



Education

# HIGH AVAILABILITY AND DISASTER RECOVERY FOR NAS DATA

Paul Massiglia  
Chief Technology Strategist  
agámi Systems, Inc.

- 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.

*As network attached storage has matured, users have entrusted increasingly critical data to it, creating requirements for protection against failures and disasters. This session will present a survey of techniques available for protecting data stored on NAS systems against loss or destruction by threats ranging from hardware and software component failure to accidental or deliberate corruption to disasters that completely incapacitate an entire data center.*

*Backup, RAID and mirroring, system fail-soft, snapshots, continuous and periodic replication, and NAS system clustering will all be discussed. For each technique, the threats against which it protects, the capital and operating costs, and the expected recovery time and recovery point objectives will be presented. The goal for the session is to give students an appreciation for the high availability and disaster protection options available for their NAS-managed data, in order to better equip them to make well-informed decisions when purchasing or defining operating procedures.*

## ➤ Data protection

- ◆ An intact copy of critical data survives disaster events
- ◆ Restoring application and client access is treated as a separate problem

## ➤ High availability

- ◆ An intact copy of critical data survives disaster events
- ◆ Procedures (usually automatic) in place to restore service to applications and clients

› *With continuous data protection (CDP), the distinction becomes less clear*

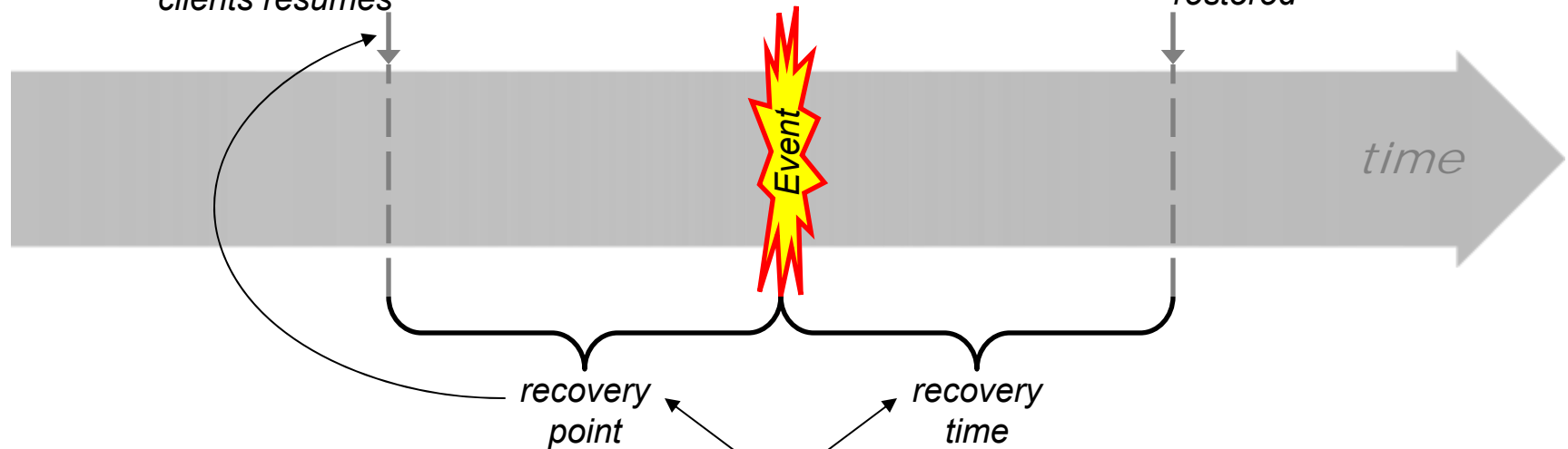
# Measuring availability

[Link to RPO details \(33\)](#)

[Link to RTO details \(32\)](#)

*“Business time” represented  
by data when service to  
clients resumes*

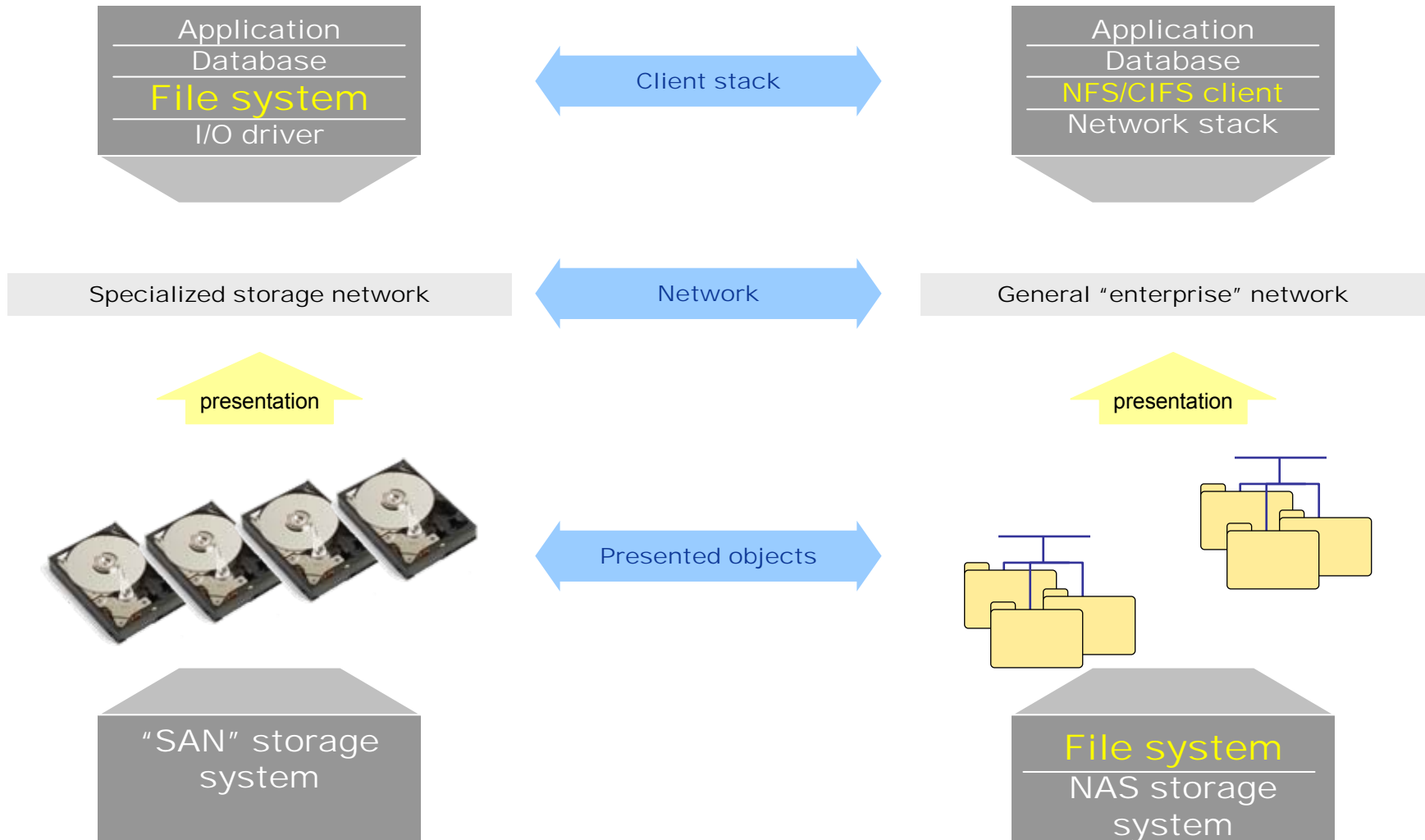
*Service to clients  
restored*



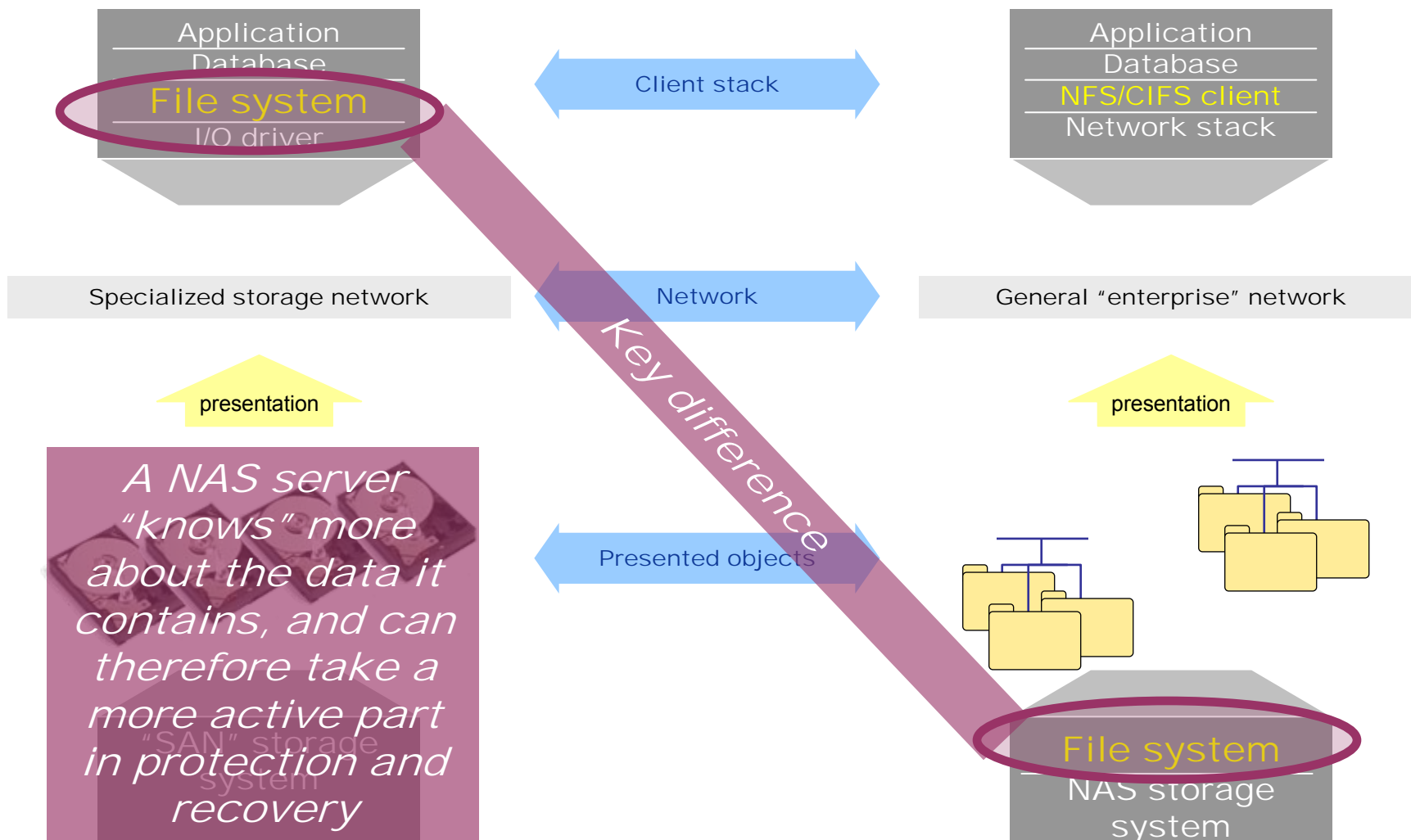
*Business objectives for these measures  
determine appropriate data recovery and  
availability techniques*

*In general,  $\downarrow RTO-RPO \downarrow \Rightarrow \uparrow \$ \$ \$ \uparrow$*

# What's unique about NAS ?



# What's unique about NAS ?



## ➤ Logical

- ◆ Equipment and facilities remain physically intact
- ◆ Data has been destroyed or corrupted

## ➤ Physical

- ◆ System failure  
*The surrounding environment is intact*
- ◆ Data center loss  
*The entire IT environment must be recreated*

## ➤ Causes

- ◆ Malice (malware or human action)
- ◆ Innocent human error
- ◆ Hardware or software fault

## ➤ General recovery strategy: “turn back the clock”

- ◆ Restore and revert to a “good” version of data

## ➤ Inherent consequences:

- ◆ Periodic copies of data sets stored in safe locations
- ◆ Updates occurring between latest recovery point and time of fault discovery are “lost” and must be recreated

# A “good” version of data is...

- **Correct**
  - ◆ Is captured prior to the corrupting event
  
- **Consistent**
  - ◆ Represents a “snapshot” of the data set
  
- **Modular**
  - ◆ Is restorable en masse or file-by-file
  
- **(Near-)current**
  - ◆ Represents a recent recovery point when restored
  
- **Persistent**
  - ◆ Is restorable after years or decades

#### persistence

Synonym for non-volatility. Usually used to distinguish between data and [metadata](#) held in [DRAM](#), which is lost when electrical power is lost, and data held on [non-volatile](#) storage (disk, tape, battery-backed DRAM, etc.) that survives, or *persists* across power

<http://www.snia.org/education/dictionary/s/>

# Techniques for protecting NAS data

```
cp -r -x * /backup_location
```

Advantages	Limitations
Nothing to buy	Bandwidth-intensive
No specific configuration or training	Platform-specific
Long-term stability of data formats and access protocols (NFS/CIFS)	Administration-intensive: data sets, backup devices, and media must be hand-selected and managed

› *In two words*

➤ Doesn't scale

➤ Is inflexible

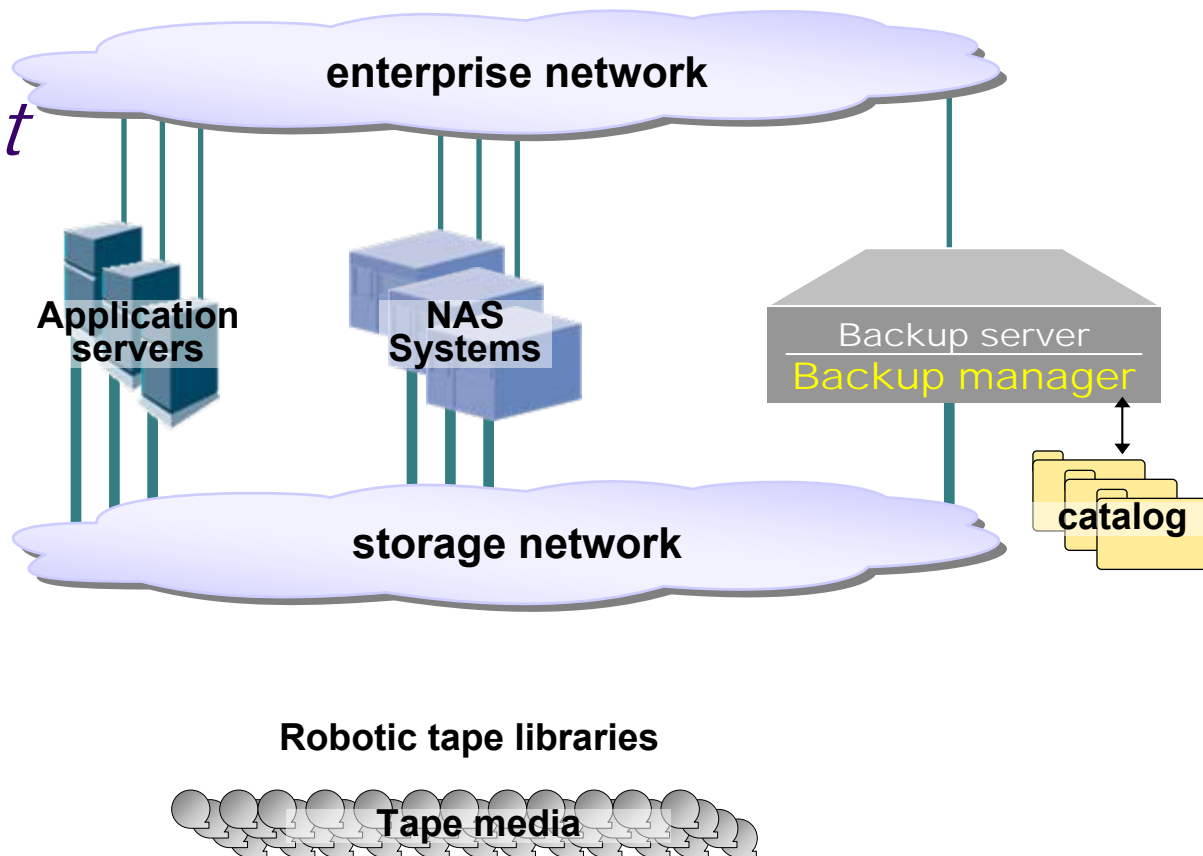
# Techniques for protecting NAS data

## Backup management software

### ➤ Why use it?

### ➤ In a word, *management*

- ◆ Data
- ◆ Schedule
- ◆ Device
- ◆ Data flow
- ◆ Media



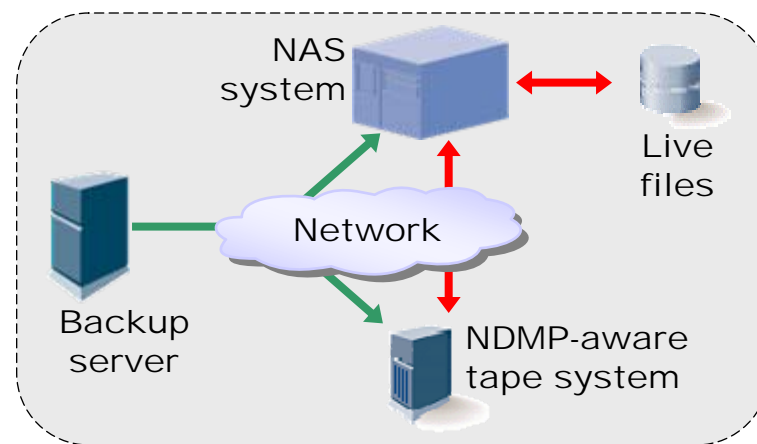
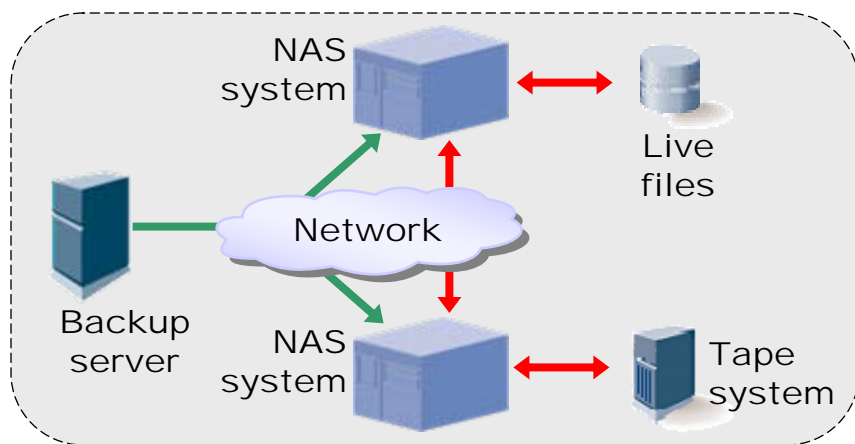
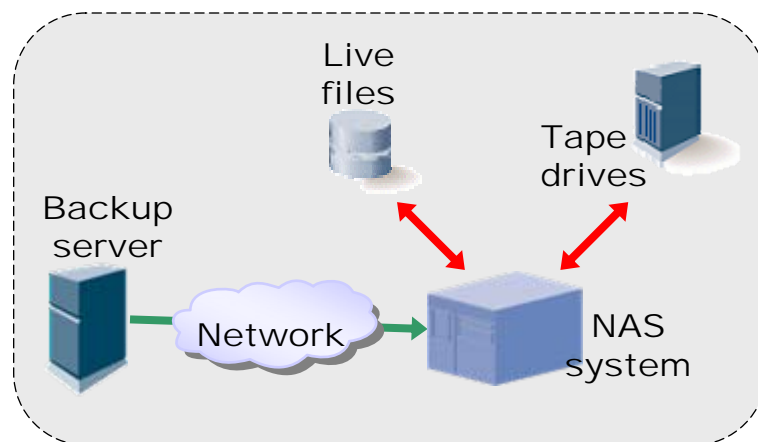
[Link to +/- details](#)

# Techniques for protecting NAS data

## Network Data Management Protocol (NDMP)

### Supported by most major backup management packages

- ◆ Reduces backup time
- ◆ Minimizes network traffic
- ◆ Enables more frequent recovery points



— Job control  
— Data flow

### ➤ Faking it

- ◆ Virtual tape libraries (VTL)

### ➤ Emulates

- ◆ Tape drives
- ✦ Media
- ✦ Robotic loader



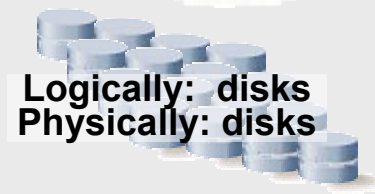
Logically: tapes  
Physically: disks

### ➤ Admitting it

- ◆ On-line storage “containers”

### ➤ Containers managed by

- ◆ backup management software
- ◆ A popular NAS application



Logically: disks  
Physically: disks

[Link to VTL details](#)

# Techniques for protecting NAS data

## Continuous Data Protection (CDP)

### ➤ Fundamental technique

- ◆ Log every transaction on a data set (a “redo log for any kind of data”)

### ➤ Fundamental benefit

- ◆ Defers specification of recovery point until restore time

### ➤ Compared to snapshots

- ◆ Snapshot:.....Pre-stored images of data taken at fixed times
- ◆ CDP:.....Image of data recreated dynamically at recovery time

### ➤ Challenges

- ◆ Definitions of “transaction” and “data set”
- ◆ Storage consumption (for log)
- ◆ Impact on application performance
- ◆ Time to identify and present a point-in-time data set image
- ◆ Integration with applications and data managers

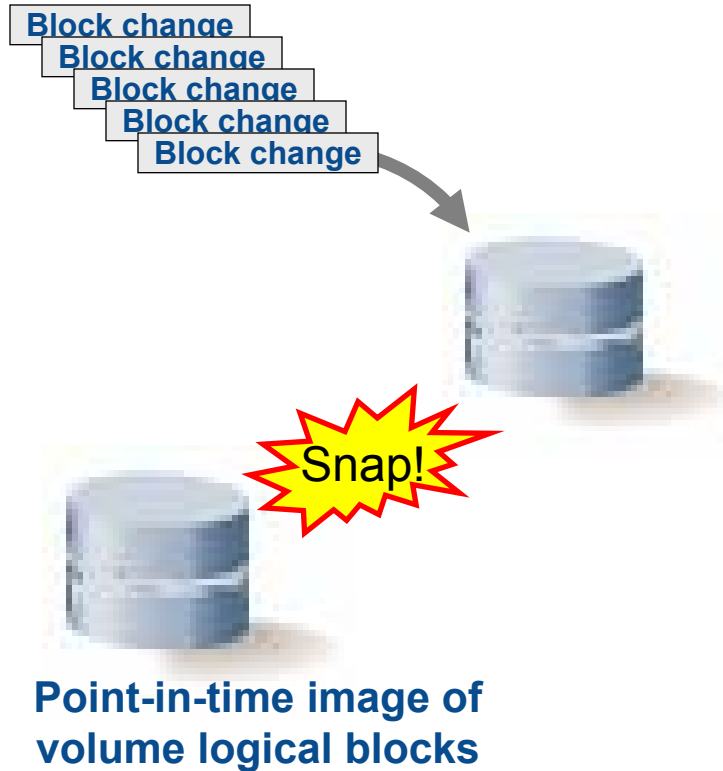
Continuous Data Protection – CDP

#### **CONTEXT [Data Recovery]**

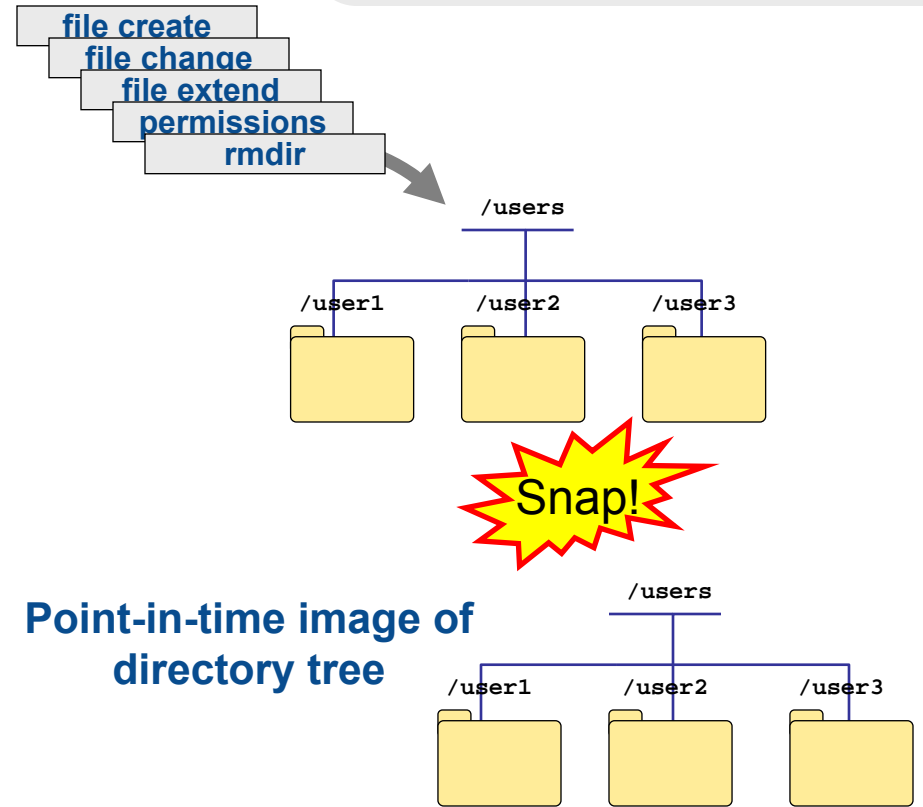
A data protection service that captures changes to data to a separate storage location. There are multiple methods for capturing the continuous changes involving different technologies that serve different needs. CDP-based solutions can provide fine granularities of restorable objects ranging from crash-consistent images to logical objects such as files, mail boxes, messages, etc.

<http://www.snia.org/education/dictionary/s/>

# Snapshots: making data “stand still”

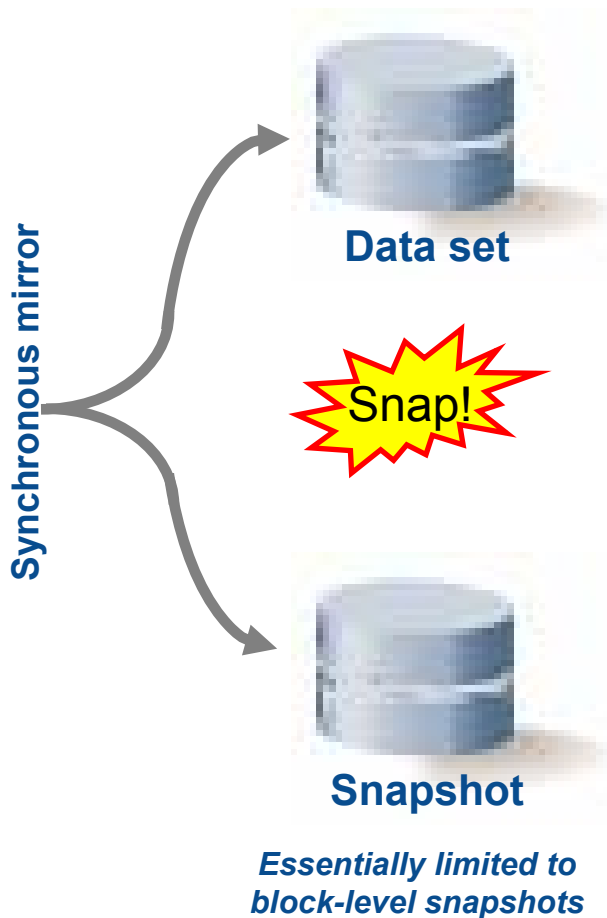


snapshot **CONTEXT [Data Recovery]**  
**[Storage System]**  
A fully usable copy of a defined collection of data that contains an image of the data as it appeared at the point in time at which the copy was initiated. A snapshot may be either a [duplicate](#) or a [replicate](#) of the data it represents.  
<http://www.snia.org/education/dictionary/s/>



[Link to +/- details \(35\)](#)

# Full-size and space-saving snapshots

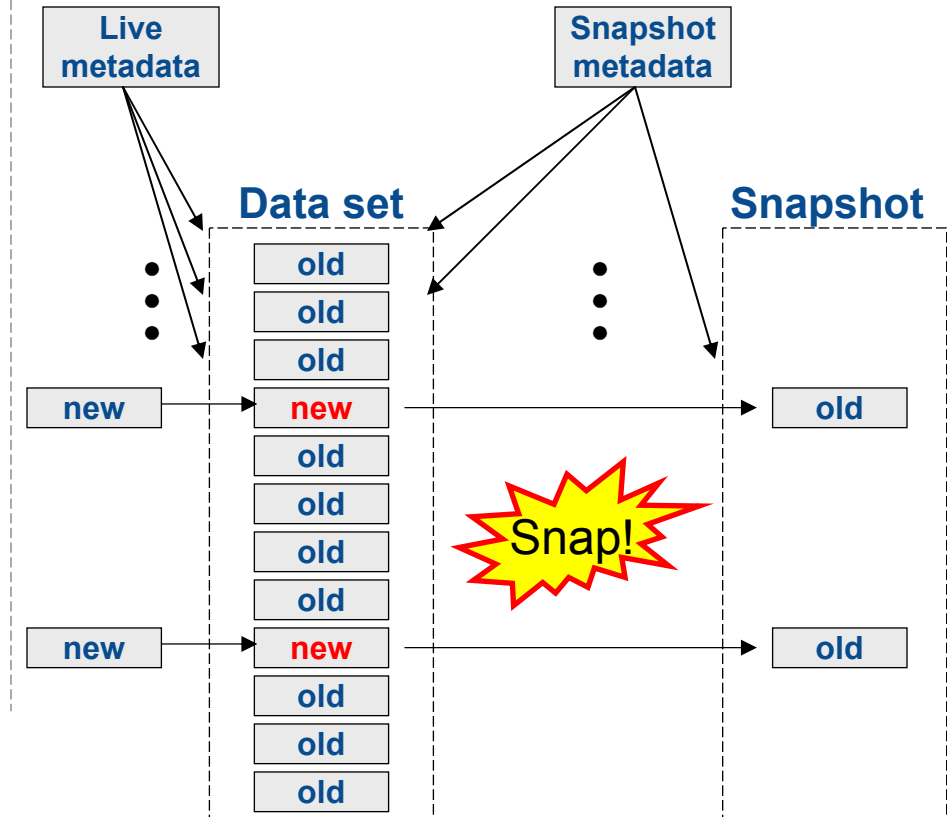


copy on write

**CONTEXT [Storage System, Backup]**

A technique for maintaining a point in time copy of a collection of data by copying only data which is modified after the instant of replicate initiation. The original source data is used to satisfy read requests for both the source data itself and for the unmodified portion of the point in time copy. cf. [pointer remapping](http://www.snia.org/education/dictionary/c/pointer-remapping)

<http://www.snia.org/education/dictionary/c/>



*Copy-on-write vs pointer mapping*

[Link to +/- details \(36\)](#)

# Which kind of snapshot is best ?

- It depends on your definition of “best”
  
- Full-copy (“split mirror”) snapshots
  - ◆ Common feature of block-level SAN storage
  - ◆ Almost no impact on production data
  - ◆ High cost in storage and synchronization
  - ◆ Protect against physical destruction of data set
  - ◆ Greatest value is off-hosting
  
- “Copy-on-write” and “pointer-mapped” snapshots
  - ◆ Space-saving (⇒frequent recovery points)
  - ◆ Some (usually minor) impact on application write performance
  - ◆ Do not protect against physical destruction of data set

# Techniques for protecting NAS data

*Backup may not be so cheap*

Conventional backup	Cost advantage	Replication
Tape drive, media, and library	←	Online storage capacity
Physical transportation	←	Replication link with adequate bandwidth and latency
Acquisition or activation of recovery site	←	Full-time recovery facility premises and staff
Cost of downtime to restore data (hours to days)	→	Seconds to minutes
Cost of data loss due to recovery point (hours to days)	→	Zero to seconds

## ➤ Logical

- ◆ Equipment and facilities remain physically intact
- ◆ Data has been destroyed or corrupted

## ➤ Physical

- ◆ System failure  
*The surrounding IT environment is intact*
- ◆ Data center loss  
*The entire IT environment must be recreated*

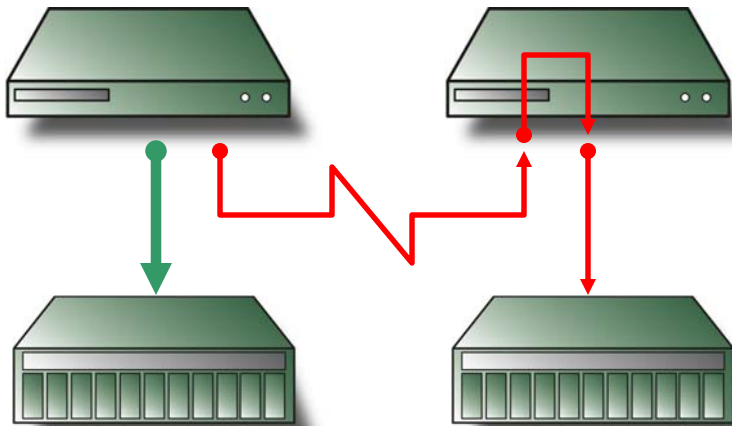
	<b>System failure</b>	<b>Data center loss</b>
<b>Definition</b>	<p>Failure beyond the reach of usual techniques like...</p> <ul style="list-style-type: none"> <li>▪ RAID/mirroring</li> <li>▪ NAS head “clustering”</li> <li>▪ Network redundancy</li> </ul>	<p>Entire IT environment incapacitated</p> <ul style="list-style-type: none"> <li>▪ Storage systems</li> <li>▪ App servers</li> <li>▪ Local clients</li> <li>▪ Connections to remote clients</li> </ul>
<b>Needed to recover</b>	<ul style="list-style-type: none"> <li>▶ Intact, accessible copy of critical production data</li> <li>▶ Connectivity to app servers</li> </ul>	<ul style="list-style-type: none"> <li>▶ Intact recovery site</li> <li>▶ Staff (including provisioning)</li> <li>▶ Hardware and connectivity for critical apps</li> <li>▶ Intact, accessible copy of critical production data</li> </ul>

- Have or acquire adequate equipment for access to critical data and applications
  
- Have or recreate a copy of critical data at the recovery location

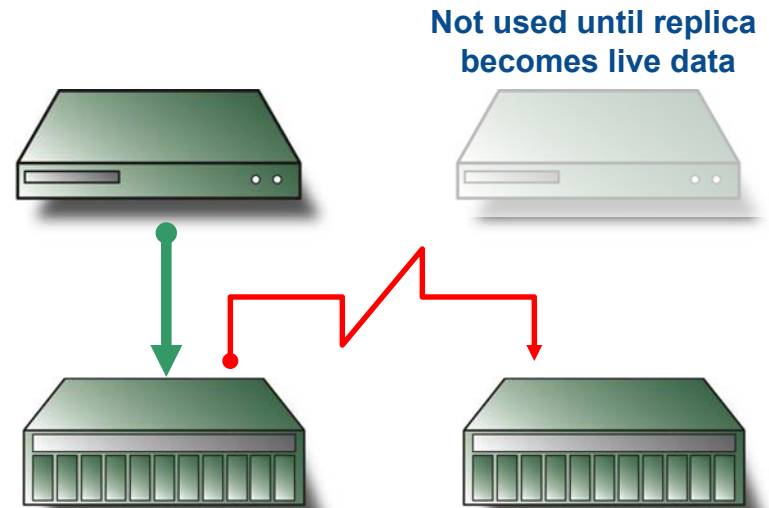
- **Have** or **acquire** adequate equipment for access to critical data and applications
- **Have** or **recreate** a copy of critical data at the recovery location
- The difference between **have** and **{ acquire }  
{ recreate }** is the distinction between high and “normal” availability

- Basic purpose: keep an up-to-date copy of a data set on separate storage resources  
*(usually attached to separate processing resources)*
  
- Use cases
  - ◆ Recovery from physical disaster
  - ◆ Data distribution/consolidation
  - ◆ Second data source for read-only applications
  - ◆ Baseline for writable “clones” of data
  
- Not just mirroring
  - ◆ Primary-secondary relationship
  - ◆ Time ordered updates to a “consistency group” of devices
  - ◆ May be asynchronous

## “Host” (application server)-based



## Storage system-based

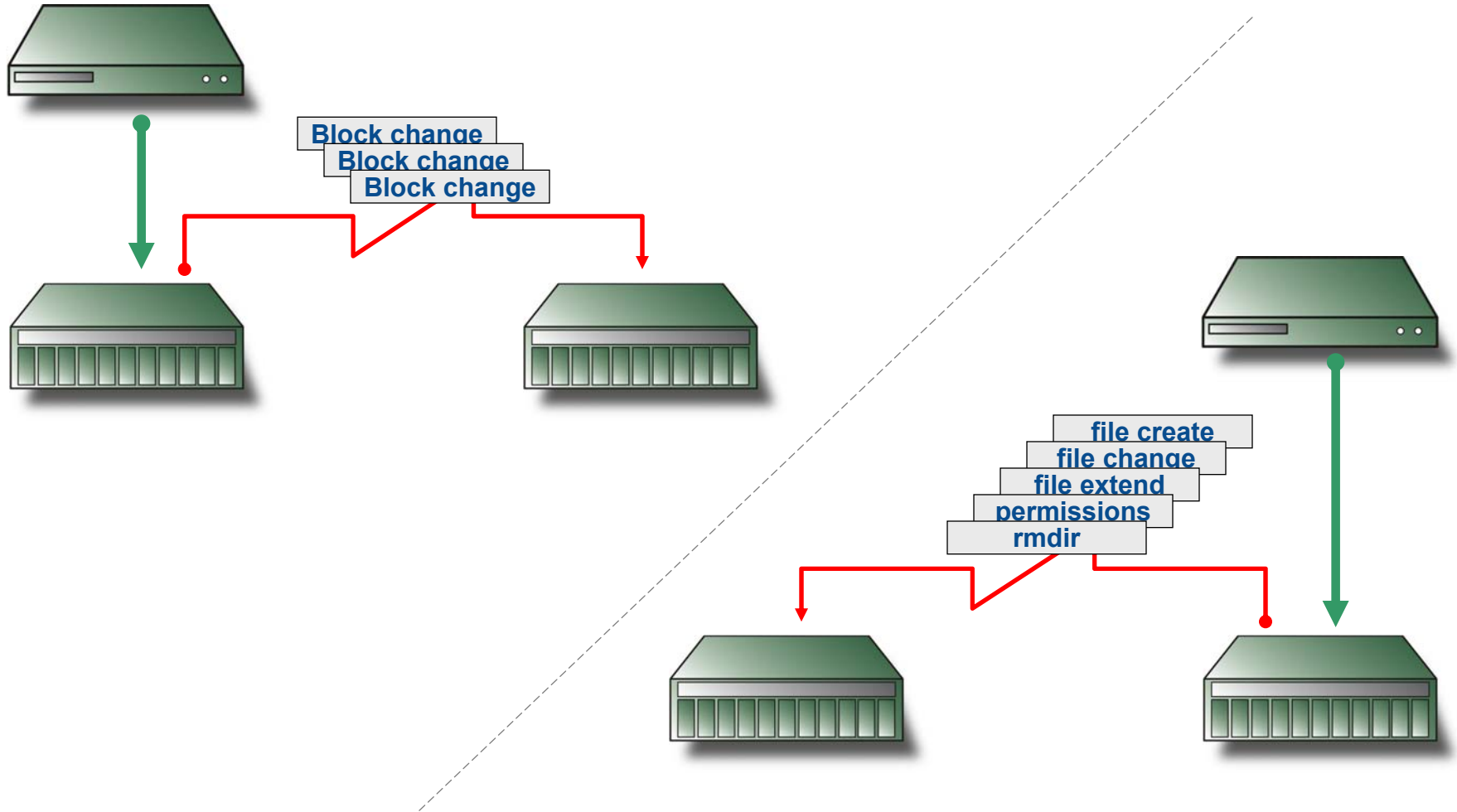


## ➤ Other variants:

- ◆ SAN switch, NAS aggregator, dedicated appliance
- ◆ All are roughly equivalent to storage system-based replication

— Client updates  
— Replication traffic

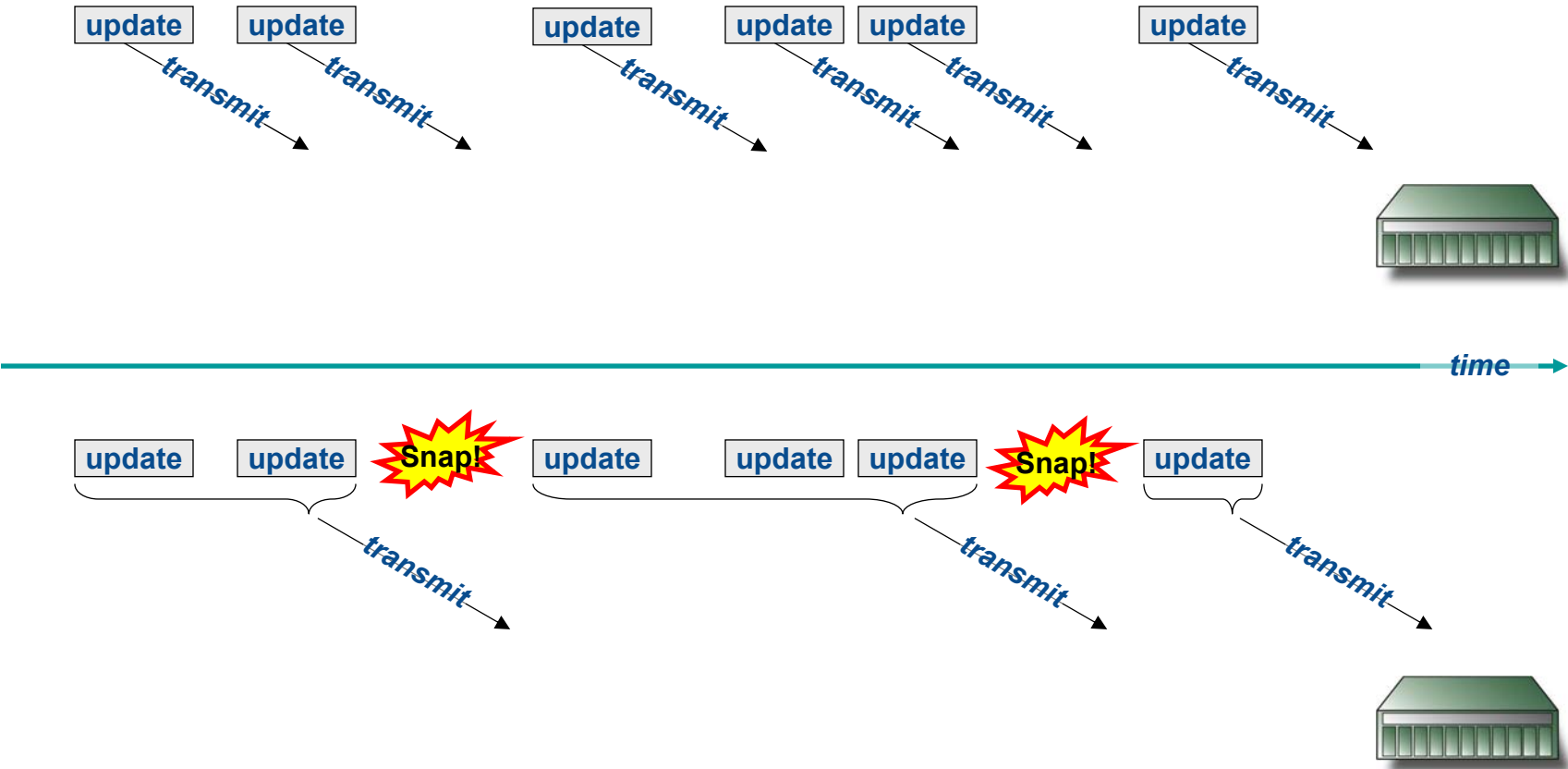
# “Block” and file-level replication



— Client updates  
— Replication traffic

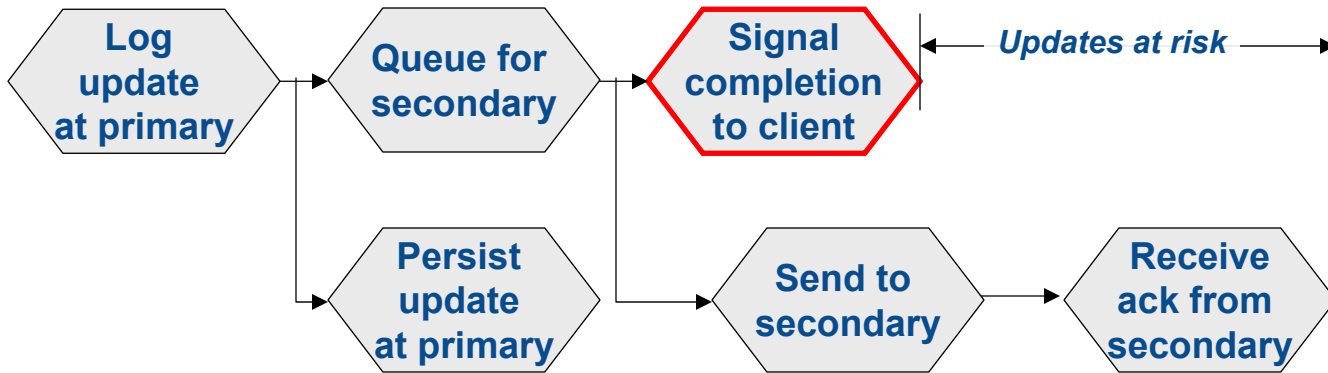
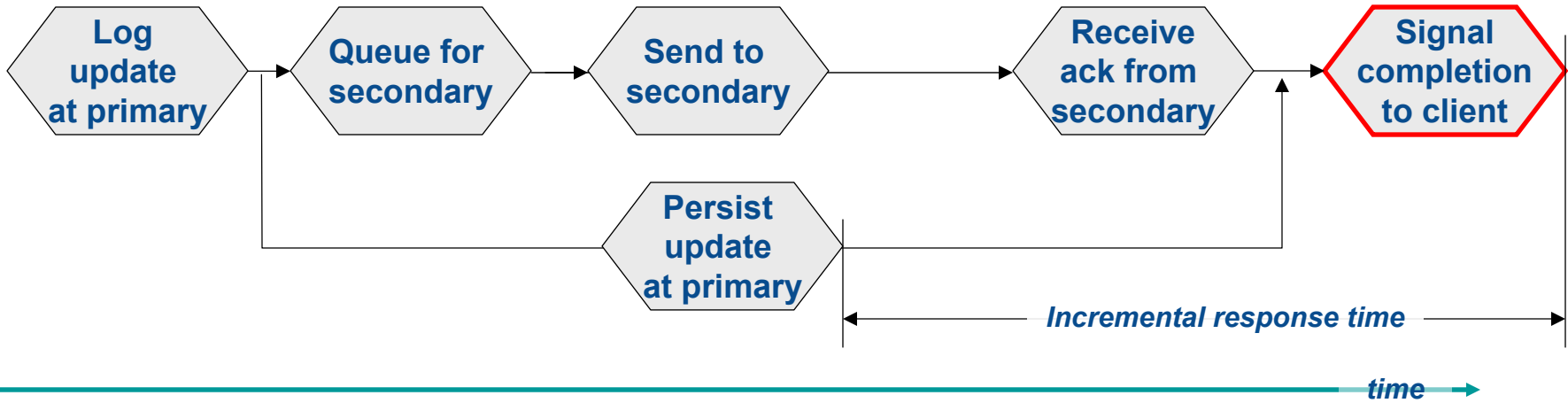
[Link to +/- details \(41\)](#)

# Real-time and batch replication



[Link to +/- details \(42\)](#)

# Synchronous vs asynchronous replication



› Variation on a theme:  
*"limited asynchronous replication"*

[Link to +/- details \(43\)](#)

# Which kind of replication is best ?

- It depends on your definition of “best”
  
- Block-level vs. file-level
  - ◆ Universal availability vs. bandwidth and readability
  
- Synchronous vs. asynchronous
  - ◆ Potential for data loss vs. client responsiveness
  
- Real-time vs batch update
  - ◆ Bandwidth vs. recovery point granularity

## ➤ Different threats→different recovery techniques

### **Logical disaster**

- ◆ Technique: “turn back the clock”
- ◆ Property: Inherently some data loss

### **Physical disaster**

- ◆ Technique: Recovery facility + replica of critical data
- ◆ Property: Cost vs. recovery point tradeoff

## ➤ Different data→different requirements

- ◆ Critical: RTO = seconds; RPO = now!
- ◆ Less-critical Cost becomes a factor

## ➤ NAS is unique because it “knows” about data structure

- ◆ Able to participate proactively in protection and recovery processes

- Please send any questions or comments on this presentation to SNIA: [trackfilemgmt@snia.org](mailto:trackfilemgmt@snia.org)

**Many thanks to the following individuals  
for their contributions to this tutorial.**

*SNIA Education Committee*

**David Dale  
Rob Peglar  
Warren Avery  
Norman Owens**

**Netapp  
Xiotech  
storagenetworking.org  
storagenetworking.org**

# Availability objectives: RTO

Recovery time objective	Requirements
Days	Prepare and staff facilities and hardware Acquire backup copy of data and restore Replay logs, restore app environment and restart apps
Hours	Restore data from onsite incremental backup Replay logs, restore app environment and restart apps
Minutes	Replay logs against data replica and activate Restart apps Restore client connections
Seconds	Detect failure (avoiding false positives) Switch data replica to live mode Switch apps to live or full-service

# Availability objectives: RPO

Recovery point objective	Recovery point is a consequence of the data preservation technique
Days	Time of newest backup copy
Hours	Time and location of newest incremental backup
Minutes	Amount of time by which data replica “lags” live data
Seconds	Zero

# Techniques for backing up NAS data

## Backup management software

Advantages	Limitations
<p data-bbox="144 368 600 486">“Set-and-forget” automation</p> <ul data-bbox="241 511 672 753" style="list-style-type: none"><li>Schedules</li><li>Device and media management</li><li>Data selection</li></ul>	<p data-bbox="967 382 1103 429">Cost</p> <ul data-bbox="1064 468 1624 586" style="list-style-type: none"><li>License &amp; maintenance</li><li>Training</li></ul>
<p data-bbox="144 786 755 839">Added-value features</p> <ul data-bbox="241 863 710 1053" style="list-style-type: none"><li>Incremental backup</li><li>Multi-streaming &amp; multiplexing</li></ul>	<p data-bbox="967 796 1638 929">Requires data to “stand still” during backup</p>
<p data-bbox="144 1096 929 1229">Application (e.g., database) integration</p>	<p data-bbox="967 1096 1715 1300">Inherently coarse-grained recovery times &amp; recovery points</p>

# Snapshots: what to snap ... “blocks” vs files

Blocks	Files
Fast creation (short time to usability)	Creation implies some metadata duplication
Block atomicity	File system operation atomicity
Unit of expansion: virtual volume	Unit of expansion: file system dependent

*NAS developers have the luxury of  
choosing either*

<b>"Split mirror"</b> (content = bit-for-bit copy)	<b>"Copy-on-write"</b> (content = changes only)
<b>Protects against device failures and media defects</b>	<b>Requires separate physical protection mechanism</b>
<b>Storage requirement: full data set size</b>	<b>Storage requirement: <math>\propto</math> change in data set size during snapshot life</b>
<b>Overhead to maintain: zero</b>	<b>Overhead to maintain: <math>\approx</math> 2-3x for every "first write"</b>
<b>Deletion cost: Resynchronization with data set</b>	<b>Deletion cost: Space reclamation</b>

Host-based	Storage system-based
Uses app server processing resources	No processing impact on app server
Arbitrary consistency groups (e.g., volumes from multiple arrays)	Consistency group limited to storage system's scope

Block replication	File replication
Sends every block update over the replication link	Sends file system operations over the replication link (uses less network bandwidth)
Replicates file system operations I/O by I/O (i.e., is “bug-compatible”)	Performs source and target file system actions independently
No context for block updates: (Replica is not usable during replication)	File system operation-atomic (Replica can be used by read-only apps during replication)

Real-time (op-by-op)	Batch (periodic snapshots)
Sends repeated ops repeatedly	Uses network bandwidth more efficiently
Replicates every primary data set state (any app that can recover from local faults is recoverable)	May not represent all primary data set states in the replica (may affect recoverability)



# Synchronous vs asynchronous replication

Synchronous	Asynchronous
Data update is at the replica before client I/O completes  (No data lost in a disaster)	Data is queued for transmission to replica before client I/O completes  (committed updates may be lost)
“Round trip” time is additive to client response time	Very little increment in client response time

## ➤ High Availability and Disaster Recovery for NAS systems

---

### ➤ Learning objectives:

- -deliver knowledge of the techniques available for protecting data stored on NAS systems against component failures, system failures, and wide-scope disasters
- -improve students' ability to select and implement NAS HA/DR solutions that are appropriate to business value of data assets

### ➤ Outline

#### ➤ I. The NAS HA/DR data protection problem space summary

- -physical vs logical failures and recovery techniques
- -protecting data vs. preserving accessibility
- -component failures and protection/recovery techniques
- -system failures and protection techniques
- -data center disasters and recovery techniques

#### ➤ II. Protecting data against component failures

- -differences between SATA, SAS, and Fibre Channel disk drives
- -what mirroring and RAID can and cannot protect against
- -mirroring and RAID metrics--inline performance, recovery time, recovery impact
- -matching the data protection technique to data requirements

#### ➤ III. System failures and data replication

- -replication vs mirroring
- -forms of replication: synchronous vs. asynchronous
- -forms of replication: continuous vs. episodic
- -restarting applications using data replicas

#### ➤ IV. Recovering NAS data from disasters

- -what's unique about NAS data in a DR context (replica asynchrony and network addressing)
- -NAS system failures (excepting disk drive failures)
- -"dual-head" NAS clusters
- -"shared nothing" NAS clusters
- -combining HA and DR--inherent costs and properties of the dual-head and shared-nothing techniques