

Make Big Data Small with Lean Data Management

Best Practices to Reduce Costs and Increase Efficiencies

WHITE PAPER



This document contains Confidential, Proprietary and Trade Secret Information (“Confidential Information”) of Informatica Corporation and may not be copied, distributed, duplicated, or otherwise reproduced in any manner without the prior written consent of Informatica.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Informatica does not accept responsibility for any kind of loss resulting from the use of information contained in this document. The information contained in this document is subject to change without notice.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Informatica software product—as well as the timing of any such release or upgrade—is at the sole discretion of Informatica.

Protected by one or more of the following U.S. Patents: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; or by the following pending U.S. Patents: 09/644,280; 10/966,046; 10/727,700.

This edition published July 2011

Table of Contents

Executive Summary	2
The Real Cost of Big Data Growth	3
Challenges in Big Applications	3
The Impact in Production Environments	3
The Multiplier Effect of Data Growth in Non-production Environments	4
Challenges in Big Data Warehouses	5
Lack of Visibility About How Data Is Used	6
Challenges in Big Application Portfolios	6
A Drain on IT Budget	6
E-discovery Costs	6
Lean Data Management: The Solution for Making Big Data Small	7
What Does Lean Mean?	7
What Is Lean Data Management?	8
Lean Applications	8
Lean Data Warehouses	10
Lean Application Portfolios	11
The Journey to Lean Data Management	12
Informatica Lean Data Management Practices in Action	14
Lean Applications: A Case Study with IKON	14
Lean Data Warehouses: A Case Study with Bank of America	14
Lean Applications and Lean Application Portfolios: A Case Study with EMC	15
Conclusion	16

Executive Summary

“Big Data”, which has become an IT industry term, means something different depending on who you ask. The basic concept of Big Data is clear: enterprise information in all its forms is becoming massive and, as a result, more difficult to manage. Enterprise data has evolved to manifest itself in unstructured, structured, semi-structured, and social media data. Whether in a Hadoop environment with petabytes of data, data warehouses containing hundreds of terabytes, or an ERP packaged system containing 5-10 terabytes, these massive data stores present the same management challenges that need to be addressed by eliminating waste, optimizing the environments, and minimizing costs associated with supporting this data.

Forrester estimates that the data in your largest databases and applications is growing 65 percent per year¹. That means the amount of data under management will more than double in the next two years. If you don't manage your data growth, you should assume that over the next two years everything you're doing now will take more than twice the time and cost than it does today. This increase will shift more of your budget to maintaining the status quo, leaving less of your budget available to take on new projects that deliver incremental value to the business.

It is clear that the Big Data challenge is real and enterprises have to deal with it. So, how do you appropriately manage information rather than become overwhelmed by Big Data? The key is having a coherent **Lean Data Management** strategy.

Lean Data Management is a sustainable approach to treating data as an asset across the enterprise. Although data is not typically an asset from an accounting perspective in the manner of a building or equipment, it can indeed be leveraged to generate business value and be used as a competitive weapon. And just as physical assets can be a liability if mismanaged, so too can data assets.

The concept of “Lean”, which is a manufacturing approach, is defined as increasing efficiency, decreasing waste, and using empirical methods to decide what matters. “The Lean concept” fits very nicely with the objectives of managing Big Data growth to reduce costs and improve efficiencies. Lean practices can be applied to all the different types of Big Data. This white paper discusses the application of Lean practices to managing structured data in application databases and data warehouses, thus the term “Lean Data Management”. In addition, this white paper discusses:

- The challenges around Big Data growth and the drivers for adopting a Lean Data Management approach
- What Lean Data Management is and why it's the right approach for managing your Big Data challenges
- How Informatica® Information Lifecycle Management solutions provide the right technology and tools for implementing Lean Data Management
- A few case studies of customers who have implemented Lean Data Management practices and the types of results they achieved

¹ Forrester Research, Inc. TechRadar: Enterprise Data Integration, February 2010

The Real Cost of Big Data Growth

Despite the fact that most of the data accumulating in databases is dormant (up to 80 percent according to some industry estimates), the idea of destroying data is scary and organizations tend to keep most of their data forever. It's also difficult because increasingly stringent compliance and e-discovery requirements mandate you hold onto the data for longer periods of time.

The issues associated with explosive data growth (Big Data) are many, and they also affect different groups within business organizations in varying dramatic and subtle ways. The most significant costs involve big transaction applications and their associated downstream copies, large data warehouses, and big application portfolios.

Challenges in Big Applications

Transactional ERP and CRM applications, such as PeopleSoft, Oracle E-Business Suite, Siebel, SAP, and other custom and homegrown applications, are the lifeblood of a business' operation. As a result, these applications are growing at a staggering rate. Business growth, expanded implementations, application upgrades, application consolidation from multiple regions and data centers or from mergers and acquisitions, as well as more stringent retention regulations, have all contributed to this data growth.

The Big Data growth problem affects not only production databases but non-production environments as well.

The Impact in Production Environments

Because transactional applications are mission-critical applications, their production environments tend to be provisioned with high-end servers and storage with the fastest performance and high-availability specifications. The highest service level agreements (SLAs) are also usually applied to these systems because thousands of users are accessing the applications and data is rapidly changing. Over time, transactions are closed and they tend to become inactive and rarely accessed. There is a tremendous amount of wasted cost, both in hard dollars and soft costs by maintaining this inactive data on the most expensive production infrastructure.

Escalating Infrastructure Costs

The most obvious cost of data growth is in increasing hardware and software infrastructure necessary to support it. Another cost associated with data growth is the requirement for more database and application software licenses because most enterprise software license prices are based on processing power or data volume. Recurring maintenance contract costs associated with these additional licenses compound the costs of supporting Big Data.

Degrading Performance and Productivity

As Figure 1 illustrates, with more data to process, application users' response time will eventually degrade, unless you add more hardware processing power. The ultimate impact of degrading application performance is loss of productivity, not only from the end users of the application but also from database and application administrators, who have to spend more time doing performance tuning.

In addition, database tuning has its limitations and even an optimized database cannot compensate for performance issues caused by sheer data volumes. The problem is not always at the database layer, but at the application layer and not much tuning can be done at that level.

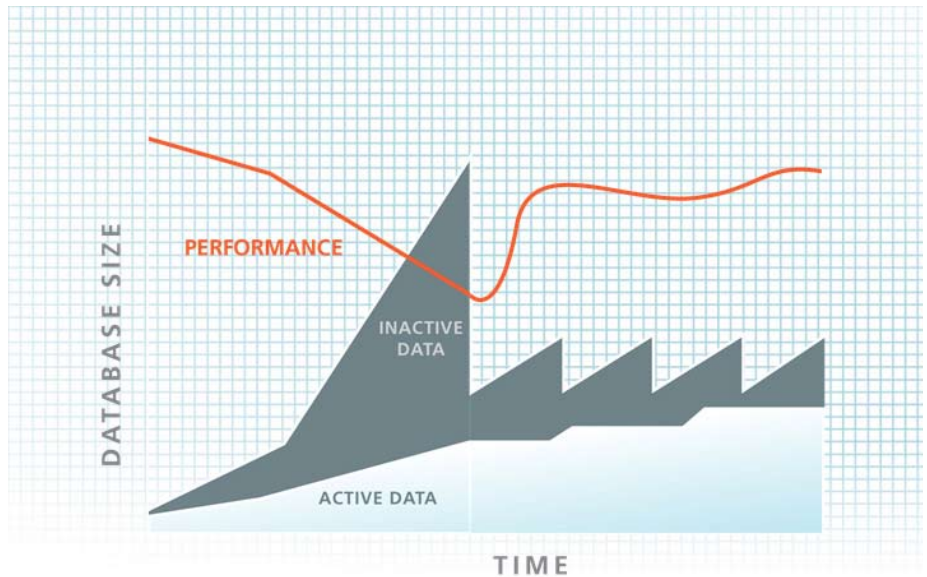


Figure 1: The impact of Big Data on performance: increasing data volumes degrade application user response time

Increased Maintenance Effort and Longer Business Outages

Backup, disaster recovery, replication, and upgrade windows increase along with data volumes. Eventually, it becomes impossible to meet backup windows or upgrade downtime over the weekend. The time it takes to get systems back online due to errors or disaster recovery can cost millions of dollars if it means downtime during operational hours. This downtime also results in an inability to respond to customers.

The Multiplier Effect of Data Growth in Non-production Environments

The cost associated with Big Data growth in production has a multiplier effect when you consider the number of full production copies replicated for non-production use such as development, testing, demo, and training purposes, as shown in Figure 2. This multiplier effect is compounded even more when you take into account the mirrors and replicated images made for each of these copies to support high availability, disaster recovery, and backups.

The issues around performance degradation, lower productivity, and increased maintenance efforts are also manifested for each of these copies. Larger production copies also mean longer times to provision environments, extended refresh times, and longer development and test lead times. The ultimate impact is in longer project delivery timelines, slower IT response time to the business, less business agility, and less competitive edge.

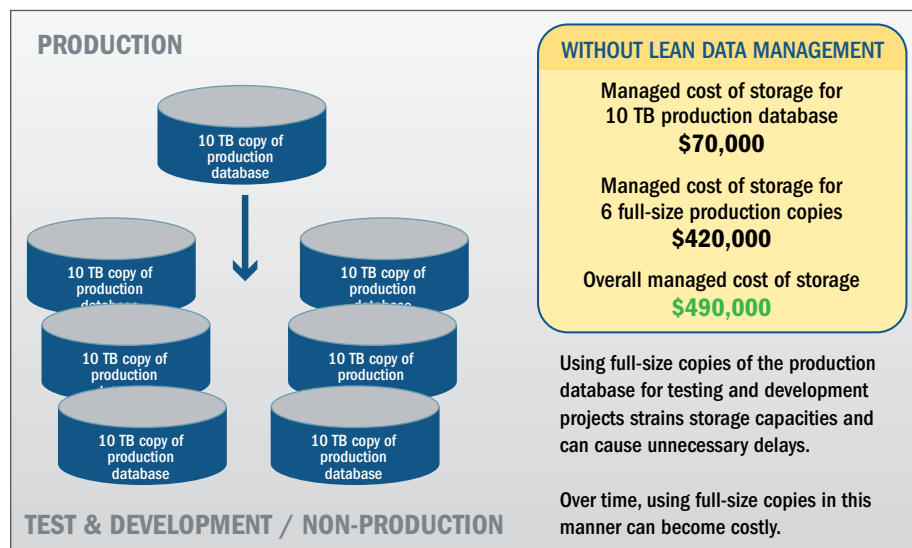


Figure 2: The multiplier effect of data growth in non-production environments

Increased Risk of Data Breach

The proliferation of data in multiple copies also increases your risk of a data breach because you have a greater amount of sensitive data scattered throughout your enterprise. The risk is even greater if your development and testing is outsourced or sent offshore. According to a recent survey from Ponemon Institute², at least half of the data copied to non-production environments may be confidential, yet these environments are seldom protected. And about 60 percent of companies use real production data for their application development and testing. As increasing amounts of sensitive data are exposed to unauthorized users, there is a higher risk of data breaches.

Challenges in Big Data Warehouses

Data warehouses experience many of the same data growth issues as transactional systems: higher infrastructure costs to support dormant data in production, large and growing maintenance windows, lower user productivity due to poor performance, any derivative data marts created, and the multiplier effect in the many non-production copies, with the associated risk of data breaches.

These issues are all more significant for data warehouses because data warehouses integrate data from multiple applications, as well as historical information for analytical reporting, so their sizes are many times larger and usually grow faster than transactional systems. Instead of a few terabytes, data warehouses typically contain tens to hundreds of terabytes. According to industry estimates, data warehouses actively use only the first years' worth of data, but maintaining historical data can easily increase the storage requirement to as much as 20 times that of the current year's data size. Based on this estimate and the massive potential size of data warehouses, the impact on cost could be staggering.

² Ponemon Institute Press Release: "Ponemon Study Shows the Cost of a Data Breach Continues to Increase", January 25, 2010.

Lack of Visibility About How Data Is Used

In transactional systems, classification of which data is used and which has become inactive is somewhat straightforward because they are dictated by the application logic and based on the status of the transactions. In data warehouses, on the other hand, whether data is used or not can only be determined by monitoring data utilization—that is, which data is being accessed, who is accessing it, and how often. Without information about which data is used and how it is used, it is very difficult for the data warehouse owner and administrators to optimize the design of the data warehouse and its supporting infrastructure. It is also difficult to prioritize among the many conflicting requests from various departments and business groups that need to access and use the data warehouse.

Without visibility about the following types of information, data warehouse infrastructure and processes will grow out of control, along with the expanded data sets that are integrated within the data warehouse:

- **Business activity and data usage:** What data is frequently used and by whom? Which data is no longer used and is dormant? What is the impact of making data changes? Who is using that data?
- **Query performance:** Which Business Intelligence (BI) users run bad ad hoc queries that impact performance? When? What reports did they run and what SQL statements were executed? What frequently used data needs optimization? What are the candidate fields for indexing?
- **Data load performance:** What critical loads are exceeding performance thresholds? What data is unnecessarily loaded that is adding overhead?

Challenges in Big Application Portfolios

The wave of mergers and acquisitions across multiple industries, plus the drive for IT modernization and consolidation, have made redundant and legacy applications stand out as a cost burden. These legacy applications are left operational for infrequent query purposes; in addition, the business may need to access the data in those applications in the future. Old systems with outdated technology make it difficult to find resources to support them. Supporting multiple systems that have different underlying technologies and platforms to serve the same function is just plain inefficient, not to mention very costly.

A Drain on IT Budget

A recent survey of organizations with more than 50 IT staff indicates that 70 percent of an organization's IT budget on average is spent on existing applications and 50 percent of those existing applications are legacy (see Figure 3). This is a significant drain on the IT budget. The cost associated with legacy applications encompasses hardware and software maintenance contracts, the staff maintaining those applications, the data center footprint, and power expenditures. This cost can be allocated to more critical new projects that are impactful to the business.

E-discovery Costs

With really old applications, if the data is not in a readily accessible format that is easily searchable and reportable, the company will be exposed to exorbitant fines due to delayed responses in the event of a legal discovery request or compliance audit. E-discovery fines are something that companies rarely anticipate and they can be significant.

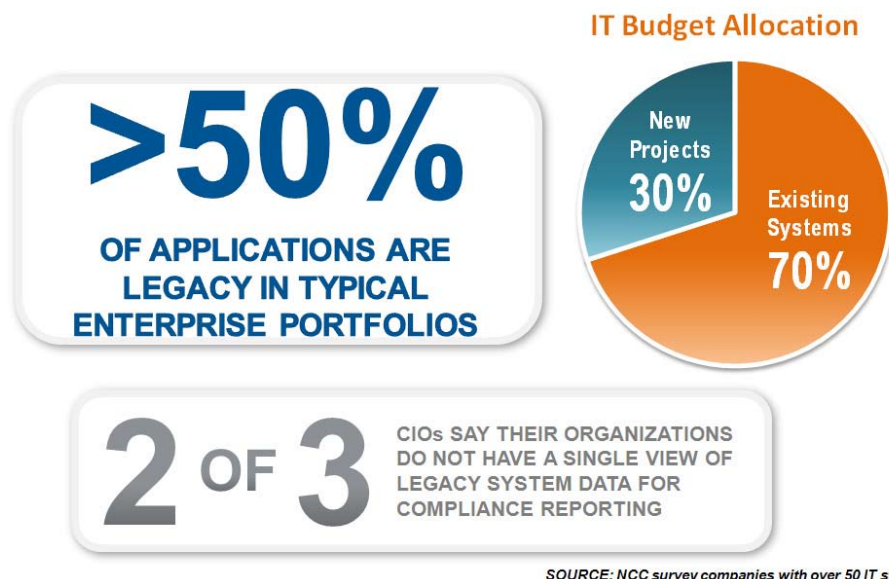


Figure 3: The Costs of Big Application Portfolios

Lean Data Management: The Solution for Making Big Data Small

So, what's an enterprise to do in light of the rapid data growth over the past several years with no slowdown in sight? The answer is that you first need to implement a comprehensive approach to managing your data using Lean Data Management practices. Lean Data Management is Informatica's approach to reducing overhead caused by Big Data and optimizing IT operational efficiencies by managing that data. This approach involves utilizing people, process, and technology.

Lean Data Management focuses on eliminating waste across transactional and analytic systems, and across production and non-production environments. By removing dormant data from applications and data warehouses, shrinking and securing non-production copies, and retiring redundant or obsolete applications, IT can dramatically lower cost, improve SLAs, and ensure proper compliance by proactively managing retention of structured data.

What Does Lean Mean?

The concept of "Lean" is defined as increasing efficiency, decreasing waste, and using empirical methods to decide what matters. Lean allows IT to align the value of information with the cost associated with managing it.

Lean principles have evolved from the manufacturing process. Lean manufacturing or lean production is a practice that considers the expenditure of resources for any goal other than the creation of value for the end customer to be wasteful and thus a target for elimination. From the perspective of the customer who consumes a product or service, "value" is defined as any action or process that a customer would be willing to pay for.

Essentially, lean principles center on preserving value with less work. Lean manufacturing is a management philosophy derived mostly from the Toyota Production System (TPS) and identified as “Lean” only in the 1990s. The TPS is renowned for its focus on reduction of the original Toyota seven wastes to improve overall customer value. The seven wastes are delay, duplication, unclear communication, unnecessary movement, incorrect inventory, defects, and lost opportunity.

What Is Lean Data Management?

According to the Data Management Association (DAMA), data management is defined as “the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise”. Lean Data Management takes this 30-year-old practice to the next level by applying lean principles such as waste elimination, automation, continuous improvement, and staff empowerment to the mix. The end result is a data driven organization with increased organizational agility to react to changes and opportunities faster than the competition.

Informatica’s approach to Lean Data Management focuses on applying application information lifecycle management (ILM) technology to three distinct areas: applications, data warehouses, and application portfolios (see Figure 4).

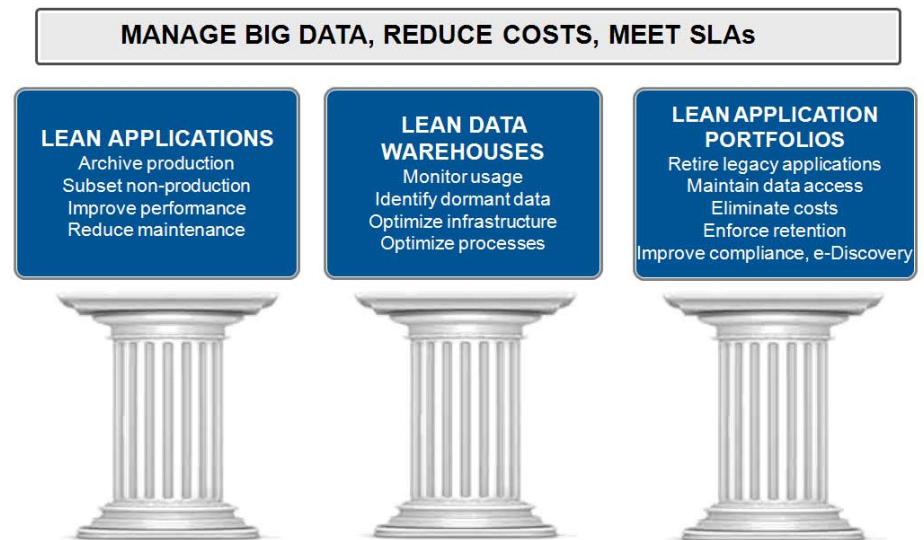


Figure 4: The three pillars of Lean Data Management practices

Lean Applications

Lean application practices apply the principles of lean to your application databases. This approach requires both inventorying and classifying your data and databases. Once these tasks are accomplished, you can prioritize those environments that will give you the biggest return on investment of time and resources in overcoming the challenges presented by Big Data.

Database archiving is one of Informatica's lean application solutions that can yield huge benefits in the struggle to overcome Big Data challenges in your application databases. To implement a database archiving solution successfully, organizations also need to establish internal processes and standards. Most importantly, retention policies must be established as mandated by regulatory and internal requirements.

Database archiving can significantly reduce the data in production in a relationally intact manner, ensuring appropriate removal and deletion of inactive data from production environments. Informatica Data Archive™ provides database archiving with seamless access to your archived data from the same original application interface, minimizing the impact on production users. Informatica Data Archive also provides tools for data growth analysis and data profiling to help automate data classification. By archiving inactive transactional data out of online transactional systems, you accomplish the goal of supporting only the mission-critical data that is currently required by production users to meet their daily responsibilities on high-end infrastructure. As a result, you can avoid the need to upgrade hardware and software licenses with transactional data growth, reduce maintenance costs, and improve performance.

In addition, by streamlining the approach to long-term data retention, you can enforce retention and disposal policies and ensure compliance to retention regulations.

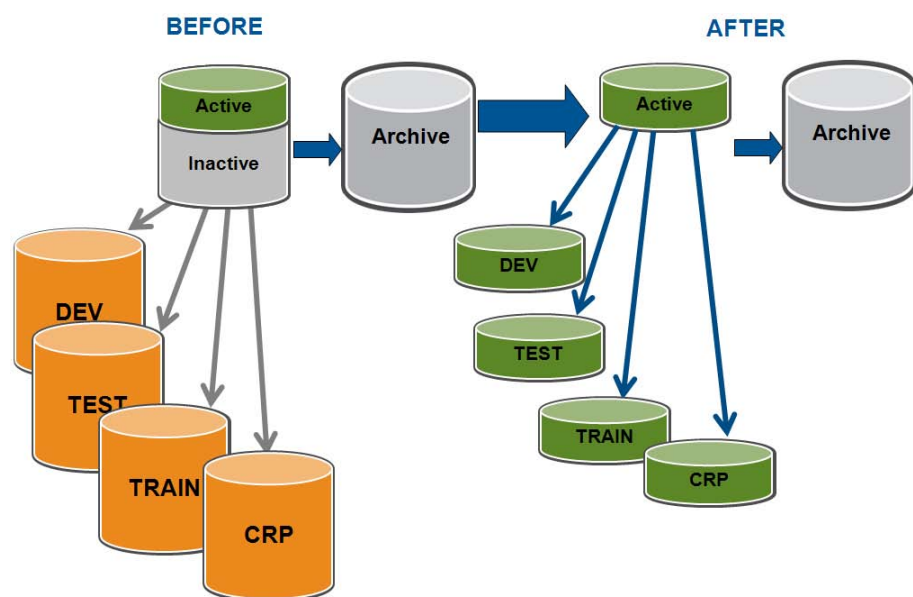


Figure 5: Managing big applications by archiving inactive data to create lean applications

Although archiving reduces the size of your production applications and any corresponding copies, as illustrated in Figure 5, creating data subsets in non-production copies can further shrink data sizes to address the Big Data challenges in non-production environments. Making subsets involves providing a smaller set of data from production based on the most relevant functional or time slices to meet the needs of the non-production users, whether they are development resources or testing teams. In addition, data masking is typically used in conjunction with data subsets to create secure non-production environments that have desensitized data (see Figure 6) - that is, sensitive data replaced with contextually accurate, but de-identified data that maintains the original characteristics. These steps will prevent data breaches in highly vulnerable non-production environments yet maintain test data quality.

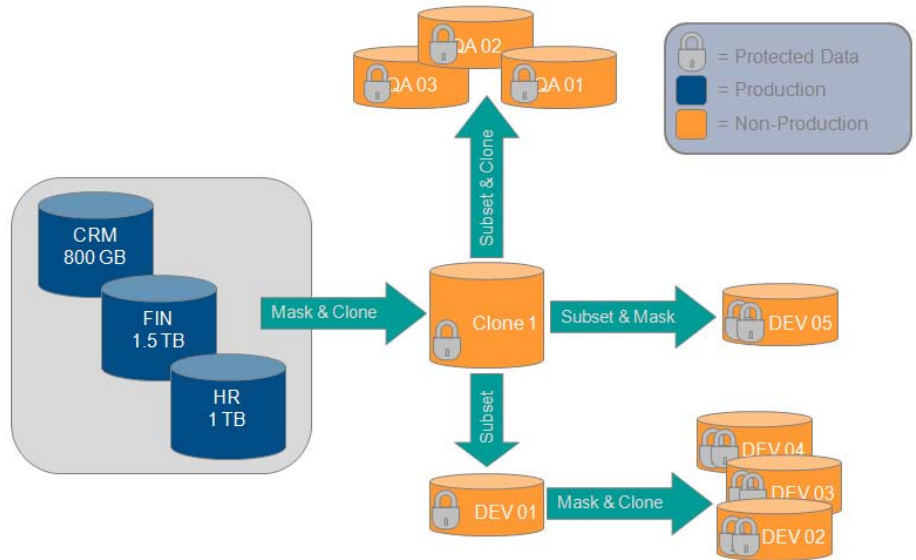


Figure 6: Managing the multiplier effect of Big Applications in non-production environments by creating subsets and masking

The combination of Informatica Data Archive for relocating inactive data from production environments and Informatica Data Subset™ to create smaller subsets of production data in non-production environments enables organizations to manage data growth in both production and non-production transactional environments and create lean applications. In addition, the application of Informatica Data Masking™ to de-identify sensitive data in non-production environments reduces the risk of exposure to unauthorized users and ensures compliance.

Lean Data Warehouses

Data warehouses can get extremely large. Informatica’s lean data warehouse practices use both the database archiving solution in production and the data subset solution in non-production instances to significantly reduce the overall footprint of production and non-production data warehouses. In addition, Informatica’s approach to lean data warehouses also brings better visibility to how the data warehouse is used across departments and applications and how it is populated.

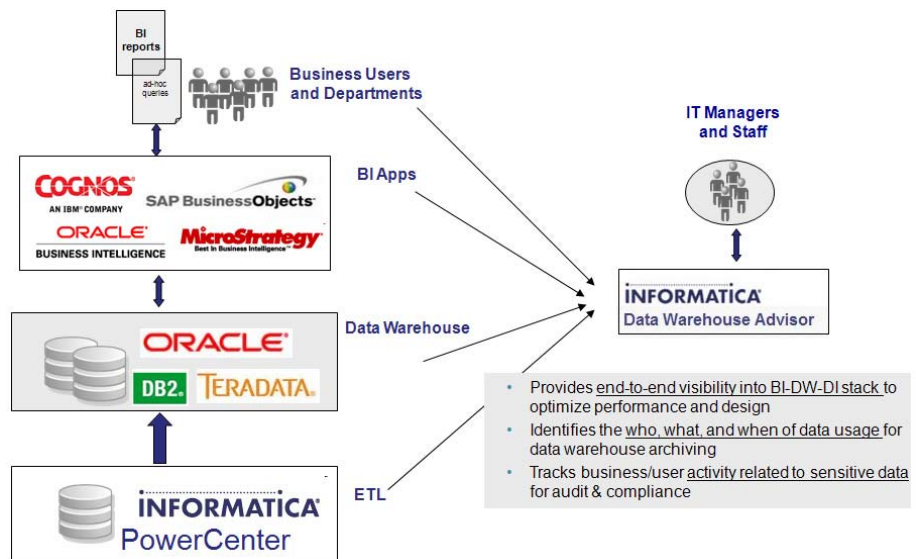


Figure 7: Manage Big Data warehouses by delivering end-to-end visibility to data warehouse usage and implementing data warehouse archives, subsets, and masking to create optimal lean data warehouses

Determining which data to archive in a data warehouse depends on which data is accessed, by whom, and the frequency of access. Informatica provides the tool for monitoring data usage to deliver end-to-end visibility about how data is being used. Informatica Data Warehouse Advisor™ monitors which departments, reports, and BI applications are using various data objects, and the types of queries they run, to optimize data warehouse design, improve performance, and manage data more efficiently. It also provides reports on the performance of batch load processes to ensure efficient execution (see Figure 7). By identifying data that’s never used, Informatica Data Warehouse Advisor can simply reduce the time to complete load processes. By identifying dormant data that’s no longer used, Informatica Data Warehouse Advisor feeds the business rules for archiving the data warehouse. The information it delivers helps justify data warehouse costs and prioritize resources based on data utilization.

Together, Informatica Data Archive, Informatica Data Subset, Informatica Data Masking, and Informatica Data Warehouse Advisor provide the technology to support lean data warehouse practices.

Lean Application Portfolios

Industry estimates indicate that as much as 20 percent of your total application costs can be reduced by retiring legacy or redundant applications. Lean application portfolio practices eliminate the overhead and costs associated with redundant and legacy applications by archiving the data out of their respective databases to a highly optimized file archive and simplifying the hardware and software stack required to retain and manage that data (see Figure 8). This optimized file archive is designed for data access and is secure, highly compressed (as much as 98 percent) to reduce storage requirements, and immutable to ensure data authenticity. Any ODBC/JDBC compliant reporting tool can be used to access the archived data. A Web-based portal supplies search and discovery allows business users to view the search result in a similar look and feel as the original application interface.

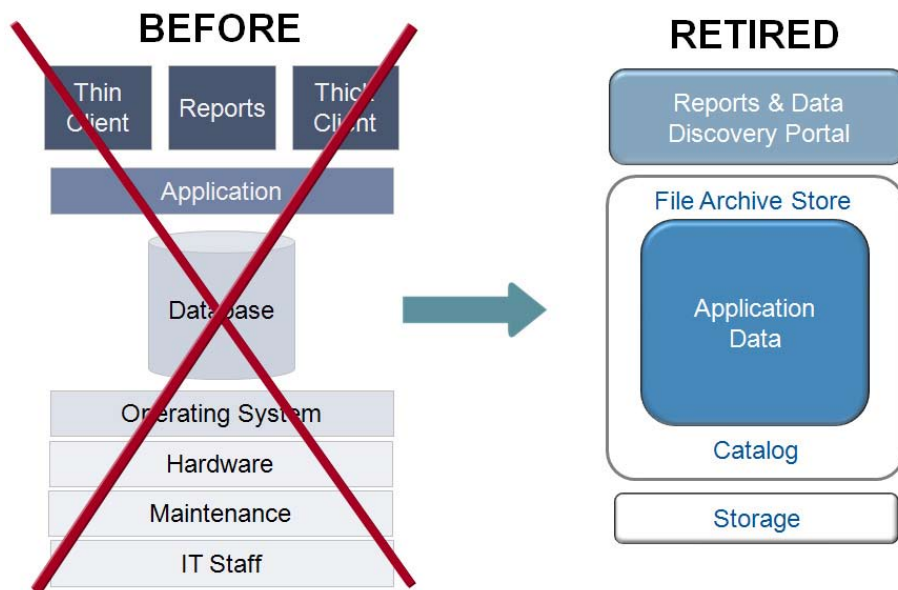


Figure 8: Reducing IT complexity and costs by retiring redundant legacy applications and making big application portfolios lean

Once data from various legacy applications is archived to a central location, you can also more effectively and consistently manage retention and disposal policies at a granular business entity level, based on regulations. Because data is now easily accessed and discovered, using standard protocols and methods, you reduce the risk of delayed responses to e-discovery requests and the potential for noncompliance.

Informatica Data Archive delivers the archiving technology and solution to support lean application portfolio practices.

The Journey to Lean Data Management

With an understanding of the three pillars of Lean Data Management practices, how does an organization that applies one or more of them evolve to the level of efficient implementation so that data becomes a competitive advantage and delivers the necessary business agility?

Figure 9 summarizes the four stages of data management approach and methodology evolution that have contributed to increasingly higher levels of operational efficiency. Hand coding, the only method available until around 1990, is still a common practice today, but it is gradually being replaced by standard tools from the past 15 years. In the second decade of 2000, more enterprise platform offerings will provide the foundation for managing data growth across all types of data, including unstructured data such as files and documents, semi-structured data such as e-mails, and structured data in databases and enterprise applications.

With the wave of IT consolidation and mergers and acquisitions, where organizations have hundreds to thousands of applications to retire, we see the emergence of the data management factories. One of the first applications of such factories is application retirement, as the next wave of data management technology combines with formal management disciplines. This wave stems from the realization that with mass retirement projects, although the source systems may be different, the technology, approach, structure, methodology, and processes are incredibly similar to each other and warrant establishing a center of excellence with highly skilled people and technology experts who can execute these retirement projects at a very high efficiency.

Management practices have also evolved from ad hoc or point-in-time projects to broad-based programs (projects of projects), to Integration Competency Centers (ICCs), and now to Lean Data Management. A major paradigm shift began early in the current century around the view of data management as a sustaining practice. The ICC encapsulates the first wave of sustainable management practices. It focused primarily on standardizing projects, tools, processes, and technology across the enterprise and addressing organizational issues related to shared services and staff competencies. The second wave of sustainable practices is the application of lean principles and techniques to eliminate waste, optimize the entire value chain, and continuously improve. The management practice that optimizes the benefits of the data management factory is Lean Data Management. The combination of factory technology and lean management practices results in significant and sustainable business benefits.

The four stages along the Lean Data Management journey lead to this question: What are the key changes that an organization needs to move from one stage to the next? The three transformations on the bottom of Figure 9 indicate the changes necessary to move to the next level of maturity.

1. The organization needs to change from independent teams working in silos to cross-functional teams that collaborate to achieve improved global business processes and other enterprise-wide solutions.
2. The organization needs to change from a focus on tactical solutions that address immediate needs (such as point to point) to viewing data management as an ongoing activity independent of the applications and source systems where the data resides. In other words, data management takes on a “life of its own” and hence becomes an enterprise strategy.
3. The organization needs to change from a top-down driven management organization to one where improvements come from front-line staff. In other words, continuous change becomes part of the organization’s culture.

As a result of these transformations in data management approach and methodologies, the perception of data also evolves along the following progression as Figure 9 illustrates:

- Unmanaged silos (data may be managed well within each silo, but there is no discipline to manage the data across the enterprise in a formal way)
- Controlled cost (improved level of maturity allows organizations to stop the bleeding and reduce operational costs)
- Shared resource (while there might be data stewards and owners of applications, critical enterprise data is treated as a shared enterprise resource and not an application-specific or function-specific resource)
- Competitive weapon (the organization manages data as an asset and leverages it for competitive advantage)

The overall notion is that organizations will achieve the greatest benefit from data if they are at the top right of the chart, where data is a competitive differentiator. This is accomplished by having a mature technology platform (factory) and a mature management practice (Lean Data Management). When you put these together, then solutions can be delivered better, faster, and cheaper (no compromise or trade-off is needed)—and the organization can respond to unexpected events or new opportunities with a high degree of agility.

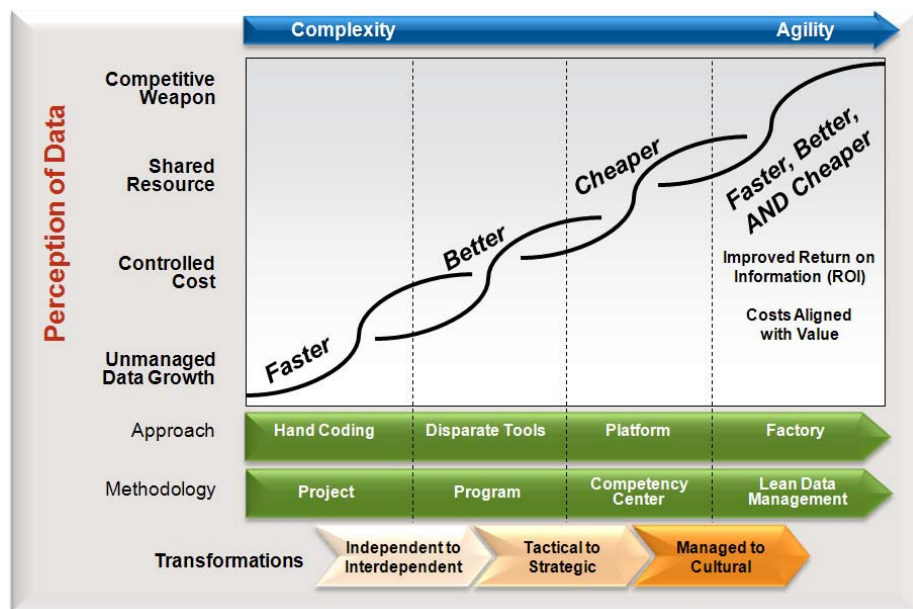


Figure 9: The evolution of data management technology and management methodology in an organization and the resulting perception of data that delivers the highest level of efficiency and agility to the enterprise

Informatica Lean Data Management Practices in Action

Hundreds of companies worldwide have relied on Informatica Lean Data Management practices to manage data growth in their enterprise. Lean Data Management and its three areas of focus—applications, data warehouses, and application portfolios—use the Informatica Application ILM product family as the underlying technology to accomplish its goals of maximizing operational efficiencies and lowering IT costs. These organizations have realized the benefits of Lean Data Management practices such as rapid return on investment (usually 6-9 months), significant IT cost reduction, greater profitability, and a more competitive edge in the marketplace. The following case studies demonstrate how the Lean Data Management practices and underlying technologies have helped three companies in each of these areas.

Lean Applications: A Case Study with IKON

The increasing demand for IKON Office Solutions document management products and services has fuelled data growth in the company's Oracle E-Business Suite applications. IKON's data growth further spiraled out of control when it upgraded to Oracle 11.5.9 when more data and functions were being managed in the application. The growth in its transaction-intensive modules, such as Order Management, Field Service, and Purchase Order, were growing at up to 2 terabytes per year. Taking into account copies of the production databases, the total database size was projected to soon reach 80 terabytes - a size beyond the company's current storage infrastructure.

IKON deployed Informatica Data Subset and Informatica Data Archive as part of its lean applications practices to manage data growth in both its production and non-production Oracle E-Business Suite environments to achieve the following:

- Reduced costs by \$1.5 million annually
- Realized a full ROI in only six months
- Saved 14 TB of storage capacity in non-production systems by creating data subsets and 4.8 TB in production systems by archiving inactive data
- Decreased the time it takes to perform a backup by 25 percent from eight hours to six hours
- Ensured that performance SLAs are met
- Replicated user-selected application data readily from production systems into lean, non-production systems
- Reduced the time and disk space required when compared with creating a full system/database copy
- Accelerated delivery cycles by decreasing test and development cycle time

Lean Data Warehouses: A Case Study with Bank of America

Bank of America's (BoFA) critical credit card services data warehouse was exploding with more than 170 TB of data and no visibility about how various businesses were using that data. User activities were generating 5 million queries a day, which, in conjunction with the sheer size of the warehouse, was adversely impacting query performance. Because it was using full size copies of production for testing, application testing and deployment were taking too long. The dilemma facing BoFA was how to classify the data warehouse data and get this runaway data growth under control in both production and non-production environments to achieve reasonable production performance and testing cycles.

As part of its lean data warehouse practices, the bank implemented Informatica Data Warehouse Advisor and archiving to monitor data usage in its data warehouse. It gained visibility regarding how data is used, who uses it, and what queries are executed to reduce the data size by archiving dormant data, optimize the data warehouse schema, improve performance, and optimize its resources against a backdrop of business growth. As a result, it was able to:

- Archive 13 TB in the first two months of production analysis
- Continue to archive data monthly with ongoing analysis
- Reduce the project timeline for delivering the data warehouse from six months to two weeks
- Improve performance by 25 percent by adding indexes based on the analysis and saved 12 weeks of effort
- Achieve a return on investment in the software in less than six months

Lean Applications and Lean Application Portfolios: A Case Study with EMC

During the past four years, EMC, the leading provider of information storage, has extended its core business and market to new areas by acquiring more than 40 software and services companies. As a result, it needed to retire 21 Oracle E-Business Suite and custom applications averaging more than 700 GB in size from these acquisitions.

In addition, EMC also has arguably one of the largest Oracle production applications in the world: an Oracle 11i customer relationship management (CRM) solution that provides 50,000 users with sales, service, marketing, and commissions management information. The Oracle E-Business Suite production application is 8 terabytes in size, but once the 19 other copies are taken into consideration, the total production and non-production database volumes account for approximately 350 terabytes of storage. The production environment was growing by 2 terabytes annually - or 50 terabytes annually once all the different non-production instances were included.

As part of its lean applications and lean application portfolio practices, EMC deployed Informatica Data Archive to cost-effectively manage growing data volumes in its Oracle E-Business Suite application and to retire its legacy applications, storing the data in its EMC Centera storage for long-term retention and compliance. To further manage data growth in its non-production environment, EMC is also planning to implement Informatica Data Subset.

As a result of implementing Informatica's lean applications and lean application portfolio practices, EMC was able to:

- Save \$3 million annually by archiving inactive data from its live Oracle E-Business Suite application
- Achieve payback within six months
- Increase performance by 50 percent
- Reduce data growth by 75 percent (2 TB to 500 GB in production, 50 TB to 12.5 TB in non-production)
- Reduce maintenance downtime by 20 percent
- Cut performance-tuning staff by 60 percent, from five to two
- Save an additional \$1 million annually by retiring legacy applications
- Anticipate savings of \$820,000 annually by deploying Informatica Data Subset in its non-production environments

Conclusion

Big Data growth impacts many facets of the organization. Transactional applications, data warehouses, and legacy and redundant applications experience the most data proliferation due to business growth and recent trends in mergers and acquisitions, IT consolidation, and modernization. The cost of Big Data growth in these areas has a profound impact not only on IT costs but also on the efficiency of business operations, agility, and response time to customers. Without adopting an enterprise strategy to proactively manage the Big Data challenges and their associated costs, organizations will find it difficult to proactively modernize their IT landscape. In addition, they will see their profitability, competitiveness, and efficiency continue to decline.

Informatica's Lean Data Management practices provide the necessary tools and technologies, along with best practices and methodologies, to solve the Big Data challenge. Informatica's approach to lean applications, lean data warehouses, and lean application portfolios helps organizations to organize and manage data growth in production and non-production environments and in transactional systems and data warehouses, as well as handle legacy and redundant applications that need to be retired. By implementing the three pillars of Lean Data Management, organizations can dramatically reduce costs, increase efficiencies, meet application SLAs, and as a result, respond to business demands more swiftly. Adopting Informatica's Lean Data Management practices will help organizations operate better, faster, and more profitably.

To get started with their Lean Data Management practices, organizations should consider analyzing the data growth trends in their enterprise applications and data warehouses using Informatica Data Growth Analyzer and identifying dormant records using Informatica Data Archive Dormant Data Analyzer. With the results from this analysis, you can focus your archive and subset strategies on the appropriate data sets. In addition, the Informatica ILM ROI Tool (see Figure 10) provides an ROI analysis that helps organizations make the business case for archiving live enterprise applications, managing test data more efficiently by creating subsets, masking sensitive data, and retiring legacy applications.

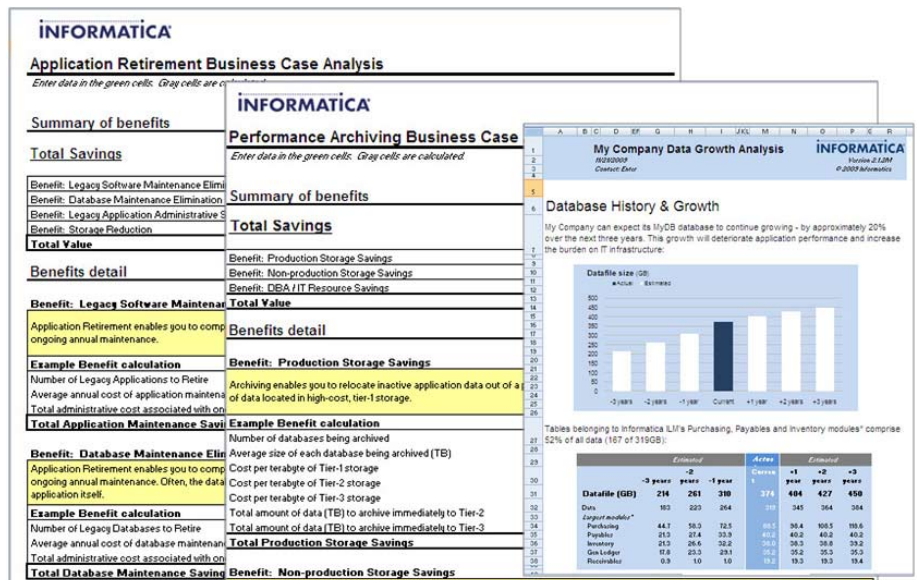


Figure 10: Identification of the largest application modules and tables by Informatica Data Growth Analyzer and the resulting ROI analysis help make the business case for implementing Lean Data Management practices

LEARN MORE

Learn more about the Informatica Platform. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.), or visit www.informatica.com.

ABOUT INFORMATICA

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world gain a competitive advantage in today's global information economy with trustworthy, actionable, and authoritative data for their top business imperatives. More than 4,200 enterprises worldwide rely on Informatica to access, integrate, and trust their information assets held in the traditional enterprise, off premise, and in the cloud.



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