16 TB of Disk Savings and 3 Oracle Applications Modules Retired in 3 Days: EMC IT’s Informatica Data Retirement Proof of Concept

Abstract

This white paper illustrates the ability to reduce the data growth challenge seen with EMC’s Oracle Applications CRM implementation via a Proof of Concept (POC) sponsored by the EMC IT organization. This POC will demonstrate the rapid retirement of three modules of an acquired company’s Oracle E-Business Suite footprint via Informatica’s Data Archive for Retirement solution.

April 2011
16 TB of Disk Savings and 3 Oracle Applications Modules Retired in 3 Days: EMC IT’s Informatica Data Retirement Proof of Concept

Applied Technology
# Table of Contents

Executive summary .......................................................... 4  
Introduction ........................................................................ 4  
  Audience ............................................................................ 4  
Data retirement needs ......................................................... 4  
Technology overview ........................................................... 5  
  Informatica Data Archive for Retirement architecture .............. 5  
  Enterprise Data Manager ..................................................... 6  
  Informatica Data Archive for Retirement Workbench user interface ........................................................................... 7  
  EMC IT’s Informatica Archive for Retirement deployment infrastructure ................................................................. 13  
EMC IT’s Proof of Concept ...................................................... 14  
  POC daily actions ................................................................ 14  
    Highlights from the POC .................................................... 14  
    Day 1 .............................................................................. 14  
    Day 2 .............................................................................. 15  
    Day 3 .............................................................................. 15  
Conclusion ........................................................................... 15  
  Acknowledgments .................................................................. 15
Executive summary

Exponential “data growth” with enterprise applications such as Oracle E-Business Suite brings challenges for IT organizations to manage the physical (storage), operational processes (replication/archive), and financial (people/technology) costs of their data explosion.

An enterprise’s data footprint can be defined as the total storage needed by the organization to fulfill its business needs for the lifecycle of its Oracle Applications implementation infrastructures, in areas such as:

- Development
- Production
- Test
- Business continuity/disaster recovery

Introduction

The purpose of this paper is to illustrate a method and toolset to retire the data growth challenges above via a Proof of Concept (POC) sponsored by the EMC IT organization. This POC demonstrated the rapid retirement of three modules of Oracle Applications (from Legato, a company acquired by EMC) utilizing Informatica’s Data Archive for Retirement solution.

Audience

This white paper is focused on the CIO, system architect, Oracle architect, storage architect, and supporting staff, focusing on Oracle Applications DBAs, server administrators, and network administrators.

Data retirement needs

EMC, like many large enterprises, has deployed enterprise-scale implementations of Oracle’s ERP and CRM solutions to enable its business in Manufacturing, Finance, Quoting, Customer Service, Professional Services, Sales, and Marketing.

Two enterprise-scale mission-critical systems support EMC’s core revenue-generating functions:

- An ERP solution, supporting 20,000 employees with 2,000 concurrent users
- A CRM solution, supporting 36,000 named users worldwide with 3,500 concurrent users. This implementation is one of the top five Oracle Applications transactional systems in the world depending upon the modules that are used.

With time, EMC’s ERP production database has grown to 2 TB in size, and the CRM production database has grown to about 7 TB. There are currently 15 instances of ERP and 19 instances of CRM databases serving to ensure efficient global application delivery and support.

As with many growth companies, EMC has acquired over 35 companies in the past 6 years, including Legato (known for the NetWorker® backup solution). Once integrated into the EMC ERP/CRM systems, Legato’s infrastructures and Oracle Applications instance/infrastructure were prime candidates to be retired and decommissioned. This POC will illustrate the rapid retirement of three Oracle Applications modules. This baseline gives EMC IT’s POC team the ability to understand the retirement of EMC Legato’s Oracle Applications instance and to benefit from the elimination of the following operating and capital expenditures:

- Software licenses
- Maintenance fees
- Hardware
- Subject matter expertise

16 TB of Disk Savings and 3 Oracle Applications Modules Retired in 3 Days: EMC IT’s Informatica Data Retirement Proof of Concept

Applied Technology
Figure 1 illustrates the direct retirement from a legacy Oracle Applications instance (Legato’s) via the Informatica Data Archive for Retirement and the ability to query the archived/retired data with data query tools such as Informatica’s Data Discovery toolset.

EMC IT’s direct from legacy production to archive

The EMC IT organization took on the challenge of a Proof of Concept to evaluate Informatica’s Data Archive for Retirement solution. This solution has two user interface components: an Enterprise Data Manager that has accelerators (Metadata) that speed the data mining of a module to be retired and to define retirement objects needed to retire in the module, and the Workbench, which controls the retirement process.

Objectives of EMC IT’s POC were the following:

- Use a copy of EMC Legato’s “11.5.8” production instance
- Retire the following Oracle Applications 11.5.8 modules using Informatica Data Archive for Retirement:
  - Accounts Receivable
  - General Ledger
  - Inventory
- Retire objects moved to an EMC Centera® device
- Use the Data Discovery feature to access the retired modules
- Demonstrate access using a JDBC client like Aqua or similar third-party reporting tools

Technology overview

The following sections describe the technology components used in this POC.

Informatica Data Archive for Retirement architecture

Figure 2 illustrates the components in the Informatica Data Archive for Retirement architecture:
16 TB of Disk Savings and 3 Oracle Applications Modules Retired in 3 Days: EMC IT’s Informatica Data Retirement Proof of Concept

Applied Technology

Figure 2. Informatica Data Archive for Retirement architecture

The archive server relocates inactive data from the production database and produces compressed temporary staging files in a designated staging area. The temporary staging files are converted into an optimized file archive format to be loaded to the file archive store by the file archive server. Reporting tools use the JDBC/ODBC interface to retrieve data from the file archive store (by communicating with the file archive server).

How does it work?

Informatica Data Archive mines the metadata stored in the database to capture information about the application tables and relationships between them. Prebuilt accelerators for prepackaged enterprise applications, such as Oracle E-Business Suite, SAP, PeopleSoft, Siebel CRM, and others, have comprehensive metadata about the business entities, the associated database tables for those entities, and the business rules that validate which transaction records are eligible for archiving. These prebuilt accelerators can be easily customized to incorporate custom modules, tables, and attributes that are typically defined for different businesses. Application accelerators can also be extended to archive and retire custom applications. The combination of mined and prebuilt application metadata allows Informatica Data Archive to relocate complete transactions and in this case, retire the entire application data, including reference data.

Informatica Data Archive provides a wizard-based interface (Retirement Workbench) to easily create archive policy definitions. In an application retirement scenario, the Archive Engine relocates all the data within identified tables and entities based on the archiving policy definition. Simulation reports enable administrators to review the effect of the retirement policy before actually moving the data. The simulation reports provide information about what modules and tables are impacted, which business rules have passed validation, and the number of rows and size of the data that will be archived. Once the archive policies are defined, Informatica Data Archive can relocate large amounts of data quickly using a high-performance and scalable engine.

Retired application data is stored in a highly compressed immutable file archive format, which significantly reduces the size of the data. Retired data can still be easily accessed through the Informatica Data Discovery portal for searching and browsing and using any third-party reporting or business intelligence tool such as Crystal Reports, MicroStrategy, Business Objects, and Cognos.

Enterprise Data Manager

Enterprise Data Manager (EDM) is a component of Informatica Data Archive, where the action of creating the retirement object via data mining and creating the retirement objects for the module to retire is defined. This software component gives the retirement team member an easy-to-use toolset for the relationship of the application module, business entities, and database tables.

Figure 3 is a screenshot of the EDM tool.
Informatica Data Archive for Retirement Workbench user interface

The Retirement Workbench is where the retirement job is defined and executed. This software component gives the retire team members an easy-to-use interface that is the “control center” for retirement activities during the POC.

Figure 4 shows a screenshot of the Workbench user interface.
Figure 5 shows how to retire the entire module via the **Insert table with where 1=1** option. This method will archive all records in the specified tables within the retired modules.

**Figure 5. Retirement module example**

**Informatica Data Archive for Retirement method**

The following are the steps used in the Informatica Data Archive for Retirement process:

1. Data mining of the retired module – Using the EDM user interface, mine the module for objects to retire, as shown in Figure 6.

**Figure 6. Data mining via Enterprise Data Manager**

2. Define retirement objects using EDM.
3. Define the source and destination of retirement – Using the Informatica Data Archive for Retirement Workbench/UI, define the source application connectivity and the file archive destination.

4. Create the security groups - Using the Informatica Data Archive for Retirement Workbench/UI, create the security groups. The security groups control which user has access to the source module and entities. Only users who have the appropriate access control, as defined in the security groups, will be able to create and execute a retirement job for that source application module and entity.
5. Create the Retirement definitions (Metro stops) using the Informatica Data Archive for Retirement Workbench/UI.

6. Schedule the archive crawler process - Using the Informatica Data Archive for Retirement Workbench/UI schedule the process to convert the temporary staging files into the highly compressed, optimized file archive, which is stored in EMC Centera.
16 TB of Disk Savings and 3 Oracle Applications Modules Retired in 3 Days: EMC IT’s Informatica Data Retirement Proof of Concept

Figure 11. Archive Crawler process via the Informatica Data Archive for Retirement UI

7. **Data Discovery** – Using the Informatica Data Discovery web-based user interface, search for specific transactions in the retired data archive.
Figure 12. Search for and view transaction records from Informatica Data Discovery
EMC IT’s Informatica Archive for Retirement deployment infrastructure

The following was the EMC POC Legato retirement deployment infrastructure:

- A deployed VMware® ESX®/ESXi™ machine. The recommended server is an eight-core CPU.
- A single file archive server should be able to handle a 20 TB (compressed) archive. This is equivalent to approximately 400 TB of source data.
- Informatica’s Enterprise Data Manager (EDM) client user interface resided on EMC IT’s POC retirement personnel laptop.

As Figure 13 illustrates, the VMware server was used to deploy the Informatica Archive for Retirement toolset.

![Informatica Archive for Retirement](image)

**Figure 13. EMC IT’s Informatica Archive for Retirement deployment infrastructure**

The following were the storage components used in the POC:

**EMC Symmetrix DMX™** - The Symmetrix DMX-4 system extends EMC’s leadership in the high-end enterprise and storage market. The DMX-4 delivers immediate support for the latest generation of disk drive technologies, Flash drives for superior performance, 4 Gb/s Fibre Channel for high performance, and SATA II for high capacity.

The DMX-4 is based on Enginuity™ 5773, which provides investment protection that delivers performance gains along with information-centric security advancements via integration with RSA enVision®. With the DMX-4 and Enginuity 5773 all replication and security activities are easy to manage with the Symmetrix Management Console (SMC).

**EMC TimeFinder®** — TimeFinder allows users to nondisruptively create and manage point-in-time copies of data (local replication). This allows operational processes, such as backup, reporting, and application testing, to be performed independently of the source application to maximize service levels, without impacting performance or availability.

TimeFinder/Clone was used in this use case. It creates highly functional, high-performance, pointer-based, full-volume copies of Symmetrix DMX volumes that can be used as point-in-time copies for data warehouse refreshes, backups, online restores, and even volume migrations.

**EMC PowerPath®** — PowerPath works with the storage system to intelligently manage I/O paths, and supports multiple paths to a logical device. In this solution PowerPath manages four I/O paths and provides:

- Automatic failover in the event of a hardware failure. PowerPath automatically detects path failure and redirects I/O to another path.
Dynamic multipath load balancing. PowerPath distributes I/O requests to a logical device across all available paths, thus improving I/O performance and reducing management time and downtime by eliminating the need to configure paths statically across logical devices.

EMC Centera — EMC Centera is an IP network storage system specifically designed to store and provide fast, easy access to fixed content. It is the first solution to offer online availability with long-term retention and assured integrity for this fastest-growing category of information.

EMC Centera is the optimal information archive for businesses and organizations that require a simplified solution to expanding amounts of fixed content. EMC Centera greatly simplifies the task of managing, sharing, and protecting all sizes of content repositories. It cost-effectively puts information online in support of new sources of revenue generation, expanded business models, and increased service levels to users and customers.

EMC IT’s Proof of Concept

The EMC POC consists of three important components:

- **POC team** – EMC IT’s Oracle Applications DBA team and Oracle Applications Development team (modules), Subject Matter Experts (SMEs), and Informatica SMEs
- **Retirement process** – Described in the “Informatica Data Archive for Retirement method” section on page 8
- **Technology** - Informatica’s Date Archive for Retirement solution (EDM and Workbench UI)

**POC daily actions**

The following were the activities done each day for the POC candidate modules:

**Step 1.** Review the EDM toolset for the Mining module/retirement objects. Use the EDM to define what data to use in the retirement process.

**Step 2.** Confirm with the development/business team what the retirement parameters are for each module.

**Step 3.** Commence the retirement via Informatica’s Data Archive for Retirement Workbench to start the retirement process. Please review the “Informatica Data Archive for Retirement method” section.

**Step 4.** Review Informatica’s Workbench (UI) retirement results.

**Step 5.** Query the retired data via Informatica’s Data Discovery tool or a third-party reporting tool (JDBC/ODBC).

**Highlights from the POC**

The following highlights the dramatic retirement of modules for each day of the POC for EMC IT.

**Day 1**

The POC team started the actions of the POC using both EDM and the Informatica Data Archive for Retirement Workbench to retire the first module, which was Accounts Payable (AP).

- VMware and virtual machine image (file archive repository) deployment (Figure 13)
- Mining of the AP module using EDM
- Creation of a retirement object for AP using EDM
- Data extraction to temporary staging files (Figure 13)
Day 2
The POC continued the retirement for the other two POC modules — Inventory (INV) and General Ledger (GL). The following was accomplished on Day 2:

- AP module data was converted to compressed, optimized file archive format
- Data stored successfully in EMC Centera
- Discover data, which is in EMC Centera, through the Informatica Data Discovery portal and JDBC connectivity was established to the file archive server
- Scheduled the retirement process for INV and GL modules via the Informatica Data Archive for Retirement Workbench

Day 3
The POC team completed the POC ahead of schedule. The following was accomplished on Day 3:

- Converted the INV and GL modules’ data to compressed optimized file archive format and stored it in EMC Centera
- Used the Informatica Data Discovery user interface to search and view the transaction records within the GL and INV modules
- Overall, compressed 17 GB of source data to 1 GB of data in the optimized file archive format and pushed the archived data to EMC Centera

Conclusion
The following was the impact of the EMC IT Informatica Data Retirement POC:

- Successful application/module retirement of EMC candidate modules (AR, GL, and INV) in a three-day POC
- Successful demonstration of data availability to an EMC Centera device
- Overall, compressed 17 GB of data to 1 GB and migrated the data to EMC Centera
- Successful access from a reporting/query tool (JDBC) to the retired data of Oracle Applications tables such as GL_BALANCES

In conclusion, EMC IT used the savings shown above (EMC’s retirement of Legato’s Oracle Applications environment of GL, AP, and INV modules) to estimate the result of retiring other designated legacy applications. The estimate is calculated by multiplying this savings by the number of legacy applications to retire at EMC. This POC illustrated that there is a potential for 100 TB of storage savings for the next 5 years, resulting in up to $1 million in cost savings.

Acknowledgments
The author would like to thank EMC IT’s POC team and IT development team, and Informatica’s POC team for assisting in the creation of this white paper.