Education

Trends in Data Protection and Restoration Technologies

Mike Fishman, EMC^2 Corporation
(Author and Presenter)

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

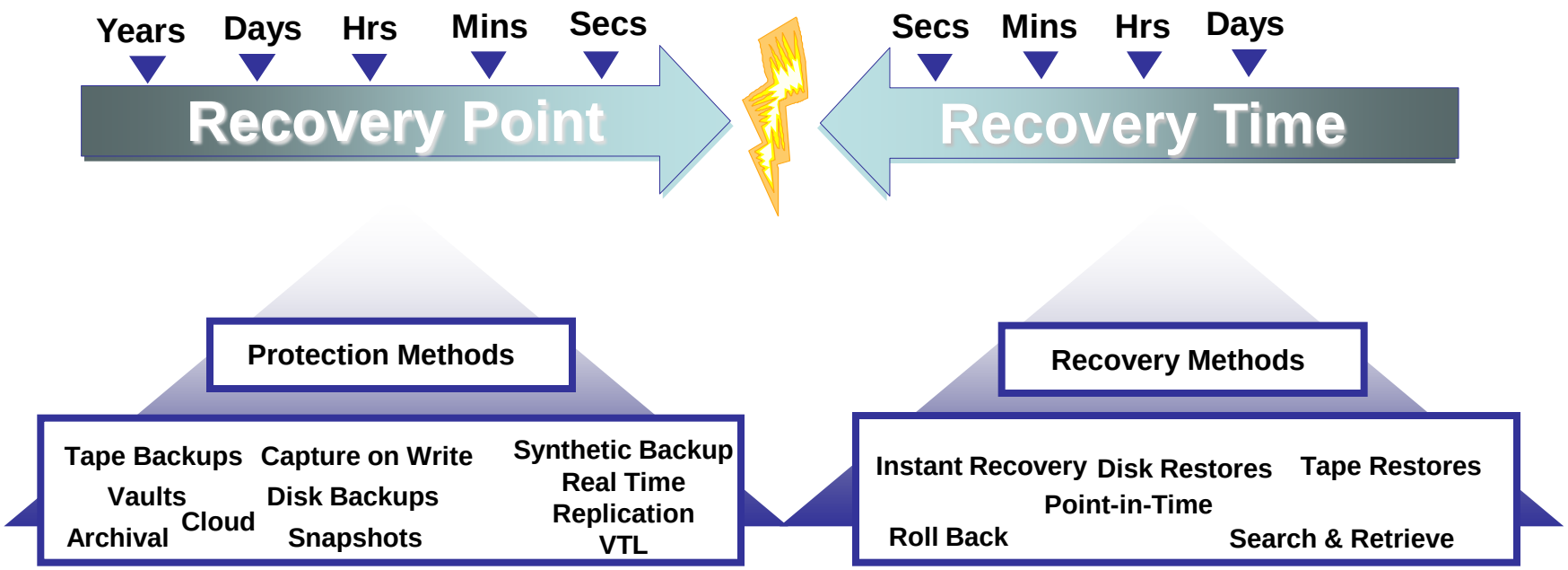
- Many disk technologies, both old and new, are being used to augment tried and true backup and data protection methodologies to deliver better information and application restoration performance. These technologies work in parallel with the existing backup paradigm,

- 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 how 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 device or location
- 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 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 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
- 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
- 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

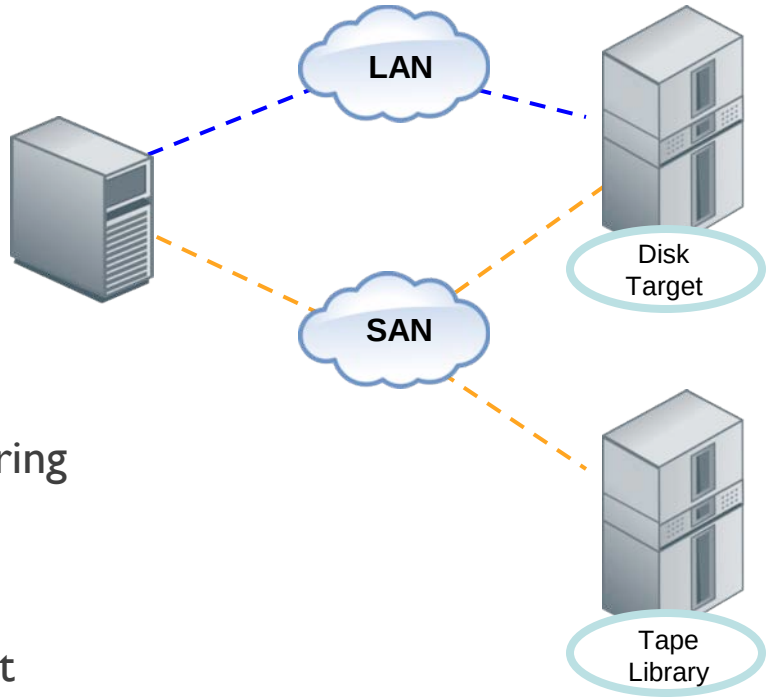
	Full Copy Snapshot	Differential Copy Snapshot
Upsides	• Minimal performance impact • Independent copy available for DR	• Less storage consumption • Often takes advantage of cheaper disk
Downsides	• High disk utilization • No GEO-redundant protection	• Performance may be impacted • Dependent on primary copy
Applications	• Disaster Recovery • Near zero backup window • Fastest restore • Valuable for data repurposing	• Backup source • Near zero backup window • Fast restore • Can help with data repurposing ◦ Beware performance impact

Snapshot Considerations

- Snaps of production storage may impact production
  - Consider performance, storage, virtual-machine impacts
- 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

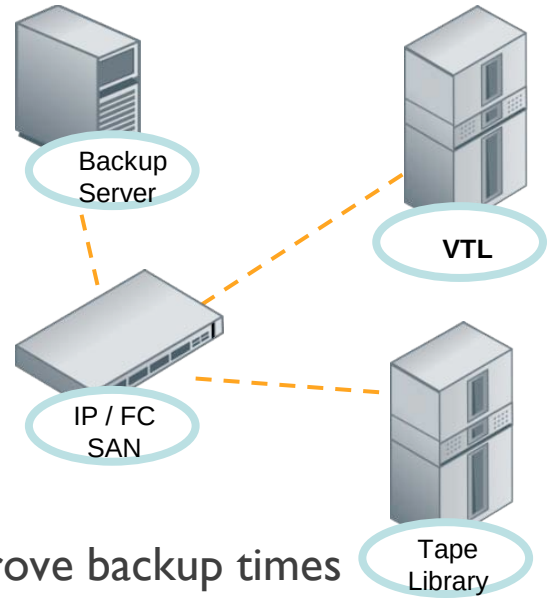
Backup to Disk (B2D)

- What
  - 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 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
  - Enables 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

VTL Best Practices

- Use as primary backup target to reduce backup window
- 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
  - 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 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

Check out SNIA Tutorial:

Advanced Deduplication Concepts

- 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
  - Trade-offs (Performance, Encryption, Compression)

Factors Impacting Space Savings

More Effective Deduplication	Less Effective Deduplication
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
Continuous business process improvement	Business as usual operational procedures
Format awareness	No format awareness
Temporal data deduplication	Spatial data deduplication

Don't forget about compression

Backup of Virtual Machines

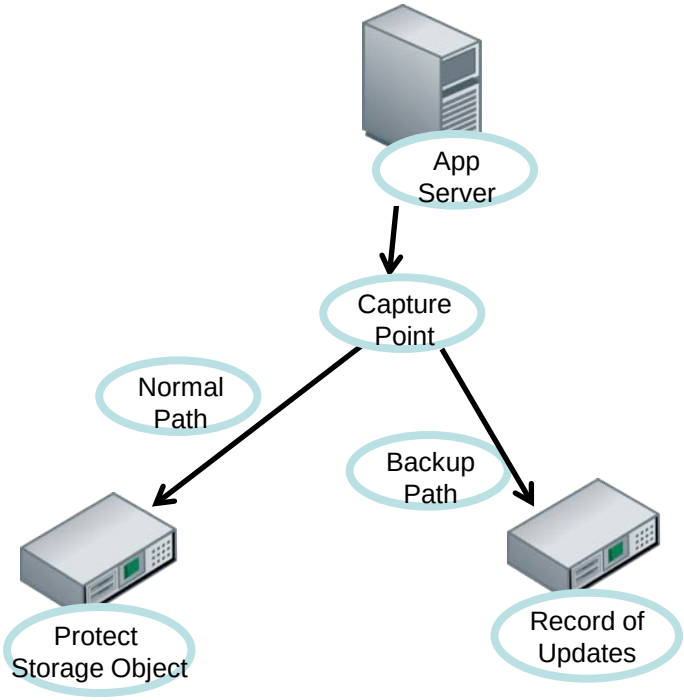
- What?
  - Backup virtual machines and hypervisors
  - Avoid resource contention (ie. Impact of backing up 20 VM's)
  - Address backup storage growth
  - Consider your recovery goals and objectives
- How?
  - Deliver granularity of "client" backup and restore
  - Improve operations by reducing backup 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
- Isolate data files that don't need backups (temporary and transient data)
- Larger configurations may require more advanced tools
- Insist on integration with hypervisor tools and API's
- Backup the hypervisor AND then backup the virtual hosts
- Deduplicate your backups
- Consider snaps, clones and copies

Continuous Data Protection

- What?
  - Capture every change as it occurs
  - Data can be collected and stored externally
  - "Roll-back" to any point in time
- Why?
  - Zero data loss, Zero backup window
  - Granular recovery to any point in time
  - Continuous protection

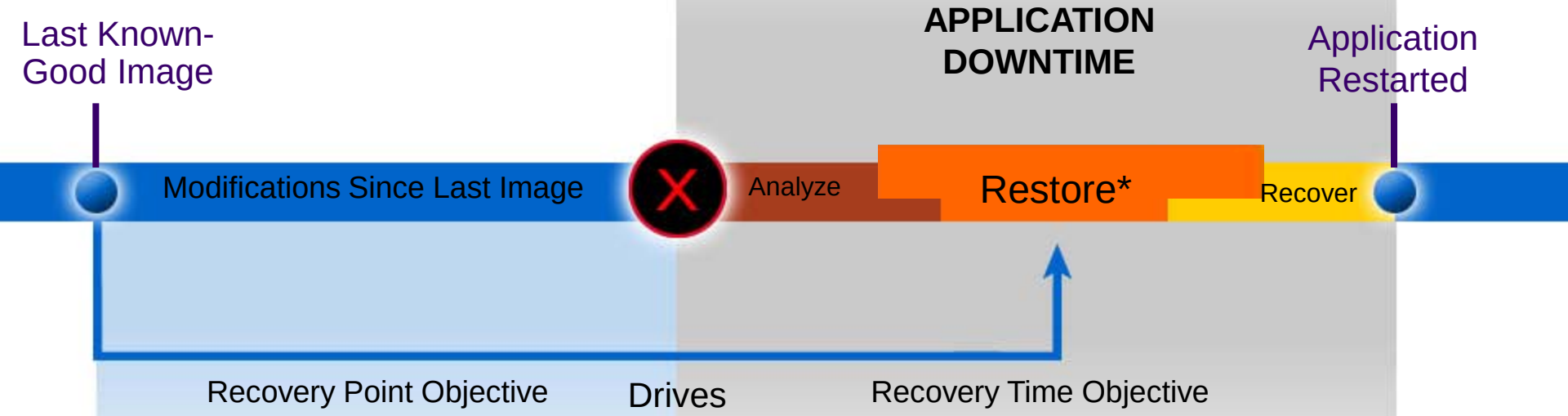


- How?
  - Block-based – changes saved at the block level
  - File-based – file level changes are logged & recoverable
  - Application-based logging is not quite CDP

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

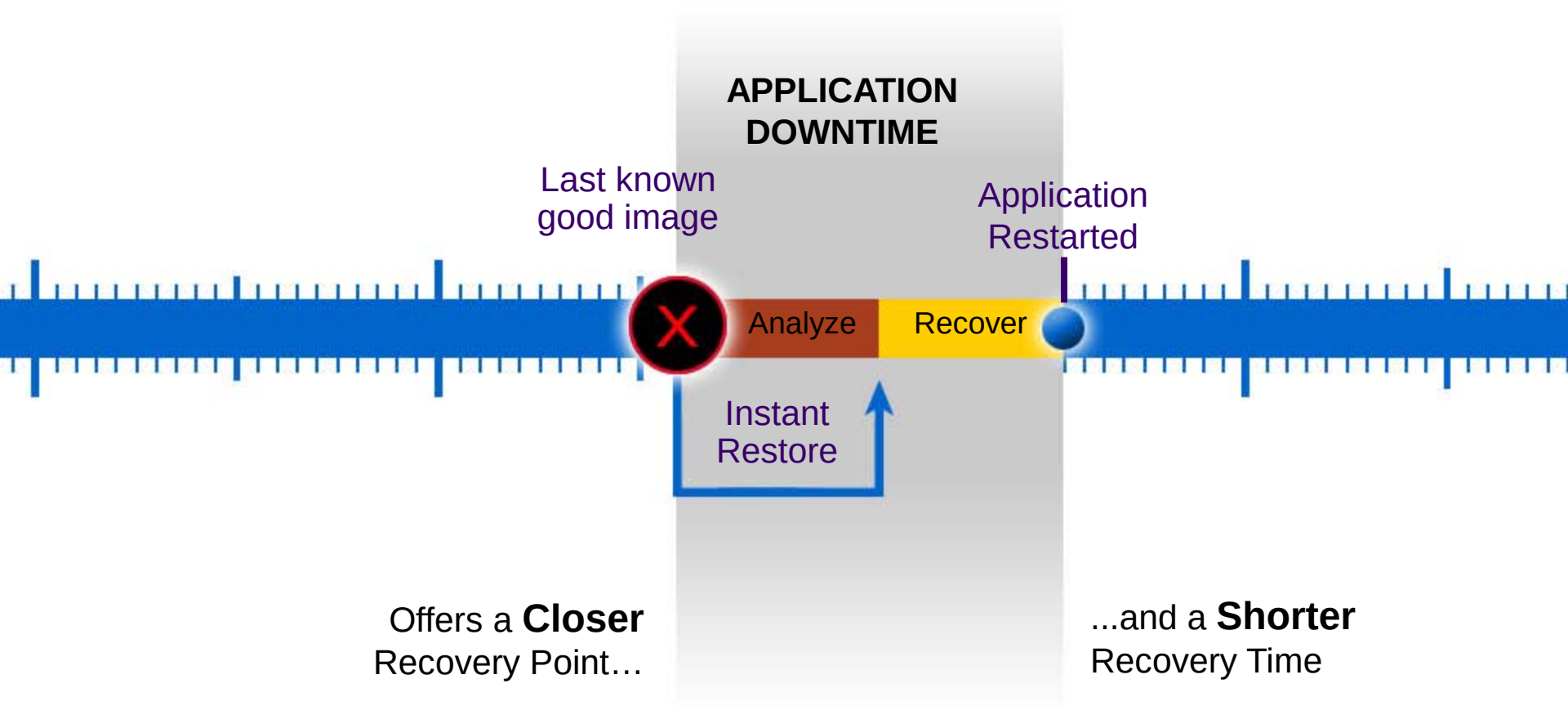
Traditional Recovery



*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

Recovery with CDP

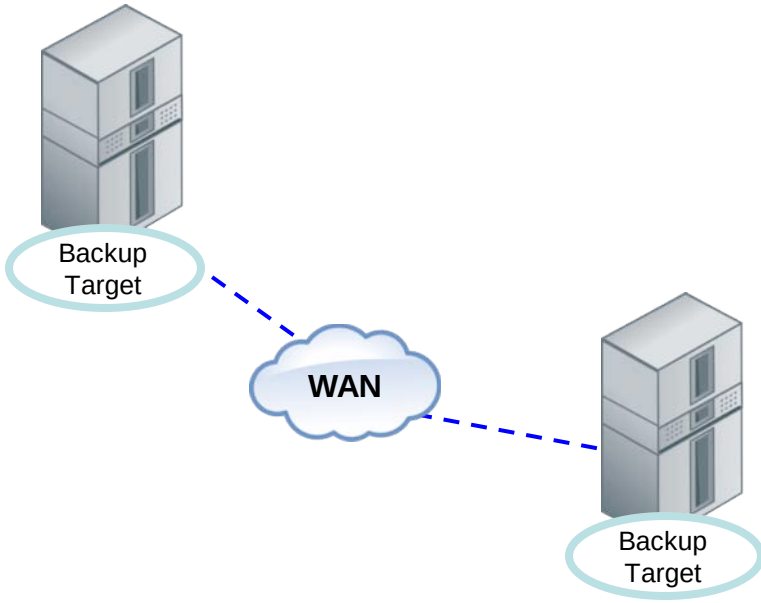


Offers a **Closer** Recovery Point…

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

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?



- Considerations
  - Network bandwidth - Leverage deduplication, network devices, "fat" pipes
  - Recovery process and resources
  - Leveraging the copies (data mining oportunities)
  - Backup versus archival

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…..

For more on Data Protection….

Hands-On LAB COMPUTERWORLD SNIA SNW

Check out the Hands-On Lab

Rethinking Archiving

Advanced Data Reduction Concepts

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: 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
Mike Fishman
Jason Iehl
SW Worth

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