EMC BACKUP MEETS BIG DATA

Strategies To Protect Greenplum, Isilon And Teradata Systems
Agenda

• Big Data: Overview, Backup and Recovery
• EMC Big Data Backup Strategy
• EMC Backup and Recovery Solutions for Big Data
  – EMC Greenplum
  – Teradata
  – EMC Isilon
Big Data – Volume, Velocity and Variety

- **1990s** (RDMBS, Data Warehouses, etc.)
- **2000s** (Content and Digital Asset Management)
- **2010s** (Hadoop, MapReduce, NoSQL, Cloud)

VOLUME OF INFORMATION

- **LARGE**
  - MEASURED IN
    - TERABYTES
      - 1TB = 1,000GB
  - MEASURED IN
    - PETABYTES
      - 1PB = 1,000TB
  - WILL BE MEASURED IN
    - EXABYTES
      - 1EB = 1,000PB

**LE: LARGE EXA焼銀**

- **SMALL**

1TB = 1,000GB

1PB = 1,000TB

1EB = 1,000PB

© Copyright 2012 EMC Corporation. All rights reserved.
Big Data Transforms Business
Every Industry Will Feel Impact

Financial
Telecom
Retail
Energy
Insurance

Government
Healthcare
Cyber Security
Advertising
Gaming

© Copyright 2012 EMC Corporation. All rights reserved.
Big Data Backup And Recovery

• Challenges
  – Data Deluge
  – Performance And Scale
  – Data Islands

• Big Data Backup
  – The Journey Has Just Begun...
End-To-End Solutions
Backup Software & Storage Designed To Work Together

Implements

- **Performance**: Up To 50% Faster Than Traditional Solutions
- **Simplicity**: One User Interface—Common Policy Management
- **Predictability**: Proven Best Practices—Over 20,000 Installations
EMC Big Data Backup & Recovery Strategy

Data Domain Provides The Foundation With Backup Software

Big Data systems -> Backup App / Utility -> Data Domain System

- CIFS, NFS, NDMP, DD Boost
- Ethernet
- Virtual Tape Library (VTL) over Fibre Channel
EMC Data Domain: Leadership and Innovation

A history of industry firsts

- **2003**: First deduplication NAS
- **2004**: First deduplication volume replication
- **2005**: First deduplication virtual tape library
- **2006**: First deduplication directory replication
- **2007**: Largest deduplication array
- **2008**: Fastest backup controller
- **2009**: Cascaded replication
- **2010**: First long-term retention system for backup and archive
- **2011**: First distributed processing
- **2012**: First inline deduplication for compliant archiving
PURPOSE BUILT BACKUP APPLIANCES
OPEN SYSTEMS + MAINFRAME CAPACITY

2011 Total Market
927,474 TB

## Data Deduplication: Technology Overview

### Store More Backups In A Smaller Footprint

#### Data Reduction Summary

<table>
<thead>
<tr>
<th></th>
<th>Logical</th>
<th>Estimated Reduction</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRIDAY FULL</td>
<td>1 TB</td>
<td>2–4x</td>
<td>250 GB</td>
</tr>
<tr>
<td>Monday Incremental</td>
<td>100 GB</td>
<td>7–10x</td>
<td>10 GB</td>
</tr>
<tr>
<td>Tuesday Incremental</td>
<td>100 GB</td>
<td>7–10x</td>
<td>10 GB</td>
</tr>
<tr>
<td>Wednesday Incremental</td>
<td>100 GB</td>
<td>7–10x</td>
<td>10 GB</td>
</tr>
<tr>
<td>Thursday Incremental</td>
<td>100 GB</td>
<td>7–10x</td>
<td>10 GB</td>
</tr>
<tr>
<td>Second FRIDAY FULL</td>
<td>1 TB</td>
<td>50–60x</td>
<td>18 GB</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2.4 TB</td>
<td>7.8x</td>
<td>308 GB</td>
</tr>
</tbody>
</table>

#### Backup Data Diagram

- **Friday Full Backup**
- **Monday Incremental**
- **Tuesday Incremental**
- **Wednesday Incremental**
- **Thursday Incremental**
- **Second Friday Full Backup**

---

© Copyright 2012 EMC Corporation. All rights reserved.
## Retain: Store More for Longer with Less

Over one year of retention in 3U of Data Domain deduplication storage

### Backup Data

<table>
<thead>
<tr>
<th>Period</th>
<th>Cumulative Logical</th>
<th>Estimated Reduction</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Full</td>
<td>1 TB</td>
<td>4x</td>
<td>250 GB</td>
</tr>
<tr>
<td>April 7</td>
<td>2.2 TB</td>
<td>8x</td>
<td>288 GB</td>
</tr>
<tr>
<td>April 14</td>
<td>3.4 TB</td>
<td>10x</td>
<td>326 GB</td>
</tr>
<tr>
<td>April 21</td>
<td>4.6 TB</td>
<td>13x</td>
<td>364 GB</td>
</tr>
<tr>
<td>April 28</td>
<td>5.8 TB</td>
<td>14x</td>
<td>402 GB</td>
</tr>
<tr>
<td>May 31</td>
<td>10.6 TB</td>
<td>19x</td>
<td>554 GB</td>
</tr>
<tr>
<td>June 30</td>
<td>15.4 TB</td>
<td>21x</td>
<td>706 GB</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>15.4 TB</strong></td>
<td><strong>21x</strong></td>
<td><strong>706 GB</strong></td>
</tr>
</tbody>
</table>
Data Integrity: Data Invulnerability Architecture

End-to-end data verification
- Checksum
- Deduplication, write to disk
- Verify

Self-healing file system
- Cleaning
- Expired data
- Defrag
- Verify

Other
- RAID 6
- NVRAM
- Snapshots

File System
Deduplication
Local Compression
RAID

Generate Checksum
Verify Data

End-to-end data verification
Re-Checksum and Compare

Verify the file system metadata integrity
Verify user data integrity
Verify stripe integrity
Network-Efficient Replication for True Disaster Recovery

95–99% cross-site bandwidth reduction

Flexible replication
- One-to-many
- Many-to-one
- Bi-directional
- System-to-system
- Cascaded

Source: Remote sites
Destination: Data Center Hub
Supports hundreds of remote sites

Data Domain DD890

Archive data
Backup data
Data Domain Boost Software

- Distributes parts of deduplication process to backup server or application clients
- Speeds backups by up to 50 percent
- Enables more efficient resource utilization
- Provides application control of Data Domain replication process
- Supports majority of backup software market and native utilities in industry leading databases
## Industry’s Most Scalable Inline Deduplication Systems

### Data Domain Software Options
- DD Boost
- DD Encryption
- DD Extended Retention
- DD Replicator
- DD Retention Lock
- DD Virtual Tape Library

### Speed (DD Boost)
- DD160: 1.1 TB/hr
- DD620: 2.4 TB/hr
- DD640: 3.4 TB/hr
- DD670: 5.4 TB/hr
- DD860: 9.8 TB/hr
- DD890: 14.7 TB/hr
- DD990: 31.0 TB/hr

### Speed (other)
- DD160: 667 GB/hr
- DD620: 1.1 TB/hr
- DD640: 2.3 TB/hr
- DD670: 3.6 TB/hr
- DD860: 5.1 TB/hr
- DD890: 8.1 TB/hr
- DD990: 15.0 TB/hr

### Logical Capacity
- DD160: 40–195 TB
- DD620: 83–415 TB
- DD640: 0.32–1.6 PB
- DD670: 0.6–2.7 PB
- DD860: 1.4–7.1 PB
- DD890: 5.7–28.5 PB
- DD990: 13–65 PB

### Usable Capacity
- DD160: Up to 3.98 TB
- DD620: Up to 8.3 TB
- DD640: Up to 32.2 TB
- DD670: Up to 55.9 TB
- DD860: Up to 142 TB
- DD890: Up to 570 TB
- DD990: Up to 285 TB

1 With DD Extended Retention software option
### Data Domain And Big Data

**Overcome Your Big Data Backup And Recovery Challenges**

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data deluge</strong></td>
<td>High speed inline deduplication</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>Up to 248 TB can be backed up in &lt; 8 hours (31 TB/hr)</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>Protects up to 65 PB of logical capacity in a single system</td>
</tr>
<tr>
<td><strong>Data islands</strong></td>
<td>Simultaneously supports NFS, CIFS, VTL, DD Boost and NDMP</td>
</tr>
</tbody>
</table>
EMC Backup Solutions For Big Data
Big Data Systems

- EMC Greenplum
- Teradata
- EMC Isilon
EMC Greenplum: Purpose-built For Big Data

- EMC Greenplum
  - A Shared Nothing, Massively Parallel Processing (MPP) Data Warehouse System

- EMC Greenplum Data Computing Appliance (DCA)
  - Built Using Standard Hardware Components
  - Modular For Easy Scaling

- Core Principle Of Data Computing
  - Move The Processing Dramatically Closer to the Data and to the People

- Fast Data Loading
- Extreme Performance & Elastic Scalability
- Unified Data Access
Platform Independence
Delivers Choice and Flexibility

- **Data Computing Appliance**
  Ideal for Production Environments

- **Software-Only**
  Ideal for Q/A or DR

- **Virtualized Infrastructure**
  Ideal for Test & Development
Greenplum Backup & Recovery Requirements

1. Backup As New Data Is Added To Database ✓
2. Backup Database Data And Configuration Info ✓
3. Recover Data To A Consistent Point In Time ✓
Efficient Greenplum Backup
With Data Domain Systems

- Only Certified Backup And Recovery Solution For Greenplum
- Flexible Deployment – NFS Or DD Boost
- 10 To 30x Reduction In Backup Storage Requirements
- Network-efficient, Encrypted Replication
- Backup All Segment Servers In Parallel
DD Boost for EMC Greenplum

- Under The DBA’s Direct Control
- Faster, More Efficient Backup And DR
- Simplified Configuration And Administration
Teradata
Teradata – Data Warehousing

• Database Software, Enterprise Data Warehousing, Data Warehouse Appliances, And Analytics

• Tape Based Backup Not Meeting Requirements
Teradata Backup Requirements

1. Teradata Backup Archive Restore (BAR) Certified

2. Coordinates Across All The Components Of The System

3. Minimizes The Time Data Stays In Read-only Mode
Efficient Teradata Backup
With Data Domain Systems

- Certified By Teradata As The Deduplication Platform Of Choice
- Significantly Enhances Database Backup And Recovery
  - Store More Backups Onsite Longer
  - Multisite Disaster Recovery
  - Ultra-safe Storage for Reliable Recovery
Teradata And EMC – The Relationship

- Teradata An EMC OEM Partner
- Teradata Provides Implementation And Customer Support Services
- Teradata Certified Data Domain systems as a BAR “Advocated Solution”
  - Best Practices Package for Backup and Recovery
  - Optimized for Teradata Environments
  - Scalability Fits EDW and ADW Needs
Teradata BAR And Data Domain

- Faster, More Efficient Backup And DR
- Teradata BAR Coordinates Fast Data Extraction
- Flexible Deployment
  - DD Boost (NetBackup) or VTL (IBM TSM and BakBone NetVault)
EMC Isilon
EMC Isilon – Scale Out Storage

- Broad Adoption
- Over 2,100 Global Customers
- OneFS - Innovative Scale Out, Operating Environment
- Industry Leading, Linearly Scalable Performance
Efficient Isilon Backup
With NetWorker And Data Domain

• Leverage NetWorker Unified Backup and Recovery Software
  – Industry Standard NDMP
  – Advanced Integration With Data Domain
  – Ability To Exclude Certain Data Types

• 3 Deployment Options
NetWorker Storage Node over IP

Option 1

• NDMP Backup Over IP To NetWorker Storage Node

• Advanced Integration With DD Boost
  – Faster, More Efficient Backup
  – NetWorker Control Of Data Domain Replication
Data Domain NDMP Tape Server over IP
Option 2

• Backup Over IP Directly To Data Domain
  – Via Data Domain NDMP Tape Server Capability

• With EMC NetWorker Or 3rd Party Backup Software
Isilon Backup Accelerator Over FC

Option 3

• Backup Over Fibre Channel To Data Domain
  - Via Isilon Backup Accelerator Node

• Leverage Existing SAN Infrastructure
Big Data Backup – We’ve Got You Covered!

- Big Data Backup And Recovery
  - Critical As Well As An Unaddressed IT Challenge
- EMC Data Domain Deduplication Storage System Ideal For Big Data Backup And Recovery
- Integrated Backup/Recovery Solutions Available For EMC Greenplum, Teradata And EMC Isilon
EMC BACKUP MEETS BIG DATA

Contact Your EMC Or Partner Sales Rep for Next Steps
Learn More About EMC Backup: emc.com\backup
Join The Conversation: www.thebackupwindow.com