

PRESENTATION DOWNLOAD: BACKUP SCHOOL 2009

# >> Backup School 2009

Presented By

**data**domain

 **IRON MOUNTAIN** DIGITAL™

 **i365**  
A Seagate Company

**Novell**

**pillar**  
DATA SYSTEMS

# Backup School 2009

W. Curtis Preston  
Founder/CTO  
Revolution For IT (RFI)

- Abstract:

This Advanced Backup School seminar presentation is guaranteed to provide you with time-tested advice and techniques that will make backup easier. Renowned backup guru W. Curtis Preston will review the basics – but will also be going into in-depth detail on the newest issues, technologies, and solutions surrounding backup.

## A Little About Me

- When I started as “backup guy” at \$35B company in 1993:
  - Tape Drive: QIC 80 (80 MB capacity)
  - Tape Drive: Exabyte 8200 (2.5 GB & 256KB/s)
  - Biggest Server: 4 GB ('93), 100 GB ('96)
  - Entire Data Center: 200 GB ('93), 400 GB ('96)
  - My TIVO now has 5 times the storage my data center did!
- Consulting in backup & recovery since '96
- Author of O'Reilly's *Backup & Recovery & Using SANs and NAS*
- Webmaster of BackupCentral.com
- Founder of Revolution For IT (RFI)

## A Little bit about RFI

- Revolution For IT (RFI) will remove the fear of the unknown when purchasing or upgrading technology
- NO revenue from vendors (AKA "partners")
- Completely unbiased technical product info
- I'm really excited about my new company and will be happy to talk your ear off about it. Look for me at the breaks!

# Optimizing Your Backup System

Things you might not know about your backup system and what to do about them

# Backup Failures

- Usually caused by mismatch of system abilities to design
- Biggest example is mismatch of tape speed to backup speed
  - Tape failure the number one cause of backup failures
  - Other failures are often tape failures masquerading as other types of failures
- *Must* solve this problem

# Calculating Tape Utilization

- Inventory tapes marked “Full” by backup software
- Calculate average full capacity (FC)
- Find tape’s native capacity (NC)
- Calculate compression ratio (CR = FC/NC=1.5:1)
- Calculate media utilization percentage
  - Total all GBs stored on tape (TGT)
  - Count all tapes with data (TC)
  - Calculate total tape capacity (TTC = TC \* NC \* CR)
  - Calculate percentage (TGT / TTC = Media Utilization)



## Increasing Tape Usage

- All: Reduce the number of pools – especially for offsite tape
- NBU: Do not allow multiple retention periods per tape
- NW: Use Full/Non-Full pools and expire Non-Fulls sooner than Fulls
- NBU/NW: Minimize number of MPX settings
- TSM: Use collocation groups instead of node-level collocation. Spend what you need to in order to get expiration & reclamation done. Start reclamation of emptiest tapes first by slowly lowering your reclamation threshold.

## More Real Money

- Stop spending money on more/faster tape drives to “solve” your problem unless you know that you are streaming the ones you have
- You’re making it worse, not better
- Stream one drive, then use another.
- If you can’t do this, you need to change your design

## Get Better Plumbing

- Move what backups you can off the LAN
  - Use LAN-free backups
  - Use virtual full backups
- Increase LAN throughput any way you can
  - TCP Offload Engine
  - Updated TCP/IP stacks (e.g. Solaris 10)
  - Jumbo Frames
  - 10 GbE (600 MB/s+) backup network
    - Don't buy another backup server, build a network for the ones you have!
    - Backbone's probably not ready, so build your own -- just like we did before Fibre Channel and GbE were ready

## Get a Better Toilet

- Tape can be great *if you keep it happy*, and that's getting harder every day
- Using disk as an intermediary staging device to tape can make it much easier to stream your tape drive
- Storing all *onsite* backups on disk is now more possible and affordable than ever before
- Before you replace your existing tape library with yet another tape library, please seek independent advice on the cost of alternative solutions.

## Macro & Micro

- Monitoring software tends to look at backups at a micro level (Apollo's backup failed!)
- Reporting software tends to look at backups at a macro level (3% of your backups failed)
- A good data protection management product should give you trends on both

## Trends to look for

- Percentages changing

- Overall backup failure %
- Backup failure % by client, server, application
- Partial backup failure
- Media/Drive Utilization

- Growth

- Growth rate should allow you to plan for capacity
- What if the rate changes?

- Consecutive failures

- Is there a client/filesystem/application that keeps failing?
- Will it be automatically noticed and escalated?

# IMPORTANT FEATURES

# Virtual Full Backups

- Available with many products
  - NetBackup Synthetic Full
  - Simpana Synthetic Full
  - NetWorker Saveset Consolidation
  - TSM Backup Sets (not quite)
- Get the power of a full without the data transfer
- Resulting full is the same as a normal full
  - Restores are the same
  - Do not need previous tapes
  - Incremental backups based on the full



# Deduplication

- To be covered more later
- Three methods
  - Client (AKA source dedupe) – NW, NBU
    - Can backup laptops
  - Media Server (can dedupe remote sites) – CV, NBU
  - Backup Server (AKA target dedupe) – TSM, AS

## CDP & Near-CDP

- Continuous incremental backup
  - No more full backups
  - No backup window
- RTO can be near-instantaneous
  - Some CDP recoveries can even be in place
- RPO can be very small (seconds to minutes)
- RPO determined by
  - Frequency of snapshots (if near-CDP)
  - Latency & throughput to remote site

## Disk Backup Targets

- Easier to stream tape drives when copying from backups sent to disk
- A *good* D2D2T system should easily be able to stream any tape drive (not all can do this)
- One reason is that randomly distributed data on source disks is serialized on backup disks
- Recoveries are also easier/faster if at least one copy of all backups is left on disk
- Enables other interesting possibilities

# Disk Staging

- Little bit of disk, lot of tape
- Backup to disk, copy/migrate to tape
- All else is the same
- Requires enough disk for one night's backups (e.g. 14%: 1/28<sup>th</sup> or 4% for full, 10% for incrementals if you do full backup once a month and incrementals daily)
- Helps backups, not restores
- Restores still come from tape
- Still requires shipping tape
- "Those who cannot remember the past are condemned to repeat it." \*

# Disk Backups

- Lot of disk, little bit of tape
- Store all onsite backups on disk
- Offsite backups can be disk or tape
- Requires Bigger shift in thinking & procedures
- Operational restores come from disk
- Requires enough disk to hold all backups (e.g. 2000%: 400% for 4 monthly fulls, 100 days of 15% incrementals = 1000%, 15 10% differentials = 150%)
- Requires deduplication to be as affordable as tape – next session

# Deduplication & Protecting Backup Data

W. Curtis Preston

# What is Deduplication

- Identification and elimination of redundant data at the subfile level
- Not the same thing as “deduplication”
  - CAS
  - Single-instance storage
  - File-level deduplication
- What it can do
  - It really can reduce the amount of disk needed to store backups from 10:1 to 400:1, depending on a number of factors
- What it can't do
  - Find redundancies where there aren't any
    - Seismic data
    - Medical Imaging
    - Between apps of different types (usually)
    - Data encrypted at the source or before the disk target
  - Back up your 20 TB database to 1 TB of disk
  - And it can't do this

## Why Deduplication?

- It makes sense
  - Why store 20 copies? (2, 3, sure. But 20?)
  - We did it on tape because we had to
  - Choose level of resiliency that you want
- It makes disk backups affordable
- It enables other great things, to be discussed later



## Reduce everything to one copy!?!?!?

- Some like the “redundancy” provided by tape
- In reality, 50% do *not* have redundant copies of their data (no copies)
- At *best* they have older backups. Will it really be OK to restore to 3 weeks ago?
- Dedupe allows you to plan for redundancy with replication, and eliminate all others
- If you still want a tape copy, make one!

## Are there really 20 copies?



A user creates a file, stores it on a file server, and it is backed up.

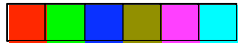


## Are there really 20 copies?



The user modifies the file by putting some new content at the end of the file. This causes the file to get backed up in the next incremental backup.

# Are there really 20 copies?



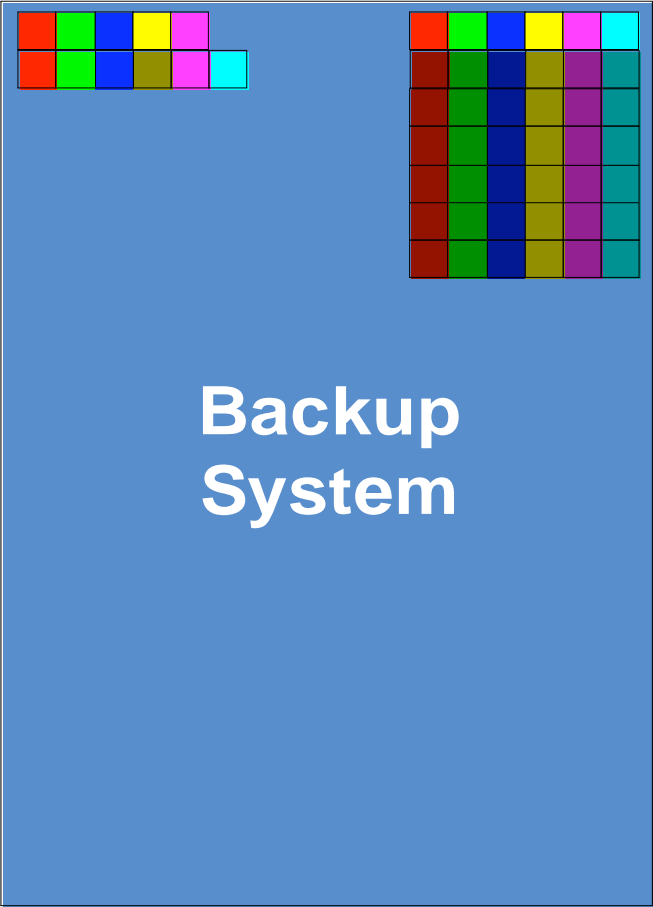
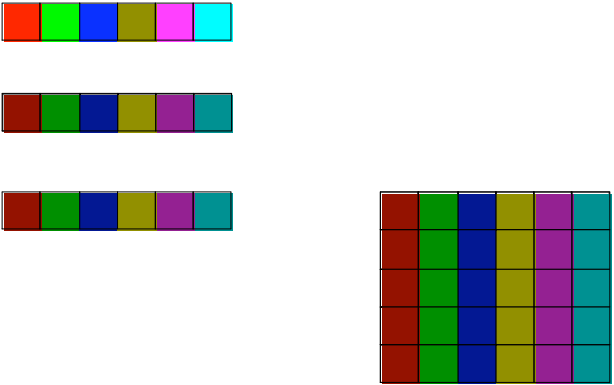
The user modifies the file by putting some new content in the middle of the file. This causes the file to get backed up in the next incremental backup.

## Are there really 20 copies?



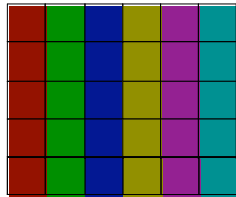
Another user creates another file,  
and it gets backed up.

# Are there really 20 copies?

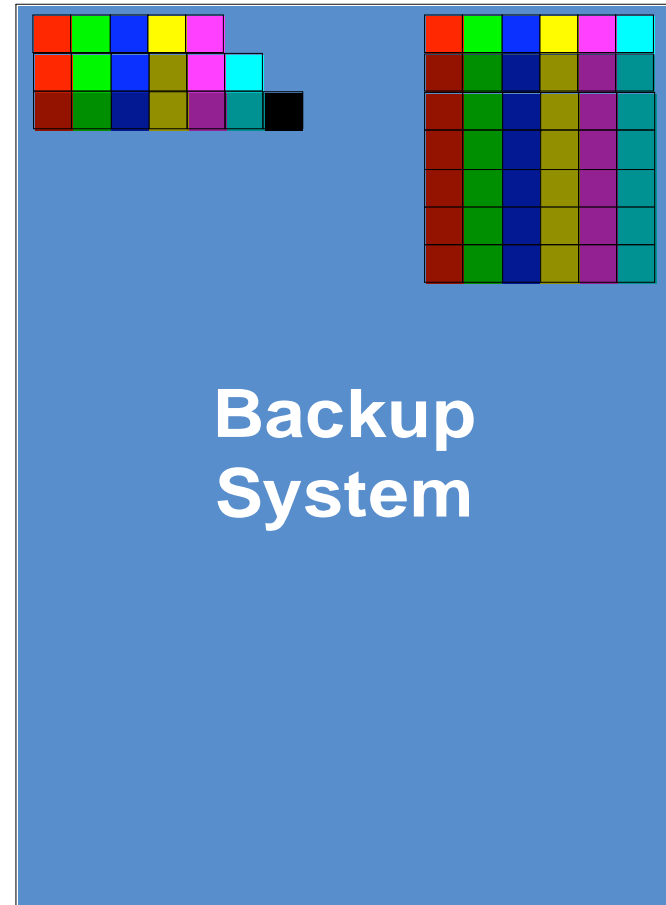


That user emails that file to a number of coworkers. That file is now in several Inboxes. The email server is then backed up.

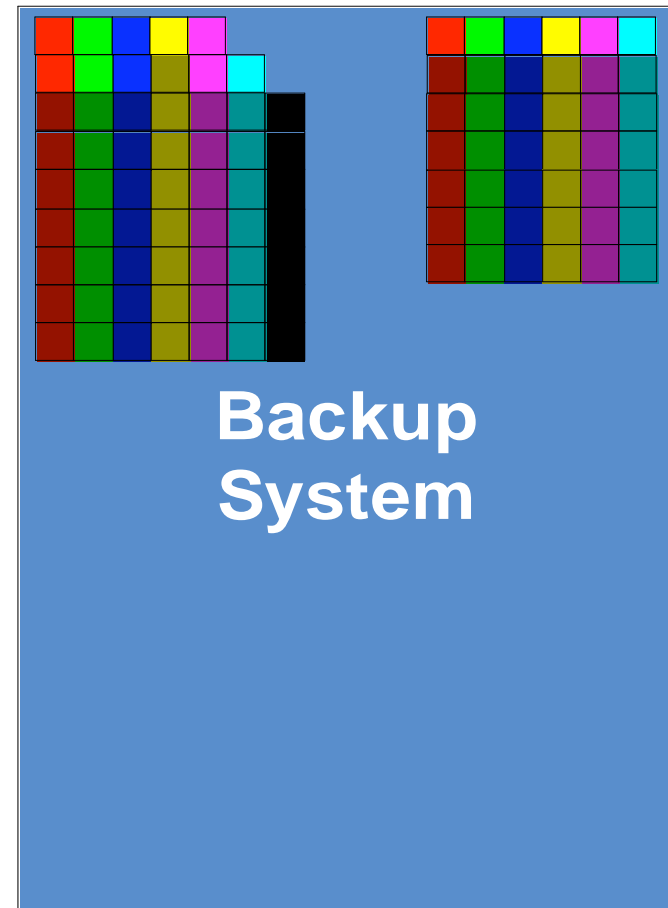
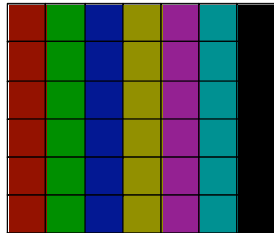
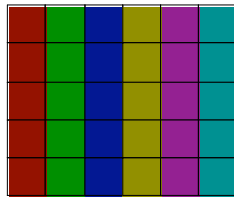
# Are there really 20 copies?



One of the recipients modifies the file and stores it in another location on the file server, causing it to show up on the next incremental backup.



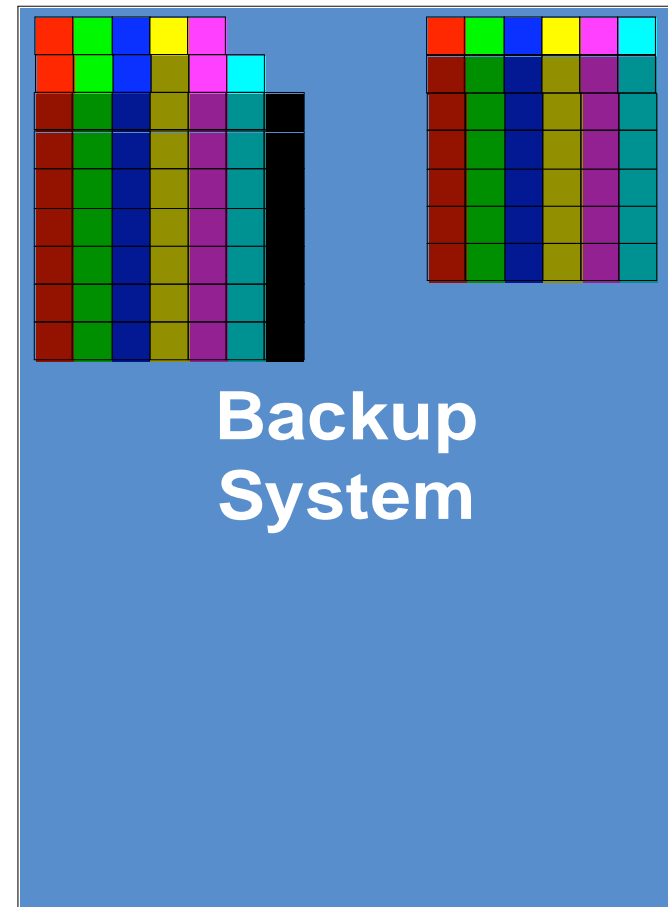
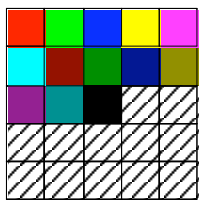
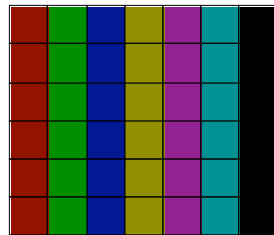
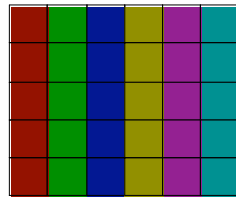
# Are there really 20 copies?



The modified file is sent to the same list of people as before, causing it to in more Inboxes. Since it's an entirely new file, it will get backed up again.

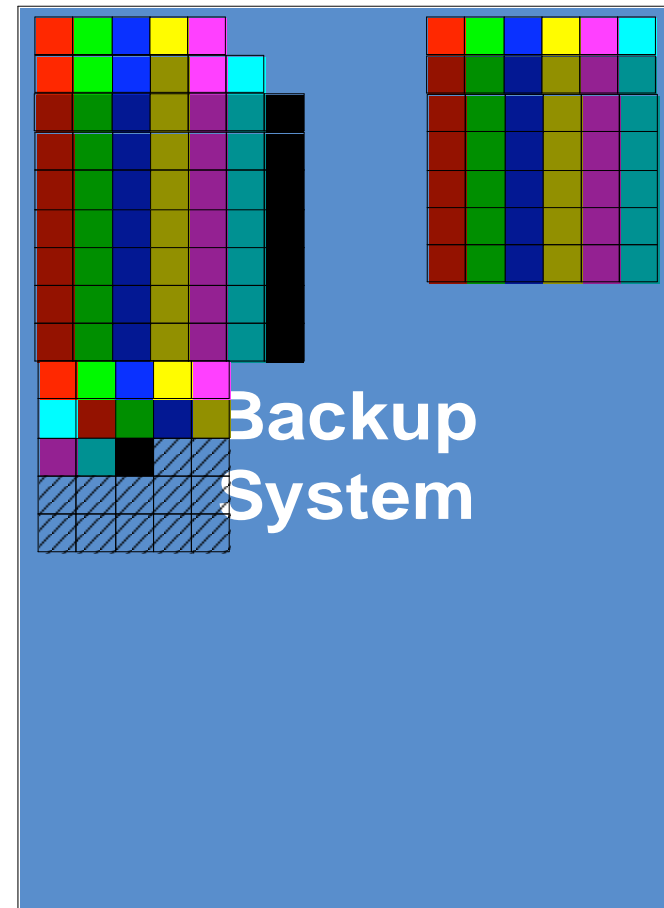
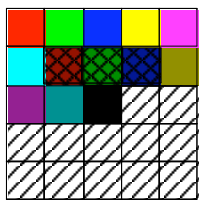
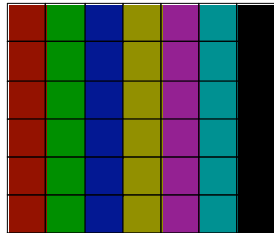
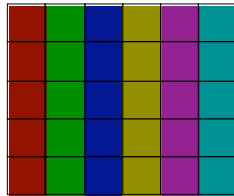


# Are there really 20 copies?



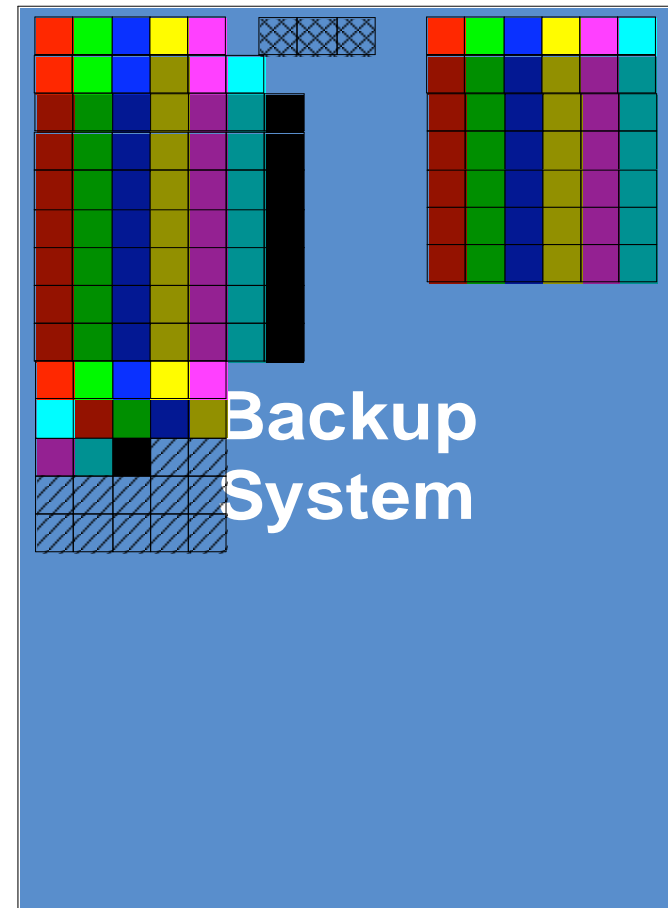
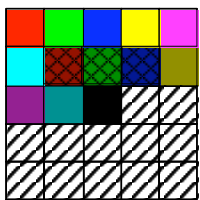
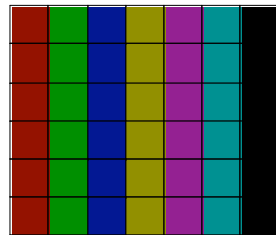
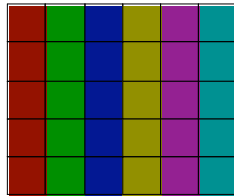
Someone performs a full backup of a database.

# Are there really 20 copies?



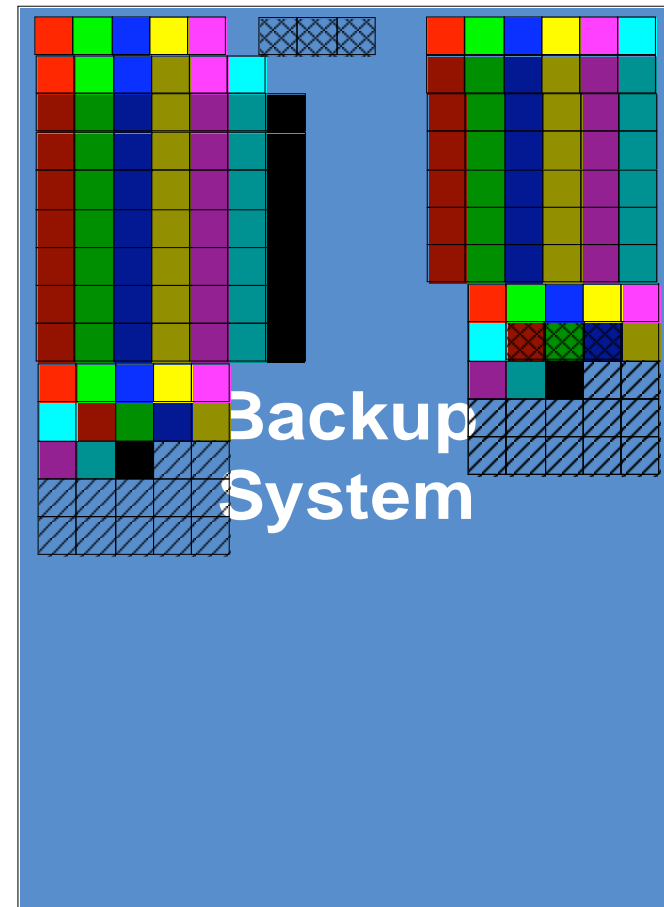
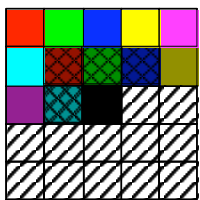
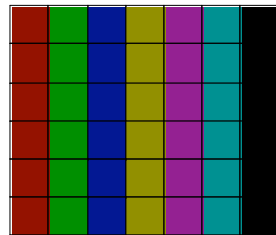
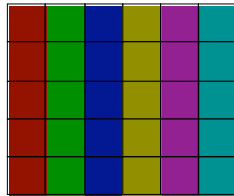
Some records are updated in the database and show up on the next incremental backup of the database.

# Are there really 20 copies?



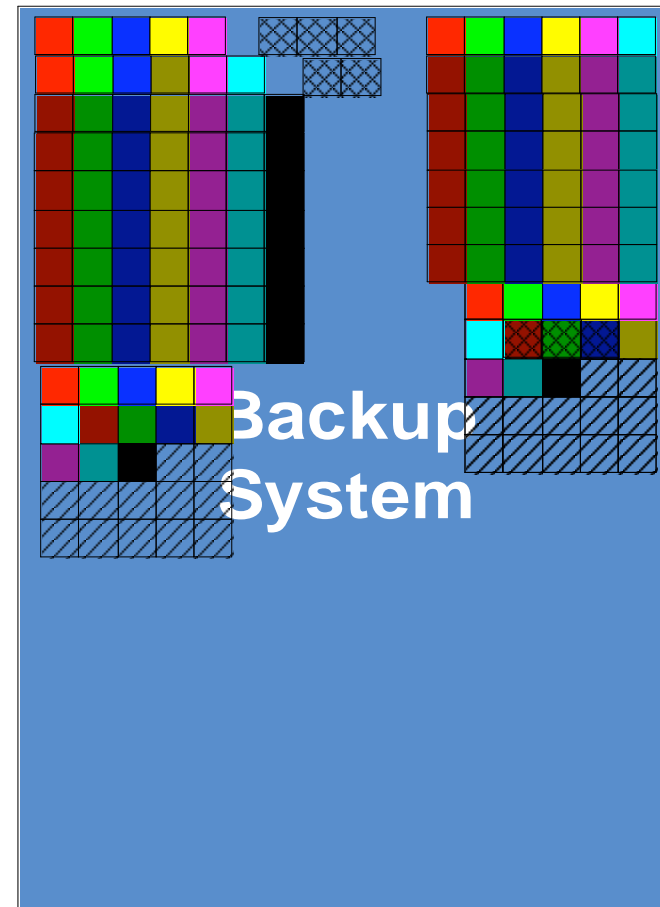
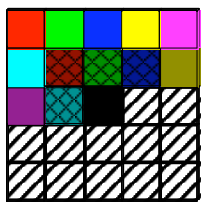
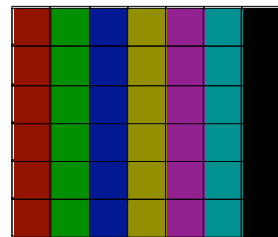
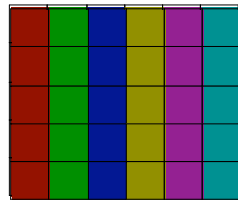
Someone performs another full backup of the database.

# Are there really 20 copies?



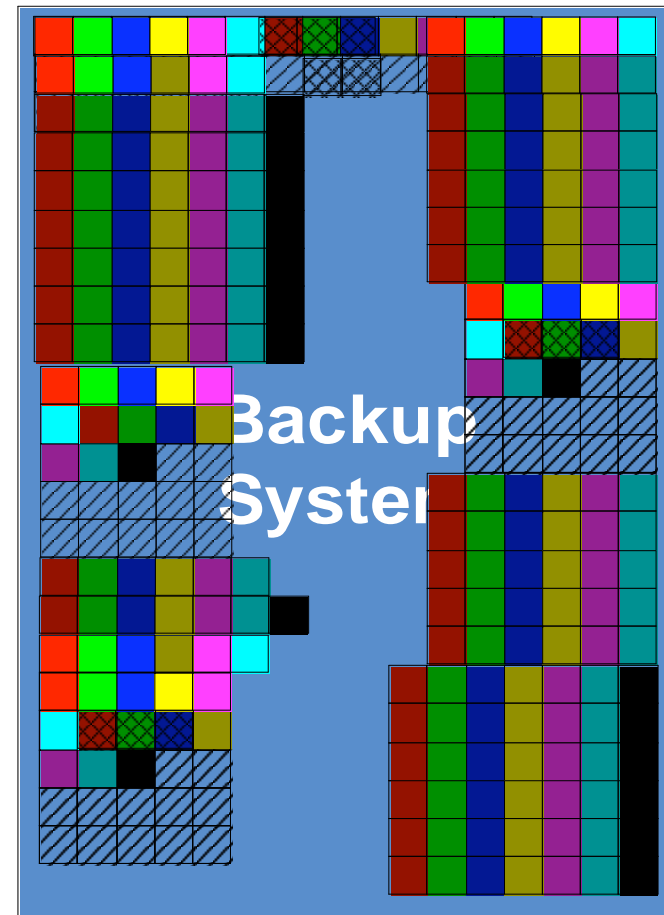
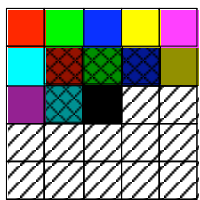
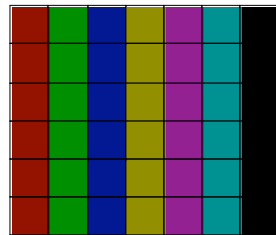
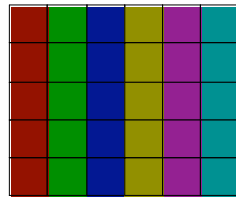
More records are updated in the database and show up on the next incremental backup of the database.

# Are there really 20 copies?



The backup system performs the next week/month's full backup

# Are there really 20 copies?



If all those backups had been stored in a deduplicated backup system, this is how much space they would have taken up.

## Your mileage WILL vary

- You really can get 10x to 400x
- It depends on
  - Frequency of full backups
  - Length of retention
  - Inter-datacenter redundancy
  - Compression or encryption enabled in your environment
  - Multiplexing (if using VTL)
  - Your chosen dedupe technology vs your data

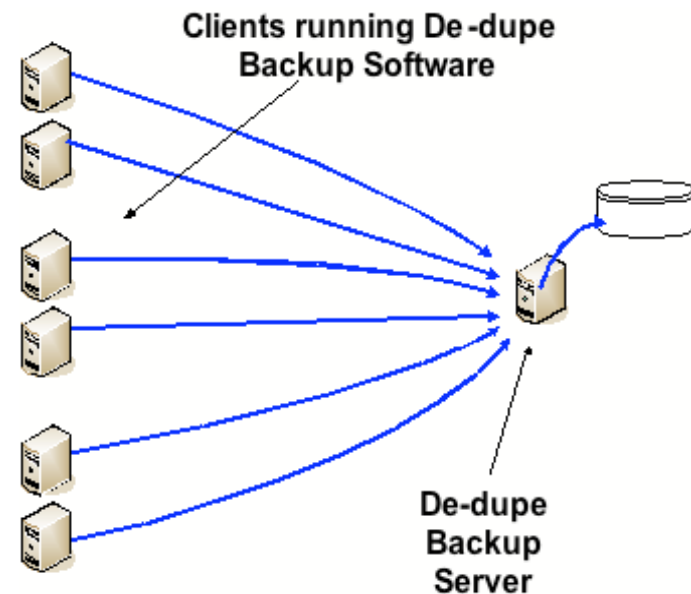
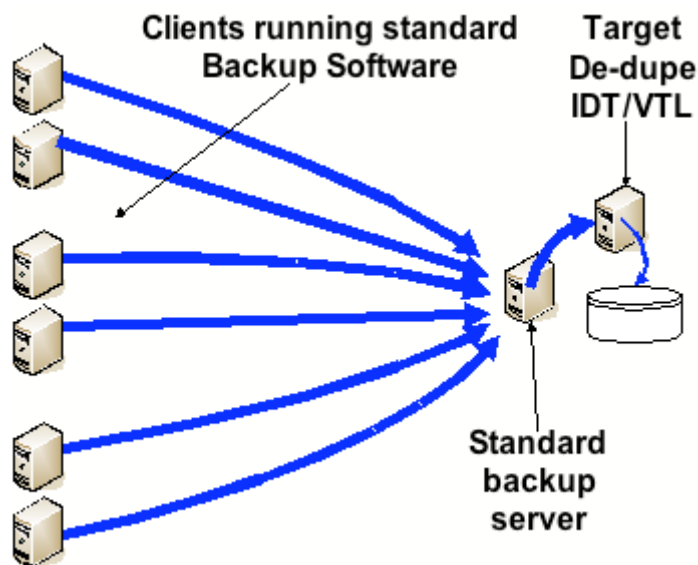
# Implementation Differences

- Source vs Target Dedupe
- Local vs Global deduplication
- Hashing vs Delta Differentials
- Post Process vs Inline



# Source vs Target Dedupe

- Target dedupe
  - VTLs, IDTs that dedupe regular backups
  - Does not reduce IP load unless replicating
  - Best for large scale datacenters
- Source dedupe
  - Dedupe software that dedupes at source
  - Reduces IP load from very beginning
  - Best for remote office branch office (ROBO)

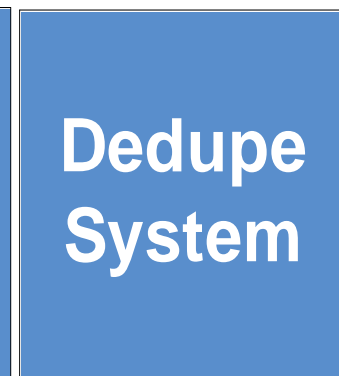
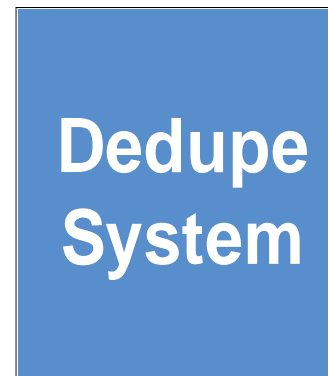


## Target dedupe from backup software?

- Some source dedupe products offer a mode of operation that is actually target dedupe (Symantec)
- Some backup software products are doing only target dedupe (CommVault, TSM)
- Be sure to test these against appliance-based dedupe
  - Probably won't win in a performance drag race
  - May be fast enough
  - May be less expensive

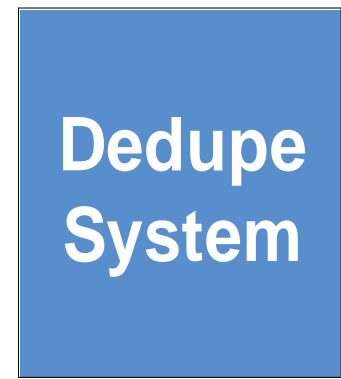
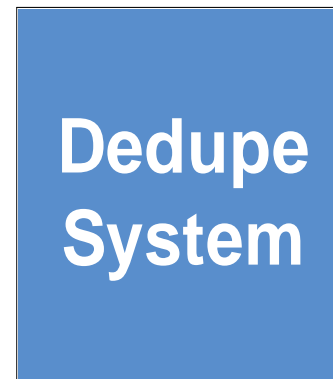
# Local Deduplication

- Each deduplication appliance is its own "dedupe domain"
- Data sent to each appliance is only deduped against other data sent to that appliance



# Global Deduplication

- Backups are deduped against all other backups, regardless of which appliance they came in
- May or may not use common storage pool



# Local or Global Deduplication

- All source dedupe uses global dedupe
- Local
  - More field tested (6+ years)
  - Requires user to allocate backup to certain appliances
  - Higher system and management cost in large environments
- Global
  - Not necessarily about comparing all data to all data
  - It's about comparing common backups, regardless of where they came from or are going to
  - Lower system and management cost
  - Requires no new processes or configurations
  - Not as field tested (Released in 2007)
- If you're backing up <10TB/night it doesn't matter
- If backing up >10TB/night , consider this feature
- Only Falconstor, IBM/Diligent (2 nodes), NEC, SEPATON have global dedupe today

## Hash-based methods

- Run each chunk of data through an algorithm (e.g. SHA-1) to get a large (e.g. 160-bit) value
- If the value matches, the chunk is the same
- Discard the original chunk
- Uses SHA-1, a well-established algorithm
- Challenges are the work required to create hashes and intelligently chunking the data

## Delta Differentials

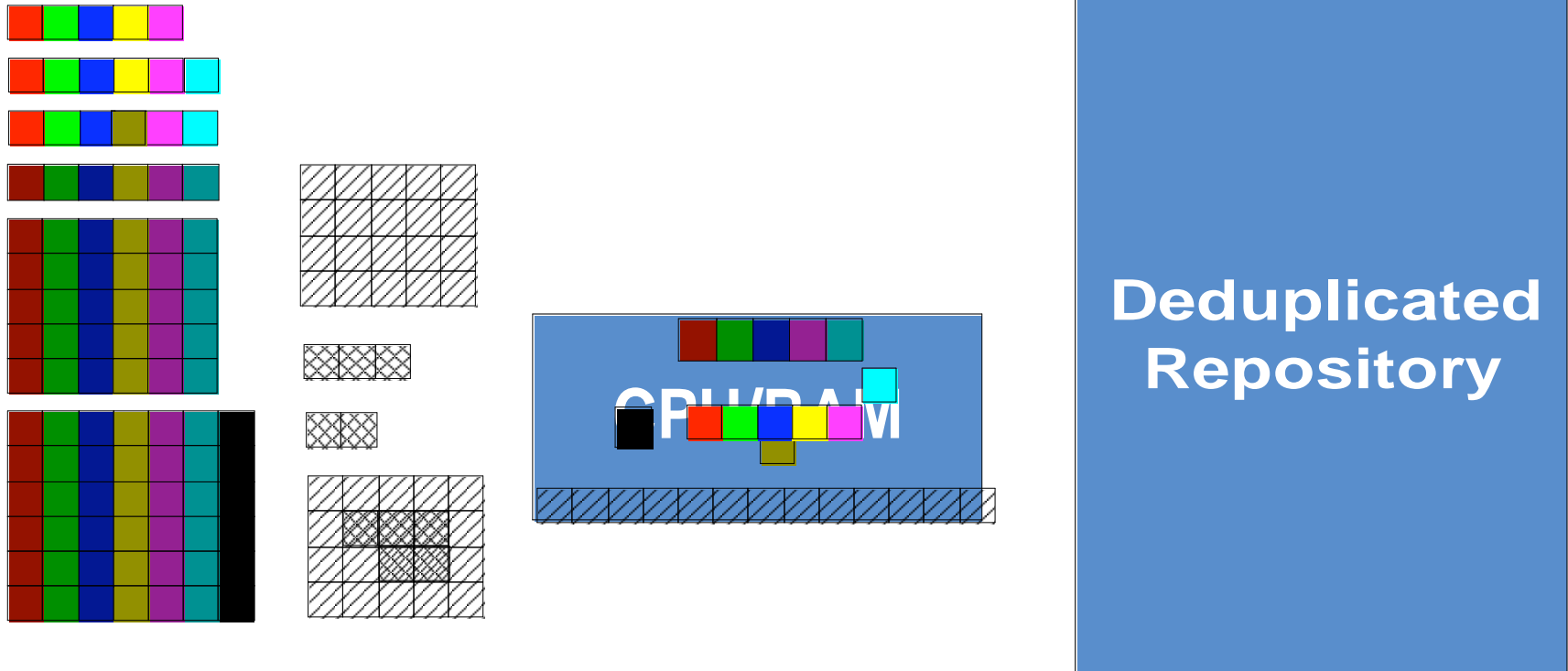
- Correlate backups to each other
  - Full content awareness
  - Partial content awareness
- Match today's backup to yesterday's backup byte-by-byte
  - Store the bytes that are new
  - Discard those that aren't
- Delta differential concept is also well-established
- Challenge is lining up backups in a format they don't control

# Hashing & Others

- Hashing
  - Most used method with most mileage
  - Some concerned about hash collisions
  - Even after applying the birthday paradox, you need 40+ Exabytes to have the chances of a hash collision appear in 50 decimal places
  - I've got more chance of dating Jennifer Garner
- Delta Differentials
  - Faster than hashing
  - No concern about hash collisions
  - Will not compare different backup types to each other, or different hosts to each other
  - You would think this would reduce overall dedupe
  - The greater level of granularity outweighs it
- Hash: Falconstor/Sun, Quantum/EMC, Data Domain, NEC, Symantec, EMC/Avamar
- DD: Diligent, Exagrid, NetApp, SEPATON/HP

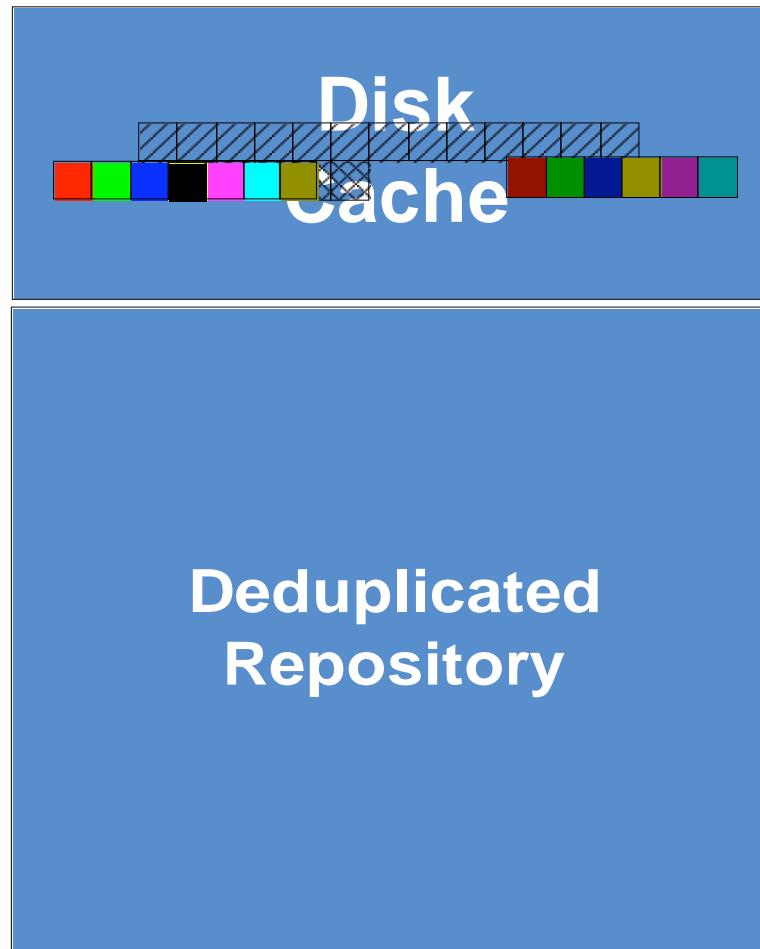
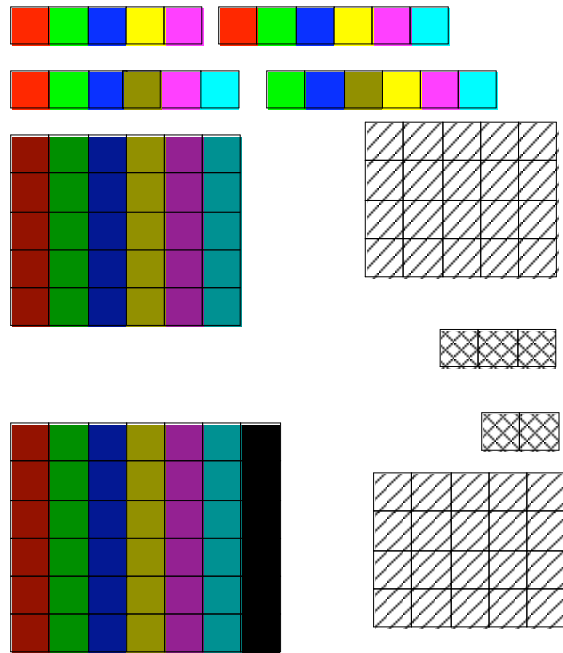


# In-line



Backups are examined for duplicate data on their initial pass through the CPU of the IDT. Unique data is written to disk; duplicate data is discarded (*never* written to disk).

# Post Process



After a given batch of backups are ingested to the disk cache, they are examined for redundant data. Unique data is kept; redundant data is deleted from disk. *(Note: Customer can usually decide to run ingest & dedupe asynchronously.)*

## Post Process vs Inline

- Inline: Simple, but not flexible
  - Replication begins as soon as data is written
  - First vendors with GA product (4-5 yrs+ to date)
  - Does not require disk for staging area
  - May slow down backup, especially very large, very fast single stream backups
- Post Process: Flexible, but more complex
  - First vendors with global dedupe (across appliances)
  - Allows for staggered implementation of dedupe
    - Initial implementation
    - Daily copy to tape before data is deduped
  - Replication begins when dedupe begins
  - Requires disk for staging area (from MBs to TBs, depending on implementation)
- Inline: Data Domain, IBM/Diligent
- PP: Exagrid, Quantum/EMC, SEPATON/HP, NetApp

# Replication

- Does replication use dedupe?
- If deduping many to one, will it dedupe globally across those appliances?
- Can I control what gets replicated and when? (e.g. production vs development)

## Data integrity: The hard questions

- What happens if the index goes bye-bye?
- How do you protect against that?
- Does it need its index to read the data?
- What do you do to verify data integrity?
- What about malicious people?
- Some dedupe vendors aren't very good at answering these questions, partially because they don't get them enough
- This is a big issue. If the structure to your data is gone – your data's gone.

# All that matters

- What are the risks of their approach?
  - Data integrity questions
- How big is it?
  - What's *my* dedupe ratio?
  - How big can it grow (local vs global)
- How fast is it
  - How fast can it backup/restore/copy *my* data?
  - How fast is replication?
- How much does it cost?
  - Pricing schemes are all over the board
  - Try to get them on even playing field
  - Also consider operational costs
    - Adding storage
    - Replacing drives (how long does rebuild take?)
    - Monitoring, etc

## Possible Advanced Uses

- Offsite backups w/o shipping tapes
- Backups with no human hands on them
- No tapes shipped = No need to encrypt tapes
- Tapeless backup
- Shorter Recovery Point Objectives

# Test everything

- Installation and configuration, including adding additional capacity
- Support – call and ask stupid questions
- Dedupe ratio
  - Must use your data
  - Must use your retention settings
- Aggregate performance
  - With all your data types
  - Especially true if using local dedupe
- Single stream performance
  - Backup speed
  - Restore and copy speed (especially if going to tape)
- Replication
  - Performance
  - Lag time (if using post process)
- Dedupe speed (if using post process)
- Loss of physical systems
  - Drive rebuild times
  - Reverse replication to replace array?
  - Unplug things, see how it handles it
  - Be mean!



## Not Ready for Dedupe?

- Consider dedupe vendors that already have a time-tested IDT/VTL, whose dedupe feature can be turned on/off
- Test these products, pick the one with the best dedupe
- Implement disk *staging* with dedupe off
- Turn it on once comfort level increases
- Requires vendor to support optional dedupe

## What about the VTL question?

- If you need  $< 200$  MB/s, it doesn't matter
- If you need more than this, then you may be considering block access
- If you need more than this on a single server, you *really* need block access
- Block access + Dedupe = VTL
- Don't believe the anti-VTL-hype. The ones with good products aren't going anywhere.

## If only there was a company...

- That didn't take ANY money from vendors (unlike all major consulting firms and every analyst firm)
- That could offer *totally* unbiased information
- That had an automated RFI system that allowed you to conduct an RFI in minutes instead of months
- That would share the results of their previous bake-offs
- That would test products in your environment if that's what you want
- Could do that for a *fraction* of what you can do it yourself for, let alone what a consulting company would charge you
- Email [curtis@backupcentral.com](mailto:curtis@backupcentral.com) if you'd like to hear more
- Ask me about it at lunch!

# Protecting Stored Data

Internal Hackers & Tape Thieves

## Routes to your Data

- The Backup Server
- The Tape SAN
- Your Own People
- The Backup Tapes

## Backup Server is all powerful

- Once you can run a script as a privileged user (which any good backup system can do), the sky is the limit to what you can do
- Things I've personally done (with permission)
  - Performed system inventories (df, ps -ef, etc)
  - Changed the root password
  - Deleted the Administrator password
  - Rebooted systems, databases
  - Remotely executed an xterm session
  - Make a portable copy of data for removal

# Responding to the Threats

## The Backup Server

- Backup Server should be the most hardened server in the data center
- Turn off all plain text access to management ports; allow only encrypted access
- Use the separation of duties concept as much as possible. If at all possible, do not give backup admin
  - Root/Administrator access
  - Unrestricted tape access
  - Encryption system access
  - Unrestricted offsite vault access



## The Tape SAN

- Use port-based zoning or port-binding and hard zoning on the tape SAN
- Turn off all plain text access to management ports; allow only encrypted access
- Change the default password!
- Separate management of the SAN as well

## Your Own People

- Do background checks (Someone hired Roger Duronio after he was indicted and told not to work in the IT industry!)
- Show them a little respect
- Give them a career path
- Look for signs of stress and malice
- You know what an at risk person looks like: they should not have access to the backup system!
- Educate them about social engineering and other hacking techniques (Have them watch Sneakers!)
- Do your best to eliminate the ability of a single person to thwart your security system

## Where Should You Encrypt?

- Wherever you can do it cheaply and still meet your requirements
- Most systems had a decent answer to making sure you don't lose your keys
- It's protecting the keys from a trusted employee where most currently don't have an answer
  - If employee sets up encryption system
  - Knows how/where system is backed up
  - Steals backup, tapes, and has drives
  - Voila! They've thwarted the system!
  - Think this through and you can stop it
- Best answer so far is quorums

# VMware & Remote Office Backups

## VMware backups

- VMware, MS Virtual Server, Virtuallron make a lot of things easier
- One thing they don't make easier is backup
- *They don't make backup harder, though*
- Taking advantage of virtualized server functionality is what complicates things
- The question, is do you want to do this?
- If only there was a movie that talked about virtualization...

Let's take the red pill and go  
inside a virtual machine

## It's more fun in The Matrix

- Easiest and most common method
- Treat VMs like PMs
- Install backup software in VM
- Back up just like any other machine
- Allows for easy single-file recovery
- Still required for application/db/email servers
- Can be licensing issues, but not usually
- Does not allow for easy bare metal recovery
- Use vmware to re-create virtual server, then
  - Use your own BMR technique, or
  - Reinstall OS, applications, then restore

## Real World choices

- Standard virtual server backup
- Dedupe or CDP backups
- Use a traditional backup product that supports VMware Consolidated Backup (VCB)
- Use a product specifically designed for VMware and back up to disk
  - ESX Ranger
  - esXpress



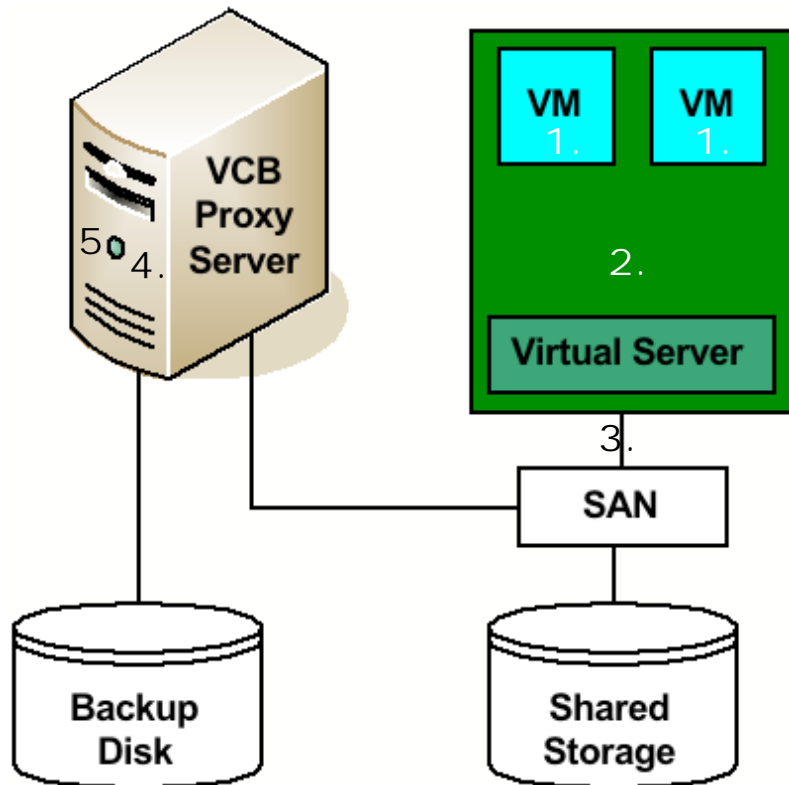
## Standard Virtual Server Backup

- Backup software running on the virtual console
- Back up the virtual disk files at this level
- Advantages
  - Simple
  - Possible license advantages (usually)
- Disadvantages
  - Full backups only
  - No file-level recovery
  - No VSS support
  - Consistency choices
    - Power off VMs during backup (consistent)
    - Power off VMs, create snapshot, power on VMs (consistent)
    - Snapshot VMs (crash consistent)
    - Cannot snapshot raw, RDM physical mode, or independent disks

## Dedupe & CDP backups of VMs

- Dedupe & CDP backups designed for physical machines can be leveraged for VMs
- Incremental-forever nature helps the “I/O problem”
- Some products specifically targeting functionality for this space

# Traditional Backup & VCB



1. VCB creates a VSS Snapshot in Windows
2. VCB creates snapshot in VMware
3. Snapshot is exported to Proxy Server
4. Snapshot is mounted on the Proxy Server as a Windows drive
5. Snapshot is backed up

- VMware Consolidated Backup allows you to mount & view VM filesystems on a proxy server and back them up
- Reduces load on VMs
- Can interact with VSS to create frozen image
- Requires shared storage
- Two methods: volume-level & file-level
- Current volume-level backup requires two-stage copy and restore (changing 2H09)
- Must be a physical proxy to make sense

## ESX Ranger

- Software running on Windows host as proxy
- Managed via GUI on proxy host
- Performs compressed full and differential backups of vmdk files (proprietary zip)
- Also supports VCB
- Uses VSS to lock Windows volumes
- Can also back up Linux volumes, but would require scripting to pause app writes
- Can mount compressed backups as drives for single-file recovery
- Perfect for bare metal recovery
- Info at <http://www.visioncore.com>
- First mover

## esXpress

- Software running on VM inside VMware
- Managed via VMware Infrastructure Console
- Performs compressed full and differential backups of vmdk files (std. gzip/lzop/zip)
- Simultaneously backs up files to zip file
- Supports encrypted backups
- Info at [http://www.esxpress.com/competitive\\_comparison.php](http://www.esxpress.com/competitive_comparison.php)
- Fast follower

## Deduplication & VMware

- All of these filesystem-based backup products that are creating full backups every day are going to create a *lot* of duplicate data
- If you had a NAS-based deduplication backup target, you would probably get 100x or greater dedupe ratio (full after full)
- NAS-based dedupe targets
  - Data Domain
  - Exagrid
  - NEC
  - Quantum

# Remote Office/Branch Office Backup & Recovery

## The ROBO Problem

- Remote offices/branch offices (ROBOs) have data, but usually have little or no IT staff
- Tape drives and VTCs are high touch & high risk
  - 97% of one customer's failures were due to failure of personnel at remote site to simply put in the tape
  - One customer's security guard who was supposed to put tape in case & ship, was actually pushing tape back in and shipping empty cartridge offsite
  - Requires vaulting service at every site
  - Even Iron Mountain uses common carriers to ship tapes between sites
- Disk backups are not easily sent offsite
  - Cannot easily replicate full-file incremental backups



## ROBO Solutions

- Eliminate tape in all ROBOS
  - Removes need for human intervention
- Sub-file incremental with no full backups
  - Reduces backup load to new data blocks
- Cross-site deduplication software/hardware
  - Reduces backup load even further by backup up multi-site data only once
- Solutions for every budget
  - There are pay-as-you-go and open-source solutions

# ROBO Solutions

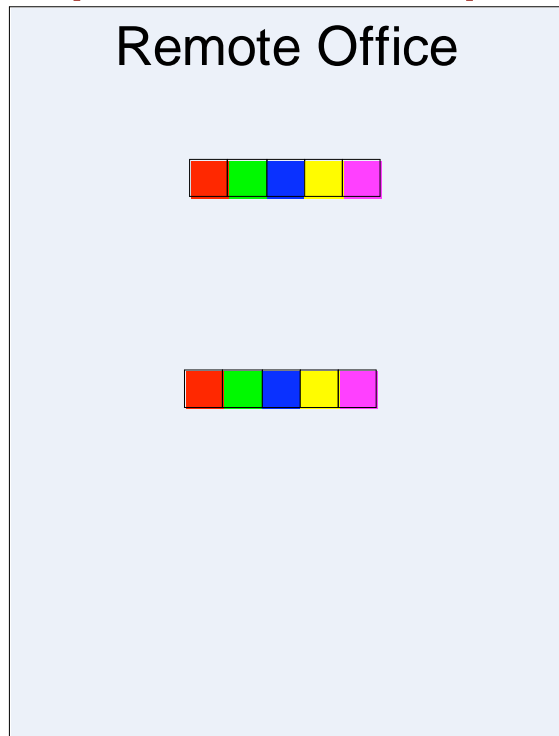
## ● Software Only

- Dedupe software only
- Continuous data protection software
- Near-continuous data protection software

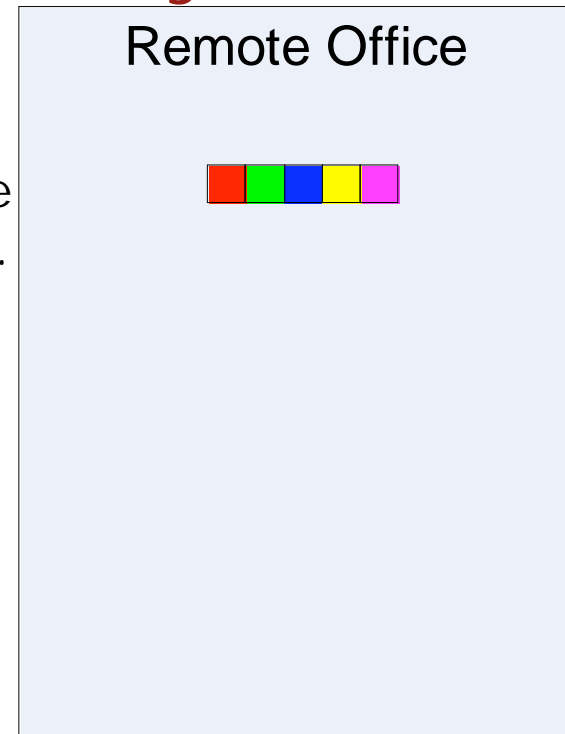
## ● Local Appliance

- Dedupe software w/local appliance
- Std. backup software to local dedupe disk target
- Local storage with near-continuous data protection

# Dedupe Backup Software Only



All restores come from the central server. Ideal for very small data or RTOs of >72 hrs.

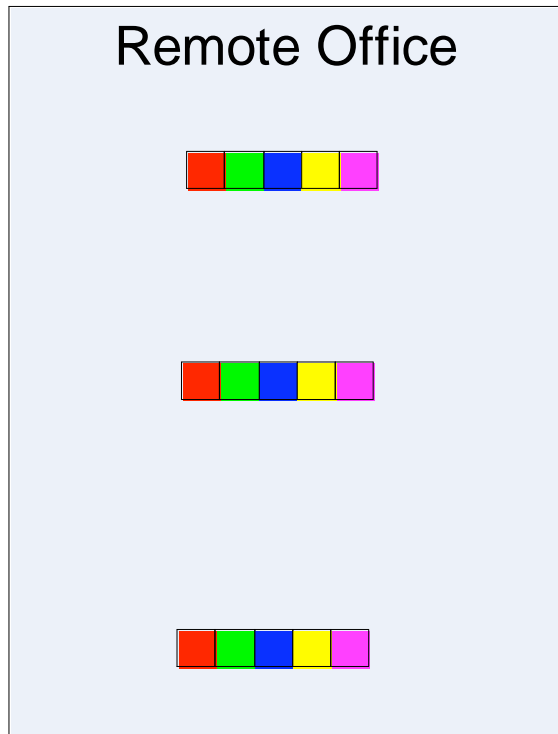


Central Server

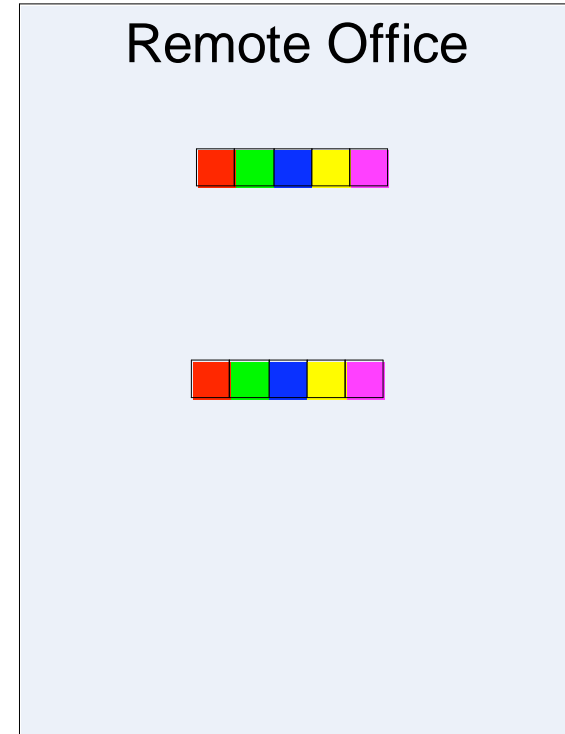
1. Creates hashes of chunks of new/changed files
2. Client checks local cache of hashes to see if it's seen the hash before.

3. If hash is new to client, it asks the server if it has seen it before
4. If so, it stores the hash
5. If not, it sends the new chunk to the server

# Dedupe Software w/Local Appliance



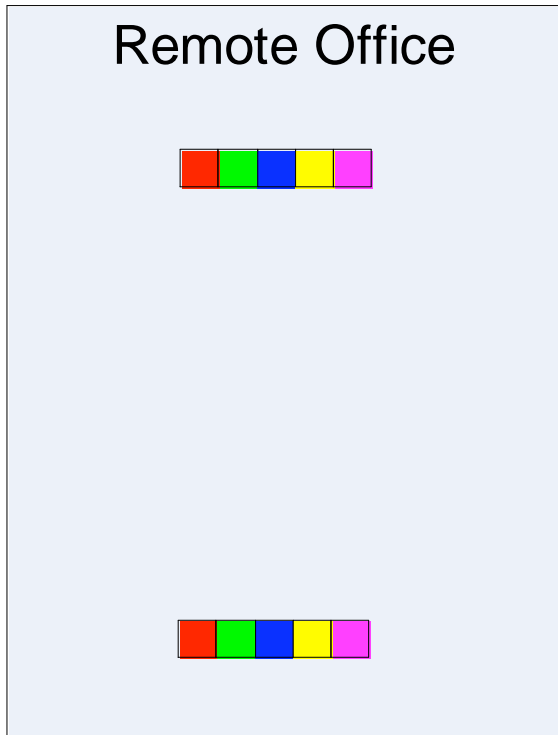
Restores can now come from a local device, while still maintaining a copy offsite for D/R



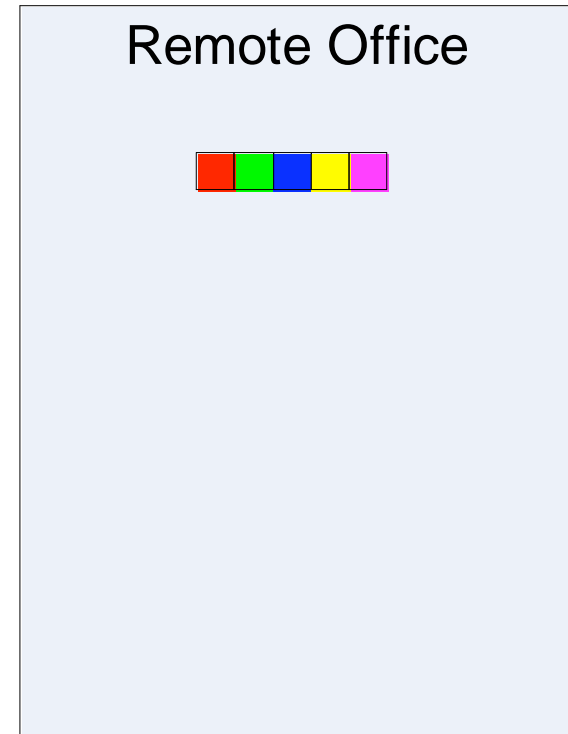
Same as software only, except each site stores a local copy of all chunks unique to that site

Designed for sites with more aggressive RTOs. Can upgrade from software-only to this

# Std. Backup Software w/Dedupe Target



Restores come from local device. Restores from remote device often require advanced configuration.



Sub-file incremental backups block up only new blocks from each file.

If same files/blocks are seen in multiple files or sites, they are sent and stored again.

## Dedupe Options

Both options send the least amount of data offsite of all ROBOS options we're discussing

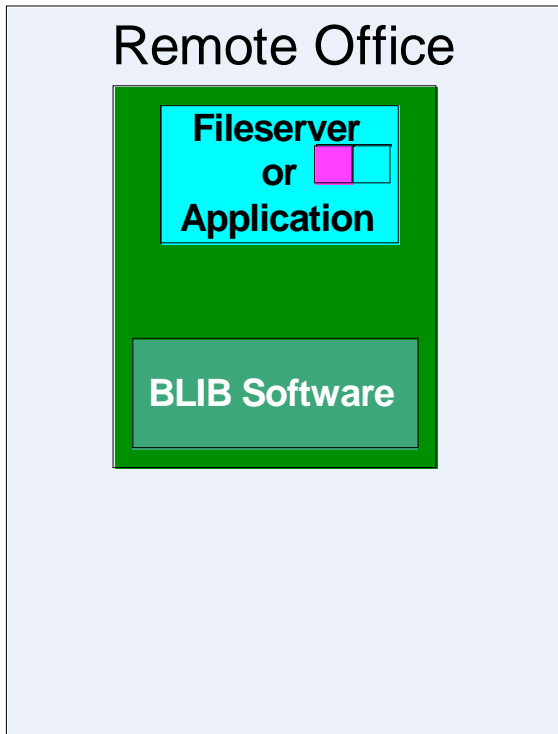
### *Source/Software Dedupe*

- Must learn different backup s/w for ROBOS
- Must pay for entire solution (s/w ,h/w)
- Only requires hardware for large ROBOS
- Current products *very* slow when copying to tape

### *Target/Hardware Dedupe*

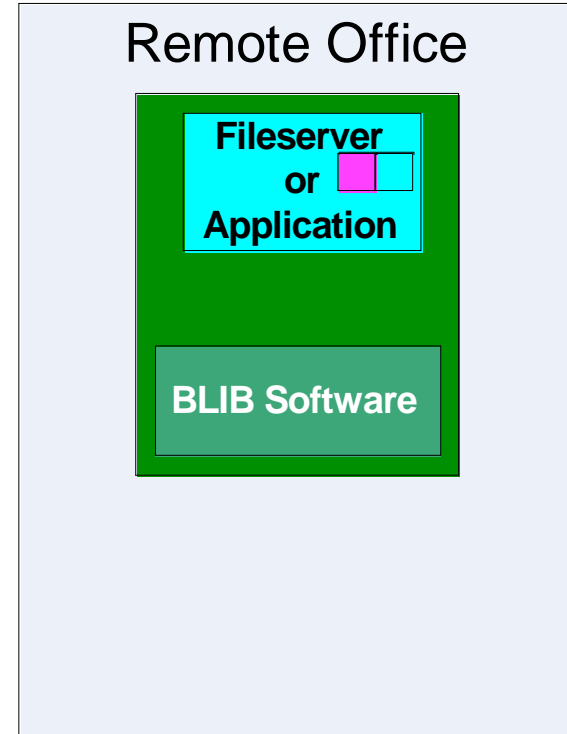
- Can use same backup s/w for datacenter & ROBOS
- Can leverage existing s/w, buy only dedupe h/w
- Requires hardware for all ROBOS
- Fast enough to stream data to tape

# CDP/Near-CDP Software



Continuous/immediate sub-file incremental. Only changed blocks are backed up.

All restores come from the central server. Ideal for very small data or RTOs of >72 hrs.



CDP backs up changed blocks *immediately*. Near-CDP backs up changed blocks on regular basis, typically every hour.



## CDP/Near-CDP Options

- Both options are incremental-forever, so they send much less data than typical backup software, making them ideal for ROBOs
- Both are being offered as option to some large backup software products

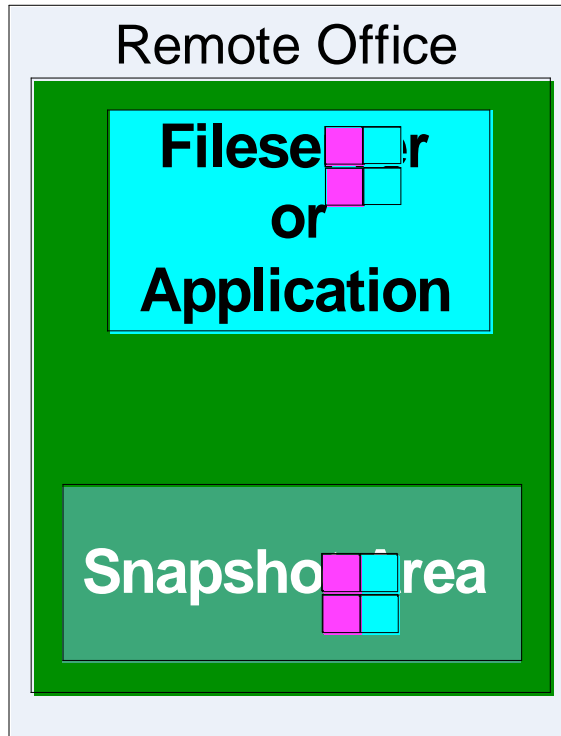
### *CDP*

### *Near-CDP*

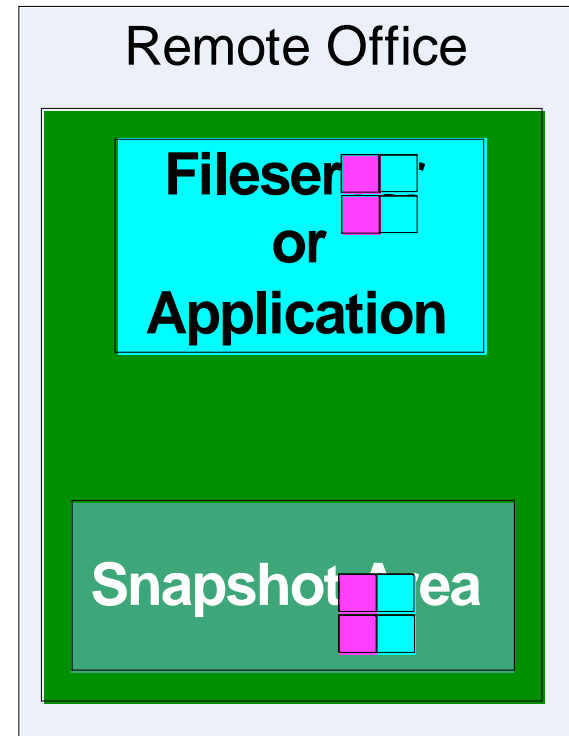
- Can meet an RPO of 0
  - Any application integration will be up to vendor
- RPO = snapshot frequency or replication window
  - Some can use Windows built-in VSS, so can interact with applications



# Storage Device w/Near-CDP



Restores can now come from a local device, while still maintaining a copy offsite for D/R



Some CDP products can store a local copy of each block



Restores can come from local or remote device

## Near-CDP Hardware

- Most tested of all the ROBO options
- Also forever-incremental
- Requires you to store your ROBO data on the Near-CDP hardware appliance, so is also typically the most expensive ROBO option
- Some of these products have integration with other applications, such as Exchange, Oracle, VMware

## PRESENTATION DOWNLOAD: BACKUP SCHOOL 2009

RESOURCES FROM OUR SPONSOR

**data**domain

- « **CASE STUDIES:** The ROI and TCO Benefits of Data Deduplication for Data Protection in the Enterprise
- « **BEST PRACTICES GUIDE:** Microsoft SharePoint Design Best Practices with Data Domain
- « **BEST PRACTICES GUIDE:** Symantec NetBackup (NBU) Design Best Practices with Data Domain

### About Data Domain

Data Domain® is the leading provider of deduplication storage systems. Over 2,500 companies worldwide have purchased Data Domain systems to reduce storage costs and simplify data management. Data Domain delivers the performance, reliability and scalability to address the data protection and nearline storage needs of enterprises of all sizes. Data Domain products integrate into existing customer infrastructures and are compatible with leading enterprise backup and archive software products. To find out more about Data Domain, visit [www.datadomain.com](http://www.datadomain.com).

Data Domain is headquartered at 2421 Mission College Blvd., Santa Clara, CA 95054 and can be contacted by phone at 1-866-933-3873 or by e-mail at [sales@datadomain.com](mailto:sales@datadomain.com).

## PRESENTATION DOWNLOAD: BACKUP SCHOOL 2009

RESOURCES FROM OUR SPONSOR



« **WHITE PAPER:** Expert Advice: Enterprise-Grade Online Backup Solves Data Protection Challenge

« **INTERACTIVE DEMO:** Server Backup and Recovery

« **INTERACTIVE DEMO:** PC and Mac Backup and Recovery

### About Iron Mountain Digital

Iron Mountain Digital is the world's leading provider of Storage-as-a Service solutions for data protection, backup and recovery, archiving, eDiscovery and intellectual property management. The technology arm of Iron Mountain Incorporated offers a comprehensive suite of solutions to thousands of companies around the world, directly and through a worldwide network of channel partners. Iron Mountain Digital is based in Southborough, Massachusetts with European headquarters in Frankfurt, Germany.

For more information, visit [www.ironmountain.com/digital](http://www.ironmountain.com/digital).

## PRESENTATION DOWNLOAD: BACKUP SCHOOL 2009

RESOURCES FROM OUR SPONSOR



« **WHITE PAPER:** 10 Steps to Safeguard Small Business Data

« **BOOK:** Data Backup for Dummies

« **WHITE PAPER:** Reducing Corporate Risk: Best-practices Data Protection Strategy for Remote and Branch Offices (ROBOs)

### About i365, a Seagate Company

i365, A Seagate Company, provides proven solutions for the protection, retention, and discovery of electronic information. Trusted by over 22,000 customers, our innovative technologies combined with flexible deployment options form solutions to solve organizations' most rigorous compliance and data management problems. Our solutions address the complexity of explosive data growth, the challenge of maintaining high availability of critical business systems, and the increasing demands of regulatory compliance and litigation. Our offerings include i365 EVault Data Protection software and SaaS for backup and recovery; i365 Retention Management solutions for data recovery, migration, restoration and data management solutions; i365 MetaLINCS E-Discovery solutions for first pass processing and advanced content analysis of electronic information; and i365 ProServ Professional Services for implementation, consulting, training, and risk assessment of the i365 portfolio.

For more information, please [www.i365.com](http://www.i365.com).

## PRESENTATION DOWNLOAD: BACKUP SCHOOL 2009

RESOURCES FROM OUR SPONSOR



« **WHITE PAPER:** A Pragmatic Approach to Server and Data Center Consolidation

« **WHITE PAPER:** Consolidated Disaster Recovery Using Virtualization -  
Affordable Workload Protection and Recovery

### About Novell

Making IT Work as One?

Through its infrastructure software and global ecosystem of partnerships, Novell harmoniously integrates mixed IT environments, allowing people and technology to work as one. Novell delivers the best-engineered, most interoperable Linux platform and a portfolio of integrated IT management solutions that helps businesses worldwide reduce cost, complexity and risk. With 25 years of experience, 5,100 partners, 54,000 customers and over US\$1 billion in annual sales, Novell is a global leader in delivering solutions that allow people and technology to work as one.

## PRESENTATION DOWNLOAD: BACKUP SCHOOL 2009

RESOURCES FROM OUR SPONSOR



- « **WHITE PAPER:** Microsoft Exchange Best Practices
- « **WHITE PAPER:** Thin Provisioning: The Secret to Achieving Operational Efficiency in the Data Center
- « **WHITE PAPER:** The Modern Virtualized Data Center

### About Pillar Data Systems

Founded in 2001, Pillar Data Systems develops Application-Aware Storage systems for midsize and enterprise organizations. With the highest utilization rates in the storage industry, the Pillar Axiom solution is the most efficient storage system on the market today. The Pillar Axiom cuts administrative time and total cost of ownership by more than 50 percent as well as provides the only storage system that can differentiate services based on application priority. Designed from the ground up as the only true Application-Aware Storage system, the Pillar Axiom allows users to match multiple application characteristics to the appropriate service levels within a single storage platform. Pillar Data Systems is privately funded by Tako Ventures, LLC, the venture arm of Larry Ellison.