



Education

Storage Considerations and Requirements for Database Archiving

Julie Lockner, Solix Technologies, Inc.

- Introduction to Database Archiving
- Database Archive Infrastructure Architecture Considerations
- Getting Started

About SNIA and the DMF

About the Storage Networking Industry Association (SNIA)

- SNIA's primary goal is to ensure that storage networks become complete and trusted solutions across the IT community
- For additional information about SNIA see www.snia.org

About the SNIA Data Management Forum (DMF)

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

DMF		
Data Protection Initiative (DPI)	Information Lifecycle Management Initiative (ILMI)	Long term Archive and Compliance Storage Initiative (LTACSI)
Defining new approaches and best practices for data protection and recovery	Developing, teaching and promoting ILM practices, implementation methods, and benefits	Addressing challenges in developing, securing, and retaining long-term digital archives

Databases Grow, Most Data is Inactive

- What percentage of your IT Budget is spent on operations and maintenance of database applications?
 - ◆ Include the multiple copies needed for production support, test & dev
 - ◆ Include daily Incremental & weekly full backups of PROD and Copies
 - ◆ MetaGroup analyst, Charlie Garry, estimated 70%

- Annual data growth is piling on more static data
 - ◆ Growth is fueled by upgrades, acquisitions, and new applications
 - ◆ Upgrading Applications can increase database size **40-60%**
 - ◆ Lack of data retention policies allows uncontrolled DB growth, all data is retained!

- After 2 years in production, how much of your data is **INACTIVE** or **AGED** – meaning “read-only” or infrequently accessed
 - ◆ How often do users actually access inactive or aged data?
 - ◆ Are classification policies in place to determine active vs inactive?

DB Data Most at Risk for Long-Term Readability

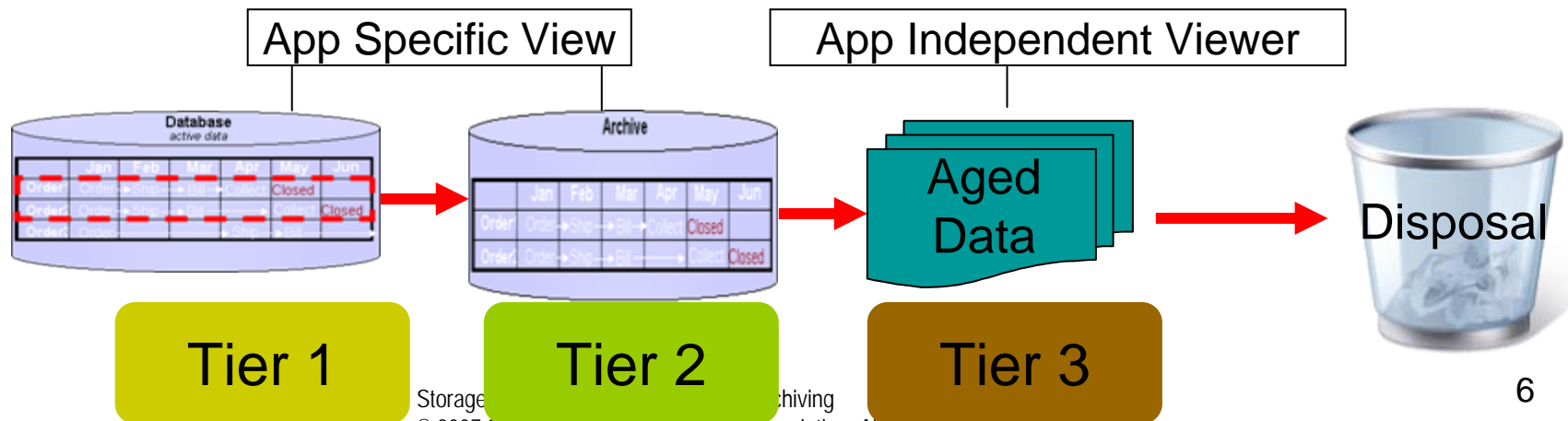
According to SNIA's 100 Year Archive Task Force Survey

“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>

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

Sample Database Archiving Configuration



Database Archiving Concepts

➤ Database Archiving or “Tiering” Characteristics

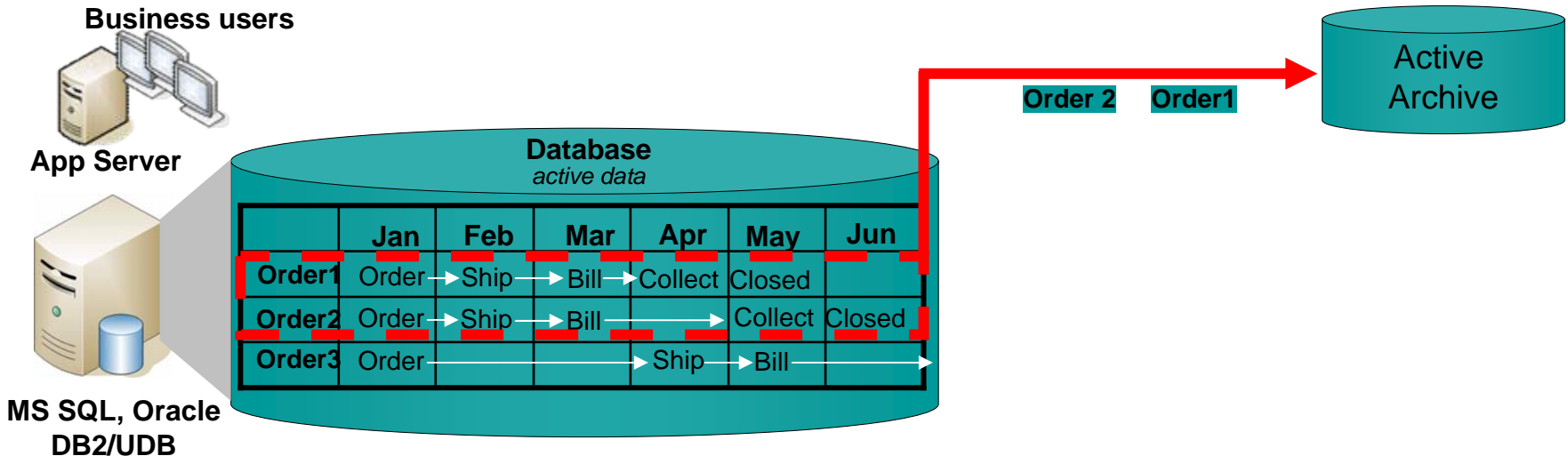
- ◆ Data is classified by a policy
 - i.e. age of data, application business logic
- ◆ Database data is moved to an online or offline archive
 - I.e. separate database, database partition, file, etc.
- ◆ Accessing Archived Database Data
 - Access requirements vary by regulation and corporate governance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

➤ Reports Archiving

- ◆ Database data is written to a report and the report is archived

➤ 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 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**

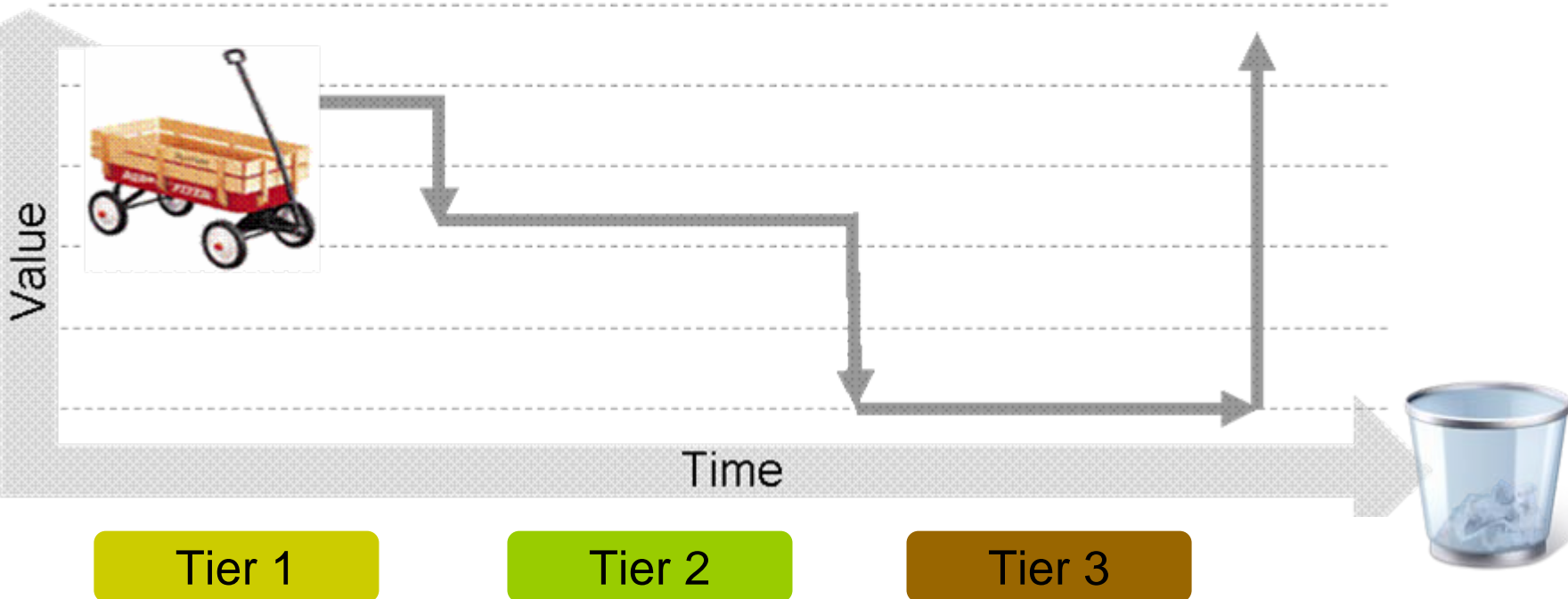
Data Classification Example: Sales Order

New Order

Product Ships

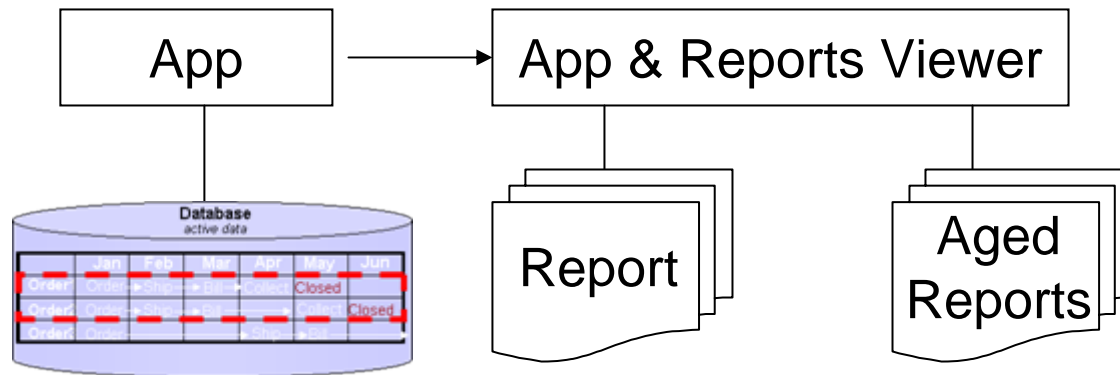
Warranty

Quality Recall



Reports Archiving is a Variation of DB Archiving

- Reports Archiving writes data to a file that can then archived



Database Archive Storage Considerations **SNIA**

- ◆ 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

- ▶ 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 SNIA

➤ 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

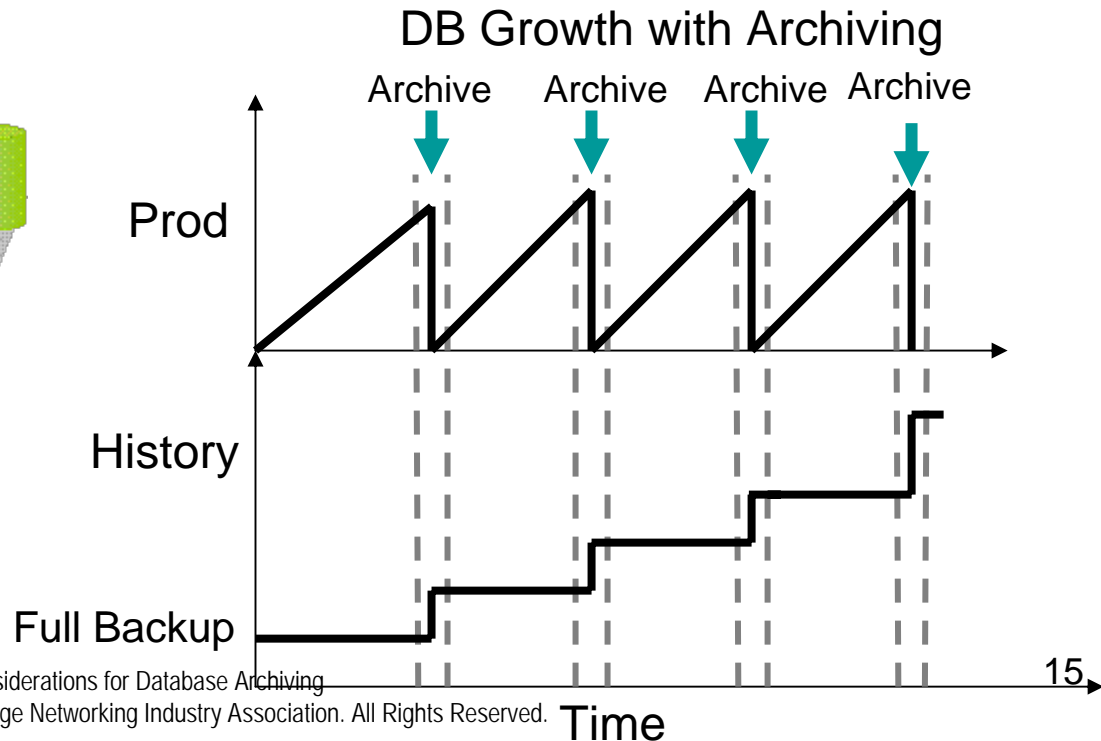
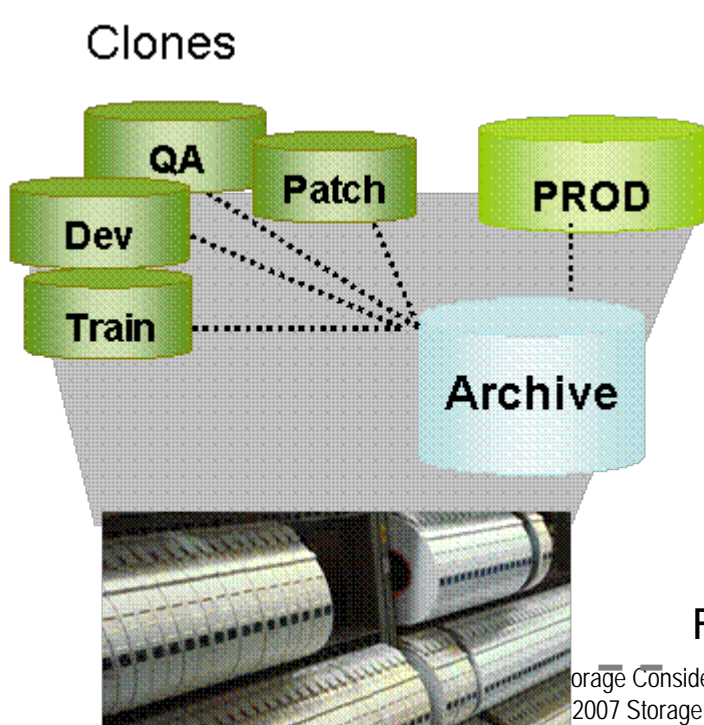
- ◆ Recovery of the production database is significantly reduced
- ◆ If the chosen DB archiving solution executes at the RDBMS or application tier, all network or block level D/R strategies should not be impacted
- ◆ Need to consider SLA's for production and the archive

➤ Replication

- ◆ Check if replicated production databases may or may not need access to the archive

Database Archiving & Backup / Restore

- 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

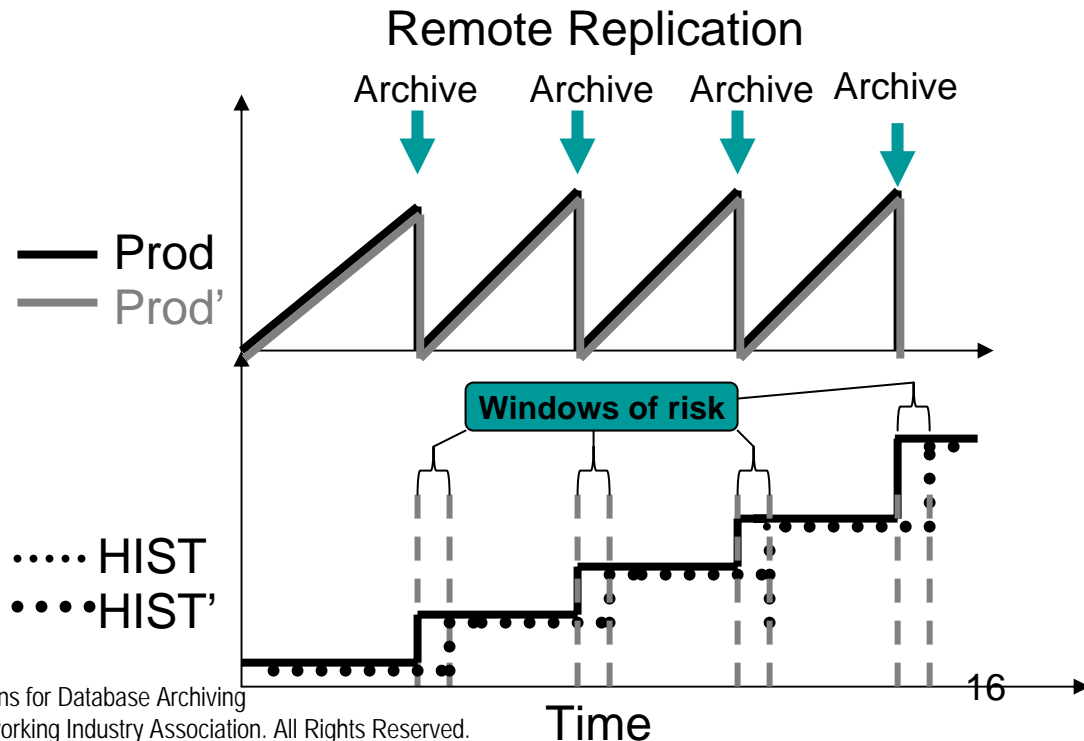
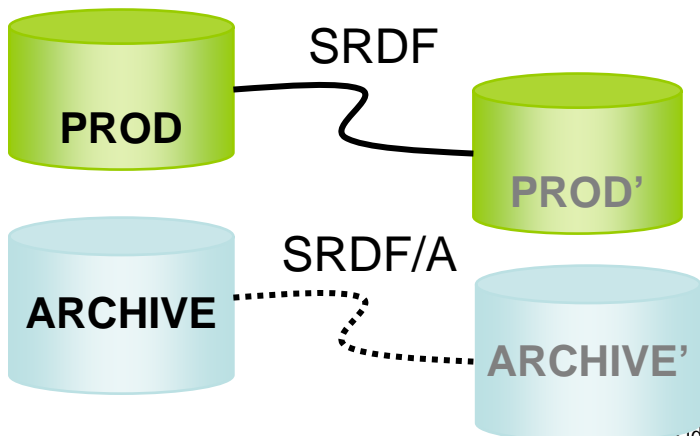


Database Archiving & D/R

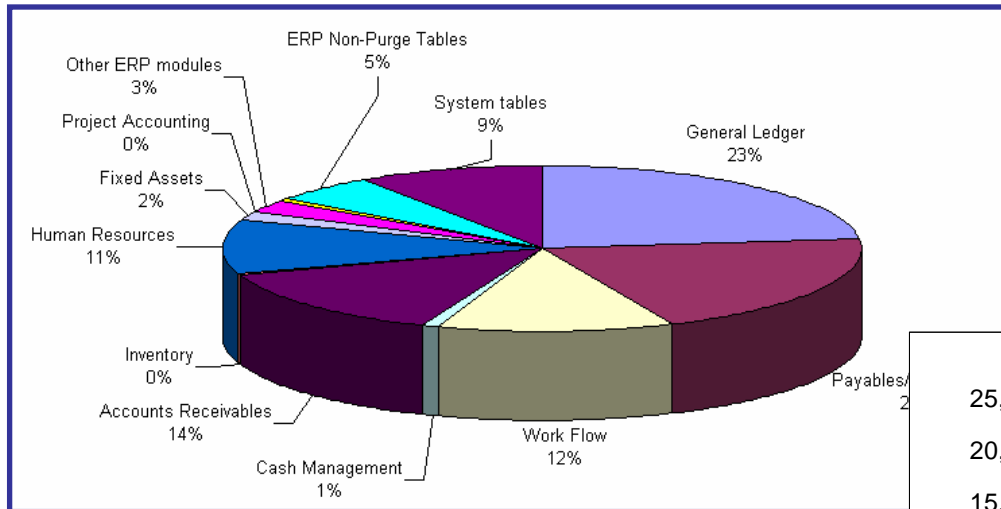
- Make sure Archiving is considered in the D/R strategy
 - ◆ Need to validate the archiving solution does not impact the D/R solution

- 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

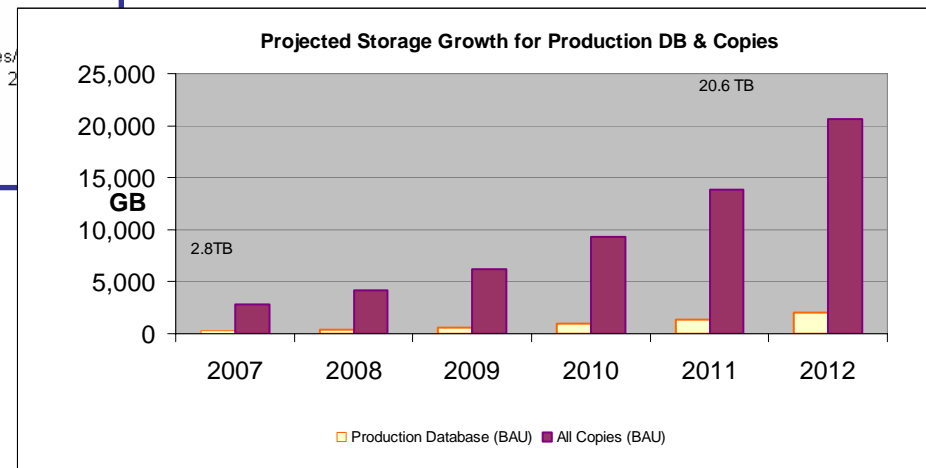


Database Classification Assessments



Distribution reports show data volumes by business process

Growth charts illustrate forecast requirements for storage



Archive Potential Impact by Application

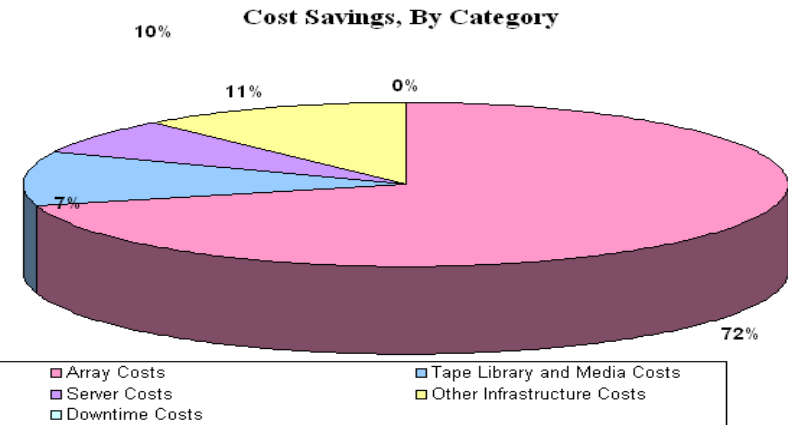
Application	Retention Policy (months)	Data Volume Before Archiving (GB)	Data Volume Eligible for Archive (GB)	Data Volume After Archiving (GB)
General Ledger	24	54	15.84	38.16
Accounts Payables	12	47.93	13.14	34.79
Workflow	1	28.42	4.38	24.04
Content Management	12	1.77	0.77	1
Accounts Receivables	12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

<u>Before</u>	<u>After</u>	=	<u>Savings</u>
236 GB	179 GB		56 GB

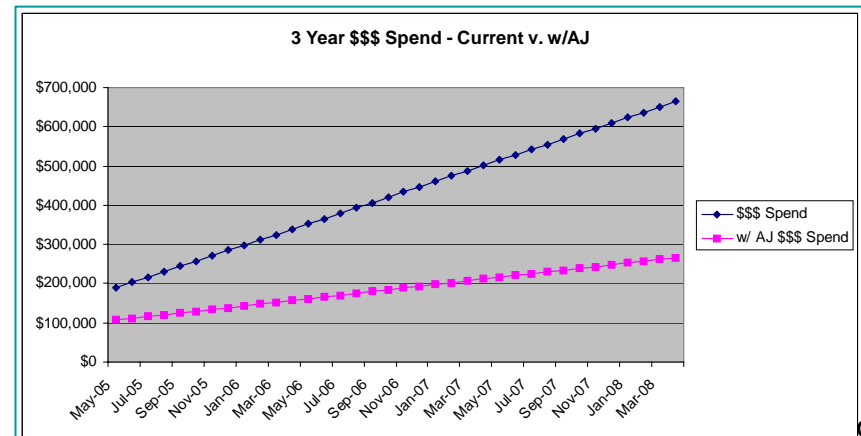
Archiving Can Be Justified with Cost Savings

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



- **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