

ISILON @GLOBUS WORLD

Patrick Combes

Senior Solution Architect for Life Sciences at EMC/Isilon

patrick.combes@isilon.com



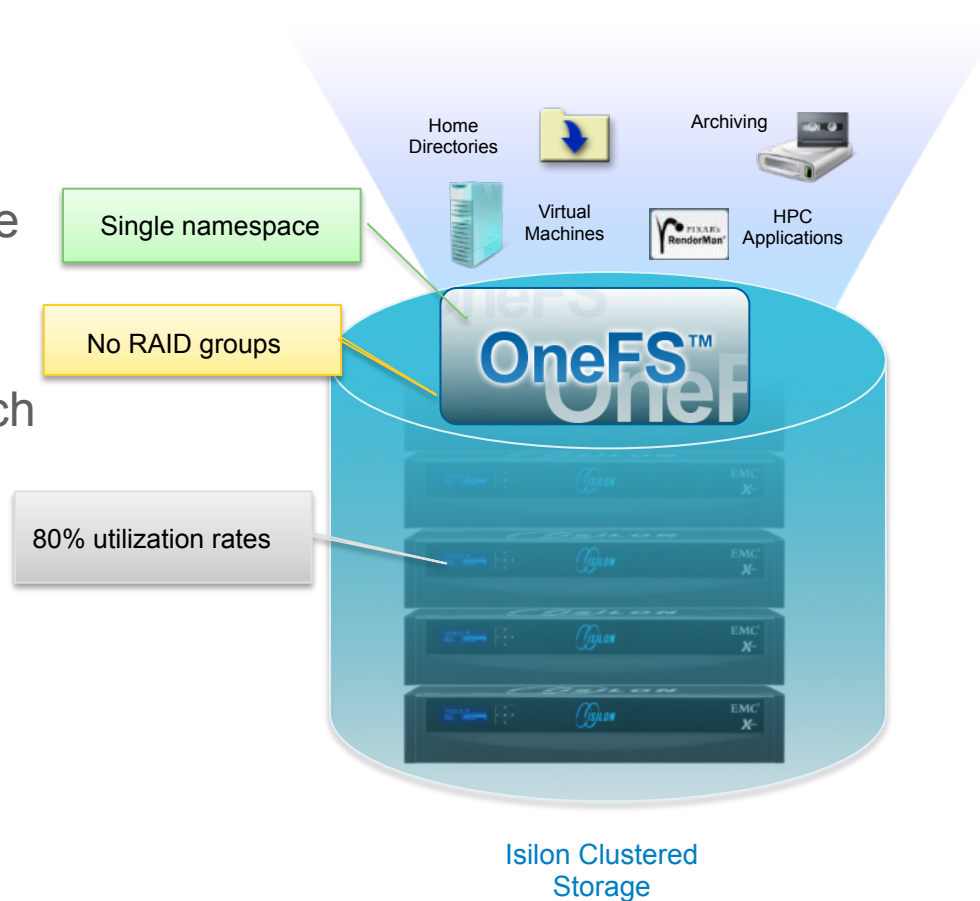
Isilon Overview

- Cluster of nodes, easily managed as a single unit, designed to fit any need around performance, scalability and connectivity
 - ▶ OneFS provides the framework for simple overall management, performance and scalability
 - ▶ SmartPools provide a way to migrate data around a cluster according to customer defined rules
 - ▶ SmartConnect is intelligent client connection software designed for optimizing performance and availability
 - ▶ Many other features available
 - ▶▶ synchronizing data (SyncIQ)
 - ▶▶ Locking and data lifecycle management (SmartLock)
 - ▶▶ Quota management (SmartQuotas)
 - ▶▶ Snapshots (SnapshotIQ)

OneFS

Isilon Clustered Storage

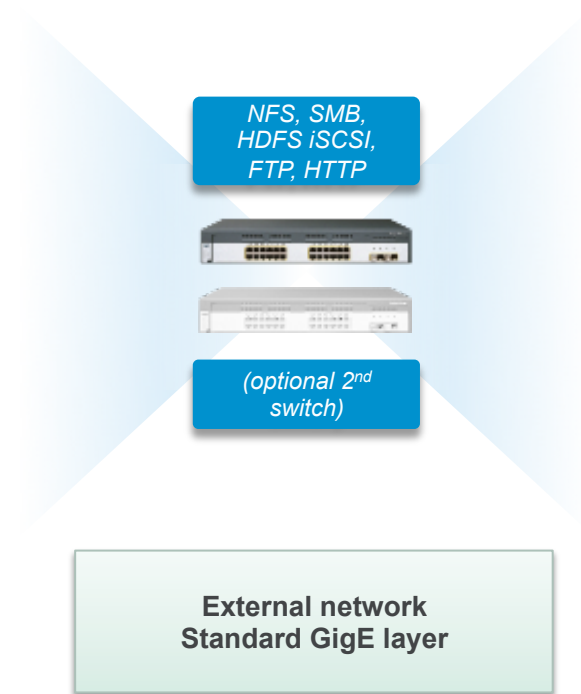
- Isilon clustered NAS storage uses a distributed file system - OneFS
- OneFS combines file system, volume manager, and raid into one unified software layer
- OneFS spans the entire cluster, which is managed as a single namespace
- Each node in a cluster is a peer
- As an Isilon cluster scales, OneFS eliminates the need to:
 - ▶ Manage separate islands of storage
 - ▶ Perform client-side management
 - ▶ Manually rebalance content by migrating data



Isilon Architecture

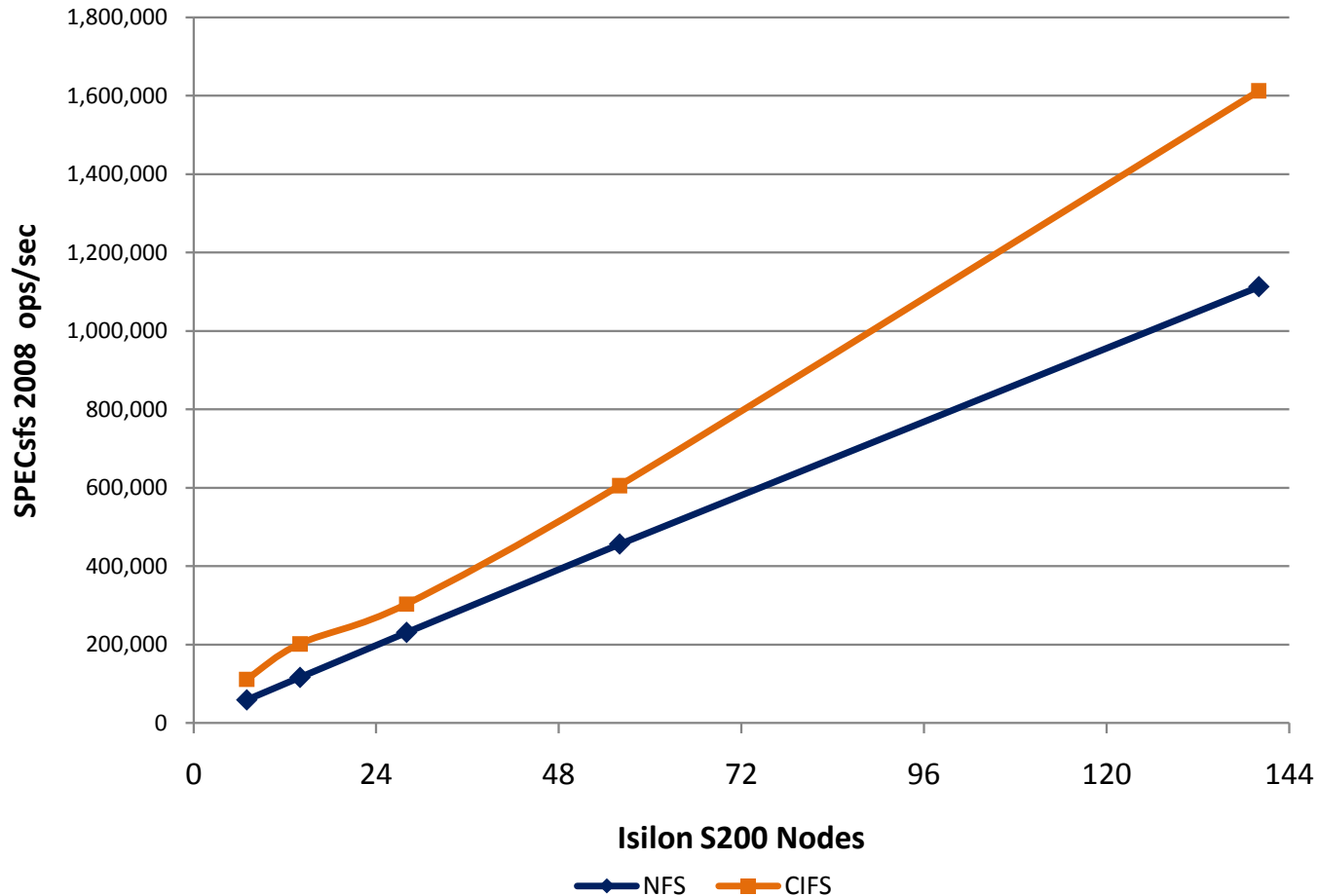
External Network

- Standard network file-sharing protocols are supported
 - ▶ NFS v3 & v4
 - ▶ SMB v1 & v2
 - ▶ HDFS
 - ▶ iSCSI
 - ▶ HTTP
 - ▶ FTP
- SmartConnect provides client connection management
 - ▶ Provides a virtual host name
 - ▶ Enables static or dynamic mapping



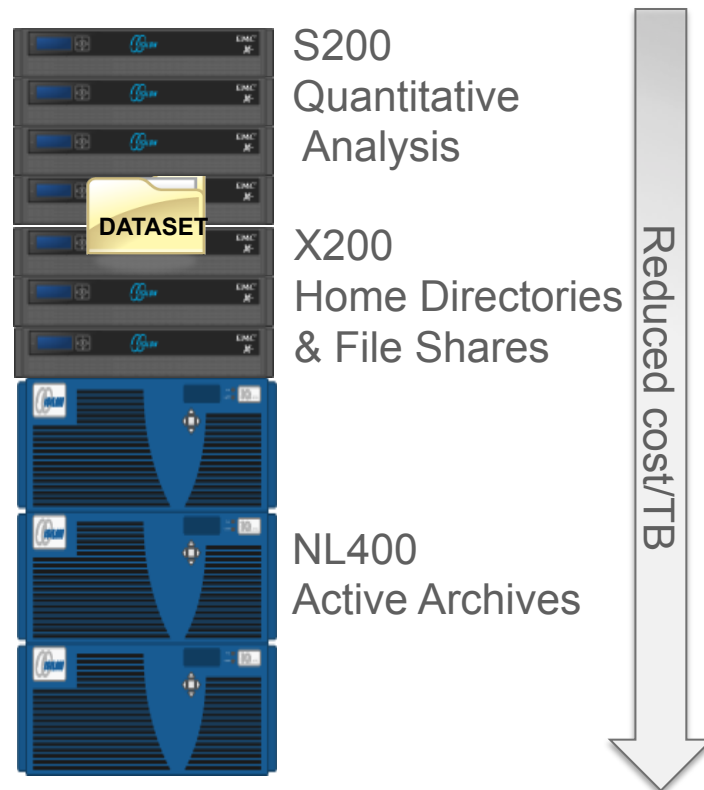
SpecSFS node scaling

Number of nodes	7	14	28	56	140
Network ports (10GbE)	7	14	28	56	140
Capacity (TB)	43.1	86.1	172.3	344.5	864.0
Memory (GB)	339.5	679	1,358	2,716	6,790
Number of disk drives	168	336	672	1,244	3,360
Node configuration	S200-6.9TB-200GB-48GB-10GBE				



SmartPools

- Policy-based, automatic data movement
 - Within a single namespace, file-system and volume
 - Without complex or risky links and stubs
- Automatically aligns the value of data with optimal storage performance
 - Tiering is transparent to end users and applications
- Increases available capacity of performance tier for active data
 - Reduces CAPEX
- Eliminates manual data migrations between performance/capacity tiers



SmartConnect

- SmartConnect is intelligent client connection software designed for optimizing performance and availability
- Can use a virtual IP failover scheme specifically designed for Isilon scale-out NAS and does not require any client side drivers
- If a node becomes unavailable, its IP addresses are automatically moved to other node interfaces in the pool, preserving NFS connections
- When the offline node is brought back online, SmartConnect rebalances the NFS clients across the entire cluster

Isilon by area in R&D

On-site data acquisition

- ❑ instrument/workstations can feed data directly into an Isilon cluster using SMB or NFS.
- ❑ Isilon supports a Global Name Space (GNS) which minimizes the need to move data around after it is copied to the cluster

Off-site data acquisition

- ❑ Data produced by other institutions or service vendors can be collected onto Isilon clusters using GlobusOnline or other standard transfer methods

Repository

- ❑ Isilon uses a Global Name Space (GNS) and expands easily to accommodate and manage large (multi-Petabyte) data repositories
- ❑ SmartPools provides a rule based system for data management

Isilon by area in R&D

Workspaces	<ul style="list-style-type: none">❑ user home directories can be staged within the same cluster giving users a space to arrange data for study and manage study metadata
Collaboration	<ul style="list-style-type: none">❑ Users within the same institution can leverage SyncIQ to replicate data between clusters❑ Isilon clusters can tie into institution-wide directory services (LDAP, AD, etc.) to provide a common user/permission framework❑ GlobusOnline can be leveraged to transfer data between institutions
HPC & Hadoop	<ul style="list-style-type: none">❑ Isilon supports quick access to data from multiple nodes via NFS/SMB and SmartConnect over multiple 10GbE connections❑ The Hadoop filesystem (HDFS) is natively supported
Archival	<ul style="list-style-type: none">❑ SnapshotIQ and SmartPools can both be used to manage the movement of data into an archive.❑ Isilon supports both near line and offline/tape-out archives.

Data in Research

- Managing data volume and the exchange of data are significant challenges in research today
 - ▶ Research & lab institutions (utilizing NGS, for example) can generate as much as a petabyte of data annually
 - ▶ Data gathering and collaboration efforts often extend across institutions and geographies
- Current state of research in the life sciences is fractured with items like sequencing and image gathering being done in one location and studies/analysis being performed at another
 - ▶ The transfer of large volumes of data remain a challenge
 - ▶ Permission management and access to data remain a challenge
- GlobusOnline has done significant work to address all of these challenges

Isilon and Globus in the Future

- To overcome these challenges, GlobusOnline and EMC Isilon are collaborating to develop research data management solution that will leverage Globus Connect Multi-User and Isilon OneFS technologies
- Deeper integration into the Isilon software stack – to enable faster and more easily automated transfers – will happen in the future
- Isilon is very excited to work with GlobusOnline to enable better, easier data transfer to facilitate future research

Isilon and Globus Genomics

- Isilon is an ideal storage fit for the new Globus Genomics service:
 - ▶ Scalable Storage – Isilon scales to fit any need and expands easily
 - ▶ Data libraries – Isilon provides an easily accessible solution for
 - ▶ Collaborators – Isilon and GlobusOnline are working together to facilitate easy and efficient data transfer between institutions

Resources

- OneFS <http://www.emc.com/collateral/hardware/white-papers/h8202-isilon-onefs-wp.pdf>
- Isilon in the Life Sciences: <http://www.emc.com/collateral/solution-overview/h11212-so-isilon-life-sciences.pdf>
- Contacts:
 - ▶ Patrick Combes – Sr. Solution Architect/Life Sciences – patrick.combes@isilon.com
 - ▶ Sasha Paegle – Sr. Business Development Manager/Life Sciences – sasha.paegle@isilon.com
 - ▶ Sanjay Joshi – CTO/Life Sciences – sanjay.joshi@isilon.com