

Active Archive

Extending File Systems to Tape:
Enabling New Methods to Manage
Explosive Data Growth

www.activearchive.com

Online Access to ALL Your Data.

Active Archive Alliance

- Founded April 27, 2010
 - Education
 - Best practices
 - Intercompatibility testing and demonstrations



<http://www.activearchive.com>

ACTIVE Archive

Online Access to ALL Your Data.

Definition of Active Archive

Active Archive provides affordable, file system-based access to storage for all created data.

An archive contains the primary copy of data, no matter how old or infrequently accessed, that can still be indexed, searched and retrieved.

Members

Atempo 

 CROSSROADS[®]



 FUJIFILM

 GRAU DATA

 QSTAR[®]
TECHNOLOGIES

sgi[®]

 SPECTRA[®]

 Designates Board of Director representation

* Joined and pending announcement – Quantum, Imation and Cleversafe

ACTIVE Archive

Online Access to ALL Your Data.

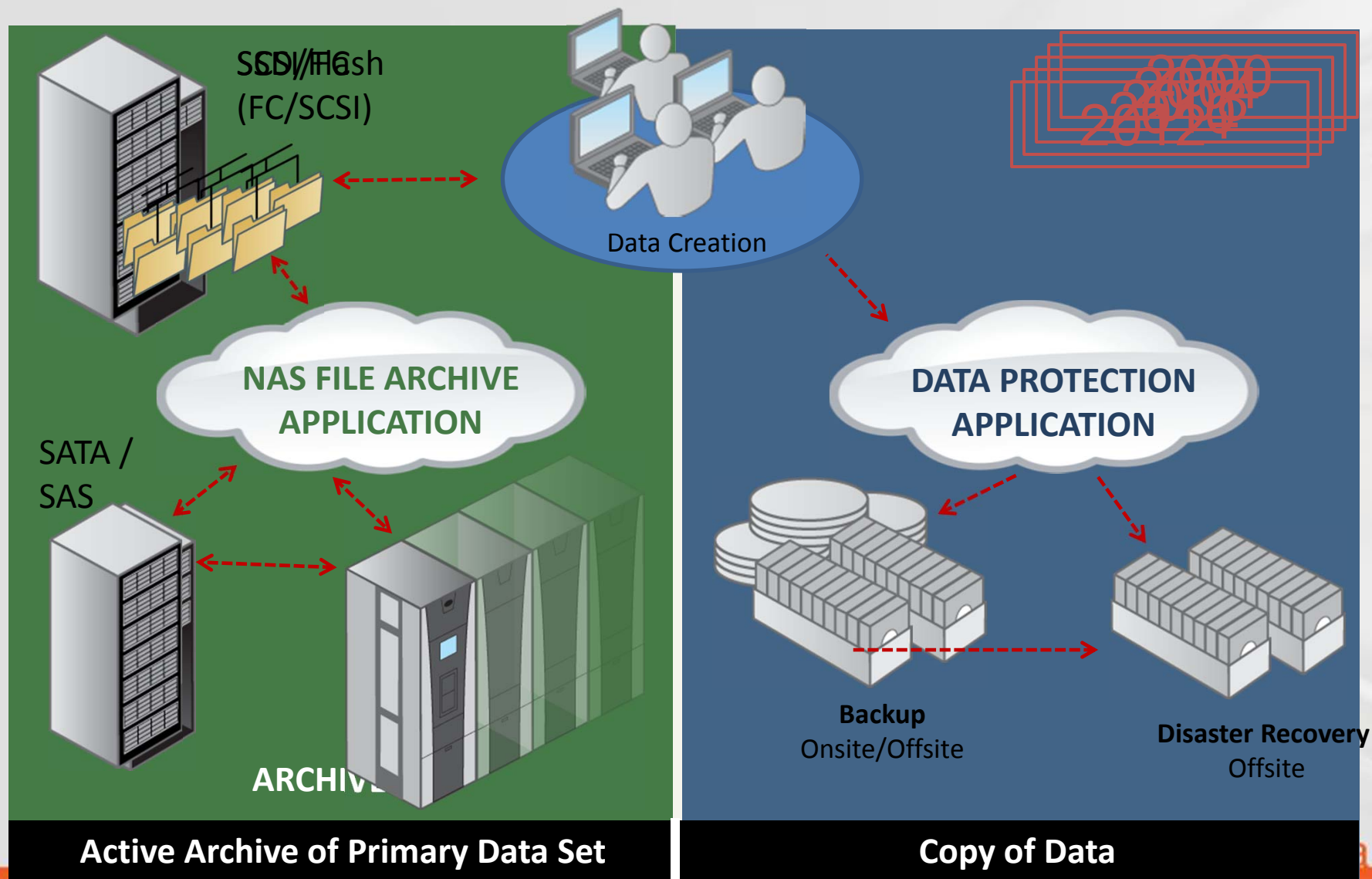
File Archive To Tape: Solves Data Center Problems

1. Offload & extend primary storage
2. Obtain business value from all stored data
3. Protect from viruses, corruption, online replication of data integrity problems



Online Access to ALL Your Data.

The Evolving Role of Tape

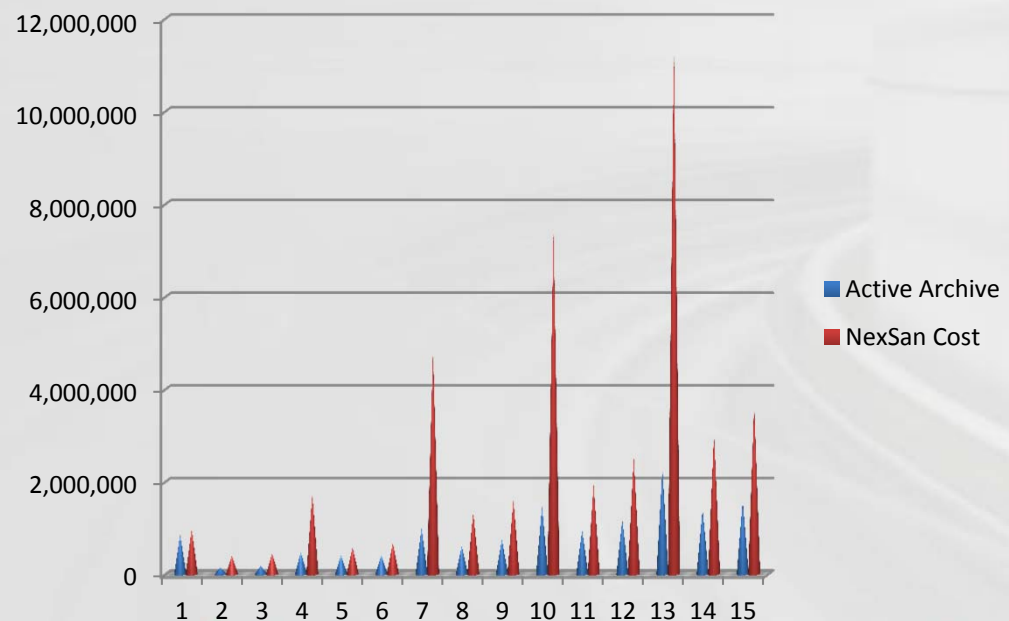


Closest Alternative to Active Archive to Tape?

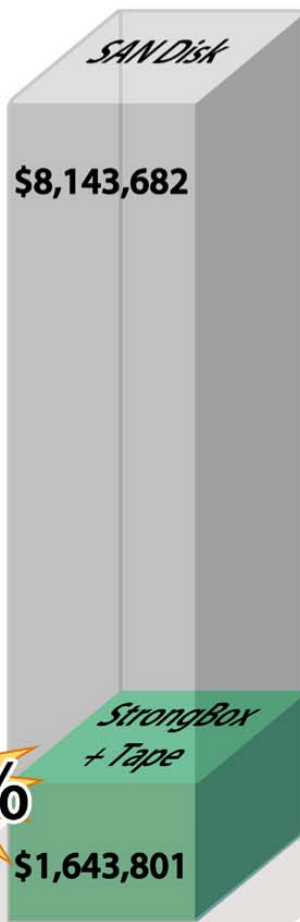
Low Cost Disk

How do you compare? Cost Over Time

15 Year - Comparative Solution	
Active Archive	NexSan Cost
\$13,597,311	\$42,389,467



Online Access to ALL Your Data.



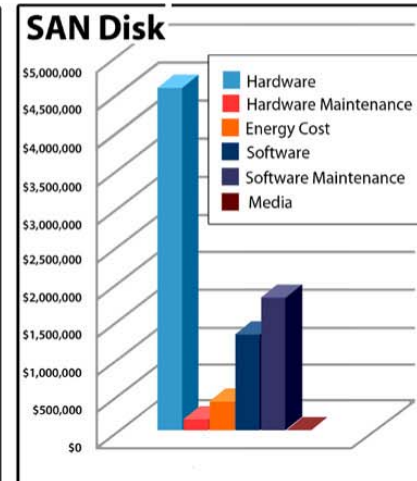
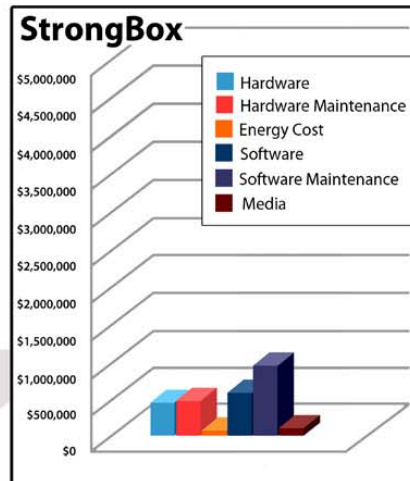
STRONGBOX

TOTAL: \$1,643,801

VS.

DISK

TOTAL: \$8,143,682



COST OF 1 PETABYTE | 2 COPIES | 10 YEARS

ACTIVE Archive

Online Access to ALL Your Data.

Hard Error Rates

- For disk: Number of bits read before the failure of a sector
- For tape: Number of bits read before the failure a bit

Device	Hard error rate in bits	PB equivalent
SATA consumer	1.00E+14	.11
SATA Enterprise	1.00E+15	1.11
Enterprise SAS/FC	1.00E+16	11.10
LTO	1.00E+17	111.02
Oracle T10000	1.00E+19	1,1102.22
IBM TS1140 Technology	1.00E+19	1,1102.22

Making Sense of the Numbers: Time to Error

What difference does an order of magnitude make? Factor the hard error rate and assume 100 devices at maximum native transfer rate for each medium:

- Consumer SATA: **2.3 Hours**
- Enterprise SATA: **21.6 Hours**
- Enterprise SAS/FC: **6.7 Days**
- LTO-5: **96.2 Days**
- IBM TS1140 / Oracle T10000: **≈15 Years**

Software Capability with “Tape as NAS”

- ✓ File system access to entire tape system
- ✓ Metadata search
- ✓ Data movement
- ✓ Data migration, compaction
- ✓ Data integrity verification
- ✓ Backup
- ✓ Encryption
- ✓ Compliance
- ✓ Power and OpEx savings

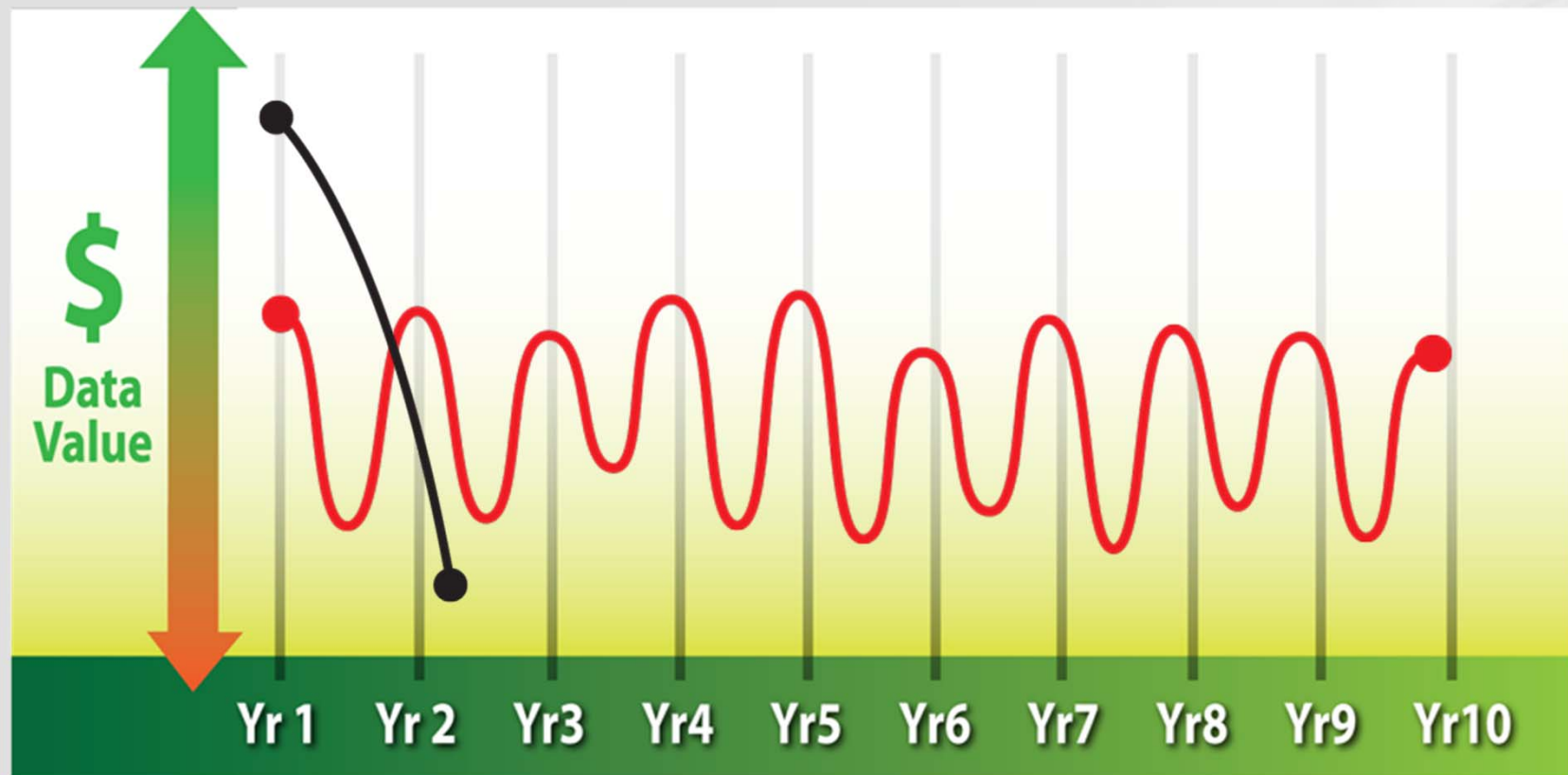
2011 INSIC Tape Applications System Report

“Tape has been shifting from its historical role of serving as a medium dedicated primarily to short-term backup, to a medium that addresses a much broader set of data storage goals, including:

- active archive (the most promising segment of market growth),
- regulatory compliance (approximately 20% - 25% of all business data created must be retained to meet compliance requirements for a specified and often lengthy period), and
- disaster recovery, which continues in its traditional requirements as a significant use of tape.”

Leveraging Tape as NAS for Active Archive of Data

The Value of Data Over Time



Backup Data ● — ●

Archive Data ● — ●

ACTIVE Archive

Online Access to ALL Your Data.

Performance and Capacity Needs Outgrowing Budget Increases

- Problem: I can't hit my backup window
 - Historical Norm: Buy more hardware
 - New Solution:
 - Actively archive unstructured file data
 - Use your hardware more efficiently with 70-80% less data to back up
- Problem: My users aren't able to access transactional data fast enough:
 - Historical Norm: Buy more SSD or FC Disk
 - New Solution: Actively create metadata upon creation and migrate less-critical data off of SSD or FC Disk to NAS tape
- Problem: My users are out of storage space
 - Historical Norm: Buy more "Primary Storage" aka NAS/SAN disk
 - New Solution: Utilize NAS tape to allow user data to be migrated to less expensive storage instead of restricting usage



Online Access to ALL Your Data.

Application Options for Active Archive “NAS Interface” to Tape

1. Crossroads StrongBox
2. FileTek StorHouse
3. SGL FlashNet
4. QStar Archive Manager
5. Atempo TimeNavigator
6. IBM HPSS
7. Seven10 EAS
8. SAM-FS / QFS
9. Quantum StorNext
10. Tape Library integrated LTFS options



Online Access to ALL Your Data.

Active Archive Applications: How to Choose?

Online Access to ALL Your Data.

Two Primary Active Archive Principles

- **There are exceptions to every rule**
 - “Agentless except for.....”
- **There is no black or white**
 - “There is only gray” and lot’s of overlap between technologies



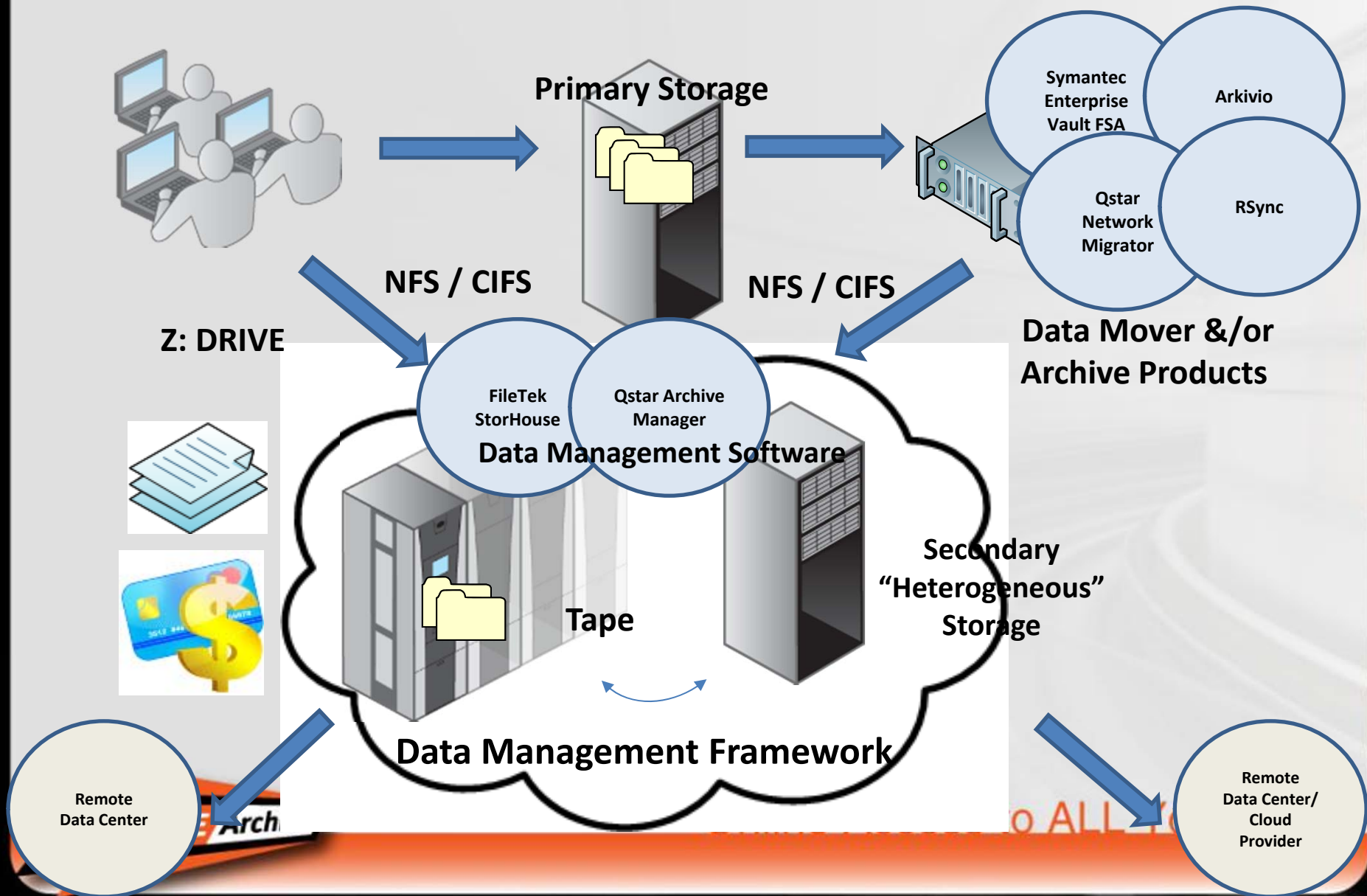
Online Access to ALL Your Data.

Let's Review

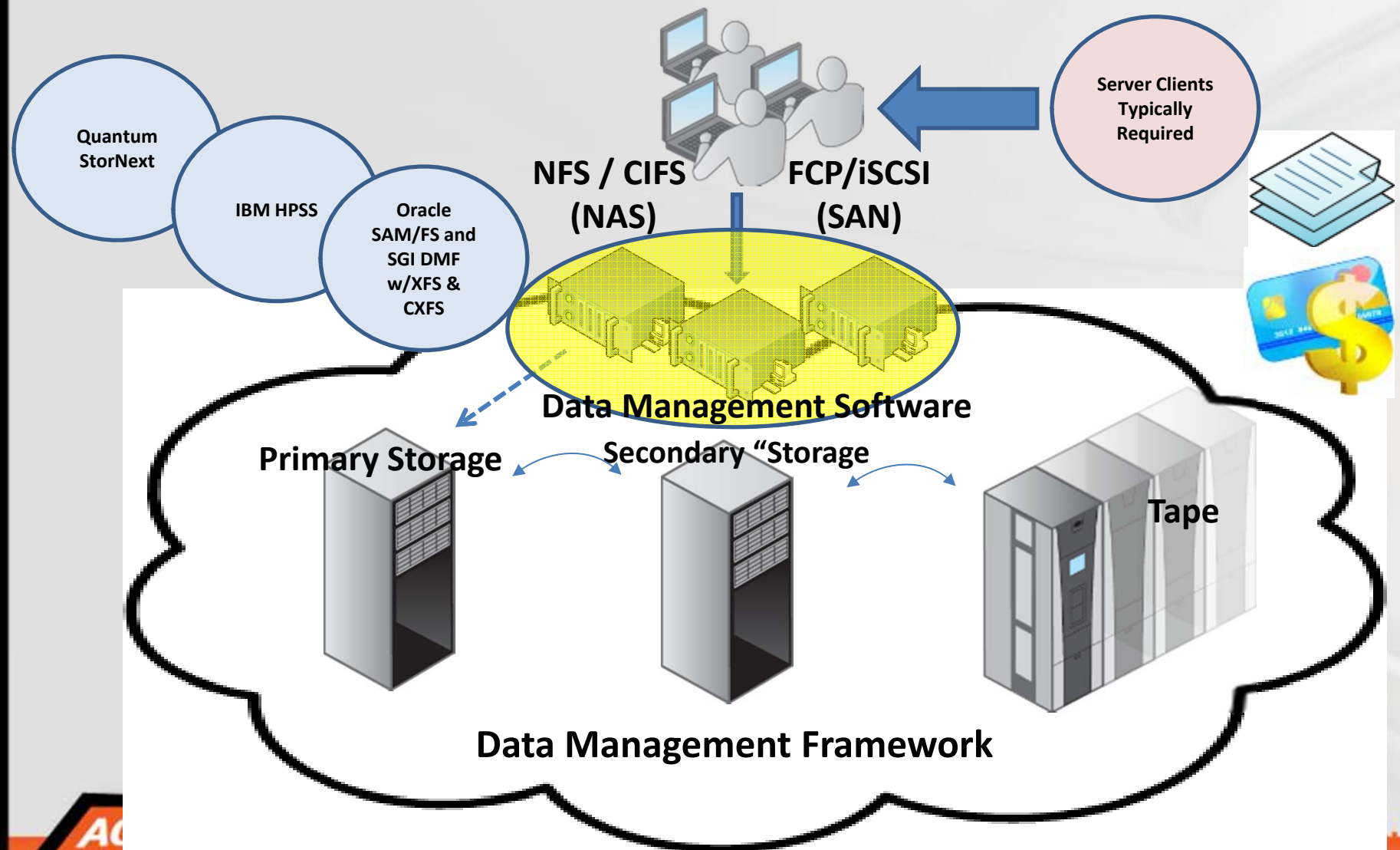
Out of Band vs. In Band Architecture

Online Access to ALL Your Data.

Active Archive Out of Band Conceptual Design



Active Archive In Band Conceptual Design



Data Storage Initiatives

- **“Big Data” Storage:** *Focus is on Performance 1st – Accessibility 2nd*
 - Data Collection
 - Data Processing
 - Data Storage
 - Data Movement
 - Data Sharing, Manipulation, Mining and Analytics
 - *Aligns with “Workflow and Application Specific” Active Archive Technologies*
- **“Just Long Term Storage”:** *Focus on Scalability 1st – Accessibility 2nd*
 - Low Cost Repository
 - Multi-Purpose
 - Regulatory or Business Driver Requirements
 - Little to NO manipulation of the data required
 - Data must be accessible if ever needed
 - Metadata is helpful for search
 - *Aligns with “General Purpose Active Archive Technologies”*

Technology Categories

- **General Purpose - Agentless**
 - **Out of Band Approach**
 - Crossroads Strongbox
 - Filetek StorHouse
 - QStar Archive Manager
 - Seven10 EAS
- **General Purpose – Agents/Clients Required**
 - **Out of Band Approach**
 - Commvault Simpana
 - ASG Atempo ADA
- **Work Flow and Application Specific - Agents/Clients Required**
 - **In Band Approach**
 - SGI DMF
 - Quantum StorNext
 - IBM HPSS



Online Access to ALL Your Data.

Active Archive Check Box Differentiators

- ☐ Unstructured
- ☐ Structured
- ☐ Out of Band
- ☐ In Band
- ☐ Agents Required
- ☐ Agentless
- ☐ Data Mover Included
- ☐ Tape Supported
- ☐ LTFS Supported
- ☐ Copy Management



Online Access to ALL Your Data.

Active Archive Check Box Differentiators

- ☐ Version Management
- ☐ Varied Retention Across Tiers
- ☐ Spinning Disk Storage Tiering
- ☐ Multi-Site Support with Replication
- ☐ Cloud Integration
- ☐ Single Name Space
- ☐ H/A Strategy
- ☐ Multi-Tenancy
- ☐ Data Profiling
- ☐ Content Classification



Online Access to ALL Your Data.

Active Archive Check Box Differentiators

- ☐ Custom Metadata
- ☐ Search
- ☐ Legal Hold
- ☐ Single Instancing
- ☐ Deduplication
- ☐ Compression
- ☐ Auto Data Deletion (tape consolidation?)
- ☐ Data Shredding
- ☐ Scale Out Architecture



Online Access to ALL Your Data.

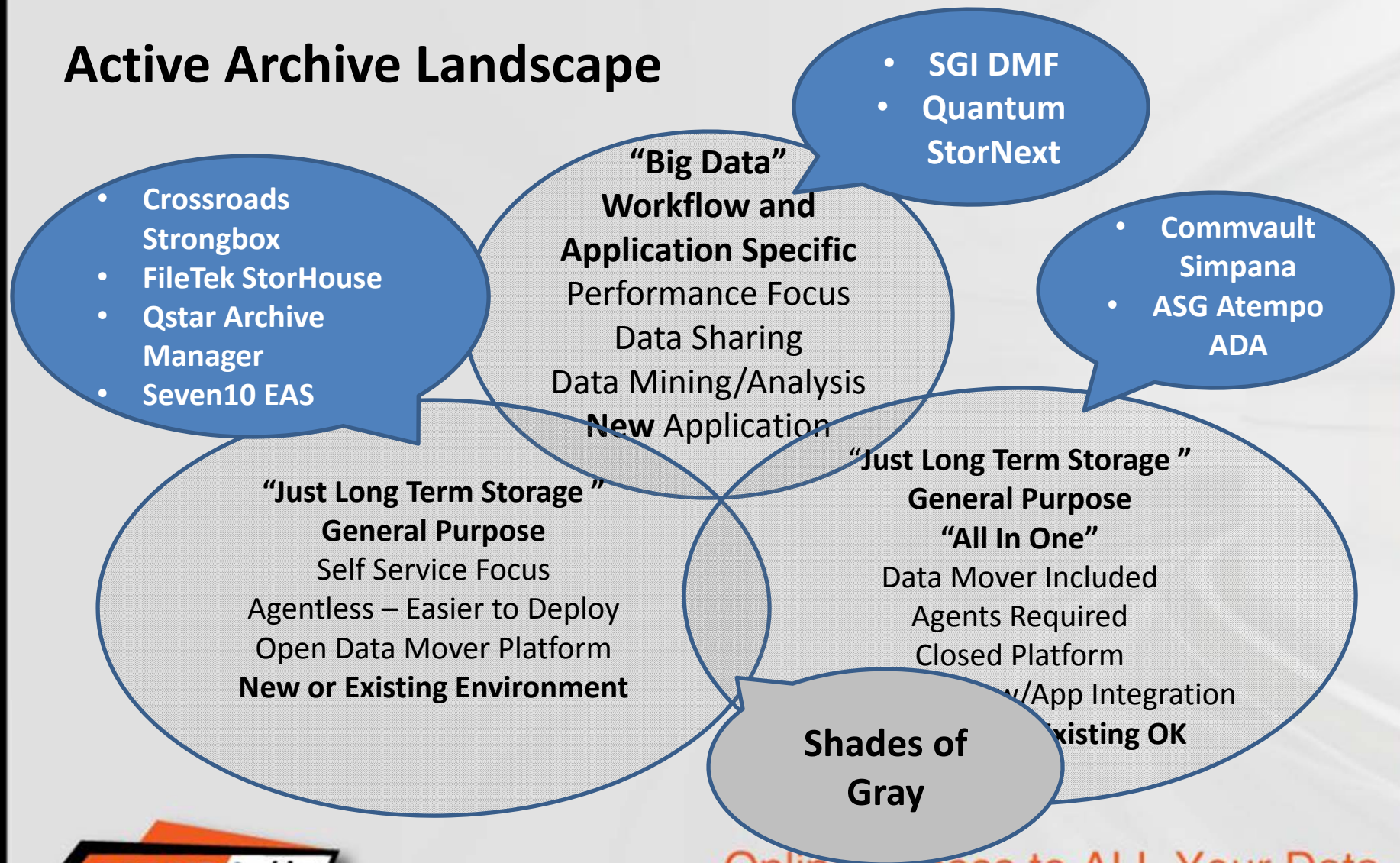
Active Archive Check Box Differentiators

- ☐ Digital Fingerprint/Self Healing
- ☐ Catalog Rebuild Capability
- ☐ Partial File Retrieval
- ☐ Reporting Capabilities
- ☐ 3rd Party Application Integration



Online Access to ALL Your Data.

Active Archive Landscape



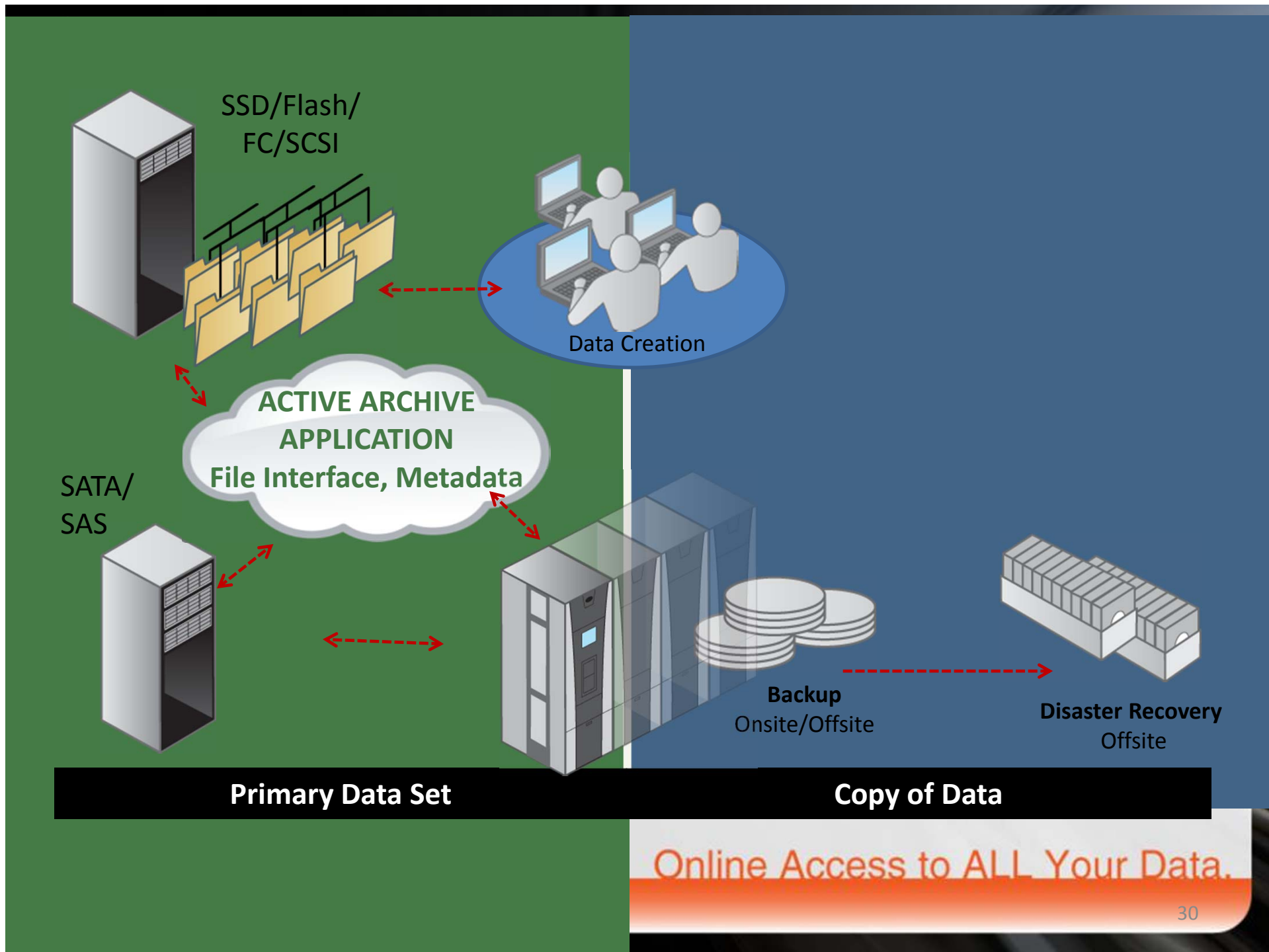
ACTIVE Archive

Online Access to ALL Your Data.

Approach to Selecting the Right Technology

- **Identify the goals of the customer:**

- What type of data will be archived?
- Is this a “long term **“general purpose”** storage repository”?
- Is this for a **“specific application or work flow”**?
- Is this for data mining and analytics?
- Is this for a specific business unit or department?
- Is it a new or legacy environment?



Questions?

mollyr@spectrallogic.com

001.720.940.5736 (Cell)

@mollyrector (Twitter)

www.activearchive.com

Next Active Archive Certification training
is Oct. 9/10 in Maryland

Online Access to ALL Your Data.