Evaluating Disk Backup with Deduplication: 
*Case Studies and Industry Information from the Frontlines*

Michael Krieger, Moderator
Robert Stevenson, TheInfoPro
Marc Crespi, ExaGrid Systems

July 30, 2009
2:00 p.m. Eastern / 11:00 a.m. Pacific
45 minutes
Today’s Agenda

- Key IT insights from TheInfoPro's latest storage study
- Impact of deduplication now and in the future
- Important considerations when evaluating a solution for disk backup with data deduplication
- What role deduplication plays in achieving cost-effective backups that are fast and reliable
- Real-world customer examples of using disk backup with deduplication to greatly improve backup operations.
- Questions
Robert Stevenson
Managing Director of Storage Research
TheInfoPro
Deduplication – Adoption Moving Forward at a Rapid Pace

TheInfoPro™ Storage Study- Wave 12
Top Storage Professionals’ Pain Points

Managing Storage Growth
Proper Capacity Forecasting
Managing Costs
Backup Administration and Management
Dealing with Performance Problems
Managing Complexity
Lack of Integrated Tools
Storage Provisioning
Data Mobility
Managing Storage Equipment
Archiving and Archive Management
Application Recoveries / Backup Retention
Vendor Management
Regulatory Compliance
Power Management

(6/10/09): F1000 Sample. n=155. * Note that due to multiple responses per interview, total may exceed 100%. Others n=3 with no answers
**Change in Storage Budgets**

*How did your 2008 Storage budget spending change compared to 2007? Compared to 2008, how do you expect your Storage budget spending to change in 2009?*

**MSE 2009 vs. 2008**

- **Increase 51 – 75%**: 0%
- **Increase 26 – 50%**: 10%
- **Increase 11 – 25%**: 20%
- **Increase 1 – 10%**: 23%
- **Decrease 1 – 11%**: 0%
- **Decrease 11 – 25%**: 10%
- **Decrease > 25%**: 57%

**F1000 2009 vs. 2008**

- **Increase 51 – 75%**: 33%
- **Increase 26 – 50%**: 49%
- **Increase 11 – 25%**: 18%
- **Increase 1 – 10%**: 18%
- **Decrease 1 – 11%**: 18%
- **Decrease 11 – 25%**: 39%
- **Decrease > 25%**: 49%

Top Storage Team Projects for 2009

- Consolidation
- Backup Redesign
- Tiered Storage Build Out
- Technology Refresh
- Virtualization Adoption
- Archiving
- Disaster Recovery
- Data Migration
- Improving Performance
- Improving Forecasting
- Expanding Replication
- New Data Center
- Securing Storage
- Thin Provisioning
- Green Storage
- Cloud Sourced Storage
- New Application Deployment
- Merger and Acquisition

(6/3/09): F1000 Sample. n=155. * Note that due to multiple responses per interview, total may exceed 100%
Present Tier 1, Tier 2, and Archive Tier Capacity

What percentage of your Storage capacity is on the following tiers?

Percentage of Storage on a Tier

Anticipated Growth of Tiers

What applications are targeted for your archive tier?

- Email
- Unstructured File Content
- Imaging and Multimedia Applications
- Non-Critical Business Applications
- Backup Environments
- Business Critical CRM, ERP and Databases/Warehouses
- Databases
- Compliance Content
- Document Management
- Long Term Application Compliance Retention

(6/5/09): Full Sample. n=74. * Note that due to multiple responses per interview, total may exceed 100%
Select Storage Backup Technology Trends at F1000

**Backup Data Reduction / Deduplication**
- In Use Now (NOT including pilots): 27%
- In Pilot / Evaluation: 8%
- In Near-term Plan (through Q2 '09): 15%
- In Long-term Plan (Q3 '09 – Q1 '10): 25%
- Not in Plan: 26%

**Online Data Reduction / Deduplication**
- In Use Now (NOT including pilots): 15%
- In Pilot / Evaluation: 6%
- In Near-term Plan (through Q2 '09): 9%
- In Long-term Plan (Q3 '09 – Q1 '10): 16%
- Not in Plan: 54%

**Disk-to-Disk (for backup to disk targets)**
- In Use Now (NOT including pilots): 74%
- In Pilot / Evaluation: 3%
- In Near-term Plan (through Q2 '09): 5%
- In Long-term Plan (Q3 '09 – Q1 '10): 8%
- Not in Plan: 10%

**Backup Virtualization Management**
- In Use Now (NOT including pilots): 30%
- In Pilot / Evaluation: 5%
- In Near-term Plan (through Q2 '09): 7%
- In Long-term Plan (Q3 '09 – Q1 '10): 12%
- Not in Plan: 46%

(5/27/09): F1000 Sample, n=156
## Backup and Recovery Technology Heat Index®

(Gauges the Immediacy of User Needs and Planned Spending for Each Technology)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Technology</th>
<th>Heat Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Backup Data Reduction / Deduplication</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>Backup Virtualization Management</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>Online Data Reduction / Deduplication</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>Email Archiving</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>Disk-to-Disk (for backup to disk targets)</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>Virtual Tape Library (VTL) for Open Systems</td>
<td>51</td>
</tr>
<tr>
<td>7</td>
<td>E-Discovery and Indexing</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>Tape Encryption</td>
<td>46</td>
</tr>
<tr>
<td>8</td>
<td>Application / Database Archiving</td>
<td>46</td>
</tr>
<tr>
<td>10</td>
<td>Continuous Data Protection</td>
<td>38</td>
</tr>
</tbody>
</table>

(5/27/09): F1000 Sample. n=156
Top 10 Storage Management Technologies in Consideration

- Virtualized Storage Provisioning / Thin Provisioning
- Information Lifecycle Management
- Continuous Data Protection
- Capacity Planning and Forecasting
- E-Discovery and Indexing
- Backup Virtualization Management
- Structured Data Classification / Categorization
- Storage Record Management / Document Management
- Application / Database Archiving
- Unstructured Data Classification / Categorization

(5/7/09): F1000 Sample. n=156
Backup Data Reduction / Deduplication – Implementation

Heat Index Rank: 1
Storage Networking
Wave 12

***Wave 7
- In Use Now: 24%
- In Pilot / Evaluation: 6%
- In Near-term Plan: 4%
- In Long-term Plan: 16%
- Not in Plan: 50%

**Wave 8
- In Use Now: 22%
- In Pilot / Evaluation: 7%
- In Near-term Plan: 15%
- In Long-term Plan: 16%
- Not in Plan: 41%

*Wave 9
- In Use Now: 9%
- In Pilot / Evaluation: 9%
- In Near-term Plan: 12%
- In Long-term Plan: 25%
- Not in Plan: 45%

*Wave 10
- In Use Now: 15%
- In Pilot / Evaluation: 15%
- In Near-term Plan: 14%
- In Long-term Plan: 30%
- Not in Plan: 25%

*Wave 11
- In Use Now: 24%
- In Pilot / Evaluation: 12%
- In Near-term Plan: 16%
- In Long-term Plan: 28%
- Not in Plan: 20%

Wave 12
- In Use Now: 27%
- In Pilot / Evaluation: 7%
- In Near-term Plan: 15%
- In Long-term Plan: 25%
- Not in Plan: 26%

*Technology was previously categorized as Deduplication. **Technology was previously categorized as De-Duplication / Capacity Optimized Storage / Single Backup Instance Store. ***Technology was previously categorized as Single Backup Instance Store Software.
What top three applications are best suited for deduplication?

- Unstructured Data
- Email Environments
- Server Virtualization Environments
- Backup Environments
- Databases
- Archiving Data
- Unstructured Environments
- Remote Office
- Legacy Databases
- Development Environments

(6/10/09): Full Sample. n=89. * Note that due to multiple responses per interview, total may exceed 100%
Deduplication Repository Size and Realization Ratio

How large is your current deduplication repository (TB)? What is your deduplication realization (e.g., 7:1, 20:1, etc.)?

Size (TB)

Ratio

Under 10
10 to 20
21 to 50
51 to 100
Over 100

Under 5:1
5:1 to 10:1
11:1 to 15:1
16:1 to 20:1
Over 20:1

Wave 10
Wave 11
Wave 12

In Summary

Consolidate Intelligently

Minimizing Organization Friction with Data Movement

Ensure Backup Scales with Storage Growth
Marc Crespi
Vice President of Product Management
ExaGrid Systems
ExaGrid Appliance

Best-in-Class Components
- Intel processors
- Seagate or Western Digital SATA drives
- LSI RAID 6 with hot spare
- ExaGrid 2nd generation software

Mix-and-match servers in a GRID for up to a 30TB full backup plus retention per GRID system

EX1000
EX2000
EX3000
EX4000
EX5000
Store Only The Bytes That Change

**Standard Disk**
- Oldest Backup: 5TB, 5TB, 5TB, 5TB, 5TB, 5TB
- Most Recent Backup: 5TB, 5TB, 5TB, 5TB, 5TB, 5TB
- Total 50TB

**ExaGrid**
- Oldest Backup: 100GB, 100GB, 100GB, 100GB, 100GB, 100GB
- Most Recent Backup: 2.5TB, 100GB, 100GB, 100GB, 100GB, 100GB
- 2x compressed
- Total 3.4TB

Oldest Backup vs. Most Recent Backup
Post-Process Provides Fastest Backups

**Fastest Backups**
- Backup directly to disk
- Utilizes unique landing zone architecture
- Post process compression and byte level data de-duplication
- No in-line processes to slow backups down
- Shortest possible backup window
Architected for Fastest Restore

90% of restores are from the latest backup

- ExaGrid stores the latest backup in its complete form ready for instant restore
- Restores from landing zone
- For restores of earlier versions byte-level changes are simply and quickly merged into the latest backup
- Allow for simultaneous restore jobs
- Allows fast offsite tape copy
- Instant DR capability provides fastest possible restores for local and off-site data copies

ExaGrid stores the latest backup in its complete form ready for instant restore.

Restores from landing zone.

For restores of earlier versions byte-level changes are simply and quickly merged into the latest backup.

Allow for simultaneous restore jobs.

Allows fast offsite tape copy.

Instant DR capability provides fastest possible restores for local and off-site data copies.
ExaGrid GRID Scalability – Capacity and Performance

ExaGrid

Virtualized GRID Architecture

Processor Head with Disk Shelves

Additional Disk Shelves
On-site and Off-site Tape Replacement

Byte-Level Delta
50 to 1 WAN Efficiency

Cross Site Protection

Secure Data Center

Primary Site Backup Job
ExaGrid System

Scheduled Data Transfer

MAN/WAN

Secondary Site Backup Job
ExaGrid System

Secure Data Center

Example:
1TB Backup
500GB (last backup)
500GB - 20GB
500GB - 20GB

Rebuilds most recent backup for Instant DR

Updated to Last Backup

On-site and Off-site Systems are Identical

For each 1TB full, WAN bandwidth of 3 mbps is required (assuming a 2% byte change rate)
Multi-site Data Center Disk Backup Topology

ExaGrid System Architecture

- 50 to 1 WAN efficiency across all sites
- Manage entire environment from single UI
- Cross-site protection available between major sites
- Reduce costs by consolidating DR copies of backup data
Unified Management Console

ExaGrid UI

- Single interface across all systems
- Simple dashboard gives quick view
Backup Job Aware Reporting

**Reports**
- Backup job view ties to backup application
- De-duplication ratio by backup job
- Replication progress and status by backup job
Customer Successes
Medina General Hospital

- 118-bed healthcare facility in Medina, Ohio.
- Employs nearly 1000 to offer 24-hour emergency services and serves more than 31,000 people a year,
- Delivers almost 900 babies annually.
- MGH’s Brunswick Campus offers a variety of medical services, such as an Immediate Care Center, an imaging center and physical and occupational therapy.
Medina General Hospital’s Challenges

- Long standing practice of backing up everything (servers, databases, files, Exchange, etc) to one server then backing that server to tape
- Full backups each weekend and ship tapes offsite for DR
- Backups exceeding window
  - Process over 12 hours to tape
  - Bad tapes would mean restart entire process
- Needed a better solution for disaster recovery
Medina General Hospital’s Solution

- Establish off-site DR to disk
- Utilize existing facilities to keep costs low
  - Operates Brunswick, OH campus 15 miles from main facility
  - Dark fiber already in place to exchange radiology files between facilities
- Needed to work with existing backup application (Symantec Backup Exec)
- Chose ExaGrid Systems Disk-based backup
Medina General Hospital’s Success and Results

- Weekly full backups reduced from 12 hours to 4
- Differential backups only take 20 minutes
- Achieved HIPAA compliance
- Bonus: ExaGrid’s ability to easily expand means MGH can add capacity as regulatory or other needs change
- Combination with Backup Exec is highly cost-effective, helping HGH IT stay within ever-tightening budgets

"Because the ExaGrid compresses our data and deduplicates it at the byte level, we are able to maximize our retention."

Michael Skrant
MIS Operations Supervisor
Medina General Hospital
Thomas & King

- 8th largest restaurant franchisee in the U.S.
- Owns and operates restaurants in Arizona, Indiana, Kentucky, Ohio and Pennsylvania
  - 89 Applebee’s restaurants
  - 6 Johnny Carino’s Italian Grill restaurants
- Over 7,500 employees
- Based in Lexington, Kentucky
Thomas & King’s Challenges

- Long backup windows
  - Full backups took 26-30 hours
- Unreliable Backups
  - No confidence in tape for restores
- Tape Management
  - Time consuming and expensive
  - Affected productivity
Thomas & King’s Solution

- Disk-based backup with data deduplication
  - Dedicated appliance, independent of SAN
- Seamlessly integrated with existing backup application
- ExaGrid system backed up to tape once a month for archival purposes
Thomas & King’s Success and Results

- **Reduced Backup Window**
  - Backup Times Cut in Half
    - Previously spent 26-30 hours on the weekend backing up data

- **More Reliable Backups**
  - Full backups completed each and every night
  - Faster restores in just seconds

- **Reduced Reliance on Tape**
  - ExaGrid backed up to tape 1x month
  - Less time spent managing and administering tapes
    - Previously spent 45 minutes per week or 1 business week per year

- **Made Backups More Efficient**
  - VMware images backed up directly from Vizioncore’s vRanger Pro
    - 3 full VM backups now reduced by over 6:1

---

“ExaGrid also fit in well with our existing backup solution, and it took away a lot of the routine, manual tasks that we ordinarily would have to do with tape.”

Jonathan DeMersseman
Manager, Infrastructure Services
Thomas & King
Why ExaGrid Disk Backup with Deduplication

- Disk that is easy to install, use and manage
- No hardware configuration or complex implementation
- Greater ability to handle rapid data growth
- Data de-duplication cuts the cost of disk by up to 95%
- Dramatically improve performance and reliability of backups and restores
- Reduce the backup window by 30% to 90%
- Better offsite strategy for DR (both with tape or tape replacement)
- Seamlessly scale as data volume or retention grows
- Better visibility via unified console and backup aware reporting
Questions?
Attendee Services

- Download a copy of today’s presentation
- Provide your feedback! Please complete our survey
- View our calendar of upcoming events
- A recorded version of this seminar will be available at www.eSeminarslive.com