SNIA

STORAGE NETWORKING INDUSTRY ASSOCIATION

EDUCATION

Disk-based Restoration Technologies

Michael Rowan, Consultant

Incipient, Inc.

Disk-based Restoration Technologies

Many disk technologies, both old and new, are being used to augment the backup paradigm to deliver better restoration performance. These technologies work in parallel to the existing backup paradigm, sometimes synergistically and sometimes completely orthogonally. This session will discuss many of these technologies in detail. This will include snapshots, including full-copy snapshots (mirror-splits), differential snapshots, and small aperture snapshots (SAS), as well as continuous data protection (CDP) in it's various forms. We will discuss both how these technologies work as well as how they can be deployed in modern heterogeneous data centers.

SNIA Legal Notices

- 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

Storage Networking Industry Association

SNIA
Storage Networking Industry Association
EDUCATION

- SNIA is the trade group for storage networks
  - "ensuring that storage networks become complete and trusted solutions across the IT community" - http://www.snia.org
  - SNIA "Dictionary of Storage Networking Terminology" is an excellent resource http://www.snia.org/dictionary
- SNIA Tutorials are on-line:
  - http://www.snia.org/education/tutorials

About SNIA and the DMF

SNIA
Storage Networking Industry Association
EDUCATION

About the Storage Networking Industry Association (SNIA)

- SNIA's primary goal is to ensure that storage networks become complete and trusted solutions across the IT community
- For additional information about SNIA see www.snia.org
- SNIA's "Dictionary of Storage Networking Terminology" is online at www.snia.org/dictionary

About the SNIA Data Management Forum (DMF)

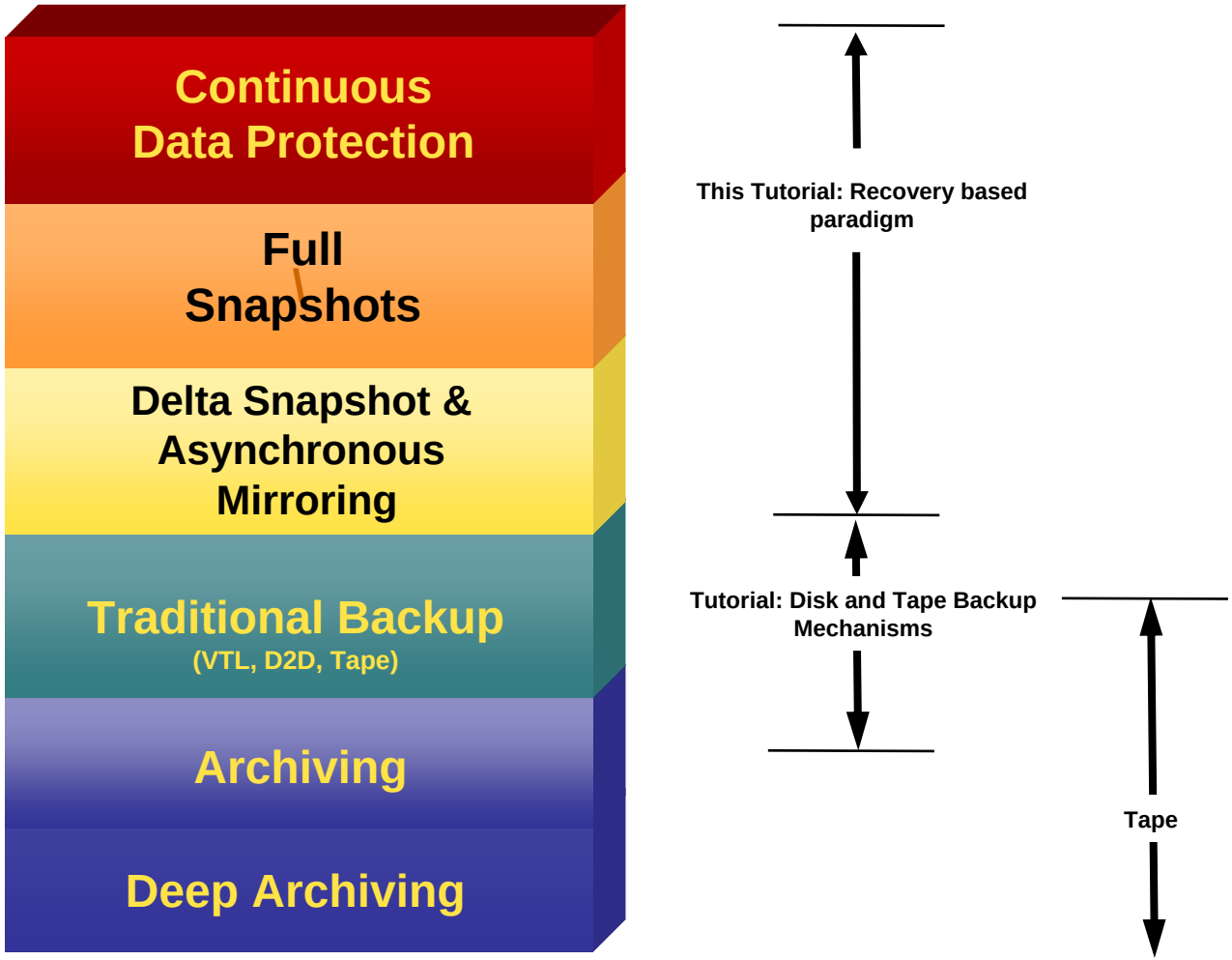
- The DMF is a sub-group of SNIA acting as the worldwide authority on Data Management, Data Protection and ILM www.snia-dmf.org
- The DMF is a collaborative storage industry resource available to anyone responsible for the accessibility and integrity of their organization's information.

DMF		
Data Protection Initiative (DPI)	**Information Lifecycle Management Initiative (ILMI)**	**Long term Archive and Compliance Storage Initiative (LTACSI)**
Defining new approaches and best practices for data protection and recovery	**Developing, teaching and promoting ILM practices, implementation methods, and benefits**	**Addressing challenges in developing, securing, and retaining long-term digital archives**

Disk-based Data Protection

- Data Protection is about **Data Availability**
  - **Backup is NOT a business requirement -- Availability is.**
  - Data protection as an extension of data availability - Restoration focused
- Disk-Assisted and Disk-based protection methods
  - Speed
    - Backup Windows
    - Recovery Time Objectives
  - Random Access
    - Data set, application object or sub-object recovery - quickly
  - ILM
    - More gradations in lifecycle (B&W -> shades of grey)
- Disk based protection does not include **"Deep Archival"**
  - Tape is not dead!

The Operational Recovery Spectrum



Agenda

- Making Protective Copies
  - Cold, Hot or Atomic: Options to understand
- Snapshots
  - Overview
  - Full Copy Snapshots
  - Differential Snapshots
    - RoW, CoW, WA: Approaches and benefits
- Continuous Data Protection
  - Overview
  - Approaches
    - Block, file, host-based, application-based, network-based
    - Implementation examples

Application Consistency

SNIA
Storage Networking Industry Association
EDUCATION

When an application is running during the "copy" process, various techniques are available to ensure data consistency

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).

It is important to understand the consistency semantics of your application so that your data protection copies are recoverable.

To Quiesce or Not?

- Cold Snapshot
  - Less complex, but backup window is downtime
- Application Consistent Snapshot
  - Application intervention
  - Application dependent
  - “Hot Backup” or “Online Backup”
- Atomic or Crash Consistent Snapshot
  - Ability to take snapshot for entire dataset at exactly same moment
  - Can be done in multiple ways
  - Recovery domain same as high availability systems

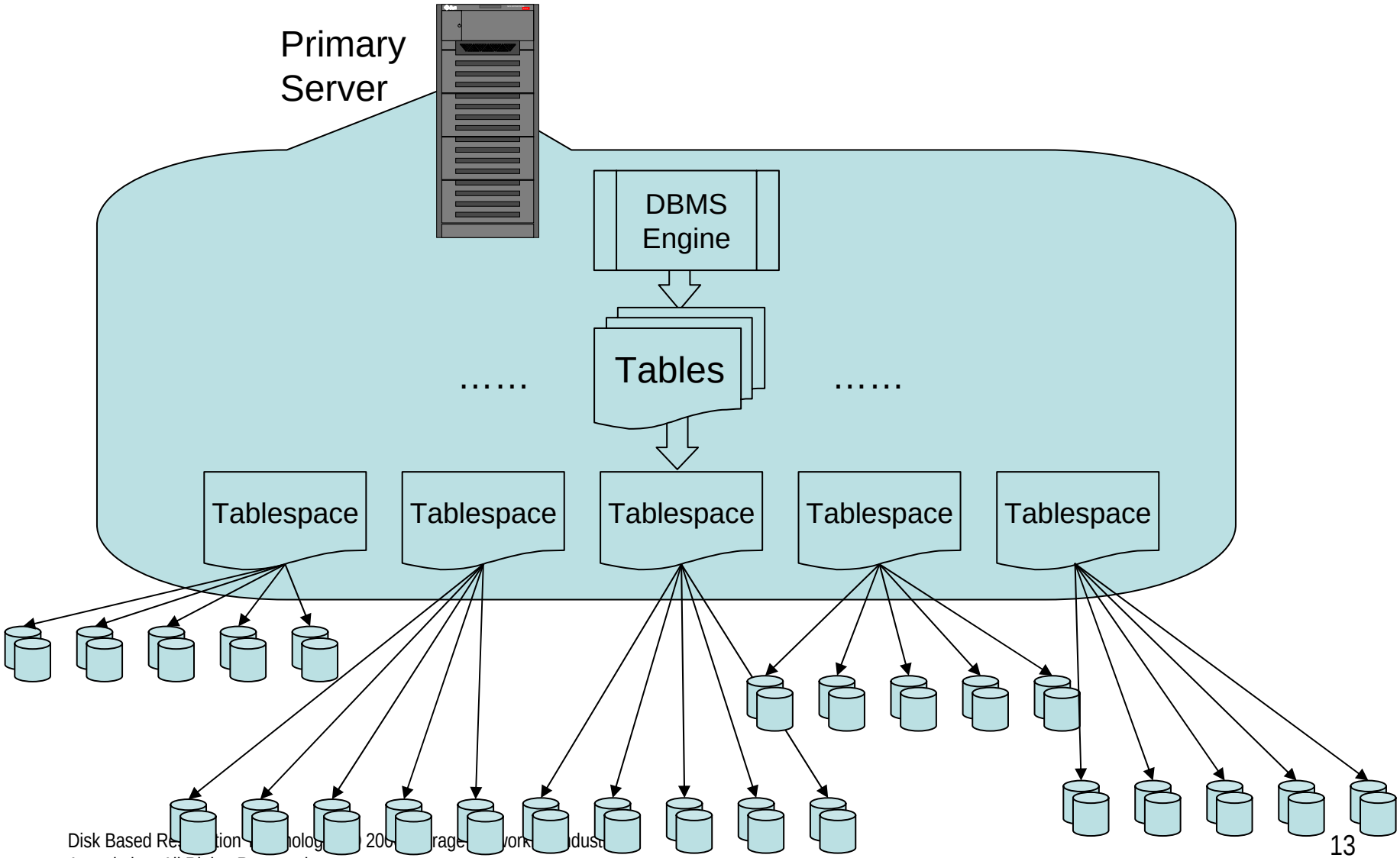
Cold Protective Copy

- Old School approach:
  - Bring application down
  - Create copy (or mark timeline as a “cold moment”)
  - Bring application back online
- Requires downtime
  - How much is dependent on technology, implementation and environment
    - Can be from seconds to hours
  - This is a true “backup window”
    - Actual downtime during backup copy operation

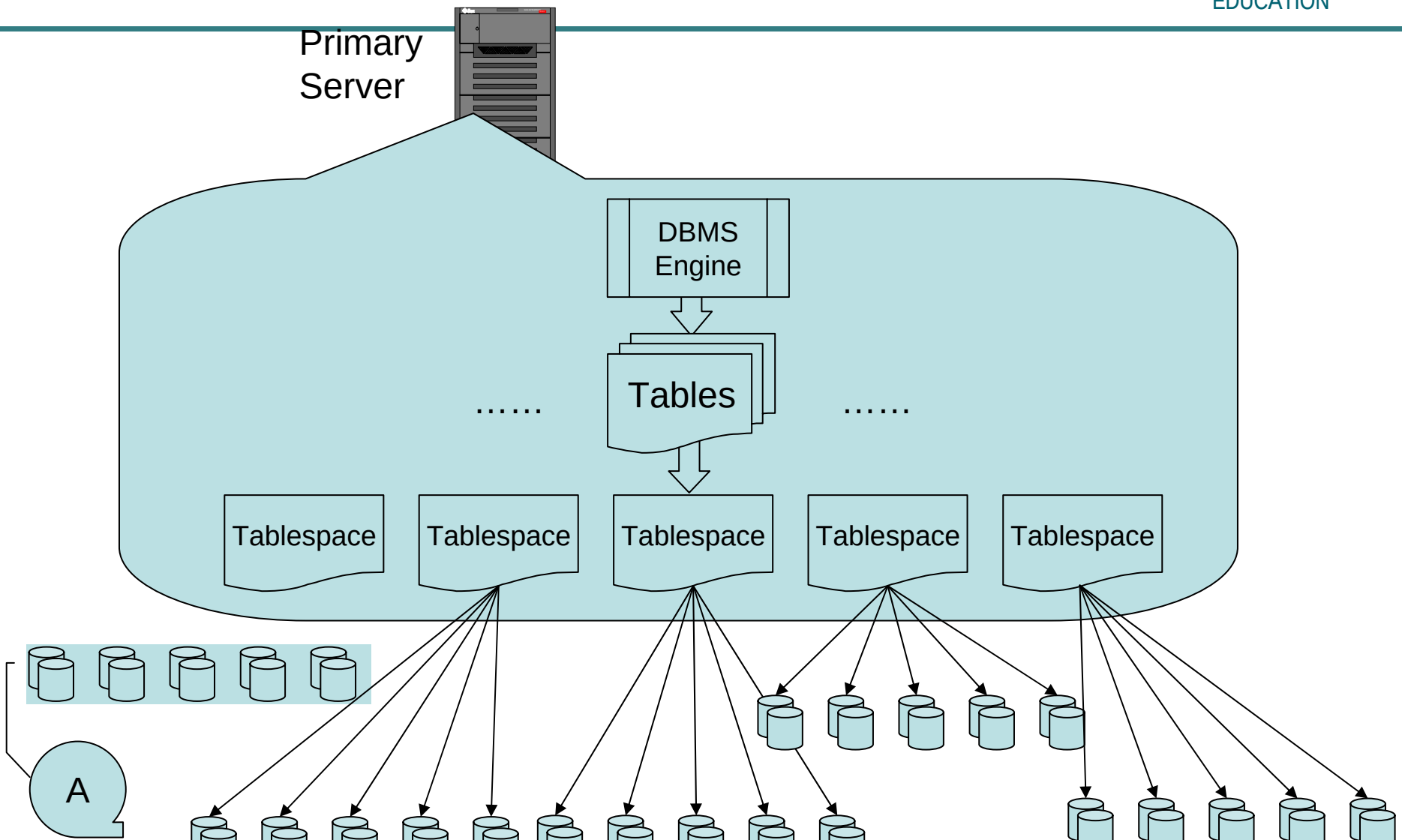
Application Assisted Protective Copy

- A “hot” or “online” backup or copy - is done while application is online and active
- Application coordination is invoked to facilitate copy
  - Often done in stages across data set or using a transaction log for shorter durations
  - Performance of application often degrades during “hot” window
- Also known as a “backup window”
  - Application isn’t offline, but isn’t performing at optimal levels
  - Application performance often continues degrading as the window expands or transaction flow increases
- Warnings
  - Not all applications support this operative mode
  - Source of complexity -- high failure rates (test often and thoroughly)

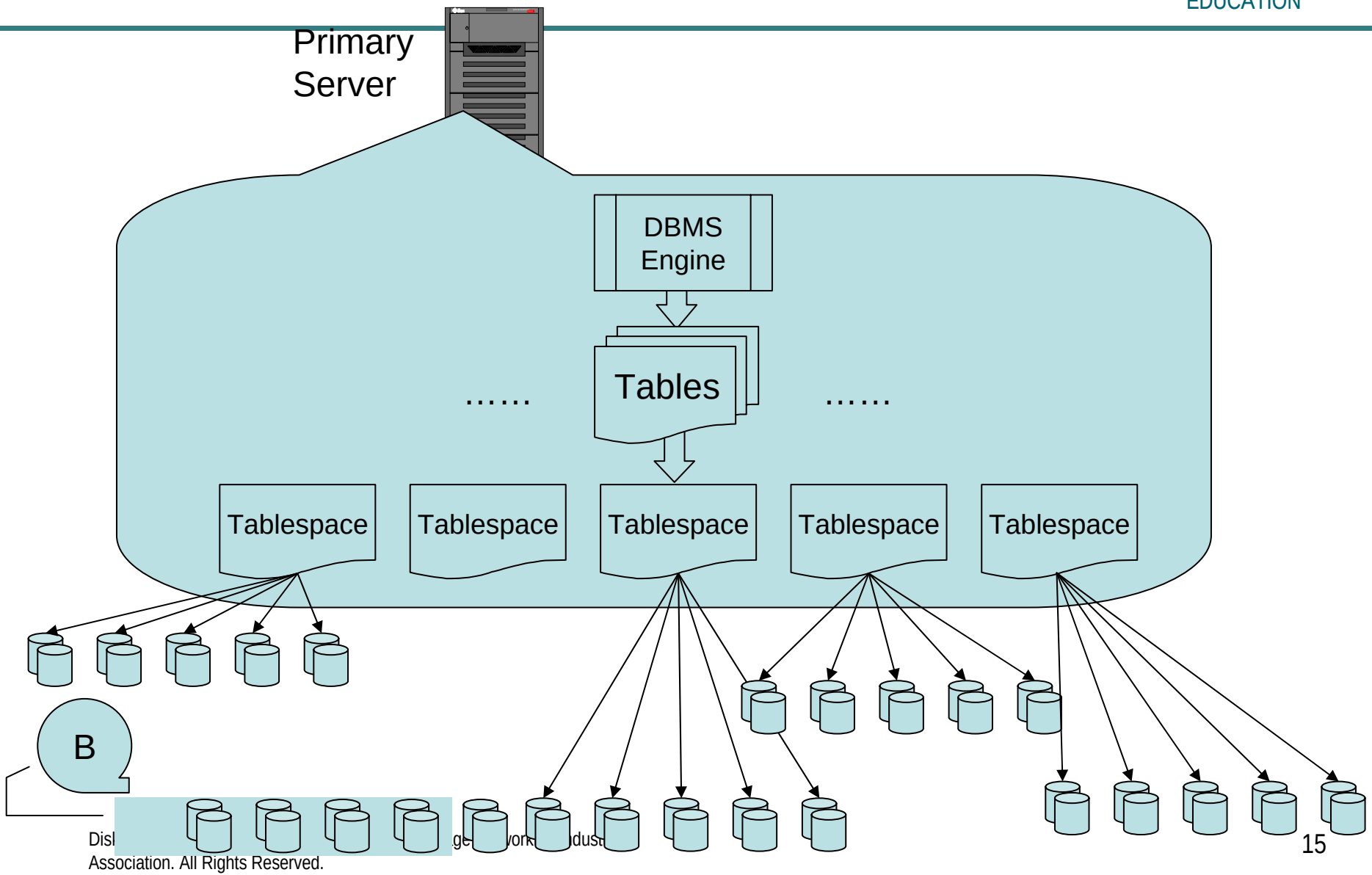
Database Example



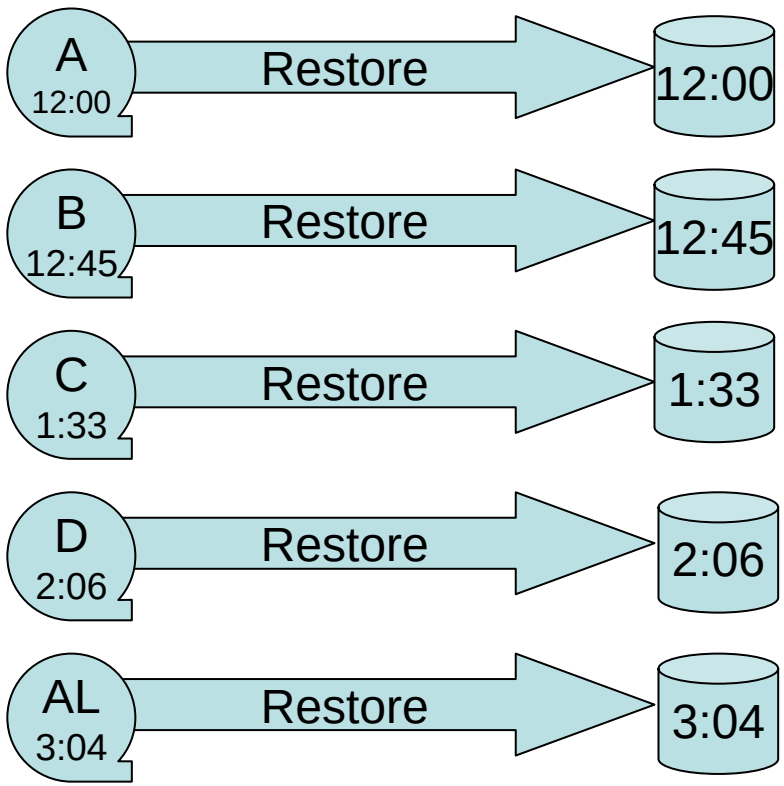
First set safe to copy…



…Second set safe to copy…

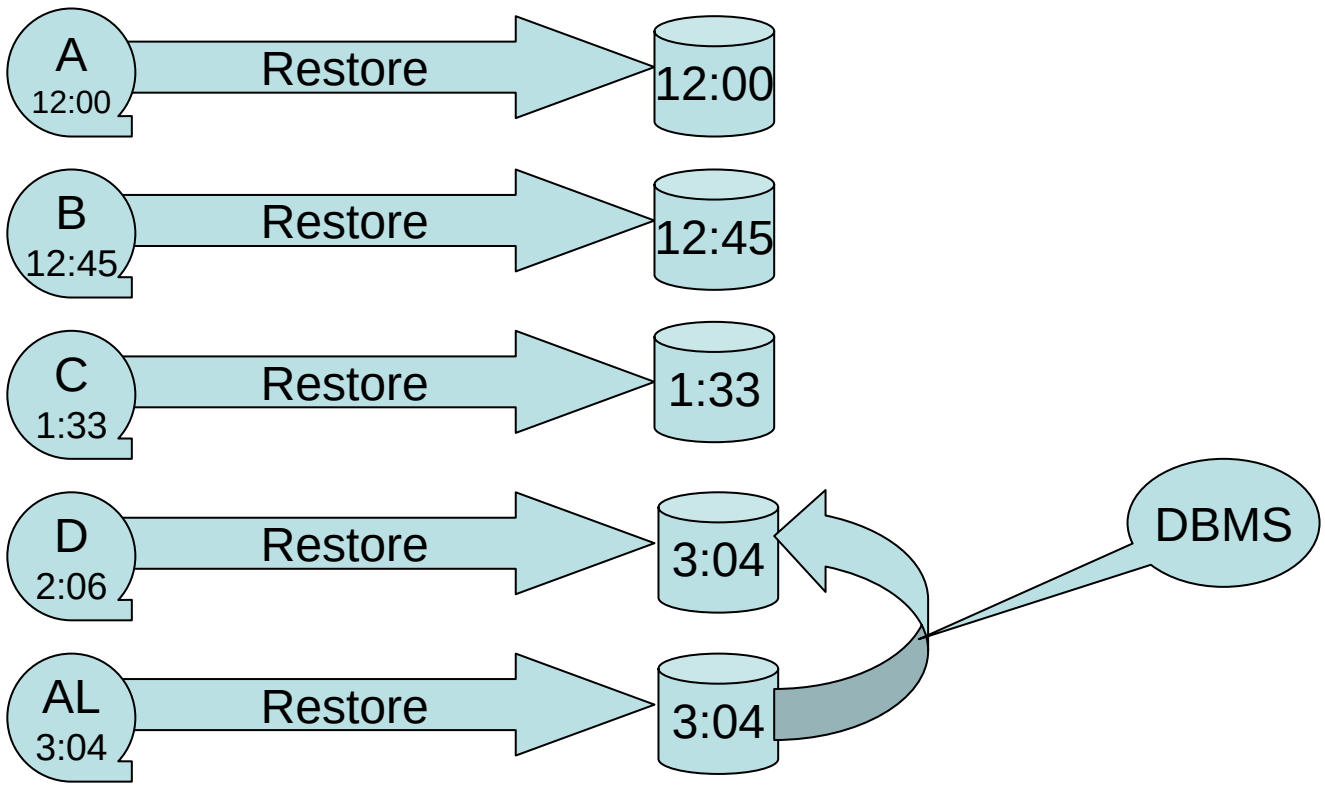


Resulting Save Set

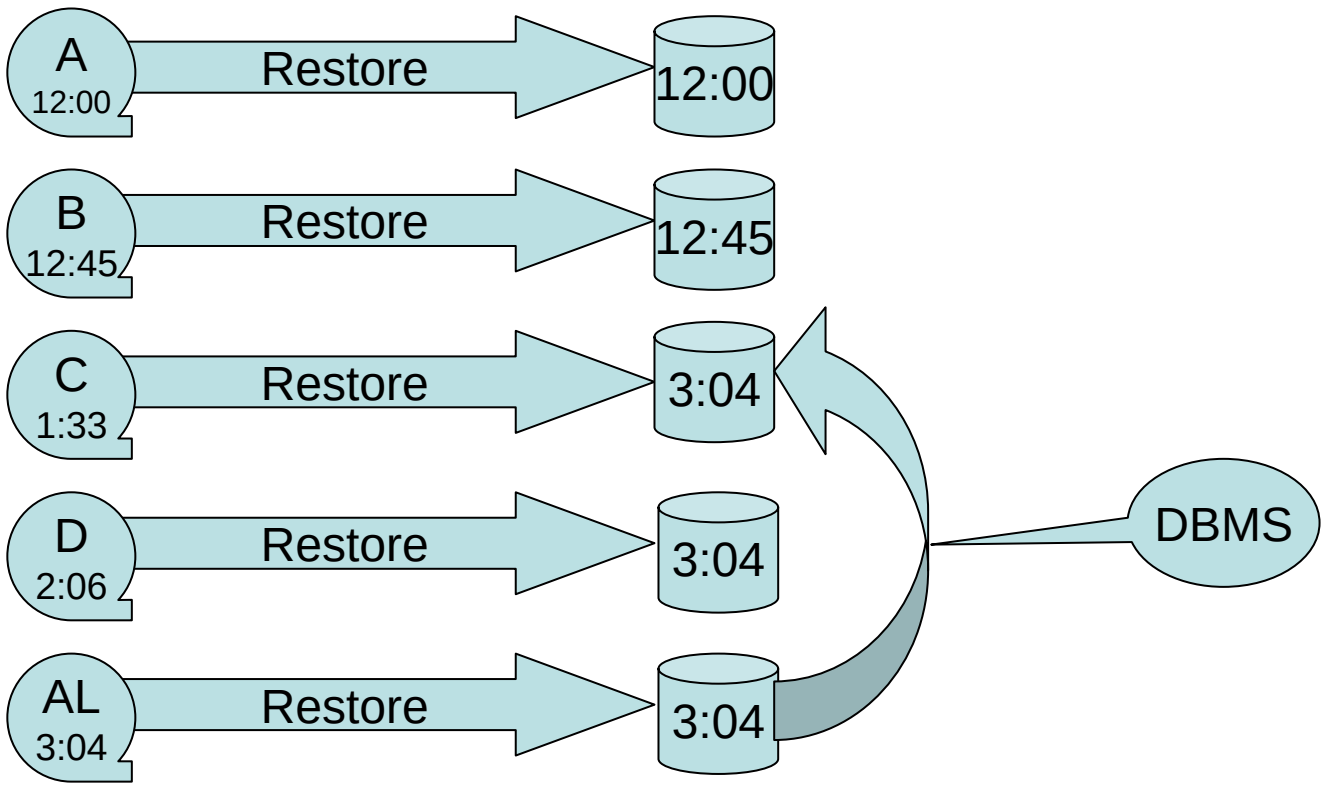


Resulting Save Set

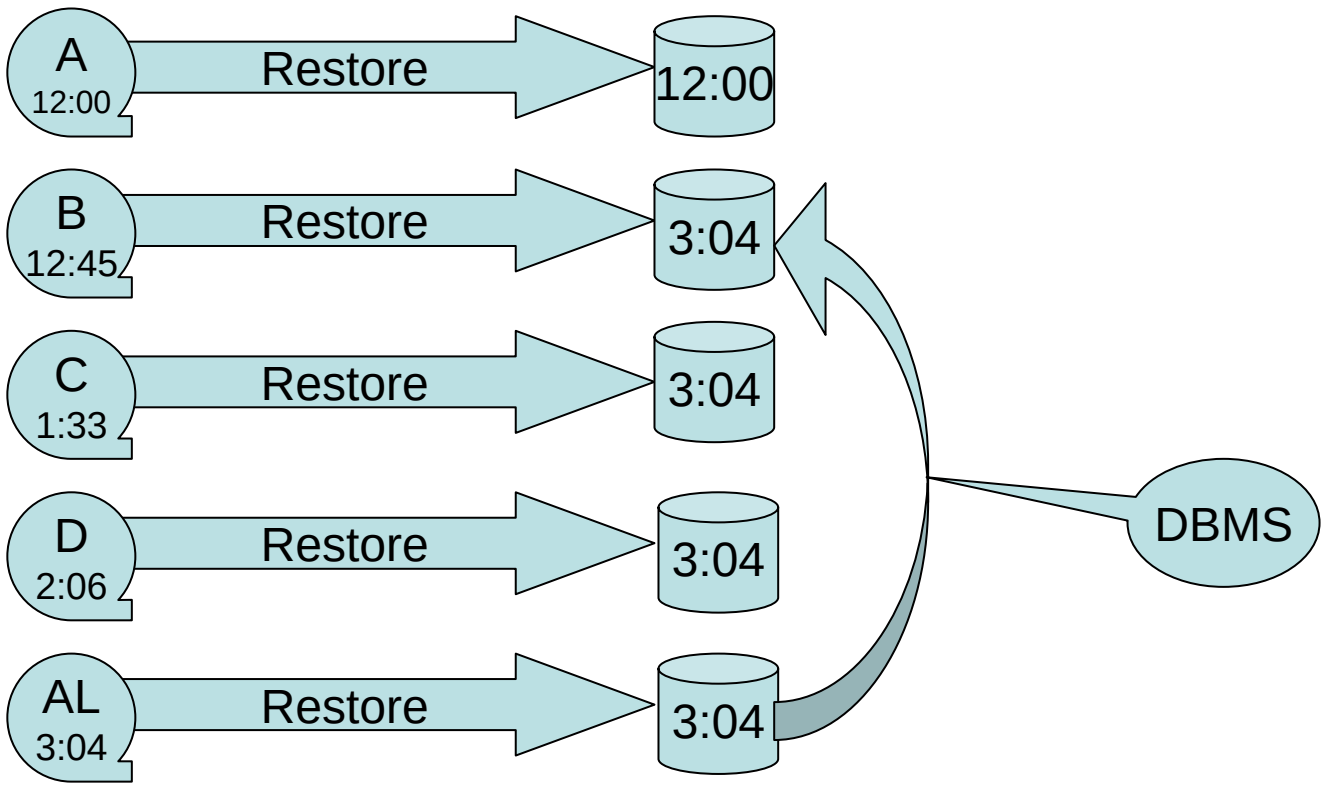
Apply the archive log to each save set, bringing all of the save sets to a synchronized point in time



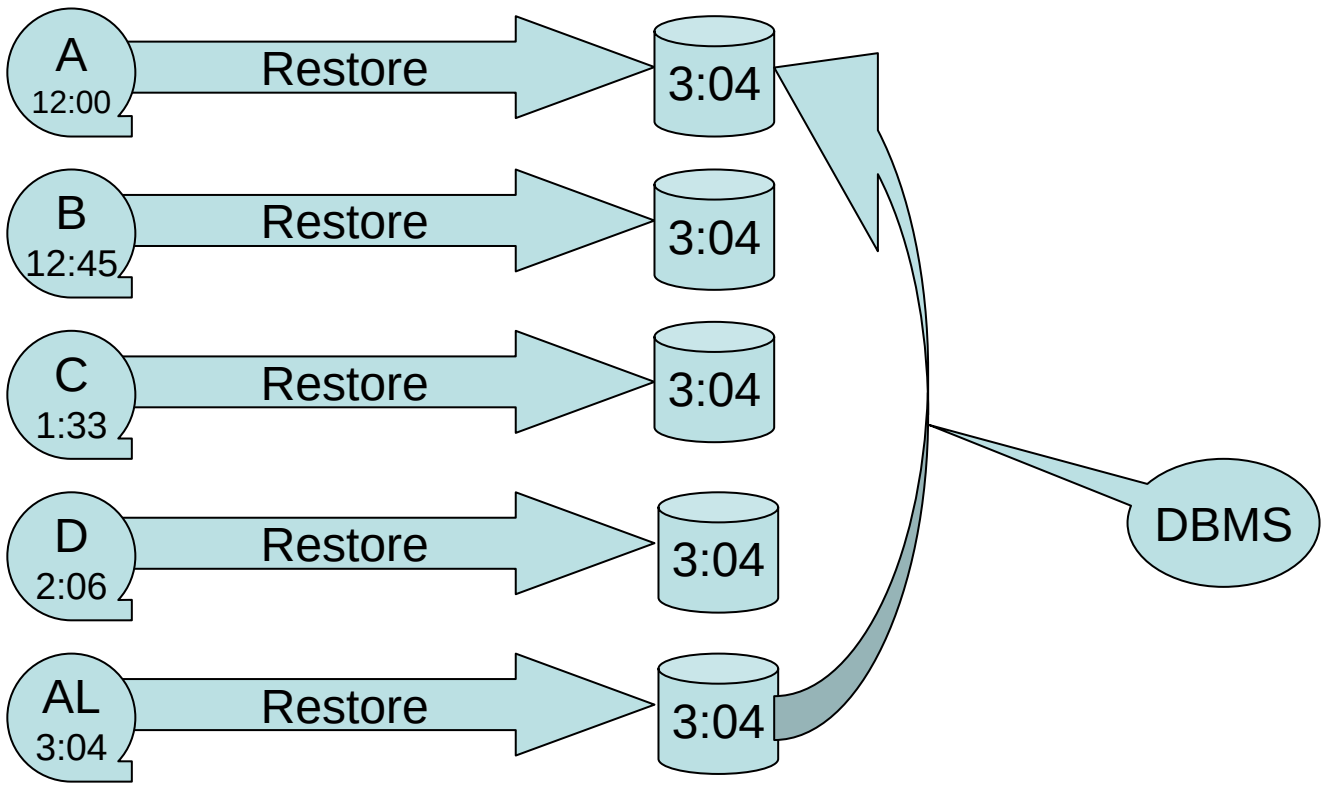
Resulting Save Set



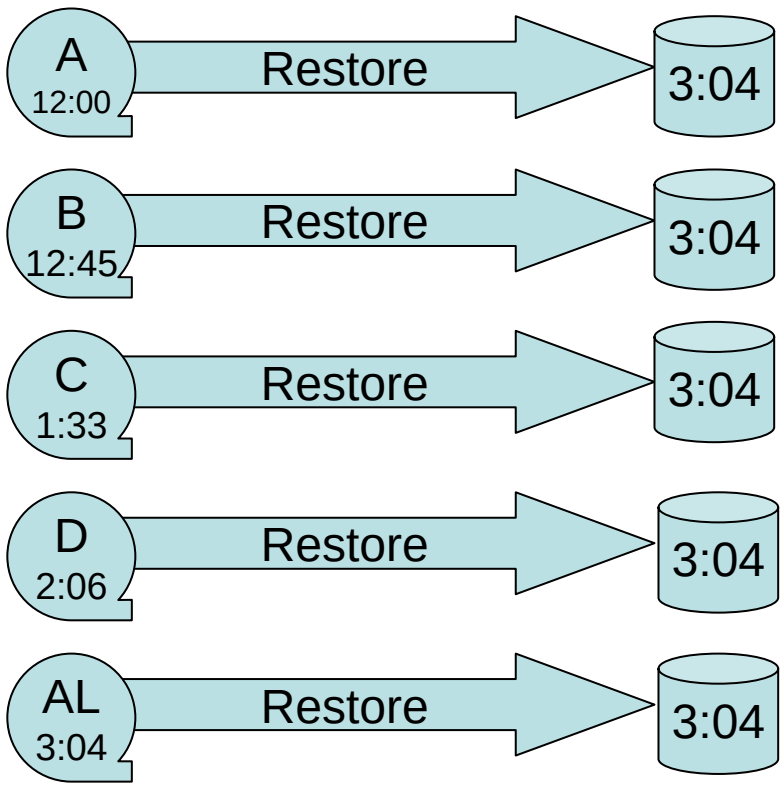
Resulting Save Set



Resulting Save Set



Resulting Save Set



Post Restoration:
It's like the lights went out
At 3:04

Atomic Protective Copy

- Atomic or Crash Consistent
- An atomic instant picture of the data
  - Application must be capable of crash recovery
  - Can it be clustered?
- No application interaction
  - Application isn't involved in the snapshot or CDP operation
  - No (snap/operation) related performance issues
  - Backup window is eliminated completely
  - Simplifies operation run-book and backups
- Warnings
  - Not all applications can do crash recovery
  - Not all disk protections can do atomic operations across a wide set of disks

Snapshots

SNIA
Storage Networking Industry Association
EDUCATION

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 fully usable copy of a defined collection of data that contains an image of the data as it appeared at the ***point in time*** at which the copy was initiated. A snapshot may be either a ***duplicate*** or a ***replicate*** of the data it represents.

www.snia.org/dictionary"

Snapshot of Networked Storage

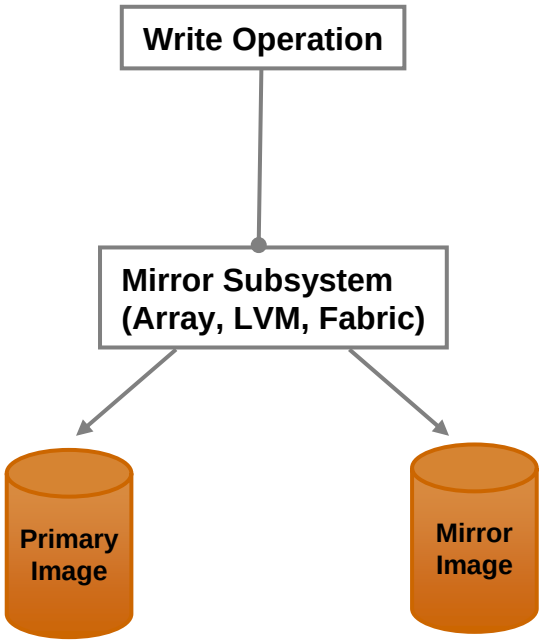
SNIA
Storage Networking Industry Association
EDUCATION

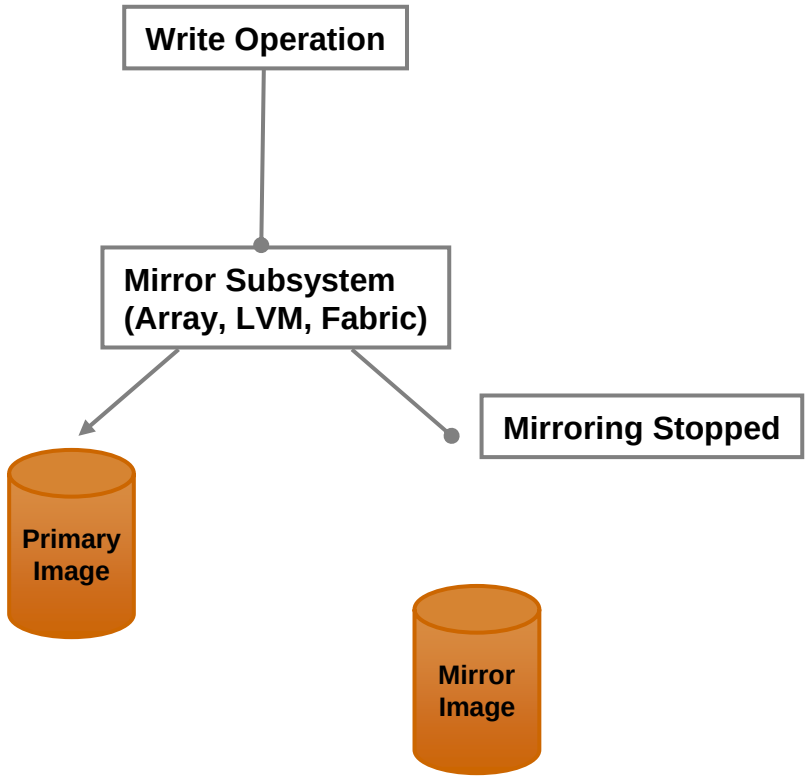
- **Terminology:** Snapshot, Checkpoint, Point-in-Time, Stable Image = Any technology that presents a consistent point-in-time view of changing data. *Many implementations exist.*
- **Why?** Allows for complete backup or restore, with application downtime measured in minutes (or less)
- Most vendors: 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

Full Copy Snapshots

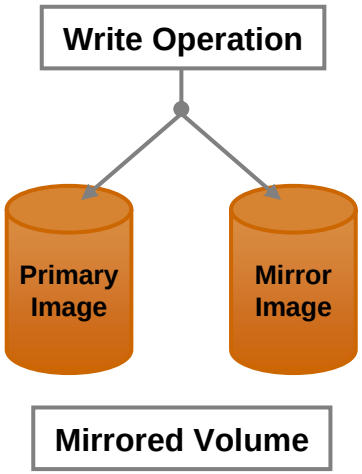
- Keep multiple RAID-1 sets of data
- Cease RAID-1 operations to one set of the disks
- Freezes the contents of that set at a specific point in time
- Can be done by a variety of sources
  - Host based logical volume manager
  - Fabric based logical volume manager or virtualizer
  - Array based functions
- Frozen copy is not dependent on current copy for content
  - Better tolerance for failure
  - Be careful: If alternate set is dependent (cage, power, director), this “independence” is compromised

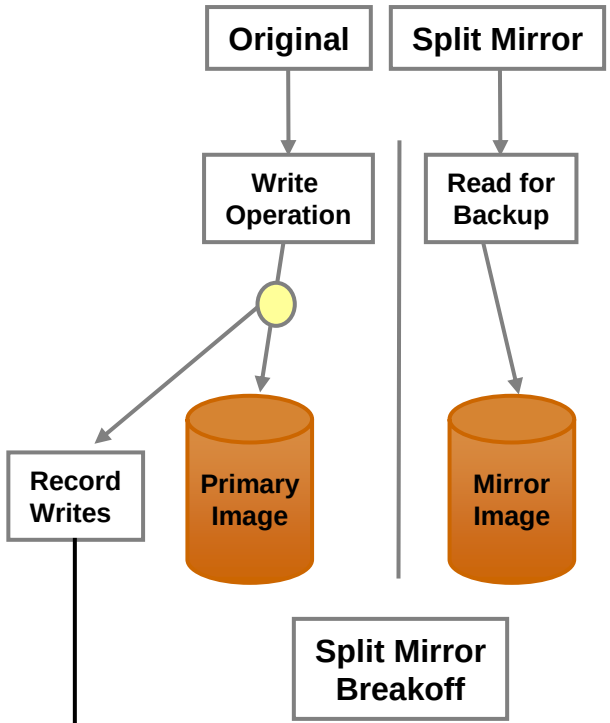
“Splitting” a Mirror



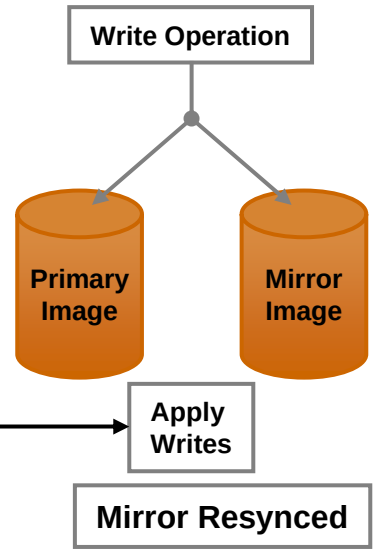


Full Split Life Cycle





1. Split the Mirror
2. Backup from the Split Mirror
3. Resync the Mirror: apply writes, or copy entire Primary Image



NOTE: If writes aren't recorded, or space for recording them is exhausted, the entire Primary Image must be copied to Resync.

DRL - Data versus Meta Logging.

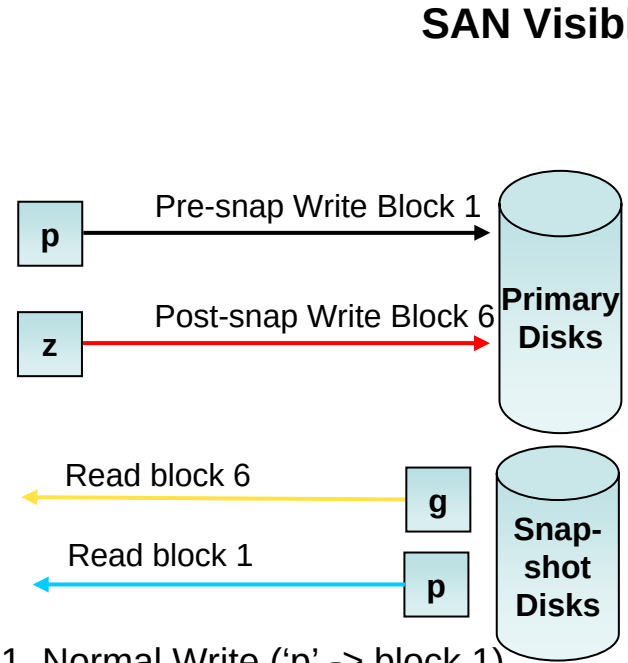
Differential Snapshots

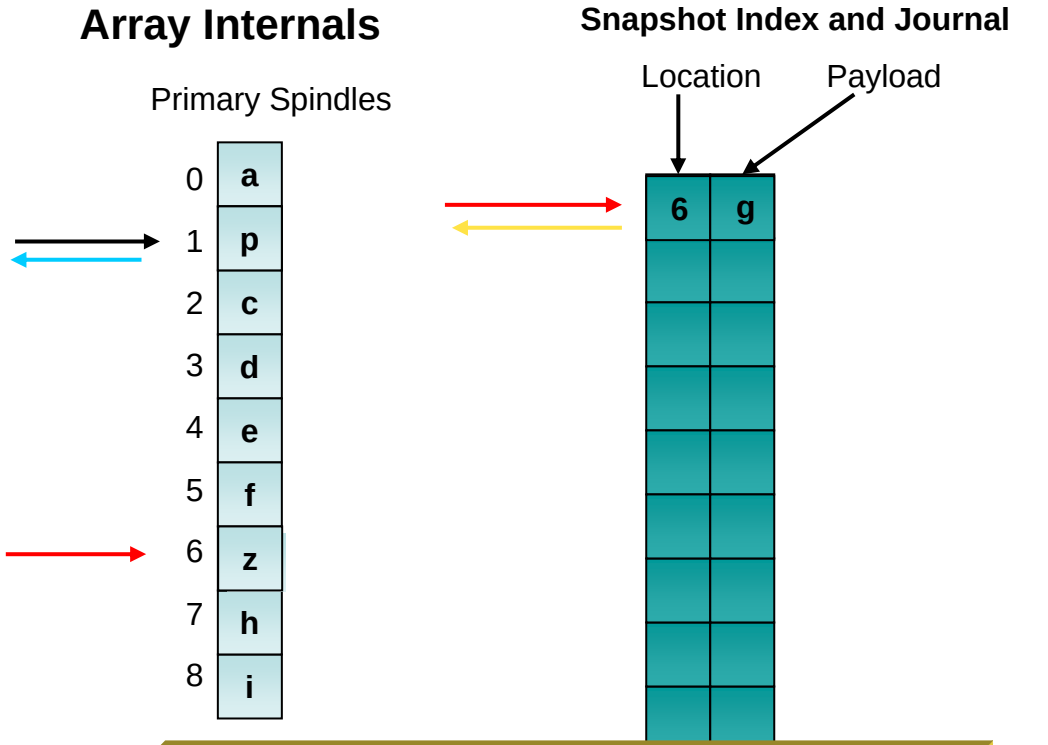
- Keeps track of “changes” to the primary copy
- Uses a combination of the “change” set and the primary disks to save/present snapshot
- Different approaches optimize for different use cases
  - Copy On Write (CoW)
  - Redirect On Write (RoW)
  - Write Anywhere (WA)

Copy On Write

- Primary disks remain current
- Whenever a Write operation arrives, it is held:
  - First the current contents of the write-destination are read in
  - The old-contents from the primary disk is saved off somewhere and indexed
  - The new write is now allowed to pass through
- Read path of current disks remains optimized
- Write path of current disks is potentially impacted
- Read/Write path of “snapshot” disks impacted

Copy on Write Snapshots





1. Normal Write ('p' -> block 1)
2. Create Snapshot
3. Write 'z' to block 6:
   - Hold write off
   - Move block 6 contents to journal ('h')
   - Allow new contents 'z' to update primary
4. Read from snapshot filters through index/journal - if it's not there, then we use the primary storage

- If snapshot is read/write, write operations to the Snapshot disk go directly to the snapshot Index/journal.
- If a block location is already there, the entry is overwritten (only need the last one)

Redirect on Write

- Primary disks are frozen
- New write operations to primary disks are stored in a journal (and indexed)
  - To read current copy, the journal is checked first
  - To read the snapshot copy, the primary disks are used
  - When snapshot is “dissolved”, write journal must be applied to primary disks to “catch up”
- Read path of snapshot is optimized
- Write path of current disks is optimized (no copy)
- Read path of current disks is potentially impacted

Write Anywhere

- All disk blocks are virtualized
  - Current disk is represented by a map to real blocks -- not directly mapped
  - Disk storage is larger than maps present
  - New writes, instead of “overwritting” blocks, are directed to free blocks
  - Maps are kept for “now” and potentially for multiple “snapshots”
  - Reference counts are kept for blocks “in-use”
- Performance doesn’t generally change primary/snapshot
- Performance can be impacted by fragmentation

Snapshot Comparison

	Full Copy Snapshot	Differential Copy Snapshot
Upsides	• No cost during "snapshot" process • Can be used for DR - independent copy	• Small storage consumption - typically 10-20% –Depends on churn • Typically can take advantage of cheaper disk
Downsides	• Massive storage cost – 1x of storage per RPO – Like disk - expensive • Often in the same disk frame – Loss of DR component • Consider re-sync time in schedules	• Performance impacts 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 –Most run-books require copy • 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

Continuous Data Protection (CDP)

SNIA
Storage Networking Industry Association
EDUCATION

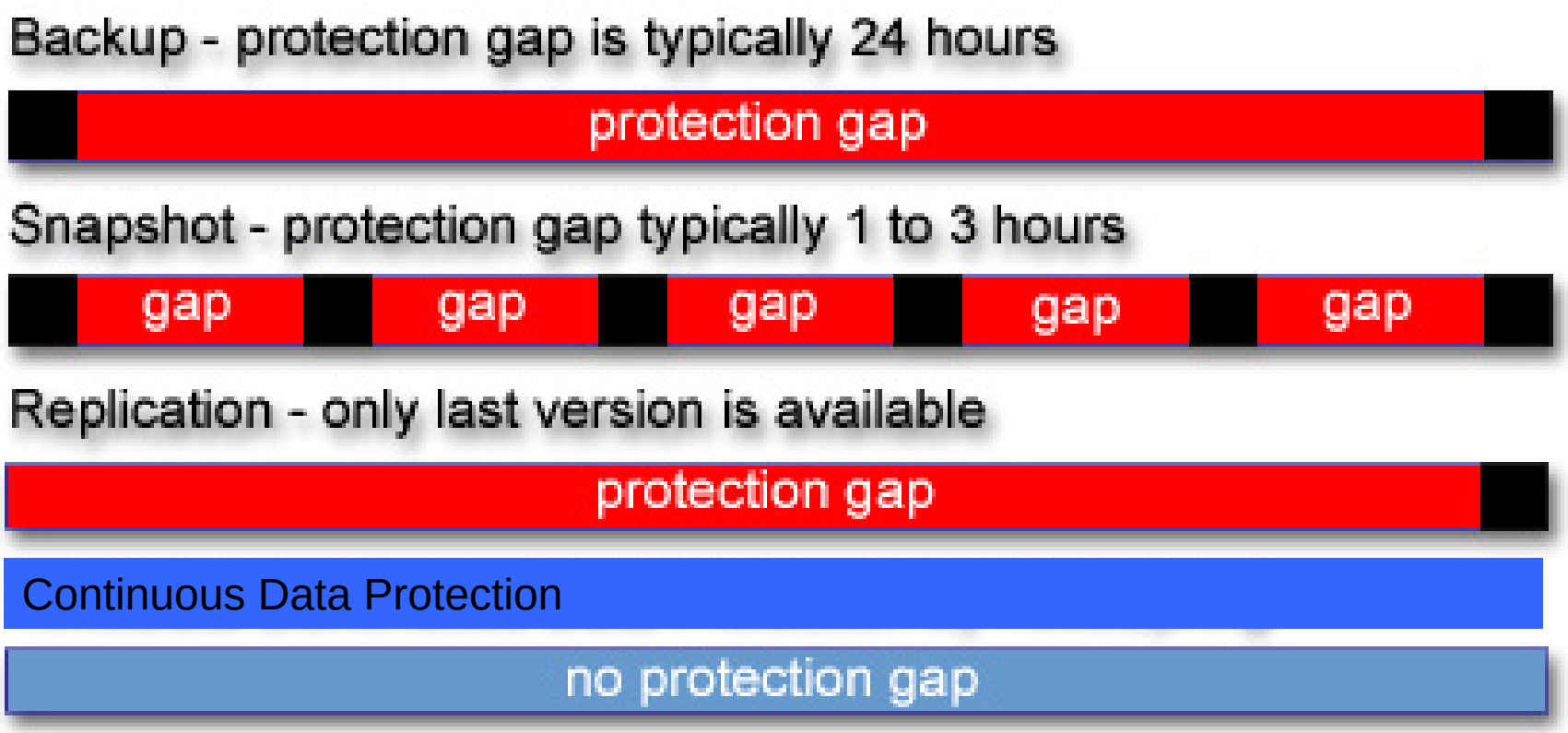
A disk based method that captures all changes, and can recreate virtualized snapshots

"Continuous data protection (CDP) is a methodology that continuously captures or tracks data modifications and stores changes independent of the primary data, enabling recovery points from any non-predetermined point in the past. CDP systems may be block-, file- or application-based and can provide fine granularities of restorable objects to infinitely variable recovery points. *Draft definition from CDP SIG, SNIA*"

Overview of CDP

- Constantly watching all write operations
- RPO is continuous -- any point in time
  - May degrade to discrete points as data ages
- RTO may be from near-instant to minutes or hours
  - Depends on recovery and deployment model
  - Time-to-data depends on finding exact moment when data in desired state, mounting the virtual volume, or file desired
  - Time-to-recover defined mostly by application recovery time
- A form of logical data protection
  - Can be delivered in multiple places (device, OS, app)
  - Can any tier of storage (primary, secondary, SATA, etc)
  - Block or file
- Protection
  - Application objects, Applications, or Federations of applications

Data Capture and RPO



Mechanics of CDP

- Methods of CDP
  - Before Imaging (CoW, RoW, Write-anywhere)
    - Save/retain copy of block contents before it's overwritten
  - After Imaging (Journaling)
    - Save new writes in order, apply against an aged replica
  - Mixed approaches
- Software monitors data flow between applications and storage
  - Host Based: software is loaded on application server
    - Typically a device driver/OS extension
    - Can be embedded in application, or an extension of application
  - Fabric based : embedded in storage infrastructure (switch, appliance, array)
- Restoration
  - Directional or Replay recovery: applying logs to a replica
  - Virtualized or Indexed recovery: Presenting a virtualized picture of result, data movement occurs after availability

Implementation

- Application level
  - Built into application
  - Inherent to application, can't extend beyond
- Host level
  - Agent or device driver loaded into OS
  - Can handle multiple applications (on the same host)
- Network/Storage level
  - Appliance, embedded in switch, or embedded in storage controller
  - Federations - typically OS and application agnostic

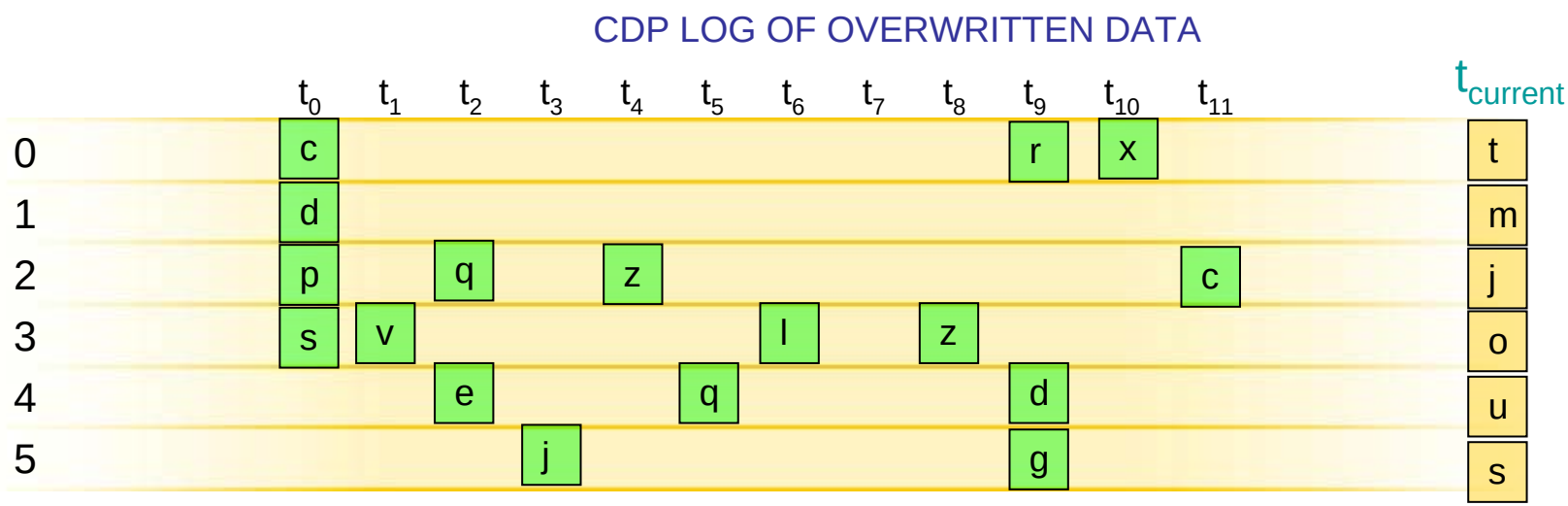
Deployment

- In-band
  - Data must pass through CDP stack on its way to primary storage
  - Host based solutions are in-band; fabric based solutions can be
- Side-band
  - Copy of data passed to CDP stack in parallel to sending it to primary storage (RW or WO mirror)
  - Fabric based solutions may do this
- Out-of-band
  - Journals or logs created by the virtualizer (LVM, switch); meta data from journals are handed to the CDP engine for processing after IO has aged
  - No production examples of this yet (see technology demos)

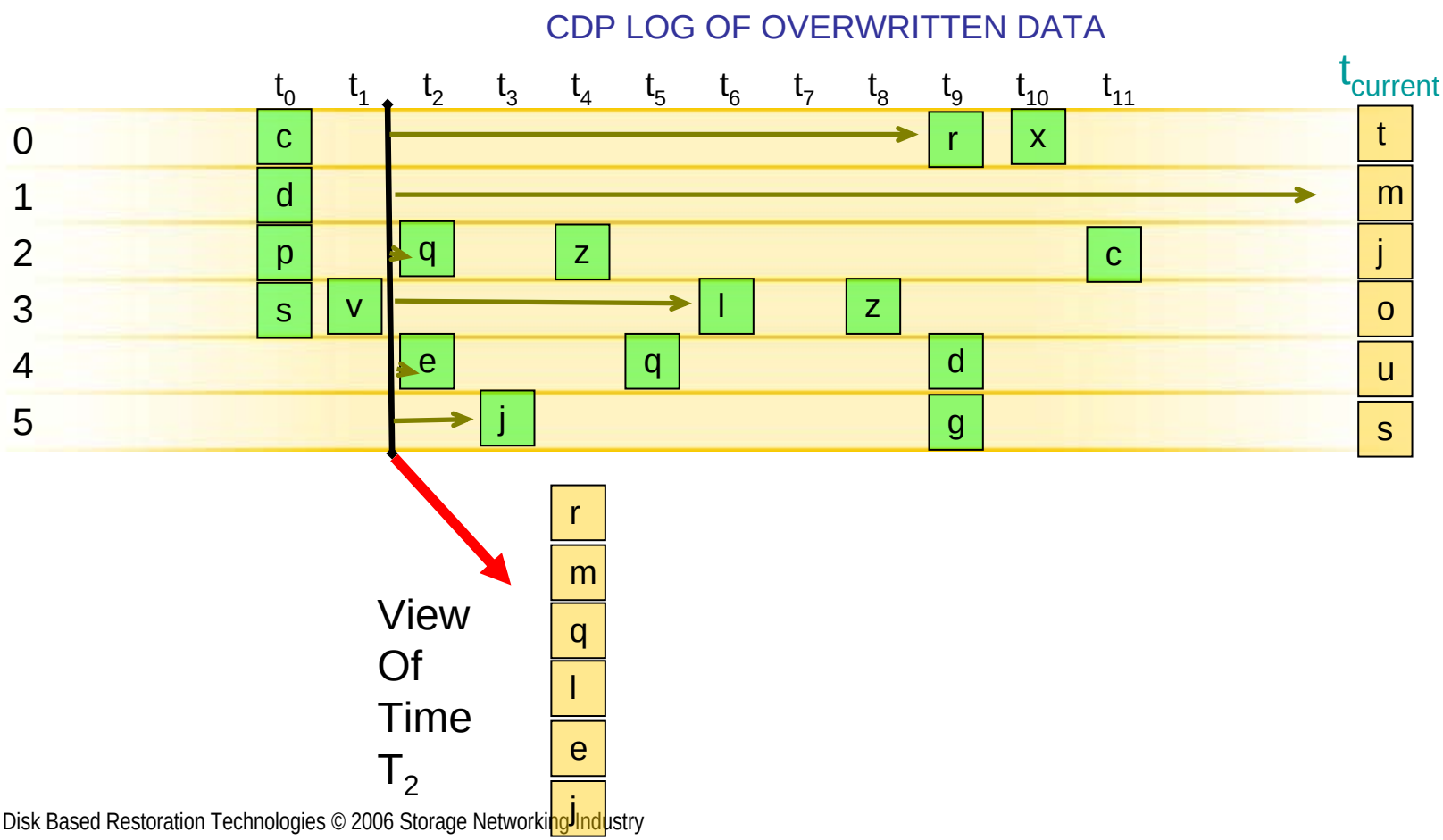
Mechanics of CDP – "Restore"

- Restore Process
  - Chose a point in time (PIT) before corruption/loss
  - Build the required image
    - May be a file or a mountable virtual disk or group of disks
    - May be part of production (fast restore), or a copy point (virtual snapshot)
  - Mount and examine the PIT
  - Use the PIT to roll back:
    - Entire file system, database or combination by:
      - Roll back the production image, or
        - » Rollback or no-copy "restoration virtualization"
      - Re-silver the production image with the selected PIT
    - Select file(s), objects, tables
  - PIT is dynamic and re-creatable -- big difference from snapshots
    - Can enable hot restore (no-copy, instant)

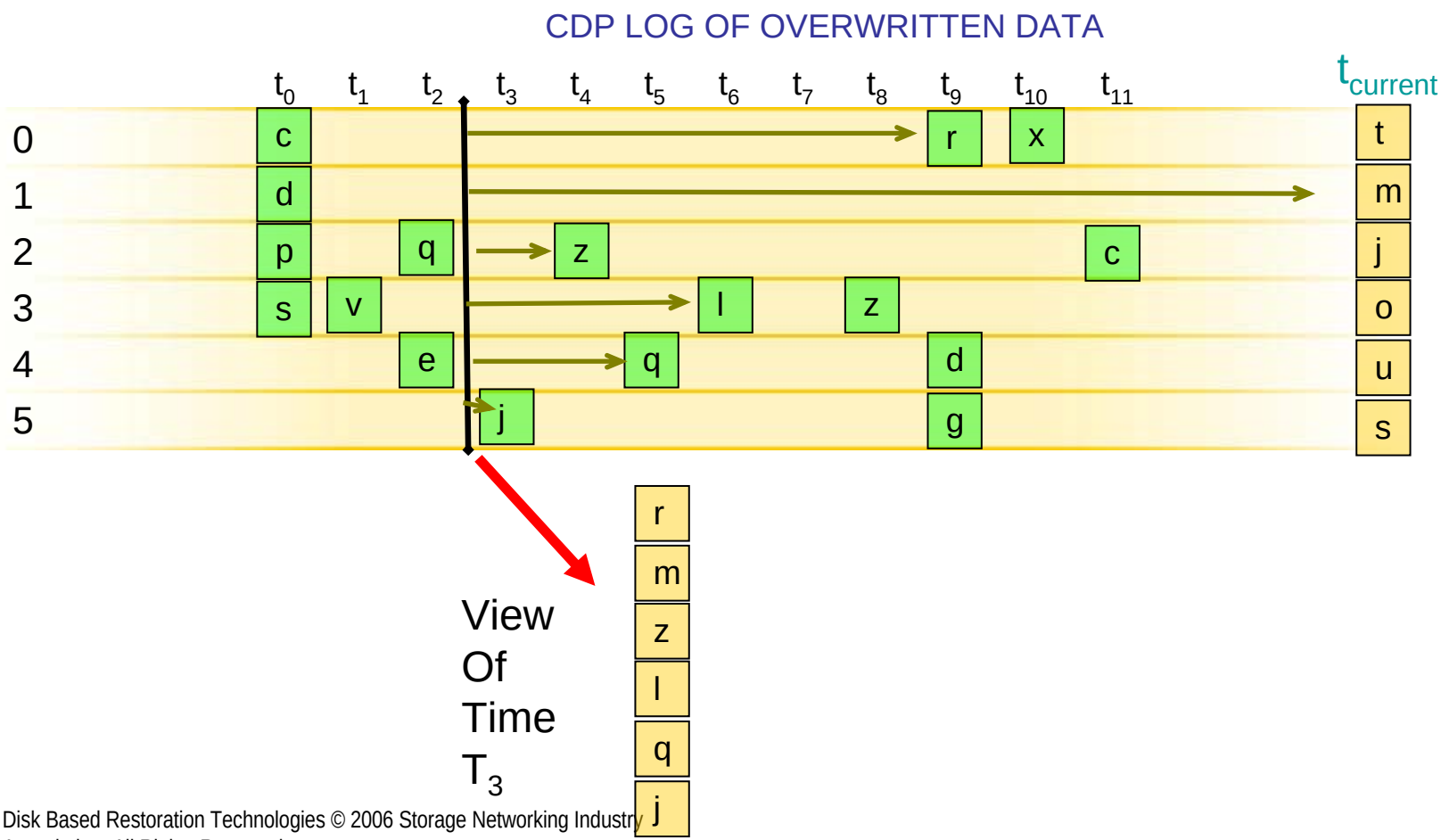
CDP Captures Everything



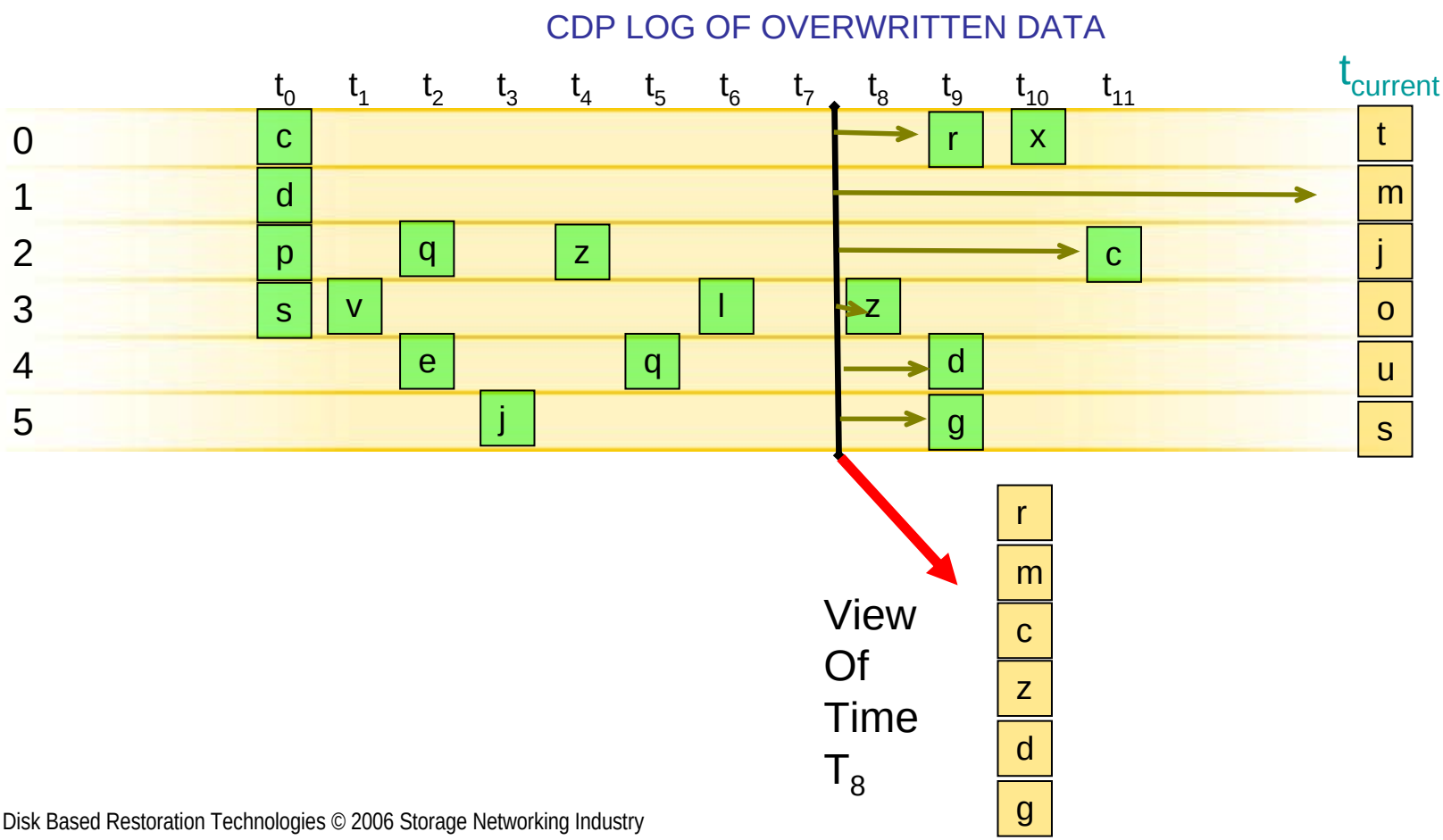
Virtual Snapshot of Time T_2



…or T_3



…Or T_8



CDP Considerations

- Requires additional disk space
  - Higher when longer retention windows used
  - Ranges from 0.2x to 2x
- Requires care in policy setting
- Where/how?
  - File system, Block/device, application
  - Application integration versus application CDP
    - Atomic recovery eliminates gap
    - Application Integration is really about ease of use/automation
- Mechanics of object based restore
- Future
  - Standardization of Business Event tracking
    - Most provide it, needs to be standardized
  - Fabric Integrated with adoption of Intelligent fabrics

Disk Backup Methods

SNIA
Storage Networking Industry Association
EDUCATION

Method	Positive Benefits	Issues
Disk Backup	**1. Part of the Backup Application itself - easy to implement** **2. Can use any disk, any type, any tier - high end, SATA, DAS, NAS, etc.** **3. Allows backup application to attain better tape utilization -- streaming.** **4. Restores are faster than tape…**	**1. Faster than tape, slower than VTL (generally)** **2. Current versions of enterprise backup applications support it** **3. Backup window issues still exist - not "instant", but much faster than tape**
Tape Emulation: VTL (*Virtual Tape Libraries)*	**1. Re-uses existing backup procedures, emulates tape, uses tape compression, tape archive made in background** **2. High performance – multiple parallel channels – consolidates backup & minimizes backup window issues** **3. Restore from disk if data is online, or from tape archive** **4. Allows more efficient use of tape media**	**1. Adds another layer of hardware complexity** **2. Adds cost of a HA disk array plus tape emulation software, connectivity and management** **3. Backup window issues still exist - not "instant", but much faster than tape**

Disk Backup Methods - 2

SNIA
Storage Networking Industry Association
EDUCATION

Method	Positive Benefits	Issues
Full Copy Snapshots: Split mirrors	**1. Fast, transparent operations** **2. At least one full copy of the original exists at all times** **3. Rapid, user-initiated file or volume recovery, redirection, and bare-metal recovery** **4. "Instant backup" from application perspective, rapid restore without losing original or instant restore**	**1. Storage costs - N times storage use (RPO's drive cost)** **2. Recovery from tape can be awkward, depending on product used** **3. Performance on re-integration of mirrors for re-silvering**
Differential Snapshots: CoW, RoW, WA	**1. Fast, transparent operations** **2. Rapid, user-initiated file or volume recovery, and redirection** **3. "Instant backup" from application perspective, rapid restore without losing original**	**1. Performance while snapshot active or when dissolved** **2. Storage repository growth can be large - highly dependent on implementation** **3. Recovery from tape can be awkward, depending on product used** **4. Scales based on change rate**

Disk Backup Methods - 3

SNIA
Storage Networking Industry Association
EDUCATION

Method	Positive Benefits	Issues
Continuous Data Protection (CDP)	**1. Continuously protects data, giving "Any Point In Time" recovery** **2. Eliminates backup window, consolidates backup, little to no impact on operations** **3. "Instant backup" and "instant or Rapid restore"** **4. Rapid, user-initiated volume or file level recovery, redirection, and bare-metal recovery**	**1. Adds a new process into the mix** **2. Costs: Additional 1.5x or more online disk capacity** **3. New market, small number of product options available** **4. Scales based on change rate**
Remote Replication	**1. Extension of full snapshots, delta snapshots or CDP (depending on implementation/vendor)** **2. File or block replication across a communications pipe synchronously or asynchronously (remote)** **3. Combines DR with corruption protection**	**1. Speed of light issues** **2. Site-to-site recovery can be awkward -- local corruption issues better dealt with locally**

Please send any comments on this tutorial to SNIA: trackdatamgmt@snia.org

Many thanks to the following individuals for their contributions to this tutorial:

Nik Simpson
Bill Pierce
Dan Tanner
Edgar St.Pierre
Alex Adamopoulos

Get Involved !

www.snia-dmf.org

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

Q&A / Feedback

SNIA
Storage Networking Industry Association
EDUCATION

Please send any questions or comments on this presentation to SNIA: trackdatamgmt@snia.org

Thanks to the following individuals and groups for their contributions / reviews.

SNIA Education Committee

Lionel Teysseire
Alan Lindsey
SW Worth
David Black
Jeff Bain
Michael Rowan
Nancy Clay
SNIA Tech Council
Richard Saunders
Charles Curtis
Gene Nagle
Michael Fishman
Gavin Cole
David Hill
SNIA "Data Management Forum"
Tom Lanzatella
Neal Watkins
Bill Webster
Kirby Wadsworth
Edgar St. Pierre
Lauren Whitehouse
SNIA Backup Tech Working Group

Get Involved !

www.snia-dmf.org

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