Lower Maintenance Costs and Sustain System Performance with the SAP NetWeaver BI Data Aging Toolkit

Bryan Katis
SAP Labs
Learning Objectives

As a result of this session, you will be able to:

- Explain data aging and Information Lifecycle Management (ILM)
- Understand how SAP NetWeaver BI supports ILM
- Know the most important aspects of implementing a data aging strategy
# Data Aging and Information Lifecycle Management

<table>
<thead>
<tr>
<th>Motivation to Develop a Data Aging Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Aging Scenarios</td>
</tr>
<tr>
<td>Near-line Storage and SAP NetWeaver 2004s BI</td>
</tr>
<tr>
<td>Data Archiving and SAP NetWeaver 2004s BI</td>
</tr>
</tbody>
</table>

## Summary
What is Data Aging?

Data warehousing is a very powerful concept for creating a unified and consistent view of the business.

In a data warehousing environment, it is typical that:

- Data is amassed and analyzed at an increasing rate.
- As time progresses, companies face the dilemma of storing more and more historical data.
- Over time, data tends to lose its “day-to-day” relevance and is therefore accessed less frequently.
- The costs associated with maintaining historical data are high.

Data aging is a strategy for managing data over time, balancing data access requirements with TCO.

Each data aging strategy is uniquely determined by the customer’s data and the business value of accessing the data.

The SAP BI solution provides alternatives for the typical “cost vs. business data availability” conundrum.
Challenges Facing the Information Lifecycle

Data growth
- Emails, attachments, Web sites, audio and video content, voice recordings ...
- Constantly increasing data volumes in BI

Direct access capabilities
- Predictable residence time for ERP data
- Long-term direct accessibility appreciated, ad-hoc analysis needs

Legal requirements
- SEC, FDA, HIPPA, SOA, GDPR, Basel II for ERP data

Data value during lifecycle
- Constantly decreasing in ERP environments
- More long lasting in BI environments

Costs
- Personal costs, technology costs, process costs

Technological innovations
- ATA disks, blue laser, etc.
- Write-once file system, NLS, etc.

The challenges cannot only be addressed by purchase of additional memory. An effective administration of the data is necessary.
ILM is not only a product

ILM is not only software

Everyone’s talking about ILM

ILM is not only a tool

ILM is a combination of processes and technologies with a goal of providing the right information at the right time in the right place with the lowest possible costs

ILM is a storage management concept that actively manages all information objects during their entire lifecycle
Archiving Requirements of the Various Layers in a BI Environment

- **BI**
  - **Operational Reporting**
  - **Strategic Analysis**
  - **Data Mining**
  - **Business Planning**

- **Persistent BI analysis horizon**
  - High-volume data sets
  - Direct availability required
  - Data archiving process

- **Persistent BI inbound layer**
  - High-volume, highly granular data
  - Direct availability not required
  - Data archiving processes

- **Description of extraction formats**
  - DART (Data Retention Tool)
  - Service API for BI Extraction

- **Classic archiving**
  - Archiving objects
  - Legal storage regulations
  - Cost reduction

- **ERP**
  - **Extracts**
  - **Relationships**
  - **Entities**

© SAP AG 2006, 8
Information Lifecycle Management at SAP – One View

Analyze business objects, determine data categories and layers, define attributes, identify parameters

Define policy for attributes and categories such as retention period, legal compliance, data extracts

Apply policies to business objects and data categories

Realize

Automation, archiving, migration, deletion, accessibility
Data Aging and Information Lifecycle Management

Motivation to Develop a Data Aging Strategy

Data Aging Scenarios

Near-line Storage and SAP NetWeaver 2004s BI

Data Archiving and SAP NetWeaver 2004s BI

Summary
Typical Data Warehouse Problems

End-User Challenges

- Making timely, informed business decisions
  - Users cannot wait for historical data to be restored
  - Transparent access to data for regular reporting and ad-hoc analysis

IT Management Challenges

- Meeting end-user data demand while managing cost
  - High costs of adding/managing online disk storage
  - High costs of backup and recovery – especially when data is accessed infrequently
  - Data protection and availability
Meta Group

“70% of all corporate data is currently located in databases”

Gartner Group

“Databases with multiple terabytes are already a reality – in the near future these will have hundreds of TBs”

“Business Applications – such as those from SAP – play a significant role in these growth in memory needs”

“Growth rate is 64%”
“The costs for data media don’t even make up a quarter of memory costs” (Giga Information Group)

“The administrative expense for 1 terabyte of memory is five to seven times as high as the memory cost itself” (Dataquest/Gartner)
Motivation for a Data Aging Strategy: Benefits

Performance
- Faster data load times
- Faster query execution times

Cost
- Storage costs: High availability, high IO disks, etc.
- Resource and Administration overhead
  - System: CPU, Memory, etc.
  - Headcount: Number of full-time employees, etc.
- Control of system growth

Availability
- Data availability – faster rollups, change runs, etc.
- System availability – less downtime for backups, upgrades, etc.
How to Avoid High Data Volumes in a DW Environment?

Is data still needed?

Yes

Can the data be aggregated?

Yes

Aggregate data!

No

Can the data be archived?

Yes

Archive data!

No

Data remains in the database!

We need to address the no longer needed AND the maybe scenario!
Data Aging Strategy Implementation – Initial Steps

Data aging is a strategy for managing data over time to balance data access requirements with TCO

- Each data aging strategy is uniquely determined by the customer’s data and the business value of accessing the data

Which tools should I consider using?

<table>
<thead>
<tr>
<th></th>
<th>Online Database Storage</th>
<th>Near-line Storage</th>
<th>BW data archiving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequently read /updated data</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrequently read data</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Very rarely read data</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Benefits of a Fundamental ILM Strategy for BI

1. **Increase Volume**
   - Manage and use even larger amounts of information more effectively
   - Information available for any time frame for ad-hoc analyses and rebuilds

2. **Reduce Resource Consumption**
   - Reduction of hardware costs for hard drive on the BI side
   - Main memory and CPU as well as costs for system administration

3. **Increase Availability**
   - Quicker, simpler software and release management in BI
   - Reduced backup and recovery times
   - Intelligent data access

4. **Optimize Performance**
   - Speed up loading processes in SAP NetWeaver BI
   - SAP NetWeaver BI Query answer times in the dialog
<table>
<thead>
<tr>
<th>Data Aging and Information Lifecycle Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation to develop a Data Aging Strategy</td>
</tr>
<tr>
<td>Data Aging Scenarios</td>
</tr>
<tr>
<td>Near-line Storage and SAP NetWeaver 2004s BI</td>
</tr>
<tr>
<td>Data Archiving and SAP NetWeaver 2004s BI</td>
</tr>
<tr>
<td>Summary</td>
</tr>
</tbody>
</table>
Include data aging early on in the blueprinting phase

- Data retention should be determined during requirements gathering
  - Determine retention for all data layers including transactional and master data
  - Transactional data should be evaluated from both a DataSource and an individual DataTarget perspective
  - Evaluate legal reporting requirements
  - Evaluate regulatory reporting and retention requirements
  - Future business data analysis requirements should also be considered

- Observation: SAP typically observes data retention of three to five years in InfoCubes and DataStore objects

- Data model sizing should be included in overall BW capacity planning
  - Data volume and growth should be determined
  - Data “change” activity profiles should also be determined
  - Frequency of data deletion and data updates should be included
  - BW capacity plans and TCO should be revisited regularly!
Data warehouse data retention should be integrated with your OLTP system’s archiving plans

- Data warehouses normally retain more historical data than operational transaction processing systems
- Do you need to archive BW data if your ERP system also archives it?
- Does your OLTP archiving strategy limit future data warehouse developments?

Info: SAP R/3 archives can be read from BW via SAP Archive Information System functionality

Keys to a successful data aging strategy development:

- Define data retention for all data
- Profile your data activity and access
- Determine the capacity impact of ongoing data storage
- If possible, determine a cost model for your data storage/access
- Choose and implement technology that does not limit your business
ILM in SAP NetWeaver BI

1. Online
   - Data persistent in the database
   - Data modeling aspects important
   - Use multiple layers to control data growth
   - Frequent cleanup necessary

2. Near-line
   - Near-line Storage (NLS)
   - ADK (Archiving Development Kit) and non-ADK-based
   - Set up proper archiving policy
   - Transparent access for reporting

3. Offline
   - Classic archiving as known in SAP BW 3.x
   - Very cheap storage medium can be used
   - ADK-based
   - No access for reporting
Online: Data Volume Minimization Strategies

**PSA deletion:** Temporary inbound data load layer, exists for recovery or reload purposes

- Retain daily loads one week, monthly retain 60 days
- Customer example: DB growing at 450 GB/month without PSA deletion, 150 GB/month with deletion

**ODS object data deletion**

- Optimize ODS Object activation performance

**Change log deletion:** It’s like a PSA

- When data consolidated in subsequent data target
- When no delta at all is provided by this ODS

**Active table**

- When data is consolidated in active table
**Data consolidation:** Eliminate redundant data

- Merge data from similar data marts
- Optimize aggregates
- Aggressively archive and delete from InfoCube and ODS object layers
- Summarize at the InfoCube and aggregate level (RRI)

**Remote access:** Use RRI, Virtual InfoProviders, UDI, etc.: Operational processing possibly might remain in OLTP system
Use Offline Archiving with MultiProvider

MultiProviders are used ...

- To combine actual data with reloaded data from archive
- Defining data target-independent queries

Data has to be reloaded before availability in reporting
MultiProviders are used for ...

- Defining data target-independent queries

BEx or Web Reporting

Multi Provider

2006

NLS Archive

2005

2004

2003
MultiProviders are used for ...

- Splitting data from one business scenario into separate basic InfoCubes (“logical partitioning”)
  - Possible partitioning criteria: Year, plan/actual, regions, business area
  - Use MultiProviders to cut large amounts of data into chunks
Benefits

- Single InfoProviders smaller, less complex, and less sparsely filled than one big InfoProvider
- (Parallel) data load into individual InfoProviders
- Better performance in administration on basic InfoProviders
- In most cases, queries are split automatically and distributed to InfoProviders (for more details on parallel execution see SAP Service Marketplace Note 629541)
- Dedicated query split for archive-related data portion as well
- Central interface for querying even if underlying data models change

Disadvantages

- Administration (with aggregates)
- Data has to be reloaded before available in reporting
- Increasing number of InfoProviders to be administrated
- Remodeling at year end (are you using a new InfoCube in the new year?)
- New process design at year end (unload, new archive object)
Single Basic InfoCube Approach with NLS

Data Archiving Process
- Archive Type
  - Near-line storage
- Selection Schema
  - Time slices
    (Archive everything older than 2 years)

Schedule: January, 01

BI accelerator

Year Change

Archive 2002 2001

Archive 2003 2002 2001

Zero Administration
Data Aging and Information Lifecycle Management
Motivation to Develop a Data Aging Strategy
Data Aging Scenarios
Near-line Storage and SAP NetWeaver 2004s BI
Data Archiving and SAP NetWeaver 2004s BI
Summary
SAP BI’s data aging toolkit:

- **Near-line data storage**
  - The relocation of data from online storage to a secondary storage system
  - Secondary storage is typically utilizing more cost-effective online storage (e.g., IDE disks) or removable media (e.g., tape, optical)
  - Seamless online BW Query access to data via direct SQL access

- **BW data archiving**
  - The relocation of data from online storage to offline storage
  - No direct access to archived data for online querying
  - Integrated with SAP’s Archive Development Kit

<table>
<thead>
<tr>
<th></th>
<th>Online database storage</th>
<th>Near-line storage</th>
<th>BW data archiving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequently read/updated data</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrequently read data</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Very rarely read data</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Where Is Archiving and Near-line Storage Applicable?

Archiving (BW 3.x)
- For analysis, archived data must be reloaded first again into the BW database
- Reduction in costs of data retention on alternative media

Near-line storage (> BW 3.x*)
- Direct access to data in alternative storage media for queries
- Performance and data retention costs to access aged data can be minimized

* Pilot project possible for BW 3.x
Near-line Storage (NLS): Basics

Access Frequency

Near-line Storage/Archives

Online Memory

Direct SQL Access

Memory Management

Offline Archive

Costs

Performance

Days Months Years

Costs

Performance
Modern NLS Memory Management System

Performance Characteristics

- Performance- and cost-optimized management to keep data in multi-layer, transparent storage systems
- Combined storage systems of disk, tape, and optical storage media in different variations
- Direct row-oriented data access – also to compress archive data using all types of storage media
- Access strategies and aging pattern for logical grouping of data
- Dynamic migration on cost-effective storage media for data files with regressive access frequency
- Automated process for backup, shadowing, mirroring, recovery, etc.
Archiving (SAP BW 3.x)

- ADK-based archiving solution for InfoCubes and ODS objects
- Cost reduction due to storing data on alternative storage media
- ADK indexes for BI Query access ineffective, index structures too inflexible

- Archived data must be reloaded into the SAP NetWeaver BI database for analysis purposes
Near-line Storage (SAP NetWeaver 2004s BI)

- SAP NetWeaver BI analyses have direct access to archived data in various storage media
- Availability of historic data while reducing costs
- Physical decoupling of frequently, less frequently, or rarely used data
- Reloading of data into the InfoCube or DataStore object only necessary in exceptional cases
The Near-line Storage Solution for SAP NetWeaver BI

Separation of frequently used (kept in database) and rarely used (in NLS) data

NLS support for InfoCubes and DataStore objects

Transparent access to “non-archived” and “archived” data for queries

Intelligent data access
  ■ Analysis/feedback data selection
  ■ High-level index in SAP NetWeaver BI
  ■ Low-level index in NLS

Open interface for certified partners

Present development partners
  ■ PBS Software – CBW®
  ■ FileTek – StorHouse®
  ■ OuterBay – LiveArchive®
  ■ SAND-Technologies – Searchable Archive®
SAP NetWeaver 2004s BI – Generic NLS Interface

Near-line Storage   Partner Solution

Data Flow
Control Flow

Analysis

Combination

Distribution and Sending

High-Level Index

DB Interface

Archiving/ Restoring

Near-line Storage Adapter

BI Database

Low-Level Index

Data Manager

InfoCube/DataStore with NLS

Near-line Storage

Robot.

Network-Attached Storage or Cost-Effective Data Medium

Tape Library

Optical Libraries

© SAP AG 2006, 38
SAP NetWeaver 2004s BI – Generic NLS Interface (cont.)

- InfoCube/DataStore with NLS
- DB Interface
- BI Database
- Analysis
- Data Flow
- Control Flow
- Distribution and Sending
- Near-line Storage Adapter
- Archiving/Restoring
- Near-line Storage
- Optical Libraries
- Network-Attached Storage or Cost-Effective Data Medium
- Low-Level Index
- Data Manager
- High-Level Index
- Robot Tape Library
- Partner Solution

© SAP AG 2006, 39
SAP NetWeaver 2004s BI: NLS-Based Archiving

Data Archiving Process

- **Archive Type**
  - Offline, ADK-based
  - Near-line, ADK-based
  - Near-line storage, standalone

- **Selection Schema**
  - Time slices (Relative archiving periods, delta oriented, compressed data only)
  - Request-oriented
  - Free selection criteria

Query Properties
Near-line storage should be read as well

Scheduling via Process Chain
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Aging and Information Lifecycle Management</td>
</tr>
<tr>
<td>Motivation to Develop a Data Aging Strategy</td>
</tr>
<tr>
<td>Data Aging Scenarios</td>
</tr>
<tr>
<td>Near-line Storage and SAP NetWeaver 2004s BI</td>
</tr>
<tr>
<td><strong>Data Archiving</strong></td>
</tr>
<tr>
<td>Summary</td>
</tr>
</tbody>
</table>
BW Data Archiving Technology

SAP BW 3.0B/3.1 Content provides the following functionality:

- BW objects data archiving
  - Selection of data based on any criteria (time- or data-specific)
  - Automatically scheduled on a periodic basis
- Data deletion (without archiving)
- Restoring (reloading) of archived data
- Consistent archiving processes for BW objects
  - “Future proof” data archives for reuse

Archive objects can be defined for the following DataTargets:

- InfoCubes
- ODS objects
- It is planned to include PSA and master data archiving in future BW releases
- IDoc archiving has been available since BW 2.0B
SAP BI data archiving is built on the Archive Development Kit (ADK). The ADK is part of the underlying WebAS technology on which mySAP applications such as SAP BW, R/3, CRM, etc., are built.

ADK Features:
- Open to CSP archiving and storage systems (example: IXOS)
- Widely used by mySAP application components
- Handles structural changes
- Platform independence (code-page, number format)
- Data compression
- Statistics and indexing support
- Administration of archiving via transaction SARA
- Provide logging function for archive administration
- Supports indexing with Archive Information System (SAP AS) (BW does not use this functionality)
- Archiving objects are transport-enabled
BW Data Archiving Technology: Process Overview

Archiving Object

Write

Delete

Read

Scheduled

Verification

BW Repository

activates

generates

InfoCube

ODS object

DataManager

Archive Administration (SARA)

Datamart Extractor

File system, CMS, HSM

ADK

© SAP AG 2006
Archiving object meta data is maintained at the DataTarget level

Generated on DataTarget activation
- Archiving object name is generated from DataTarget name

Administration and execution
- Accessed via link from Admin Workbench or transaction code SARA
- Maintain variants for archive execution
  - Data selection criteria, Archive processing and scheduling options
- Scheduling and monitoring of archiving sessions
- Management of archive runs
- Information systems/statistics

1:1 Relationship

© SAP AG 2006
BW Data Archiving Technology: Archiving Process

Archiving Process:

1. Select DataTarget to be archived
2. Configure and generate archive object
3. Configure archiving process
4. Test archiving process
5. Archive data
6. Validate archive? (Yes/No)
7. Move archive data to storage subsystem (optional)
8. Delete Data
9. Successful? (Yes/No)
10. Test archiving process
11. Configure archiving process

© SAP AG 2006, 47
BW Data Archiving Technology: Object Definition

InfoCube Example

Populated after activation!
Time Slots option:
Time characteristics to be used for selection criteria for archiving. “Relative time” archiving can be achieved using Time Slots.

Protected Archived Areas option:
Ability to lock archived “time selections” from further data load activations. Archive data would have to be reloaded first before new data is activated. Note: Only available for Time Slots archiving of ODS objects.

Record selection option:
Possible to choose fields used for selection criteria for archiving. All DataTarget fields are available to be chosen.

ODS Object Examples
BW Data Archiving Technology: Object Definition (cont.)

Archive will create a new file when max file size is reached.

Number of data records written to a file before a new file is created.

Defines field layout for archive file.

Available (unassigned) fields for record.
Logical path for archive file defined via transaction FILE

Only required if you have an archive storage subsystem (e.g., IXOS, Filenet, Storhouse, etc.)

Determines how the subsequent data deletion job will be scheduled and executed
**BW Data Archiving Technology: Menu Access**

In the context of SAP BW Data Archiving, the image illustrates the menu access for managing data targets. Specifically, it shows how to access the Archive Administration feature under the 'Data Targets' menu.

The screenshot displays the Administrator Workbench with the Modeling view selected. Within the Data targets section, the LH Archive Load menu item is highlighted, indicating a possible menu path or configuration option for data archiving.

This image is relevant for users who need to manage data archiving policies and procedures within their SAP BW environment, emphasizing the importance of menu navigation for accessing specific tools and settings.
“Write to archive file” and “Data deletion” are separated

Archived data files are read for verification
- Completeness
- Accessibility

One-step deletion if all session files are verified (different from OLTP archiving)
- Using selection criteria of archiving session
- Note: NOT the contents of the archive file like SAP R/3 archives
- ODS objects: Selective deletion from active table
- InfoCube: Selective deletion from fact tables
- Optimized deletion strategies

Adjust or rebuild aggregates – determined by data volume

Note: DataTarget is locked for data loads until data is deleted or the archive run is invalidated
The write process archiving parameters

- Relative or absolute time-slot archiving

**Relative Time Restriction**
- Only Sentences Older Than: 3 Months
- Only Complete: Fiscal Years
- Exclude Archived Areas
- Fiscal year/period: 001 / 2001

**Absolute Time Restriction**
- Fiscal year variant: K4
- Fiscal year/period
  - from: [ ]
  - to: [ ]

**Flow control**
- Test Run
- Archive
- Archive and Delete

**Archiving Session Note**
- Testing archive of Fiscal Periods!
Administration of the Archiving Process

The write process archiving parameters

- Selective criteria archiving (non-time slot)

<table>
<thead>
<tr>
<th>Restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiscal year/period</td>
</tr>
<tr>
<td>Posting period</td>
</tr>
<tr>
<td>Fiscal year variant</td>
</tr>
<tr>
<td>Fiscal year</td>
</tr>
<tr>
<td>Statist. curr</td>
</tr>
<tr>
<td>Base Unit</td>
</tr>
<tr>
<td>Country</td>
</tr>
<tr>
<td>Region</td>
</tr>
<tr>
<td>EAN/JPC</td>
</tr>
<tr>
<td>Order</td>
</tr>
<tr>
<td>Distribution Channel</td>
</tr>
<tr>
<td>Status indicator</td>
</tr>
<tr>
<td>Ship to Customer</td>
</tr>
</tbody>
</table>

Restrictions:
- Ability to restrict archiving via selection criteria for all fields selected during archive object definition; individual values, ranges, exclude values, etc.

Flow control:
- Test Run
- Archive
- Archive and Delete

Archiving Session Note: Test of archiving with selective criteria
InfoPackage is extended by an option “archive selection”
- Selection options for available archive sessions and files
- Only full extraction supported
- Archive files are scanned with selection criteria of the request

Reload to original DataTarget is possible but not recommended

Reload recommendation:
- Extract to a copy of original DataTarget instead
- Use MultiProvider to combine remaining data with reloaded data
Data Aging and Information Lifecycle Management

Motivation to Develop a Data Aging Strategy

Data Aging Scenarios

Near-line Storage and SAP NetWeaver 2004s BI

Data Archiving and SAP NetWeaver 2004s BI

Summary
Summary

1. The “Healthy” System
   Don’t start thinking about data archiving when your system is about to crash!

2. Timely Planning
   Proactive preparation for sustainable system performance

3. Interdisciplinary Process
   Data archiving requires a large amount of coordination between IT and those who are responsible for applications
The central hub for the SAP technology community

- Everyone can connect, contribute and collaborate—consultants, administrators and developers
- Focus around SAP NetWeaver and SAP xApps

High quality of technical resources

- Articles, how-to guides, weblogs, collaborative areas, discussion forums and downloads, toolkits and code-samples

A collaboration platform, not a one-way street

- SAP experts from customers, partners and SAP

SDN is powered by SAP NetWeaver™

- Built on the SAP Enterprise Portal
- Featuring collaboration capabilities of SAP Knowledge Management
7 Key Points to Take Home

TCO can be dramatically decreased for your EDW by implementing a data aging strategy

Successful data warehouse implementations are cost-effective solutions that continue to drive the business

It is imperative that you consider data retention early on during the blueprinting phase of your project

Data retention requirements should be routinely assessed and adjusted based on business demands and costs

SAP NetWeaver BI provides a rich set of tools for implementing your data-aging strategy

SAP NetWeaver BI’s Near-line Storage solution provides a compelling, cost-effective alternative to online database storage for less frequently accessed data

Data archiving is a very flexible and powerful tool for managing infrequently accessed data in your data warehouse
Questions?

Q&A

Bryan Katis

bryan.katis@sap.com