Storage Considerations for Database Archiving

Julie Lockner, Vice President
Solix Technologies, Inc.
The material contained in this tutorial is copyrighted by the SNIA.

Member companies and individuals may use this material in presentations and literature under the following conditions:
- Any slide or slides used must be reproduced without modification.
- The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.
Abstract

Storage Considerations for Database Archiving

- According to the SNIA Archive Task Force 100 Year Archive Requirements Survey, over 81% of respondents feel database data is the most at risk for long-term accessibility. In order to maintain access to application data stored in databases over longer data retention periods, unique storage requirements and considerations need to be taken. The way in which you might implement ILM or tiered storage is different when applied to structured content. This tutorial will discuss archiving strategies and long term implications of how database data should be stored for different access requirements. Several use cases will be discussed.
Introduction to Database Archiving

Database Classification Examples

Database Archive Storage Considerations

Getting Started
About SNIA and the DMF

About the Storage Networking Industry Association (SNIA)  
www.snia.org
- SNIA’s primary goal is to ensure that storage networks become complete and trusted solutions across the IT community

About the SNIA Data Management Forum (DMF)  
www.snia-dmf.org
- The DMF is a sub-group of SNIA acting as the worldwide authority on Data Management, Data Protection and ILM
- The DMF is a collaborative storage industry resource available to anyone responsible for the accessibility and integrity of their organization’s information.

<table>
<thead>
<tr>
<th>DMF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Protection Initiative (DPI)</strong></td>
</tr>
<tr>
<td><strong>Information Lifecycle Management Initiative (ILMI)</strong></td>
</tr>
<tr>
<td><strong>Long term Archive and Compliance Storage Initiative (LTACSI)</strong></td>
</tr>
<tr>
<td>Defining new approaches and best practices for data protection and recovery</td>
</tr>
<tr>
<td>Developing, teaching and promoting ILM practices, implementation methods, and benefits</td>
</tr>
<tr>
<td>Addressing challenges in developing, securing, and retaining long-term digital archives</td>
</tr>
</tbody>
</table>
What is your Database TCO?

- What is your annual database growth
  - Upgrading Applications can increase database size **40-60%**
  - Mergers, acquisitions, and new application implementations add to size
  - Adding data, Business As Usual

- After 2 years in production, how much of your data no longer accessed?
  - How often do users actually access inactive or aged data?
  - When does data becomes read-only?

- How much are you spending on managing your database applications?
  - Include storage and servers needed for production support, test & dev copies
  - Include daily Incremental & weekly full backups of PROD and Copies
  - MetaGroup analyst, Charlie Garry, estimated 70%

- How much could you save if you implemented Database Archiving?
  - Reduced storage and server requirements & backup media volumes
  - Improved operational efficiencies and application performance

- How to get started?
According to SNIA’s 100 Year Archive Task Force Survey, Database Data is Most at Risk for Long-Term Readability

“100 YrATF Analysis: The industry’s focus on retaining unstructured data and email has left an important gap. According to these respondents, databases (including all enterprise applications running on databases) are at far more risk than any other type of information.”

http://www.snia-dmf.org/100year/index.shtml
Introduction to Database Archiving

- Classify data into “Active”, “Inactive” and “Aged” categories based on business processes and value to the business
- Relocate archive eligible transactions to online/offline archive
- Maintain appropriate levels of access to archived data

Sample Database Archiving Configuration

1. Classify data into “Active”, “Inactive” and “Aged” categories based on business processes and value to the business.
2. Relocate archive eligible transactions to online/offline archive.
3. Maintain appropriate levels of access to archived data.

---

Tier 1

Tier 2

Tier 3

Disposal
Database Archiving Concepts

- Database Archiving or “Tiering” Characteristics
  - Data is classified by a policy
    - i.e. age of data, application business logic
  - Database data is moved to an online or offline archive
    - i.e. separate database, database partition, file, etc.
  - Accessing Archived Database Data
    - Access requirements vary by regulation and corporate governance policies

- Reports Archiving
  - Database data is written to a report and the report is archived
Data Classification is App Specific

- Database Classification Characteristics
  - Transactions may be updated until “closed”
  - “Closed” may not a function of time
  - A transaction may reside in multiple tables with relational dependencies
  - Application context may be required to access archived data
Classification for Archiving

Classification is driven by the specific application

- **Online Transaction Processing Application Examples**
  - **General Ledger**
    - Constraint = Closed books, journals, balances
    - Data retention policy = 2 years + current quarter
  - **Sales Order**
    - Constraint = Closed sales orders
    - Data retention policy = warranty period expires

- **Data Warehouse Application Examples**
  - **Object Data Stores** older than 3 years
  - “Last accessed” older than 5 years

Levels of Classification for Database Data

- **Active** = High Value, High Accessibility, High Performance
- **Inactive** = Medium Value, Medium Accessibility
- **Aged** = Lower Value, Low Accessibility
Data Classification Example: Sales Order

- **Tier 1**
  - New Order
  - Product Ships
- **Tier 2**
  - Warranty 1 Year
- **Tier 3**
  - Quality Recall
- **Disposal**
Reports Archiving writes data to a file that can be archived as unstructured content.
Collaborate with Line of Business Users to determine archive data access requirements

- Application Users with different patterns of access can impact classification policies
- Database Archive Containers with different I/O requirements can impact archive access retrieval rates

Storage Tiers for the Archive Container should be mapped to the expected usage patterns and archive container

- i.e. CAS may not be a good tier for a Database archive container but would be excellent for a file or report archive container
Collect End User Access Requirements

End User Access Requirements should be collected prior to designing the database archive infrastructure to answer questions such as:

- What is the needed Performance/Access speed of the archive?
- How much data will actually be archived?
- Do I need to store the archive data on WORM media?
- Does the archive data need to be accessible from the native application?
DB Archiving Architecture Considerations

- **Backup & Restore Processes**
  - Archiving to a central archive removes data from prod & all copies
  - Consider backup and restore procedures to include the archive

- **Disaster Recovery**
  - Recovery of the production database is significantly reduced
  - If the chosen DB archiving solution executes at the RDBMS or application tier, all network or block level D/R strategies should not be impacted
  - Need to consider SLA’s for production and the archive

- **Replication**
  - Check if replicated production databases may or may not need access to the archive
- Older transactions are archived in batch on a periodic basis
  - Full backup of History before and after an archive batch

- Production database continues to add new transactions
  - Continue daily incremental and weekly full backups on production
  - Full backup of Prod before and after an archive batch
Database Archiving & D/R

- Make sure Archiving is considered in the D/R strategy
  - Need to validate the archiving solution does not impact the D/R solution

- Validate if the archive needs to have the same Recovery SLA
  - Different Replication schedules can impact risk associated with data loss

Example: Risk of Data Loss when using different replication schemes
Database Classification Assessments

Get started – an ERP Example

Find out where the data is growing the fastest

Use growth charts to illustrate forecast requirements for storage and servers

Projected Storage Growth for Production DB & Copies

Storage Considerations for Database Archiving
© 2008 Storage Networking Industry Association. All Rights Reserved.
### Archive Potential Impact by Application

Meet with Line of Business to Determine Data Access Requirements and Data Retention Policies

<table>
<thead>
<tr>
<th>Application</th>
<th>Retention Policy (months)</th>
<th>Data Volume Before Archiving (GB)</th>
<th>Data Volume Eligible for Archive (GB)</th>
<th>Data Volume After Archiving (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Ledger</td>
<td>24</td>
<td>54</td>
<td>15.84</td>
<td>38.16</td>
</tr>
<tr>
<td>Accounts Payables</td>
<td>12</td>
<td>47.93</td>
<td>13.14</td>
<td>34.79</td>
</tr>
<tr>
<td>Workflow</td>
<td>1</td>
<td>28.42</td>
<td>4.38</td>
<td>24.04</td>
</tr>
<tr>
<td>Content Management</td>
<td>12</td>
<td>1.77</td>
<td>0.77</td>
<td>1</td>
</tr>
<tr>
<td>Accounts Receivables</td>
<td>12</td>
<td>32.54</td>
<td>11.9</td>
<td>20.64</td>
</tr>
<tr>
<td>Inventory</td>
<td>12</td>
<td>0.11</td>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td>Human Resources</td>
<td>12</td>
<td>25</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Fixed Assets</td>
<td>12</td>
<td>4.12</td>
<td>0.4</td>
<td>3.72</td>
</tr>
<tr>
<td>Project Accounting</td>
<td>12</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Other ERP Apps</td>
<td>95</td>
<td>6.07</td>
<td>0</td>
<td>6.07</td>
</tr>
<tr>
<td>Custom Tables</td>
<td>95</td>
<td>0.97</td>
<td>0</td>
<td>0.97</td>
</tr>
<tr>
<td>ERP Non-Archive Tables</td>
<td>95</td>
<td>12.75</td>
<td>0</td>
<td>12.75</td>
</tr>
<tr>
<td>System Tables</td>
<td>95</td>
<td>22</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td><strong>Totals ---&gt;</strong></td>
<td></td>
<td><strong>235.78</strong></td>
<td><strong>56.43</strong></td>
<td><strong>179.35</strong></td>
</tr>
</tbody>
</table>

**Before** 236 GB **After** 179 GB **Savings** 56 GB
Reduced data volumes can drive savings in:

- Disk (prod and clones)
- Tape (backups)
- Servers & CPU
- Reduced downtime
- Archiving to tiered storage

### Total Cost of Ownership Comparison

#### 3 Year $$$ Spend - Current v. w/AJ

- Current $$$ Spend
- w/ AJ $$$ Spend
Database Archiving Reduces Risk

* Compliance Rules Better Accommodated Through Classification
  - Simplifies enforcing data retention & archiving policies
  - Archiving to an online active archive keeps data online and available to the native application while providing archive benefits

* Archiving Reduces Technology Obsolescence Risk
  - Active Archiving requires keeping underlying technologies current
  - Encapsulated archiving maintains application independence

* Archiving Keeps Data Volumes Manageable
  - Improves recoverability & application performance
  - Lowers TCO
Please send any questions or comments on this presentation to the SNIA: trackdatamgmt@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

SNIA Education Committee

Michael Fishman
Jeff Porter
Julie Lockner

How to get involved

- Find a passion
- Join a committee
- Gain knowledge & influence
- Make a difference

www.snia-dmf.org