INFORMATICA®
The Data Integration Company™

9 Steps to Successful Information Lifecycle Management:

Best Practices for Efficient Database Archiving

WHITE PAPER

This edition published June 2009

Table of Contents

A leading manufacturer of electronic test tools and software needed to dramatically improve the response time of an inventory on-line application. Archiving inventory data produced immediate performance improvement and gave the businesspeople relief from ever-increasing performance problems.

Executive Summary

Organizations that use prepackaged ERP/CRM, custom, and third-party applications are seeing their production databases grow exponentially. At the same time, business policies and regulations require them to retain structured and unstructured data indefinitely. Storing increasing amounts of data on production systems is a recipe for poor performance no matter how much hardware is added or how much an application is tuned. Organizations need a way to manage this growth effectively.

Over the past few years, the Storage Networking Industry Association (SNIA) has promoted the concept of Information Lifecycle Management (ILM) as a means of better aligning the business value of data with the most appropriate and cost-effective IT infrastructure—from the time information is added to the database until it can be destroyed. However, the SNIA does not recommend specific tools to get the job done or how best to use tools to implement ILM.

This white paper describes why data archiving provides a highly effective application ILM solution and how to implement such an archiving solution to most effectively manage data throughout its life cycle.

Exponentially Increasing Data Volumes

Organizations that employ prepackaged enterprise and CRM applications, such as Oracle, PeopleSoft, and Siebel, as well as custom and third-party applications face mushrooming data volumes. The SNIA estimates that many large organizations had an average compound storage growth rate of 80 percent from 1999 to 2003. To make matters worse, the volume is growing at near exponential rates. In fact, IDC research shows that digital information will grow from 281 exabytes in 2007 to nearly 1,800 exabytes in 2011, which is compound annual growth rate of almost 60 percent.[1]

Where Does This Growth Come From?

As enterprise application vendors expanded and improved their applications in the late 1990s to make their applications truly enterprise-grade solutions, organizations expanded their use of these applications throughout their enterprise. As a consequence, these organizations have had exponential transactional data growth. Rarely, if ever, did they delete data. Organizations have continued to add new applications, further increasing the amount of data they generate. Moreover, with the advent of the Internet, more users than ever have been demanding access to the business systems that IT supports. These additional business users continue to add to the transaction data growth problem.

At the same time that data volume has been growing, it has become increasingly difficult for organizations to purge data. Organizations have increasingly adopted conservative data retention policies to address the threat of potential future litigation. Regulations such as the Health Insurance Portability and Accounting Act (HIPAA), Sarbanes-Oxley (SOX), SOX for Japanese Companies (J-SOX), Basel II in Europe, and many others require organizations to retain business data indefinitely.

As data volumes have grown, the time and effort necessary for end users and database administrators to perform essential tasks on production systems has increased. End users find that data entry responsiveness declines and reports take longer to run. Database backups are slower. And essential administrative tasks such as upgrading applications or applying software patches become more time consuming.

[1] IDC, *The Diverse and Exploding Digital Universe, An Updated Forecast of Worldwide Information Growth Through 2011*, March 2008

Inadequate Solutions

Until recently, organizations responded to growing databases by purchasing additional storage and processing hardware, tuning application code, or using vendor-provided purge routines. Yet, no matter how much hardware they added, database sizes continued their upward march. This meant organizations found themselves continually increasing hardware outlays at a time when shrinking budgets limited the resources IT had available to throw at the problem.

When tuning application code, DBAs discovered that tuning was most effective the first time while successive tunings offered diminishing returns.

Some enterprise application and CRM vendors have offered solutions that purge and/or archive data. However, these solutions are inadequate for a number of reasons.

- These routines were implemented inconsistently across modules, increasing the training and testing required; for example, an estimated 15 percent of Oracle modules come with purge routines; of this 15 percent, only 50 percent of Oracle modules come with both purge and archive routines; the remaining other modules have neither. Another example of a business application is Seibel, which has no archiving routines.
- Because organizations need to retain data, a purge routine that deletes data entirely is not a viable option.
- The limited number of software vendor archiving routines that remove data from production systems are often inflexible. They do not provide extensible business rules or the ability to accommodate customizations. This can result in both an inadequate amount of data and the wrong data being archived and therefore failing to meet the data management objectives of an organization's overall application ILM strategy.
- To achieve buy-in from end users, organizations need to continue to make historical data available to users and allow them to access it seamlessly along with production data. Yet when organizations archive data using ERP vendor routines, end users typically must run separate reports on the live and the archived data.

The Solution: Application Information Lifecycle Management (ILM)

More recently, industry analysts and experts have found that the solution to managing exploding data volumes lies in the fact that the value of individual data items changes over time. As just one example, organizations running distribution applications may occasionally need to access old inventory transactions. However, most of this inventory data is no longer required for day-to-day business operations. Through a process called application Information Lifecycle Management (ILM), organizations can move less frequently accessed data from production systems to second-line storage to reduce costs and improve performance—all while satisfying retention, access, and security requirements.

The Storage Networking Industry Association (SNIA) defines ILM as "policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost-effective IT infrastructure from the time information is conceived through its final disposition."

Specifically, application ILM encourages organizations to:

- Understand how their data has grown
- Monitor how data usage has changed over time
- Predict how their data will grow
- Decide how long data should survive
- Adhere to all the rules and regulations that now apply to data

Benefits of an application ILM solution include:

- Improving application performance by eliminating unnecessary data from the production database
- Reducing total cost of ownership (TCO) by lowering hardware costs, reducing storage costs and reducing DBA support time
- Enabling regulatory compliance

The largest wireless company in the United States could not complete month-end processing due to growing fixed asset data. Archiving fixed asset data not only allowed reports that had been dropped from the month-end processing to complete but also allowed a complex asset revalidation process to be completed as part of a major business merger.

Archiving: A Best-Practices Approach to Implementing Application ILM

While the SNIA defines what an ILM system should accomplish, it does not specify any particular technology for implementing application ILM. Archiving is one approach that can be particularly effective—if organizations follow archiving best practices to ensure the optimal management of data during its life cycle.

The nine archiving best practices

1. Understand your data growth trends
2. Determine your success criteria
3. Establish a data retention policy
4. Select a solution with prepackaged business rules
5. Customize the business rules, as needed
6. Test the business rules
7. Create user access policies
8. Ensure restoration
9. Follow a time-tested methodology

1. Understand Your Data Growth Trends

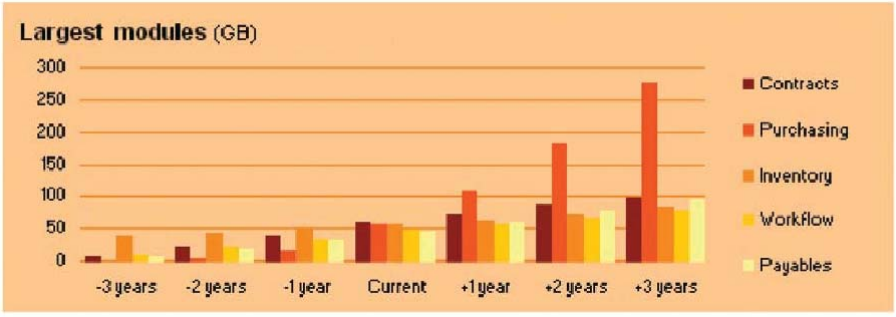
As organizations grow, adjust their business strategies, or undergo mergers and acquisitions, their data volumes expand and storage requirements change. To plan their archiving strategy most effectively, organizations need visibility into the resulting data growth trends.

A best-practice archiving solution will include tools to enable the organization to evaluate where data is currently located as well as which applications and tables are responsible for the most data growth. Organizations must perform this evaluation on an ongoing basis to continually adjust their archiving strategy as necessary and maximize the ROI for these archiving efforts.

One example of a solution that enables the evaluation of data growth is the data growth analysis tool, a feature of the Informatica® Application Information Lifecycle Management products, shown in Figure 1. This tool takes a snapshot of an application database and determines how data is distributed across different modules. The data growth analysis tool examines historical data to determine how the database has grown over time. Sophisticated algorithms use this trending information to predict future growth. The data growth analysis tool also enables administrators to calculate the ROI for different archiving alternatives to help organizations determine the best way to structure their archiving efforts.

Figure 1: Tables Belonging to Global Industries' Contracts, Purchasing, and Inventory Modules Make Up 32 Percent of All Data (170 of 532 GB):

	Estimated			***Actual***	*Estimated*		
	-3 years	**-2 years**	**-1 year**	**Current**	**+1 year**	**+2 years**	**+3 years**
Datafile (GB)	**286**	**482**	**699**	**1,025**	**1,322**	**1,667**	**2,073**
Data	149	250	362	532	686	865	1,076
*Largest Global Industries modules**							
Contracts	5.0	22.2	39.5	59.5	72.5	85.6	98.7
Purchasing	0.2	2.3	14.0	55.1	106.3	179.8	275.6
Inventory	37.9	42.5	47.1	55.0	62.4	71.6	82.7
Workflow	9.1	21.3	33.5	47.5	57.0	66.6	76.3
Payables	6.9	18.7	31.4	44.9	60.2	77.3	96.1



2. Determine Your Success Criteria

To define the most appropriate archiving strategy, organizations must determine their objectives. Some organizations will emphasize performance, others space savings, still others will specifically need to meet regulatory requirements. Examples of archiving goals may include:

- Improve response time for on-line queries to ensure timely access to current production data
- Shorten batch processing windows to complete before the start of routine business hours
- Reduce time required for routine database maintenance, backup, and disaster recovery processes
- Maximize the use of current storage and processing capacity and defer the cost of hardware and storage upgrades
- Meet regulatory requirements by purging selected data from the production environment and providing secure read-only access to it
- Archive before upgrade to reduce the outage window required by the upgrade

3. Establish a Data Retention Policy

Once an organization understands its environment and success criteria, it must classify the different types of data it wishes to archive. As one example, in a general ledger module, an organization may decide to classify data as balances and journals. In an order management module, an organization may classify data into different types of orders such as consumer orders or business orders or perhaps orders by business unit.

Organizations can then create data retention policies that specify criteria for retaining and archiving each classification of data. These archiving policies must take into account data access patterns and the organization's need to perform transactions on data. For example, a company may choose to keep one year of industrial orders from an order management module in the production database, while choosing to keep only six months of consumer order data in the production database. Another example is an organization could choose to keep nine months of data for its U.S. business unit while at the same time keeping three months of information for its U.K. operations, which could be dictated by different policies for accepting returns.

Data retention policies must also maintain consistency across modules, where appropriate. For example, when archiving a payroll module, organizations will want to coordinate retention policies with those of the benefits module because data for both of these modules is likely to contain significant interdependencies. Another example of the requirement is to have a consistent data retention policy that involves the inventory, bill of materials, and work in process modules across a typical manufacturing organization.

The archiving solution an organization chooses must therefore be flexible enough to accommodate separate retention policies for different data classifications and to enable them to modify these policies as requirements change.

Figure 2: offer examples of retention policies for different enterprise application solutions and modules.

Oracle Data Retention Policies

Module	Production Data Retention	Archiving Frequency
General Ledger (GL)	Current + 1 Full Fiscal Year	Annually
Fixed Assets (FA)	Current + 1 Full Fiscal Year	Annually
Global Accounting (AX)	Current + 1 Full Fiscal Year	Annually
Project Accounting (PA)	Current + 1 Full Fiscal Year	Annually
Accounts Payable/Purchasing AP/PO	6 Months	Monthly
Accounts Receivable (AR)	6 Months	Monthly
Contracts (OKC)	6 Months	Monthly
Inventory (INV)	3 Months	Monthly
Order Management and Shipping (ONT/WSH)	6 Months	Monthly
Bill Of Materials/Cost (BOM/CST)	3 Months	Monthly
Work In Process (WIP)	3 Months	Monthly
Payroll (PAY)	Current + 1 Year	Annually

PeopleSoft Data Retention Policies

Module	Production Data Retention	Archiving Frequency
NA Payroll	Current + 1 Year	Annually
Base Benefits, Benefits Administration, HCM	Current + 1 Year	Annually
Time & Labor, Expenses	6 Months	Monthly
General Ledger	Current + 1 Full Fiscal Year	Annually
Asset Management	Current + 1 Full Fiscal Year	Annually
Project Costing	Current + 1 Full Fiscal Year	Annually
Accounts Receivable	6 Months	Monthly
Billing	6 Months	Monthly
Purchasing	6 Months	Monthly
Order Management	6 Months	Monthly
Production Management	6 Months	Monthly
Inventory	3 Months	Monthly

Siebel Data Retention Policies

Module	Production Data Retention	Archiving Frequency
Opportunities	3 Months	Monthly
Quotes	3 Months	Monthly
Orders	6 Months	Monthly
Invoices	6 Months	Monthly
Quality – Change Requests	6 Months	Monthly
Service Requests	6 Months	Monthly
Activities	3 Months	Monthly
Messages	3 Months	Monthly

4. Select a Solution with Prepackaged Business Rules

The number one concern for organizations implementing a data growth management solution is to ensure the integrity of the business application. Thus, the process of archiving must take into account the business context of the data as well as relationships between different types of data. Data management is rendered even more complex because transactional dependencies are often defined at the application layer rather than the database layer. This means that a data growth management tool cannot simply reverse engineer the data model at the time of implementation. And any auto-discovery process is bound to be insufficient because it will miss all of the relationships embedded in the application. These rules and relationships can become quite complicated in large prepackaged products, such as Oracle E-Business Suite, PeopleSoft Enterprise, and Siebel CRM, which may have tens of thousands of database objects and a large number of integrated modules.

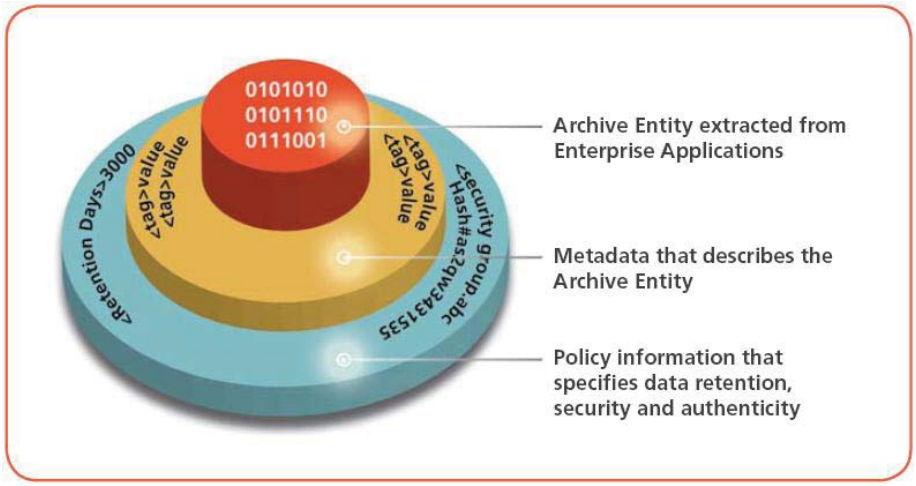
Figure 3 illustrates an example of a prepackaged business rule for Oracle applications that prevents the data management software from archiving an invoice if it is linked to a recurring payment.

Figure 3: Prepackaged Business Rules with Exceptions

Module/Application:	Accounts Payable
Category:	Business Rules for Invoices
Rule:	Invoice is a recurring payment
Exception Description:	This rule validates that the invoice reflects a recurring payment. Invoices set up as a recurring payment are not purged. Oracle requires that all recurring invoices related to a recurring invoice template be retained.
Exception Workarounds:	Recurring invoices are generated by a recurring invoice template. It may be functionally legitimate to purge all recurring invoices related a recurring invoice template if all the recurring invoices in that template are purgeable and the recurring invoice template has no periods remaining.

Successfully archiving data in these solutions requires an in-depth understanding of how the application defines a database object—that is,i.e., where the data is located and what structured and unstructured data needs to be related—and the set of rules that operate against the data. Most in-house developers have a difficult time reverse engineering the data relationships in complex applications. A best-practices archiving solution includes prepackaged business rules that incorporate an in-depth understanding of the way a particular enterprise solution stores and structures data. By choosing a solution with prepackaged rules, organizations save the time and effort of determining which tables to archive.

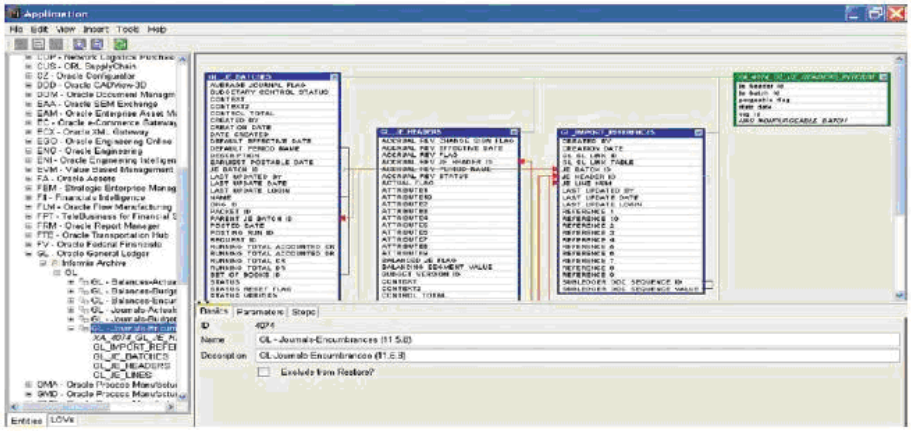
Figure 4: Data Growth Management Archive Object



5. Extend the Business Rules

Since not every ERP or CRM customer runs all of its applications the way the vendor envisions, an archiving solution must also allow organizations to modify and customize the prepackaged archiving business rules. For example, despite the fact that the primary business rule in figure 3 does not allow the archiving of recurring invoices, a custom archiving rule does allow recurring invoices to be archived when all of the recurring invoices in an invoice template are archivable. A best-practices solution should include a graphical developer toolkit, such as the one below, figure 5, from Informatica that resembles standard database design tools and makes it easy to modify the prepackaged archiving rules.

Figure 5: Graphical user interface for customizing archiving templates and business rules in Informatica Data Archive



6. Test the Business Rules

Once the organization has developed business rules, it needs to test them by simulating what will happen when data is actually archived. A best-practices solution provides simulation reporting, see Figure 6, that shows database administrators exactly how many records a given archiving policy will remove from the production system and how many will remain because the ERP classifies them as an exception. For example, in Figure 3, invoices representing recurring payments are not archived. Using simulation reporting, database administrators can iteratively adjust their archiving policy to meet their archiving objectives.

Figure 6: Simulation Reporting

Candidate Summary Report

Generated 02-Mar-2007 11:05:49 EST

Report Parameters

DB Instance: PROD Cycle: Payables Archive - 24 Month Cycles20060302 105441 Entity: AP - Invoices

Period To 29-FEB-2004

Table	Operating Unit	Purgeable Count	NonPurgeable Count	Total
XA_4103_AP_INVOICES_INTERIM	Progress UK	60	85	145

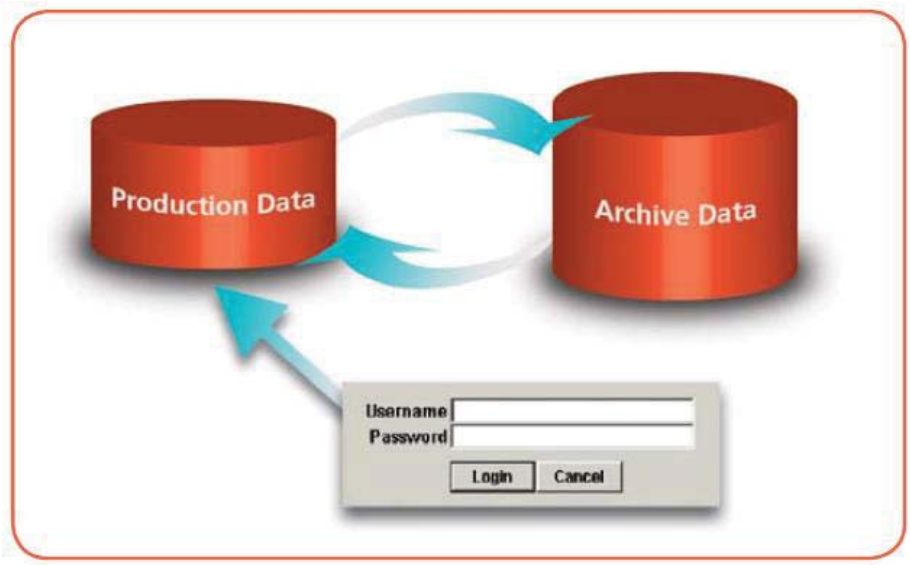
Exception Name	Count
Accounting event not purgeable	0
Check(s) match to non-purgeable invoice(s)	0
Distribution line(s) not purgeable	62
FA distribution not purgeable	5
Invoice has PA distribution line	1
Invoice has invalid Payment Status	80

7. Create User Access Policies

Many organizations will want to control which users access historical data in the archive and which data access method—screen, report, or query—they can use to access the data.

A best-practices solution will allow organizations to configure user access policies that specify which users are authorized to access historical data and which reports they are able to use. An example of this policy appears in figure 7 in which the user is authorized to see both current production and historical archived data through one seamless access view.

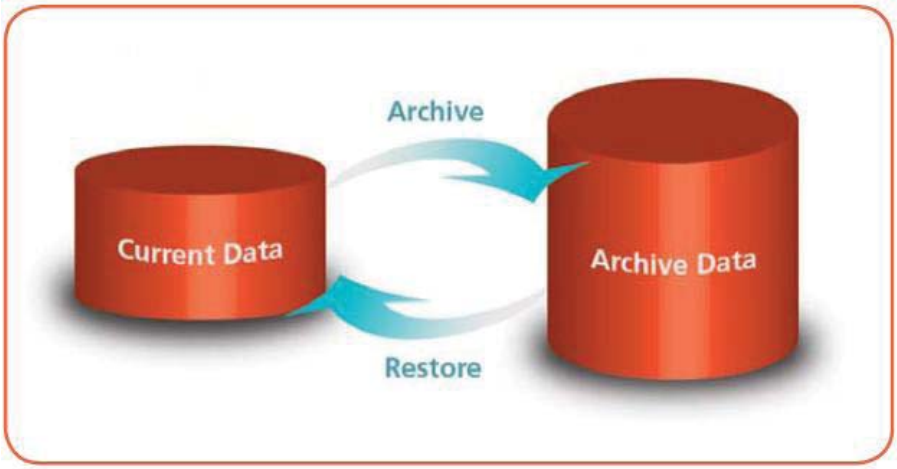
Figure 7: Seamless Data Access: Users Access Archived Data Through the Existing Production Applications Interface



8. Ensure Restoration

A restoration capability functions as an insurance policy should specific transactions need to be modified after archiving. Only by having such a restoration capability can most organizations convince business users that it is safe to implement an archiving solution.

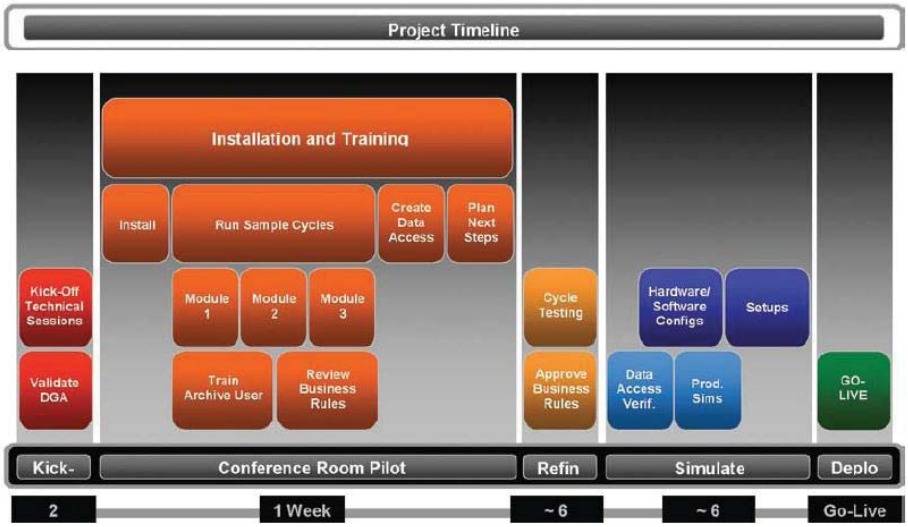
Figure 8: Archived Data Can Be Restored



9. Follow a Time-Tested Methodology

No organization wants its implementation—no matter how customized—to be on the bleeding edge of experimentation. It wants to be sure that the vendor it works with has seen and addressed the types of challenges likely to arise during an implementation. Therefore, organizations should choose a vendor that has developed an implementation methodology for complex archiving solutions that meets the outlined business objectives and has been successfully applied over a large number of implementations. Figure 9, illustrates a project plan and timeline for successfully implementing an archive solution.

Figure 9: Archive Sample Project Plan



Conclusion

Today, the size of production databases is growing exponentially. At the same time, growing numbers of regulations mean that organizations must retain their data indefinitely. Therefore, organizations need an application ILM solution. An application ILM solution will allow organizations to store data in the most appropriate IT infrastructure as it moves through its life cycle. Archiving offers an appropriate technology for implementing application ILM. Organizations that succeed in implementing a best-practices archiving solution will improve application performance by eliminating unnecessary data from their production database, reducing TCO by lowering hardware costs, and enabling regulatory compliance.

ABOUT INFORMATICA

Informatica enables organizations to operate more efficiently in today's global information economy by empowering them to access, integrate, and trust all their information assets. As the independent data integration leader, Informatica has a proven track record of success helping the world's leading companies leverage all their information assets to grow revenues, improve profitability, and increase customer loyalty.

INFORMATICA®

Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

6956 (06/10/2009)