The Changing Role of Data Protection In a Virtual World

Gene Nagle, BridgeSTOR
Thomas Rivera, Hitachi Data Systems

SNIA Legal Notice

- 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.

About the SNIA DPCO Committee

- This tutorial has been developed, reviewed and approved by members of the Data Protection and Capacity Optimization (DPCO) Committee which any SNIA member can join for free
- The mission of the DPCO is to foster the growth and success of the market for data protection and capacity optimization technologies
  - Online DPCO Knowledge Base: www.snia.org/dpco/knowledge
  - Online Product Selection Guide: http://sniadataprotectionguide.org
- 2013 goals include educating the vendor and user communities, market outreach, and advocacy and support of any technical work associated with data protection and capacity optimization

Check out these SNIA Tutorials:

- **Understanding Data Deduplication**
- **Advanced Data Reduction Concepts**
- **Deduplication's Role in Disaster Recovery**

Hands-On LAB SNW

The ABCs of Big Data - Analytics and Bandwidth for Your Content

Abstract

The Changing Role of Data Protection

- This session examines the many changes occurring in Information Technology today that affect data protection methodologies as it seeks to answer questions important to IT managers, such as:
- Does virtualization of servers and storage help or hinder Data Protection?
- How do snapshots fit in with other aspects of Data Protection?
- Does traditional centralized, batch-oriented backup still have a place in overall Data Protection design?
- Should protecting structured & unstructured data be treated differently?
- What are the best practices for integrating cloud storage into an existing Data Protection system?

Some Definitions

- **Data Protection:** Assurance that data is not corrupted, is accessible for authorized purposes only, and is in compliance with applicable requirements.
- **Continuous Data Protection (CDP):** A class of mechanisms that continuously capture or track data modifications enabling recovery to previous points in time.
- **Big Data**: A characterization of datasets that are too large to be efficiently processed in their entirety by the most powerful standard computational platforms available.

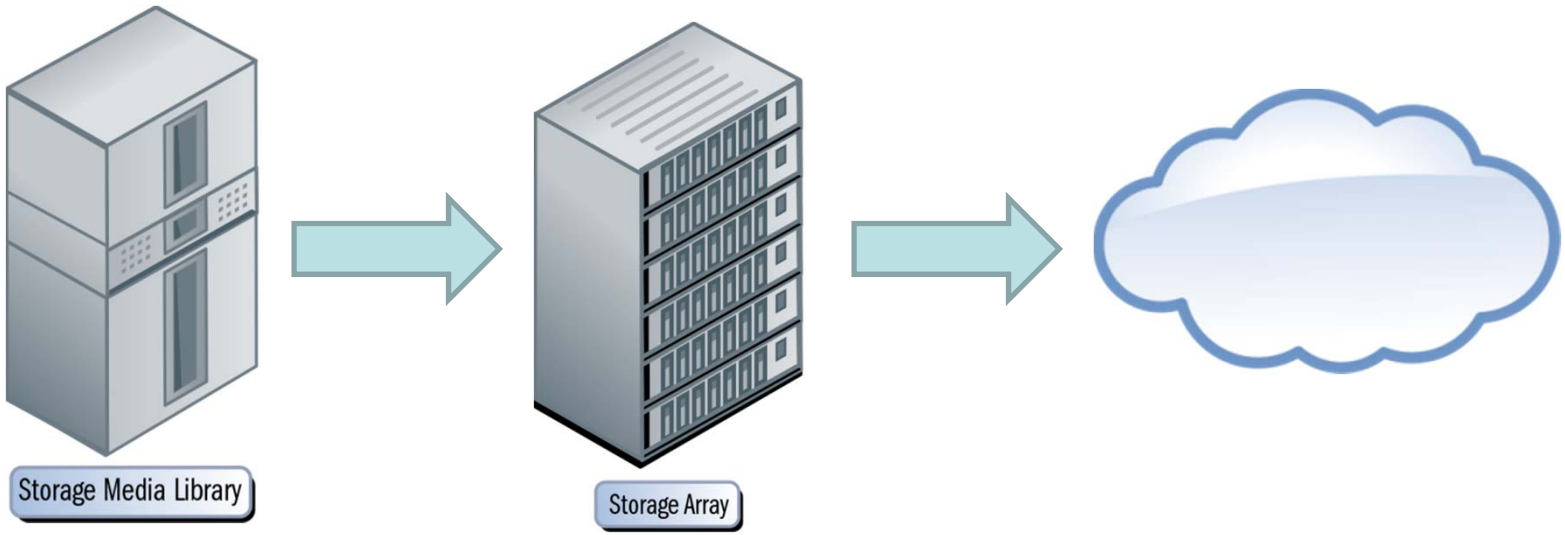
Some Definitions (Cont.)

- **Restore Point Objective (RPO):** The maximum acceptable time period prior to a failure or disaster during which changes to data may be lost as a consequence of recovery.

- **Restore Time Objective (RTO):** The maximum acceptable time period required to bring one or more applications and associated data back from an outage to a correct operational state.

Decisions, Decisions….

"I moved my data protection from tape to disk + tape, now do I just need to move to disk + cloud?"



Sorry, it's more complex than that….

The Forces of Change

- Virtualized Infrastructure
- Tighter Data Protection Requirements
  - Application-based RPO and RTO
  - New Service Level Agreements (SLAs)
  - Archiving & long-term retention requirements
- Big Data
  - Many traditional backup tools inappropriate
- Cloud Computing & Cloud Storage
  - Cloud storage often includes data protection
- Primary Storage with Built-In Data Protection
  - Snapshots & versioning
  - Self healing

The Virtual Environment

- Traditional backup should be “VM-aware”
  - Managing agents on every VM is costly
- Hypervisors can provide snapshot capability
  - Advisable for some applications, not for others
- Media servers can be VMs
- Virtualized storage can provide thin provisioning
  - Block-Level incremental storage
- Virtualized storage can include cloud-based storage
  - True transparency often requires local cache

More Use of Snapshots

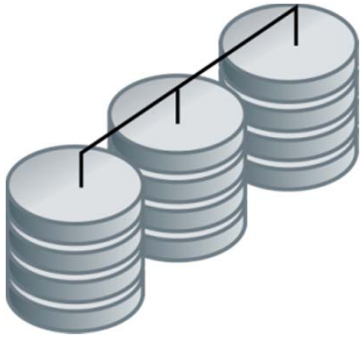
- Major improvement in RPO
  - Backup to disk improved RTO, but not RPO
- Can have side effects on applications
  - Performance
  - Disk space requirements
- Can present some challenges
  - Snapshots provide only logical data protection, not physical
  - Open file & consistency issues - if not application aware
  - Application aware agents can add cost, complexity, management

Snapshots in the Data Protection Mix

- Used with traditional backup
  - Eliminates need for backup window
- Used with Near-CDP
  - Very common practice now for email, databases
- Used with Replication for Disaster Recovery
- Virtualization brings snapshot decisions
  - File system level vs. hypervisor level
  - Hypervisor level often adequate for unstructured data
  - Use aplication-aware snapshots for structured data

More Need for High Availability

- HA drives changes in Data Protection
- Continuous Availability vs. Highly Protected
  - RPO & RTO = zero
  - RAID 6, RAID cluster
  - RPO & RTO very low
  - Frequent snapshots



- Traditional RAID techniques are becoming less used
  - Large HDDs cause long rebuild times
- HA can be merged with Disaster Recovery
  - Requires synchronous replication to DR site

More use of Public Cloud

- Huge CAPEX benefits
  - Backup as part of the deal –
    - Software-as-a-Service (SAAS)
    - Cloud storage with HA & data protection
  - Virtual backup appliance in the cloud
  - Significant savings for Disaster Recovery

- Latency often acceptable for secondary storage
  - Sometimes works for primary, unstructured data
- When cloud is an alternative - Compare existing storage costs to cloud offering(s): $/GB/month

Being Careful about Cloud

- Hybrid systems with local caching or local backup targets becoming common
  - Take advantage of cloud economies and flexibility

    But
  - Adds another layer of management

- HA may require difficult cross-cloud mirroring
- DR may require restores to cloud PAAS
- Service provider risks
  - Be careful about SLAs
  - Be mindful about putting “all eggs in one basket”

More use of Archiving

- Regulatory requirements
- Potential for reducing primary data and data protection loads
- Frequency of restore points determine how it's done
  - Users prefer every version of a file – but for how long?
  - Block-level deduplication reduces versioning redundancy
- Access time requirements also determine how it's done
  - Users tend to over-estimate their "need for speed"
  - Deduplicated disk vs. tape
    - Tape still a good choice for long term retention

Management Challenges

- Moving targets
  - Transitioning to virtual infrastructure
    - Causes “shell game” for data protection admins
- Consolidation of storage functions
  - Primary with HA + snapshots & versioning -- traditional backup still needed?
    - Can work if snapshots on a separate device
- More decisions on where to perform DP
  - At source, local media server, hypervisor, or at cloud
- Changing responsibilities in governance
  - Primary + backup in one cloud service or one device – who has responsibility?

The Bottom Line: $$$

- Application-Based tiers of data protection
  - May add efficiency, but -
  - Makes a single pane of glass very complex
- Reduce data protection OPEX and CAPEX by
  - Use automated snapshots with separate local disk storage for immediate restore in place of daily backup
    - May or may not vary the frequency based on applications
    - Use hypervisor when possible
  - Capacity optimization, especially for secondary storage
  - VMs for media servers, gateways
  - Use public cloud for capacity optimized DR
  - Use tape for archiving, long-term retention

Check out these other SNIA Tutorials:

Protecting Data in the Big Data World

Advanced Data Reduction Concepts

Hands-On LAB COMPUTERWORLD SNIA SNW

Data Protection, Business Continuity and Disaster Recovery – New Technologies

Attribution & Feedback

The SNIA Education Committee thanks the following individuals for their contributions to this Tutorial.

Authorship History

Data Protection and Capacity Optimization Committee, March, 2013

Gene Nagle
Thomas Rivera
Mike Dutch
Kevin Dudak
David Hill
Tom Sas
Gideon Senderov
Alan Yoder

Please send any questions or comments regarding this SNIA Tutorial to ***tracktutorials@snia.org***