# Trends in Data Protection and Restoration Technologies

Jason Iehl / NetApp

Author:
SNIA - Data Protection & Capacity Optimization (DPCO) Committee

# 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
- 2016 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 at:** www.snia.org/education/tutorials

- **Intro to Data Protection: Backup to Tape, Disk & Beyond**
- **Advanced Data Reduction Concepts**
- **Data Protection in the Big Data World**
- **Data Protection vs Privacy: How EU legislation is changing**
- **Intro to Encryption**

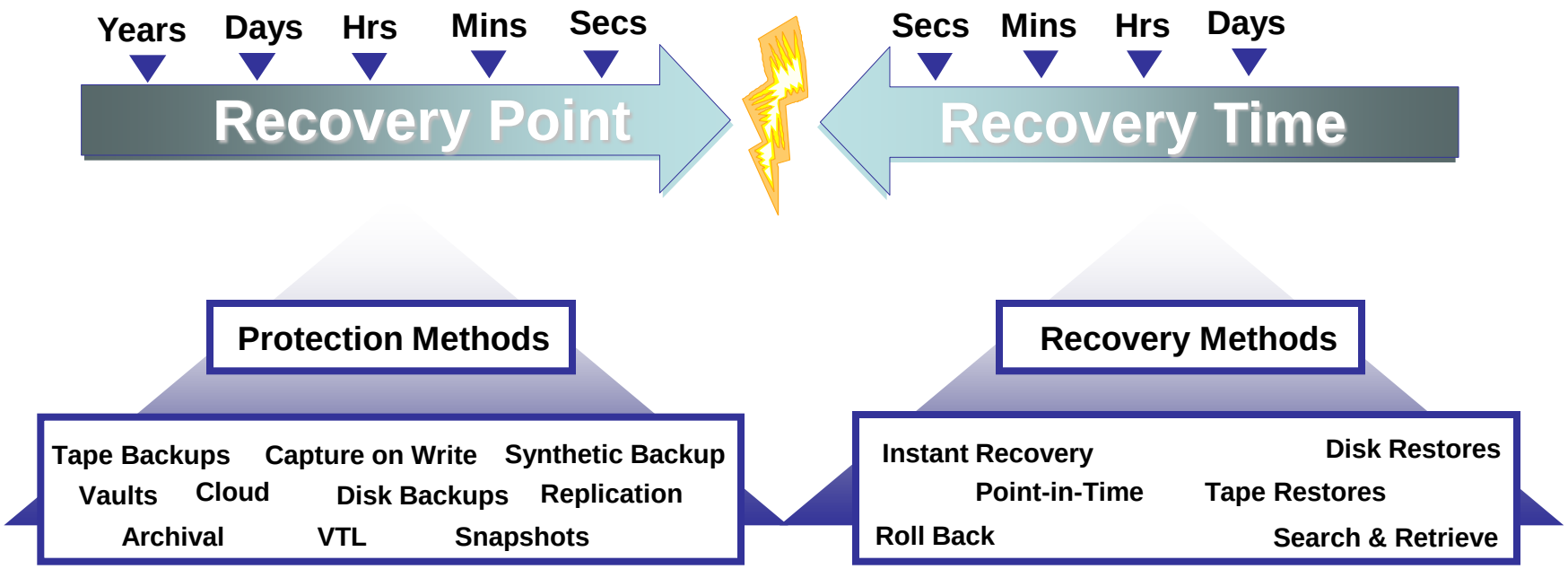
# Abstract

- Many disk technologies, both old and new, are being used to augment tried and true backup and data protection methodologies to deliver better information and application restoration performance; These technologies work in parallel with the existing backup paradigm

- This session will discuss many of these technologies in detail; Important considerations of data protection include performance, scale, regulatory compliance, recovery objectives and cost; Technologies include contemporary backup, disk based backups, snapshots, continuous data protection and capacity optimized storage, as well as cloud services.

- This tutorial will cover how these technologies interoperate, as well as best practices recommendations for deployment in today's heterogeneous data centers & you will:
  - Understand legacy & newer storage technologies that provide advanced data protection
  - Compare and contrast advanced data protection alternatives
  - Gain insights into emerging data protection technologies

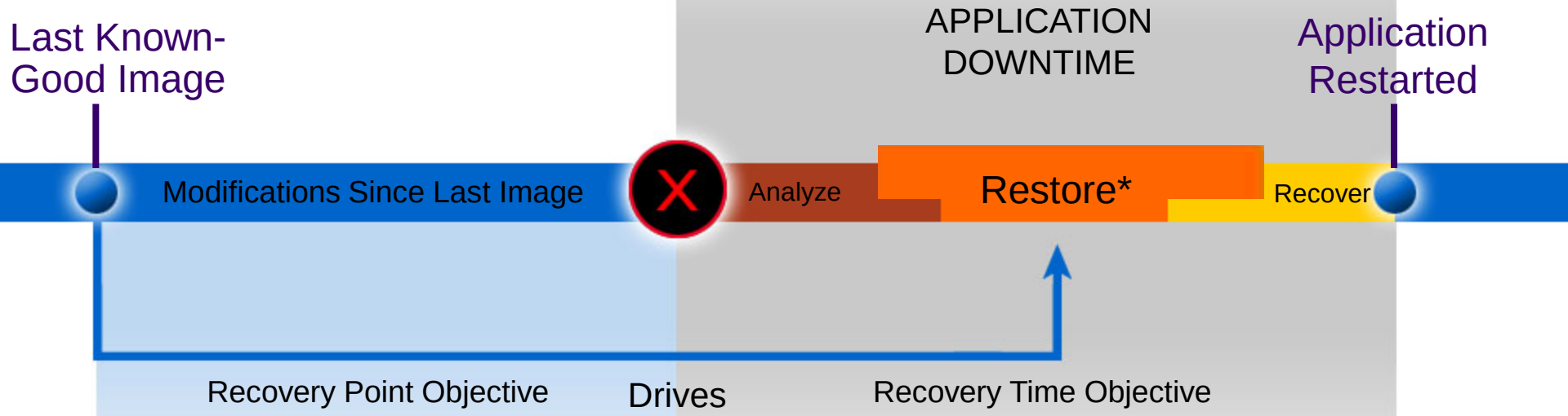
# Data Protection Based on Recovery




## Enabling Technologies

| | | | | | |
|---|---|---|---|---|---|
| Tape & Automation | Cloud | Deduplication | Replication | Snapshots | Backup to Disk |

# Application Recovery




*10TB restore from backup ~= 4 hours from disk, 12.5 hours from tape

Recovery from snapshots reduces "restore" but older point in time snap causes increase in recovery time

# Application Consistency

- What?
  - When an application is running during the "copy" process
  - Various techniques are available to ensure data consistency
- Why?
  - Much like the "open files" issue when backing up a file system that is in use, applications (like databases, messaging systems, etc) allow for different approaches to capturing a holistic picture of the applications data during a copy process, such as a snapshot, a split-mirror, or continuous data protection (CDP)
- Considerations
  - It is important to understand the consistency semantics of your application so that your data protection copies are recoverable

# Consistency – Offline/Cold Backup

- Shutdown the application / database
  - Guarantees  application consistency
  - All cache data copied to disk; all transactions closed
  - Optional: database consistency check
- Backup to another disk / tape
  - OR create a snapshot
- Optional: move / delete the transaction logs
  - Frees up disk space
  - Enables incremental backup based on transaction logs
- Start the application
- Optional: backup the snapshot to another device or location
- Recycle the snapshot
  - Keep the last # snapshots (snapshot rotation)

# Crash Consistent

- Create a snapshot while the application runs
  - Restore quality of application consistency after a system crash
  - Most applications / databases can survive system crashes
    - But some don't and some not always
  - Recovery cannot be guaranteed
  - Note: Some applications may force a database consistency check
- Use cases
  - 7 x 24 operations -> no backup window
  - Virtual Machine backup without agent or service API
  - Application lacks online backup mode feature
  - No resources for transaction logging during backup
  - Snapshots enable more points in time
  - Combine with consistent backups

# Consistency - Online

- Database(s) are in “backup mode“ during backup
  - Data files don‘t change while in backup mode
  - Changes during backup happen in the cache and go into logs
  - After backup all changes are applied to the data files
  - Optional: backup of the transaction logs & delete logs afterwards
  - Optional: ongoing log file backup after database backup -> CDP
- Consistent search index
  - All databases need to go into the backup mode
  - Must ensure consistency across all nodes
- Use cases
  - 7 x 24 operations -> no backup window
  - Guaranteed & fully supported consistent recovery

# Data Protection & Data Management

## Data Management

- Information classification
- Information valuation ($$$)
- Information lifecycle management
- Data indexing

## Tiered Storage

- Primary
- Secondary
- Archive
- Backup
- Cloud

## Data Protection

- Disk-Assisted and Disk-based protection methods
- Array and storage network based data protection
- Object based Archival
- Snapshots and replication
- Tape based data protection
- Backup to Virtual Tape
- Backup to Disk
- Backup of Virtual Machines

# SLAs: 85/95/5 Retention & Recovery Model

- 85% of restores are local
  - These are the faster restores
    - Leverage snapshots where available
    - End user restores, project restores
  - Single file, email, table space, complete database leveraging logs
- Up to 95% from D2D
  - Usually D2D with Replication
  - Can be onsite or offsite depending on RPO/RTO requirements
- Everything else or 5% from Tape and/or Cloud
  - Example: eDiscovery, system recovery
  - In most cases restores to different/new set of disks
- Tape is leveraged for long term and infinite retention
  - Best practice to keep offsite or third location

# Backup of Virtual Machines

- What?
  - Backup virtual machines and hypervisors
  - Avoid resource contention (i.e., Impact of backing up 20 VMs)
  - Address backup storage growth
  - Consider your recovery goals and objectives
- Why?
  - Deliver granularity of “client” backup and restore
  - Improve operations by reducing backup and restore times
  - Eliminate redundant data
  - Reduce complexity
- Considerations
  - VM Clones are not “backups”
  - Native backup tools
  - Integration with hypervisor
  - Capacity, performance, scale

# Virtual Machine Backups: Best Practices

- Consider hypervisor-based backup tools for “simple” environments
  - Available from backup and storage vendors
  - Larger configurations may require more advanced tools
- Consider snapshots, clones and copies
- Isolate data files that don’t need backups
  - Temporary and/or transient data
  - Focus on Users and Applications
- Insist on integration with hypervisor tools and APIs
- Backup the hypervisor AND THEN back up the virtual hosts

# Snapshots

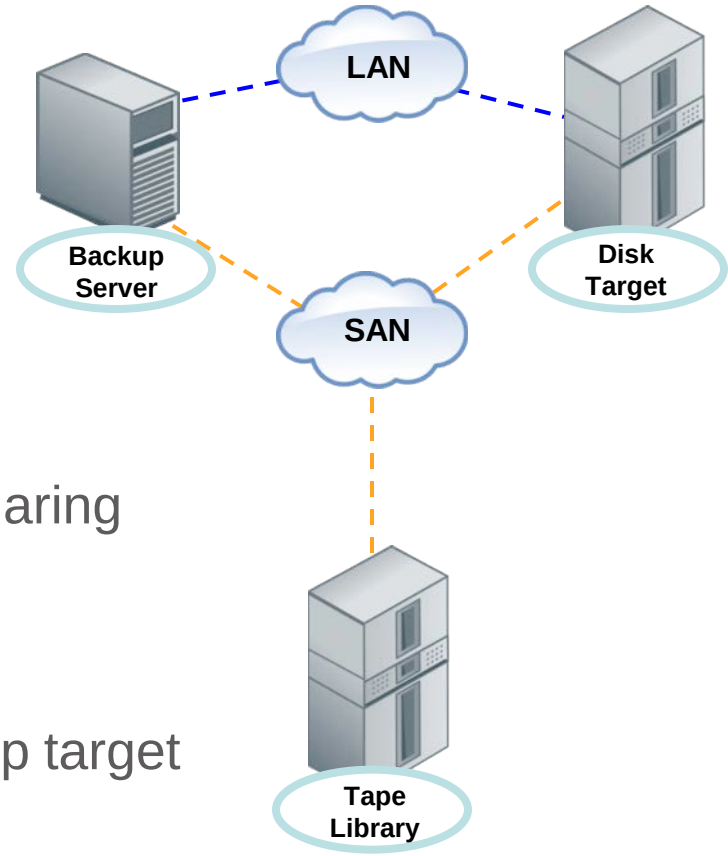
- What?
  - A disk or VM based “instant copy” captures the original data at a specific point in time; Snapshots can be read-only or read-write
  - Also known as Checkpoint, Point-in-Time, Stable Image, Clone
  - Any technology that presents a consistent point-in-time view
- Why?
  - Allows for complete backup or restore
    - With application downtime measured in minutes (or less)
  - Most vendors: Image only = (entire Volume)
  - Backup/Restore of individual files is possible
    - If conventional backup is done from snapshot
    - Or, if file-map is stored with Image backup

# Snapshot Considerations

- Snaps of production storage or VM may impact production
  - Consider performance, storage, virtual-machine impacts
- Retention policy impact
  - Number of copies retained
  - Recovery granularity
  - Meeting off-site protection via distance replication
- On-array, off-array, VM-based alternatives
- Cost trade-offs and information classification
- May combine snapshots with replication

# Backup to Disk (B2D)

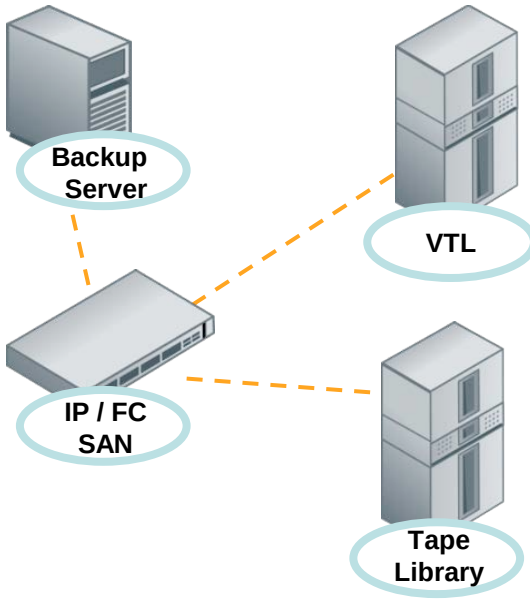
- What
  - Replace tape targets with disks
  - Perform snapshot-based backups
  - Recover from data on-line versus on-truck
- Why?
  - Leverage disk reliability and performance
  - Simplify backup process, reduce device sharing
  - Improved security
- What to watch out for
  - Bottlenecks, which may NOT be the backup target
  - Potential for added complexity and training
  - Advanced features may increase SW costs
  - Backup window issues may still exist
  - Meeting offsite or off-line backup requirements



# Virtual Tape Library (VTL)

## What?

- Virtual Tape Libraries emulate traditional tape
- Easy to deploy and integrate
- Reduce / eliminate tape handling




## Why?

- Improved performance and reliability
- Reduced complexity
- Unlimited tape drives reduce device sharing, improve backup times
- Leverage technologies such as remote replication, deduplication

## Considerations

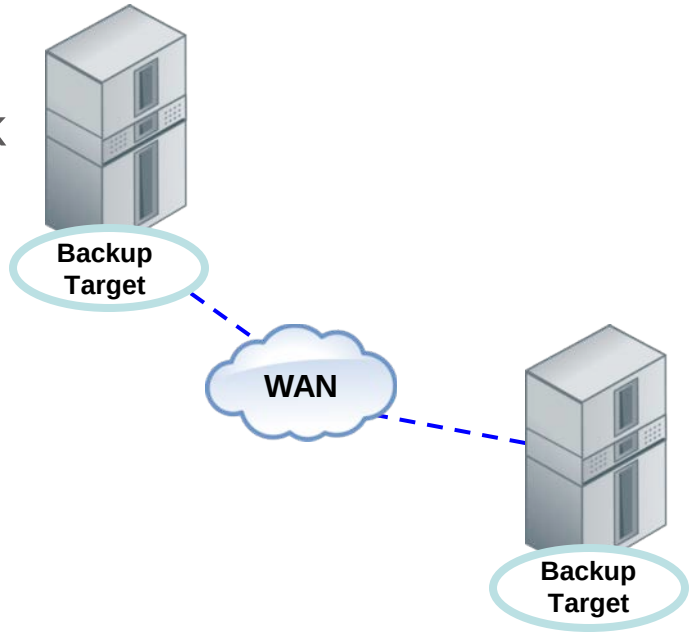
- Avoids mechanical and performance limitations of tape
- Integration with physical tape
- Consider total aggregate speed as well as speed per-drive

# Data Deduplication

- What?
  - The process of examining a dataset or I/O stream at the sub-file level and storing and/or sending only unique data
  - Client-side SW, Target-side HW or SW, can be both client and target
- Why?
  - Reduction in cost per terabyte stored
  - Significant reduction in storage footprint
  - Less network bandwidth required
- Considerations
  - Greater amount of data stored in less physical space
  - Suitable for backup, archive and now more often primary storage
  - Enables lower cost replication for offsite copies
  - Store more data for longer periods
  - Beware 1000:1 dedupe claims – Know your data and use case
  - Tradeoffs (Performance, Encryption, Compression)

# Backup Replication

- What
  - Multiple copies of backup data
  - Copy tape-to-tape, disk-to-tape, disk-to-disk
  - Replication can be network based
- Why
  - Off-site data protection
  - Media consolidation
  - Increased security
  - Leverage the Cloud / Cloud Storage




- Considerations
  - Network bandwidth - Leverage dedupe, network devices, “fat” pipes
  - Recovery process and resources
  - Leveraging the copies (data mining opportunities)
  - Backup versus archival

# Encryption

- What
  - Converting data to an encrypted form so that only authorized users can access to the data using the appropriate decryption key
- Why
  - Sensitive data
  - Confidential data
  - Restricted data
  - Top Secret
  - Etc.
- Considerations
  - Encryption can be accomplished at various points (Application, file system, HBA, Network, Storage Array, Disk Drive, Tape Drive, etc.)
  - Key management is one of the biggest issues with encryption
    - Key Loss, Disclosure of Keys, Key Availability, etc.

# Public Cloud & Private Cloud

- What
  - Backing up data to a cloud service provider
    - Various cloud models available
- Why
  - Less cost than buying infrastructure (in the case of outsourcing)
  - Eliminates costs and complexities of data protection schemes
- Considerations
  - Who owns the data?
  - Security of the data?
  - What is the data availability rating?
  - How will the data be recovered? (e.g., cloud, physical media, WAN)
  - How fast can I recover my data in the event of a disaster?

# Data Protection Summary

- Data growth requires us to plan for tomorrow
  - Investigate data and information management technology
- Information value determines data protection levels
  - Stop protecting employee home movies, last years news
  - Not all data assets are created equal
- Architecture
  - Applications are not all the same
  - Understand your networks, hosts, applications
  - PLAN ahead – Avoid reactionary thinking
- Do your homework
  - SNIA offers seminars, classes, workshops…..

# Additional Resources

- Related tutorials
  - Advanced Data Reduction Concepts
  - Managing Backup and Recovery in Today's Agile, Complex and Heterogeneous Data Centers
  - The Changing Role of Data Protection in a Virtualized World
  - Protecting Data in the Big Data World
  - Retaining Information for 100 Years
- Visit the Data Protection and Capacity Optimization Committee (DPCO) website http://www.snia.org/forums/dpco/
- DPCO – Data Protection Best Practices White Paper

  http://www.snia.org/forums/dpco/
- DPCO – Online Product Selection Guide

  http://sniadataprotectionguide.org/

# Attribution & Feedback

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

## Authorship History

**Original Author: SW Worth / 2001**

**Updates:**

**DPCO Committee, 9/2010**
**DPCO Committee, 2/2012**
**DPCO Committee, 8/2012**
**DPCO Committee, 2/2013**
**DPCO Committee, 8/2013**
**DPCO Committee, 3/2015**
**DPCO Committee, 4/2016**

## Additional Contributors

**David A. Chapa**
**Kevin Dudak**
**Mike Dutch**
**Mike Fishman**
**Larry Freeman**
**David Hill**
**Jason Iehl**
**Tom McNeal**
**Gene Nagle**
**Ronald Pagani**
**Rob Peglar**
**Thomas Rivera**
**Tom Sas**
**Gideon Senderov**
**Paul Talbut**
**SW Worth**

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