

A decorative graphic consisting of multiple parallel, wavy lines in various colors (purple, blue, orange, grey, yellow) that flow from the left side of the slide towards the right, creating a sense of movement and depth.

# **Introduction to Data Protection: Backup to Tape, Disk and Beyond**

Jason lehl, NetApp

- ◆ 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.**

- Introduction to Data Protection: Backup to Tape, Disk and Beyond
- Extending the enterprise backup paradigm with disk-based technologies allow users to significantly shrink or eliminate the backup time window. This tutorial focuses on various methodologies that can deliver efficient and cost effective solutions. This includes approaches to storage pooling inside of modern backup applications, using disk and file systems within these pools, as well as how and when to utilize Continuous Data Protection, deduplication and virtual tape libraries (VTL) within these infrastructures.
- Learning Objective:
  - ◆ Get a basic grounding in backup and restore technology including tape, disk, snapshots, deduplication, virtual tape, and replication technologies.
  - ◆ Compare and contrast backup and restore alternatives to achieve data protection and data recovery.
  - ◆ Identify and define backup and restore operations and terms.

# 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](http://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**

# Backup to Tape, Disk and Beyond

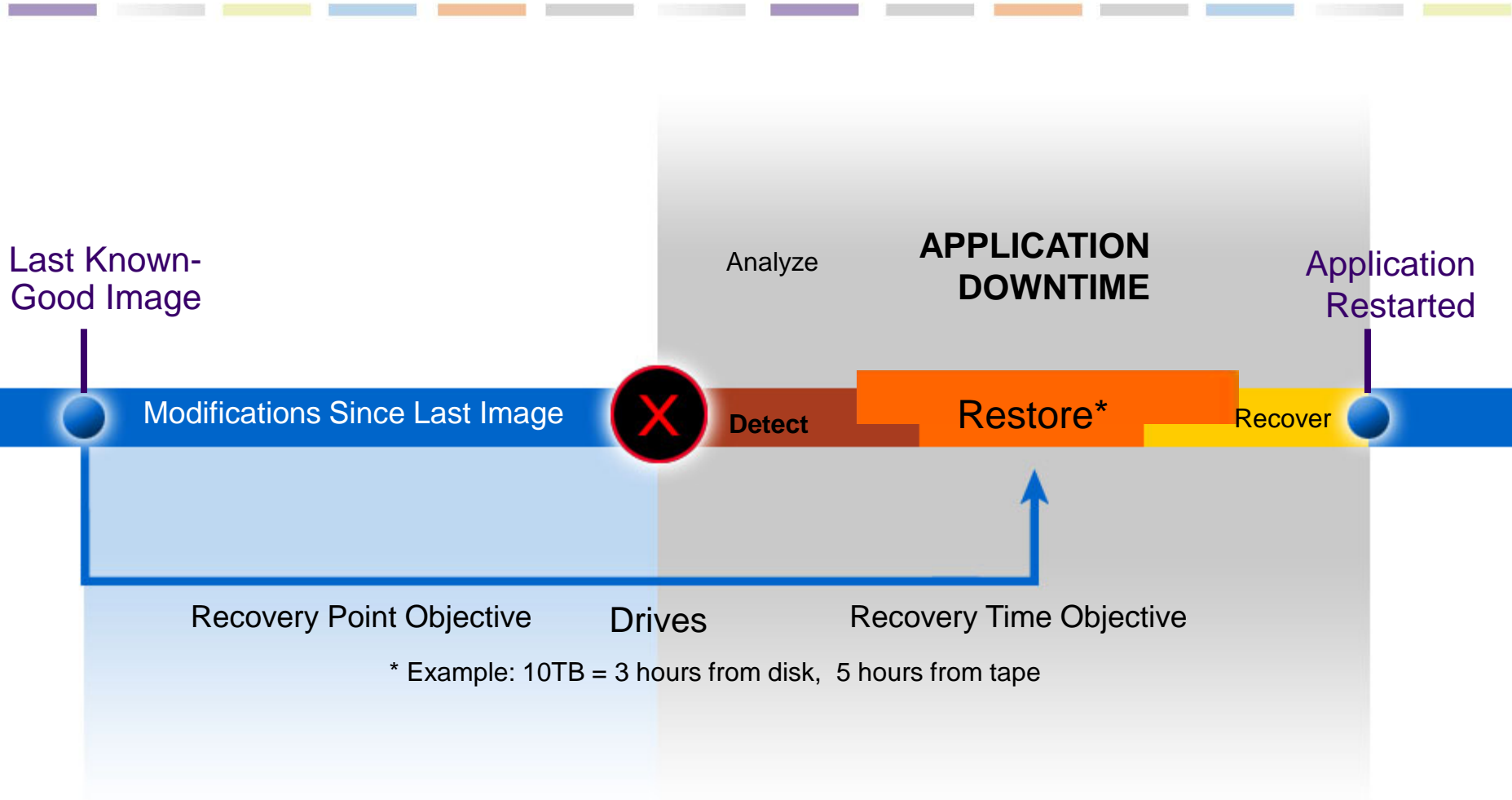
- Fundamental concepts in Data Protection
- Overview of Backup Mechanisms
- Backup Technologies
- Appendix

- ◆ Data protection is about data availability
- ◆ SNIA definition of Data Protection: Assurance that data is not corrupted, is accessible for authorized purposes only, and is in compliance with applicable requirements.
- ◆ There are a wide variety of tools available to us to achieve data protection, including backup, restoration, replication and disaster recovery.
- ◆ It is critical to stay focused on the actual goal -- availability of the data -- using the right set of tools for the specific job -- within time and \$ budgets.
- ◆ Held in the balance are concepts like the value of the data (data importance or business criticality), budget, speed, and cost of downtime.

# The Process of Recovery

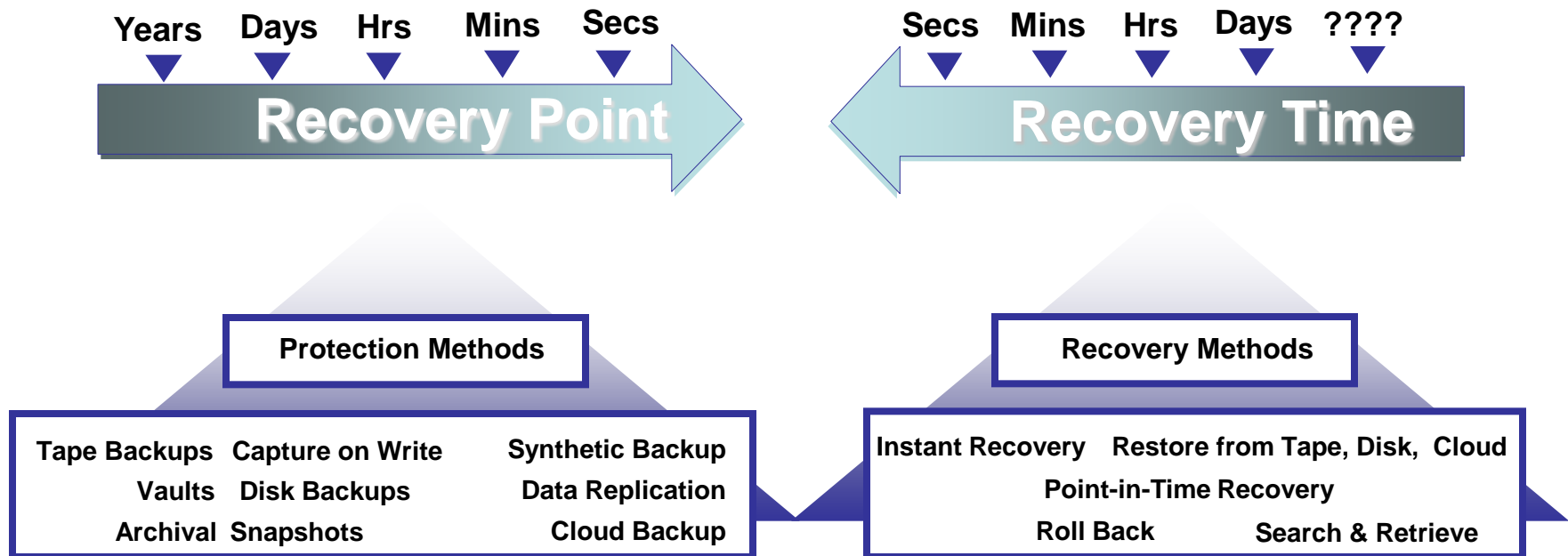
- ◆ **Detection**
  - ◆ Corruption or failure reported
- ◆ **Diagnosis / Decision**
  - ◆ What went wrong?
  - ◆ What recovery point should be used?
  - ◆ What method of recovery should be used -- overall strategy for the recovery?
- ◆ **Restoration**
  - ◆ Moving the data from backup to primary location
  - ◆ From tape to disk, or disk to disk, or cloud to disk; Restore the lost or corrupted information from the backup or archive (source), to the primary or production disks.
- ◆ **Recovery – Almost done!**
  - ◆ Application environment - perform standard recovery and startup operations
  - ◆ Any additional steps
    - > Replay log may be applied to a database
    - > Journals may be replayed for a file system
- ◆ **Test and Verify**

# Traditional Recovery





# Protection Based on Recovery



## ❖ Cold

- ◆ Offline image of all the data
- ◆ As backup window shrinks and data size expands, cold backup becomes untenable.
- ◆ Cheapest and simplest way to backup data

## ❖ Application Consistent

- ◆ Application supports ability to take parts of the data set offline during backup
- ◆ Application knows how to recover from a collection of consistent pieces.
- ◆ Avoids downtime due to backup window.

## ❖ Crash Consistent or Atomic

- ◆ Data copied or frozen at the exact same moment across the entire dataset.
- ◆ Application recovery from an atomic backup similar to a application failover.
  - › Rebuilding may be needed
- ◆ No backup window.



**Check out SNIA Tutorial:  
Trends in Data Protection**

# Data Protection Design Trade-offs

## ◆ Assessing your priorities

- ◆ Backup Window
  - › Shorten or eliminate
- ◆ Recovery Time Objective (RTO)
  - › Speed of recovery
  - › What is the cost of application downtime?
- ◆ Recovery Point Objective (RPO)
  - › Amount of data loss
  - › How far back in time to recover data?
- ◆ Move data offsite for DR or archive

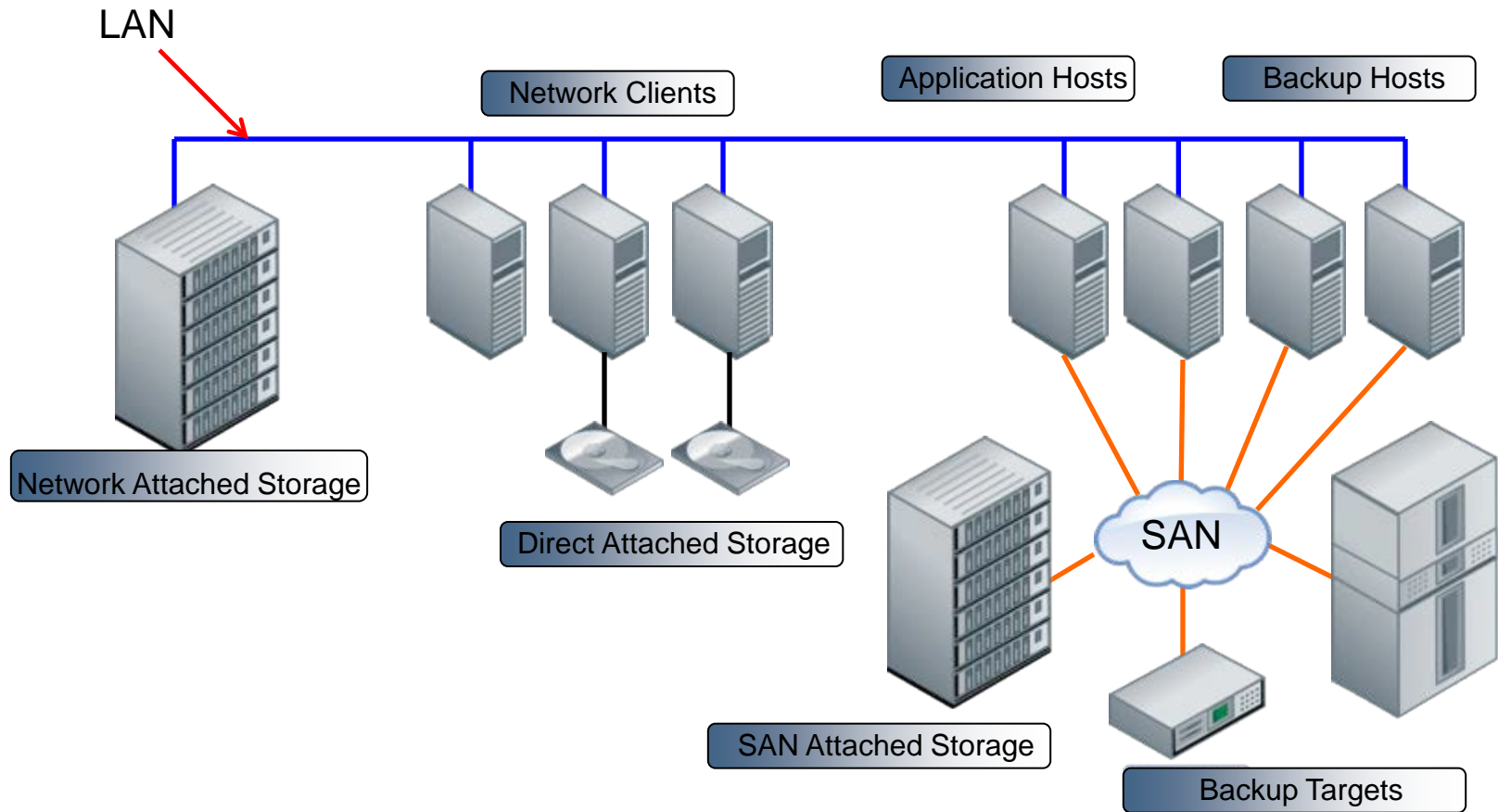


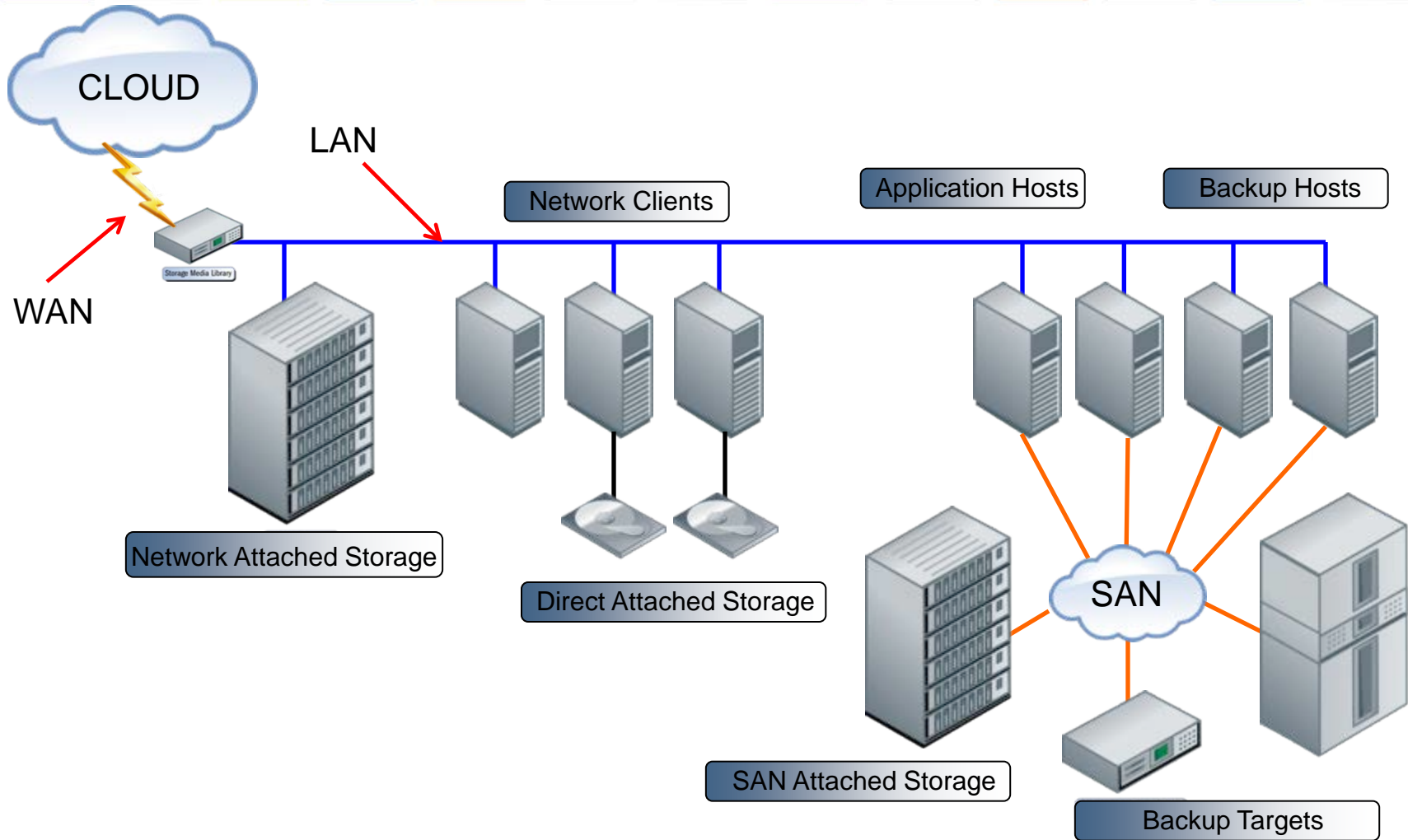
## ◆ There are trade-offs everywhere

- ◆ Newer technology improves but may not eliminate trade-offs
  - › Cost, downtime, business impact,
- ◆ Need to identify the priority order, and establish SLA targets for each type of data
  - › What is the cost of lost application?

# Backup to Tape, Disk and Beyond

- Fundamental concepts in Data Protection
- Overview of Backup Mechanisms
- Backup Technologies
- Appendix





# Backup Topology Components

## ➤ Backup Server

- ◆ Typically single point of administration
- ◆ Owns the Metadata catalog
  - › Must protect the catalog

## ➤ Storage Node or Media Server

- ◆ Collects the data from the Agent
- ◆ Read and writes to a secondary storage device

## ➤ Agent

- ◆ Manages the collection of the data and Metadata
- ◆ Traditional thin client or modern intelligent client

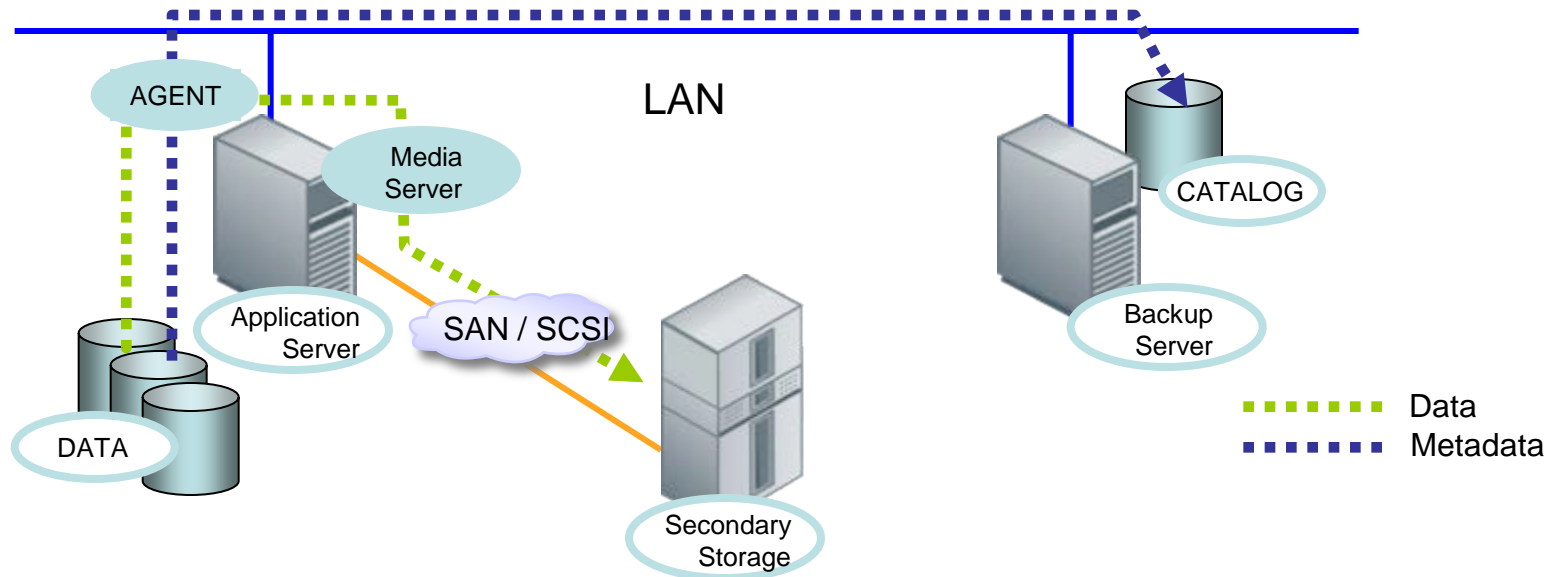
## ➤ Application Server

- ◆ Server that owns (produces) the data
- ◆ Maybe structured or unstructured data

## ➤ Secondary Storage

- ◆ Target media (destination) for the backup data

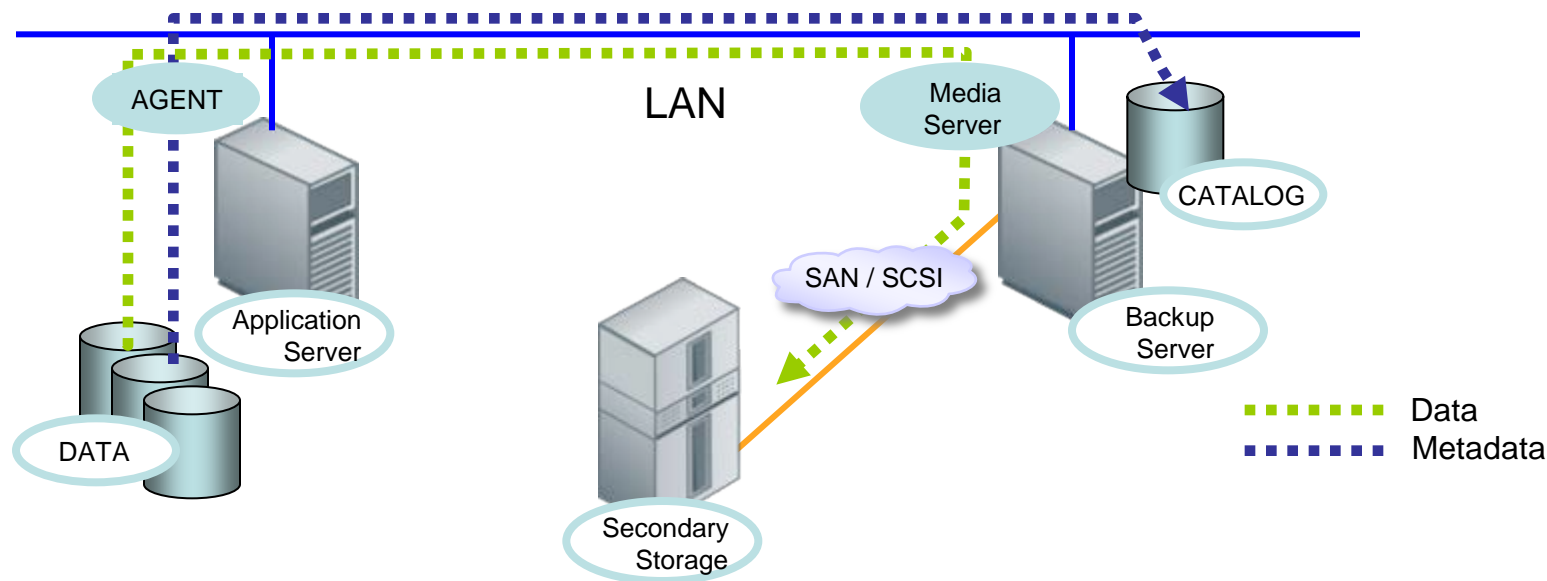
# Local Data Mover for Performance



- Sometimes known as LAN-Free backup
- Application server reads and writes the data locally
  - ◆ Application server acts as a media server
  - ◆ Storage is accessible by the application server
- Minimal LAN impact.
- Significant application server impact.



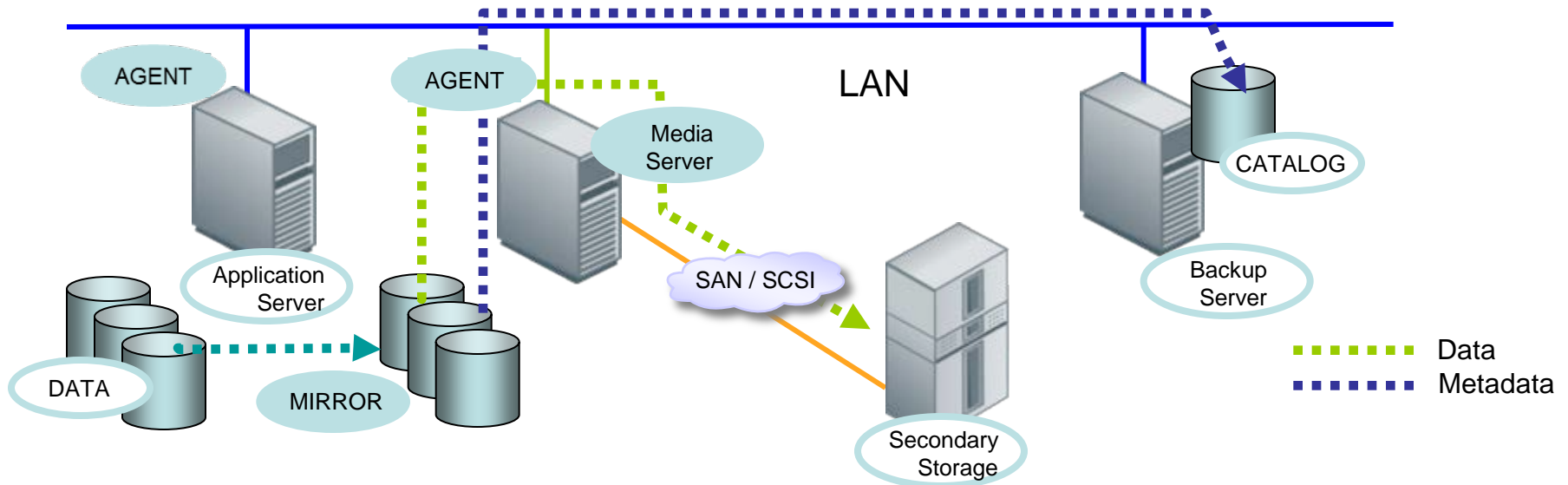
# LAN Backup



## ➤ Backup server receives data and Metadata from application server across the LAN

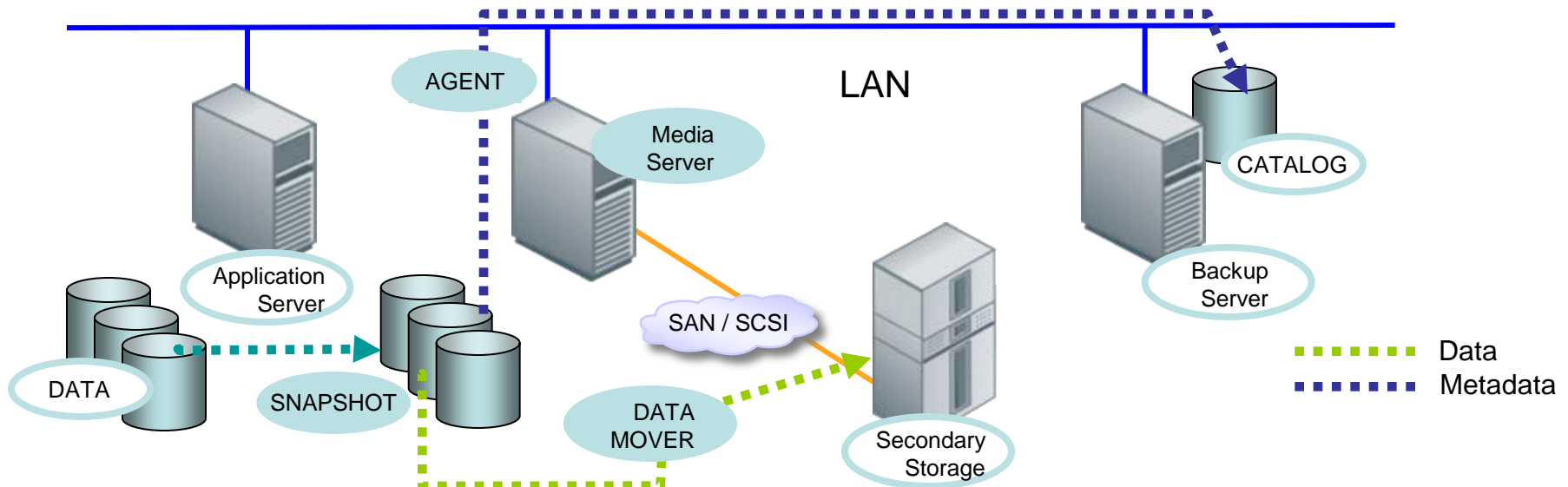
- ◆ LAN is impacted by both backup and restore requests
- ◆ Application server may be impacted by storage I/O
- ◆ CIFS, NFS, iSCSI, NDMP, or vendor specific

# (Application) Server-free Backup



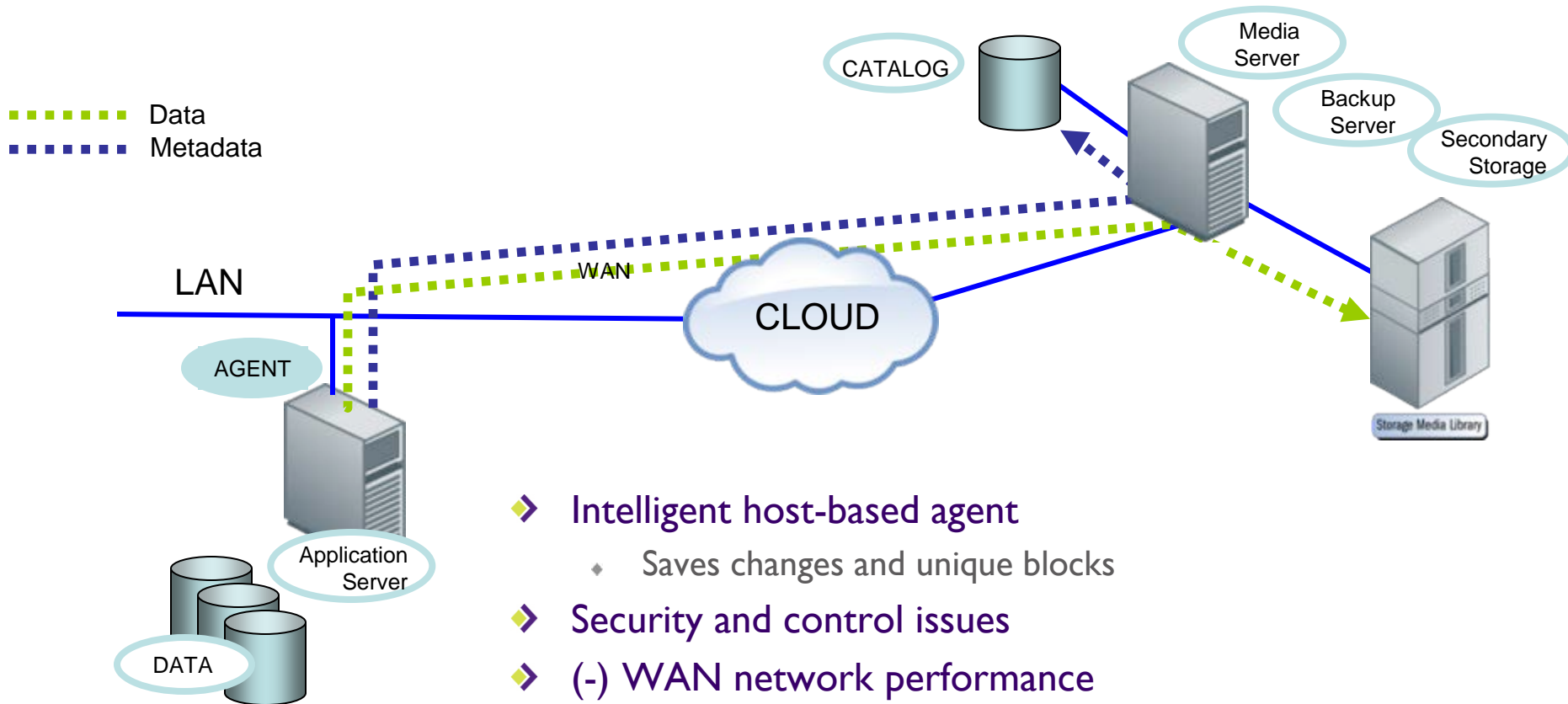
- The application server allocates a snapshot/mirror of the primary storage volume to a media server that delivers the data over the LAN or SAN
  - ◆ Media server must understand the volume structure
    - › Mirror: Application server impacted when creating the mirror
    - › Snapshot: Application server impacted by volume access
  - ◆ Metadata over the LAN to the backup server

# Server-free (Server-less) Backup



- ◆ Backup server delegates the data movement and I/O processing to a “Data-mover” enabled on a device within the environment
  - ◆ Network Data Management Protocol (NDMP)
    - › NDMP is a general open network protocol for controlling the exchange of data between two parties
  - ◆ SCSI Extended Copy (XCOPY or “Third-Party Copy”)
    - › Metadata still sent to the backup server for catalog updates
    - › Much less impact on the LAN

# CLOUD Backup



- Intelligent host-based agent
  - ◆ Saves changes and unique blocks
- Security and control issues
- (-) WAN network performance
  - ◆ Can use local cache to mitigate (“hybrid cloud”)
- (+) Low CAPEX
- (+) Off-site protection

# Traditional Backup Schedules

## ➤ Full Backup

- ◆ Everything copied to backup (cold or hot backup)
  - › Full view of the volume at that point in time
- ◆ Restoration straight-forward as all data is available in one backup image
- ◆ Huge resource consumption (server, network, tapes)

## ➤ Incremental Backup

- ◆ Only the data that changed since last full or incremental
  - › Change in the archive bit
- ◆ Usually requires multiple increments and previous full backup to do full restore
- ◆ Much less data is transferred

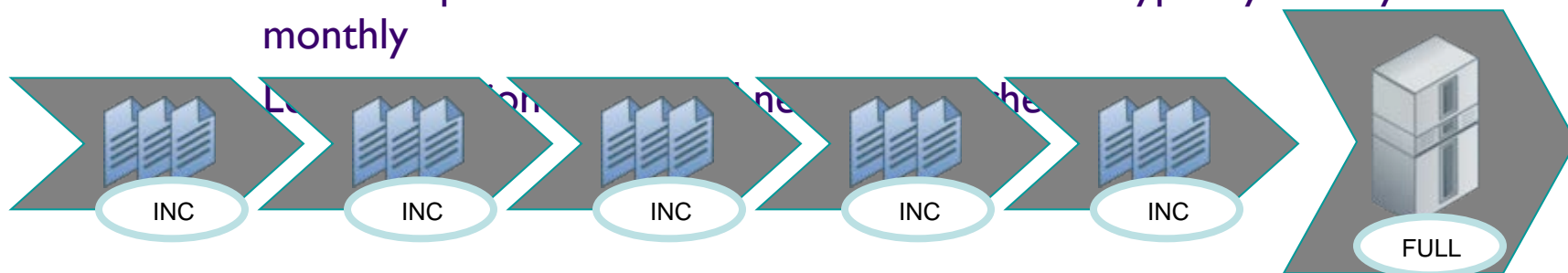
## ➤ Differential backup

- ◆ All of the data that changed from the last full backup
- ◆ Usually less data is transferred than a full
- ◆ Usually less time to restore full dataset than incremental

# Synthetic Backup & Incremental Forever

## ➤ Synthetic Full Backups

- ◆ Incremental backups are performed each day
  - › Full backups are constructed from incrementals typically weekly or monthly



## ➤ Incremental Forever

- ◆ Incremental backups are performed every day
- ◆ Primary backups are often sent to disk-based targets
- ◆ Collections of combined incrementals used for offsite copies
  - › Usually consolidate images from clients or application and create tapes
  - › May construct synthetic full in the cloud

# What Gets Backed Up and How

## ➤ File-level backups

- ◆ Any change to a file will cause entire file to be backed up
- ◆ Open files often require special handling SW
  - Open files may get passed over – measure the risks
- ◆ PRO: Ease of BU and restore CON: Moves tons of data

## ➤ Block-level backups

- ◆ Only the blocks that change in a file are saved
- ◆ Requires client-side processing to discover changed blocks
- ◆ PRO: Smaller backups, Less network impact, Faster
- ◆ CON: Client-side impact, increased complexity

## ➤ Client-side backups

- ◆ Intelligent agent monitors changes and protects only new blocks
- ◆ Agent enables advanced technology, granular backups and user policies
- ◆ Deduplication can enable network efficiency, reduce BU data volume
- ◆ PRO: Efficiently distributes work CON: Complex client/server

# Backup to Tape, Disk and Beyond

- Fundamental concepts in Data Protection
- Overview of Backup Mechanisms
- **Backup Technologies**
- Appendix



# Introduction to Tape

- Sequential access technology
  - ◆ Versus random access
- Can be removed and stored on a shelf or offsite
  - ◆ Disaster Recovery
  - ◆ Encrypted, Archived for compliance?
  - ◆ Reduce power consumption
- Media replacement costs
  - ◆ Tape life, reusability
- Performance and Utilization
  - ◆ Can accept data at very high speeds, if you can push it
  - ◆ Streaming and multiplexing
- Typically Managed by backup and recovery software
  - ◆ Controls robotics (Inventory)
  - ◆ Media management



**Tape is not Dead!**

# Tape Based Backup: Considerations

- ◆ Tape drives run faster than most backup jobs – Is this good?
  - ◆ Matching backup speed is more important than exceeding it
  - ◆ Avoid shoe-shining
- ◆ Slower hosts can tie up an expensive drive
  - ◆ It's a shame to waste a drive on these hosts.
- ◆ Slower tapes can tie up expensive (important) servers.
  - ◆ It's a shame to let the tape drive throttle backup servers
  - ◆ Slow backup can impact production servers as well
- ◆ Replacing your tapes may not solve your backup challenges
  - ◆ A well designed backup architecture is the best answer
- ◆ If backup target speed is your issue:
  - ◆ Consider multiplexing – Good for backup, not-so-good for restore
  - ◆ Consider alternates such as virtual tape (VTL) or D2D2T.
- ◆ Security, security, security.....

# Introduction to B2D/D2D

## ➤ What?

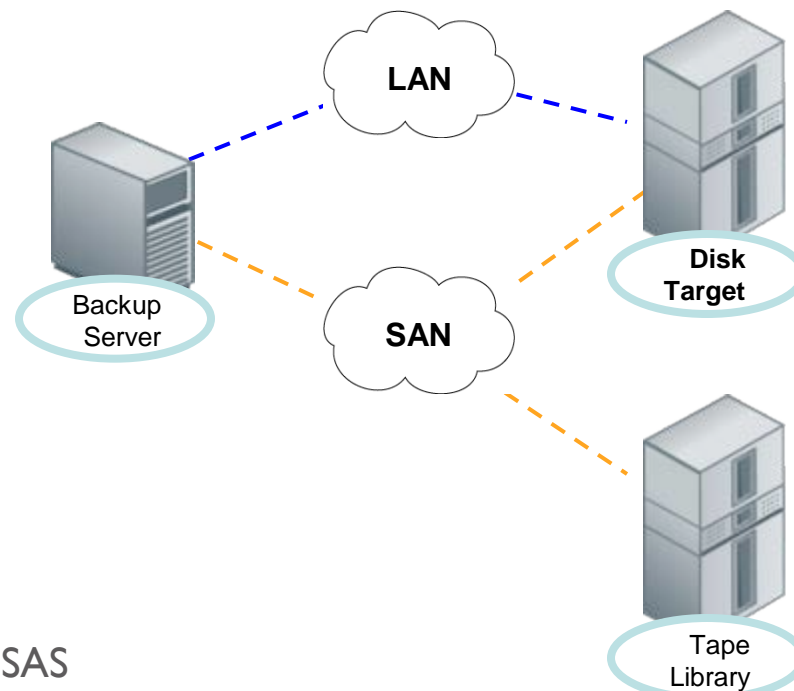
- ◆ Backup to Disk / Disk to Disk Backup
- ◆ Disk as a primary backup target

## ➤ Why?

- ◆ Performance and reliability
  - Reduced backup window
  - Greatly improved restores
  - RAID protection
  - Eliminate mechanical interfaces
- ◆ Eliminate (tape) multiplexing
- ◆ More effective sharing of backup targets

## ➤ Considerations

- ◆ Fibre Channel Disks versus SATA versus SAS
  - I/O random access vs. MB/s sequential
- ◆ SAN, NAS or DAS
- ◆ VTL or mirroring
- ◆ Consider a mix of Disk and Tape (D2D2T)



# Introduction to VTL

## ➤ What:

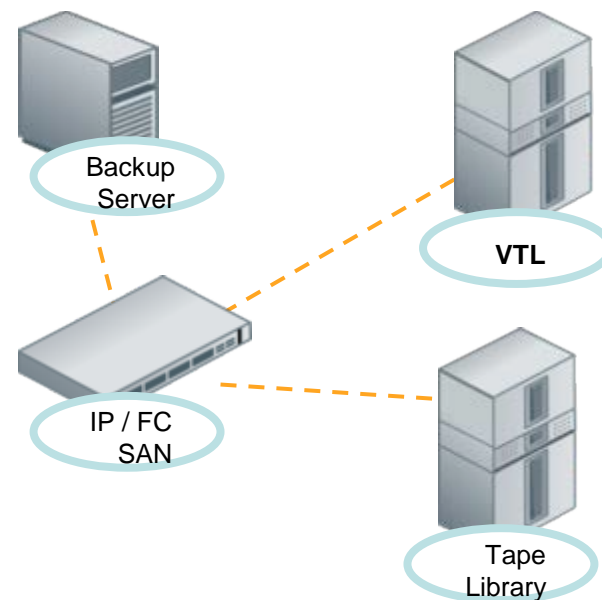
- ◆ Virtual Tape Libraries emulate traditional tape
- ◆ Fits within existing backup environment
- ◆ Easy to deploy and integrate
- ◆ Reduce / eliminate tape handling

## ➤ Why:

- ◆ Improved performance and reliability (see B2D)
- ◆ Reduced complexity versus straight B2D or tape
- ◆ Unlimited tape drives reduce device sharing, improve backup times
- ◆ Enables technologies such as remote replication, deduplication

## ➤ Considerations:

- ◆ Easy to manage in traditional backup software environment:
- ◆ Can extend the life of current physical tape investment



# Introduction to CDP

## What:

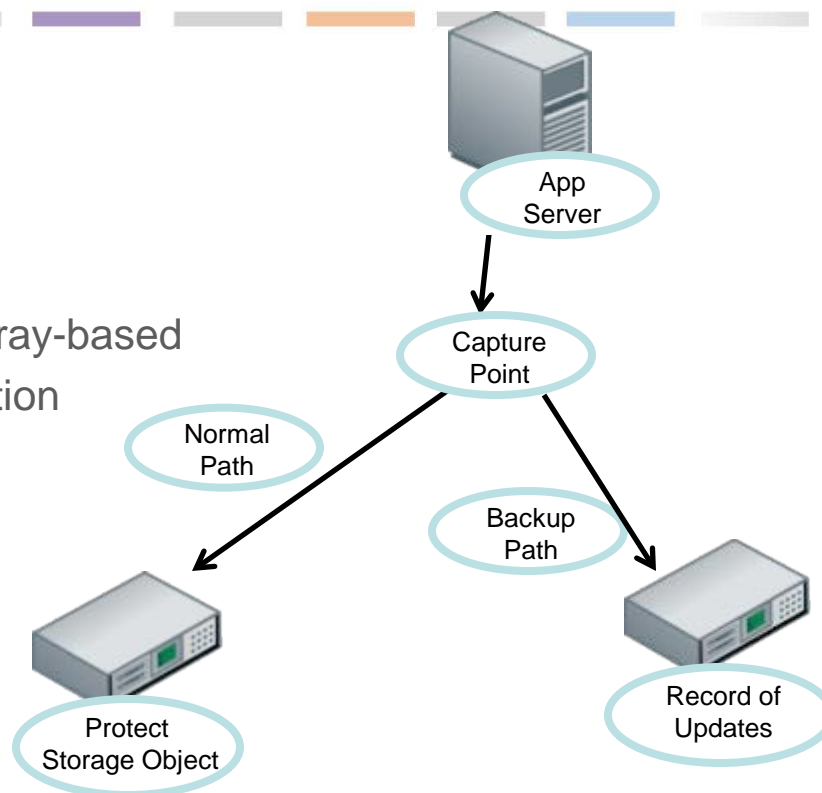
- ◆ Continuous Data Protection
- ◆ Capture every change as it occurs
- ◆ May be host-based, SAN-based, array-based
- ◆ Protected copy in a secondary location
- ◆ “Roll back” to any point in time

## How:

- ◆ Block-based
- ◆ File-based
- ◆ Application-based

## Why:

- ◆ Implementations of true CDP today are delivering zero data loss, zero backup window and simple recovery. CDP customers can protect all data at all times and recover directly to any point in time.
- ◆ “Near CDP” (Snapshots, checkpoints) may also help but will not catch every change



# Contrasting CDP and Replication

- Replication is not CDP (Synchronous)
  - ◆ Replica or mirror is a single PIT copy of the data
  - ◆ Multiple replicas plus logs can create multiple points in time
  
- Snapshots are not CDP (Asynchronous)
  - ◆ Data loss possible if crash or corruption happens between snaps
  - ◆ Snapshots frequently to same system as primary
  - ◆ Lack continuous index with embedded knowledge of relationship of data to files, folders, application and server
  
- Backups (even multiple backups) are not CDP:
  - ◆ Schedule frequency
  - ◆ Database logging can provide additional granularity but still not CDP

# Introduction to Snapshots

## ➤ What?

- ◆ A disk based “instant copy” that 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
- ◆ Often handled at the storage level
  - › May be done at application server, hypervisor, and/or in cloud

## ➤ Why?

- ◆ Allows for complete backup or restore
  - › With application downtime measured in minutes (or less)
- ◆ May be able to be combined with replication
- ◆ 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

	<b>Full Copy Snapshot</b>	<b>Differential Copy Snapshot</b>
<b>Upsides</b>	<ul style="list-style-type: none"> <li>◆ Minimal performance impact</li> <li>◆ Independent copy available for DR</li> </ul>	<ul style="list-style-type: none"> <li>◆ Less storage consumption</li> <li>◆ Often takes advantage of cheaper disk</li> </ul>
<b>Downsides</b>	<ul style="list-style-type: none"> <li>◆ High disk utilization</li> <li>◆ No GEO-redundant protection</li> </ul>	<ul style="list-style-type: none"> <li>◆ Performance may be impacted</li> <li>◆ Dependent on primary copy</li> </ul>
<b>Applications</b>	<ul style="list-style-type: none"> <li>◆ Disaster Recovery</li> <li>◆ Near zero backup window</li> <li>◆ Fastest restore</li> <li>◆ Valuable for data repurposing</li> </ul>	<ul style="list-style-type: none"> <li>◆ Backup source</li> <li>◆ Near zero backup window</li> <li>◆ Fast restore</li> <li>◆ Can help with data repurposing                             <ul style="list-style-type: none"> <li>◆ Beware performance impact</li> </ul> </li> </ul>



# Introduction to Data Deduplication

## ➤ What?

- ◆ The process of examining a data-set 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 (maybe) 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
- ◆ Multiple performance trade-offs

# Factors Impacting Space Savings

<b>More Effective Deduplication</b>	<b>Less Effective Deduplication</b>
<b>Data created by users</b>	<b>Data captured from mother nature</b>
<b>Low change rates</b>	<b>High change rates</b>
<b>Reference data and inactive data</b>	<b>Active data, encrypted data, compressed data</b>
<b>Applications with lower data transfer rates</b>	<b>Applications with higher data transfer rates</b>
<b>Use of full backups</b>	<b>Use of incremental backups</b>
<b>Longer retention of deduplicated data</b>	<b>Shorter retention of deduplicated data</b>
<b>Continuous business process improvement</b>	<b>Business as usual operational procedures</b>
<b>Format awareness</b>	<b>No format awareness</b>
<b>Temporal data deduplication</b>	<b>Spatial data deduplication</b>

**Don't forget about compression**

# Next Steps in Data Protection

- Choose the appropriate level of protection
  - ◆ Assess risk versus cost versus complexity
  - ◆ Include your “customers” in your decisions
- Match RPO, RTO goals with technology
  - ◆ Consider resources required to support your decisions
  - ◆ Consider centralized versus distributed solutions
- Performance is **ALWAYS** a consideration
  - ◆ Assess your system today for strengths and weaknesses
  - ◆ A new box or new SW may NOT be the answer
- When in doubt, call in the experts

# Where to Get More Information

## ➤ Related tutorials

- ◆ Active Archive – Data Protection for the Data Center
- ◆ Advanced Deduplication Concepts
- ◆ Trends in Data Protection and Restoration Technologies
- ◆ Understanding Data Deduplication
- ◆ Retaining Information for 100 Years

## ➤ Visit the Data Protection and Capacity Optimization Committee website <http://www.snia.org/dpco/>

## ➤ DPCO online Product Selection Guide

- › <http://sniadataprotectionguide.org/>

- Please send any questions or comments on this presentation to SNIA: [tracktutorials@snia.org](mailto:tracktutorials@snia.org)

**Many thanks to the following individuals  
for their contributions to this tutorial.**

**- SNIA Education Committee**

**SNIA Data Protection &  
Capacity Optimization Committee  
SNIA Tech Council  
Nancy Clay  
Rob Peglar  
Gene Nagle**

**Mike Fishman  
Jason lehl  
Mike Rowan  
SW Worth  
Joseph White  
Thomas Rivera**

- **Data Protection and Capacity Optimization Committee:**  
<http://www.snia.org/dpco/>