Rethinking Archiving: Exploring the path to improved IT efficiency and maximizing value of archiving solution investments

Marshall Amaldas & Brad Nisbet, IDC

Authors: Marshall Amaldas & Brad Nisbet, IDC
The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.

Member companies and individual members may use this material in presentations and literature under the following conditions:

- Any slide or slides used must be reproduced in their entirety without modification
- The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.

Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.

The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
Abstract

Archive has traditionally been viewed as a low-priority process of moving historical information to cold storage media. An increased burden on IT from explosive growth in digital information and growing regulatory/compliance mandates have increased the need to revisit archive strategies.

This session will discuss:

- Sources and drivers of digital information growth
- Information management challenges, symptoms and root causes
- Why, What & How framework for narrowing down archive solutions strategies
- How cloud-based archiving can help meet your information management requirements
- Key considerations when choosing a cloud service provider

© 2012 Storage Networking Industry Association. All Rights Reserved.
Rethinking Archiving: Exploring the Path to Improved IT Efficiency and Maximizing Value of Archiving Solutions Investments

© 2012 Storage Networking Industry Association. All Rights Reserved.
Amount of data created is **doubling** every two years

Digital universe to grow from 1.8 ZB in 2011 to 7.9 ZB in 2015
Data is being retained much longer

Extended Retention Periods

- <2 years: 10% structured, 22% unstructured
- 2 to 7 years: 39% structured, 21% unstructured
- 7 or more years: 50% structured, 27% unstructured

*IDC 2011 Quick Poll Survey

7 or more years:
- 50% firms retain structured data
- 27% firms retain unstructured data
Factors Driving Extended Data Retention

Traditional Factors

• Government Regulation and Compliance
• Industry Guidelines and Best Practices
• Legal Hold - Support of eDiscovery

Important Emerging Use Case

• Information reuse
  • Access to historical information
  • Archiving as a business/service
  • Business analytics

Check out “Managing Big Data Hands-On Lab”
Heavy unstructured contents consumes storage resources

Data Retained > 7 years

More unstructured data is retained longer

Lack of an expiration date and inability to enforce if defined

- 26% firms retain image content indefinitely
- 1 out of 5 firms reported retaining audio and video files indefinitely

*IDC 2011 Quick Poll Survey
Emerging & Future Info Mgmt Challenges to lookout for – SharePoint & Big Data

72% of organizations use SharePoint
Source: IDC Study, 2010

23% Storage admins identify proliferation of SharePoint instances as a top file based storage management issue
Source: IDC #228824

Big Data to grow at 62.6% CAGR from 2,624 EB in 2011 to 29,796 EB 2015
Symptoms to look out for

Top FBS Challenges

- Dealing with growing number of file servers and/or NAS filers in datacenters
- Controlling growth and protecting data in file servers at remote offices
- Controlling proliferating SharePoint instances
- Proliferating storage growth associated with rich content applications (video, imaging)
- Managing resource utilization and load balancing performance across multiple systems
- Implementing consistent disaster recovery strategies for file-based data

Storage and IT Infrastructure Cost
Data Protection & Recovery Problems
Performance Penalty
Inefficient Search and expensive eDiscovery

Source: IDC’s 2011 Trends in File-Based Storage Survey
Some Root Causes

- Retain everything mentality - Even archiving unwanted information is wasteful
- Using **Backup** for meeting long-term retention needs is common (57% according to IDC 2011 Study)
- Many firms do not use any specific solution and the data is **archived in place**
- Longest retention period often gets applied
- Disparate IM tools for managing different data complicates Information management
Choosing your archiving solution

Why this needs to be archived?
- Storage Efficiency
- Performance Optimization
- Compliance
- Legal
- Reuse

What is the content that needs to be archived?
- Email, Instant Messages
- File System
- SharePoint
- Transactional Data/Databases
- MGD
What else?

 Alternate & Complementary Use Cases

- Internal Discovery
- Email Continuity and Security
- Work flow optimization
- Integration with other applications

 Information and Risk management

- Policy alignment with DLP and Email monitoring solutions
- Support for future information streams
- Growing Emphasis on Privacy (e.g. right to be forgotten)
- Enterprise application social media retention and access

Benefits of raising “What Else?” question

- Turn a mandatory cost center process useful to profit center activities.
- Justification from multiple stakeholders and better ROI.
How? – Some Alternatives

- **Archive In Place: Do Nothing Strategy**
  - Extremely wasteful of storage assets
  - No central control and manual process
  - Approximately 1 in 5 organizations do this

- **Backup as Archive: Path Of Least Resistance**
  - Conflicting priorities
  - Loss of context (especially in tape)
  - Inefficient search & retrieval
  - Approximately 1 in 2 organizations do this

- **Application specific: Built-in archiving**
  - Overhead of siloed information mgmt
  - Fall short of advanced compliance needs
  - No single index: redundant search
  - Approximately 1 in 10 organizations do this

---

**When appropriate?**
- Digital data growth is minimal
- Cost of implementing and maintaining archive is greater
- Data retained for compliance, not for litigation readiness.
- Metadata based indexing satisfies access & retrieval criteria
- Only a subset of digital data needs to be retained and accessed.
- Regulatory, discovery and review requirements minimal
- Archiving purely for efficiency reasons
Dedicated Archiving Solutions: Key Capabilities to look for in Archival Storage Systems and Software

- **Scalability:** Lower upfront investment and incremental capacity expansion without expensive and time consuming data migration. In software solutions explore capacity based licensing/pay as you go models.

- **Performance:** flexibility to provide throughput for less predictable archival retrievals without conflict.

- **Efficiency:** take full advantage of proven efficiency technologies such as deduplication and policy-based tiering for both. Differentiate between active archiving and long-term retention and identify the role of tape in archiving.

- **Compliance readiness:** retention enforcement capability for data and tools to aid IT administrators make informed policy decisions.

- **Common data management** foundation for backup and archive that provides flexibility in the placement of data

- **Don’t Forget the Cloud…**
Cloud Information Management

Cloud Data Protection
Backup

Cloud Business Continuance
Replication
Disaster Recovery

Cloud Archive
Compliance
Analytics

Cloud File Services
File-Sharing
Synchronization
Collaboration

Target Market

Personal
SOHO
Enterprise

Service Location/Architecture

Public Cloud (External)
Hybrid
Private Cloud (Internal)
Growing requirement within organizations to consolidate “secondary” forms of storage
- Backup & Archive on converging path
- “Hybrid” environments a long-term transition

Verticals Expertise
- Regulation will drive expertise in industry-specific compliance
- Evolving business initiatives (analytics/Big Data) will drive integration with industry-specific applications
The Transition to Cloud Archive

Transition to Cloud will be evolution:
- Hybrid solutions that span internal & external back-end storage will evolve to address:
  - Security & Performance concerns
  - Integrated operations & risk associated with managing multiple locations
  - Search extensibility
  - Contracts – SLAs - Liabilities
Cloud-based Archive expands from email to content

Source: IDC, 2011
Q. How significant are the following challenges/obstacles in your planned or continued use of a Public Cloud-based Archive Service?

- Security
- Reliability of service
- Pricing
- Response times
- Longevity of service provider
- Performance on ingestion
- Breadth of content types
- Address regulatory/compliance
- Service Provider 'lock-in'
- eDiscovery expertise

**Control**
- Data security
- Data location
- Data retrieval

**Reliability**
- Performance
- Availability

**Certainty**
- Fragmented
- Vendor lock in
- Where to start?

Closing comments

- Invest in technologies that help **maximize value of existing investments**

- Archiving is no more just an insurance – recognize the **reuse and analytics** value of historical information

- **Not all data is the same** and hence need to be understood and treated differently

- Get used to the idea of **disposing information** – Conventional “keep it all” mentality is no longer feasible

- Don’t fear the **transition to the cloud** – suppliers recognize the opportunity to enable a long transition period
Many thanks to the following individuals for their contributions to this tutorial.
- SNIA Education Committee

Joseph White
Nancy Clay
Laura DuBois

Send any questions or comments on this presentation to SNIA: tracktutorials@snia.org