



Education

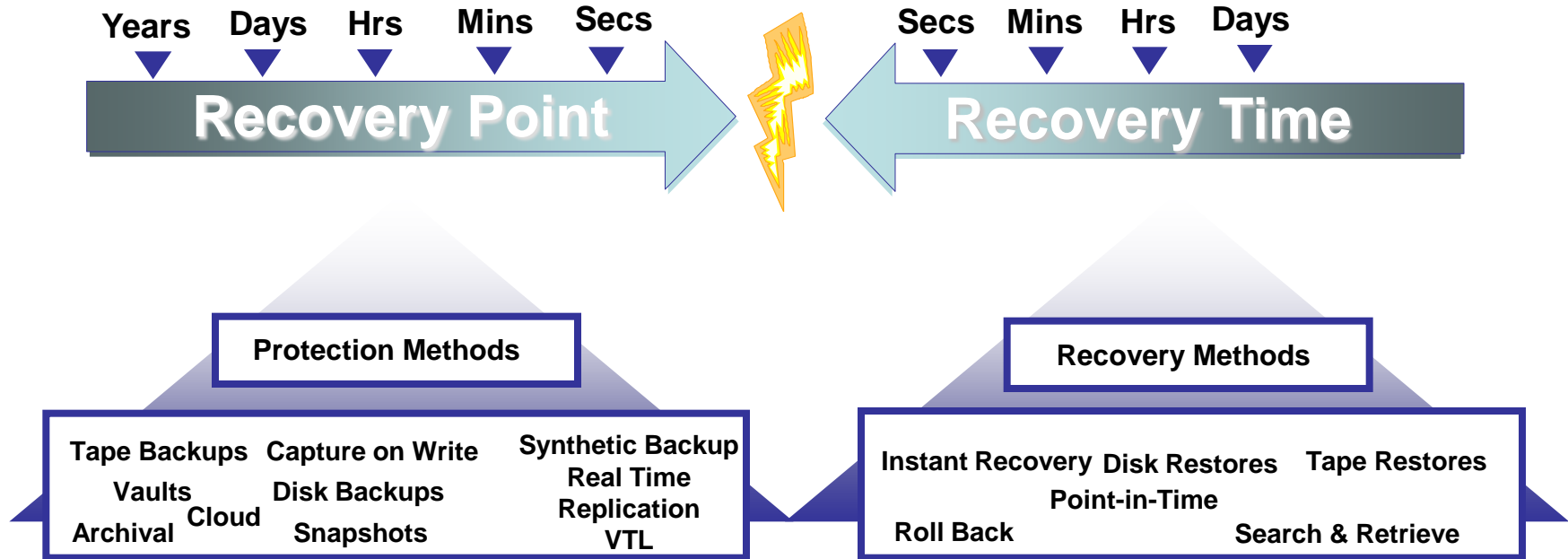
# Trends in Data Protection and Restoration Technologies

Mike Fishman, EMC<sup>2</sup> Corporation

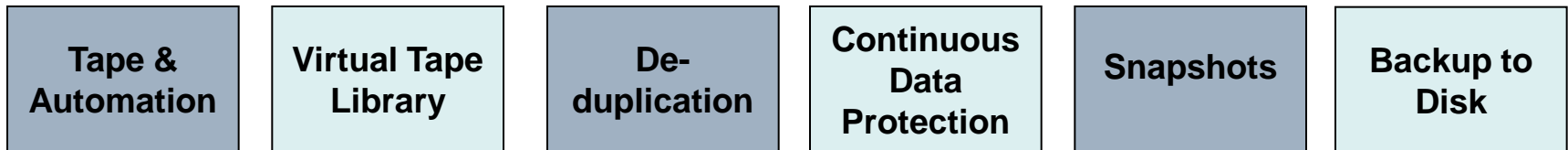
- ❖ The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- ❖ Member companies and individual members may use this material in presentations and literature under the following conditions:
  - ◆ Any slide or slides used must be reproduced in their entirety without 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.**

- 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



## ➤ 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

- 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

- 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



**Check out SNIA Tutorial:  
Trends in Application  
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 Devices

## ➤ Data Management

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

## ➤ Tiered Storage

- ◆ Primary
- ◆ Secondary
- ◆ Archive
- ◆ Backup
- ◆ Cloud

## ➤ 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

	<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>

- 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

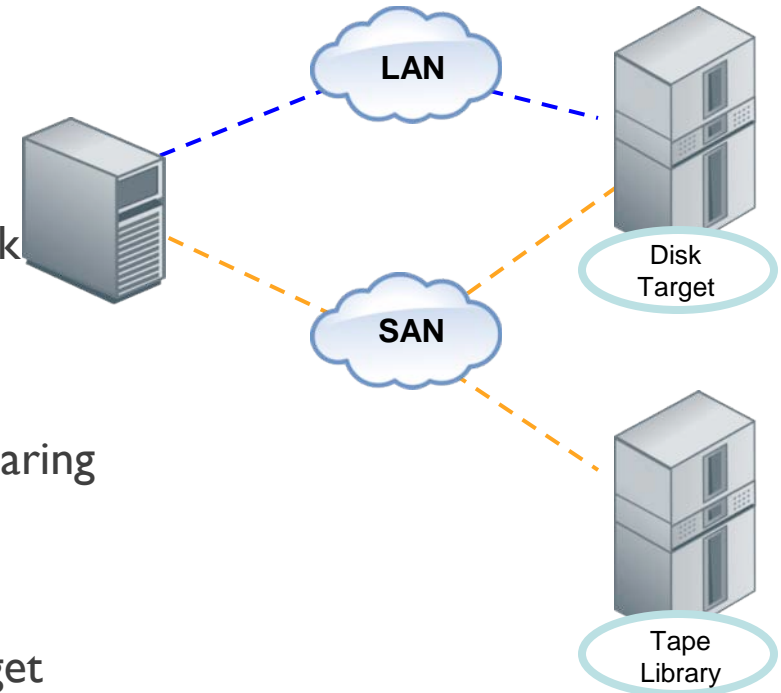
- ◆ Replace tape targets with disks
- ◆ Perform snapshot-based backups
- ◆ Recover from data on-line versus on-truck

## ➤ 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 backup requirements



# Virtual Tape Library (VTL)

## ➤ What?

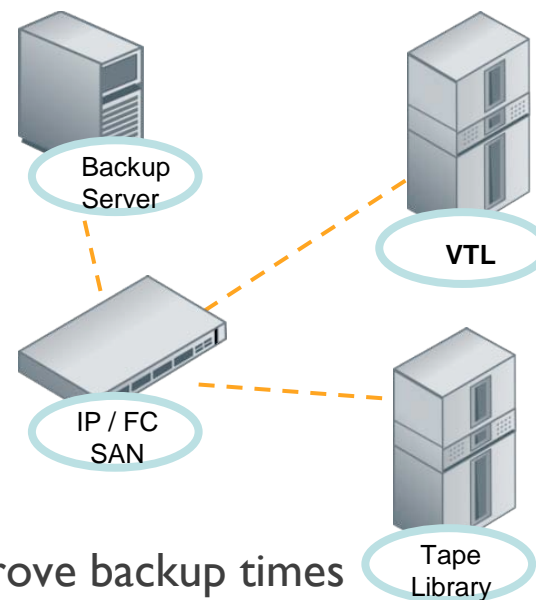
- ◆ Virtual Tape Libraries emulate traditional tape
- ◆ Easy to deploy and integrate
- ◆ Reduce / eliminate tape handling

## ➤ Why?

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

## ➤ Considerations

- ◆ Integration with physical tape
- ◆ Consider total aggregate speed as well as speed per-drive



- 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
- ◆ 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

## ➤ What?

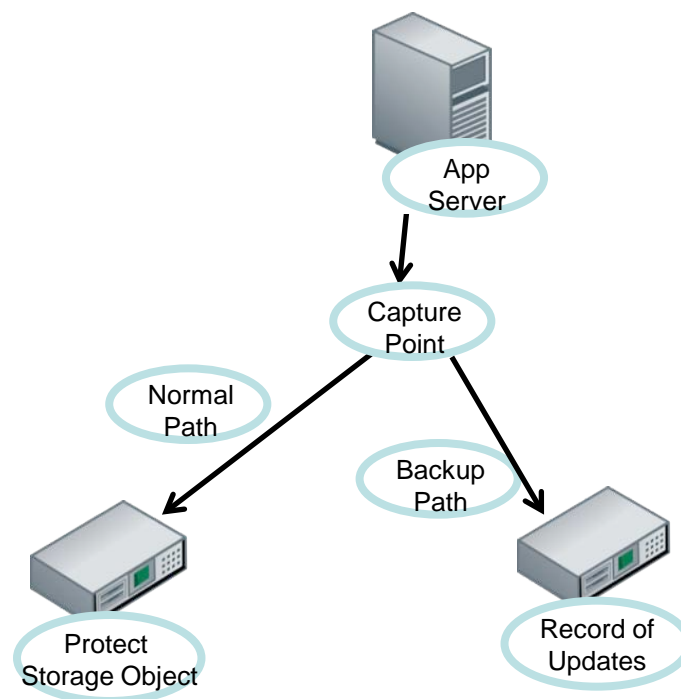
- ◆ Capture every change as it occurs
- ◆ Data can be collected and stored externally
- ◆ “Roll-back” to any point in time

## ➤ Why?

- ◆ Zero data loss, Zero backup window
- ◆ Granular recovery to any point in time
- ◆ Continuous protection

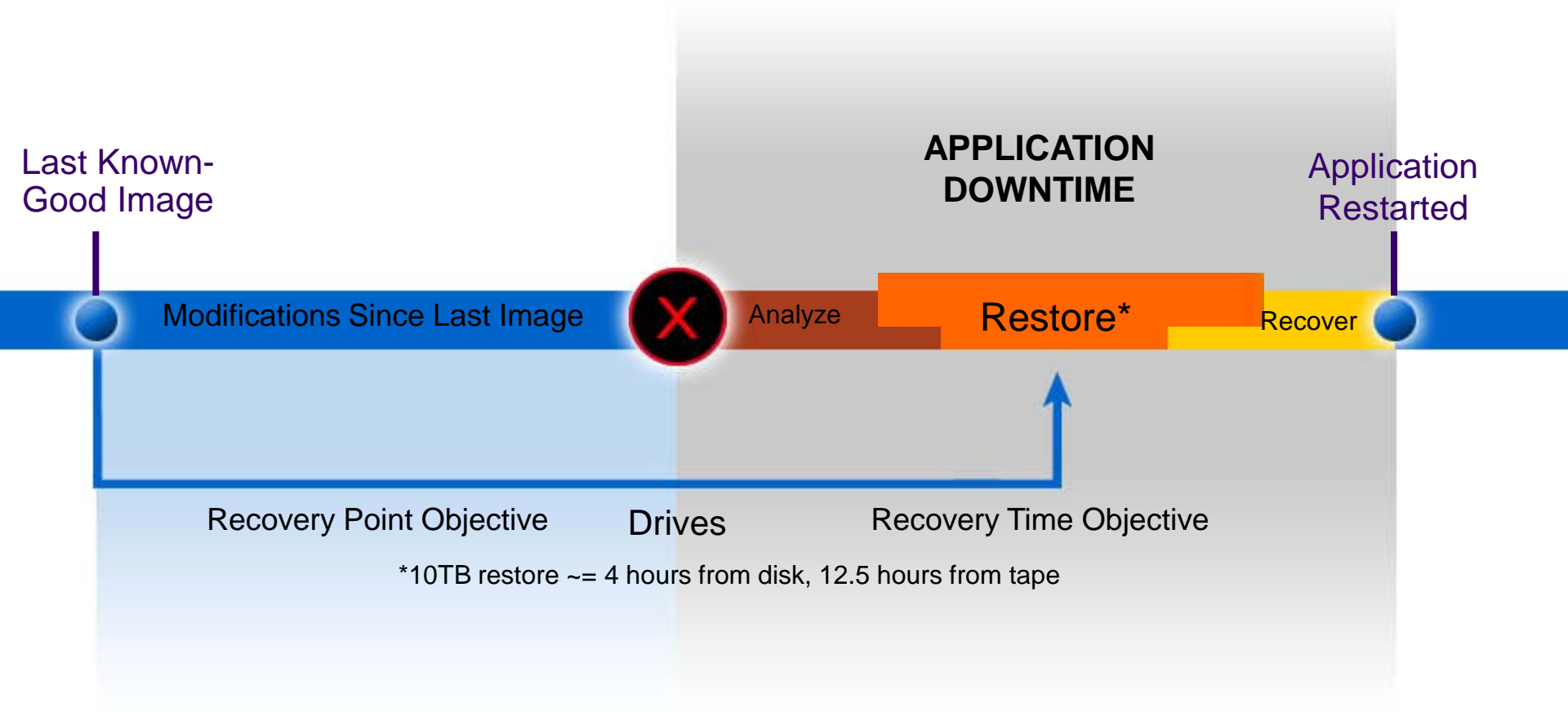
## ➤ How?

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

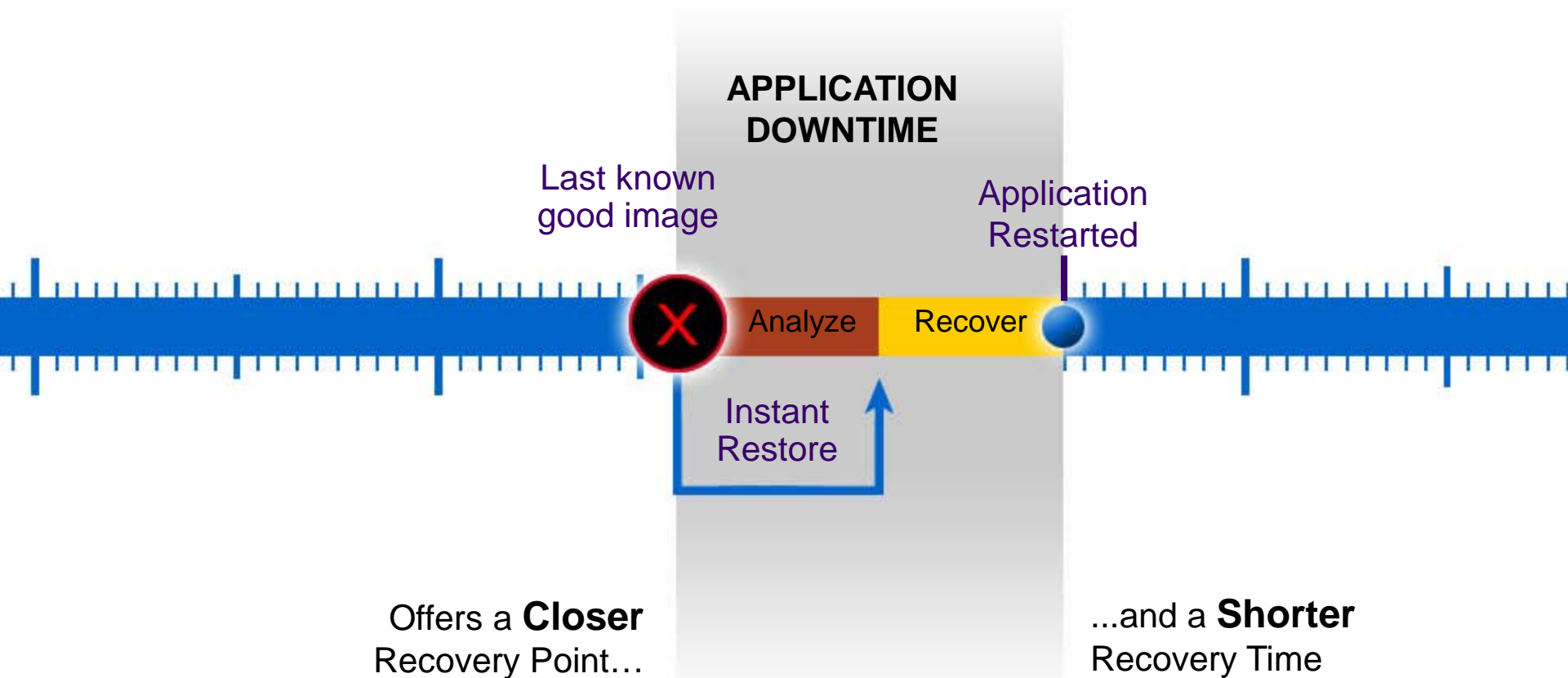


- Replication is not CDP (Synchronous)
  - ◆ Replica or mirror is a single PIT copy of the data
  - ◆ Multiple replicas plus logs can create multiple points in time
  
- Snapshots are not CDP (Asynchronous)
  - ◆ Data loss possible if crash or corruption happens between snaps
  - ◆ Snapshots frequently 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



# Recovery with CDP



## ➤ What

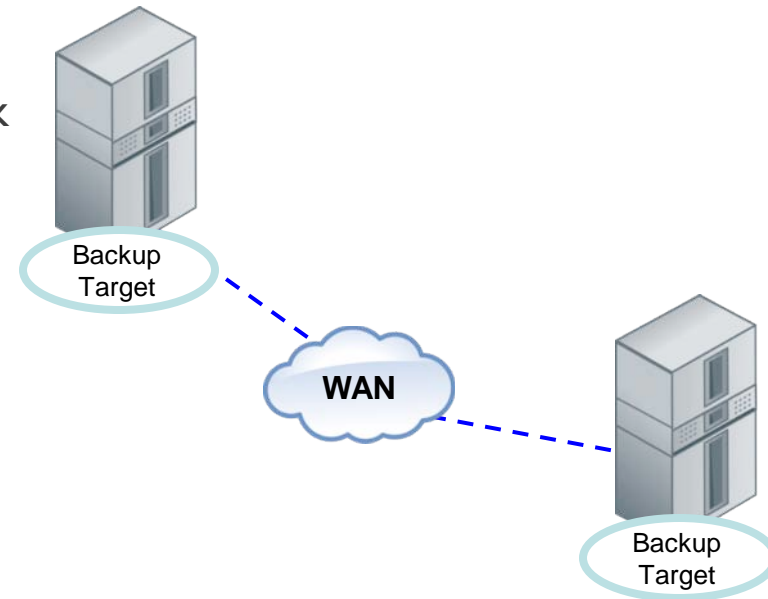
- ◆ Multiple copies of backup data
- ◆ Copy tape-to-tape, disk-to-tape, disk-to-disk
- ◆ Replication can be network based

## ➤ 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
- ◆ 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**  
**Enterprise Content Management, Cloud Storage**  
**Advanced Deduplication Concepts**  
**Deduplication of Primary Data Panel Discussion**

- Please send any questions or comments on this presentation to SNIA: [trackdatamgmt@snia.org](mailto: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 lehl  
SW Worth**

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