis, results

- HDFS requires locality aware storage (namenode)
- Data transfer slows time to results
- Different frameworks & analytics tools use data differently

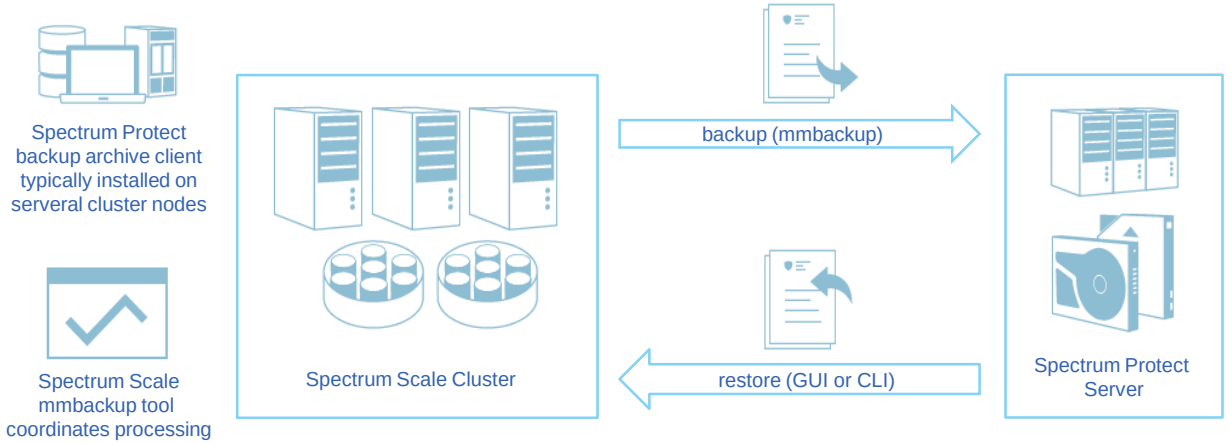
HDFS Transparency

- Map/Reduce on shared, or shared nothing storage
- No waiting for data transfer between storage systems
- Immediately share results
- Single ‘Data Lake’ for all applications
- Enterprise data management
- Archive and Analysis in-place

➔ Analyze object and file data without copying into HDFS



Backup Of Large Spectrum Scale File Systems



Function
▪ Massive parallel filesystem backup processing ▪ Spectrum Scale mmbackup creates local shadow of Spectrum Protect DB and uses policy engine to identify files for backup ▪ Spectrum Protect backup archive client is used under the hood to backup files to Spectrum Protect Server ▪ Spectrum Protect restore (CLI or GUI) can be used to restore files

➔ Use any backup program to backup file, object and Hadoop data
➔ Use Spectrum Protect to benefit from mmbackup and SOBAR to backup and restore huge amounts of data

Distributed workflows – University Queensland

Problem statement

- “At the moment if you store some data at this remote data centre in the cloud storage, and then you want to manipulate that on campus, you have to manually copy it, and then work on it locally and then send it back, or delete it, or whatever else, to manage it.”
- “The problem has been that until now, QRIScloud and the on-campus data centres have operated as separate storage silos”
- “This has made the process of storing and moving data from one location to another a less-than-user-friendly process for already busy researchers”

Customer Value

- “improve data access between its main campus and its off-site data centre”
- “The distributed file system will allow the working set of data that is currently being accessed by researchers to be cached on campus using IBM’s GPFS product (also known Spectrum Scale) without any additional user involvement.”
- “If you want to touch some data remotely in the cloud at Polaris, you can access that data using this distributed file system. If you want to access that data in campus, it appears automatically and gets moved without the user’s involvement. So that lessens the burden,”
- “A major additional benefit for users working with data stored at Polaris is that bandwidth bottlenecks between the campus and Polaris are less likely to cause a slowdown in their work.”

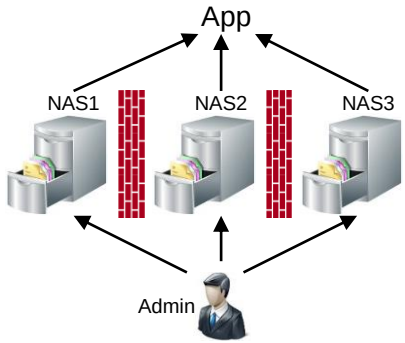
Customer feedback

- “GPFS is a parallel system, so it works well on high-performance computers, but it’s also a distributed file system. And they’re pretty smart guys [at IBM] that have worked out a lot of things over the years. It’s an old product, but it’s been constantly updated.”

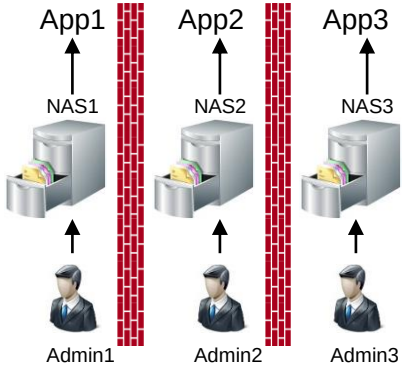
Source: http://www.itnews.com.au/news/qld-uni-debuts-its-own-data-storage-fabric-417506

Contrasting NAS Consolidation Opportunities

NAS Capacity Consolidation



NAS Filer Consolidation



Same Customer pain point …

Wants to consolidate multiple NAS filers on one large scale-out NAS system

… but different scenario

Single high-end NAS filer is not capable to provide storage capacity or throughput for one or a few applications

➔ Potentially good workload

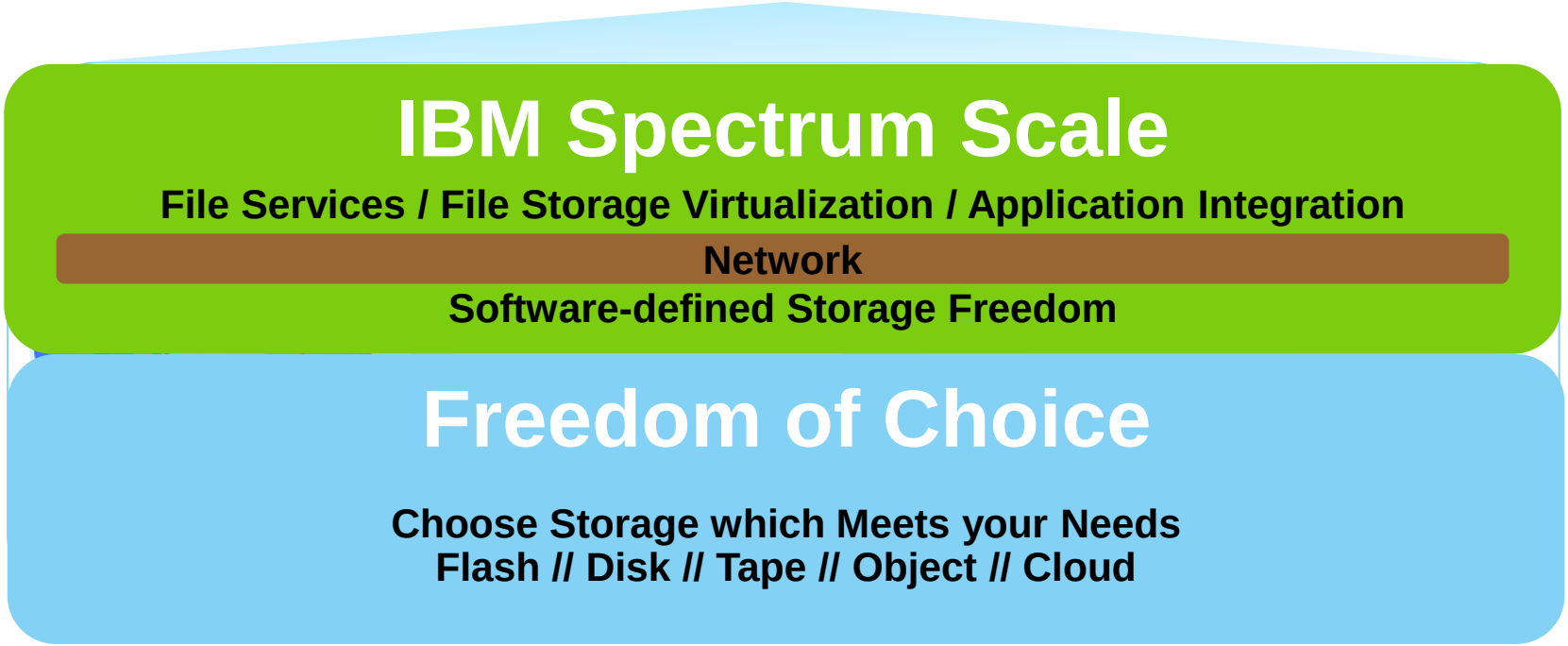
… but different scenario

Multiple NAS filers to provide storage capacity for broad range of applications and user groups

➔ Potentially dangerous workload

Examples – Application Acceleration

Core HPC | **File-based Workflows** | **Application Acceleration** | **Virtual Infrastructure**

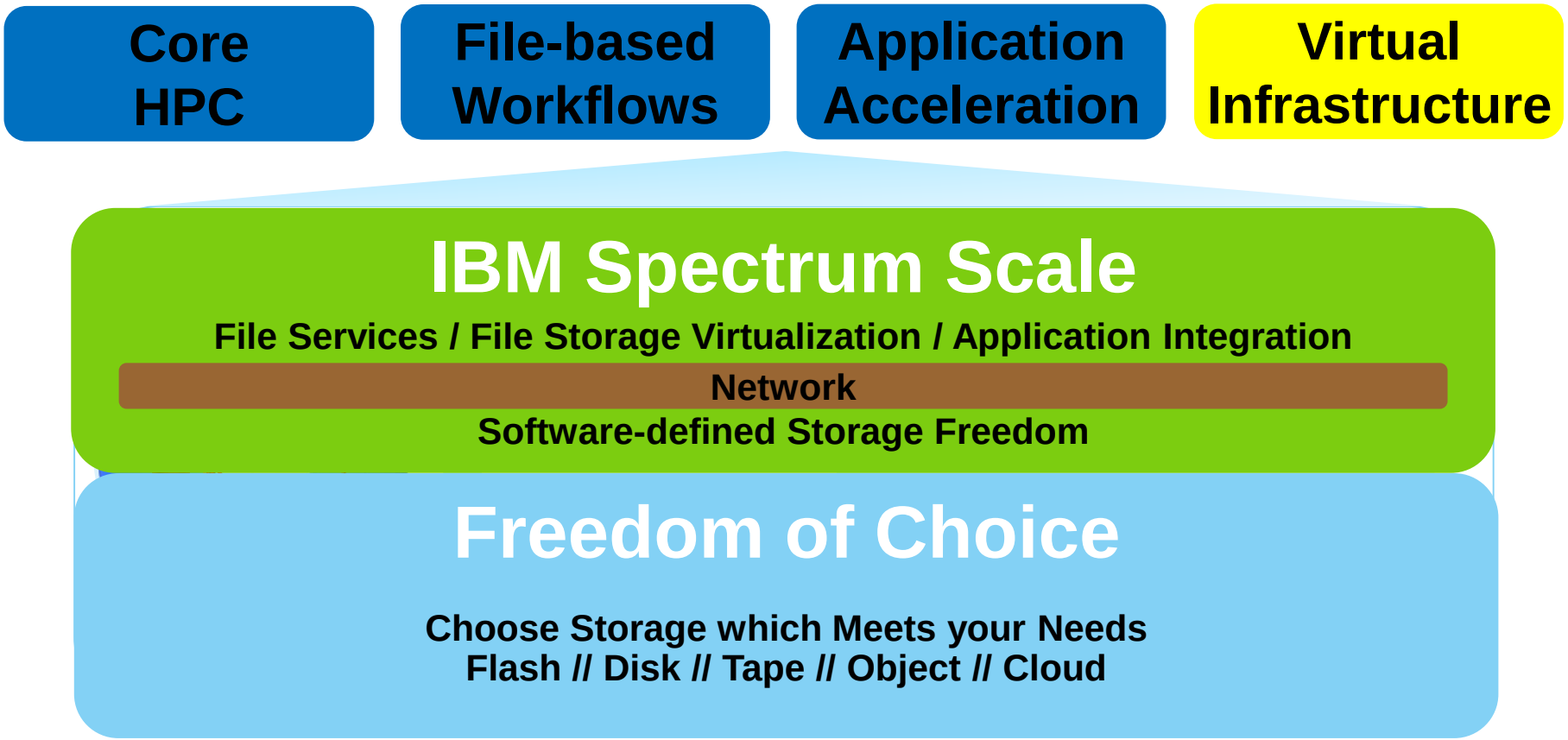


Application Acceleration

- The mature, fast and flexible file system accelerates a broad range of applications.
- Same applications have been covered in earlier user group meetings:
  - SAP HANA on IBM Elastic Storage Server (ESS)
    http://files.gpfsug.org/presentations/2017/Ehningen/04_-_SSUG17DE_-_Volker_Fischer_-_SAP_HANA@Bosch,_An_Infrastructure_Story.pdf
  - SAS on Spectrum Scale
    http://files.gpfsug.org/presentations/2017/Ehningen/06_-_SSUG17DE_-_Adrian_Immler,_Marco_Pighetti_-_Performance_Optimization_for_Risk_Management.pdf
  - Spectrum Protect on Spectrum Scale
    http://files.gpfsug.org/presentations/2017/Ehningen/11_-_SSUG17DE_-_Andre_Gaschler,_Stefan_Schaedeli_-_Flexible_Storage_for_Spectrum_Protect.pdf
- Which applications do you run on Spectrum Scale?

  ➔ Whiteboard

Examples – Virtual Infrastructure



Spectrum Scale as Infrastructure Service for OpenStack

BEAR

BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

Introducing #BEARcloud ... our private cloud for research

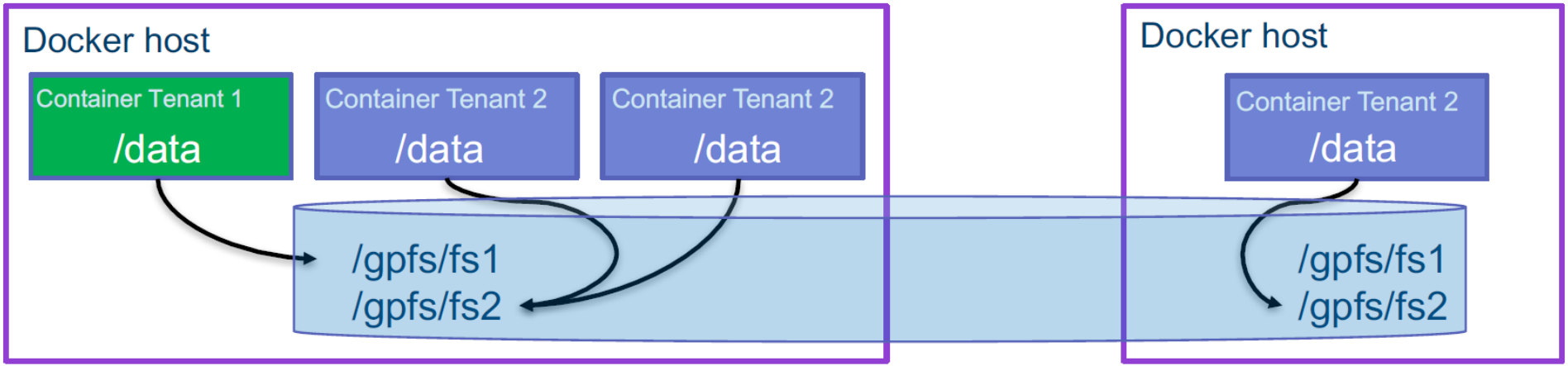
Published on October 22, 2016

Simon Thompson
Research Computing Infrastructure Architect at University of ...

17 4 16

https://www.linkedin.com/pulse/introducing-bearcloud-our-private-cloud-research-simon-thompson

Spectrum Scale and Docker



Details

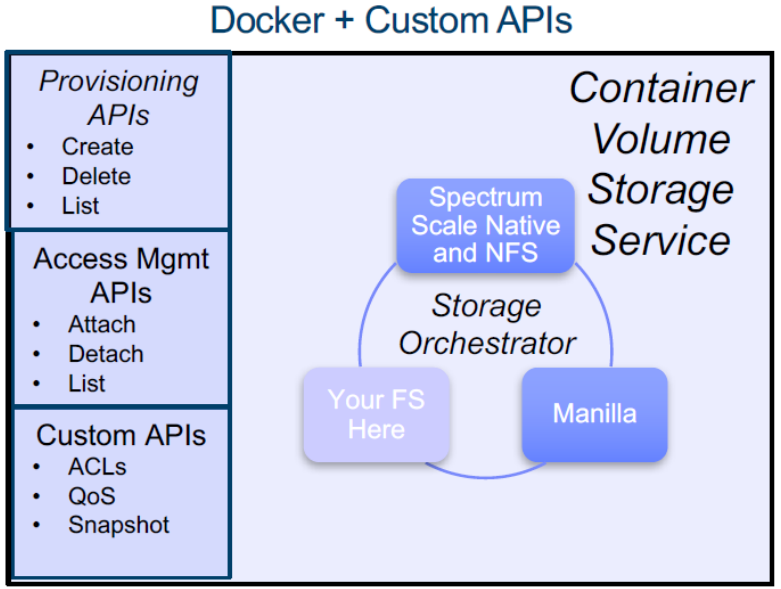
- Single or separate userid namespaces between containers and hosts
- Data sharing across containers and hosts
- All POSIX commands supported from container

A Few Benefits

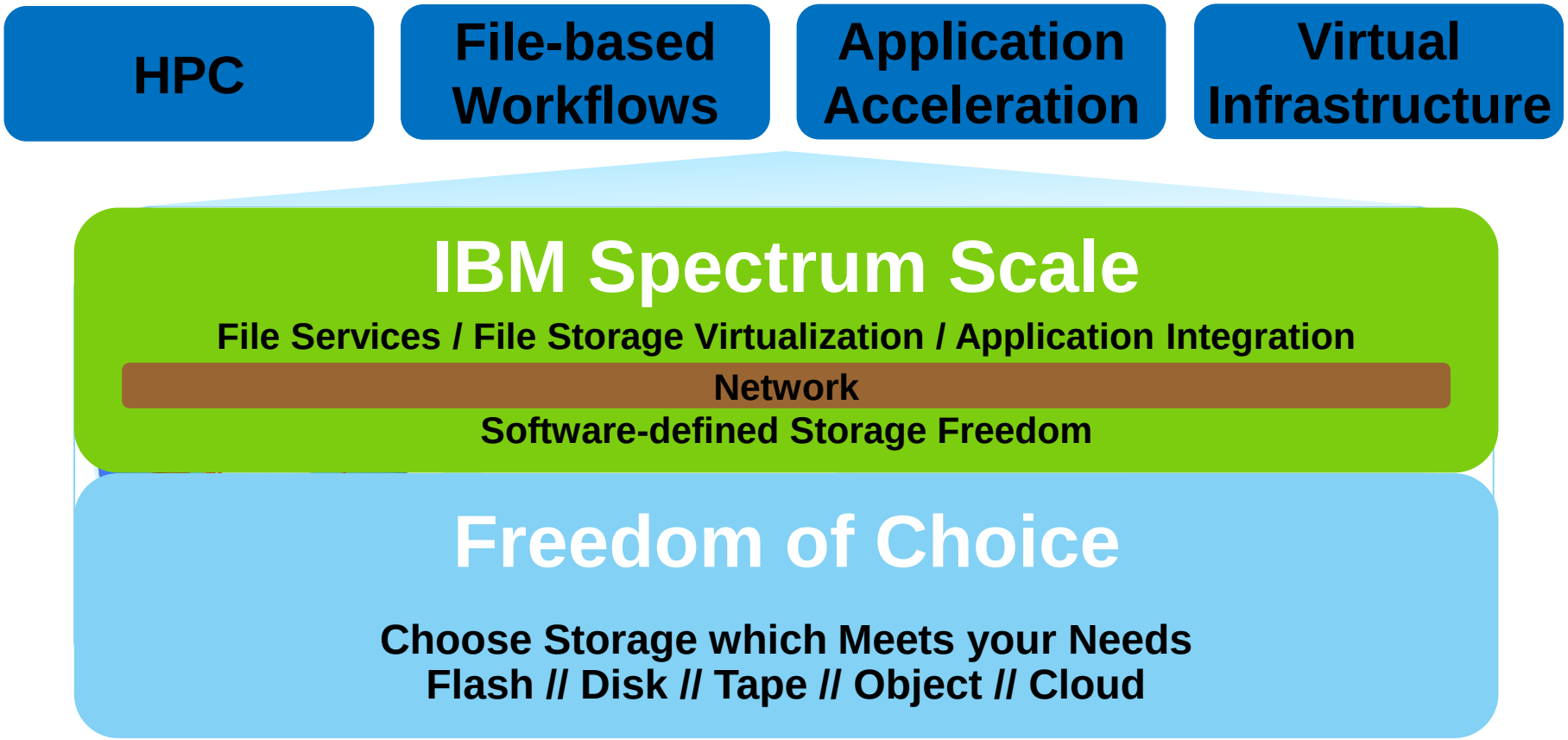
- Multi-tenant access
  - Container can only access its volumes
  - Allow root access in container without allowing root access to file system
  - ACLs can add an extra level of security
- Native client performance

Spectrum Scale Container Volume Service Architecture
Provisioning persistent storage for Docker and Kubernetes

- **Storage volume drivers** abstract storage complexities while exposing key capabilities
- Docker and Kubernetes have different mechanisms for provisioning storage
- Build a single service to provision volumes for that can be shared across all frameworks
- Support a variety of features, existing or new volumes, Quota, etc
- Currently working with Spectrum Computing
- Plan is to open-source



Miscellaneous



Spectrum Scale User Group

IBM Spectrum Scale

The Spectrum Scale User Group is free to join and open to all using, interested in or integrating Spectrum Scale.

Independent of IBM. IBM supports with speaker.

Join the User Group activities to meet your peers and get access to experts from partners and IBM.

http://www.spectrumscale.org/uk-may-2016-group-report/

Major meetings 2017:

- German User Meeting	**Mar 8+9,2017**	**Ehningen, Germany**
- US User Group Meeting	**Apr 4+5,2017**	**Berkeley, USA**
- Australia User Group Meeting	**Apr 27, 2017**	**Sydney, Australia**
- UK User Group Meeting	**May 9+10, 2017**	**Manchester, UK**
- User Group Meeting @ ISC17	**Jun 19, 2017**	**Frankfurt, Germany**
- User Group Meeting @ SC17	**Nov 12, 2017**	**Orlando, USA**

Web page: http://www.spectrumscale.org/
Presentations: http://www.spectrumscale.org/presentations/
Mailing list: http://www.spectrumscale.org/join/
Contact: http://www.spectrumscale.org/committee/

Agenda – Manchester – May 9th, 2017

10:00 - 10:30 Welcome - Simon Thompson (UG), Michael Daubman (IBM)
10:30 - 11:20 Spectrum Scale Update - Ulf Troppens (IBM)
11:20 - 11:40 ESS Update - Christopher Maestas (IBM)
11:40 - 12:00 Ellexus: Catching rogue jobs before they overload shared storage: and I/O profiling case study - Rosemary Francis
12:00 - 13:00 Lunch and Networking
13:00 - 13:20 Customer Talk - The UK Research Data Facility - Kieran Leach (EPCC)
13:20 - 13:40 DDN: AFM Case Studies - Vic Cornell
13:40 - 14:30 Spectrum Scale Cloud Deployments - Christopher Maestas (IBM)
14:30 - 14:50 Licensing - Carl Zetie (IBM)
14:50 - 15:20 Coffee and Networking
15:20 - 15:40 Arcastream: Automated data curation and discovery with Spectrum Scale and Arcastream - Jez Tucker
15:40 - 16:30 Problem Determination - Simon Lorenz (IBM)
16:30 - 17:20 Best Practices - Olaf Weiser (IBM)
17:20 - 17:50 Q&A - General
17:50 - 18:00 Wrap-up - Simon Thompson (UG)

Agenda – Manchester – May 10th, 2017

09:00 - 10:00 (Parallel Sessions)

1) Integration with Spectrum Protect for Backup - Fabian Kulhl (IBM)
2) Docker Support / Update on metadata sizing - Olaf Weiser (IBM), Indulis Bernsteins (IBM)
3) CES Authentication / Spectrum Scale Security (Meet the Devs) - Sandeep Ramesh (IBM)

10:00 - 11:00 (Parallel Sessions)

1) Introduction to AFM and Use Cases - Nils Haustein (IBM)
2) Object, SwiftHLM and OpenStack (cloud) - Sandeep Ramesh (IBM), Simon Lorenz (IBM)
3) GUI und REST API (Meet the Devs) - Alexander Wolf (IBM)

11:00 - 12:00 (Parallel Sessions)

1) Automation of Storage Services - Nils Haustein (IBM)
2) Container and Virtualisation support (Cloud) - John Lewars (IBM)
3) HDFS Support (Meet the Devs) - Piyush Chaudrhary (IBM)

12:00 - 13:00 Lunch and Networking

13:00 - 13:20 Mellanox: RoCE and Spectrum Scale - RDMA support and performance over Ethernet - Darren Harkins

13:20 - 13:40 Customer Talk - AFM Migration – The road to perdition - Mark Roberts (AWE) & Laurence Horrocks-Barlow (OCF)

13:40 - 14:00 Seagate: Accelerating Spectrum Scale with an Intelligent IO manager - Ray Coetzee

14:00 - 14:20 Customer Talk - TBC

14:20 - 14:50 Coffee and Networking

14:50 - 16:00 Research Topics - Sven Oehme (IBM)

16:00 - 16:30 Wrap-up and closing - Simon Thompson (UG), Michael Daubman (IBM)

Agenda – Poll Break-out Sessions

09:00 - 10:00

1) Integration with Spectrum Protect for Backup - Fabian Kulhl (IBM)
2) Docker Support / Update on metadata sizing - Olaf Weiser (IBM), Indulis Bernsteins (IBM)
3) CES Authentication / Spectrum Scale Security (Meet the Devs) - Sandeep Ramesh (IBM)

10:00 - 11:00

1) Introduction to AFM and Use Cases - Nils Haustein (IBM)
2) Object, SwiftHLM and OpenStack (cloud) - Sandeep Ramesh (IBM), Simon Lorenz (IBM)
3) GUI und REST API (Meet the Devs) - Alexander Wolf (IBM)

11:00 - 12:00

1) Automation of Storage Services - Nils Haustein (IBM)
2) Container and Virtualisation support (Cloud) - John Lewars (IBM)
3) HDFS Support (Meet the Devs) - Piyush Chaudrhary (IBM)

Customer Feedback

"I want meaningful alerts that don't cause alert fatigue. You can't tell the difference between a client leaving a cluster and a quorum node leaving a cluster."

"If we can't monitor something, we can't roll it out."

"What is going on with my GPFS system?"

"Our ops team is looking at dashboards all day. If something doesn't flash in red or come up on their monitoring console, they're not going to see it."

"This is an art that you learn from experience."

"What I really need is to be able to track down the rogue user who is bogging down the entire system."

"One of the things that's really lacking in GPFS is constant monitoring."
"There are tens of thousands of components that could break at any given time."

“When I come in to work each morning, give me a dashboard that surveys the entire Infrastructure landscape and tells me instantly if my day is going to be great or if it is going to pieces.”

2016 Development Priorities

Every year we define a set of goals

- Based mainly on client feedback and market opportunity
- Target is to achieve them within the year

Focus areas

- Problem determination
- Documentation
- Security
- Defect backlog

Functional enhancements

- Improvements for Big Data
- More flexibility for GNR

Hills – Problem Determination

1. An IT administrator who monitors Spectrum Scale can be made aware of the health of his Spectrum Scale components in one cluster, from a single place.

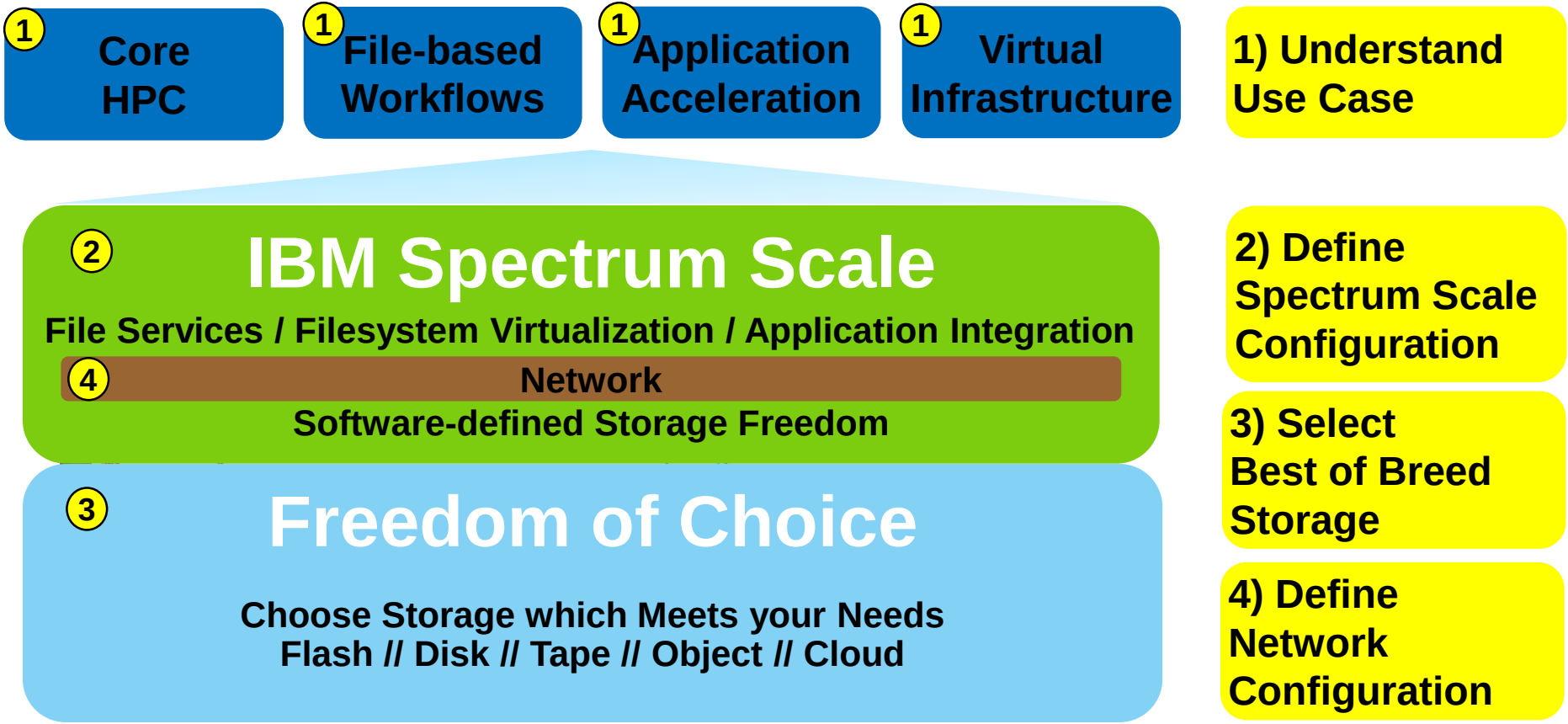
2. An IT Administrator, can perform self-service problem determination by utilizing provided guidance or automated solutions to problems, without contacting IBM Support.

3. An IT Administrator, can pre-check/check Spectrum Scale and its operating environment to avoid potential problems after initial installation or when changes are made, from a single tool.

Selected new features and functions

- Problem Determination Enhancements (Talk by Simon Lorenz)
- Diagnosing network problems in IBM Spectrum Scale with mmnetverify (Blog)
  https://developer.ibm.com/storage/2017/02/24/diagnosing-network-problems-ibm-spectrum-scale-mmnetverify/
- Extension of the problem determination capabilities in IBM Spectrum Scale 4.2.2 (Blog)
  https://developer.ibm.com/storage/2017/01/10/extension-of-the-problem-determination-capabilities-in-ibm-spectrum-scale-4-2-2/
- Hortonworks as a Strategic Partner (Breakout Session by Piyush Chaudhary)
  https://hortonworks.com/press-releases/ibm-delivers-new-platform-help-clients-address-storage-challenges-massive-scale/
- REST API (Break-out Sessions by Alexander Wolf-Reber)
- GUI Enhancements (Break-out Sessions by Alexander Wolf-Reber)
- IBM Spectrum Scale Security (Break-out Sessions by Sandeep Patil)
  https://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/redp5426.html?Open
- HPC enhancements for CORAL (Talk by Sven Oehme)
  http://files.gpfsug.org/presentations/2016/SC16/11_Sarp_Oral_Gautam_Shah_Spectrum_Scale_Enhancements_for_CORAL_v2.pdf
- What is new in Spectrum Scale 4.2.2 (Presentation)
  http://files.gpfsug.org/presentations/2016/SC16/04_Scott_Fadden_Spectrum_Scale_Update.pdf

Engagement Process



Engagement Process for Repeatable Engagements

1. Core HPC
1. File-based Workflows
1. Application Acceleration
1. Virtual Infrastructure

2. **IBM Spectrum Scale**

File Services / Filesystem Virtualization / Application Integration

4. Network

Software-defined Storage Freedom

3. **Freedom of Choice**

Choose Storage which Meets your Needs
Flash // Disk // Tape // Object // Cloud

Blueprint

1) Understand Use Case

2) Define Spectrum Scale Configuration

3) Select Best of Breed Storage

4) Define Network Configuration

Hills – Repeatable Engagements

1 Art, the seller, can give the winning proposal(*) to a client for genomic medicine workload that gives the data scientist faster time to results compared to competition.

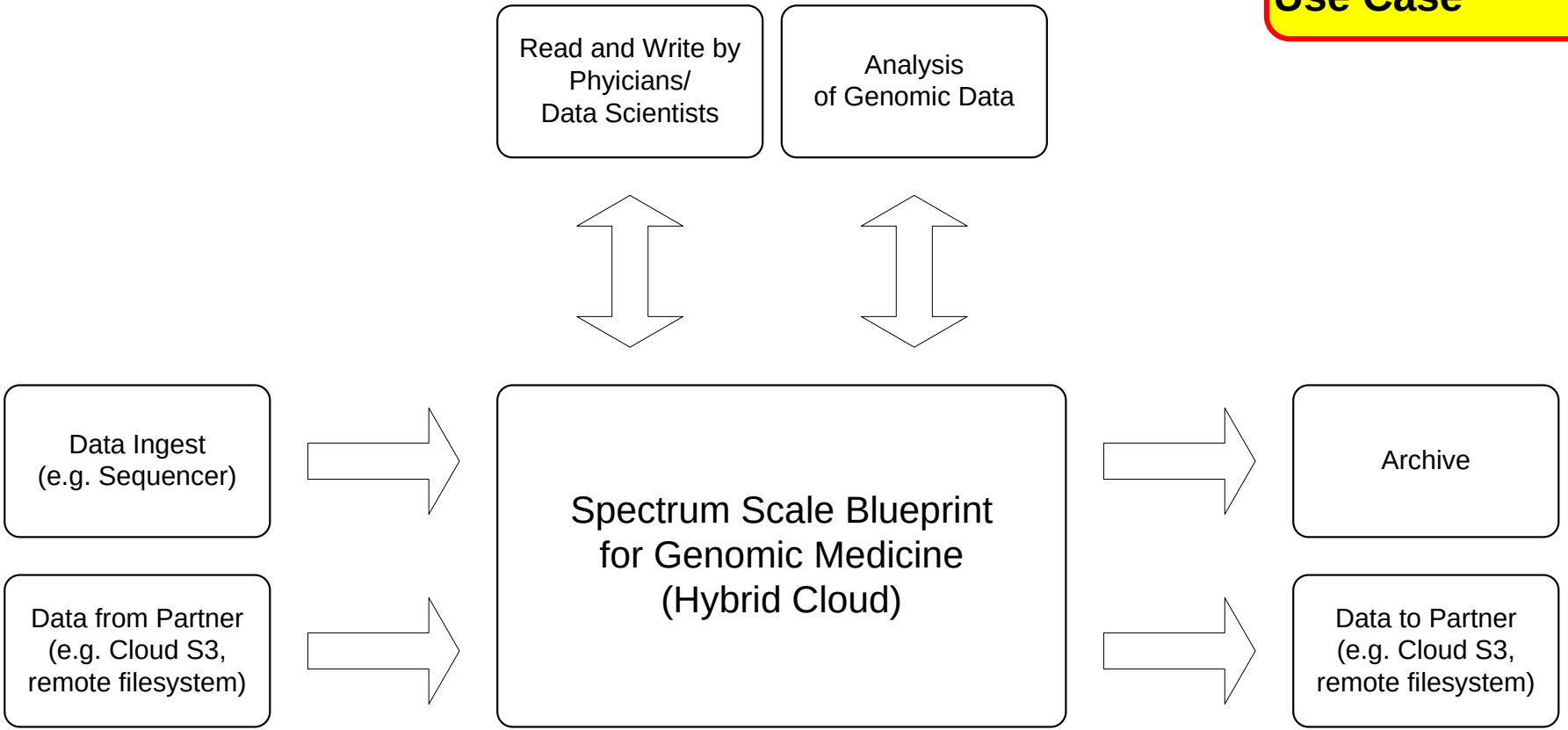
(*) Metric + win/loss reports

2 Aidan, the IT architect, can create a Spectrum Scale based solution architecture that meets genomic medicine workload performance requirements without consulting from IBM R&D.

3 Aidan, the IT architect, can integrate a Spectrum Scale based solution that meets their genomic workload performance requirements into an existing infrastructure without consulting from IBM R&D, without service disruption to the data scientist.

Blueprint for Genomics Medicine 1.0 – System Context

1) Understand Use Case



Keep in mind

- Spectrum Scale inherits all the benefits and maturity of GPFS and continues to support the needs of GPFS customers, in particular for HPC.
- Spectrum Scale is a file virtualization layer, which accelerates file- and object-based workflows, and reduces the costs for owning and accessing data on file.
- Spectrum Scale accelerates selected applications and simplifies the management of their underlying storage.
- Spectrum Scale is adding support for virtualized frameworks such as OpenStack, Docker and Kubernetes
- Get involved in Spectrum Scale User Group activities
  (e.g. speaker at the next User Group Meeting)!
- Volunteer as reference!
- Enjoy this User Group Meeting and make it an active exchange of thoughts and experience!

Legal notices

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectually property rights, may be used instead.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER OR IMPLIED. IBM LY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. IBM makes no representations or warranties, ed or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 1 0504- 785
U.S.A.

Information and trademarks

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.