Effective Data De-duplication in a Virtualized Environment

Sal Fernando
Chief Technical Architect
In Summary: Why EMC for VMware – 5 Reasons

- Simple, easy-to-use solutions that integrate with and extend all VMware advanced functions (e.g., DRS, Storage VMotion, SRM)
- Flexibility for iSCSI, FC, and NFS – every protocol VMware needs = no risk, no sacrifices, no protocol wars
- Proven scaling, proven replication, proven availability, proven tier 1 app solutions
- Unique capabilities in VMware environments:
  - Backup built for VMware
  - VDI solutions – from 1 image to 10,000 in minutes
  - Change control and end-to-end virtual-to-physical management
  - Virtual appliances
  - Joint VMware/Exchange/SQL/Oracle/SAP solutions
- Net – more customers choose EMC for VMware

“For virtual servers, networked storage solutions are more heavily weighted toward EMC storage. In previous years, storage attached to virtual servers was highly captive relative to the server hardware purchase.”

  — Source: IDC’s Server Virtualization 2007 Multiclient Study, Dec 2007
  * EMC and VMware were sponsors of this study

Some Statistics

Backup, DR is expensive – and it fails!

“75% of storage management is backup and recovery yet 30% of all data recovery instances fail due to botched backups”

Forrester Research

Your customer environment is growing

“Storage & Infrastructure Growth rates 50%->80% per annum”

Gartner & confirmed by EMC & Partners via Classification of customer environments

Your customer has no information about how data is created in their environment

“Dormant data 50-96%”

“Duplicate files >20%”

“Illegal files >15%”

“Duplicate block level data being backed up to 80%”

“Storage utilization is unbalanced – some apps run out and need new storage whilst others are under-utilized”

“Your customers think tape is cheap”

“Most don’t have shared storage”

EMC & Partners via Classification of customer environments

Budgeting issues

Your customer is wasting energy (electricity) and paying for it

Their archival (long term retention) policy most likely won’t work

Only 10-15% of their CPU population is utilised

Their environment isn’t secure

Nearly) Everyone!
EMC Portfolio

Strengthen and Enable Information Infrastructure

LEVERAGE
- documentum
- casl Once
- captiva
- X-HIVE

OPTIMIZE
- Astrum
- VMware
- Relativity
- INTEGRITY
- PROACTIVITY
- nLayers
- akimbi

PROTECT
- Legato
- dantz
- RSA
- Avamar
- Authentic
- Verid

STORE
...
1. “Unlimited storage”
2. Reduced real-estate for IT
3. Increase service levels (DR/BC)
4. Faster network
5. Security
6. Compliance
7. Increase IT’s ROI
EMC Avamar

Enabling Technology Transforms Data Protection

- Global data de-duplication defuses the explosion of backup data
  - Radically reduce the media, network, and time required for backup
- Alter the fundamental economics of disk versus tape
  - Accelerate shift to disk as primary medium for backup
- An alternative to archaic IT processes (shipping tapes for disaster recovery)
  - Automated, encrypted remote copy over existing WANs
Avamar Solves Customer Challenges

- Remote office/branch office backup
  - Achieve data protection across existing IP WANs
  - Centralize management
- Bandwidth-constrained backup
  - Accelerate backup by reducing amount of data transferred
  - Meet shorter backup windows
- VMware backup
  - Reduce backup impact on physical machines
  - De-duplicate within and across virtual machines
Tangible Customer Results

**Before Avamar**
- 50 TB primary backup, 8 TB per daily incremental, weekly total = 98 TB

**With Avamar**
- 0.5 TB daily backups, 3.5 TB cumulative per week
- 70-hour staged full backup window reduced to four hours
- Avamar wins Morgan Stanley CIO Innovation Award

“Avamar has a game-changing solution. Through their innovative technology, we have been able to rethink our backup, recovery, and replication infrastructure, providing Morgan Stanley with better local and remote recovery at a greatly reduced TCO.”

— Guy Chiarello, CTO/CIO, Morgan Stanley
Tangible Customer Results

Before Avamar

- 73 remote offices, backup to local, direct-attached tape drives
- No local IT staff
- Daily backup required five hours per site
- Six hours to restore entire server

With Avamar

- Four hours to backup all 73 offices via existing WAN (56k–T1) to central Avamar server
- 45 minutes to restore entire server—files restored in seconds
- Centralized management and control

“Avamar enabled us to reduce administrative support requirements by 80 percent, reduce backup windows by 90 percent, and recover lost files and servers in minutes rather than hours.”

— Mike DePhillip, Virginia DMV
Storage Consolidation

Before

‘G’ drive

Utilization: 88%
Dormant Data: 73%
Duplicate Data: 5%

3 yr tape backup: 64 TB

After

Tier 1 – High Perf: 7%
Tier 2 – Dormant ➔ Cheaper: 20%
Tier 3 – Others: 0%

15 TB ➔ Less tape physically more manageable

3 yr tape backup: 15 TB

= 93,000 music CDs… HMV Singapore believe they have on display 10,000 CD/DVD’s on all 4 floors in their Orchard store !!

IT operations ➔ Higher value IT work
VIRTUALIZED servers working as ONE COMPUTER

1. Power to spare
2. Synchronized computing at DR site
3. Higher availability
4. Manageability
## What Users Can Expect

<table>
<thead>
<tr>
<th>Storage</th>
<th>Pool Computing</th>
<th>Network</th>
<th>IT Security</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ Unlimited storage</td>
<td>✔ Shorter time to enable IT services</td>
<td>✔ Regulatory needs</td>
<td>✔ Secure end-user computing</td>
</tr>
<tr>
<td>✔ Secure &amp; always there</td>
<td>✔ Rep offices no longer need to ‘manage’ IT</td>
<td>✔ Secure &amp; full service remote access</td>
<td>✔ BCP readiness</td>
</tr>
<tr>
<td>✔ More ‘G’ ‘H’ drive</td>
<td>✔ High performance computing</td>
<td>✔ Secure wireless - anywhere</td>
<td>✔ Robust IT processes to safeguard corporate information systems</td>
</tr>
<tr>
<td>✔ Faster access</td>
<td>✔ Better end-user PC support</td>
<td>✔ Faster network performance</td>
<td>✔ Complies with possible regulatory req</td>
</tr>
<tr>
<td>✔ Search engine</td>
<td>✔ Synchronized DRP computing power / capability</td>
<td>✔ Redundancy → BCP</td>
<td>✔ Anti Spy &amp; SPAM ware solutions</td>
</tr>
<tr>
<td>✔ Trace / Audit</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Copyright 2007 EMC Corporation. All rights reserved.
**Financial**

<table>
<thead>
<tr>
<th>Metric</th>
<th>2008</th>
<th>2009 (Proj)</th>
<th>2010 (Proj)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost avoidance from hardware savings</td>
<td>$400k</td>
<td>$1470k*</td>
<td>$500k</td>
</tr>
<tr>
<td>Cost savings from Real Estate (Server rack space)</td>
<td>$288k</td>
<td>$366k</td>
<td>$504k</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$688k</strong></td>
<td><strong>$1836k</strong></td>
<td><strong>$1004k</strong></td>
</tr>
</tbody>
</table>

* Includes one time cost avoidance from purchasing 97 DR servers

**Operational Excellence**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downtime required for system maintenance</td>
<td>Required downtime</td>
<td>No downtime</td>
</tr>
<tr>
<td>Effort required to generate inventory &amp; utilization reports</td>
<td>Days</td>
<td>minutes</td>
</tr>
<tr>
<td>Reduce staff effort required to create Test and QA environments</td>
<td>Weeks</td>
<td>minutes</td>
</tr>
</tbody>
</table>

**User Experience**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity</td>
<td>Quota-based</td>
<td>Unlimited</td>
</tr>
<tr>
<td>High Availability</td>
<td>4</td>
<td>104</td>
</tr>
<tr>
<td>Responsiveness to business requirements (server provisioning)</td>
<td>4 weeks</td>
<td>1 hour</td>
</tr>
</tbody>
</table>

**IT Experience**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained on VMware</td>
<td>0</td>
<td>2*</td>
</tr>
<tr>
<td>Servers to administrator</td>
<td>52:1</td>
<td>72:1</td>
</tr>
</tbody>
</table>

* 2 more to be trained in 2008/9