

# **Data Archiving** - Solutions, Challenges, Considerations & 3rd Party Tools

**SOAIS**

Putting Customer First

# Agenda

---

- ❖ Introduction
  - ❖ Case Study
  - ❖ Setup
  - ❖ Data Archiving Process
  - ❖ Limitations/Challenges
  - ❖ Alternate Methods of Data Archiving
  - ❖ Q&A
-

# Introduction

---

- ❖ Data Archiving
- ❖ Why Data Archiving ?
- ❖ Factors to be considered while performing Data Archiving
- ❖ PeopleSoft Data Archiving
- ❖ Advantages of PeopleSoft Data Archiving

# Data Archiving

- ❖ Data Archiving is the process of <sup>1</sup> removing <sup>2</sup> selected data records from operational databases that are not expected to be referenced again and <sup>3</sup> storing them in an archive data store <sup>4</sup> where they can be retrieved if needed.

# Data Archiving

- 1. Removing:** Deleting Data from the operational database. Here, when data moves into the archive state, query and access is no longer anticipated to be required.
- 2. Selecting:** Selectively choose particular pieces of related data for archival... not the whole database, not an entire table or segment.
- 3. Storing:** Archived data is separate and independent from the production systems from which it was moved.
- 4. Retrieving:** The purpose may be external, in the form of a **lawsuit** or to support a **governmental regulation**; or the purpose may be internal, in the form of a new business practice or requirement. At any rate, the data needs to be readily accessible in a reasonable timeframe without requiring a lot of manual manipulation.

# Why Data Archiving ?

- ❖ Companies invest **millions of dollars** each year in maintaining and upgrading business critical applications that rely on complex relational databases.
- ❖ These databases collect **increasing amounts of data** for business operations and decision-making.
- ❖ Overloaded databases degrade performance and limit the availability of the comprehensive capabilities these applications were designed to deliver.
- ❖ Ironically, most of this data is stored online in production databases but is rarely accessed. The solution of this problem has led to the concept of **Data Archiving**.

## Factors to be considered while performing Data Archiving

---

- ❖ **Type of Data:** Determine the kind of data your organization generates.
- ❖ **Volume of Data:** Next, determine how much data you have to store. This will become important when you evaluate the storage capacity of different solutions.
- ❖ **Mobility Needs:** Once you know the type and volume of data you want to store, ask yourself how you plan to use this data in the future..
- ❖ **Inventory Control:** In order to develop a comprehensive and effective archiving and backup plan, you need to develop an inventory of your IT resources.

## Factors to be considered while performing Data Archiving

- ❖ **Flexibility Needs:** Beyond the idea of data mobility is the issue of flexibility. Do you need continuous access to your data once it's stored? Are your data storage needs short term, long term or a combination of both? The best solution for storage is usually a combination of products that address an organization's various needs.
- ❖ **Backup Schedule:** Your storage plan should include a schedule to carry out data backup on a regular, timely basis. Depending on the volume of data generated, location and the critical nature of the data, the schedule can be set to either a continuous daily, weekly or monthly timeline.



# PeopleSoft Data Archive Manager

---

- ❖ In any enterprise application, the ability to purge and archive transactional data is critical to data management. You need to have consistent methods to archive transactional data before your database increases to unmanageable sizes.

**PeopleSoft Data Archive Manager** provides an **integrated** and **consistent framework** for archiving data from PeopleSoft applications.

# Advantages of PS Data Archive Manager

- ❖ **Used for Business Analytics:-** Helping make business decisions on huge history data without impacting the core application.
- ❖ Time reduction in **Backup** and **recovery**.
- ❖ **Improves Performance:-** Both Online and Batch Process
- ❖ **Saves Cost** by helping having smaller database instances.
- ❖ Reduce high cost storage options in production systems through reduced online data volumes.
- ❖ **Increase productivity** among application users and application administrators.
- ❖ **Automating** Archiving and Purge processes
- ❖ **Selectively** Archive data that you no longer need.
- ❖ Provide **run time parameters**

# Agenda

---

- ❖ Introduction
  - ❖ **Case Study**
  - ❖ Setup
  - ❖ Data Archiving Process
  - ❖ Limitations/Challenges
  - ❖ Alternate Methods of Data Archiving
  - ❖ Q&A
-

# Case Study

---

- ❖ A Big Manufacturing giant has **Payroll for North America** implemented and pays around 45000 employees every month. Data keeps accumulating in their Core tables every time there is a calc causing serious performance issues on the database. There were more than 100000 time card entries provided into the system every week.

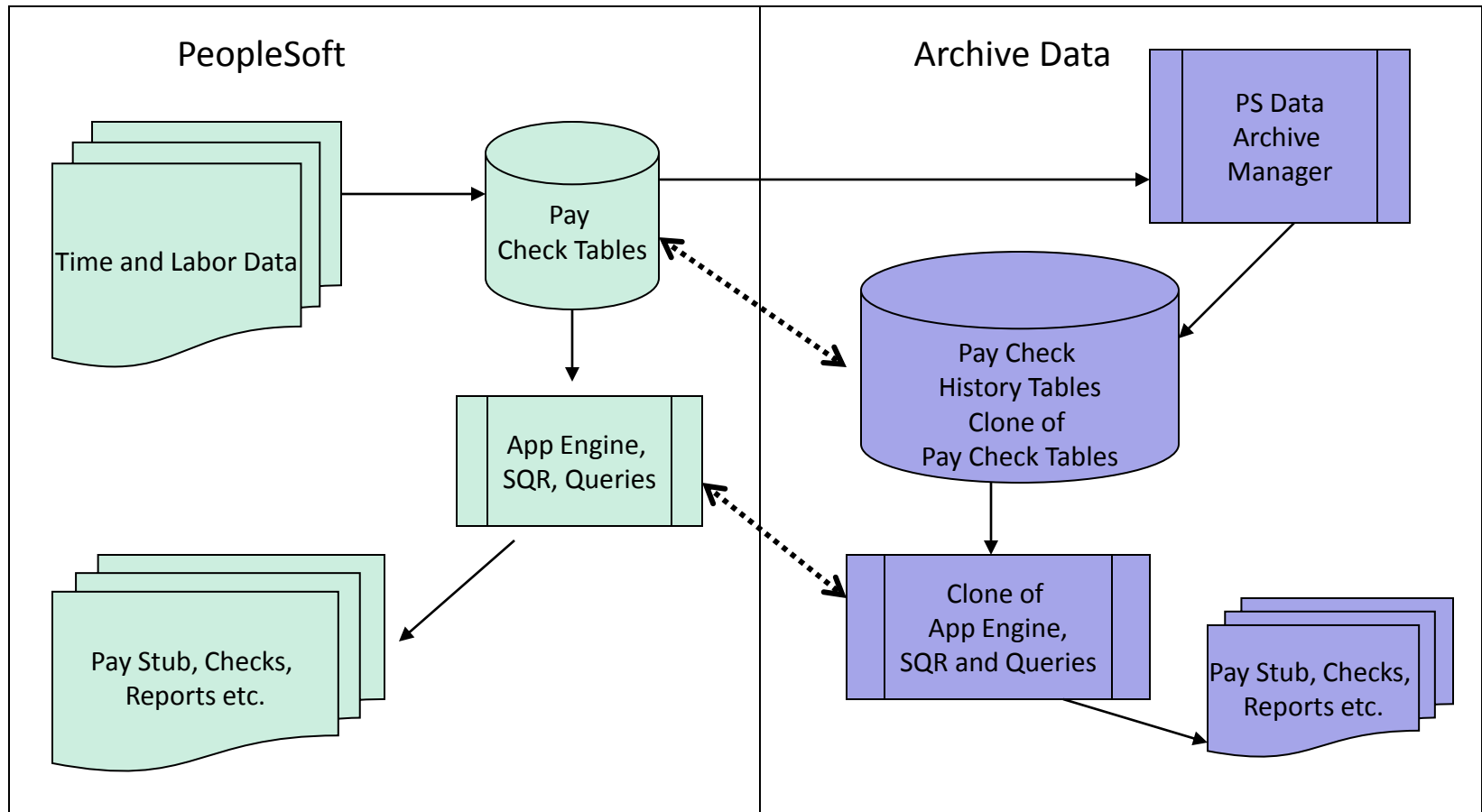
# Case Study

---

## ❖ Challenge:

- **Cobol processes** were taking a lot of time to execute and Database administrators had a tough time maintaining the databases. They had to frequently index the tables, constantly monitor for peak load, concurrent users, Backup and recovery was taking a long time etc.... They could not drop tables because of legal requirements and data had to be stored for more than 7 years.
- **Legal Requirements** – Some states have more than 7 years of legal **data retention** requirement for all Pay Check related details. So history data could not be deleted for a period of seven years.

# Case Study



# Agenda

---

- ❖ Introduction
  - ❖ Case Study
  - ❖ **Setup**
  - ❖ Data Archiving Process
  - ❖ Limitations/Challenges
  - ❖ Alternate Methods of Data Archiving
  - ❖ Q&A
-

# Setup

- ❖ Basic Terminologies
- ❖ Building the history table
  - The example in this presentation uses the **JOB** table which is used to record an employee's job history data such as actions taken, department, job code, location, and salary history..
  - Over time, the **JOB** table can grow quite large, however you probably want to keep all of the information it contains (for auditing purposes) which is why archiving provides the best solution.
  - One should first establish the business rules regarding the data that you want to archive before developing the archive process. In our example, the business rule will be to archive any data in **JOB** where the Employees are Terminated/Retired on or before 01-January-1960. This is a *very simple* business rule, and your rules will vary in complexity.



# Basic Terminologies

---

- ❖ **Archive object** which are the table(s) you want to archive.
- ❖ **Archive query** which is used to determine the criteria for the set of data (rows) to archive.
- ❖ **Archive template** to combine the archive objects and queries.
- ❖ **Archive data to history process/job** which archives data to history.
- ❖ **Restore query** to restore archived rows from history.
- ❖ **Archive auditing** to track what has been archived.

# Basic Terminologies

---

- ❖ The **base table** is the key transactional table you will be archiving. It should contain all the keys by which any other tables you are archiving can be reference if you are archiving more than one table.
- ❖ The **history table** is the table you will be storing the archived information.

# Building the history table

- ❖ The first step is to create the history table to store the archived data from the **JOB** transactional table.
  
- ❖ The history table must have exactly the same structure as the base table with the addition of the **PSARCHIVE\_SBR** sub-record to form the key structure. This sub-record includes the fields:
  - **PSARCH\_ID**
  - **PSARCH\_BATCHNUM**

# Building the history table

Num	Field Name	Type	Key	Ordr	Dir	CurC	Srch	List	Sys	Audt	Default
1	PSARCHIVE_SBR	SRec					No	No	No		
	PSARCH_ID	Char	Key	1	Asc		No	No	No		
	PSARCH_BATCHNUM	Nbr	Key	2	Asc		No	No	No		
2	EMPLID	Char	Key	3	Asc		Yes	Yes	No		PER_ORG_ASGN.
3	EMPL_RCD	Nbr	Key	4	Asc		Yes	Yes	No		
4	EFFDT	Date	Key	5	Desc		No	No	No		%date
5	EFFSEQ	Nbr	Key	6	Desc		No	No	No		
6	PER_ORG	Char	Alt		Asc		No	Yes	No		PER_ORG_ASGN.
7	DEPTID	Char	Alt		Asc		No	Yes	No		
8	JOBCODE	Char	Alt		Asc		No	Yes	No		
9	POSITION_NBR	Char	Alt		Asc		No	No	No		
10	SUPERVISOR_ID	Char	Alt		Asc		No	Yes	No		
11	HR_STATUS	Char					No	No	No		'A'
12	APPT_TYPE	Char					No	No	No		'0'
13	MAIN_APPT_NUM_JPN	Nbr					No	No	No		
14	POSITION_OVERRIDE	Char					No	No	No		'N'
15	POSN_CHANGE_RECORD	Char					No	No	No		'N'
16	EMPL_STATUS	Char					No	No	No		'A'
17	ACTION	Char					No	No	No		
18	ACTION_DT	Date					No	No	No		%date
19	ACTION REASON	Char					No	No	No		

- ❖ In this example, the history table will be called **JOB\_HST**.

# Agenda

---

- ❖ Introduction
  - ❖ Case Study
  - ❖ Setup
  - ❖ **Data Archiving Process**
  - ❖ Limitations/Challenges
  - ❖ Alternate Methods of Data Archiving
  - ❖ Q&A
-

# Data Archiving Process

---

## 1. Create the Archive Object

- ❖ Manager Archive Objects are used to define the Base tables and non base tables.
- ❖ Add a new archive object. The archive object in this case will just be the name of the **base table** - JOB. Enter the archiving record (base table **JOB**) and the **history record JOB\_HST** as shown. Also give the archive object a description and check the base table check box.

# 1. Create the Archive Object

Navigate to:

PeopleTools > Data Archive Manager >  
Manage Archive Objects

## Manage Archive Objects

Archive Object: JOB

Description:

PeopleSoft Job Data

Records in Archive Object				Find   View All   	First  1 of 1  Last
<u>Base Record</u>	<u>'Archiving Record</u>	<u>Description</u>	<u>'History Record</u>		
1	<input checked="" type="checkbox"/>	<input type="text" value="JOB"/>	 EE Job History	<input type="text" value="JOB_HST"/>	  

## 2. Create the Archive Query

### ❖ Follow the below process:

- A. Create a new query.
- B. In the records tab, find the record **JOB** and add this record to your query. You typically use the base table to determine the criteria for archiving.
- C. In the query tab, check **EMPLID**, **EFFDT**, **ACTION**, **ACTION\_REASON**, **ACTION\_DT** the fields.
- D. In the criteria tab:
  - i. Add a criteria where **ACTION\_DT < 1960-01-01** and **Action in ('RET','TER')**



## 2. Create the Archive Query

Navigate to:

Reporting Tools > Query > Query Manager

Records Query Expressions Prompts Fields **Criteria** Having View SQL Run

Query Name: ARCHIVE\_JOB Description: PeopleSoft Job Data

Add Criteria Group Criteria Reorder Criteria

Criteria	Logical	Expression1	Condition Type	Expression 2	Edit	Delete
	▼	A.EFFDT - Effective Date	Eff Date <=	Current Date (EffSeq = Last)	Edit	[-]
AND	▼	A.ACTION - Action	in list	('RET','TER')	Edit	[-]
AND	▼	B.ACTION - Action	equal to	A.ACTION - Action	Edit	[-]
AND	▼	B.EFFDT - Effective Date	Eff Date <=	Current Date	Edit	[-]
AND	▼	C.ACTION - Action	equal to	A.ACTION - Action	Edit	[-]
AND	▼	C.ACTION_REASON - Reason Code	equal to	A.ACTION_REASON - Reason Code	Edit	[-]
AND	▼	C.EFFDT - Effective Date	Eff Date <=	Current Date	Edit	[-]
AND	▼	A.ACTION_DT - Action Date	less than	1960-01-01	Edit	[-]

Customize | Find | First 1-8 of 8 Last

Save Save As New Query Preferences Properties New Union Return to Search

## 2. Create the Archive Query

- ❖ This query returns rows from **JOB** on or before **1960-01-01**.
- ❖ View the query SQL and run your query to ensure it works. You may find that your query result set is too large to display and warning appears - this is safe to ignore.
- ❖ Click on the properties link, name your query **ARCHIVE\_JOB** and set the description to Archive PeopleSoft Job Data. Set the query type to archive and the owner to public.

## 3. Create the Archive Template

### ❖ Follow the below process:

- A. Create a new archive template with the ID **JOB**. The ID is limited to **8** characters.
- B. Specify the description *PeopleSoft Job Data*
- C. Select your archive object (**JOB**) and check the base object check box.
- D. Select your archive query.

## 3. Create the Archive Template

- E. Save the template.
- F. This simple example doesn't use a restore query. So leave the *template allows selective restoring of data from history* check box unchecked.
- G. We won't need to use pre and post custom application engine programs. However you can create your own application engine programs to run before and after the archive process to perform additional custom steps. Leave the AE processes section empty.

## 3. Create the Archive Template

Navigate to:

PeopleTools > Data Archive Manager >  
Manage Archive Templates

### Manage Archive Templates

Archive Template: JOB      Description: PeopleSoft Job Data

Archive Template Objects		Find   View All	First	1 of 1	Last
Base Object	Archive Object				
<input checked="" type="checkbox"/>	JOB	PeopleSoft Job Data			<input type="button" value="+"/> <input type="button" value="-"/>

Selective Archiving Queries		Find   View All	First	1 of 1	Last
Query Name	Description				
ARCHIVE_JOB	Archive PeopleSoft Job Data				<input type="button" value="+"/> <input type="button" value="-"/>

Template Allows Selective Restoring of Data From History

Selective Restoring Queries		Find   View All	First	1 of 1	Last
Query Name	Description				
					<input type="button" value="+"/> <input type="button" value="-"/>

AE Processes			Find   View All	First	1 of 1	Last
Archive Process	Pre AE Program	Post AE Program				
						<input type="button" value="+"/> <input type="button" value="-"/>

## 4. Managing Archived Jobs

### ❖ Follow the below steps:

1. Add a new Archive run control ID - **ARCHIVE\_JOB**.
2. Specify the archive template - **JOB**.
3. Select any one of the below **4 Process Types** to manage the archived jobs.
  - i. **Selection:** This will first be testing a straight copy from the base table to the history table.
  - ii. **Delete:** Used to delete data from transaction tables. Data rows will be deleted from the transaction tables only if they've already been archived in the history tables.

## 4. Managing Archived Jobs

### ❖ Follow the below steps:

- iii. **Rollback:** This will copy data from history tables back to transaction tables.
  - iv. **Remove from History:** Used to delete data from the history tables.
- 4. Specify the archive query as the selective query **ARCHIVE\_JOB**, if you would have selected **Process Type** as **Selection**. For **Remaining Process Types** use the batch number criteria and press the prompt on the Batch Number field. Select the appropriate batch number. If you have only run the archive once, then this will be batch number 1.

---

## 4. Managing Archived Jobs

5. Leave commit processing to Set-based Processing.
  6. Check the audit row count checkbox. This allows you to see the number of rows archived by your run in the audit archiving page.
  7. Save and run the process. This calls the data archive manager **PSARCHIVE** application engine program.
-



## 4. Managing Archived Jobs

Navigate to:

PeopleTools > Data Archive Manager >  
Archive Data to History

Run Control ID: ARCHIVE\_JOB

[Report Manager](#) [Process Monitor](#)

Archive Template	
*Archive Template:	<input type="text" value="JOB"/> <input type="button" value="Search"/> PeopleSoft Job Data
Archive Process	
*Process Type	<input type="text" value="Selection"/> <input type="button" value="v"/> copying from on-line tables to history tables
Selection Criteria	
<input checked="" type="radio"/> Selective Query	<input type="text" value=""/> <input type="button" value="v"/>
<input type="radio"/> Batch Number	<input type="text" value=""/>
Commit Processing	
<input type="radio"/> Commit at End	
<input checked="" type="radio"/> Set-Based Processing	
<input type="radio"/> Row-Based Processing	
<input type="checkbox"/> Audit Row Count	

## 4. Managing Archived Jobs

❖ You should also **Navigate to:**

**PeopleTools > Data Archive Manager >**

**Audit Archiving**

❖ Check that your archive worked and that the correct number of rows are displayed. The view details button shows you the SQL used by the **PSARCHIVE** application engine to perform the archive.

# Agenda

---

- ❖ Introduction
  - ❖ Case Study
  - ❖ Setup
  - ❖ Data Archiving Process
  - ❖ **Limitations/Challenges**
  - ❖ Alternate Methods of Data Archiving
  - ❖ Q&A
-

# Limitations

---

## 1. Index Limitations

- i. The database platform may have a limitation on the number of columns that an index can contain. Some have a restriction of 16 columns for an index. If the table that you want to archive already has **16** keys, then you can't add other keys (**PSARCH\_ID** and **PSARCH\_BATCHNUM** from **PSARCHIVE\_SBR** sub-record) to the corresponding history table.
  
- ii. To solve this problem, you can create the history table with the **PSARCH\_ID** and **PSARCH\_BATCHNUM** as non-key fields.

# Limitations

---

## 2. Data Limitations

- i. For *Oracle databases only*, due to platform and meta-SQL restrictions, Data Archive Manager does not support archiving of records with **LONG**, **IMAGE**, or **ATTACHMENT** columns if you have not performed a data type switch. If you have performed a data type switch, there are no limitations. The selection process (inserting data from the online records to the history records) will result in the loss of the long, image, or attachment columns in the history record.
  
- ii. However, this restriction applies only to templates archived using set-based processing. Long, image, and attachment data are archived to history records (and back to the transactional records) if the template is archived using row-based processing.

# Limitations

---

## 3. Online Access:

For the same PSFT instance there cannot be simultaneous online access to both **archived** and **current data**.

## Challenges

- ❖ **PSARCHIVE** batch process creates queries to delete select and insert which cannot be controlled or customized so we could not improve the performance of few process.
  - Since this is not a daily or weekly process it did not create much of an impact.
  - With Version **8.49** Oracle has tuned the process to run slightly better

# Agenda

---

- ❖ Introduction
  - ❖ Case Study
  - ❖ Setup
  - ❖ Data Archiving Process
  - ❖ Limitations/Challenges
  - ❖ **Alternate Methods of Data Archiving**
  - ❖ Q&A
-

# Alternate Methods of Data Archiving

- ❖ Below are some of the alternate methods of archiving data:
  - **Data Mover Scripts:**  
Creating Data Mover scripts to Export/Import the data. This is a very simple process to backup the data in PeopleSoft.
  - **Flat Files:**  
Archiving data directly from the DB into csv, xls file formats.
  - **Application Engine Programs:**  
By using the File Layouts, we can archive data.
  - **HP data archiving manager:**  
HP data archiving manager addresses online simultaneous access to archived and current data from same PSFT instance by building a layer.



# Agenda

---

- ❖ Introduction
  - ❖ Case Study
  - ❖ Setup
  - ❖ Data Archiving Process
  - ❖ Limitations/Challenges
  - ❖ Alternate Methods of Data Archiving
  - ❖ Q&A
-

# Q&A



# About SOAIS

---

SOAIS is a provider of Enterprise IT and Process outsourcing solutions. Since its inception SOAIS has expanded at a tremendous pace and has garnered customers from both mid-market segment and Fortune 100 companies. We have experience in managing ERP applications as well as in providing high value services around packaged enterprise applications such as PeopleSoft and Oracle. Our experience in the business process outsourcing area fully extends our services footprint to provide end to end enterprise wide solutions.

***Please feel free to clarify queries or provide feedback on this presentation at <http://www.soais.com/askexpert.html>***

---