Overcoming Backup & Recovery Challenges in Enterprise VMware Environments

Daniel Budiansky
Enterprise Applications Technologist
Data Domain

Dan Lewis
Manager, Network Services
USC Marshall School of Business
Introduction

Market challenge: Server proliferation

- Major administrative overhead
- Significant operational expense, especially for distributed enterprise

Solution: Server consolidation

- Improved efficiency
- VMware virtualized infrastructure now de facto approach to eliminate server growth and control costs
- New challenges for data protection and DR

Managing risk

- VMware and Data Domain provide a joint solution for the data protection and disaster recovery of virtualized server environments
Challenge: Managing and Protecting Server Environments

Storage infrastructure and bandwidth costs

- Backup data > original data
- Replication of active system volumes for high availability is bandwidth intensive and costly
- Replication of backup data is inefficient

Datacenter DR Facility

Replicate 1TB Primary Data
Replicate 10TB Backup Data
Expensive HA Only
Impractical
Challenge: Meeting Recovery Objectives

Tapes on trucks, high-cost DR solutions

Sending tapes offsite introduces risk:

- Most offsite tapes not stored at intended recovery location but local to where they originated
- Over-the-wire data protection solutions can be cost prohibitive or not scale to meet needs of whole datacenter
At the Same Time, Storage is Evolving

Storage 1.0
- PRIMARY
- TAPE

Storage 2.0
- PRIMARY
- SATA & RAID
- TAPE

Storage 3.0
- PRIMARY
- SATA & RAID & DEDUPE
- TAPE
Virtualization enables simplified data protection and DR through the encapsulation of systems as files.

However, there is often a significant storage requirement in order to capitalize on these data protection benefits (i.e. image-level VMDK backups).

Interestingly:

▶ 37% of organizations said the amount of data they need to protect increased after implementing server virtualization*

While Gartner estimates that data deduplication for backup workloads is used in 7% of all backups today, it is a technology that is experiencing extremely fast adoption.*

-- Gartner

By 2012, de-duplication will be applied to 75% of backups.**

-- Gartner

IT managers looking to boost storage efficiency next year will embrace online storage services, push de-duplication in the data center and adopt solid-state disk drives to help fuel hardware consolidation strategies and green initiatives.***

-- IDC

One-third of organizations with over 25TB of data cite that physical footprint and energy efficiency has become a more important purchasing consideration of secondary storage systems****

-- ESG

*Source: Gartner, Emerging Technology Analysis: Primary Data Deduplication, Storage Software and Hardware Technologies, April 1, 2009

**Source: Gartner, Predicts 2008: Emerging Technologies Make Storage Hardware and Software More Effective, December 21, 2007


Solution: Using Deduplication to Protect VMDK Files

VMDK File Structure
- At least 1 per virtual machine
- Large, 2GB+ File
- Unique structure encapsulates entire filesystem

Data Domain Deduplication Benefits
- Dramatic reduction in protected VMDK file size – average **40x-60x reduction**
- Disk-speed protection and reliability
- Data Domain Replicator Software option provides network-efficient disaster recovery over existing networks
Data Domain Basics

2U
- 2 - 6 ports
- 10 and 1 Gb Ethernet; 4 Gb Fibre Channel
- RAID-6
- 5.4 to 35.3 TB usable capacity with shelves
- 1 TB or 500 GB 7.2k rpm SATA HDD in shelf
- File system
- NVRAM
- N+1 Fans, Power Supplies

Gateway to:
- HDS, Fujitsu, EMC, IBM, HP, Sun, 3PAR, Nexsan, Pillar, Xiotech

Easy integration with existing environment
Data Deduplication: Under the Hood

View from Backup Software: CIFS/NFS, VTL, OST

First Full Backup
A B C B D

Incremental
A B E

Second Full Backup
A B C F D E

DD OS Storage:
Redundancies pooled, compressed

= Unique variable-length segments
= Redundant data segments
= Compressed unique segments stored

Generalized approach:
- Application independent
- Protocol independent
- File pathname independent
- Block address independent
Global Deduplicated Replication for Disaster Recovery

True DR; lowers WAN costs; improves SLAs

95-99% Cross-site Bandwidth Reduction

Source: Remote Sites

Destination: Datacenter Hub
Supports hundreds of remote sites

Global Dedupe Effect: new deduped data sent once, even if at 90 sites
Maximize Disaster Recovery Benefits

**Problems with Tape for VMware DR**
- Time required to write and read large VMDK files to tape
- Burden and expense of managing tape rotation
- Delay in locating tapes
- Risk of lost tapes, failed tapes and lost data
- Environmental impact of transporting tape media between sites

**Eliminate Tape Dependency for Disaster Recovery**
- 99% reduction in bandwidth required to replicate VMDK files
- Rapid rebuild of virtual machines
  - Recover the entire virtual machine – system state, configuration, file system and application data
  - Enable system rebuild quickly and easily into dissimilar server equipment
  - Mix of physical and virtual systems all recoverable from Data Domain deduplication storage

---

![Diagram showing data and file management processes](image-url)
The VMware and Data Domain Joint Solution

Data Domain Deduplication Storage

- Massive data retention
- Works especially well in VMware environments
- VMs protected quickly with efficient replication
- Data Invulnerability Architecture
- Seamless integration with leading enterprise data protection applications
USC Marshall School of Business

- Los Angeles campus
- IT develops & delivers application-based services

Data Protection Infrastructure Overview

- Pre-VMware/Data Domain Environment
  - Traditional tape-based data protection infrastructure
  - Weekly backups of 15TB
    - 5.5TB full backups
  - Agent-based backups from individual machines to tape
  - Tapes transported offsite by Iron Mountain
VMware Implementation

Key driver for virtualization

- Improved foundation for DR
  - Virtualized 100 servers

Data protection post-virtualization

- Initially still used traditional tape for backup
  - Agent-based backups from individual machines to tape
  - Backup/restore performance issues
- Then moved to disk-based backups
  - Still direct attached

“Like many organizations, we migrated to a virtual environment first, then encountered new storage challenges”
Challenges With Conventional Disk

“We were buying too much disk…”

- Needed to meet retention policies
- Still locally attached and going from disk to tape, then tapes offsite:
  - Logistical issues persisted
  - Unmanageable storage
  - Still sinking lots of man hours into tape management
  - Still had concerns over tape loss

Drivers for deduplication with replication

- SAN replication too expensive
- Ease of administration
- Safe replication without excessive bandwidth cost
Data Domain Solution: Phase 1 Implementation

Implemented 2 Data Domain systems & replication

- Datacenter & Arizona co-lo facility
- Vizioncore vRanger Pro for VMs

- First conducted pilot for local backups
  - Non-disruptive, 20 minutes to install pilot and initiate local backups

- Implemented replication
  - Inline deduplication (vs. post-process)
    - Raw data never hits the disk
    - Replication initiated at backup inception (faster DR)
Implemented 2 additional Data Domain systems & replication, scaled Arizona co-lo (DR) facility

- College of Letters, Arts & Sciences
- Office of the Provost
Some Lessons Learned

Agent-based backups maybe feasible without dedupe -- but not system restores
  ▶ Much more feasible with Data Domain & vRanger

Throttling capabilities of deduplication systems are key
  ▶ Able to tune bandwidth to existing cost metrics, still replicate very quickly

Inline deduplication enables rapid time-to-DR synch
  ▶ Replication begins at inception of backup process
  ▶ Quick recovery of VMs from replica
Benefit Summary

Resulting benefits

- No change to infrastructure required
- Fast DR synch: Essentially replicating real time
- Reduced footprint, cooling, management
  - Got 1½ racks worth of equipment into single 3U space
- Eliminated tape transport services
- Automation led to reduced man hours
  - 2 full days per week of administration time recaptured
- Retention went up to 6 weeks
Common Themes Across Industries

Solutions must be:
- Simple
- Reliable
- Cost-effective

Problems do not rest in one vertical
- Explosive growth of data
- Server sprawl
- Aging and archaic infrastructure
VMware virtualized infrastructure has become the *de facto* approach used to eliminate explosive server growth, control management costs, and reduce downtime.

**Data Domain backup to disk storage provides:**
- Compelling solution for today’s data protection and DR challenges for virtualized server environment
- It enables: massive data retention, efficient replication, data invulnerability, and seamless integration

**Contact your local sales team for more information:**
sales@datadomain.com