PASIG – Tiered Storage for Archiving Basics

Digital Preservation 101 Session

Philippe Deverchère
EMEA Storage CTO

September 16, 2014
Archiving Brings more than just Archiving

Storage Requirements after 1 year to Protect 100 TB of Raw Data

Volume at the end of Year 1: 3.2 PB

- Raw Data [120 TB]
- RAID Overhead [24 TB]
- Daily Incrementals [37 TB]
- Weekly Fulls [1.54 PB]
- Monthly Fulls [1.33 PB]
- Yearly Fulls [120 TB]

Retention Policy for backups
- Daily Incrementals: 1 month
- Weekly Fulls: 90 days
- Monthly Fulls: 1 year
- Yearly Full: forever

Assumptions
- RAID Overhead: 20%
- Daily Change Rate: 1%
- No data archiving
Archiving Brings more than just Archiving
What if you manage to **archive 80%** of your data?

**Retention Policy for backups**
- Daily Incrementals: 1 month
- Weekly Fulls: 90 days
- Monthly Fulls: 1 year
- Yearly Full: forever

**Assumptions**
- RAID Overhead: 20%
- Daily Change Rate: 1%
- No data archiving

Volume at the end of Year 1: **3.2 PB**

- **Raw Data** [120 TB]
- **RAID Overhead** [24 TB]
- **Daily Incrementals** [37 TB]
- **Weekly Fulls** [1.54 PB]
- **Monthly Fulls** [1.33 PB]
- **Yearly Fulls** [120 TB]

**Volume at the end of Year 1:** 3.2 PB
Archiving Brings more than just Archiving
What if you manage to **archive 80%** of your data?

*Volume at the end of Year 1: 634 TB*

- **Raw Data** [24 TB]
- **RAID Overhead** [5 TB]
- **Daily Incrementals** [7 TB]
- **Weekly Fulls** [308 TB]
- **Monthly Fulls** [266 TB]
- **Archived or Tiered Data** x 2 [192 TB]

**Retention Policy for backups**
- Daily Incrementals: 1 month
- Weekly Fulls: 90 days
- Monthly Fulls: 1 year
- Yearly Full: forever

**Assumptions**
- RAID Overhead: 20%
- Daily Change Rate: 1%
- No data archiving
# Concepts

## Understanding Purpose, Data Type and Access Pattern

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Data Protection</th>
<th>Disaster Recovery</th>
<th>Data Archiving</th>
<th>Data Tiering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short term protection of data for operational recovery (corruption, deletion, local incident)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Type</td>
<td>Dynamic data in production – focus on data, not the operating environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access Pattern</td>
<td>File, share or volume might be restored after outage / corruption / deletion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Concepts

## Understanding Purpose, Data Type and Access Pattern

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Data Protection</th>
<th>Disaster Recovery</th>
<th>Data Archiving</th>
<th>Data Tiering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Short term protection of data for operational recovery (corruption, deletion, local incident)</td>
<td>Overall protection of an IT production to keep functioning in the event of a disaster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Type</td>
<td>Dynamic data in production – focus on data, not the operating environment</td>
<td>Dynamic data in production – focus on the operating environment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access Pattern</td>
<td>File, share or volume might be restored after outage / corruption / deletion</td>
<td>Full data sets must be available with minimal RTO and RPO after disaster</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Concepts

## Understanding Purpose, Data Type and Access Pattern

<table>
<thead>
<tr>
<th></th>
<th>Data Protection</th>
<th>Disaster Recovery</th>
<th>Data Archiving</th>
<th>Data Tiering</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Short term protection of data for operational recovery (corruption, deletion, local incident)</td>
<td>Overall protection of an IT production to keep functioning in the event of a disaster</td>
<td>Long term preservation of data for business operations and compliance</td>
<td></td>
</tr>
<tr>
<td><strong>Data Type</strong></td>
<td>Dynamic data in production – focus on data, not the operating environment</td>
<td>Dynamic data in production – focus on the operating environment</td>
<td>Fixed content with ongoing business or compliance value</td>
<td></td>
</tr>
<tr>
<td><strong>Access Pattern</strong></td>
<td>File, share or volume might be restored after outage / corruption / deletion</td>
<td>Full data sets must be available with minimal RTO and RPO after disaster</td>
<td>Individual files or tables are retrieved as needed</td>
<td></td>
</tr>
</tbody>
</table>
## Concepts

### Understanding Purpose, Data Type and Access Pattern

<table>
<thead>
<tr>
<th></th>
<th>Data Protection</th>
<th>Disaster Recovery</th>
<th>Data Archiving</th>
<th>Data Tiering</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Short term protection of data for operational recovery (corruption, deletion, local incident)</td>
<td>Overall protection of an IT production to keep functioning in the event of a disaster</td>
<td>Long term preservation of data for business operations and compliance</td>
<td>Relocation of infrequently accessed data to reduce cost</td>
</tr>
<tr>
<td><strong>Data Type</strong></td>
<td>Dynamic data in production – focus on data, not the operating environment</td>
<td>Dynamic data in production – focus on the operating environment</td>
<td>Fixed content with ongoing business or compliance value</td>
<td>Mostly static content with ongoing business or compliance value</td>
</tr>
<tr>
<td><strong>Access Pattern</strong></td>
<td>File, share or volume might be restored after outage / corruption / deletion</td>
<td>Full data sets must be available with minimal RTO and RPO after disaster</td>
<td>Individual files or tables are retrieved as needed</td>
<td>Individual files or tables are automatically retrieved as needed</td>
</tr>
</tbody>
</table>
## Business Advantages

### Why a sound strategy on DP, DR, Archiving and Tiering is important

<table>
<thead>
<tr>
<th>Data Protection</th>
<th>Disaster Recovery</th>
<th>Data Archiving</th>
<th>Data Tiering</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ensure data availability and integrity</td>
<td>• Ensure business continuance</td>
<td>• Meet legal and corporate regulations</td>
<td>• Reduce TCO</td>
</tr>
<tr>
<td>• When properly done, protect against a wide range of threats</td>
<td>• Enforce Business Impact Analysis</td>
<td>• Minimize risks of data loss</td>
<td>• Bring business operational improvements</td>
</tr>
<tr>
<td></td>
<td>• Streamline business processes</td>
<td>• Enforce sound data classification</td>
<td>• Unlock the value of previously-untapped archived corporate data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Decrease backup costs</td>
<td></td>
</tr>
</tbody>
</table>
## TCO Evaluation

### Influence of Production/Archive and Disk/Tape Ratios

<table>
<thead>
<tr>
<th>Production / Archive Ratio</th>
<th>Approach</th>
<th>Backup</th>
<th>Archive</th>
<th>Total</th>
<th>Relative TCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>Disk Only</td>
<td>8.0</td>
<td>0.0</td>
<td>8.0</td>
<td>1,00</td>
</tr>
<tr>
<td>100%</td>
<td>Disk + Tape</td>
<td>6.0</td>
<td>0.0</td>
<td>6.0</td>
<td>0,75</td>
</tr>
<tr>
<td>80%</td>
<td>Disk Only</td>
<td>6.8</td>
<td>0.5</td>
<td>7.3</td>
<td>0,92</td>
</tr>
<tr>
<td>80%</td>
<td>Disk + Tape</td>
<td>4.8</td>
<td>0.1</td>
<td>4.9</td>
<td>0,61</td>
</tr>
<tr>
<td>30%</td>
<td>Disk Only</td>
<td>3.8</td>
<td>1.9</td>
<td>5.7</td>
<td>0,71</td>
</tr>
<tr>
<td>30%</td>
<td>Disk + Tape</td>
<td>1.8</td>
<td>0.2</td>
<td>2.0</td>
<td>0,25</td>
</tr>
</tbody>
</table>

**Backup**
- Number of copies of production data created for backup on disk: 8
- Number of backup copies on disk in a mixed environment: 5
- Number of backup copies on tape in a mixed environment: 3
- Number of backup copies of disk archived data: 2
- Number of backup copies of disk archived data in a mixed environment: 0
- Number of backup copies of tape archived data in a mixed environment: 0
- Relative cost of ownership between disk and tape for backup: 1

**Archive**
- Number of archive copies on disk in a disk only environment: 3
- Ratio of disk versus tape to hold archived data in a mixed environment: 0.2
- Number of archive copies on disk in a mixed environment: 1
- Number of archive copies on tape in a mixed environment: 2
- Relative cost of ownership between disk and tape for archive: 0.9
TCO Evaluation

Influence of Production/Archive and Disk/Tape Ratios

Relative TCO [production/archive ratio]