White Paper
September 2009

IBM Information Management software

Control application data growth before it controls your business

Contents

Executive summary

Information is the lifeblood of business. Corporate strategy and planning, operations, decision-making, marketing, selling, customer satisfaction and business growth all hinge on accurate, available information. In their quest for this accuracy and availability, in the 1990s and early part of this century, companies invested significantly in enterprise resource planning (ERP), customer relationship management (CRM) and other applications to collect information from a variety of sources to support business operations. This information has been extremely valuable over the years.

As time has passed, the information collected has grown by leaps and bounds. Business growth, mergers and acquisitions, data retention requirements and the Data Multiplier Effect have resulted in a complex data landscape characterized by massive data volume and hidden relationships.

Like any other asset, data must be well managed if it is to continue to deliver business value. This is becoming more difficult given unchecked data growth and distribution. What might once have been viewed as an effective strategy for managing enterprise application data, however, is probably no longer able to protect organizations from the negative effects of runaway data growth.

What are these effects? The most obvious are complexity and risk, along with storage and management costs. Data that has accumulated at a significant rate is highly distributed over multiple applications, databases and platforms creating complicated data relationships that are difficult to define, understand, control and house. Also, managing years of historical data in production databases can impact service levels and disaster recovery initiatives. Expanding online and batch processing windows, along with routine maintenance tasks, takes much more time and can reduce application availability significantly.

This white paper describes some of the ways that organizations can address data growth issues with the power of integrated data management. Integrated data management helps support your information governance initiatives, simplify your IT infrastructure, support business continuity and increase the business value of your enterprise applications.

Companies need capabilities for identifying data assets and relationships, assessing data growth and implementing tiered storage strategies – capabilities that integrated data management can provide. It is important to classify enterprise data, understand data relationships and define service levels. Database archiving has proven effective in managing continued application data growth especially when it is combined with data discovery.

In particular, the combination of IBM® InfoSphere™ Discovery and the IBM® Optim™ Data Growth Solution provides proven capabilities for classifying, controlling and managing enterprise data cost-effectively throughout the data life cycle. With the capabilities that IBM InfoSphere and IBM Optim offer for aligning integrated data management with business objectives, your organization will be better positioned to archive data effectively, improve service levels, reduce risk and control costs.

What has driven application data growth?

Many factors have caused enterprise application data to accumulate at an almost unbelievable rate. And this growth is expected to continue. IDC estimates that 45 GB of data currently exists for each person on the planet: that's a mind-blowing 281 billion gigabytes in total. While a mere 5 percent of that data will end up in enterprise data servers, it is forecast to grow at a staggering 60 percent per year, resulting in 14 exabytes of corporate data by 2011[1]. So, what are some of the underlying factors that have driven application data growth?

Organic business growth

Organic business growth is the growth that occurs when companies increase output and enhance sales, bringing about profits generated in the company. This growth also generates data—lots of it—as a natural part of the daily business activities related to heightened productivity and more sales. Data from ERP and CRM systems plays a big role in this growth, but new business intelligence, planning, analysis and performance management applications are also significant players in the data explosion. Put all of these factors together, and it is quite possible that your organization's systems are collecting millions of new transactions with countless new relationships every day!

Mergers and acquisitions

Mergers and acquisitions have also driven data growth. Over the past decade, businesses increasingly have been using mergers and acquisitions to expand and increase market share. As a result, the organization inherits the applications of the acquired company—and all of the associated enterprise data.

Data Multiplier Effect

Data duplication has significantly contributed to growth statistics. Cloning or copying a production database to support various other functions in an organization or for application development and testing initiatives is a common business practice. Organizations maintain several backup copies of critical data or implement mirrored databases to provide assurance against data loss. Then there are the disaster recovery plans that require data duplication to store critical data in an alternate location. All of this duplication has created what is known as the "Data Multiplier Effect."

As data is duplicated, storage and maintenance costs increase proportionally. The Data Multiplier Effect is the result of multiplying every gigabyte (GB) of data in a production database by the number of replicated copies. The resulting figure represents the company's total data burden—and it is usually a higher number than managers expect and often not even calculated as part of data growth measures because it is not "new."

For example, suppose that a modestly sized production database contains 200 GB of data. When the production database is copied for backup, disaster recovery, testing, development and quality control, the total data burden increases from 200 GB to 1200 GB, or 1.2 terabytes (TB). This is illustrated in Figure 1.

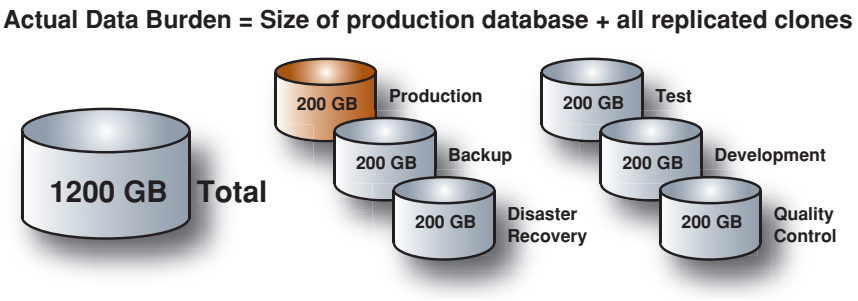

Figure 1. The Data Multiplier Effect is the result of multiplying every GB of data in a production environment by the number of replicated copies.

Data retention regulations

Compounding data growth challenges are the global data retention regulations that require the management and storage of different types of data for extended periods of time. Corporate policy or government regulations often require that data remain easily accessible for specified periods of time and sometimes specify disposal requirements. Some of these policies and laws are intended to preserve the accuracy and transparency of business records to prevent accounting scandals. Companies might need to protect corporate interests by retrieving historical records to satisfy audit inquiries and resolve claims. Retention requirements carry strict penalties for non-compliance.

In the U.S., some of the most commonly recognized regulations are HIPAA (the Health Insurance Portability and Accountability Act), which requires that healthcare organizations retain original medical records for a minimum of five years and in many cases, decades, and the Sarbanes-Oxley Act, which requires that corporate accountants retain specific records for a minimum of five years. There are many others—in 2008, there were 10,000 federal, state and local regulations governing the preservation of data—and that was before the financial crisis. More are sure to follow.

These regulations are designed to specify the appropriate standards of care for managing and retaining corporate information. Organizations are required to maintain critical data, keep it accessible and satisfy the demand for documented authenticity and accuracy of their records. This has contributed significantly to data growth.

What are the consequences of uncontrolled data growth?

Your applications are acquiring more data than ever before. On top of years of collecting data from transaction-intensive and customer-facing applications, you now have business intelligence and business analytics applications adding to the accumulation and creating more and new data relationships. Because you rely on these applications to support your business operations, managing an ever-increasingly complex data landscape is absolutely necessary for controlling costs, improving customer satisfaction and enhancing decision making. Otherwise, your business operations, your profitability and your ability to maintain a competitive advantage will be adversely affected.

The potential consequences of uncontrolled data growth can affect all areas of your enterprise, making it difficult for you to complete critical business processes and meet your business objectives. Without an effective strategy for managing enterprise application data, it becomes more difficult to address critical business issues, including meeting information governance requirements, controlling IT infrastructure costs and ensuring business continuity.

Increased IT complexity, risks and costs

Accumulating vast amounts of application data adds complexity, risk and cost to your business. As the number of applications that use and access data grows into "enterprise systems," a complicated data ecosystem is created with numerous connections and relationships, many of which are hidden or not understood. This has created a whole new level of complexity and created an additional challenge for managing data—if you don't understand your data, you can't control it.

In addition, unchecked data growth slows transaction processing and increases the time it takes to perform routine maintenance. An overburdened IT infrastructure trying to access data that is widely distributed in applications, databases, systems and platforms places a strain on capacity, slows operations and jeopardizes customer satisfaction. Boosting processing capacity and tuning the database can provide some relief, but only temporarily.

The cost of storing your data also increases substantially. The Data Multiplier Effect is in full force. Not only must you expand storage capacity for transaction processing and business intelligence, but you must also increase capacity for all cloned environments. Purchasing additional storage in general is a small percentage of the storage costs. Managing the storage is where the real money is spent and it leaves many companies in a perpetual game of catch-up. An organization can spend millions of dollars annually on overall storage costs. None of which addresses the root cause of the problem—rapid, uncontrolled data growth.

Diminished service levels

Maintaining databases that contain years of historical data spread all over your organization slows response time and impedes access to current information. With too much information to sift through, routine reports take longer to run. Business analysis queries also require more time to complete, limiting your ability to make accurate, timely decisions.

Continued uncontrolled data growth affects functional processes, such as financial period closeouts or delivery scheduling. Service levels decline, eroding the customer loyalty you have worked so hard to develop. Backup and recovery windows are stretched to the point where system availability is seriously threatened. Upgrade and migration projects are more likely to cause costly business disruptions. Without a method for managing ongoing data growth, it becomes increasingly difficult to achieve your business objectives.

Inadequate disaster recovery

Disaster recovery is a high organization priority especially in the wake of the man-made and natural disasters of this decade. When disaster strikes, the key strategy is to get your most important systems operational as quickly as possible. But when you have overloaded databases, all of the data (both current and historical) must be restored just to resume processing for today's transactions. When historical information is stored with current information, the time to complete the recovery process can increase by hours or even days.

How can you take control of application data growth?

Your enterprise applications simply must deliver measurable business value. So, how can you manage data growth and capitalize on your investment? The answer rests in understanding your data and its relationships and then aligning continuous control of your application data with your business objectives. Both are possible with integrated data management.

Integrated data management aligns application data management with business objectives, helping to optimize performance, control costs and reduce risks. It includes data discovery and archiving capabilities, so you can identify data and its relationships in disparate systems, segregate historical from current data and store it securely and cost-effectively.

The following capabilities are necessary for effective integrated data management:

- **Assess application data growth and tiered storage strategies.** With proven capabilities for assessing your data, you can easily identify the applications where data is accumulating the fastest. Gaining a complete understanding of which areas are accumulating the most information allows you to apply the most effective storage strategy. Assessing your data enables you to be proactive with integrated data management so you can identify and address potential problems before they impair business results.
- **Identify data assets and discover data relationships.** You cannot manage data if you do not understand it, so it is critical that you first document your existing data landscape using data discovery. Data discovery analyzes data values and data patterns to identify the relationships that link disparate data elements into logical units of information, or "business objects" (such as customer, patient or invoice). These business objects provide essential input for archiving.
- **Classify data and define service levels.** After you have discovered and analyzed your data relationships and business objects, you can classify each type of data record or business object and define the appropriate level of performance for each type. For example, current transactions are the highest priority, often requiring sub-second throughput. Data with a lower service requirement can be safely moved to a somewhat slower, but less expensive environment.

- **Initiate archiving as a best practice.** Archiving capabilities are an essential part of an integrated data management solution for controlling data growth. Database archiving enables you to segregate and remove historical data safely from the production environment, freeing valuable capacity for priority business needs. When you implement a policy-driven archiving strategy, based on the age, usage and status of the data, you are able to manage each class of data according to its unique service requirements.

- **Store enterprise data cost-effectively.** Storing archived data according to its evolving business value is a logical component of an integrated data management strategy. Take three-tier classification, for example. Current transactions are maintained in high-speed, primary storage. Reporting data is relocated to mid-tier storage. Reference data is retained on a secure WORM (Write Once, Read Many) device, keeping it available if an audit request should arise. When you implement a tiered storage strategy, you can reclaim capacity and maximize the value of your existing storage infrastructure. Storage alone is only part of the answer. The combination of an archiving and storage strategy ensures that the right data is in the right place, retaining access to the information in a way that will comply with regulatory requirements.

- **Promote data accessibility.** Decision makers must have access to data, whether it is current or historical. With a comprehensive integrated data management solution, those decision makers can access the right information at the right time. Authorized business users must be able to query and browse all active, inactive and reference data. Reliable access makes it possible to generate reports and respond quickly to audit and discovery requests. If additional business processing becomes a requirement, you must be able to restore archived transactions in bulk or individually as required.

- Manage data retirement. It is imperative that organizations control the retirement process to help ensure that data is eliminated at the end of the required retention period. Effective integrated data management makes it possible to automate both retention and retirement, while still offering the capability to remove data manually. Applying suitable and secure methods for data disposal allows you to prevent your information assets from becoming liabilities.

IBM delivers proven integrated data management capabilities

You invest millions of dollars in your enterprise applications and the supporting infrastructure to promote optimal operating performance, improve decision-making and gain a competitive advantage. IBM provides integrated data management solutions that can help you derive the most business value for your enterprise.

IBM InfoSphere Discovery

IBM InfoSphere Discovery prepares you with the information you need about your data so that you are ready for archiving. InfoSphere Discovery automates the identification and definition of data relationships in the complex, heterogeneous environments prevalent in IT today. Covering every kind of data relationship—from simple to complex—InfoSphere Discovery provides a 360° view of your data assets by analyzing the data itself.

Without an automated process to identify data relationships and define business objects, organizations can spend months performing manual analysis—with no assurance of completeness or accuracy. InfoSphere Discovery, on the other hand, automatically and accurately identifies relationships and defines business objects in a fraction of the time required using manual or profiling approaches.

Standard relationships are defined directly to your database catalog. In more complicated cases, however, data relationships are not clearly visible. Application-defined relationships, for instance, are not contained in the database catalog but instead are enforced through the processing logic of the application itself. In other situations, complex business rules might have been applied to transform source data as it was moved to a target system.

InfoSphere Discovery analyzes the data values and patterns from one or more sources to capture these hidden relationships and bring them clearly into view. Applying heuristics and sophisticated algorigthms, it performs a full range of data analysis techniques: single-source and cross-source data overlap and relationship analysis, advanced matching key discovery, transformation logic discovery and more. It accomodates the widest range of enterprise data sources: relational databases, hierarchical databases and any structured data source represented in text file format.

Because it provides accurate representation of business objects to be archived, InfoSphere Discovery is the perfect complement to the IBM Optim Data Growth Solution. For example, it identifies both database-defined and application-managed

relationships for the archive business object. In addition, it automatically groups database tables into business objects (Figure 2) that form the basis of Optim Data Growth service definitions and requests.

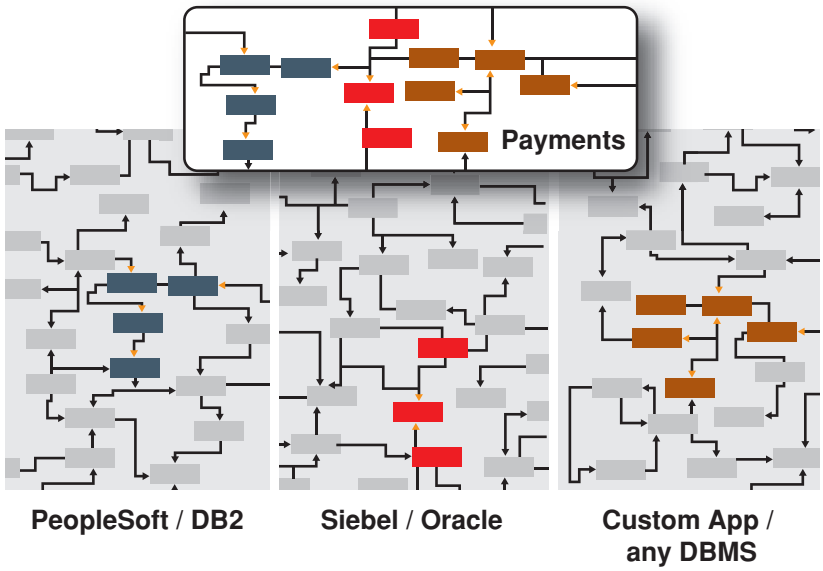

Figure 2. InfoSphere Discovery automates the grouping of tables into business objects that form the basis for IBM Optim service definitions and requests.

When you use InfoSphere Discovery and Optim Data Growth together, your organization can quickly and easily implement best-practice archiving capabilities, reducing storage requirements and delivering substantial cost savings.

IBM Optim Data Growth

The IBM Optim Data Growth solution is a scalable enterprise solution that can help meet your evolving needs. From small to large organizations, from single applications to global business centers, data management is streamlined using a consistent proven archiving strategy.

Optim Data Growth helps you manage application data growth in your enterprise, which, in turn, can improve application service levels, mitigate risk and control costs. With Optim Data Growth, you can:

- Simplify integrated data management to accelerate business-critical projects.
- Provide open access to current and archived data.
- Complete easier upgrades.
- Reduce storage management costs.
- Profit from superior application performance and availability.

By streamlining your application databases, Optim Data Growth provides what you need to deliver exceptional service to business units and profit from the superior performance and availability. Key business processes that are imperative to your organization are completed on time. And with continuous access to your data, decision makers get the information they need, when and how they need it.

Optim Data Growth also helps organizations control the costs of managing data growth. It simplifies complex infrastructures so that key IT processes require fewer resources (Figure 3). With Optim Data Growth, you can reclaim storage capacity, reduce the cost of maintaining storage and take advantage of a cost-effective data

growth management strategy. Optim Data Growth supports all major enterprise databases, operating systems and the most popular ERP and CRM applications in use today, along with your custom applications.

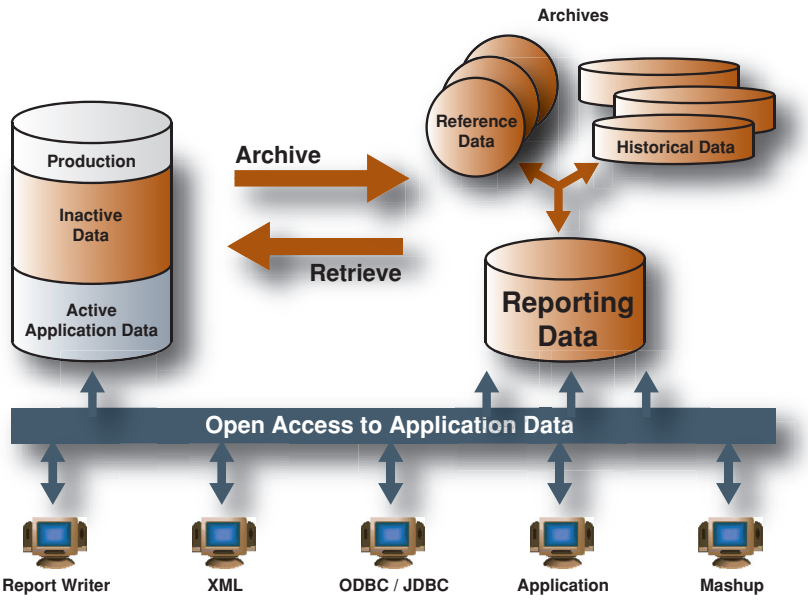

Figure 3. With Optim Data Growth, you can segregate inactive application data from current activity, safely remove it to a secure archive and keep it accessible with or without the use of the original application.

Managing enterprise data throughout its life cycle with IBM solutions

With InfoSphere Discovery and Optim Data Growth, IBM provides capabilities for controlling your application data from requirements to retirement. By implementing a proven integrated data management strategy, you take command of your mission-critical data throughout its entire life cycle and realize measurable benefits for your organization:

- **Discover.** Before archiving and storage initiatives can begin, you must first have a deep understanding of your existing data assets. Most organizations execute their business processes over multiple, interrelated applications, databases and platforms. And without explicit information to describe how data elements are related across these systems, managers have limited insight into critical information assets. InfoSphere Discovery automates the complex process of discovery, providing broader coverage and greater accuracy than manual data inspection methods could begin to offer. It uses sophisticated data analysis techniques to capture both database-defined and application-managed relationships that make up the business object so that it is ready for archiving.
- **Archive.** Database archiving is a recognized best practice. Once separated from current activity, inactive application data is safely moved to a secure archive. Optim Data Growth manages your data at the transaction, or business object level—such as service orders, payments or shipments—enabling you to define the policies and criteria for moving these transactions into an archive. With ongoing archiving, you can manage continued data growth, reduce costs, simplify maintenance, speed disaster recovery and ultimately reverse the Data Multiplier Effect.
- **Store.** With Optim Data Growth, you determine the appropriate storage location for each class of application data, based on its business value and access requirements. Utilizing a tiered storage strategy, you reclaim capacity by storing only current transactions in the high-performance production environment. Your reference and reporting data can be stored safely in near-line or offline storage. Tiered storage strategies organize your data cost-effectively, based on its business value, and help you manage your rapid data growth.

- Access. Optim Data Growth delivers access to the information you need, when and how you need it. With capabilities to query, browse and generate standard or custom reports, you can respond quickly and accurately to audit or discovery requests. Optim Data Growth supports a broad range of methods to access archived data. Application-based access offers a consolidated view of current and historical information with the existing application interface. With "self-help" data access, users can continue to rely on existing skills and spreadsheet and reporting tools. Application-independent access utilizes industry standards, methods and reporting tools to access archived business transactions without impairing online transaction processing performance and enables application retirement efforts.

- **Retire.** A comprehensive retention compliance strategy must include a means for disposal. With Optim Data Growth, you can control and automate data retirement when retention periods expire. You minimize the risks associated with retaining records longer than the stipulated time period. Whether you choose to automate the process for increased efficiency or manually select archived transactions for retirements, you can retain only necessary information in primary databases and archives.

Managing data growth – real world examples

Many clients have realized measurable benefits from implementing InfoSphere Discovery and Optim Data Growth to effectively manage enterprise data. One client, one of the world's foremost credit card companies, had watched its reward program database grow from a single database to 50 databases over the course of many years. As a result, global changes to the reward program now must span all 50 databases, making updates and changes extremely difficult and slow, if not impossible.

The company deployed InfoSphere Discovery software to accelerate the process of data map discovery and to automate the discovery of the complex relationships and transformations that exist between the older systems and the master data system. Using Discovery reduced the time for the first set of mappings from an estimated time of 560 person-hours to only 54 person-hours and reduced elapsed time from over 26 weeks to only 2 weeks of work.

For another client in the construction, mining and services industry, data growth was negatively affecting the performance of its ERP system. Response time for online transaction processing had declined noticeably, and longer batch processing windows made it more difficult to meet service levels for month-end close and other financial business processes. These issues were having a serious impact on the success of the business. Compounding the problem was a planned upgrade to the latest version of the ERP system. With Optim Data Growth, the company was able to reduce the database size and accommodate target goals for completing the upgrade and production cutover in a single weekend. Improved application service levels now support business users and daily operations and fully integrated archiving capabilities aligned with business policies manage application data.

Similarly, the European subsidiary of a global client that manufacturers electronic and high-technology products for personal and institutional use needed to integrate and consolidate data and processes with other European entities to improve service levels and operational efficiencies. Using the Optim Data Growth solution, the client reduced database size by 30% and increased application availability by archiving historical transactions to shorten time to complete 19,000 daily batch processes by 75%.

Conclusion: Manage enterprise data throughout its life cycle with IBM solutions

With the InfoSphere Discovery and Optim Data Growth solutions, IBM provides capabilities for controlling your application data from requirements to retirement. By implementing a proven integrated data management strategy, you take command of your mission-critical data throughout its entire life cycle and realize measurable benefits for your entire organization:

- Reduction of storage and maintenance costs
- Accelerated time to value for critical initiatives like data integration, governance and archiving
- The defining of logical groupings of data that serve as essential archiving inputs
- The satisfaction of information regulatory requirements
- Simplified IT infrastructure
- Support for business continuity
- Increased business value for your enterprise applications

About IBM InfoSphere solutions

Business optimization requires information that is accurate, complete, in context and actionable. Achieving this level of trusted business information requires transforming, reconciling and maintaining information and delivering it in real time to the people, processes and applications that need it. IBM InfoSphere products offer the breadth of capabilities required for complete management and delivery of trusted business information to ensure competitive success.

About IBM Optim integrated data management solutions

IBM Optim Integrated Data Management Solutions offer proven, integrated capabilities to manage enterprise application data from requirements to retirement. With IBM Optim solutions, teams can share data artifacts (like models, policies and metadata) to align data management with business goals and improve collaboration. Today, organizations of all types use IBM Optim solutions to improve performance, streamline database administration, speed application development and enable effective governance. IBM Optim integrated data management solutions deliver better business outcomes, at lower cost, with less risk, while providing capabilities that scale across enterprise applications, databases and platforms.

For more information

InfoSphere Discovery complements the value of IBM Information Management solutions, including the IBM Infosphere and IBM Optim product portfolios. To learn more, contact your IBM sales representative or visit the following Web sites:

ibm.com/software/data/ips

ibm.com/software/data/optim-solutions

IBM Software Group
Route 100
Somers NY 10589
U.S.A.

Produced in the United States of America
08-09

Endnotes

1 http://www.information-management.com/infodirect/2009_129/data_management_archiving_storage_disaster_recovery-10015658-1.html

IMW14039-USEN-01