



Education

# Trends in Data Protection and Restoration Technologies

Jason Iehl, NetApp  
Andreas Schwegmann, HP

# SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - ◆ Any slide or slides used must be reproduced without modification
  - ◆ The SNIA must be acknowledged as 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 nor should be construed as legal advice or opinion. If you need legal advice or legal opinion please contact an attorney.
- The information presented herein represents the author's personal opinion and current understanding of the 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.**

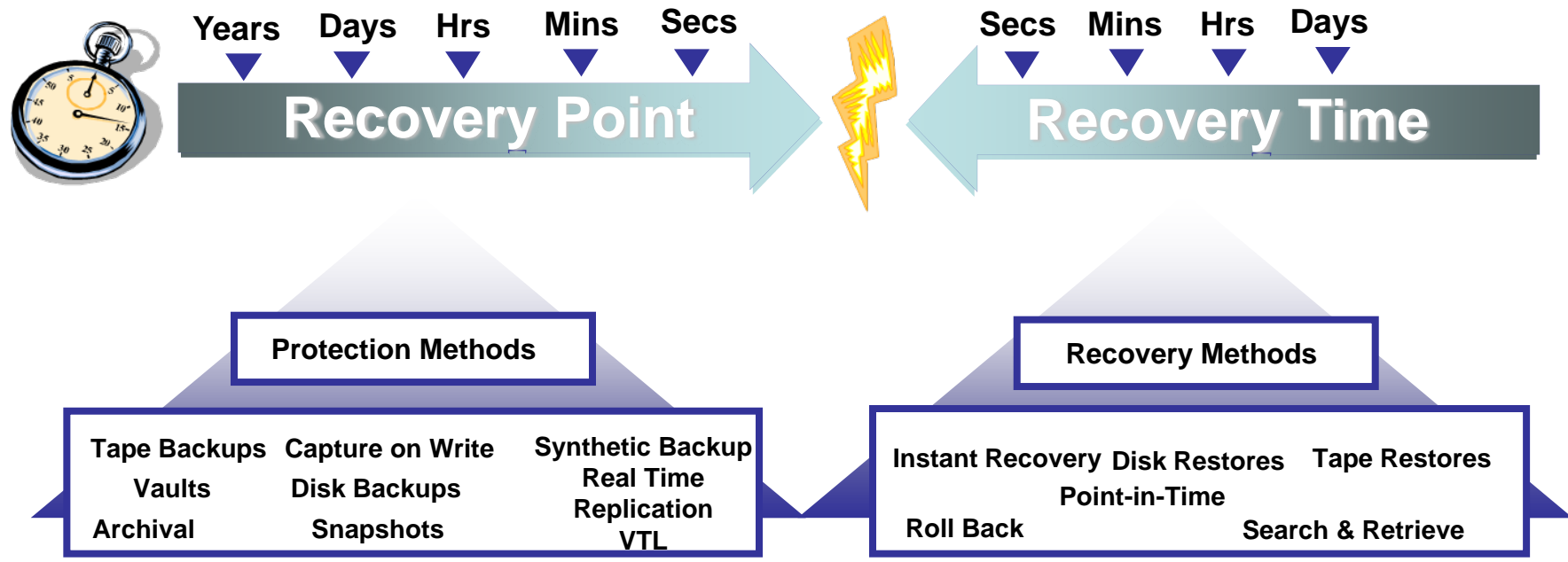
Many disk technologies, both old and new, are being used to augment tried and true backup and data protection methodologies to deliver better information and application restoration performance. These technologies work in parallel with the existing backup paradigm,

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 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



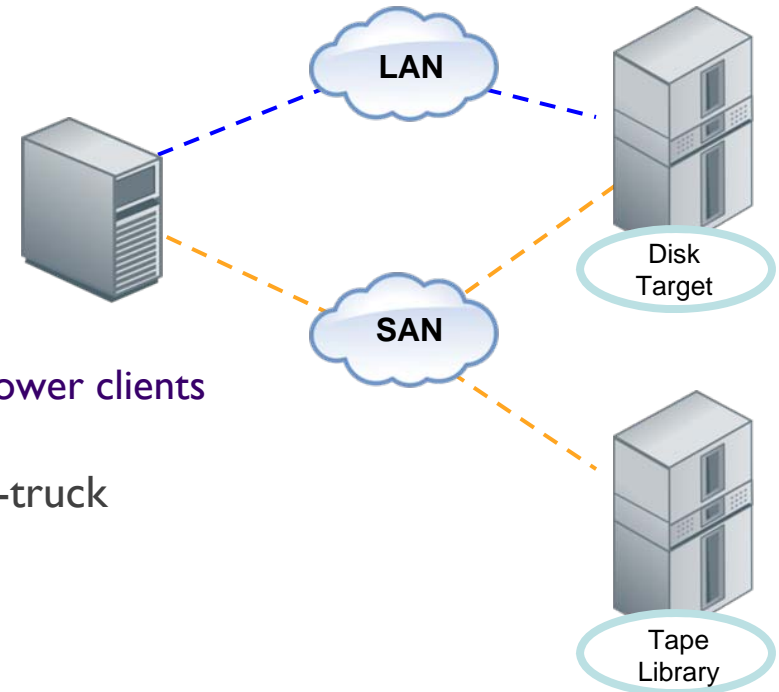
## Enabling Technologies



# 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”

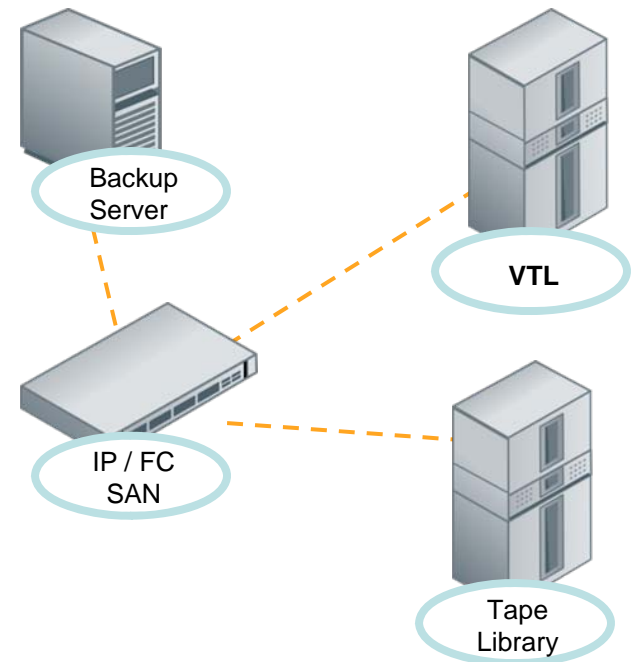
# Virtual Tape Library (VTL)

## 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
  - Backup savesets are highly redundant



# 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

# What are Snapshots

A disk based “instant copy” that captures the original data at a specific point in time. Snapshots can be read-only or read-write.

“A type of **point in time copy** that preserves the state of data at an instant in time, by storing only those blocks that are different from an already existing full copy of the data.”

***[www.snia.org/dictionary](http://www.snia.org/dictionary)***



# Snapshot of Networked Storage

## Terminology:

- Snapshot, 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

	<b>Full Copy Snapshot</b>	<b>Differential Copy Snapshot</b>
<b>Upsides</b>	<ul style="list-style-type: none"> <li>◆ No cost during “snapshot” process</li> <li>◆ Can be used for DR - independent copy</li> </ul>	<ul style="list-style-type: none"> <li>◆ Less storage consumption - typically 10-20%               <ul style="list-style-type: none"> <li>◦ Depends on change rate</li> </ul> </li> <li>◆ Typically can take advantage of cheaper disk</li> </ul>
<b>Downsides</b>	<ul style="list-style-type: none"> <li>◆ Massive storage cost               <ul style="list-style-type: none"> <li>◦ 1x of storage per RPO</li> <li>◦ Like disk - expensive</li> </ul> </li> <li>◆ Often in the same disk chassis               <ul style="list-style-type: none"> <li>◦ Loss of DR component</li> </ul> </li> <li>◆ Consider re-sync time in schedules</li> </ul>	<ul style="list-style-type: none"> <li>◆ Performance may be impacted while snapshot exists               <ul style="list-style-type: none"> <li>◦ Multiple implementations to optimize performance impact</li> <li>◦ Most vendors don't offer multiple implementations - pick at onset</li> </ul> </li> <li>◆ Leverages main copy - not DR capable</li> </ul>
<b>Applications</b>	<ul style="list-style-type: none"> <li>◆ Disaster Recovery</li> <li>◆ Near zero backup window               <ul style="list-style-type: none"> <li>◦ 24x7 operations</li> </ul> </li> <li>◆ Faster restore               <ul style="list-style-type: none"> <li>◦ Can do no-copy restore</li> </ul> </li> <li>◆ Can help with data repurposing</li> </ul>	<ul style="list-style-type: none"> <li>◆ Backup source</li> <li>◆ Near zero backup window               <ul style="list-style-type: none"> <li>◦ 24x7 operations</li> </ul> </li> <li>◆ Fast restore               <ul style="list-style-type: none"> <li>◦ copy based by definition</li> </ul> </li> <li>◆ Can help with data repurposing               <ul style="list-style-type: none"> <li>◦ Beware performance impact</li> </ul> </li> </ul>

# 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
- ILM – Are you snapping old, un-used data

## 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



**Check out SNIA Tutorial:  
Understanding Data Deduplication**

# Deduplication: What to Consider

- Factors that will impact your results:
  - ◆ Different applications or data types
  - ◆ Bandwidth and latency
  - ◆ Policies and Methodologies
  - ◆ Data Protection Overhead
  - ◆ Compression and Encryption
- Global Deduplication and Scope
- Deduplicated Data Resiliency
- Scalability
  - ◆ Capacity
  - ◆ Performance

# Backup: Factors Impacting Space Savings

<b>Factors associated with higher data deduplication ratios</b>	<b>Factors associated with lower data deduplication ratios</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>Wider scope of data deduplication</b>	<b>Narrower scope of data deduplication</b>
<b>Continuous business process improvement</b>	<b>Business as usual operational procedures</b>
<b>Smaller segment size</b>	<b>Larger segment size</b>
<b>Variable-length segment size</b>	<b>Fixed-length segment size</b>
<b>Format awareness</b>	<b>No format awareness</b>
<b>Temporal data deduplication</b>	<b>Spatial data deduplication</b>

# Continuous Data Protection

## What:

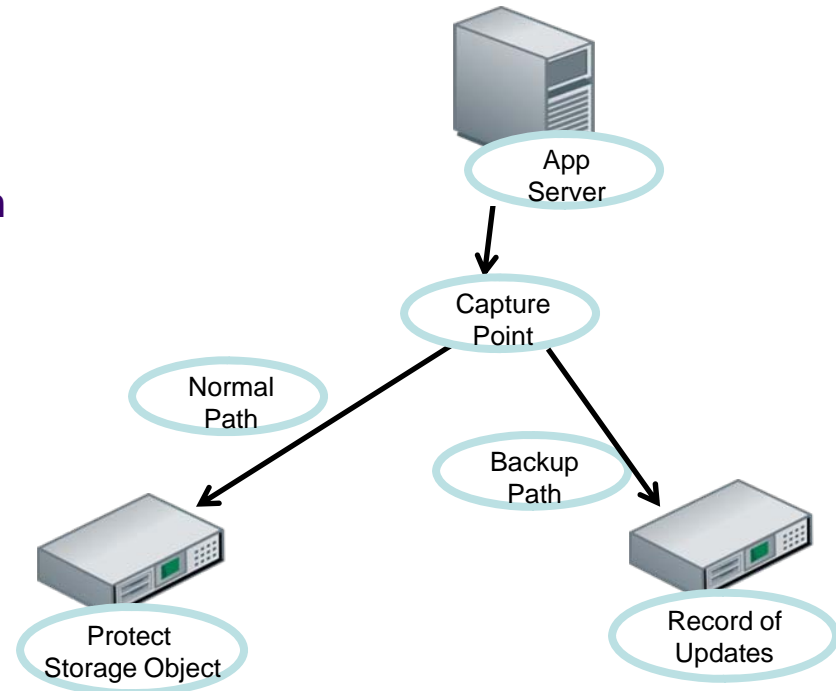
- Capture every change as it occurs
- Protected copy in a secondary location
- Recover to any point in time

## How:

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

## Why:

- Zero data loss, Zero backup window
- Simple recovery
- CDP protects data at all times
- Granular recovery - directly to any point in time

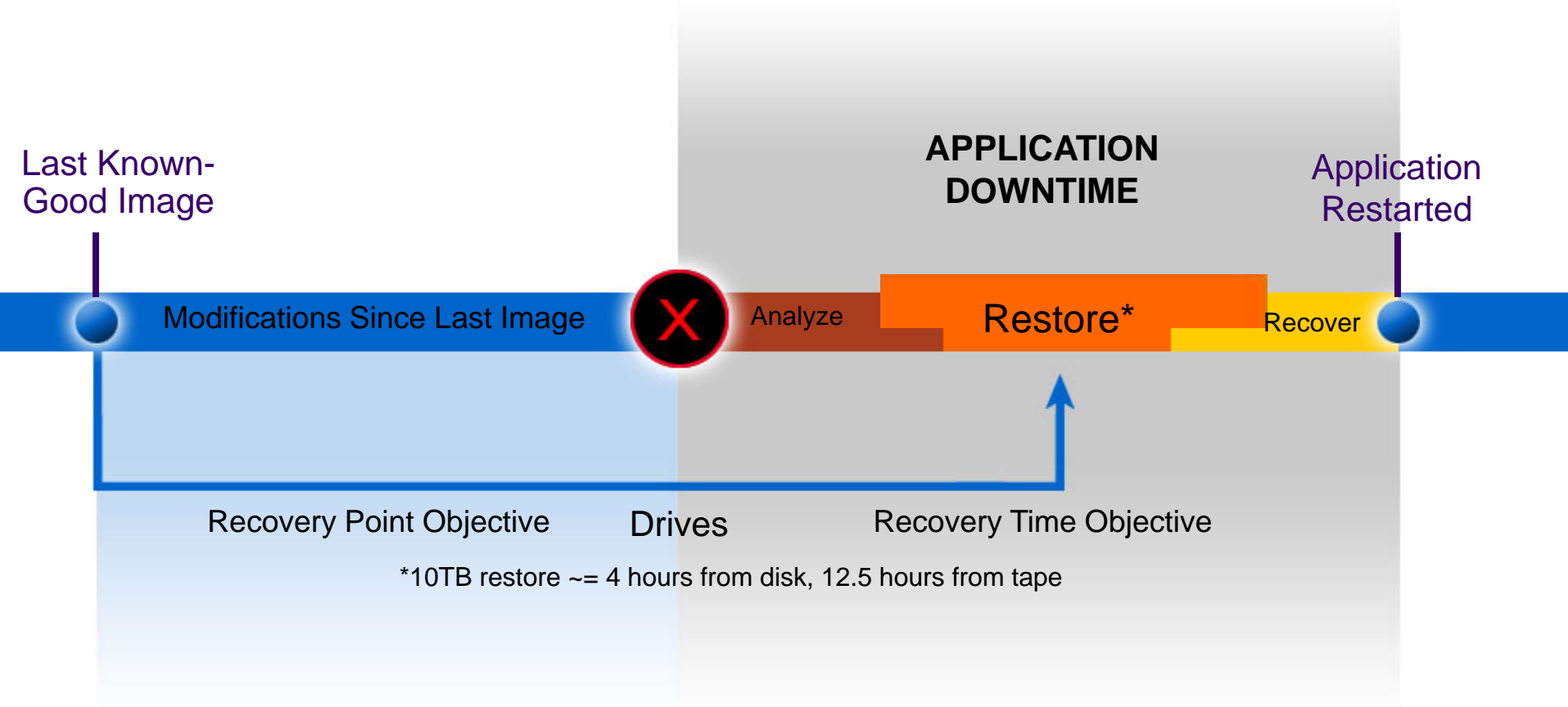


# The CDP (Continuous) Difference

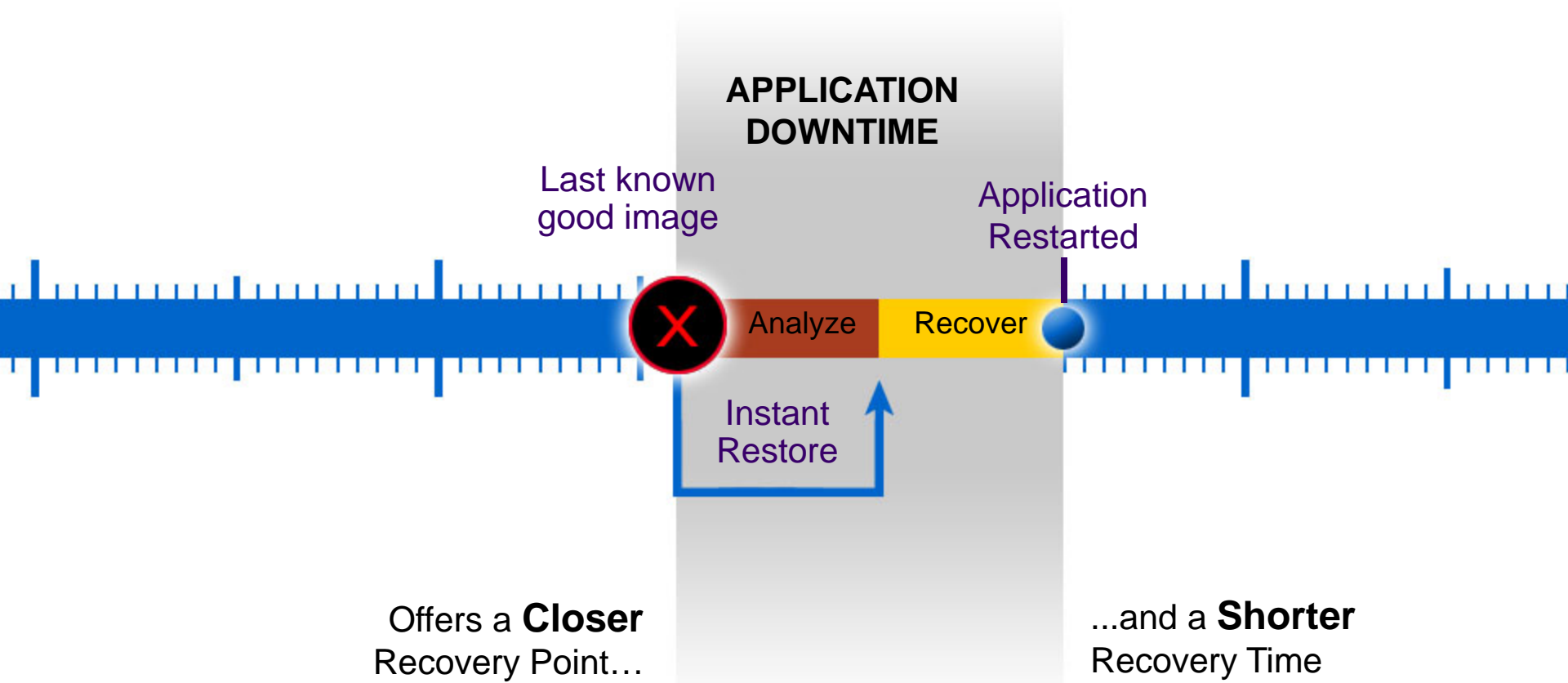
- Replication is not CDP:
  - ◆ Maintains only a current copy of the data (Synchronous)
  - ◆ May be combined with some snapshot capabilities
  
- Snapshots are not CDP:
  - ◆ Snapshots are scheduled events (Asynchronous)
    - › Data loss possible if crash or corruption happens between snaps
    - › 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



# Recovery with CDP



# Data protection summary

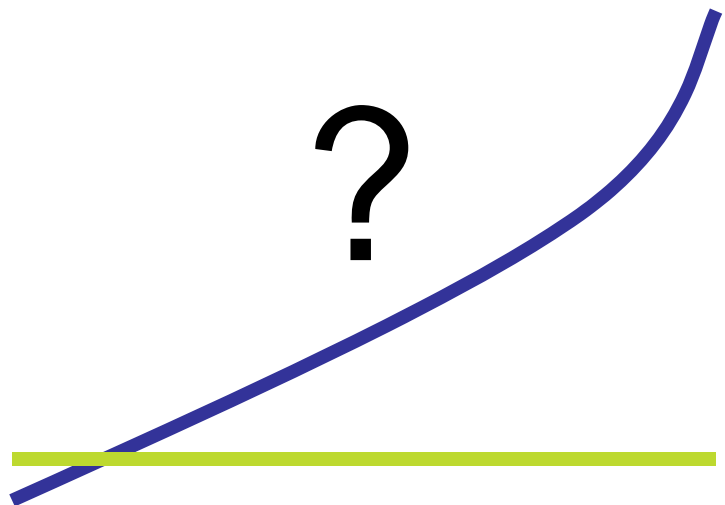
- Data growth requires us to plan for tomorrow
  - ◆ Investigate data and information management technology
- Information value determines data protection levels
  - ◆ Stop protecting employee home movies, last years news
  - ◆ Not all data assets are created equal
- Architecture
  - ◆ Applications are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

# Application Recovery



- Challenges and trends
- What to recover and by whom?
- Architectures
- Application backup
  - ◆ Consistency
  - ◆ Backup window
- Interfaces
- Recovery
  - ◆ Application entirely
  - ◆ Single items
- Virtualization
- Conclusion

# Application Trends



- Blue line - accelerated data volume growth & complexity growth = the bullets on this slide
- Green line – IT budget is flat

## ➤ Accelerated data growth

- ◆ Cheaper storage
- ◆ More performance
- ◆ Long term retention
- ◆ Distributed data -> duplicate data

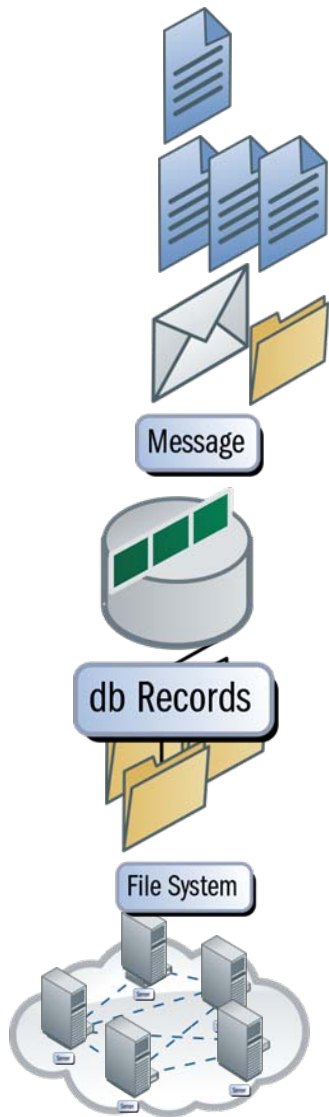
## ➤ Complexity growth

- ◆ Scale up & out
- ◆ Different data structures & type in one application
- ◆ High availability
- ◆ Virtualization
- ◆ Search
- ◆ Security
- ◆ Features

# Application Recovery Trends

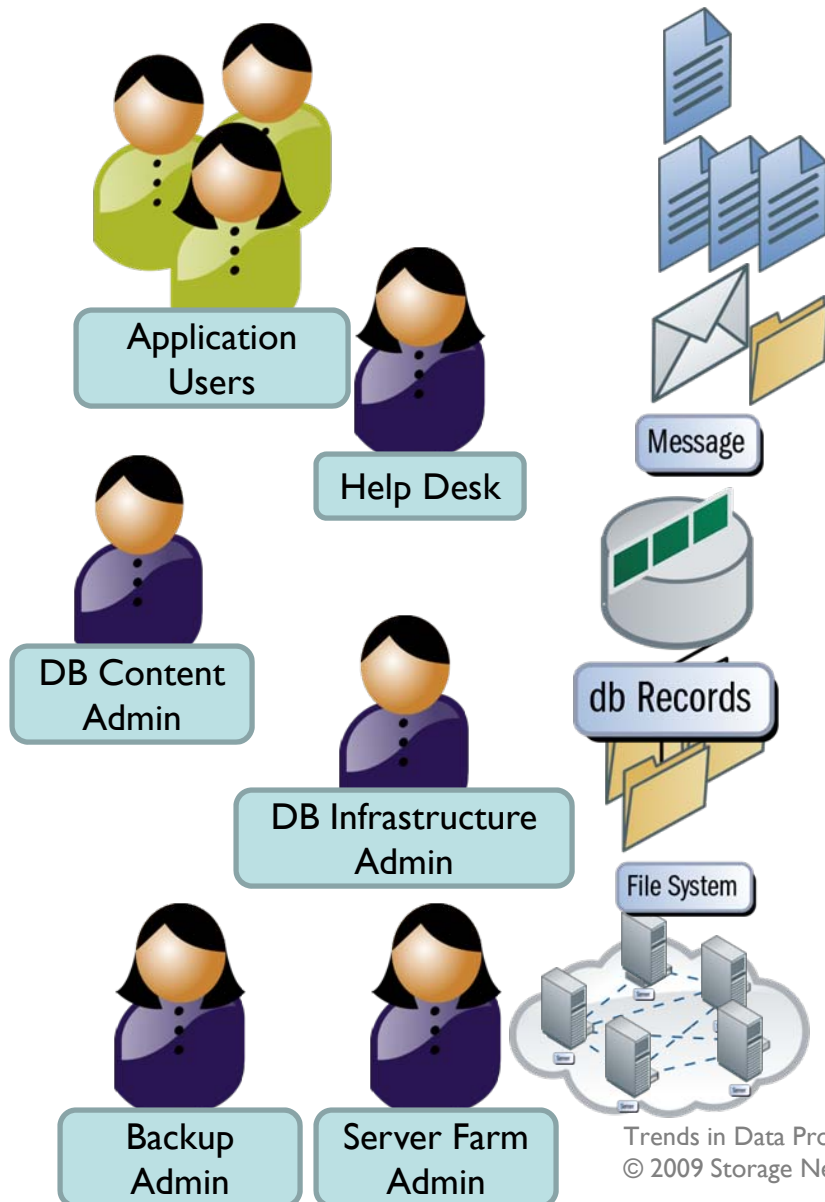
- Recovery to any point in time
- Snapshot based recovery
- Seamlessly link disk with tape
- Data reduction
  - ◆ Incremental and consolidation of incremental backups
  - ◆ Deduplication, primarily of repeated full backups
- Recovery automation
  - ◆ The know how to recover the application is build into the backup software
- Server farm down to single document, mail, ...
- Different user roles

# What to Recover?



- Single items / end user domain
  - ◆ Entry, record, transaction
  - ◆ Document, e-mail, blog
  - ◆ Calendar, tasks, contacts
  - ◆ Table, list, tree, folder, wiki
  - ◆ E-mail box, user site
- Database
  - ◆ Files system or raw disk
  - ◆ Cache
- Search Index & Services
- Application configuration
- Binaries, OS, configurations
- Server farm

# Who Does What?



- Different users groups use different user interfaces
  - ◆ Users and administrators don't want more tools, they want to manage from their tools.
- Security
  - ◆ User A should not be able to see data from user B.
  - ◆ The administrator should not be forced to break into the end user privacy.
  - ◆ DB administrator might not have the right to restore from backup.
- Processes
  - ◆ Application administrators might not be connected with backup administrators.



# Application Architectures

- Single server, single database
- Server farm with multiple databases of different types
- Database across databases – search index
- Databases with links into the file system
  - ◆ Using standard files
  - ◆ BLOBs
- Cluster: 2 –N nodes
  - ◆ Load sharing, fail over, majority node, ...
- Replication
  - ◆ Active and passive databases
- DAS, SAN

# Full & Incremental Backup

- Full backup – file based or block based
  - ◆ Database
    - › Data files = “tables“ mapped to files
    - › Control files to find data files & log files
    - › Transaction log files, optional: move / delete
  - ◆ Trend features
    - › Files, BLOBs – typically used for storing large files outside of the DB
    - › Search Index, services, encryption keys, ...
- Incremental backup – changes since last backup
  - ◆ File based
    - › Transaction log backup and move / delete
    - › Files: File system incremental backup of single instance file store
  - ◆ Changed blocks
- Differential backup – changes since last full backup
  - ◆ Higher deduplication rate than incremental backup

# 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 N snapshots
  - ◆ Snapshot rotation

# Crash Consistent Backup

- 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

# Consistency - Online Backup

- 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

# Application Backup Interfaces

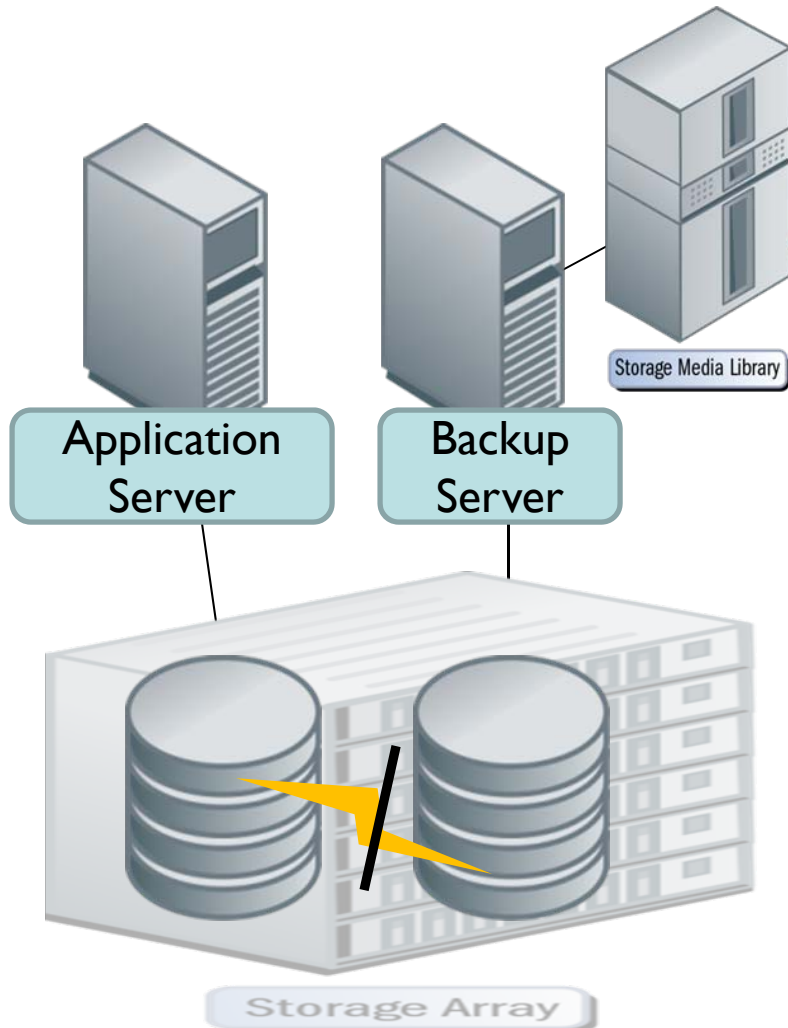
- Application specific tools via GUI, CLI
- General purpose API
- Streaming backup API
  - ◆ Direct copy
    - > Access to in-memory copy of data, cached by the application
    - > Minimizes redundant memory copies
  - ◆ Incremental backup
    - > Access to changed blocks / pages or transactions
  - ◆ Optional features
    - > Granularity below database level
    - > Compression
    - > Encryption
  - ◆ Sequential access is optimal for streaming media

# Volume based Backup API

- Creates application consistent volumes ready for backup
- Use cases
  - ◆ Copy the entire volume via snapshot
  - ◆ Copy all files needed to recover the application
  - ◆ Incremental backup
    - Changed blocks
    - Changed files
  - ◆ Feature set might be different compared to streaming API
  - ◆ Backup to disk & restore from disk
- Trend: volume based backup

Source: Antal Nemes, Hermes Softlab

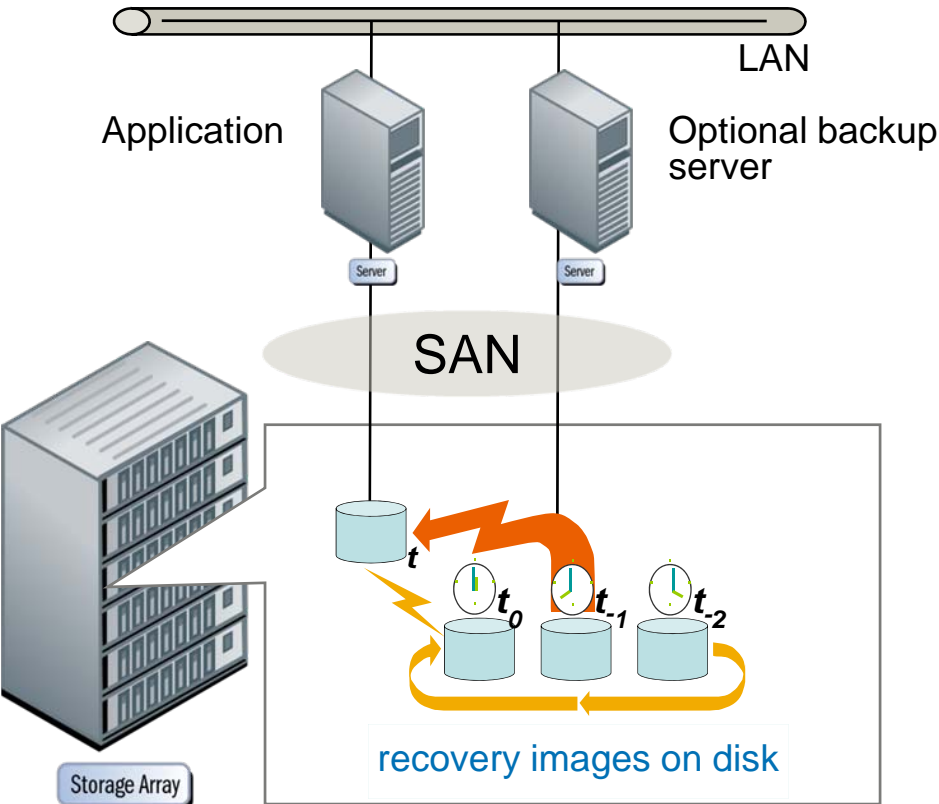
# Reduced Performance Impact



- Separate backup server
- Backup from full copy snapshot (mirror)
  - ◆ Application switches into backup mode
  - ◆ Split the mirror
  - ◆ Back to normal mode
  - ◆ Separate backup server copies the data from split mirror
  - ◆ Resync the mirror after backup
    - › Copies changed blocks only



# Application Recovery from Snapshot SNIA

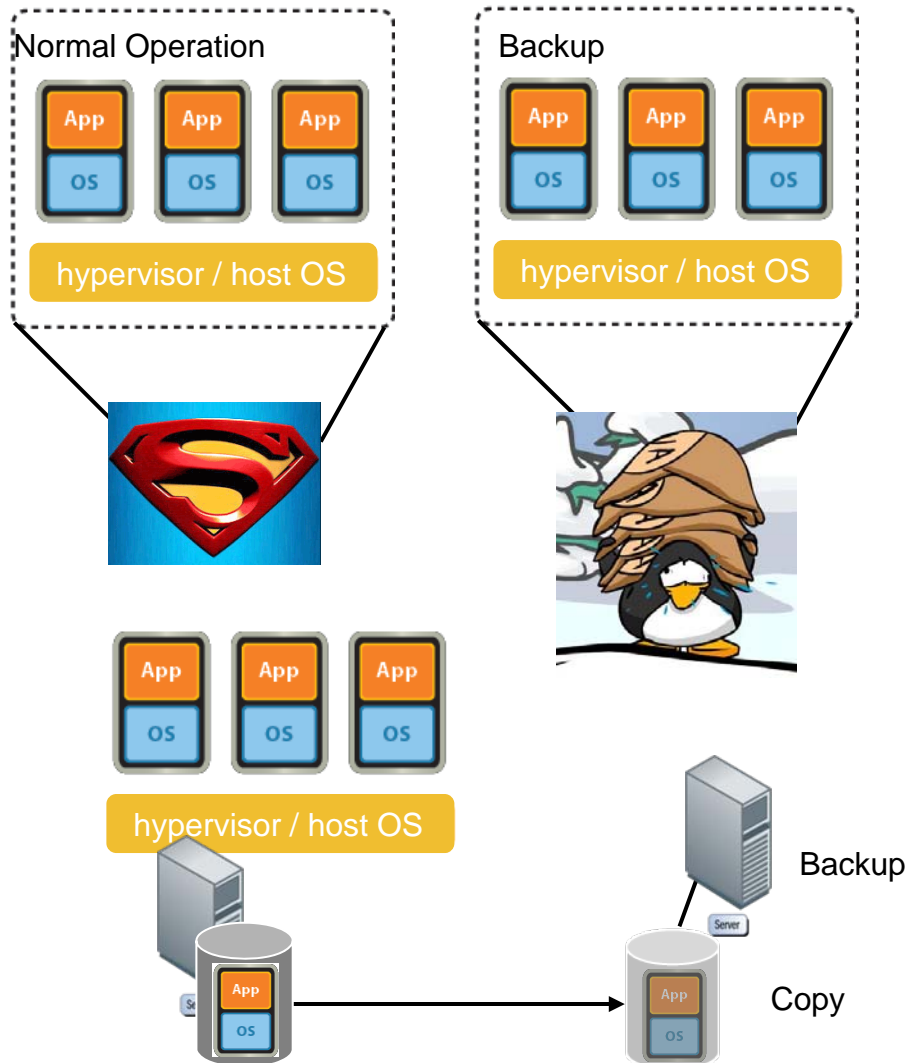


- Application shutdown
- Optional incremental transaction log backup
- Switch to selected snapshot
  - ◆ Instead of restore from tape
- Transaction log roll forward from backup or original disk up to the most recent point in time
- Application back online

# Single Item Recovery Options

- Dumpster, 2nd level dumpster, versioning, archive
- Full blown recovery environment & copy back
- The application can be used to extract single items from backup
  - ◆ Copy database from backup & mount as recovery database
    - › Needs extra space and time to copy the entire database
  - ◆ Mount the database from the backup directly into the application
- Open the backup database with a separate tool & extract
- Extract single items directly from the backup
  - ◆ Catalog of all single items during backup needed
- Single item recovery from single item backup
  - ◆ Needs a separate backup

# Server Virtualization



- Resources shared among different applications
  - ◆ Normal load spreads evenly across day / week / month
  - ◆ Backup load is exception
- Resources on physical server often not enough for backup load
- Offload backup via dedicated physical machine
  - ◆ Utilize replication



- Application specific solutions needed
  - ◆ There is no one size fits all
  - ◆ Hybrid solutions might be needed
- High complexity - many options
  - ◆ Who are your customers?
  - ◆ What are the requirements?
    - › Backup window, RPO, RTO
  - ◆ What are the application environments?
  - ◆ What are your use cases?
  - ◆ What is your budget?
  - ◆ What are the Service Level Agreements?

Please send any comments on this tutorial to SNIA at:  
[trackdatamgmt@snia.org](mailto:trackdatamgmt@snia.org)

The DMF would like to thank the following individuals for their contributions to the development of this tutorial:

SNIA Data Protection Initiative  
SNIA “Data Management Forum”  
SNIA Tech Council  
Nancy Clay  
Rob Peglar

Mike Fishman  
Jason Iehl  
Mike Rowan  
SW Worth  
Andreas Schwegmann



It's easy  
to get  
involved  
with the  
DMF !

- Find a passion
- Join a committee
- Gain knowledge & influence
- Make a difference

[www.snia.org](http://www.snia.org)

# Thank you for your feedback

## Questions and Answers