Education

File Systems for Data Protection

Windsor Hsu, Ph.D.
EMC

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.

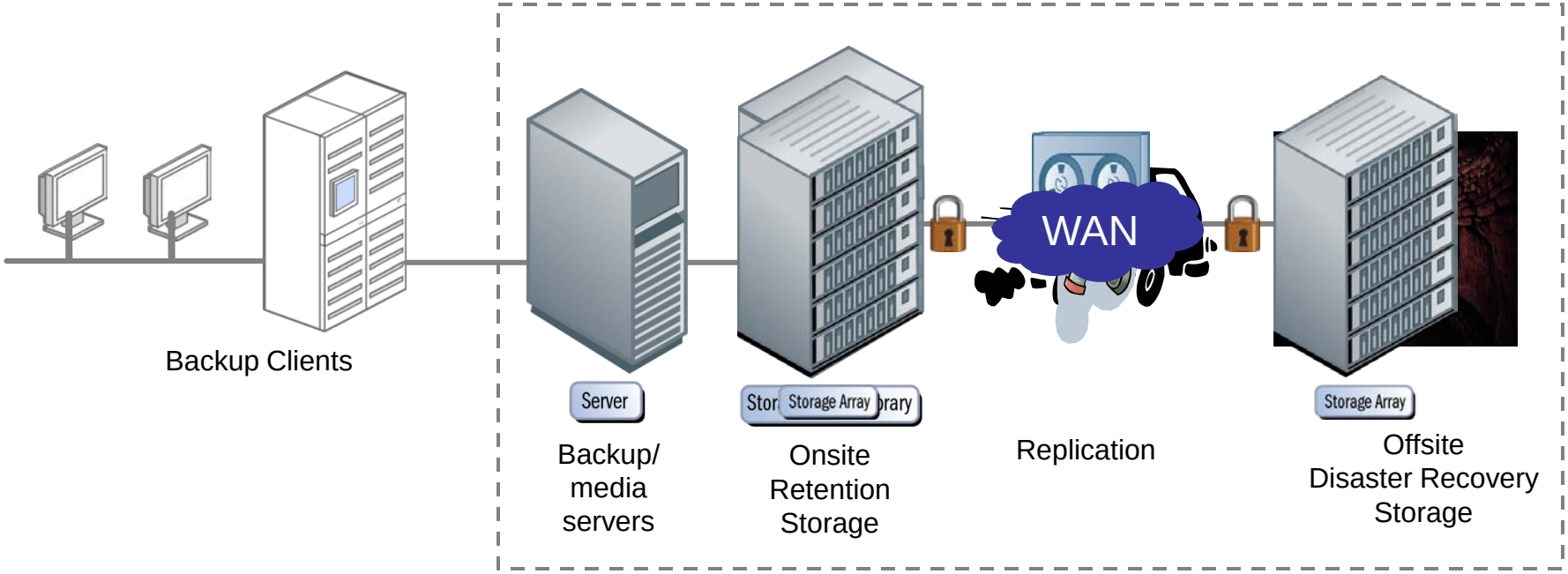
  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Abstract

File Systems for Data Protection

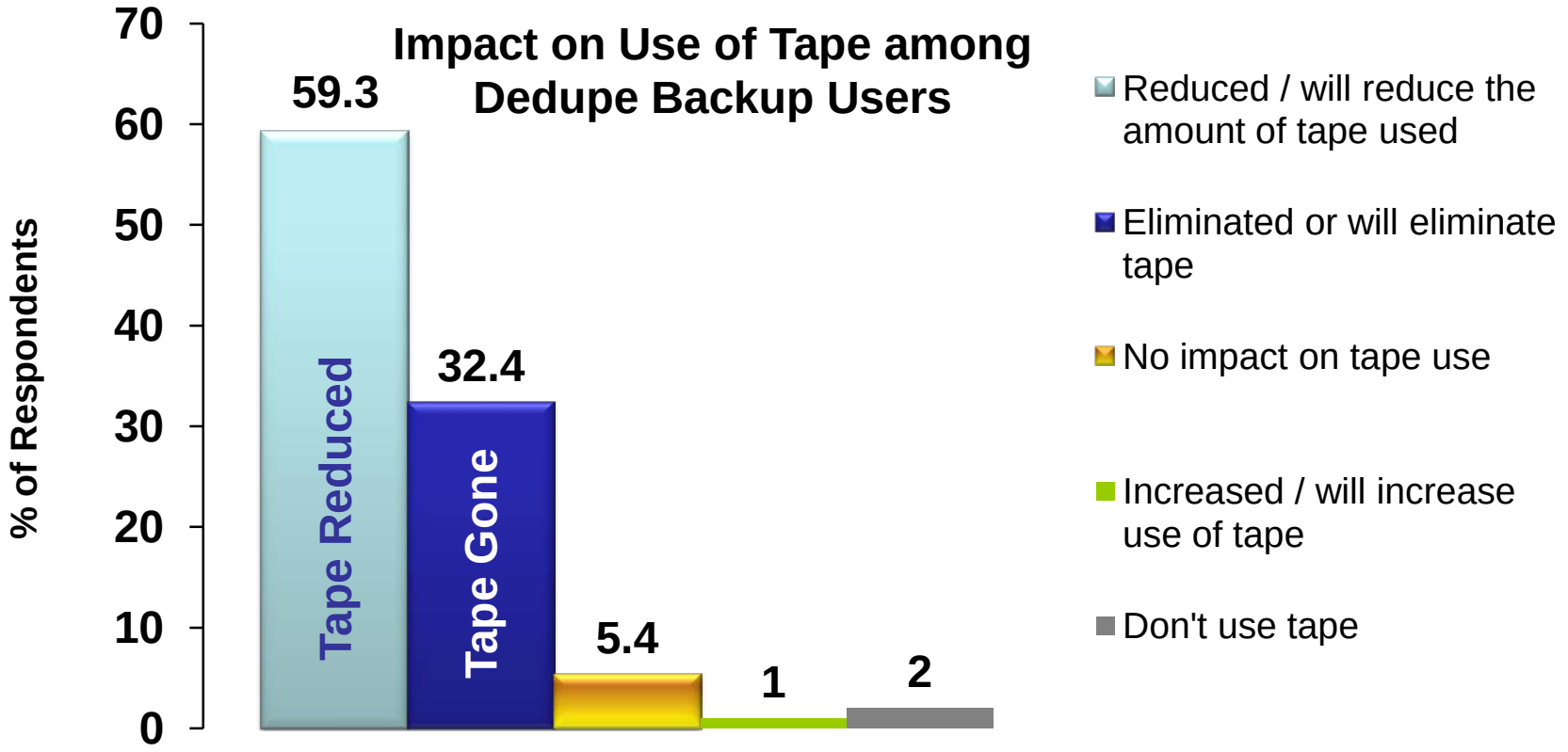
- There is a strong industry trend to adopt disk-based platforms for data protection. But not all such systems were designed to satisfy the requirements for data protection. This tutorial will explain some of the differences between a general purpose file system and a file system designed for data protection. Participants will
  - appreciate that file system targeted at data protection has unique requirements
  - understand at an overview level the different techniques for satisfying these requirements and the operational implications
  - become equipped to make informed decisions about selecting a file system for data protection

Data Protection Transformation



- Deduplicating storage systems take role of tape libraries
- Replication of reduced data for WAN Vaulting

Tape Use Decline

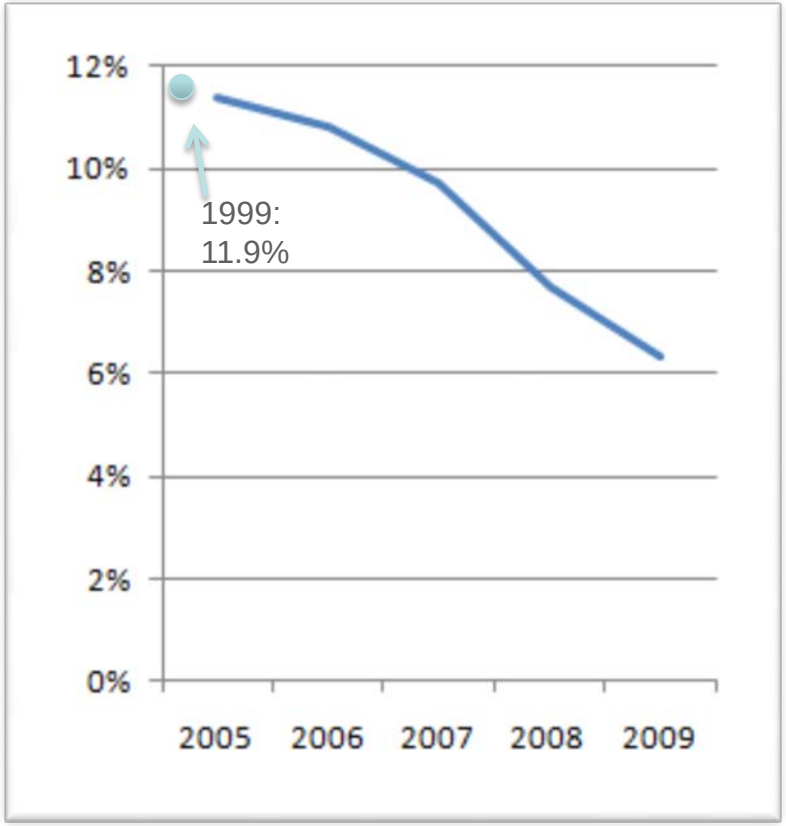


Note: Notes: N=204, data is not weighted, Base=All respondents who have seen/expect to see specific results with tape reduction/elimination
Source: IDC Deduplication Study, November 2009

Tape Automation Market Decline

- Tape use in re-evaluation
- Restore is moving to disk
- DR will be from WANs

Tape Automation Factory Revenue
% of External Storage Factory Revenue



Source: IDC Tape Automation Worldwide Forecast 2010-2014, April 2010

Protection Storage versus Primary Storage

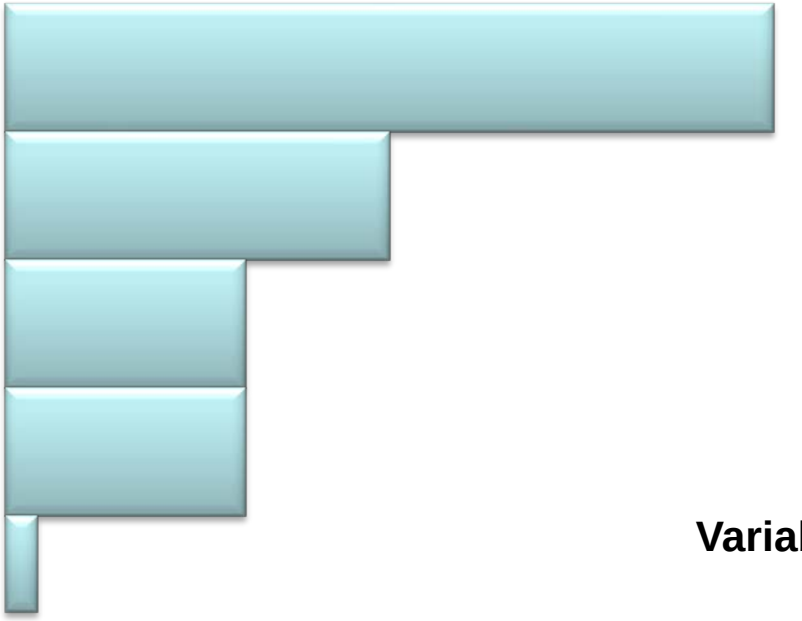
	Primary Storage	Protection Storage
Workload	Continuous random accesses Mostly reads Lots of meta-data accesses	Large batches of accesses Mostly writes Few meta-data accesses
Cost	Dollars / IOPS	Dollars / TB
Performance	Latency and throughput, IOPS	Sequential throughput
Safety	Backup to protection storage	Storage of last resort

Storage for Data Protection

- Cost (Low $/TB)
  - Compete with tape
  - Think Tier Z
- Performance (High Sequential Throughput)
  - Finish backups within shrinking backup window
- Scale and Scalability
  - Effectively handle large and growing data sets
- Safety
  - Must work, last line of defense against data loss

Cost: Low $/TB

Data Reduction



Regular Storage Array
1:1

LZ Compression
~ 2:1*

Single Instance Storage
~ 3:1*

Fixed Block
~ 3:1*

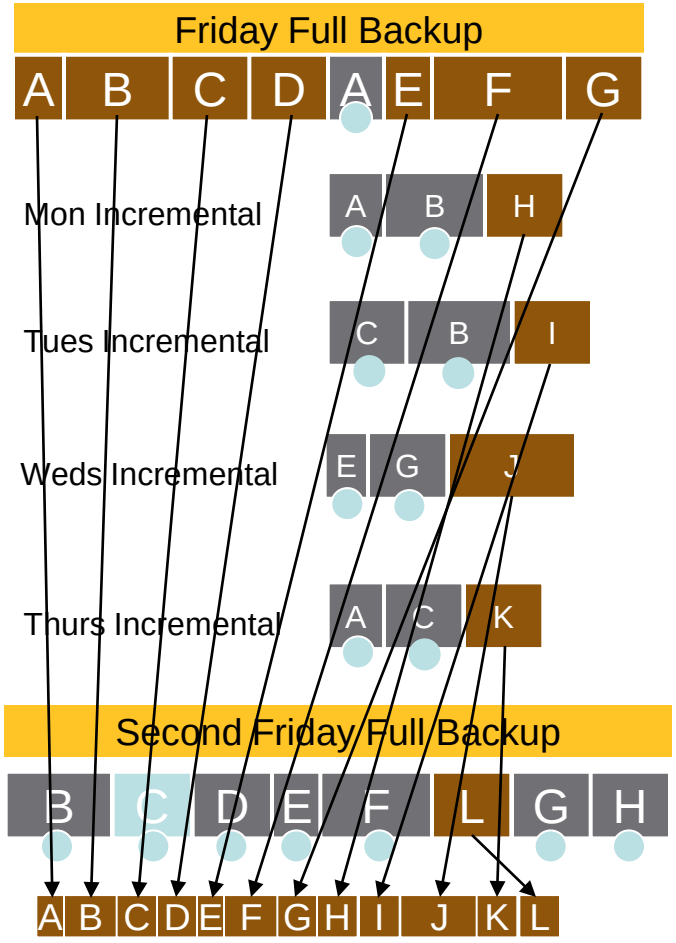
Variable-size Segment (Data Deduplication)
~ 20:1*

For backups, data deduplication achieves massive data reduction and savings in

- ***Replication WAN Bandwidth***
- ***Power***
- ***Heat***
- ***Cooling***
- ***Management***

****Typical data reduction rate for backups; actual rate depends on data and algorithm***

Data Deduplication: Technology Overview



Backup Data	Logical	Estimated Reduction	Physical
FRIDAY FULL	1 TB	2–4x	250 GB
Monday Incremental	100 GB	7–10x	10 GB
Tuesday Incremental	100 GB	7–10x	10 GB
Wednesday Incremental	100 GB	7–10x	10 GB
Thursday Incremental	100 GB	7–10x	10 GB
Second FRIDAY FULL	1 TB	50–60x	18 GB
TOTAL	2.4 TB	7.8x	308 GB

Data Deduplication: Store More for Longer with Less

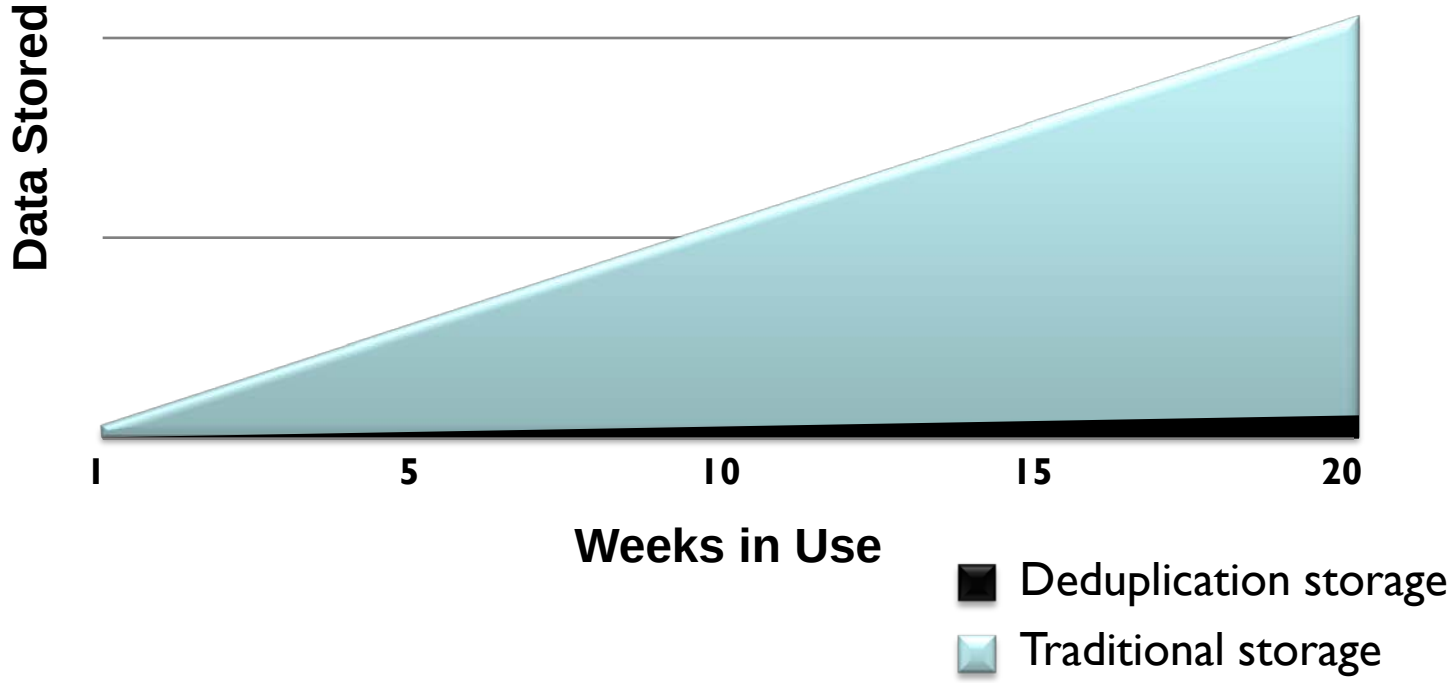
	Backup Data	Cumulative Logical	Estimated Reduction	Physical
	First Full	1 TB	4x	250 GB
Week 1	April 7	2.4 TB	8x	308 GB
Month 2	May 31	12.2 TB	17x	714 GB
Month 3	June 30	17.8 TB	19x	946 GB
Month 4	July 31	23.4 TB	20x	1,178 GB
	TOTAL	23.4 TB	20x	1,178 GB

Rule of Thumb:
Physical capacity needed for 3 months of data protection ~ logical size of data protected

Data Deduplication: Store More for Longer with Less

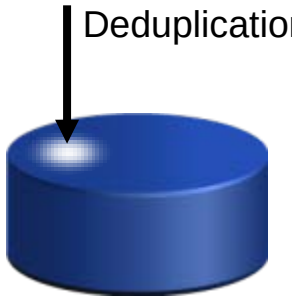
Education SNIA

10–30 times less data stored versus
fulls + incrementals with typical retention policies



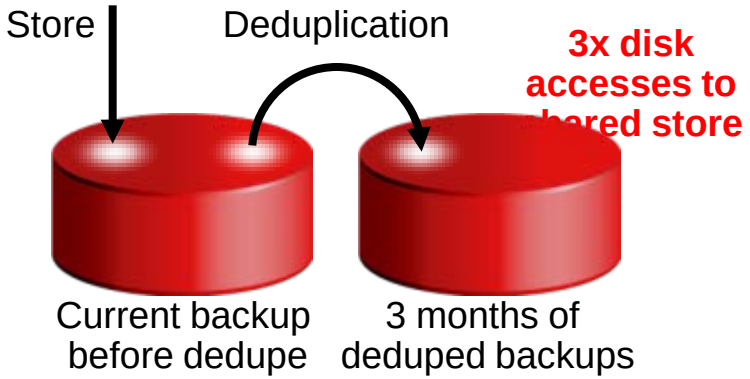
Inline vs Post-Process Deduplication

INLINE
Deduplication Before Storing



- Other activities unimpeded
  - Predictable
  - Simpler

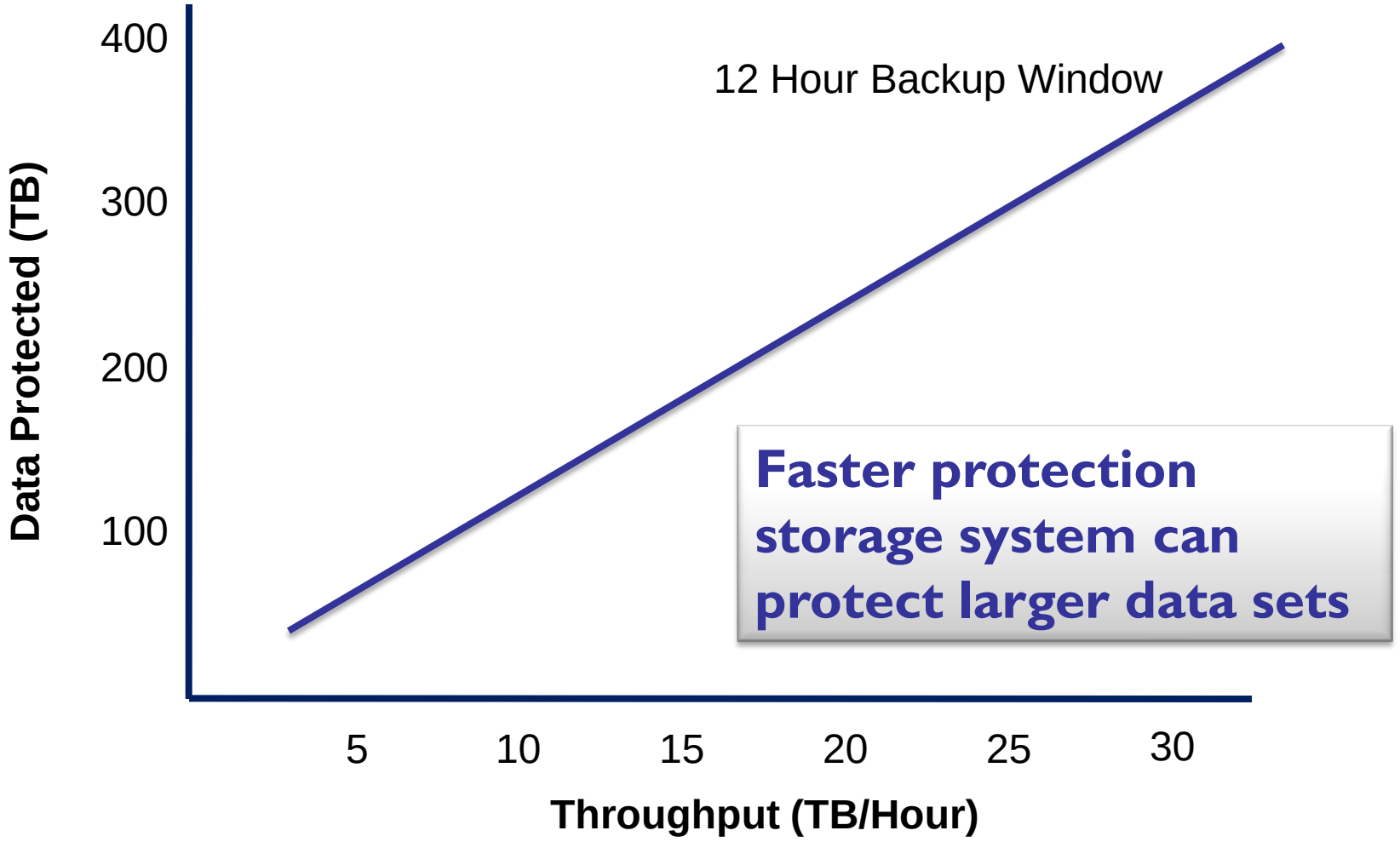
POST- PROCESS
Deduplication After Storing



- The more processes, the more resource contention
  - Copy to tape: Too slow to stream tape
  - Recovery: Unpredictability in meeting SLA
  - Replication: Poor time-to-disaster-recovery
  - Deduplication: Delayed if interleaved with backup/restore
- More administration to fight these issues

Performance: High Sequential Throughput

Speed Matters



Challenge in Fast Deduplication

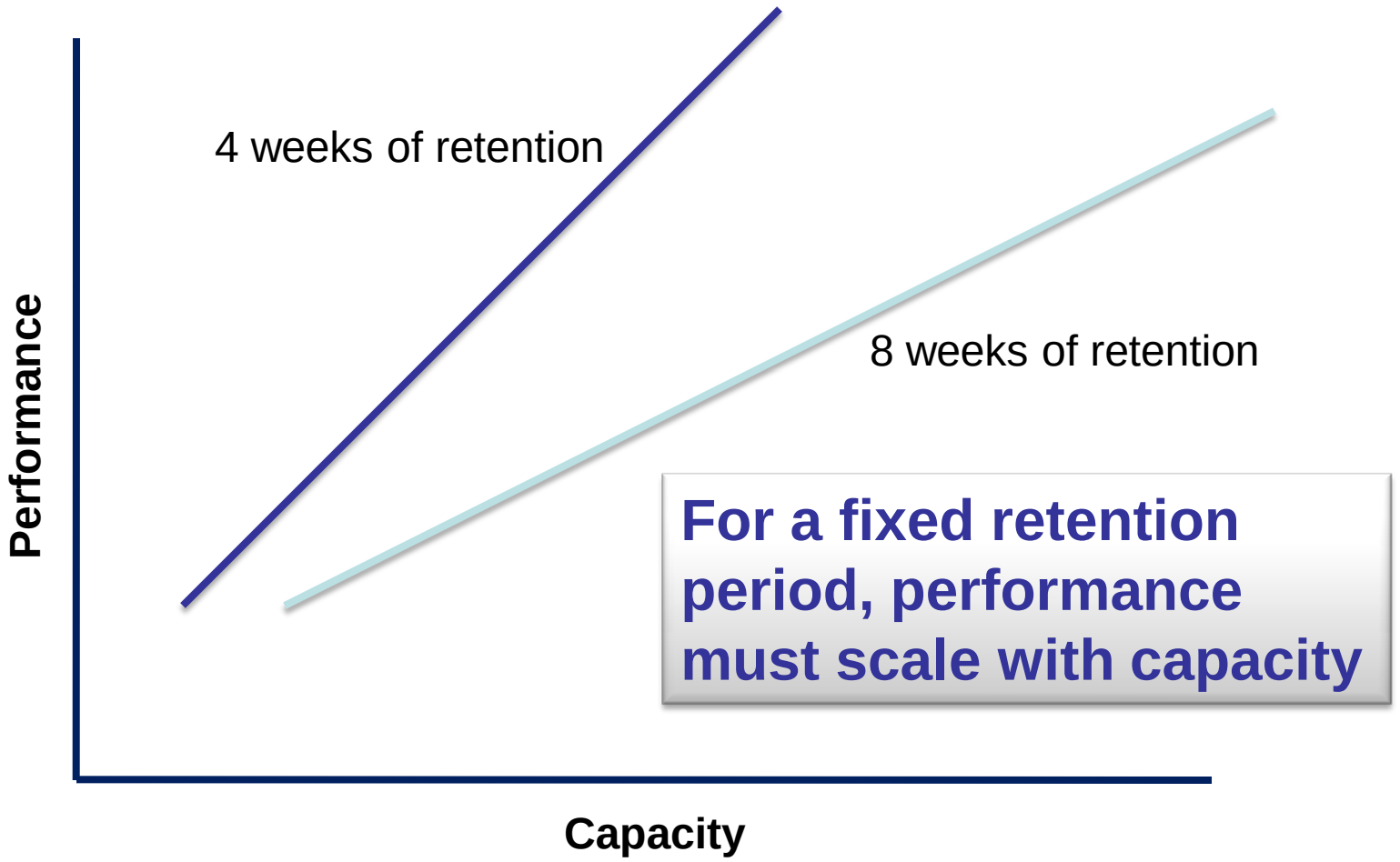
- Bottleneck is in looking up index to determine whether segment is already in system
  - Each disk: 1 MB/sec
    - E.g. for a 7.2K RPM SATA drive: <120 seeks/sec
    - 120 seeks/sec with 8 KB segments: 0.96 MB/sec/disk

How to Dedupe Fast?

- Parallel disk lookups
  - Need 1000 disks to go 1 GB/sec
- Fast disks for index of segments already in system
  - Use FC disks, short-stroke disks, use SSDs, etc
- Index in RAM
  - RAM needs to be ~1% of storage
- Cache index in RAM
  - Needs temporal or spatial locality to work
- Stream Informed Segment Layout (SISL)
  - Segments tend to arrive in same order so store segments in order received

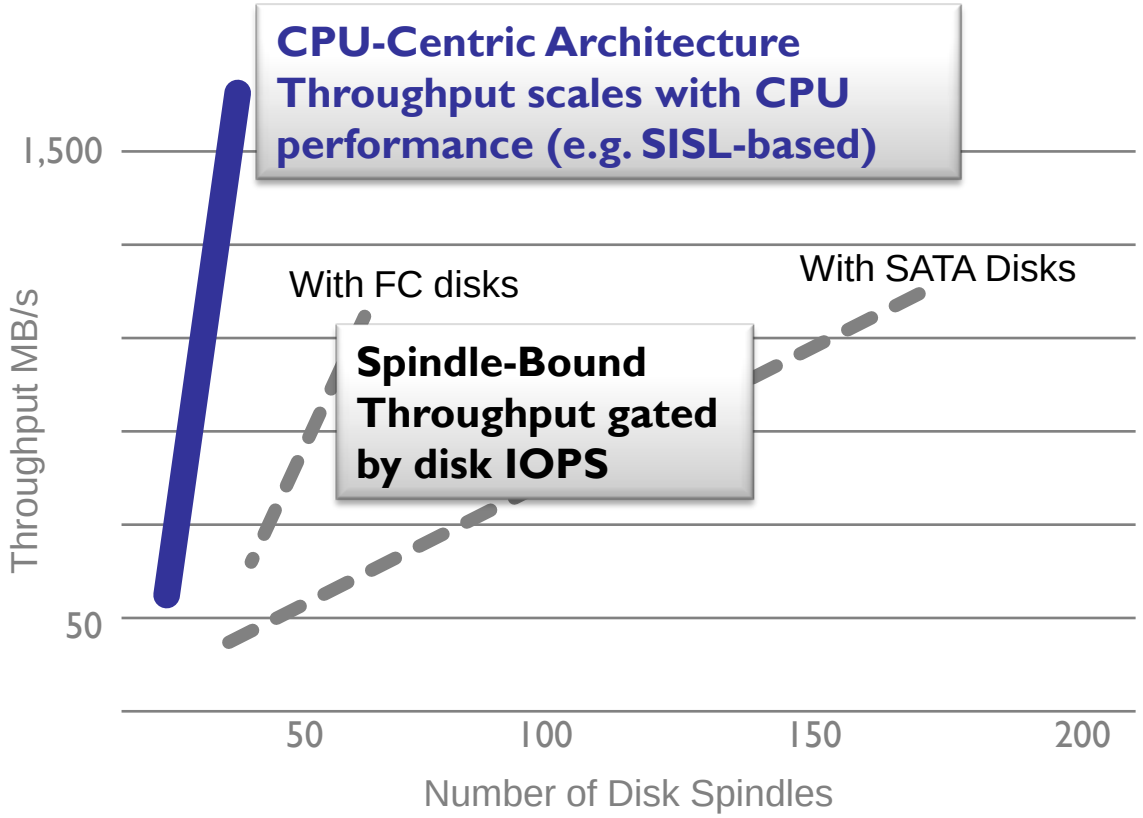
Scale and Scalability

Proportionate Performance Scaling



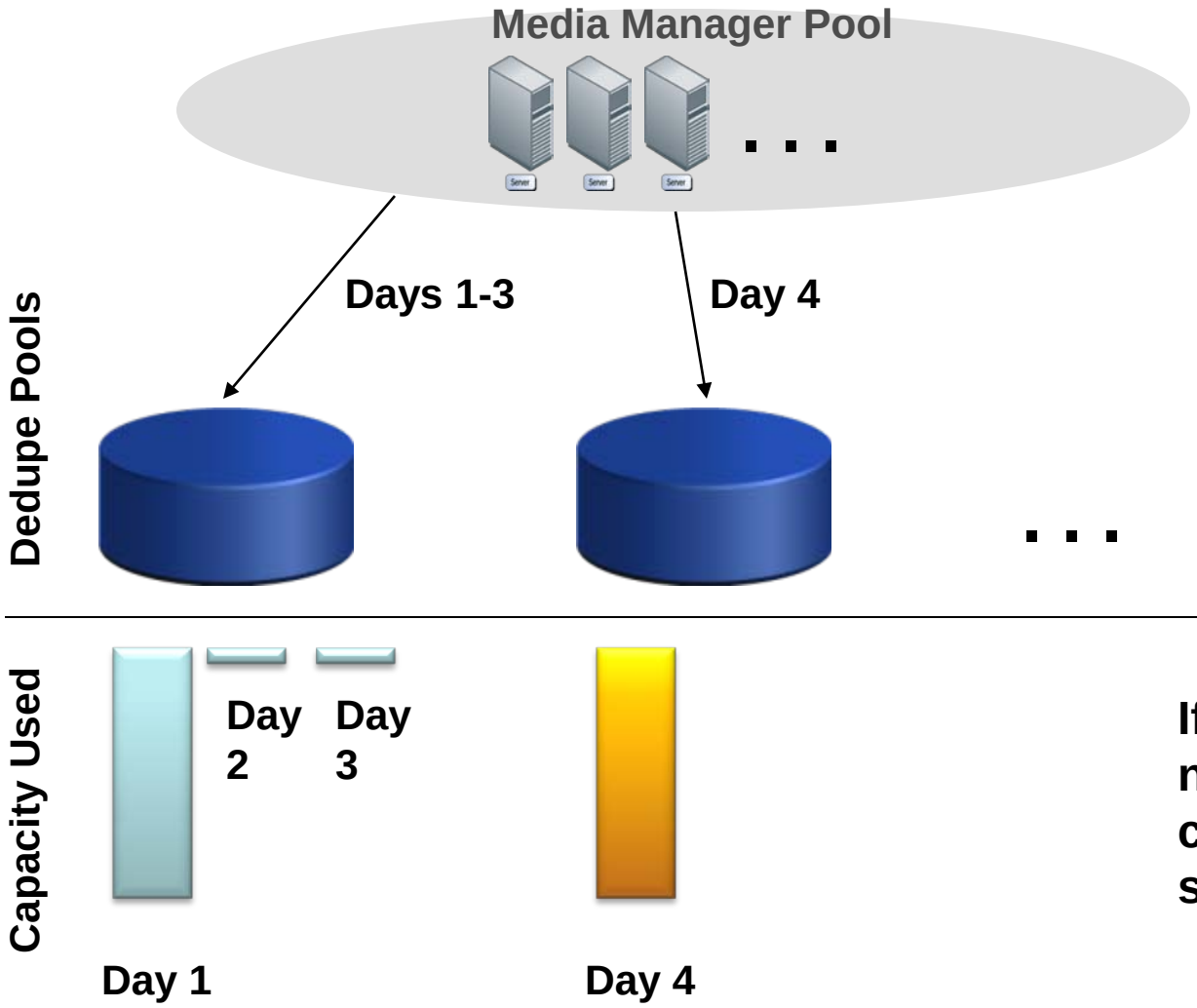
For a fixed retention period, performance must scale with capacity

CPU-Centric versus Spindle-Bound



- Data sets are growing rapidly
- CPU-centric architecture provides required scalability

Challenge of Large Consolidated Data Centers



If backup rerouted to new dedupe domain, capacity required is same as first full

Large Dedupe Domain

- Best practice: Send same data set consistently to same dedupe domain
- Challenge: As data set grows, difficult to rebalance workloads among dedupe domains
- Solution: Large dedupe domain
- Results:
  - Simplified capacity management
  - Higher storage utilization
  - Higher backup success rate

Safety

Storage of Last Resort

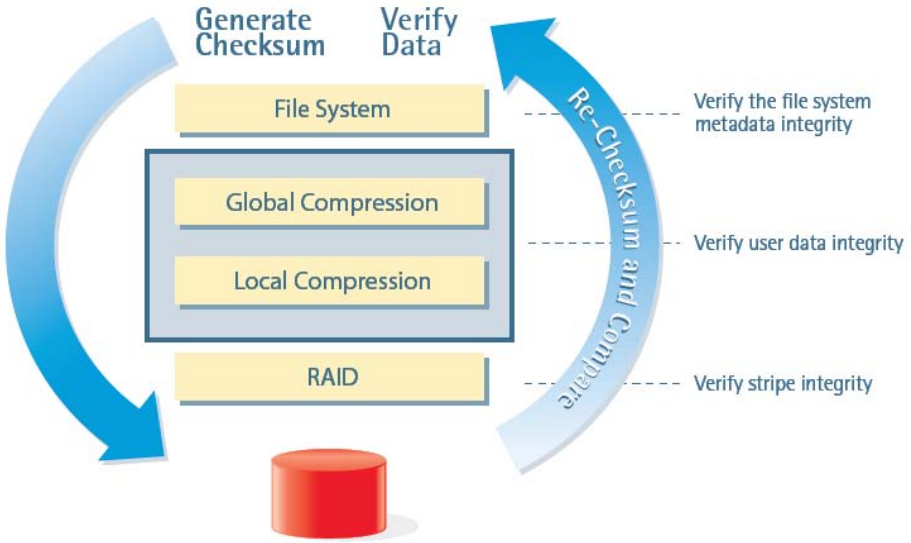
- When primary storage fails
  - Restore from backup
- When backup fails
  - No copy to restore from
  - Data is lost

Design for Data Protection

- Make sure data is stored correctly when originally backed up
- Make sure new backups do not put old backups at risk
- Make sure old backups stay correct and recoverable
- Be prepared for the worst, design the system for recoverability
- Maintain a copy in another location

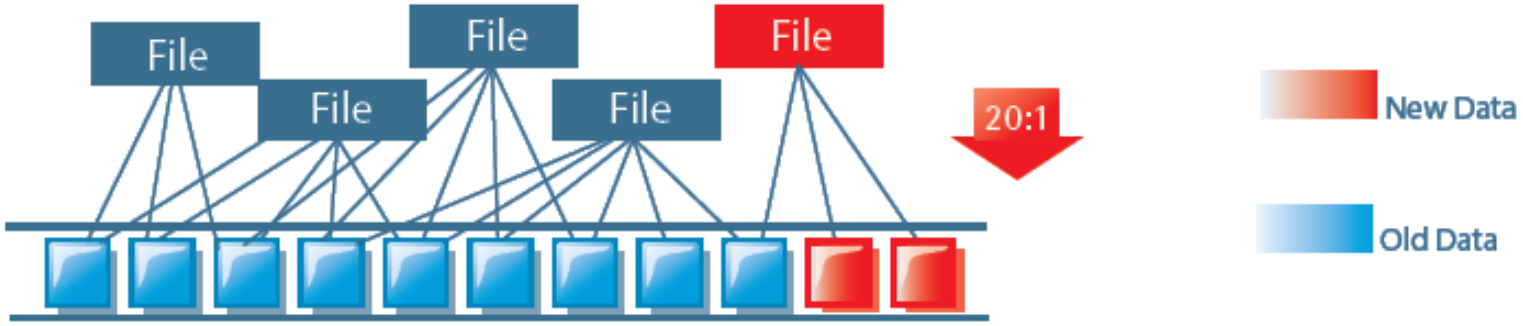
End-to-End Verification

- Test recoverability asynchronously after backups
  - File system consistency
  - Data integrity on disk
- Primary storage can't verify after write
  - It would be too slow
  - Primary storage discovers problems during restore



The end-to-end check verifies all file system data and metadata

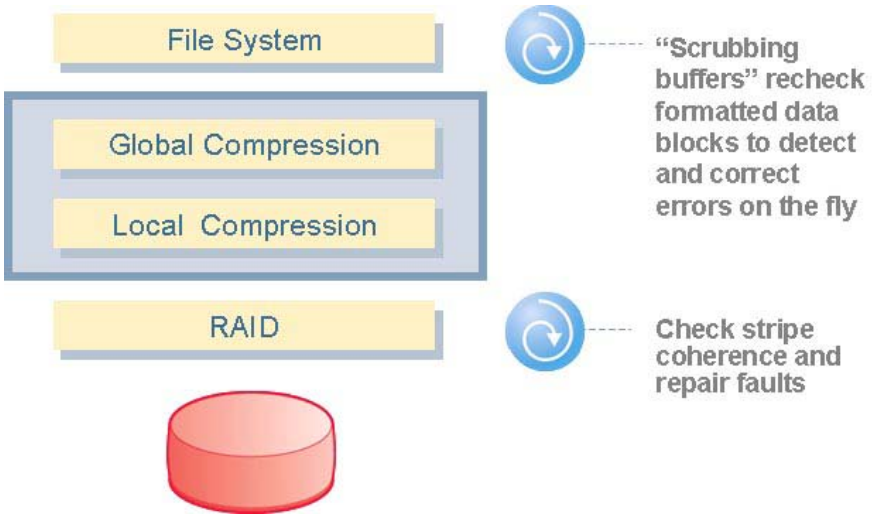
Fault Avoidance and Containment



- Overwriting good data with new data puts previous backups at risk
- Use log-structured file system
- Eliminate partial-RAID-stripe writes

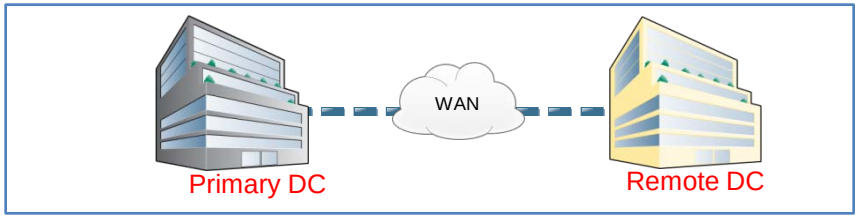
Fault Detection and Recoverability

- Verify all data read is correct
- Verify all data, even those not read, remain correct
- Ensure all metadata structures can be rebuilt
- Ensure file system check (FSCK) is fast when needed

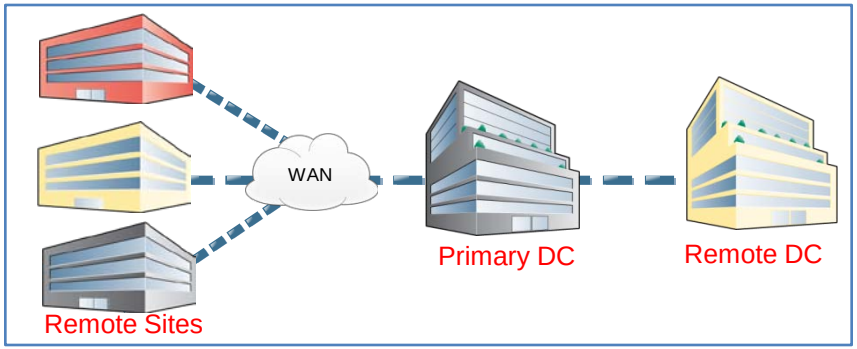


Flexible and Efficient WAN Vaulting

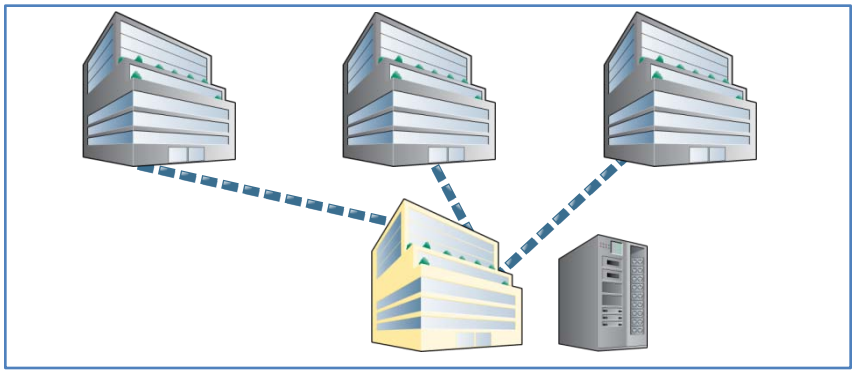
Data Center (DC) Disaster Recovery



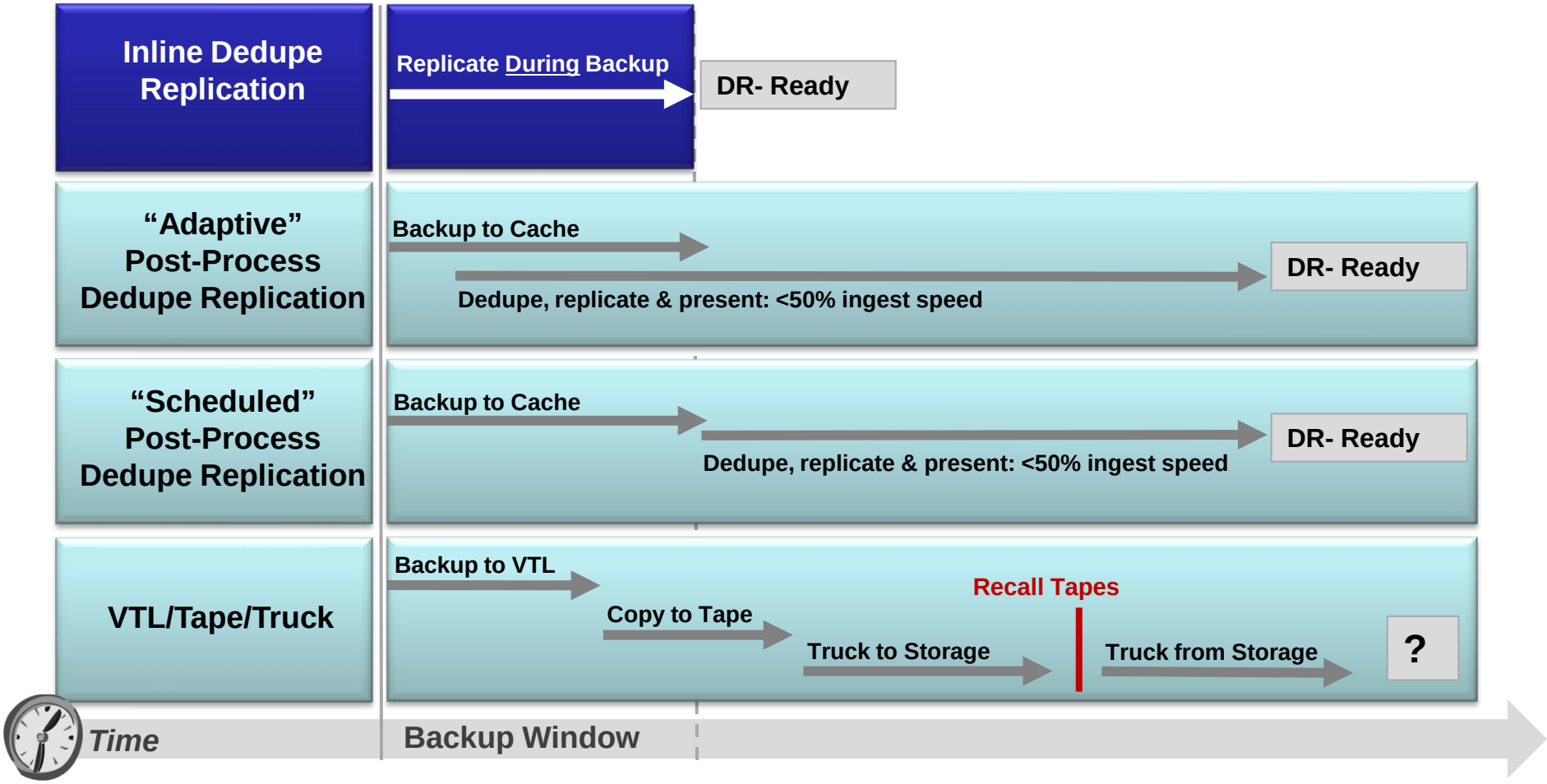
Remote Office Data Protection



Multi-site Tape Consolidation



Time to DR-Ready



Summary

- File system for data protection is different from file system for primary storage
- File system for data protection must offer
  - Fast and effective data reduction
    - CPU-centric inline data deduplication
  - Scalable performance
  - Large deduplication domain
  - Resiliency fitting storage of last resort
  - Efficient and flexible replication

Q&A / Feedback

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

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

- SNIA Education Committee

Windsor Hsu
Hugo Patterson
Joseph White
Nancy Clay