

Store Smarter: Enterprise Active Archiving

A STRATEGIC COMPONENT OF ENTERPRISE DATA MANAGEMENT

WHITE
PAPER
April 2002



Enterprise Active Archiving

Transaction-intensive, customer-facing applications, as well as high volume CRM and ERP applications are collecting and storing more data and doubling and tripling in size faster than ever before. To solve this problem of database explosion, enterprises should consider active archiving. Active archiving optimizes database performance while maximizing existing enterprise storage capacity. Going well beyond the traditional definition of archiving, active archiving allows companies to select and remove precise sets of rarely used data or “active reference data” from a production database, store that data in an archive, and keep it “active” for easy access when ever it is needed. This capability dramatically reduces database overload, allowing companies to reduce storage requirements, improve application response time and reallocate current capacity to support more users and transactions.

Contact:

Princeton Softech

111 Campus Drive

Princeton, NJ 08540-6400 USA

Phone: 800.457.7060 (US, Canada, Puerto Rico)

Phone: 609.627.5500

Fax: 609.627.7799

info@princetonsoftech.com

Contents

Coping with the Data Explosion	1
Storing Smarter with Active Archiving	5
Adding Strategic Value Across the Enterprise.....	8
Achieve Your Critical Business Initiatives	11
Conclusion: Why Store Smarter?	13
References	14

Coping with the Data Explosion

Success in today's customer-driven and highly competitive business environment depends on your company's ability to better understand and respond to customer needs. Collecting the right data is essential. Transaction-intensive, customer-facing applications, as well as high-volume CRM and ERP applications, are collecting and storing more data, doubling and tripling in size faster than ever before.

Because these applications rely on complex relational databases, managing continuous database growth is absolutely necessary for controlling costs, improving customer satisfaction and enhancing decision support to ensure long-term success.

What's Driving the Data Explosion?

The rapid growth of the Internet and eBusiness, the increase in online transaction processing and the expansion of large databases that support customer-facing applications have contributed significantly to the data explosion. This unprecedented growth is driving increasing demand for data storage and data management solutions.¹

Additionally, there is an increasing need to maintain easy access to historical information for business or regulatory requirements. For example, a company may need to protect its interests by retrieving historical financial transactions to satisfy customer inquiries and resolve claims. In other cases, corporate policy or government regulations dictate that data must remain accessible for years after it is collected.

For many organizations, accelerating data growth and the need to store historical information are putting a real strain on enterprise IT system capacity. As a result, data storage is quickly becoming the biggest expense in enterprise IT budgets, responsible for up to 30 percent of capital expenditures, despite the fact that the cost of traditional disk storage is dropping.²

¹ "Behemoth Databases Push the Limits," by Joseph McKendrick, *Database Trends and Applications*, September 2001.

² "The Defogger: How to Beat the High Cost of Storage," by Dylan Tweney, *Business2.0*, July 2001.

Why Is Data Growth a Problem?

Database growth degrades performance and limits the availability of mission-critical applications and data warehouses. To maintain service levels, enterprises are spending millions of dollars on hardware (storage and CPU) upgrades, software upgrades and maintenance fees. This tactical approach – to what is really a strategic problem – provides only a temporary solution.

Accelerating database growth is expected to continue for years. A September 2001 *Database Trends and Applications* study shows that many databases are crossing into the double digits in terms of terabyte size. Within a few years, experts predict that the largest databases in the world will contain 100 TB of data. Today, IT professionals are evaluating and deploying new data management strategies to improve the effectiveness of their current systems and to prepare for future escalation in storage requirements.

Overloaded relational databases limit productivity and increase costs. Even worse, companies risk frustrating or losing customers because of unresponsive eBusiness or customer care applications. In addition, this growth can severely impact IT organizations, extending well beyond increasing storage costs. Larger databases take longer to load, unload, search and reorganize. Response time slows to the point where maintaining traditional service levels can become nearly impossible.

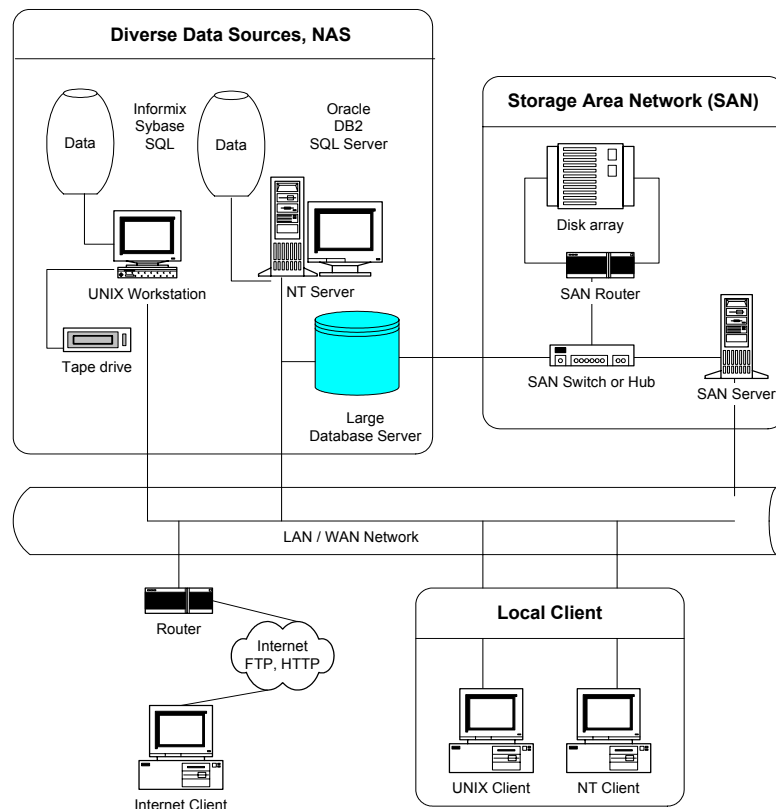


Figure 1. Databases and Storage Solutions are Overloaded

Why Have Companies Been Reluctant to Archive Data?

Despite the need to reduce database size and the potential benefits of archiving to improve performance and availability, IT organizations have been reluctant to remove data from production databases because accidentally deleting essential data could bring mission-critical systems to a halt. This is particularly true when data is stored in a relational database, where it is normalized across hundreds of tables, interconnected by hundreds of relationships. To further complicate matters, these relationships may be managed by the application rather than by database enforced referential integrity (RI) rules.

Another major concern is the need to quickly locate and access data once it is removed from the production database and stored in an archive. Responding to audits, lawsuits, government or security investigations – as well as answering customer questions – may require access to archived data and possibly restoring it to production.

While most applications provide excellent features for collecting, validating and updating data, few include tools to safely archive and remove seldom-referenced data from production. Even if data could be safely archived, any subsequent requirement to access or restore it can pose a major challenge.

To minimize risk, organizations continue to store all of their relational data in the production database and address performance and availability issues with intensive database tuning and capacity upgrades. Concerns about accidentally deleting essential data and accessibility issues are valid. Until there is a proven solution that can safely archive data and provide easy access, it may seem safer to allow continued database growth rather than risk a failed archiving process.

Active Archiving Provides a Long-Term Solution

Unlike traditional archiving, active archiving is a proven technology that safely archives and removes precise subsets of rarely used data from complex relational databases with 100 percent accuracy. Ongoing active archiving keeps databases streamlined and operating at peak performance providing a cost-effective, long-term solution to the problem of accelerating database growth.

Although some applications offer a rudimentary data purging functionality, none deliver the full set of active archiving capabilities needed to store and access transactional data. Streamlining these databases is critical to improving customer satisfaction, enhancing decision support and managing costs.

Active archiving is an effective strategy that increases the value proposition of storage and database technologies including SAN, NAS, HSM and data warehousing. Active archiving works within the framework of these technologies to improve database performance and availability. In addition, because active archiving supports both distributed and mainframe databases, it can be implemented across an enterprise.

Active Archiving Delivers Strategic Benefits

Implementing active archiving delivers many strategic benefits:

- Archiving and safely removing seldom-used data from growing databases significantly improves performance and availability of data-intensive applications.
- Improving performance and availability increases revenue opportunities and generates internal and external customer satisfaction.
- Deferring or eliminating the need for expensive upgrades can save millions of dollars in IT expenses.
- Researching and selectively restoring archived data enables organizations to quickly respond to customer inquiries, legal issues, audits and government regulations.
- Implementing routine active archiving for data-intensive applications across the enterprise can deliver a rapid return on investment.
- Storing smarter with active archiving can complement and enhance the value proposition of other industry leading storage solutions.
- Active archiving can significantly reduce the time and resources needed to rebuild the database environment when a disaster strikes. As a result, it takes less time to recover and resume the operation of critical systems.

Storing Smarter with Active Archiving

Princeton Softech's Active Archive™ Solutions help companies store smarter by providing proven technology for implementing an effective active archiving strategy. These solutions rely on Princeton Softech's Relationship Engine™, a unique technology that *understands* and remembers the data relationships. As a result, data can be archived and restored with 100 percent accuracy, no matter how many tables or relationships are involved.

Archive and Remove Relational Data Safely

Active archiving safely archives and removes subsets of related data from a production database – complete and referentially intact. The active archiving process saves not only the data, but also the metadata that describes tables, columns and relationships used to create the archive. With this information, users can always restore data in its business context, weeks or years after it has been archived, even if the data model has changed.

Powerful, yet safe, Princeton Softech's Active Archive Solutions are “referentially aware.” The Selective Delete™ feature allows users to remove all or just a portion of the archived data from the production database safely and accurately. Optionally, users may choose to archive and defer the delete for a later time allowing users to verify the archived data before removing it from the production database.

Maximize Storage Utilization

Implementing an active archiving strategy with Princeton Softech's Active Archive Solutions directly addresses the problem of database growth and can be particularly effective for maximizing storage utilization. With active archiving, IT organizations have been able to archive and remove up to 65 percent of their production data in their first archive.

Regularly scheduled active archiving provides a repeatable process for managing database growth on a go-forward basis. This strategy frees tremendous processing power and storage capacity for other business-critical applications.

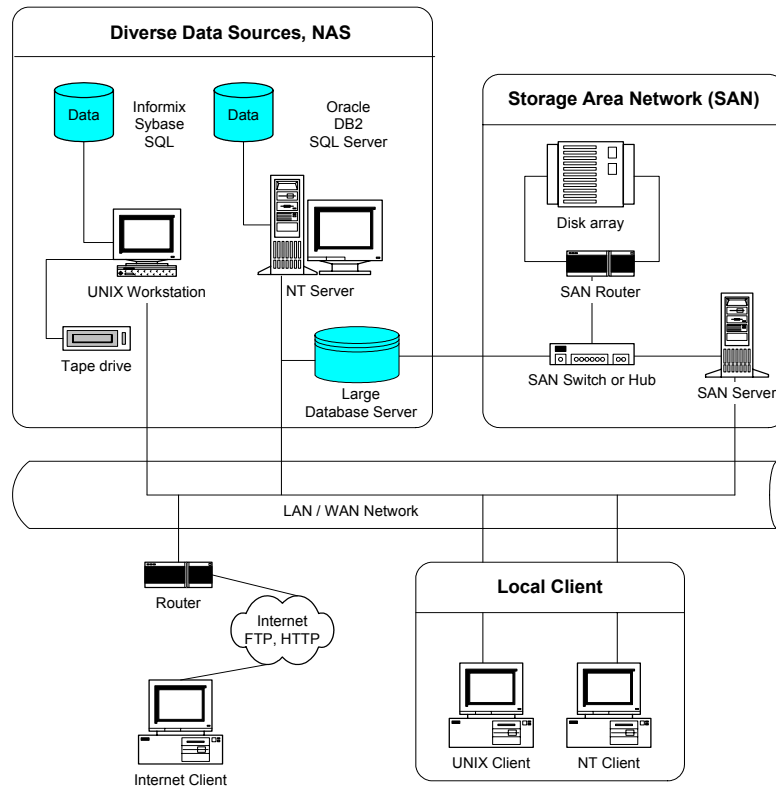


Figure 2. Active Archiving Reduces Database Overload and Maximizes Storage Utilization

Store Data on the Most Cost-Effective Medium

Archived data can be stored on the most cost-effective medium and compressed to further optimize storage utilization. In addition, Archive Files can also be stored on a virtual tape system (VTS), which emulates a group of standard tape drives.

The Archive Files can be staged on a tape media and placed in the cache when data is referenced during an "active" period, then returned to tape after the active period has passed. Together, active archiving and a VTS provide a powerful solution for streamlining overloaded databases and keeping data available for easy access on the most cost-effective storage media.

Improve Performance and Availability

The traditional solutions for improving database performance – including upgrades, intensive tuning and custom software – only alleviate the problem temporarily. Unlike these short-term “fixes,” active archiving solves this problem for the long-term.

After the initial archive, regularly scheduled active archiving optimizes database performance on a continuous basis and can dramatically shorten the time it takes to access data, process transactions, and run maintenance utilities. Instead of the one-time gain that is realized from tuning and upgrades, active archiving can become a routine part of scheduled application database maintenance.

With active archiving, backup and reorganization processing takes less time because there is less data in the production database to process. Batch processing and availability are greatly improved, especially with applications that process a large percentage of data.

Maintain Easy Access to Archived Data

As the number of Archive Files increases over time, searching and managing archived data can become a significant maintenance challenge. To solve this problem, Active Archive Solutions provide a user-managed Archive Directory for organizing, locating and maintaining any number of Archive Files.

A powerful global search capability makes it easy to search numerous Archive Files to find specific data, with no impact on the current production system. Users can also browse an Archive File directly, searching for data based on simple or complex criteria. Results are presented on an interactive display that shows data in its relational or business context. It takes less time for users to find answers to specific questions, and they can even generate reports, without restoring a single row to the production database.

Restore Data to Production if Needed

No archiving solution would be complete without the ability to selectively restore archived data when the need arises. But restoring data can be extremely difficult because archived data from hundreds of tables may need to be reassembled and restored based on complex relationships.

With Princeton Softech's Selective Restore™ capability, users can restore data by specifying any type of criteria, including 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 if only a small subset of the data is needed, as is typically the case.

Additionally, users can locate and restore specific data from any number of Archive Files, referentially intact and complete. Users can restore data to the production database, or to a separate database for review and reporting purposes. If the data model has changed, users can restore archived data to the new model or restore the data exactly as it was to a different database, which is created automatically.

Adding Strategic Value Across the Enterprise

Today, many IT organizations deploy enterprise data management technologies to provide high-speed access to data, improve data backup and reduce the cost of operations. Because active archiving optimizes storage utilization without hardware and software upgrades, it is a key component for enhancing the value proposition of other storage and data management technologies. An important benefit of active archiving is the capability to work with current or planned data management technologies and data warehouses.

Manage SAN/NAS Data Overload

Storage Area Networks (SANs) and Network Attached Storage (NAS) solutions can store data efficiently and provide high-speed database access. Because SAN and NAS systems do not remove data from production databases, data volumes will continue to grow rapidly. Active archiving works within the framework of your SAN and NAS environment to streamline relational databases on an ongoing basis, maximizing your storage investment for the long-term.

Storage virtualization and pooling configurations are an integral part of your SAN and NAS strategy. Archive Files are saved on the available network storage where they can be easily accessed, browsed, used for reporting purposes, and even restored when needed. In all instances, file access is performed through the existing virtualization and pooling configurations.

Improve Effectiveness of HSM Solutions

Hierarchical Storage Management (HSM) is designed to reduce storage and administration costs while keeping data available.³ This technology automatically manages files based on the last access date, migrating files of a certain age or type to a near-line storage device or a tape storage library. After the storage system administrator defines the migration rules, HSM manages file migration and de-migration automatically.

HSM and active archiving both address the problem of data growth by moving infrequently used data to more cost-effective storage devices. However, active archiving is designed to handle data stored in a relational database, while HSM is best suited for other types of data such as document files, bit maps and video clips.

³ “Hierarchical Storage Management,” by Luz Rink, Beyond Technology.

HSM technology alone is not well suited for managing relational database tables because they are physically stored as files at the table or data set level. HSM performs data migration based on the last time a particular database table was accessed. Because it is likely that some data in any given table will remain “hot” (such as current customer data), most tables will remain in production on the network server, degrading application performance and limiting availability.

Companies that have already deployed HSM understand the benefits of “staged” data management. Active archiving removes rarely used data from a relational database and stores it in an Archive File. HSM migration rules can be defined to manage “warm” and “cold” Archive Files and to store archived data on the most cost-effective medium. By combining HSM and active archiving, IT organizations will benefit from both a streamlined database and HSM managed Archive Files.

Streamline Data Warehouses

Typically, data flows from an operational database to the data warehouse. After operational data is filtered, translated and transformed to be useful for decision support, it is loaded into a centralized integrated data warehouse.

Archiving data from your operational database before it is moved to a data warehouse is a smart strategy because it allows you to preserve your business data in its “original” context, completely and referentially intact.⁴ The ability to access this data in the event of an audit could mean the difference between passing and failing. In addition, because active archiving includes metadata, it is possible to recreate the “original” data model if necessary. By archiving and removing data from the operational database, IT organizations can improve performance and availability of the operational system.

Most successful data warehouses are characterized by rapid data growth. However, as more users rely on data warehouses for decision support to run day-to-day business operations, more queries are executed, and slow response time and degraded performance can become major issues. The same active archiving strategy that works with operational systems can also be used to keep data warehouse performance and response time at acceptable levels.

Active archiving is the only safe way to remove “dormant” data from the data warehouse and still keep it accessible. This capability frees tremendous processing power to improve response time and overall performance. With active archiving on a regular basis, data warehouses can experience the same benefits available to the operational system. This approach provides a single archiving strategy across all database applications.

⁴ “The Audit and Archive Database,” by Michael Scofield, *DM Direct*, August 1998.

Implement New Software Releases

Depending on the size of the database, upgrading customer-facing and data-intensive applications can be a time consuming process, often taking applications out of service for extended periods of time until the installation of the upgrade or new release is complete.

One sure way to reduce the transition time is to archive data before installing the upgrade. By reducing the amount of rarely used data in the application database before upgrading to a major new application release, there will be far less data to migrate and implementing new upgrades and software releases can be accomplished much faster, minimizing any application downtime.

Implement Enterprise Active Archiving

Princeton Softech's Active Archive Solutions are designed to support distributed and mainframe environments allowing companies to implement active archiving across the enterprise.⁵ Most companies implement active archiving using a phased approach, starting with a single, data-intensive application. As active archiving continues on a regular basis, Active Archive Solutions provide tools that help determine the optimum amount of data to archive and remove. This systematic approach can then be extended to other mission-critical applications across the enterprise.

Active Archive Solutions are database independent, allowing IT organizations to manage archived data across the leading relational database platforms (Oracle, DB2/UDB, Sybase, SQL Server, and Informix).⁶ Users can archive and restore data from one database platform to another. Relational data from multiple databases can be archived together to create one or more enterprise active archives. In addition, if a database management platform changes, Active Archive Solutions will continue to support your active archiving strategy through the transition.

⁵ "Princeton Softech," by Richard L. Ptak, *Hurwitz View*, November 1, 2001.

⁶ "Archiving Data — A Quiet Market With Significant Potential," by Lou Agousta, *Ideabyte*, Giga Information Group, October 9, 2001.

Achieve Your Critical Business Initiatives

Active archiving allows companies to achieve their critical business initiatives by providing a comprehensive strategy that delivers benefits throughout the enterprise. For example, CIOs and strategic planners can realize improved return on investment across the enterprise. Customer service managers can rely on the availability of critical customer-facing applications. Storage managers can reduce data storage requirements. IT staff on the front lines can improve daily database performance.

Increase Customer Satisfaction

With so many customer-oriented applications that rely on complex database transactions, maintaining service levels is more important than ever. Competition is pressuring companies to improve response time. Now that customers are just a click away from the nearest competitor, five-nine uptime (99.999 percent) has become the new benchmark.

With active archiving, IT organizations can ensure that applications are operating at peak performance, keeping customers and users satisfied. And by dramatically shortening batch windows, active archiving can improve the availability of mission-critical applications.

Save Millions of Dollars in Upgrades and Maintenance Fees

In the past, IT professionals could count on a linear increase in capacity to maintain the service levels of growing databases. Capacity expansion and database tuning provide interim solutions, and as production databases expand, expensive upgrades are needed more often.

When IT organizations choose upgrades as an interim solution, they often spend millions of dollars in hardware costs and software license fees. What's more, as databases continue to grow, upgrades yield less and less return on investment and merely postpone the need for a long-term solution.

Active archiving provides an effective long-term solution with the potential to save millions of dollars in deferred or eliminated upgrade expenses. By safely archiving, removing, and storing rarely used production data on the most cost-effective medium, active archiving is key to controlling costs and maximizing return on investment.

Manage Risk and Comply with Data Retention Regulations

Typically, much of the stored application data will never be needed again. However, risk management, data retention and legal requirements mandate that companies keep historical data readily accessible for several years, just in case it is necessary to respond to:

- Internal and external audits
- Government or industry regulations
- Legal summons and court orders
- Customer service issues

Archived data can be researched and browsed easily without restoring a single row. If necessary, a complete Archive File or small subsets of archived data can be selectively restored to the production database or an alternate database. With active archiving, historical data is readily available when needed. Companies are better able to satisfy legal and business requirements and protect against lawsuits and costly fines.

Improve Business Continuity Strategies

Today, companies are making disaster recovery a higher priority. Most disaster recovery plans include backup and restore procedures for computer and telecommunications systems. It is also important to include processes and procedures for managing, maintaining and restoring operational data needed for critical systems in the shortest time possible.

By keeping operational databases streamlined, active archiving enhances disaster recovery plans, allowing companies to recover operational data and rebuild alternate databases in significantly less time. After the critical systems are operational, a phased recovery plan can be used to restore archived data based on the business priorities of the data.

In addition, many disaster recovery plans have service level agreements (SLAs) requiring that the recovery plan be completed within a specified timeframe. Active archiving will help IT organizations meet these requirements by reducing the time needed to complete the recovery of critical systems.

Conclusion: Why Store Smarter?

How long can you afford to delay active archiving before accelerating database growth cripples your operations? Complex relational databases that support customer-facing and transaction intensive applications as well as CRM, ERP and other mission-critical applications are doubling and tripling in size each year. Princeton Softech's Active Archive Solutions offer a cost-effective, long-term solution to what industry experts agree is one of the biggest issues facing IT today:

- Active archiving on an ongoing basis improves application performance and availability. With the potential to save millions of dollars in deferred upgrade expenses, this proven technology can maximize your return on investment.
- Active archiving adds strategic value to your company's enterprise data management solutions. By working within the framework of these technologies to manage database growth, active archiving optimizes performance, availability and storage utilization across the enterprise.

These improvements provide valuable benefits for your company. Customer satisfaction increases as response time and uptime improve. Your organization's ability to respond quickly and accurately to business or legal questions is greatly enhanced.

In the event of a disaster, active archiving will result in a faster time to the recovery and operation of critical systems. And because Active Archive Solutions are highly effective and easy to use, your IT staff will save countless hours of their valuable time.

Princeton Softech is the leader in the active archiving market. Companies throughout the world recognize the benefits of enterprise active archiving to improve performance and availability and to achieve a greater return on investment.

With more than 2000 enterprise customers in over 30 countries, Princeton Softech is ready to support your enterprise active archiving needs. Our archiving experts will work closely with your IT team to develop an effective active archiving strategy specifically for your environment.

References

1. "Archiving Data – A Quiet Market With Significant Potential," by Lou Agousta, *Ideabyte*, Giga Information Group RIB-102001-00076, October 9, 2001, <http://www.gigaweb.com>.
2. "Behemoth Databases Push the Limits," by Joseph McKendrick, *Database Trends and Applications*, Volume 15, Number 9, September 2001.
3. "Cheap Space," by Jackie Cohen, *Computerworld*, July 30, 2001, http://www.computerworld.com/storyba/0,4125,NAV47_STO62600,00.html
4. "Data archiving improves database performance, decreases storage networking complexity," by Alan Radding *Storage Networking World Online*, November 12, 2001, http://www.snwonline.com/tech_edge/data_arch_11-07-2001.asp
5. "Database Archiving Strategies and Technologies," by S. Laufer, *Gartner Group Research Note*, Gartner Group Select Q&A, QA-08-8738, August 3, 1999, <http://www.gartnergroup.com>.
6. "Database Housecleaning: No Longer a Luxury," by Richard Evans, *META Delta* Meta Group Enterprise Data Center Strategies File 944, October 9, 2000, <http://www.metagroup.com>.
7. "Hierarchical Storage Management," by Luz Rink, *Beyond Technology*, September 8, 2000, <http://www-1.ibm.com/servers/eserver/series/whpapr/hsm.html>
8. "Intelligent, The Big Storage Picture," by Arun Taneja, *Intelligent Enterprise*, May 24, 2001, <http://www.intelligententerprise.com>.
9. "Managing Database Growth: Generalized RDBMS Archiving," by J. Rubin, *Gartner Group Research Note*, Gartner Group Strategic Data Management Document #T-14-5687, October 10, 2001, <http://www.gartner.com>.
10. "Princeton Softech," by Richard L. Ptak, *HurwitzView*, Hurwitz Group, November 1, 2001, <http://www.hurwitz.com>.
11. "The Audit and Archive Database," by Michael Scofield, *DM Direct*, August 1998, <http://www.dmreview.com/master.cfm?NavID=198&EdID=916>.
12. "The Defogger: How to Beat the High Cost of Storage," by Dylan Tweney, *Business 2.0*, July 2001, <http://www.business2.com/articles/mag/print/0,1643,12077,FF.html>

princetonsofttech.com

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

