Data Growth Paradigms in the 21st Century

Think Big and Archive Securely with IBM® InfoSphere™ Optim™ Solutions for System z®

TechTarget Custom Media

Data Growth Paradigms in the 21st Century

Table of Contents

Introduction 3
Your Evolving Business and Data Growth Requirements 4
What Technology and Storage Platforms Do You Use? 6
The Impact of Legislation and Regulations on Storage Strategies 7
Managing the Costs of and Risks to Your Data 8
Kryder's Law: It's Alive and Well 8
Thinking "Inside the Box" – How IBM Addresses the Data Growth Challenge 9
Going Forward – The Time is Right 10
How IBM® Gets You There 12
Benefits of These Advances 13
Sources 14
Conclusions 14

TechTarget Custom Media

Introduction

Few organizations today have all the answers when it comes to managing data growth. Typically, enterprises have vast amounts of data stored in a variety of locations. Some data, such as customer information or patient health records, is regularly used in production, while other data, such as historical financial records, is archived for occasional or future use. Overall, research firms like Gartner and IDC agree that data will grow exponentially in the coming years.

How important is this data to your organization? Chances are it's the lifeblood of your company – without it, the enterprise simply cannot function. When was the last time you examined your data growth and archiving activities? Increasingly, managing the growth of your data and archiving patterns is an important part of planning your technology investments and strategy, as well as ensuring compliance with regulations, standards and best practices.

Finally, how well are you optimizing your organization's data to improve system performance? Monitoring systems to optimize their performance is key. Data optimization not only ensures that your data is effectively managed, but that your systems are performing at their most efficient levels as well. These activities are necessary for you to achieve the greatest return on investment from your technology infrastructure.

Your Evolving Business and Data Growth Requirements

Organizations that use specialized enterprise-class systems for activities like manufacturing, research and development, payroll and human resources all face the challenges of managing increasingly larger volumes of data. Each organization must address several issues, including managing increasing costs, ensuring satisfactory application performance, and ensuring that data retention and compliance requirements are met.

From a cost perspective, the belief that storage is a low-cost remedy no longer holds. While it's an easy strategy to simply buy more storage (e.g., a few additional terabytes of capacity) for your enterprise-level systems, what about the costs to manage that additional capacity? What about data that is not being regularly used for daily production? What happens to that data? Does it sit on a storage system, taking up space that could otherwise be used by more frequently accessed data? And what if your investments in added storage result in the need for more database administrators and other staff to manage all that data? Quickly you'll discover that data growth carries with it additional responsibilities.

Data growth without proper management could result in reduced or degraded application performance. Too much data in production will inevitably soak up processor cycles that could be used to increase production application performance. The solution may not be as simple as adding more drives or database administrators to reconfigure and optimize databases. Are you prepared to invest more overhead for questionable performance improvement?

The above situation begs the question: What can be done with data that is not being used regularly in production systems? Where can it go so that it is readily available when needed but also protected from unauthorized access? The ideal solution is to retain and archive unused data safely in protected storage areas, according to established data retention and archival policies. Once the data is safely stored in a secure area, it can be made readily available for audits or other research and discovery activities.

Thus, the true cost of data storage has many components – there is much more to consider than just disk storage. Aside from being a potential threat to your budget, the results of not

dealing with storage issues proactively and prudently can include poor employee performance, an inability to respond to customer queries quickly and unsatisfactory audit findings.

Figure 1 depicts an information supply chain addressing a typical data management life cycle found in many organizations. This supply chain of information flows throughout an organization. Unlike a traditional supply chain, an information supply chain is a many-to-many relationship. In an information supply chain, the same data,, the same data about a person can come from many places. For example, he or she may be a customer, an employee or a partner, and information about that person may appear in many different reports and applications. Different systems may subsequently redefine the information. This means that integrating information, ensuring its quality and interpreting it correctly are crucial when using the information to make better decisions. Information must be turned into a trusted asset and governed carefully to maintain quality over its life cycle. Such requirements make data archiving core to information life-cycle management and a key component of an organization's information governance strategy.

This supply chain of information flows throughout an organization. Unlike a traditional supply chain, an information supply chain is a many-to-many relationship. In an information supply chain, the same data,, the same data about a person can come from many places. For example, he or she may be a customer, an employee or a partner, and information about that person may appear in many different reports and applications. Different systems may sub-

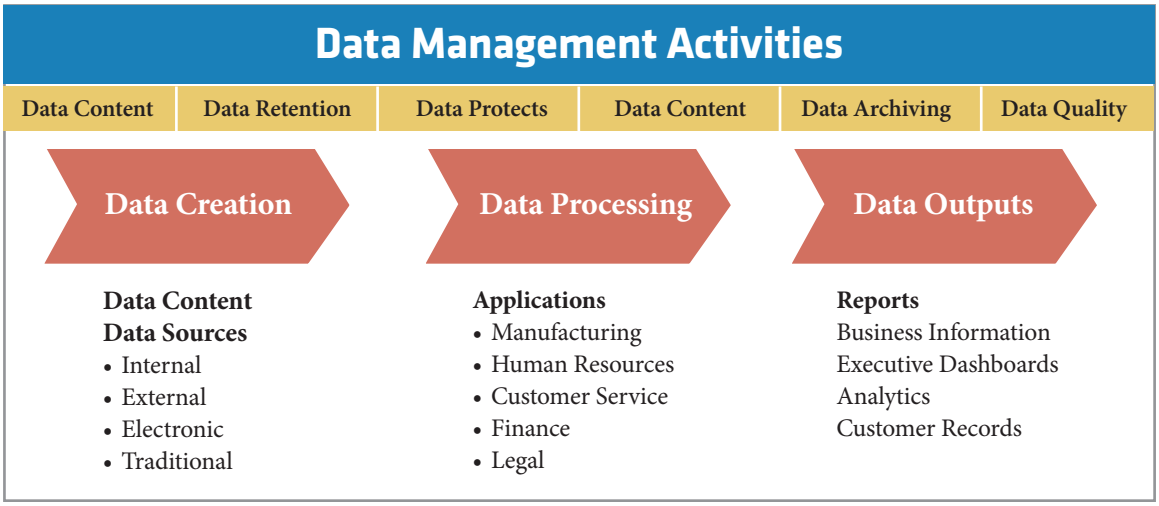

Figure 1

sequently redefine the information. This means that integrating information, ensuring its quality and interpreting it correctly are crucial when using the information to make better decisions. Information must be turned into a trusted asset and governed carefully to maintain quality over its life cycle. Such requirements make data archiving core to information life-cycle management and a key component of an organization's information governance strategy.

The systems that achieve these objectives must be cost effective and easy to maintain, and they must perform well for the workloads they need to handle, even as information volumes continue growing at extraordinary rates.

What Technology and Storage Platforms Do You Use?

In most organizations, there may be hundreds or even thousands of different systems. Information can enter these systems from many sources (e.g., transaction systems, operational systems, external information sources) and in many formats (e.g., data, content, streaming). Wherever data comes from, there are important relationships across the many data sources. Ultimately, organizations manage all this information in their systems using a structured and coordinated process.

Enterprises are generally made up of many technologies and platforms:

- **Storage platforms** – These include online, near-line and offline systems. Depending on where the data is in its life cycle, it may be stored on multiple platforms and be related across each of them.
- **Operating systems** – Most organizations have data stored across multiple platforms and operating systems, such as IBM® z/OS, Linux, Unix and Windows.
- **Database management systems** – Most organizations have one or more database management systems with data running on the same database platform supported by multiple operating systems. However, assuming there are different DBMSs, it is essential that data life-cycle solutions support heterogeneous environments.
- **Specialized applications** – Many organizations adopt a combination of ERP and CRM systems, such as SAP, PeopleSoft and Siebel, along with their own custom applications. However, systems are not stand-alone – they are often integrated and share information, which means they must be managed in a consistent manner.

- **Advanced data management** – In addition to requiring fundamental storage and archiving processes, data growth has encouraged the development of specialized capabilities such as discovery and application retirement.

A prudent objective is to have a single, scalable, heterogeneous information life-cycle management solution that includes a data retention policy. The policy should state that data must be archived and stored in a safe location, where it cannot be damaged. The policy should also specify the length of time the information is to be archived, provide guidance for destroying the information when time limits have been exceeded and detail provisions for handling the data during litigation.

The Impact of Legislation and Regulations on Storage Strategies

Today, it's not enough to simply establish your own policies and procedures for data storage and management. The need to comply with data retention regulations specified in legislation including Sarbanes-Oxley, Gramm-Leach-Bliley and HIPAA has formalized the need to retain data long after its use.

Speaking of data retention, an ideal data archival and retrieval capability would enable data to be retrieved for litigation and legal holds, for example, even if the original applications no longer exist. Additional benefits of a data archiving capability include lower costs by removing data from production systems, lower risks by keeping all data archived in a safe location, and increased performance of existing applications.

Your data archiving program should make it possible to archive data using a suitable space-saving and efficient format, store associated metadata and business objects as well as the data, reduce the need to reintroduce archived data to a production database, and provide all the associated output from the data. From a legal perspective, you now should access to your archived data without the need to use a DBMS.

Managing the Costs of and Risks to Your Data

The cost of managing storage can be three to 10 times the cost of procuring it, according to IDC, which estimated that in 2010, organizations spent an average of about $1.1 billion on storage.

Experience has shown that database administrators often spend up to 80% of their time each week on hardware capacity-related performance issues. One IBM® customer had about 19,000 batch processes that took 250 hours to run, equating to more than 10 days. Archiving helped the customer reduce that time by over 75%.

One key issue to address is how long your organization should keep data in its production systems. According to IDC, at least 50% of companies are keeping data in production for seven or more years. Some 57% of companies are leveraging their backup copies for what they call "data retention." That's a lot of data to sift through if you have an e-discovery request.

As organizations retain greater volumes of data in an attempt to stay ahead of their competition, a more proactive approach to data management is needed to ensure that their data remains accessible and trusted.

Kryder's Law: It's Alive and Well

Most people who deal with technology are familiar with Moore's Law, which states that the number of transistors that can be embedded in an integrated circuit doubles approximately every two years (updated to18 months today).

In the world of storage, there is an equivalent known as Kryder's Law, based on the work of Mark Kryder, a former Seagate Technology executive. Kryder stated that the density of information on hard drives increases by a factor of 1,000 every 10.5 years, meaning storage density doubles every 13 months. In a 2009 study, Kryder proposed that if hard drives continue to progress at their current pace, in 2020, a two-platter, 2.5-inch disk drive will be capable of storing more than 14 terabytes of data – at a cost of about $40.

Recognizing that current storage technology has not achieve such performance predictions – annual improvement in hard drive density in the past several years has been closer to 30 to 40 percent – other approaches to storing and managing data are needed.

Thinking "Inside the Box" – How IBM® Addresses the Data Growth Challenge

While there are many platforms containing data strewn across numerous databases, IBM® estimates that 95% of Fortune 1000 companies store their data on the IBM® System z® platform. According to literature and reports on System z®, the platform offers very high scalability, dynamic workload balancing, security and quality-of-service performance. System z® gives customers the ability to consolidate and manage data on hundreds and thousands of Linux-based servers (Linux on z/IFL LPARs) as well as data on AIX on Power, Linux on System x and Microsoft Windows applications using zBX blades. The System z® Enterprise Unified Resource Manager manages all of these environments, making it a cost-effective platform on which to store and manage your data. From a software perspective, IBM's® InfoSphere™ Optim™ portfolio of solutions addresses cross-platform, cross-database and cross-application requirements. According to Gartner, IBM® is the market leader for data archiving from both revenue and the number of customers in 2010. InfoSphere™ Optim™ solutions for the System z® platform help clients manage data growth through its lifecycle. Let's examine how this is done.

Locating data scattered across the enterprise – You need to know how the data you want relates to other enterprise data and where the hidden relationships are located. IBM's® InfoSphere™ Optim™ Discovery locates the data and its relationships as information comes into the enterprise.

Data growth management – The daily challenges of managing the data life cycle are intensified by the growth of data volumes. IBM's® InfoSphere™ Optim™ Data Growth solutions for System z® provide intelligent archiving techniques so that infrequently accessed data does not impede application performance but is still readily accessible. IBM® InfoSphere™ Optim™ solutions for z/OS provide a Data Growth module to reduce hardware, storage and maintenance costs. Two key benefits are the ability to archive complete business objects across multiple systems and support for distributed platforms and multiple databases.

Application retirement – While many enterprise applications need to be upgraded, consolidated and eventually retired, the data may not face the same fate. Many organizations today must still deal with redundant or legacy applications. Leveraging the InfoSphere™ Optim™ solution for application retirement with archiving best practices, clients can ensure application-independent access to business-critical archived data for long-term data retention, long after an application's life expectancy.

When you implement a system, your hardware requirements will be based on the current and future performance requirements of the business. As the amount of data increases, system performance may slowly decrease. Many businesses address this by investing more hardware to return performance to an acceptable level. But this is typically a cycle of more data, more hardware, more data, more hardware. Then what happens? Fortunately, there are alternatives to this methodology.

For one, archiving becomes a strategic activity, whereby inactive application data from current production systems is segregated and safely moved to a secure archive. Your databases then reclaim capacity, which improves application performance and availability. IBM's® InfoSphere™ Optim™ Data Growth and application archiving solutions for the System z® platform establish distinct service levels for each class of application data, such as current data, reporting data and historical data. Policy-driven archive processes let you specify the business rules for archiving, such as archiving all closed orders that are two or more years old. The IBM® InfoSphere™ Optim™ Data Growth solution can typically compress archived data by 50 percent to 90 percent, thereby making it even more attractive to archive inactive data.

Going Forward – The Time is Right

Before you launch a data growth management initiative, you must discover and define where the data in question lives and how it's all related across your many applications as well as the enterprise.

The following steps will help you address your data growth challenges:

1. Examine your current storage infrastructure, including costs and risks – Determine how well your current storage environment accommodates your data storage requirements.

2. Review your storage, data retention and archiving policies – If your analysis shows no serious inconsistency between data growth and existing storage resources, your current policies and resources may be sufficient. If it becomes clear that data growth is currently outstripping your resources and will do so in the future, your policies may need to be updated.

3. Align your storage infrastructure with your business needs – It's very important to synchronize your storage resources with your current and future business needs. This should be an ongoing activity, especially if your organization is experiencing strong growth.

4. Ensure that your data archiving processes are optimized – If you determine that older data that must still be retained is not appropriate in your production systems, update your archiving procedures to accommodate what may be a new or different paradigm for your storage policies.

5. Ensure that your storage infrastructure improves application performance – If you can relocate older and unused data to a safe archival site, you should begin to see performance improvements. You must be able to archive all relevant aspects of the older data, including metadata, which is data that provides descriptive information about one or more aspects of the data in question.

Additional considerations

In their role as the building blocks of data records, business objects can be a variety of elements, ranging from customer records to tables in a database to metadata. Inclusion of business objects ensures that the records being analyzed are valid and confirm that a transaction occurred, and that the integrity of all related data is maintained.

Another important aspect of data capture is federated object support, which means capturing a complete business object from multiple related applications, databases and platforms. Federated data capture ensures that your data management operations accurately reflect a complete, end-to-end business process.

When archiving data, is it acceptable to consume the same amount of storage that was used in the production systems, or is there a need to significantly reduce the storage space use for the archived data?

Beyond technological concerns, consider the financial implications of a data growth management program. For example, a Toshiba subsidiary increased application availability by archiving historical transactions using IBM's® Optim™ Data Growth solution for Oracle.

When considering a data growth management program, be sure to factor in your current operating costs for data storage and management, available budgets for data growth and potential savings that can be realized from increased application performance with minimal change in production environments.

How IBM® Gets You There

IBM® can help your organization with data growth management with its System z® platform. It enables you to manage data growth without having to add more and more physical boxes. InfoSphere™ Optim™ Data Growth solutions for the System z® platform can help you in several ways:

Reduce Costs: Infrequently used data can be archived from production environments and stored on less expensive Tier 2 storage – either databases or efficiently compressed flat files to save even more storage space – yet still be accessible for legal purposes.

Improve Performance: With less data, your applications will perform better, searches and batch processes will run faster, and backup processes will run more efficiently. Less data to convert to the upgraded version also means less time the application is offline.

Minimize Risks: Not only does intelligently archiving data out of production systems allow for data retention compliance, but it also supports a better long-term solution for storing the application data, providing application-independent access to that data.

The InfoSphere™ Optim™ Data Growth solution for the System z® and other platforms:

- Helps you identify and locate data records for archiving from across a varied IT environment
- Quickly and efficiently archives data while maintaining the original business context when the data is stored
- Ensures that data retention policies are implemented across the enterprise
- Facilitates archival data access using a variety of access protocols
- Is compatible with a variety of packaged and custom enterprise-level systems in a diversified environment

Benefits of These Advances

The InfoSphere™ Optim™ Data Growth solution for System z® solves your data growth challenges at the source – by managing your enterprise application data. InfoSphere™ Optim™ helps you archive historical transaction records, which controls data growth and improves application performance. Historical data is archived securely and cost effectively, and can be easily accessed for analysis, audit or e-discovery requests. With less data to analyze, reporting speeds are reduced and mission-critical business processes are completed on time. Having a defined policy for managing the retention requirements for historical data is a requirement for enterprise governance frameworks to ensure compliance with regulatory mandates.

InfoSphere™ Optim™ manages application data at the business object level, preserving both the relational integrity of the data and its original business context. Each archived record represents a historical reference snapshot of business activity, regardless of its originating application. In cases where the originating application has been retired or is no longer available, InfoSphere™ Optim™ offers application-independent access to archived transactions.

Benefits of Using InfoSphere™ Optim™
Reduces data storage and archiving costs by up to 50%
Streamlines application upgrades by reducing data conversion time – up to 66%
Increases application uptime by up to 30%
Results in total cost efficiency from decommissioning legacy applications of approximately $135,000 per year
Reduces Tier 1 storage costs by leveraging Tier 2 storage for archived data

Sources:

Forrester TEI Study for InfoSphere™ Optim™

Forrester Research, "The ROI of Packaged Apps Instance Consolidation," May 2008

IBM: http://www.ibm.com/software/data/optim/manage-data-growth/

IBM: http://www.ibm.com/software/data/db2imstools/solutions/data-governance.html

Conclusion

Recognizing that your organization is likely to face data growth issues in the coming years, the time is right to evolve your data management activities to respond to this exponential data growth and manage it well. IBM's® InfoSphere™ Optim™ Data Growth solution for System z®, along with its solutions for application retirement and data discovery, can make the difference by increasing the performance of applications and business processes that consume data, lowering operational costs through more controlled storage use and reducing risk exposure by ensuring compliance with industry and government regulations concerning data retention, access and availability.