Education

Trends in Data Protection and Restoration Technologies

Jason Iehl, NetApp

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.

Abstract

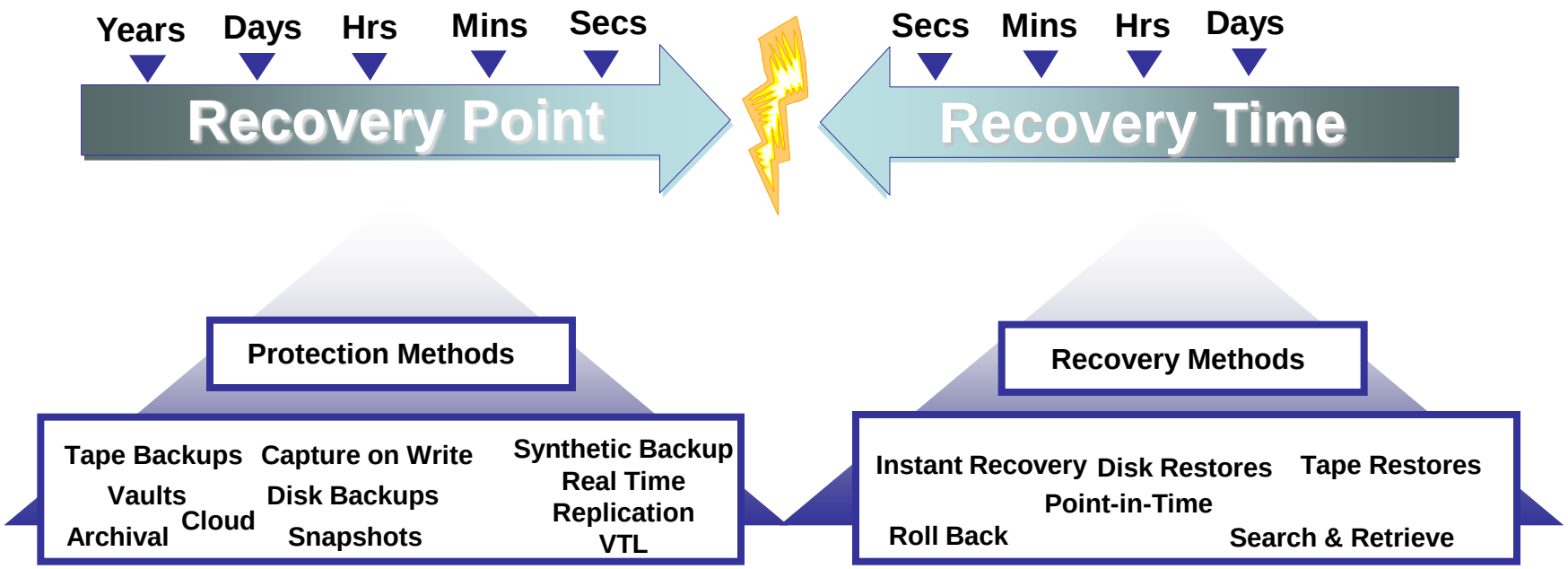
- Many disk technologies, both old and new, are being used to augment tried and true backup and data protection methodologies to deliver better information and application restoration performance. These technologies work in parallel with the existing backup paradigm,

- 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.

- Detail of these technologies interoperate will be provided as well as best practices recommendations for deployment in today's heterogeneous data centers.

  - Understand legacy and contemporary storage technologies that provide advanced data protection
  - Compare and contrast advanced data protection alternatives
  - Gain insights into emerging DP technologies.

Protection Based on Recovery



Enabling Technologies

Tape & Automation	Virtual Tape Library	De-duplication	Continuous Data Protection	Snapshots	Backup to Disk

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 mirror-split, or CDP protection).
- Considerations
  - It is important to understand the consistency semantics of your application so that your data protection copies are recoverable

Consistency - Offline 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 disk space
  - Enables incremental backup based on transaction logs
- Start the application
- Optional: backup the snapshot to another disk or tape
- Recycle the snapshot
  - Keep the last # snapshots
  - Snapshot rotation

Crash Consistent

- Create a snapshot while the application runs
  - Application consistency has the same quality as after a system crash
  - Most applications / databases can survive system crashes
    - But some don ‘t and some not always.
  - Recovery can not be guaranteed
- 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
    - Might reduce the risk
  - 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
  - Across the server farm
- Use cases
  - 7 x 24 operations -> no backup window
  - Guaranteed & fully supported consistent recovery

Data Protection & Data Management

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

Data Management

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

Tiered Storage

- Primary
- Secondary
- Archive
- Backup
- Cloud

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
  - Any technology that presents a consistent point-in-time view of changing data. Many implementations exist.
- 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 Comparison

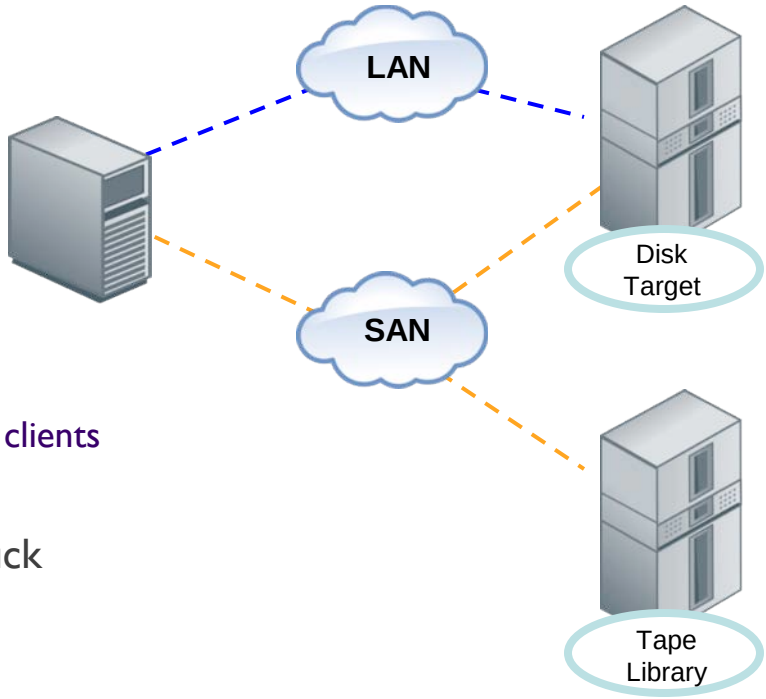
	Full Copy Snapshot	Differential Copy Snapshot
Upsides	◆ No cost during "snapshot" process ◆ Can be used for DR - independent copy	◆ Less storage consumption - typically 10-20% ◆ Depends on change rate ◆ Typically can take advantage of cheaper disk
Downsides	◆ Massive storage cost ◆ 1x of storage per RPO ◆ Like disk - expensive ◆ Often in the same disk chassis ◆ Loss of DR component ◆ Consider re-sync time in schedules	◆ Performance may be impacted while snapshot exists ◆ Multiple implementations to optimize performance impact ◆ Most vendors don't offer multiple implementations - pick at onset ◆ Leverages main copy - not DR capable
Applications	◆ Disaster Recovery ◆ Near zero backup window ◆ 24x7 operations ◆ Faster restore ◆ Can do no-copy restore ◆ Can help with data repurposing	◆ Backup source ◆ Near zero backup window ◆ 24x7 operations ◆ Fast restore ◆ copy based by definition ◆ Can help with data repurposing ◆ Beware performance impact

Snapshot Considerations

- Snaps of production storage may impact production
  - Depending on use case
- Snap recovery tools may not be as mature
- Retention policy impact
  - Number of copies retained
  - Recovery granularity
  - Meeting off-site protection via distance replication
- On-array versus off-array alternatives
- Cost trade-offs and information classification
- Can you combine snapshots with replication

Backup to Disk (B2D)

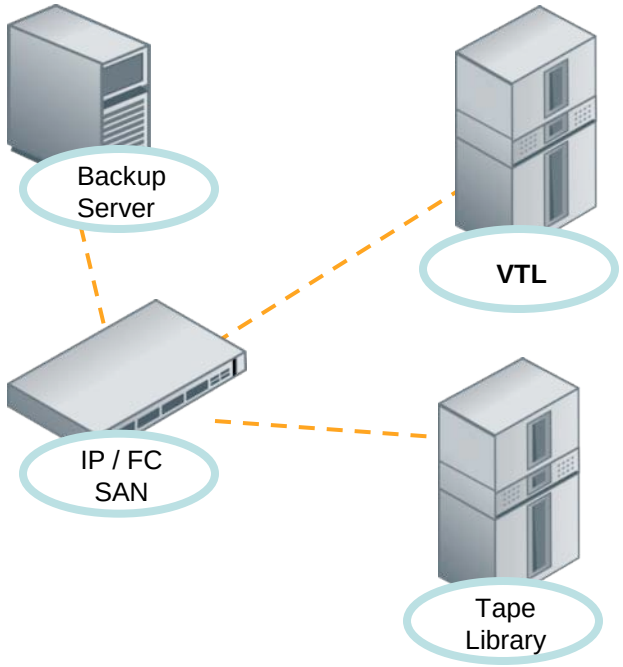
- What and Why?
  - Backup target is a LUN or a share
    - Easy to implement
  - Any type of disk or connection
    - FC, SATA, SAS, DAS, NAS, etc
  - Disk-based Snapshots
    - Reduce impact on production host
    - Improve tape streaming when backing up slower clients
  - Single/Volume Restores are faster than tape
  - Remote and local backups on-line versus on-truck



- What to watch out for
  - Bottlenecks may NOT be the backup target
  - Disks can be faster than tape, but when?
  - More difficult and possibly slower than VTL (generally)
  - Level of enterprise backup application support
    - May charge more for advanced B2D functionality
  - Backup window issues may still exist - not "instant"

Virtual Tape Library (VTL)

- What?
  - Looks like Tape, Acts like tape
  - Fits within existing backup environment
  - Easy to deploy and integrate
  - Takes advantage of current processes
  - Reduces tape media handling



- Why?
  - Improved speed and reliability
    - Exceeds speed of high end tape
    - No fast-forward, no rewind,
    - High performance without multiplexing
    - Enables faster access to data, faster restores
    - No detached tape leaders or mechanical failures
  - Watch out:
    - Integration with physical tape
    - Consider total aggregate speed as well as speed per-drive

VTL Best Practices

- Use as primary backup target to reduce backup window
  - Leverage existing tape as backup target where appropriate
- Add storage to enable additional recovery time objectives
- Follow physical tape configuration and sharing rules
  - Match virtual drives per connection
  - Don't mix tape and disk on same ports
  - Use the right OS driver
- Tape redeployment
  - Eject process, controlled by the backup software
  - Backend tape creation
    - Cloning or vaulting or tape copies – leveraging the backup server
    - VTL directly creates tapes
- Offsite requirements
  - Bandwidth, connectivity, time to complete tape copies

Data Deduplication

- What?
  - The replacement of multiple copies of data – at variable levels of granularity – with references to a shared copy in order to save storage space and/or bandwidth
- Why?
  - Many data sets contain a high degree of redundant data
    - Example Repetitive Full Backups
  - Reduction in cost per terabyte stored
  - Enables storage of greater amounts of data
  - Significant reduction in storage footprint
  - Reduce power and cooling costs
- Considerations
  - More data stored on fewer disks
  - Enables low cost replication
    - Offsite copies
    - WAN Optimization
  - Reduce backup, archive and primary storage
  - Benefit from D2D backup
  - Possible increase in workload
    - IO and Performance

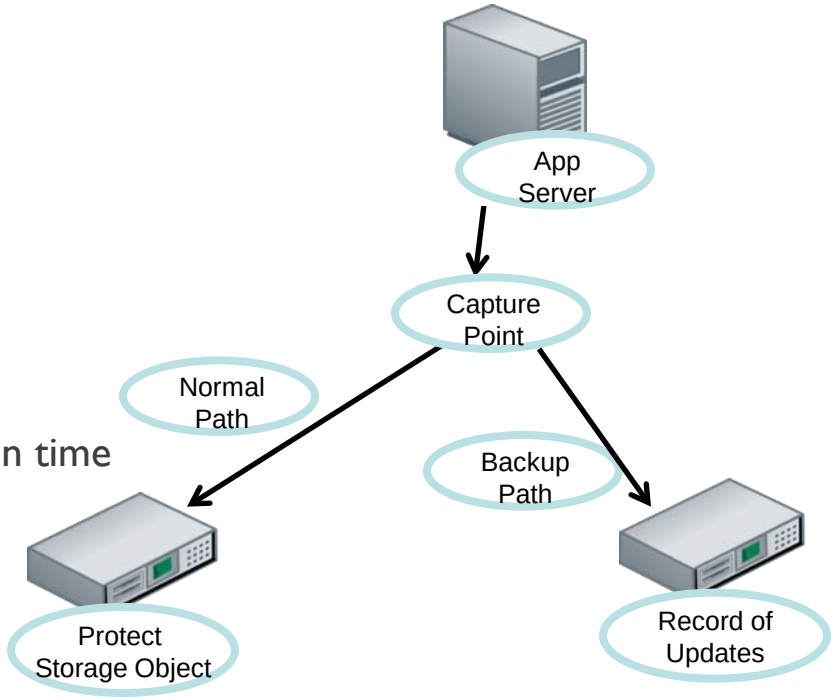
Factors Impacting Space Savings

Factors associated with higher data deduplication ratios	Factors associated with lower data deduplication ratios
Data created by users	Data captured from mother nature
Low change rates	High change rates
Reference data and inactive data	Active data, encrypted data, compressed data
Applications with lower data transfer rates	Applications with higher data transfer rates
Use of full backups	Use of incremental backups
Longer retention of deduplicated data	Shorter retention of deduplicated data
Wider scope of data deduplication	Narrower scope of data deduplication
Continuous business process improvement	Business as usual operational procedures
Smaller segment size	Larger segment size
Variable-length segment size	Fixed-length segment size
Format awareness	No format awareness
Temporal data deduplication	Spatial data deduplication

Don’t forget about compression

Continuous Data Protection

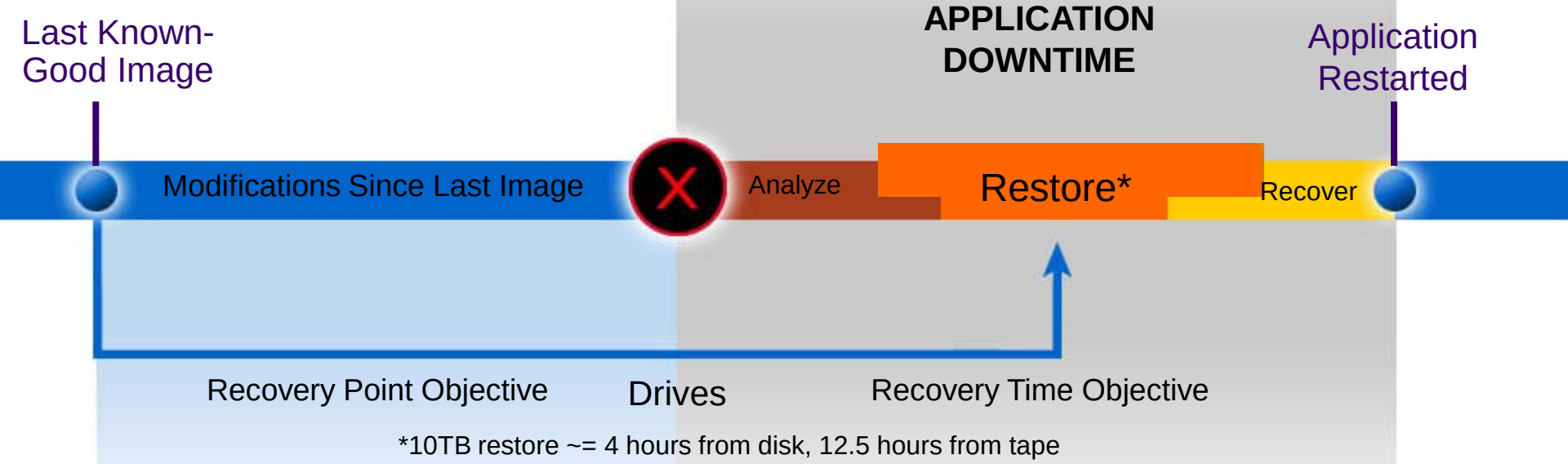
- What?
  - Capture every change as it occurs
  - Protected copy in a secondary location
  - Recover to any point in time
- Why?
  - Zero data loss, Zero backup window
  - Granular recovery - directly to any point in time
  - Simple recovery
  - CDP protects data at all times
- How?
  - Block-based
  - File-based
  - Application-based



The CDP (Continuous) Difference

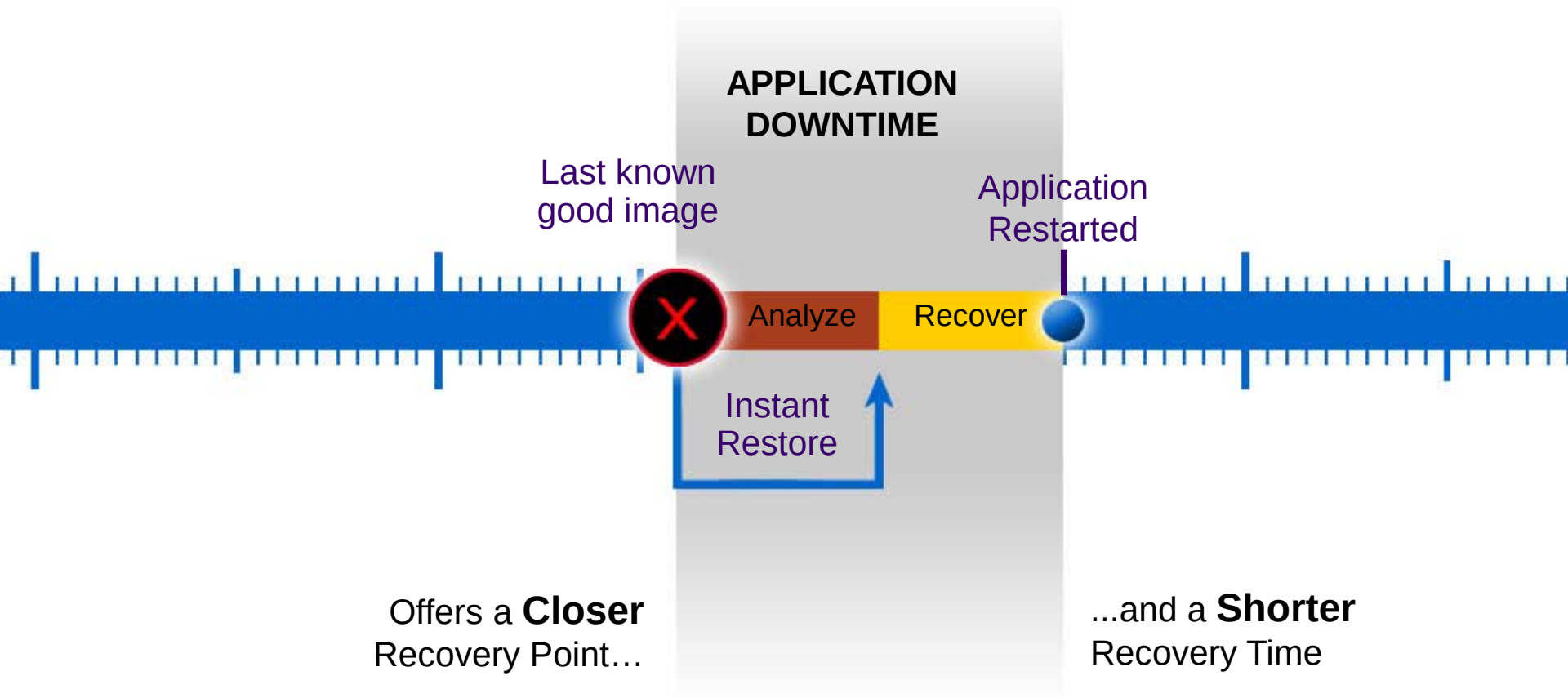
- Replication is not CDP:
  - Maintains only a current copy of the data (Synchronous)
  - May be combined with some snapshot capabilities
- Snapshots are not CDP:
  - Snapshots are scheduled events (Asynchronous)
    - Data loss possible if crash or corruption happens between snaps
    - Snapshots frequently to same system as primary
    - Lack continuous index with embedded knowledge of relationship of data to files, folders, application and server
- Scheduled events are not CDP:
  - Scheduled backup processes
  - Log collection for database style applications, rolling transactions forwards or backwards

Traditional Recovery



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

Recovery with CDP

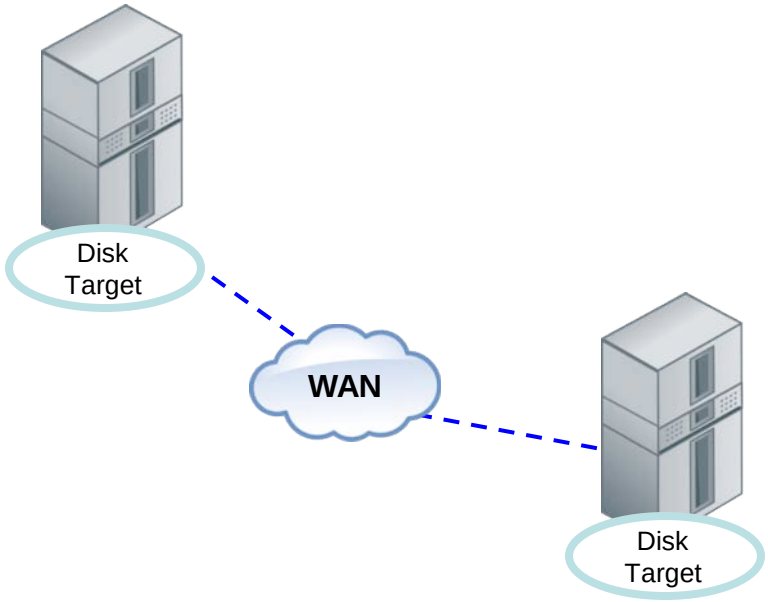


Offers a **Closer** Recovery Point…

...and a **Shorter** Recovery Time

Backup Replication

- What and Why?
  - Replicate backup images or snapshots offsite
  - Reduce amount of tape consumed
  - Quicker access to data for restore
    - Depends on bandwidth & size of restore
  - Increased security
  - Extend to the Cloud?
- Considerations?
  - How much bandwidth is available
    - Size of pipe at both sides
    - Amount already used for business applications
  - What is the process of recovery?
    - Is there a need for a backup server
  - Can the data be leveraged for other tasks?
    - Disaster Recovery & Dev/Test
  - How many copies are at primary location
  - May require extra storage for long term retention



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…..

Refer to the Hands-On Lab

Hands-On LAB COMPUTERWORLD SNIA SNW

Check out the Hands-On Lab

Enterprise Content Management, Cloud Storage

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: trackdatamgmt@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
Mike Fishman
Jason Iehl
Mike Rowan
SW Worth
Andreas Schwegmann

- Data Protection and Capacity Optimization Committee: http://www.snia.org/forums/dpco/