Redefining Backup for Your Virtual Environment with Data Duplication

2009 EMC Forum
Today's IT Reality

- A global economic recession is forecasted for 2009
- The economic environment is leading indicator of technology spending
  - 71 percent of CIOs anticipate flat or decreasing IT spending budgets
  - 2009 IT budgets in developed countries set to decline by 12 percent; globally by 9 percent

"Our IT spending indices dipped further to new lows in our latest survey, deep in contraction territory."

— Goldman Sachs

GDP has tended to be a coincident indicator of U.S. tech spending, with tech investment slowing more deeply vs. GDP in the last 2 recessions

U.S. GDP vs. tech capital investment (nominal), 1970-2010E

---

Source: Goldman Sachs IT Spending Survey, 3/9/2009
Source: IMF World Economic Outlook Report, 11/6/2008

EMC CONFIDENTIAL—INTERNAL USE ONLY
Tighter budgets are likely to accelerate the adoption of server virtualization.

“Total cost of ownership (TCO) reductions will be a key driver of the acceleration in server virtualization deployment as CIOs are forced to cut capital spending and reign in management, administrative and power/cooling costs.”

— Goldman Sachs

Source: Goldman Sachs IT Spending Survey, 3/9/2009
Why Virtualization in Today’s Environment

Top 3 reasons for virtualization

1. Reduce infrastructure cost
   - Achieve significantly higher resource utilization by pooling common infrastructure
   - Reduce number of servers and related IT hardware and the associated real estate, power and cooling requirements

2. Reduce datacenter operating cost
   - Improve operation efficiency, increase flexibility and responsiveness
   - Reduce time spend on provisioning, configuration, monitoring and maintenance

3. Minimize downtime and risk reduction
   - Eliminate planned downtime and recovery quickly from unplanned outages
   - Increase application availability and improve business continuity

“The current environment has moved virtualization toward the top of the priority list for CIOs.”
— Merrill Lynch

Source: Merrill Lynch CIO Survey, 10/28/08
Virtualization changes the IT paradigm…backup must evolve to deliver even greater consolidation and value.

**Old Paradigm**

**Physical Environment**: Low overall server utilization and plenty of bandwidth for backup

- **Server A**
  - CPU Utilization: 20%
- **Server B**
  - CPU Utilization: 20%
- **Server C**
  - CPU Utilization: 20%

20 percent resource utilization

**New Paradigm**

**Virtual Environment**: High overall server utilization and little bandwidth for backup

- **Virtual Server A**
  - CPU Utilization: 80%
- **Virtual Server B**
  - CPU Utilization: 80%
- **Virtual Server C**
  - CPU Utilization: 80%

80 percent resource utilization
Why Source-based Data Deduplication for VMware Backup

Traditional Backup Solutions

- VM 4
- VM 3
- VM 2
- VM 1

Tape

Moves up to **200%** of primary data per week

- Longer overlapping and simultaneous backup windows
- Redundant backup across virtual machines (OS, systems files, etc.)
- Performance impact on shared resources (memory, NIC, CPU)

Source Deduplication Backup Solution

- VM 4
- VM 3
- VM 2
- VM 1

Server

Moves as little as **2%** of primary data per week

- Deduplicates before backing up data
- Reduces data moved by up to 95%
- Significantly reduces overhead on shared resources
EMC’s Definition of Deduplication

“The process of detecting and identifying the unique data segments within a given set of information, enabling the elimination of redundancy when stored or moved.”

Before:
- Data Set 1
- Data Set 2
- Data Set 3
  
  total segments = 39

After:
- Unique segments = 6
How Data Deduplication Works

1. **First instance**
   - March 2009
   - Table of Contents
   - Introduction
   - Report Organization
   - Executive Summary
   - Findings
   - Highlights
   - Assessment Findings
     - Server and Data Amounts
     - Amount of Data Growth
     - Time of Data on Server

2. **Duplicate instance**
   - March 2009
   - Table of Contents
   - Introduction
   - Report Organization
   - Executive Summary
   - Findings
   - Highlights
   - Assessment Findings
     - Server and Data Amounts
     - Amount of Data Growth
     - Time of Data on Server

3. **Modified instance**
   - April 2009
   - Table of Contents
   - Introduction
   - Report Organization
   - Executive Summary
   - Findings
   - Highlights
   - Assessment Findings
     - Server and Data Amounts
     - Amount of Data Growth
     - Time of Data on Server

Only unique data segments are backed up

Data already backed up, so only a unique ID pointer is stored (20 bytes)

New data segment identified and backed up

Unique data stored on disk, available for immediate recovery
Introduction to EMC Avamar

Revolutionizes backup with global, source data deduplication

- Reduces the size of backup data at the source, enabling fast, daily full backups across existing physical and virtual infrastructure
  - Reduces daily backup times by up to 10x
  - Reduces daily network bandwidth impact by up to 500x
- Deduplicates across sites and servers for maximum efficiency
  - Reduces total disk backup storage by up to 50x
- Simplifies consistent, multi-site control via centralized web-based management and at-a-glance dashboards
- Cost-effectively store full backups on disk for extended period of time
  - Encrypted replication across sites; reducing/eliminating reliance on tape
Avamar deduplicates at the optimal location and level of granularity

- **Optimal location:** Reduces data at the source
  - Duplicate data never traverses congested shared resources
  - Data backed up is reduced from ~200 percent to ~2 percent of primary weekly
  - Significantly reduces contention for shared resources
  - Deduplicates within and across virtual machines and/or VMDK (virtual machine disk format) files

- **Optimal granularity:** Sub-file, variable-length segments
  - VMDK is one large file; any changes cause an incremental/full traditional backup
  - Fixed-length segment deduplication fails due to frame offset
  - Sub-file, variable-length segment deduplication finds changes anywhere in VMDK
  - Dramatically cuts the amount of data backed up daily
  - Significantly reduces daily backup time

- **Ideal for the protection of virtualized environments**
  - Enables fast, secure backups over existing virtual infrastructure
  - Permits greater server consolidation and maximum value from virtualized infrastructure
Avamar Benefits for VMware Backup

- Up to 95 percent reduction in data moved
- Up to 90 percent reduction in backup times
- Up to 50 percent reduction in disk impact
- Up to 95 percent reduction in NIC usage
- Up to 80 percent reduction in CPU usage
- Up to 50 percent reduction in memory usage
- All backups are stored as “virtual full backups,” ready for immediate restore
- Maintains effective consolidation ratios without overtaxing CPU utilization
Avamar Client Backup Solutions

VMware Guest OS Backup

Avamar client software runs directly on each virtual machine

vStorage APIs for Data Protection

Avamar client software runs on the proxy server
Avamar Backup for VMware Guest

- Avamar agent resides inside each virtual machine
- Deduplicates data within the virtual machine, as if they were physical servers
- Moves minimal backup data
  - Reduces resource contention and accelerates backups
- Provides file-level restore for Windows, Linux, and Solaris
Backup for VMware Guest: Avamar vs. Traditional Backup

**CPU Usage**

**Network Usage**

**Disk Usage**

---

**Disk Usage**

- Traditional
- Avamar
Avamar and vStorage API Backup

- Avamar agent resides on the proxy server
- Deduplicates within and across VMDK files
- Supports both file (Windows only) and image-level backup
- Avamar replication provides disaster recovery for backed up VMDKs
Avamar Server Backup Solutions

Avamar Software
Deployed on qualified, industry-standard servers

Avamar Data Store
Fully integrated hardware/software product with single point of support

Avamar Virtual Edition for VMware
Single-node Avamar server running as a virtual machine
**Fully integrated software/hardware product**

- Complete EMC backup and recovery product
  - Avamar backup and recovery software with integrated source/global data deduplication
  - EMC-certified hardware—fully configured and delivered

- Built-in high availability
  - Avamar RAIN technology
  - Spare node, RAID
  - Redundant power distribution

- Simplifies
  - Purchase (single vendor, certified hardware)
  - Deployment (minimizes onsite setup)
  - Service (single vendor support)

**Avamar Data Store Gen2 doubles deduplicated backup capacity**
Avamar Virtual Edition for VMware

Avamar server software deployed as a virtual appliance

- Industry’s first deduplication virtual appliance for backup, recovery, and disaster recovery
- Leverages existing servers and storage
  - Can utilize existing iSCSI, SAN or DAS disk storage
- Replication (of applications and storage) eliminates shipping tapes
  - Replicate between virtual Avamar servers and physical Avamar servers
- Facilitates rapid, cost-effective deployment and return on investment
- Supports vMotion for deployment flexibility
- Up to two Avamar Virtual Edition virtual appliances per ESX server for scalability
Extend VMware to Remote Facility
Data Protection

Exchange

File and Print

DNS

Avamar Virtual Edition

Traditional Data Protection

APP OS

APP OS

APP OS

Avamar VM OS

ESX Server

Hardware

Resource Pool

Backup Server

Tape Library

Tape Media

Offsite Tape Storage
Replication for Remote Facilities
Disaster Recovery

Avamar Virtual Edition

- APP OS
- Avamar VM OS

ESX Server
Hardware
Resource Pool

Distributed Remote Facility with VMware

Replication for Disaster Recovery

Avamar Virtual Edition

- APP OS
- Avamar VM OS

ESX Server
Hardware
Resource Pool

Avamar Data Store

= Avamar Software Agent
Avamar Solutions for VMware

Remote Offices with VMware

Avamar agent moves data to Avamar Virtual Edition for VMware; data is then replicated to the corporate data center.

Remote Offices without VMware

Avamar agent moves data to a physical Avamar Single Node server; data is then replicated to the corporate data center, OR an Avamar agent can back up directly to the data center over the WAN.

VMware Data Center with Guest-Level Backup

VMware Data Center with vStorage API Backup
Avamar Success Story: Corporate Express
Time shortened, costs reduced for backup in VMware deployment

Before Avamar

- Storage demands were rapidly increasing
- Tape library was reaching slot capacity and upgrading was not ideal due to age and maintenance costs
- Needed to control costs and simplify data management
- Backup and disaster recovery was time consuming

With Avamar

- Reduced stored data by more than 50%, from 92 TB to 44 TB
- Achieved significant financial savings
- Enabled disk-based backups to be completed in 30 minutes, compare to 6 hours in the past for tape
- Reduced restoration times for business-critical data from 24 hours to minutes

“We were blown away by the simplicity of the management interface and the comprehensive capabilities offered by Avamar. After carrying out a proof of concept, we clearly understood the benefits Avamar would bring to our business.”

— Mark Jones
Technology Infrastructure Mgr.
Introduction to EMC Unified Management for Backup

**EMC NetWorker and Avamar Integration**

- Central management of traditional and deduplication backup in one solution
  - Single client
  - Common administration—policies, scheduling, monitoring
  - Full indexing and browse of backup data
  - Common recovery interfaces
- Supports file systems and applications
  - Microsoft Exchange, SQL, SharePoint, Hyper-V
  - Oracle
  - Physical and virtual (VMware, Hyper-V)
- Supports both traditional and deduplicated (source and target) backup
Why EMC NetWorker—
Complete Backup and Recovery from EMC

- Centralized control of traditional and next-generation backup
  - Combining today’s technologies with tomorrow’s in a common framework

- Industry-leading global data deduplication
  - Reduces backup storage by up to 50 times and data moved by up to 500 times—ideal for VMware, Hyper-V, and SharePoint environments

- Enterprise performance and security
  - Nothing securely backs up and recovers like NetWorker

- Broad backup to disk
  - EMC Disk Library integration, replication, deduplication, continuous data protection (CDP), and NAS backup to disk

- Better recoverability
  - Disaster and granular recovery to meet the right service level agreements with the right recovery
  - Single-step recovery from snaps, disk, tape
  - Future-proof open tape format with better recoverability from damaged tape media
NetWorker and Avamar Integration

How it works
- Integrated client is a standard feature of NetWorker
- NetWorker represents the Avamar Data Store as a deduplication node
NetWorker and Avamar Integration

How it works

- Integrated client is a standard feature of NetWorker
- NetWorker represents the Avamar Data Store as a deduplication node
- Deduplication is enabled via NetWorker Management Console
- Save process gathers metadata (dehydrated save set) and sends it to the NetWorker server and data to the deduplication node
- On recovery, dehydrated save set is restored, required deduplication data is requested, and full dataset is rehydrated at client
NetWorker and VMware Protection

- Solutions for every stage of VMware adoption
  - Granular and application-consistent virtual machine backup at the guest level
  - Image-level backup at the ESX Console operating system level
  - Snapshot-based backup with VMware Consolidated Backup (VCB)

- Choose traditional or deduplication backup
  - Traditional operations deliver VMware protection with disk and tape
  - Source-based deduplication reduces impact of backup—ideal for virtual machine backup

- Support for VMotion, Distributed Resource Scheduler (DRS)
  - Ensure protection even if virtual machines move
EMC Data Protection Advisor (DPA)

Industry-leading data protection management solution

- Collects data from across the infrastructure
- Provides a central console for monitoring, alerting, analysis, trending and reporting
- Transforms disparate data into actionable business information
- Reporting that is easily customizable

Benefits

Lowers costs—improves efficiency, automates tasks, avoids purchases

Improves compliance—lowers risks, ensures critical data protected

Manages change—track and enforce configurations
DPA Virtualization Option

**Improve visibility and management for virtualized hosts**

**Why support VMware**
- Data protection includes system performance and availability
- Virtualization adds management complexity
- Shared resources mean that co-located systems can have a collective impact

**What it does**
- Simplifies visibility and management
  - Monitoring/alerting, troubleshooting, optimization, capacity planning and reporting

**What it collects**
- Collecting data from Virtual Center or ESX servers, including
  - System configuration, hardware and software
  - System resource consumption and growth
  - Server location of host images
  - Config, config changes
  - Network interfaces

**Note:** data collected from VMware provides a more complete picture when the Backup Option is also used
EMC Solutions for VMware Backup

Avamar

- Complete backup, recovery, and dedupe solution
- Source/global deduplication at Guest, VCB, or Service Console
- Fast, daily full backups
- Single-step recovery
- Integrated high availability (RAIN)
- Avamar Virtual Edition for VMware

NetWorker

- Industry-leading backup and recovery software
- Integration with Avamar—and other EMC backup and data protection technologies

Data Protection Advisor

- Monitor data protection infrastructure
- Troubleshoot specific failures and issues
- Trend and analyze capacity, SLAs, more
- Plan by analyzing environment-wide usage
- Predefine, customizable reporting
## EMC Global Services End-to-End Virtualization Capabilities

### Consulting
- EMC VMware Authorized Consultant (VAC)
  - Virtualization Strategy Development
  - Operational Readiness
  - VMware Plan and Design
  - Virtual Desktop Plan and Design
  - Business Continuity for VMware
  - Virtualization of Enterprise Applications

### Solutions
- EMC Backup and Recovery for VMware
- VDI Solution
- Solutions for Microsoft Exchange, Share Point
- Solutions for Oracle

### Technology Deployment
- EMC Capacity Assessment for VMware
- EMC Backup Assessment for VMware

### Managed Services
- Residency Services
- Solutions for Microsoft Exchange, Share Point
- EMC Design and Implementation for VMware SRM

### Education
- EMC VMware Authorized Training Consultant (VATC)
- Solutions for Oracle
EMC Services – accelerating the benefits of virtualization

- **EMC and VMware partnership**
  - Solidifies EMC’s position as a key VMware strategic partner with common goal of providing joint customers with proven solutions
  - Single-point of contact backed by highly skilled service professionals and engineers
  - Enables customers get to the next stage within their VMware lifecycle

- **Certified and field-tested virtualization expertise**
  - Experienced in demanding virtual infrastructure implementations
  - Adept at integrating key IT best practices into your virtual infrastructure
  - Skilled in delivery of end-to-end virtualization solutions spanning storage, networking, business continuity, backup and archiving

- **Extensive infrastructure consulting experience**
  - Hundreds of large-scale global IT initiatives
  - Holistic, information-centric approach
  - Comprehensive services across infrastructure technologies and enterprise IT operations
Why EMC for VMware Data Protection

- Proven, leading-edge technology for VMware infrastructure
  - Industry Leading Source-based data de-duplication provides answer for VMware backup

- Choices—next generation backup, traditional or both combined

- Experts in backup
  - Proven in thousands of customer deployments

- Full suite of backup services, solutions, and software
  - Comprehensive technology and delivery partnerships

- Flexible business partner
  - Simplifies management overhead by being prime contractor
  - Entire lifecycle deployment, or point engagements