The POWER of EMC ISILON To Transform eResearch

Greg Rogers, RTM
EMC ISILON Australia
Charles Sevior, CTO
EMC ISILON Asia-Pac
Agenda – Topics to be covered

1. Isilon eResearch customer references - Charles
2. Isilon NAS brief overview - Charles
3. Aspera integration into Isilon OneFS – JC
4. Hadoop analytics integration with Isilon - Charles
5. Isilon OneFS software / hardware roadmap – Charles
6. Next steps and partnerships - Greg
EMC Isilon Established Leadership in Healthcare, Life Sciences, eResearch
Challenge

- Using a traditional NAS system that struggled to support the huge amounts of input/output demands on the 400 CPUs in our computing infrastructure
- Imminently outgrowing their infrastructure

Solution

- X-Series, NL-Series
- SmartPools
- SmartQuotas

Results

- Now supports some 4,000 CPUs, which can handle the heavy I/O and data analysis demands of C2B2’s research
- Supports the storage needs of three additional entities within the Columbia University network

"After switching to Isilon, we no longer had to worry that our system couldn’t handle our research demands. We knew that we could independently scale capacity and performance, so that we buy only what we need, when we need it."

JOHN LOWELL WOFFORD
Director IT Services

Columbia University Center for Computational Biology and Bioinformatics ("C2B2")
3TIER

Turns Big Data Into Accurate Renewable Energy Intelligence

Challenge

- Traditional RAID array couldn’t scale to meet the demands of its big data workflow
- Complicated management
- Escalating costs

Solution

- X-Series
- NL-Series
- SmartPools

Results

- Single file system and point of management for multiple performance tiers
- Unifying all operations on a single, shared pool of storage
- Simplifying Big Data management

“Before Isilon, our team had to do everything from manual data migrations to mapping directory paths, which simply wasn’t sustainable for a business growing as quickly as ours. With Isilon, everything is simple. We’ve eliminated data management and storage headaches, freeing up resources to focus on the services that deliver real value to our clients.”

PAUL ENGLISH
Director of IT at 3TIER
RENCI
Renaissance Computing Institute of the University of North Carolina

- Turning Big Data into Insight in the lab and therapy in the clinic
- Knowledge based medicine programs for epilepsy & prostate cancer
- Secure medical workspace
- Informatics for Genetic Sequencing (IGS)
- Blending traditional cluster computing with iRODS and Hadoop analytics

eResearch Challenges
User requirements for RDSI?

- Growing data sets – mostly unstructured
- Reliable storage integrity / preservation
- Manage migration issues – active and cold data
- Leverage MapReduce and Hadoop analytics
- Ease of use and simple management
- Trust / Security / Immutability
- Sharing / Collaboration / Cloud – REST API
EMC Isilon
Unique Differentiators
Challenges arising from standard NAS

- Utilisation rates vary across islands
- QoS headroom
- Performance considerations

= Storage Efficiency 50% - 65%
EMC Isilon – NODE-Based Architecture

- **Servers/Clients**
- **Clients**
- **Compute Nodes**

### Clustered Storage Layer
- **NFS, SMB, HTTP, FTP, HDFS**
- **Facility Ethernet Network**

### Ethernet Layer

### Intra-cluster Communication Layer

**Private cluster InfiniBand network**
Focus on the data not the storage
**AutoBalance:** Automated data balancing across nodes reduces costs, complexity and risks for scaling storage

- AutoBalance migrates content to new storage nodes while system is online and in production
- NO manual intervention
- NO reconfiguration
- NO server or client mount point or application changes
- Eliminates “Hot Spots”
Data Layout - protection and throughput

- Data is striped across the nodes
  - Not across disks
  - FEC not RAID

- Data breakdown:
  - 8 KB blocks
  - 16 blocks per stripe unit
  - 128 KB stripe width per drive
With n+1 protection, data is 100% available even if one drive or one node fails.

Fastest rebuild time, and with Isilon, the more nodes in the cluster, the faster drive rebuild time.
EMC Isilon – Storage Never Obsolete

“No Node Left Behind”

- Generations of Isilon nodes can coexist
  - Heterogeneous nodes mix and match
  - Storage investment protected
- Rapid introduction of the latest technology
  - Faster, denser, greener storage each year
  - New capabilities like Hadoop and MobileIQ
  - No server, network or application changes
- Push-button node retire – SmartFail
  - Storage older than 5 years is a waste of space!

*No More Data Migration. Keep up with Moore’s Law*
Isilon + Aspera
Store & Move Big Files

JC Diomard Arrazau
Aspera
EMC Isilon with integrated Aspera FASP
The Benefits of Isilon + Aspera

- Aspera and Isilon have partnered for many years to create a predictable, non-disruptive, high-performance wide area file and data delivery solution specifically for moving large data sets over long distances at the fastest possible speed.

- Aspera’s cluster-aware high-speed transfer software coupled with EMC Isilon OneFS operating system and scale-out NAS architecture is a powerful combination for distributed storage environments.
The Benefits of Isilon + Aspera

- Maximum cluster-to-cluster performance over campus and wide-area networks
- Massive concurrency
- Easy scalability of WAN transfer performance
- No single point of failure
- Single point of management
- Web services and industry-standard API support
Next-Gen Analytics

Key Features
- Hadoop 2.0
- Native HDFS Support
- Pivotal HD Support
- Simultaneous HDFS 1.0 & HDFS 2.0 Support
- Distributed NameNode

Benefits
- Support For Open Hadoop Applications
- No Single Point Of Failure
- Improved TCO
**HDFS: Standard Hadoop Cluster**

1. **Step 1:** Data is copied into the Landing Zone

2. **Step 2:** Data is copied from landing zone onto a node in the Cluster

3. **Step 3:** Data is locally copied from a node onto the Cluster (3 times)

4. **Step 4:** MapReduce Job is Run

**Network Protocols:**
- SMB, NFS, HTTP, FTP

**Network Resources:**
- Name node
- Compute
- Data
- Landing Zone Servers

**Applications:**
- Decision Support Databases
- OLAP
- Log Files

**User/WEB Click data**
**HDFS: Integrated Isilon and Hadoop**

- **Step 1:** Much or all of the Data lives on the Isilon Cluster
- **Step 2:** Jobs are run

**SMB, NFS, HTTP, FTP, HDFS**

- **User/WEB Click data**
- **Decision Support Databases**
- **OLAP**
- **Log Files**

**Isilon**

Hadoop Cluster

- MAP Reduce
- MAP Reduce
- MAP Reduce
- MAP Reduce

**Name node**

**Data node**

**User/WEB Click data**

**Decision Support Databases**

**OLAP**

**Log Files**

**SMB, NFS, HTTP, FTP, HDFS**

**Isilon**

**Step 1:** Much or all of the Data lives on the Isilon Cluster

**Step 2:** Jobs are run

**Isilon**

**Hadoop Cluster**

- MAP Reduce
- MAP Reduce
- MAP Reduce
- MAP Reduce

**Name node**

**Data node**
ARCHITECTURE FOR HIGH PERFORMANCE COMPUTING WITH INTEGRATED ANALYTICS

ISILON – CAUDIT presentation to RDSI

© Copyright 2013 EMC Corporation. All rights reserved.
Future Directions
OneFS 7.0 - Mavericks

**Scalability**
- **Performance**
  - Improved Throughput
  - Reduced Latency
- **Capacity**
  - SnapshotIQ File Clones
  - 20 PB File System Ready
- **Extensibility**
  - REST Platform API

**Enterprise**
- **Multi-Tenancy**
  - Authentication Zones
  - Role Based Administration
- **VMware**
  - Reduced I/O Latency
  - VAAI, VASA, SRM
- **Security and Archive**
  - SmartLock v2 (SEC17a-4)

**Management**
- **Data Protection**
  - SyncIQ: Failover/Back
  - SnapshotIQ File Clones
- **Automated Provisioning**
  - Improved Efficiency
  - Higher Reliability
- **Availability - NDU**
  - Upgrade - 2 Minutes

ISILON – CAUDIT presentation to RDSI

© Copyright 2013 EMC Corporation. All rights reserved.
OneFS 7.1 - Waikiki

**Efficiency Scale**
- Deduplication
  - Post process, block level deduplication
  - Increased data efficiency
  - Multi-level directory configuration
  - Dry-run dedupe savings estimation tool
- SmartPools Enhancements
  - CLI and API management
  - RBAC integration
  - User defined node-pools

**Enterprise**
- Security and Archive
  - CEE support. Varonis first partner for Audit
  - Improved RBAC Coverage
  - Encrypted Nodes
- Backup
  - Faster incrementals and incrementals forever
- Manageability & Extensibility
  - RBAC improvements to API
  - Broaden pAPI coverage incl. SyncIQ AND JobEngine
  - ESRS Gateway Integration

**Performance**
- SyncIQ & Snapshots
  - “Continuous Mode” Replication
  - ~1 minute sync granularity
- Job Engine
  - Improved impact control
  - Multiple simultaneous jobs
  - Enhanced per-worker resource monitoring
- Breaking boundaries
  - 3 simultaneous background jobs

ISILON – CAUDIT presentation to RDSI
Growth In Applications

Traditional Applications

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>83M</td>
</tr>
<tr>
<td>2016</td>
<td>141M</td>
</tr>
</tbody>
</table>

70% growth from 2012 to 2016

Next Gen Cloud Applications

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>6M</td>
</tr>
<tr>
<td>2016</td>
<td>48M</td>
</tr>
</tbody>
</table>

700% growth from 2012 to 2016

Sources: IDC, Gartner, AWS Workload Estimates
Isilon’s OneFS Operating System

Strategic Vectors

- Making Compute the Bottleneck
- A Giant Leap in Capacity
- Around-the-Clock, Around-the-Globe
- Absolute Freedom and Control
OneFS Storage

Isilon, The Big Data Platform

NEXT GEN APPLICATIONS

VERTICAL DATA CENTRIC APPLICATIONS

ISILON – CAUDIT presentation to RDSI
<table>
<thead>
<tr>
<th>Area</th>
<th>Strength</th>
</tr>
</thead>
</table>
| Onsite data collection | • Multi-protocol support for instruments & workstations  
• CIFS (SMB), NFS, FTP, HTTP                                                                 |
| Repository        | • SmartPools (Automatic storage tiering)  
• GNS (Global Namespace, with metadata acceleration)                                                                                   |
| Workspaces        | • SmartQuotas (Policy-based storage management)  
• Home Directories                                                                                                                       |
| Collaboration     | • Aspera FASP protocol integrated into OneFS  
• Aspera Enterprise Server and Connect Server hosted within the cluster                                                                   |
| HPC/Hadoop        | • HDFS (native support)  
• 10GbE (2/node)  
• SmartConnect  
• Concurrent File Access support (OneFS 7.0)                                                                                                |
| Archival          | • SnapshotIQ  
• WORM – immutable writes  
• NL Series                                                                                                                                  |