

Archiving Complex Enterprise Databases

IMPROVING APPLICATION PERFORMANCE, AVAILABILITY, AND RELIABILITY

W H I T E
P A P E R

April 2001



Contents

Archiving: Opportunity and Impact	1
Archiving Relational Data: The Issues	3
Developing an Archiving Strategy: Alternatives	5
Creating an Active Archive	7
Removing Data from Production	9
Researching and Restoring Archived Data	10
Active Archiving Benefits	12

Archiving: Opportunity and Impact

Companies make major investments in technology when the pain associated with “business as usual” reaches an unbearable level. In IT, that pain is often aggravated by the relentless evolution of advanced computing and the need to embrace and exploit new paradigms, such as the Internet and eBusiness. One area where IT executives are looking for a breakthrough is in managing complex enterprise databases that keep getting larger and larger.

Industry analysts and leading computing publications report that databases are growing at an unprecedented rate, with some very large databases doubling, even tripling in size in a year’s time. Enterprise data management and storage problems are becoming a top priority for today’s IT executives.

In addition, corporate databases are growing so large that just maintaining traditional service levels is a demanding challenge for many IT organizations. CIOs face an increasing demand for 100% availability of mission-critical applications, 24/7 eBusiness support, and scalable access to enterprise data on a variety of platforms.

So, what is the “business as usual” response to the problem of database growth?

Responding to “Business as Usual”

For most companies, the solution to explosive database growth is to upgrade servers and acquire more storage capacity. Faster, more powerful processors can speed access to information, while the underlying database continues to grow. However, this approach is rapidly losing its viability for the following major reasons:

- **Performance**

In the past, IT could count on a linear increase in capacity to keep the service levels of a growing database within an acceptable tolerance. However, with the advent of eBusiness and the Internet, the demand for high performance is more critical than ever, requiring more frequent and larger increases in capacity to satisfy the demand.

- **Availability**

More data means more maintenance. The corporate data explosion has squeezed backup and reorganization windows to the point where data availability is seriously threatened.

- **Reliability**

Increased competition and 24/7 “open-for-business” initiatives require improved reliability to keep mission-critical applications up and running. However, databases that continue to expand can negatively impact application reliability.

- **Cost**

Hardware upgrades can be exceedingly expensive. IT organizations often spend millions of dollars in hardware and software license fees when they implement initiatives to expand capacity.

These factors challenge the perception that expanding capacity is the most effective way to deal with the problem of database growth. Many IT organizations are beginning to realize that capacity upgrades are an increasingly expensive tactical approach to a strategic problem.

Archiving — A Viable Solution

Today, CIOs are leveraging archiving technology to address the fundamental issue of runaway database growth. Streamlining mission-critical databases is possible by transferring outdated or inactive data to an archive. Data archives can be quickly researched and data can be easily restored when necessary. In the meantime, application performance, availability, and reliability are greatly improved, and planned increases in capacity can be deferred — often saving millions of dollars.

Archiving Relational Data: The Issues

When database size increases relentlessly, the performance of mission-critical applications deteriorates because larger databases take more time to load, unload, search, reorganize, index, and optimize. Response times increase. Access to decision-making information becomes more difficult. Service levels decline.

While most applications have excellent strategies for validating and updating data, few have an orderly method for safely removing inactive data from production when it is no longer needed. Even if data could be archived, any subsequent need to access or restore it poses a major challenge.

Accessibility Issues

Many companies realize that they must be prepared for any request to access inactive data. For example, a company may need to protect its interests with the ability to produce records of old financial transactions if the need arises. In other cases, companies must comply with government or legal retention and record keeping requirements for specific types of data.

However, the major fear is that once data is no longer in the active database, it may be difficult to quickly locate and restore when needed. As a result, companies find it easier to continue with “business as usual” and just let production databases continue to grow.

Technical Issues

In addition, there are significant technical challenges involved in archiving production data. This is particularly true when the data is stored in a relational database, such as DB2, Oracle, Sybase, SQL Server, or Informix. Often, data is normalized across hundreds of tables that are interconnected by at least that many relationships. To complicate matters, these relationships may be managed by the application rather than by database referential integrity (RI) rules.

When copying data to an archive, the IT staff is typically concerned about selecting a complete and accurate set of related rows to transfer. In addition, they worry about inadvertently removing production data that may still be needed by an application, or worse yet, accidentally corrupting the database without an easy way to restore the data. So again, it may seem easier to let a large database continue to grow. No one wants to risk breaking a database that works, particularly when strategic applications are at stake.

Opportunity and Challenge

Accessibility and technical issues represent both opportunity and challenge. Safely streamlining large databases with an effective archiving strategy can help IT organizations make substantial improvements to performance, availability, and reliability.

The right archiving strategy can also yield significant cost benefits by reducing or eliminating the need for increased capacity and safely preserving inactive data for potential reuse. But, to properly archive data, the following questions need to be addressed:

- With dozens or hundreds of normalized tables in a database, how do you select the precise and complete set of related data to archive?
- With many application-managed relationships in a database, how do you create referentially intact sets of data spanning many related tables before copying data to an archive?
- After data is archived, how do you accurately remove only the precise set of related data from production, leaving all other production data intact?
- What if you need to locate specific data from among several archives? How can you research archives quickly and easily to find the data you need?
- How can you “selectively restore” data from one or several archives rather than reloading entire sets of archived data?
- When restoring data, how can you ensure that all related rows, spanning dozens or even hundreds of tables, are restored correctly?
- What if the production data model changes between the time data has been archived and the time it needs to be restored? How can you restore data and preserve the integrity of your relational data model?

Certainly, even with adequate time and resources, an archiving application would be difficult to design, develop, and implement — and this is particularly true when your production data is housed in a large complex relational database.

Developing an Archiving Strategy: Alternatives

Because database growth affects so many companies, many IT executives are already convinced that an effective archiving strategy makes sense, and they are actively searching for viable solutions. Some of their requirements include an automatic way to streamline databases during routine database maintenance, an efficient method for tracking archive operations, and a fast and accurate way to browse and selectively restore data from the archives when necessary. There are several alternatives.

Capacity Expansion

As noted earlier, increasing capacity is not really a solution, but rather a means of living with the problems associated with ever-expanding databases. Many IT organizations are reaching the point where a true archiving strategy can yield a far greater payback than capacity upgrades — both in improved performance and reduced cost.

Intensive Database Tuning

Often, a DBA can squeeze better performance from a database by reorganizing the database more frequently, adding an index, implementing partitioning, or even de-normalizing parts of the database (though this last option may create costly ripple effects on application maintenance). However, as the database continues to grow, the effort yields less return and only postpones the inevitable need for an effective long-term solution.

In-House Archiving

Faced with the need to archive data and the limited number of proven archiving tools in the marketplace, many IT organizations initiate projects to develop an in-house solution. However, after the magnitude of the undertaking begins to impede other high priority internal programs, these projects are usually shelved or dramatically curtailed, yielding only partial and often inadequate benefits.

Developing an internal archiving solution diverts highly skilled IT resources from contributing to the direct business focus of the enterprise. The IT project staff must not only write and debug software to archive subsets of data with 100% accuracy, they must also design and develop a means for researching and restoring archived data. Then there is the problem of managing a growing number of archives. In addition, the software must be customized for every new application and every change to the production data model.

Rather than devoting additional time, budget, and personnel resources to continually maintaining, supporting, and enhancing an internal archiving solution, most senior IT managers choose to wait for archiving software products to emerge on the open market.

Off-the-Shelf Archiving

In the long run, implementing off-the-shelf archiving software is a cost-effective solution because it directly addresses the problem of database growth without diverting internal resources from IT business initiatives. With the right archiving solution, current and future applications and databases can be merged into an enterprise archiving strategy. In addition, a full set of tools may be available for all of the attendant support and maintenance tasks — tracking archives, research activities, and selectively restoring data.

Active Archiving

Traditionally, companies have considered that archiving data is only useful if they never expect to need that data again. This concept has prevailed because of the inherent difficulty in researching and restoring archived data. However, Princeton Softech is introducing a new concept — active archiving.

With active archiving, infrequently used data (reference data) can be “set aside” or removed from production and still be “close at hand” (active). Mission-critical data remains in the production database, available at peak performance, and active reference data is easy to research and selectively restore when needed. Active archiving is a major advance because it preserves the business context of archived data and yet keeps that data available for continued use.

Princeton Softech’s Active Archive Solutions™ represent proven technology and are the first in today’s marketplace specifically designed for complex enterprise databases. Archive for DB2™ and Archive for Servers™ offer complete solutions for resolving the problems associated with streamlining large relational databases. And for cross-platform flexibility, Archive for Servers supports a variety of DBMSs including DB2-UDB, Oracle, SQL Server, Sybase, and Informix.

In addition to archiving data — referentially complete and intact — these solutions provide built-in capabilities to satisfy the requirements for researching and selectively restoring archived data with 100% accuracy, regardless of changes to the data model. And there is no need to write, debug and maintain complex custom programs.

Creating an Active Archive

Princeton Softech's Archive for DB2 and Archive for Servers provide proven archiving capabilities for both the mainframe and client/server environments. These Active Archive Solutions rely on Princeton Softech's Relationship Engine™, a patented technology that always extracts precisely the right data with 100% accuracy — no matter how many tables or relationships are involved — and guarantees to keep that data referentially intact. Active archiving is accomplished in an effective two-step process.

Define the Data to Archive

First, users create an Access Definition that specifies the tables and relationships needed to create the referentially intact set of data to archive. These specifications can include selection and date criteria and user-defined indexes to speed researching and restoring archived data at a later time. Users can also specify "archive actions" to have more control over archive processing.

For example, a company may want to archive all information relating to customers who have been inactive for at least one year. The archive must include all identifying information as well as order and payment history, and the data must be extracted from many interrelated tables.

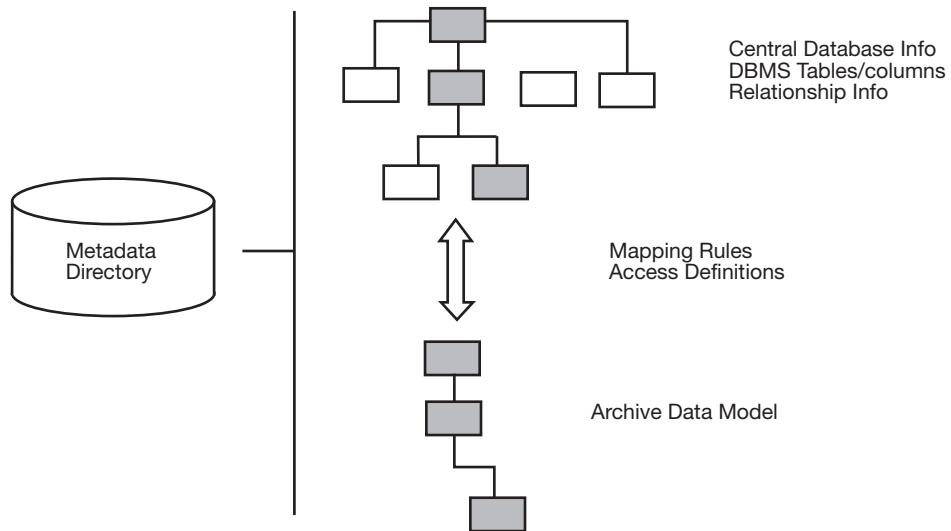


Figure 1. Define the Sets of Related Data to Archive

Copy Data to an Archive

In the second step, the Access Definition is used to copy data to an archive file. During this process, the indexes specified to track data in an archive, and if necessary to find it later on, are updated automatically.

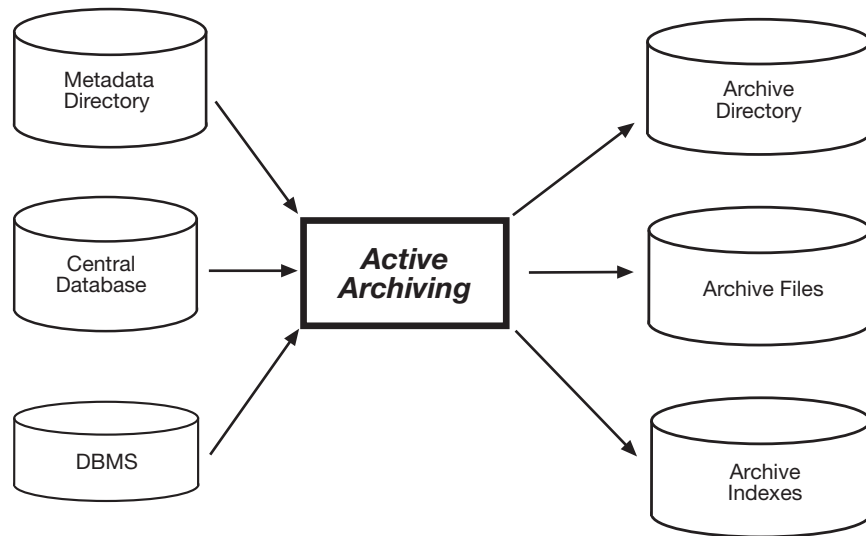


Figure 2. Creating an Archive

Active archiving saves not only the data, but also the metadata describing tables, columns, and relationships used to create the archive. With this information, users can always restore data in its business context — weeks or years later — even if the data model changes over time.

Removing Data from Production

Typically, IT organizations are reluctant to delete production data from complex relational databases, mainly because accidentally deleting essential data could bring mission critical systems to a halt. This is one reason why it is often considered safer to let databases grow than to risk corrupting the data using a delete program that does not guarantee accurate results.

Princeton Softech's Active Archive Solutions address this problem by providing users with a choice of several powerful yet safe relational delete capabilities. Because Princeton Softech's products are "referentially aware," users can remove small portions or entire sets of related data quickly and accurately. In addition, these solutions use standard database facilities for all operations. DBMS security features are never bypassed.

Immediate Delete after Archive

As part of the archive process, users may choose to remove data from production immediately after the data is archived — in a single process request. As an added safety measure, users can review a list of tables first, to verify and ensure that the desired data is removed from production, before beginning the delete process.

Deferred Delete

Deferred Delete™ is a proven relational delete capability that offers advantages when users want to verify the data before it is deleted, or want to include delete processing in their routine database maintenance tasks. The Deferred Delete can be used at any point after an archive is created successfully. The relational delete process uses the specifications defined in the Access Definition to remove data from production, complete and referentially intact.

Selective Delete

With Selective Delete™, users can remove all or just a specific portion of the archived data from production, providing for maximum flexibility during the delete process. For example, suppose an IT organization decides to archive all order and payment history for customers who have inactive accounts over the last year. And, they want to retain master account information for each customer in the production database, including names and addresses. The archive can be created to include all identifying data for each customer, and Selective Delete can be directed to remove only the related rows from the order and payment tables.

Researching and Restoring Archived Data

After data is archived and removed from production, IT organizations are concerned that years (or maybe only days) later, they will be unable to locate or restore that data in a useable form. In addition, archives proliferate over time; searching and managing them can be another maintenance challenge.

Managing Archive Files

Princeton Softech's Active Archiving Solutions provide a user-managed Archive Directory for grouping and maintaining any number of archives. Familiar file maintenance features allow IT organizations to set customized standards for managing archive files, making the data easy to research and readily accessible.

Researching Archived Data

Users can take advantage of a powerful browsing capability that takes the worry out of finding specific data among numerous large archive files. Users can leverage the archive indexing capability to quickly search archive files for the desired information.

Often, questions about archived data can be answered without restoring a single row. Users can browse an archive file directly, searching for data based on simple or complex selection criteria. Results are presented on an interactive display that shows data in its relational or business context. Users can research archive files with no impact on the current production system.

Restoring Archived Data

No archiving solution would be complete without the capability for restoring archived data when the need arises. But data recovery is often difficult because archived data from hundreds of tables may need to be reassembled and restored based on complex relationships. Then, the right set of rows must be identified and selectively restored. To complicate this scenario, as years pass, database models can change making any restore effort more challenging.

With Princeton Softech's Relationship Engine technology, users can locate specific data in any number of archives and restore it — referentially intact and complete. Archive for DB2 and Archive for Servers support a variety of options, including the ability to restore data to the production database or more commonly to a separate database for review and reporting purposes.

Selective Restore

With Selective Restore™, users can restore data using specific search criteria that may differ from the criteria used to create the archive. This means that users can restore subsets of data “selectively.” It is not necessary to restore an entire archive to obtain only a small portion of the needed data.

For example, suppose transaction and payment data is archived based on a period of activity, such as the date of the last transaction is older than two years. Then, seven years later, suppose that the company needs to restore data for customers from New Jersey. Selective Restore allows users to quickly find and restore transaction data only for the New Jersey customers. In this example, the data was archived using date criteria and restored using state criteria.

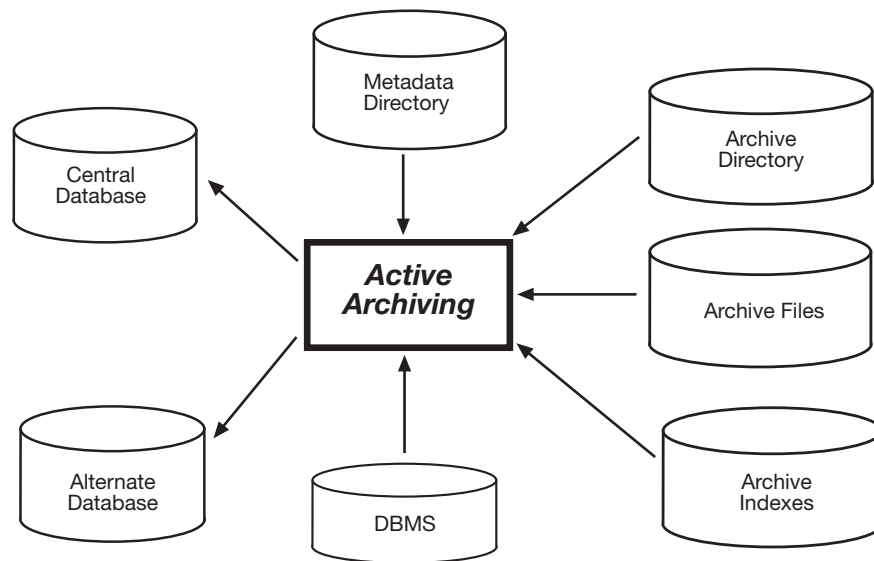


Figure 3. Restoring Data from Archives

Active Archiving Benefits

Princeton Softech's Active Archive Solutions address a critical operational need for enterprises with large and complex relational databases. Inactive data can be moved to archives in a precise, accurate manner. Databases can be streamlined for peak performance. Valuable capacity is no longer required to support tables and indexes containing inactive data. Archives can be browsed and data can be selectively restored as needed.

With Archive for DB2 and Archive for Servers, companies can maximize ROI in a number of ways by:

- Storing archive files on the most cost-effective medium.
- Deferring capacity upgrades — with the associated expensive hardware and software license fees to realize significant savings.
- Reducing IT budget requirements for maintaining production database expenses.
- Maintaining acceptable service levels by minimizing production database searches and improving response times.
- Improving availability by reducing the amount of time needed to backup and reorganize data.
- Eliminating the expense of writing and debugging complex archive programs and the additional expense of creating one-time programs for restoring archived data.

As production databases continue to grow, IT organizations can realize many benefits by implementing an effective active archiving strategy. For the large-scale committed enterprise, Princeton Softech's technology provides proven solutions that make these benefits possible.

princetonsofttech.com

111 Campus Drive
Princeton, NJ 08540-6400
Toll free **800.457.7060**
Phone 609.627.5500
Fax 609.627.7799

