February 23, 2011

Your Enterprise Data Archiving Strategy

by Noel Yuhanna
for Application Development & Delivery Professionals

FORRESTER® Making Leaders Successful Every Day

February 23, 2011

Your Enterprise Data Archiving Strategy

Data Archiving Is A Flexible Technology That Delivers Many Use Cases

by Noel Yuhanna
with Mike Gilpin, Adam Knoll, and Alissa Anderson

EXECUTIVE SUMMARY

With growing data volume, increasing data security breaches, and complex application-performance issues, most enterprises today face significant data management challenges. Keeping inactive data online not only creates security risk but also increases infrastructure and database cost. Some enterprises are literally throwing away data to minimize complexity and lower these costs. Although there are several better options available to control data growth, data archiving is the best long-term solution to significantly improve application performance and reduce the cost of servers and storage. Data archiving supports many other requirements, including retiring applications, managing test data, improving database manageability, performing data sharing, and ensuring regulatory compliance. Application development professionals and information architects should consider building an enterprisewide data archiving strategy for large databases and applications.

TABLE OF CONTENTS

NOTES & RESOURCES

Forrester interviewed 24 vendor and user companies, including BP, EMC, HP, IBM, Informatica, and Oracle.

Related Research Documents

"Loss Of Historical Financial Data Triggered This Application Consolidation Program"
March 8, 2010

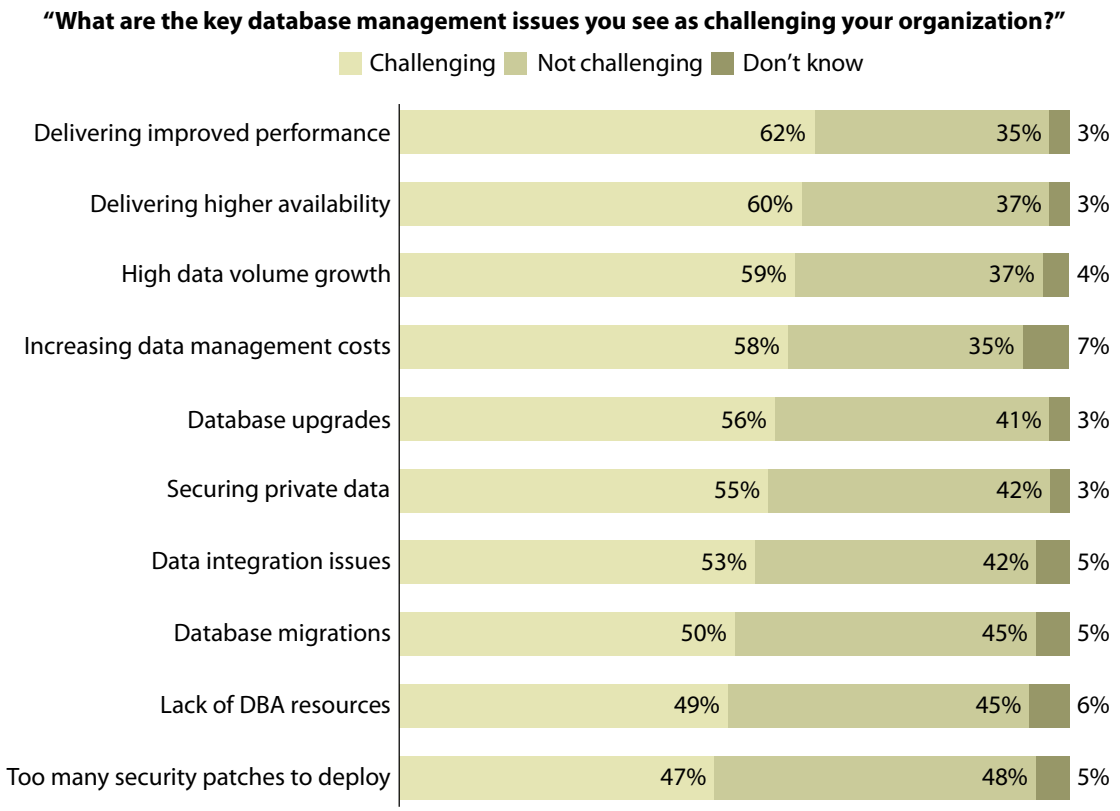
"Archiving: Finding Data In 2050"
July 1, 2008

"Database Archiving Remains An Important Part Of Enterprise DBMS Strategy"
August 13, 2007

UNSTOPPABLE DATA EXPLOSION IS PUTTING PRESSURE ON ENTERPRISES

Globalization, mergers and acquisitions, and new types of applications such as Web 2.0, radio frequency identification (RFID), social media, real-time business intelligence (BI), web services, and XML applications generate huge amounts of structured and unstructured data in today's enterprises. Most large enterprises have petabytes of data stored in various data repositories across the organization — and this is likely to grow to exabytes in the coming years. As a result, enterprises store more data every year in production systems. With these increasing data volumes come increasing costs as well as increasing challenges in securing and managing online data to deliver high performance and availability. Forrester's November 2010 Global Database Management Systems Online Survey evidenced this trend: Respondents indicated that the top three challenges they face are delivering improved performance, delivering higher availability, and dealing with high data volume growth (see Figure 1).

Figure 1 Data Volume Growth Remains A Major Concern For Most



Base: 194 enterprise database management professionals

Source: November 2010 Global Database Management Systems Online Survey

58169 Source: Forrester Research, Inc.

Large Databases Can Dramatically Increase Cost And Manageability Challenges

Today, most large enterprises have at least a terabyte-sized online transaction processing (OLTP) database in production to support critical applications, while some have databases of more than 15 terabytes (see Figure 2). Despite cheaper storage, a multiterabyte database remains challenging. Large databases require more effort in tuning, backup, migrations, upgrades, and integration. Some enterprises are even throwing away unwanted data (when not required for compliance) to reduce data growth and minimize complexity. Although data for critical online transaction processing (OLTP) applications grows on average at a rate of between 10% and 24% annually, for very large applications it typically doubles every 18 months (see Figure 3). Large databases exacerbate challenges regarding:

- **Application performance.** A terabyte-sized database often degrades application response time, unless data is cached or running on extremely large symmetric multiprocessing (SMP) or scaled-out cluster servers. Disk input/output (I/O) remains the biggest bottleneck on databases, and more data often means more delay in accessing data from storage subsystems. Despite innovation in database technology over the years, larger databases are still more difficult to tune.

- **Database management.** Although database management technology has improved, large databases still require more effort in backups, migrations, upgrades, and general administration.

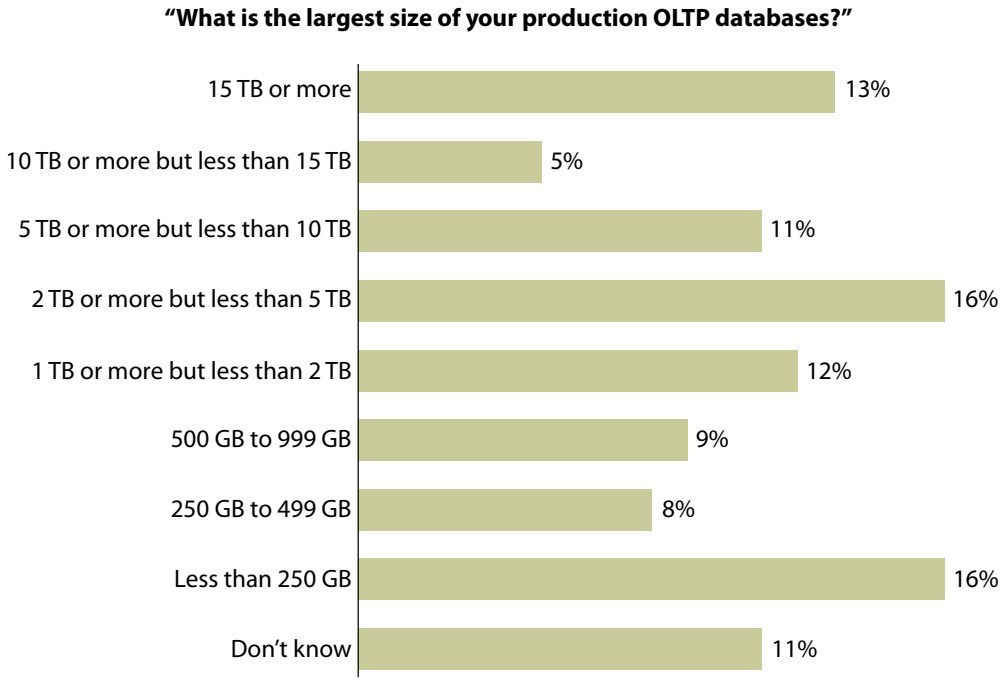
- **Application availability.** Based on client feedback, Forrester finds that smaller databases tend to deliver higher service levels than multiterabyte-sized databases.

- **Growing replicas.** Growth in transactional production databases drives growth of replicas in nonproduction environments for test, development, quality assurance (QA), staging, and training. Today most enterprises maintain an average of five to seven copies of each production database. Therefore, each large production database actually drives a footprint five to seven times larger.

- **Data security.** Keeping sensitive data online in production databases after it is no longer needed increases the risk of its exposure, so database operators should remove it. This includes old customer information such as credit card numbers, Social Security Numbers, and medical records as well as information about terminated employees. Large databases often retain sensitive data longer than necessary, increasing risk, so be sure to review their data retention on a regular basis.

- **Operational cost.** Inactive data in production systems adds significant cost, requiring more central processing units (CPUs), memory, and storage. Databases that are more compact require fewer resources and less management effort.

Figure 2 A Majority Of Enterprises Have A Terabyte-Sized Database

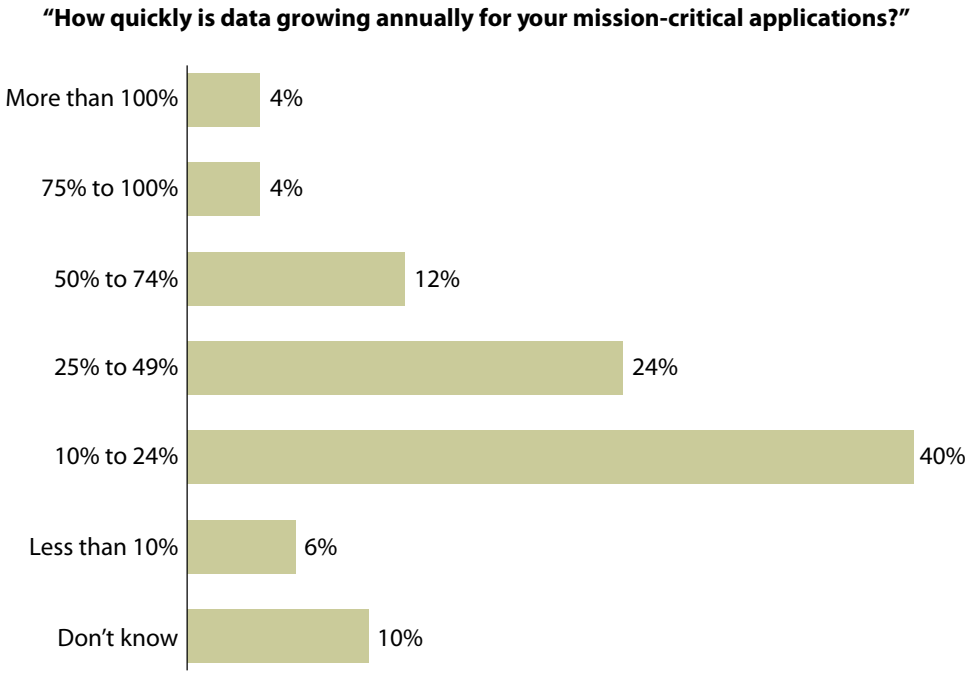


Base: 194 enterprise database management professionals
(percentages may not total 100 because of rounding)

Source: November 2010 Global Database Management Systems Online Survey

58169 Source: Forrester Research, Inc.

Figure 3 Data Growth Continues Across Various Applications



Base: 194 enterprise database management professionals

Source: November 2010 Global Database Management Systems Online Survey

58169 Source: Forrester Research, Inc.

Several Options Exist To Control Data Growth — But Some Have Limitations

Given all the problems of unbridled database growth, just adding more CPU resources, memory, and storage to handle data growth in production systems is clearly shortsighted. Although it may be a workable short-term solution, it wastes money over the long haul compared with more-cost-effective solutions. Short of implementing a data archiving solution, some cost-effective options for controlling data growth include:

- **Compressing production data.** Some database management system (DBMS) products have begun to offer the option of data compression, which can reduce database sizes by 50% or more, depending on the data type. Although database compression is free in some DBMS products, other vendors such as Oracle charge for it. While compression can be a good option to overcome data growth, it may come with tradeoffs such as demanding more system resources and requiring CPU and memory upgrades.

- **Deleting unwanted data entirely.** Most databases contain unwanted data — such as activity logs, federated data, or duplicated data — which typically accounts for 15% or more of the

overall database size. Removing this unwanted data can limit data growth to some extent, but doing so often requires a deep understanding of the data model and application impact.

- **Using tiered storage on production systems.** Top DBMS vendors such as IBM, Microsoft, and Oracle offer the ability to use tiered storage by allowing the database to use different storage tiers for different tablespaces within the same database. This allows some tables to reside in high-speed storage area network (SAN) storage and others in lower-cost storage tiers. This approach can reduce storage cost but often requires careful planning and more effort in managing production databases.

DATA ARCHIVING REMAINS THE BEST OPTION TO TAME PRODUCTION DATA GROWTH

Data archiving is the best strategy for controlling data growth in production OLTP applications in both the short and long term. It delivers a powerful data management framework that includes data movement, storage, security, and purging of inactive data. Forrester estimates that 75% of the data stored in large mission-critical OLTP applications is typically inactive, rarely accessed by any user, process, or application. We also estimate that 90% of all data access requests in production OLTP applications are serviced by new data — often data that is less than a year old. Keeping data online for two years in production and moving the rest to an archival repository helps keep databases more manageable while continuing to meet transactional applications' needs. Database archiving is a subset of data archiving; it focuses mainly on moving, archiving, and purging data in production databases.

Database archiving is:

> *Moving inactive data in production databases to an archival repository for long-term retention, while maintaining transactional integrity and regulatory compliance at all times.*

Data archiving moves inactive data and files from production systems' high-cost storage to lower-cost storage and servers for a few years before permanently storing it on tapes and other media for long-term retention. Understanding the logical database design and relationships among tables and objects remains critical. Although one could write scripts to archive data, these scripts are complex to develop and maintain for larger applications or when supporting an enterprisewide data archiving strategy. Knowing what data is inactive is a critical requirement when implementing data archiving; it's also crucial to know the data retention period, which varies depending on state and country regulations.

Data Archiving Supports Many Use Cases Beyond Solving Problems Of Database Size

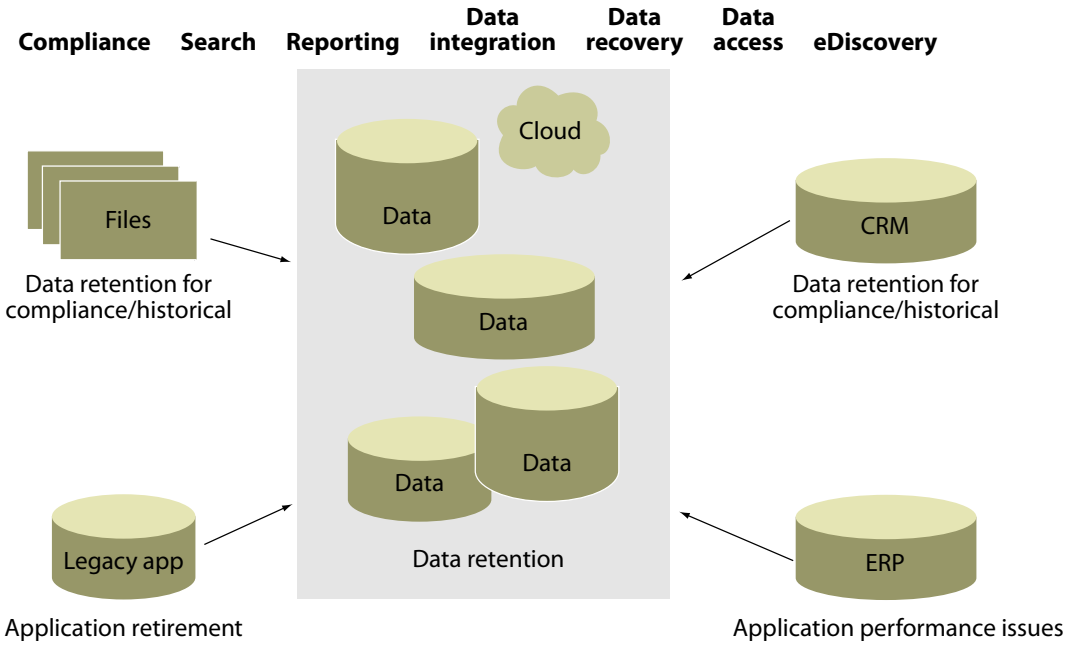
Data archiving does more than fix data-growth issues affecting application performance, cost, and manageability (see Figure 4). It's an essential part of your strategy for managing data through all stages of its life cycle, including data movement, data security, searching, and purging. Data archiving supports many use cases beyond addressing the problems stemming from database size, including:

- **Storing data for long-term compliance requirements.** A key use case for a data archiving initiative is storing data for long-term retention to meet compliance requirements. Growing regulatory compliance requirements such as the Sarbanes-Oxley Act (SOX), Payment Card Industry Data Security Standard (PCI DSS), and the Gramm-Leach-Bliley Act (GLBA) have driven increased archiving in the past three years because of mandates that enterprises retain data for the long term.

- **Retiring legacy applications.** As enterprises retire legacy applications, they must retain most of these applications' data for compliance and historical purposes. Storing this data on mainframes or high-end servers can be very expensive, while putting data on tape makes it difficult to access. Data archiving offers a viable option for low-cost data storage that allows data to remain easily accessible to users and processes.

- **Enabling information sharing of inactive data.** Unlike tapes or flat files, archived data can be accessed via business intelligence (BI) tools such as IBM Cognos, Oracle Hyperion, and SAP BusinessObjects standard structure query language (SQL) queries as well as via Microsoft Open Database Connectivity (ODBC) or Oracle Java Database Connectivity (JDBC) using popular programming environments such as .NET and Java. Archives can centralize inactive data across applications to deliver a unified view of historical data or to support predictive analytics.

Figure 4 Data Archiving Offers Many Use Cases



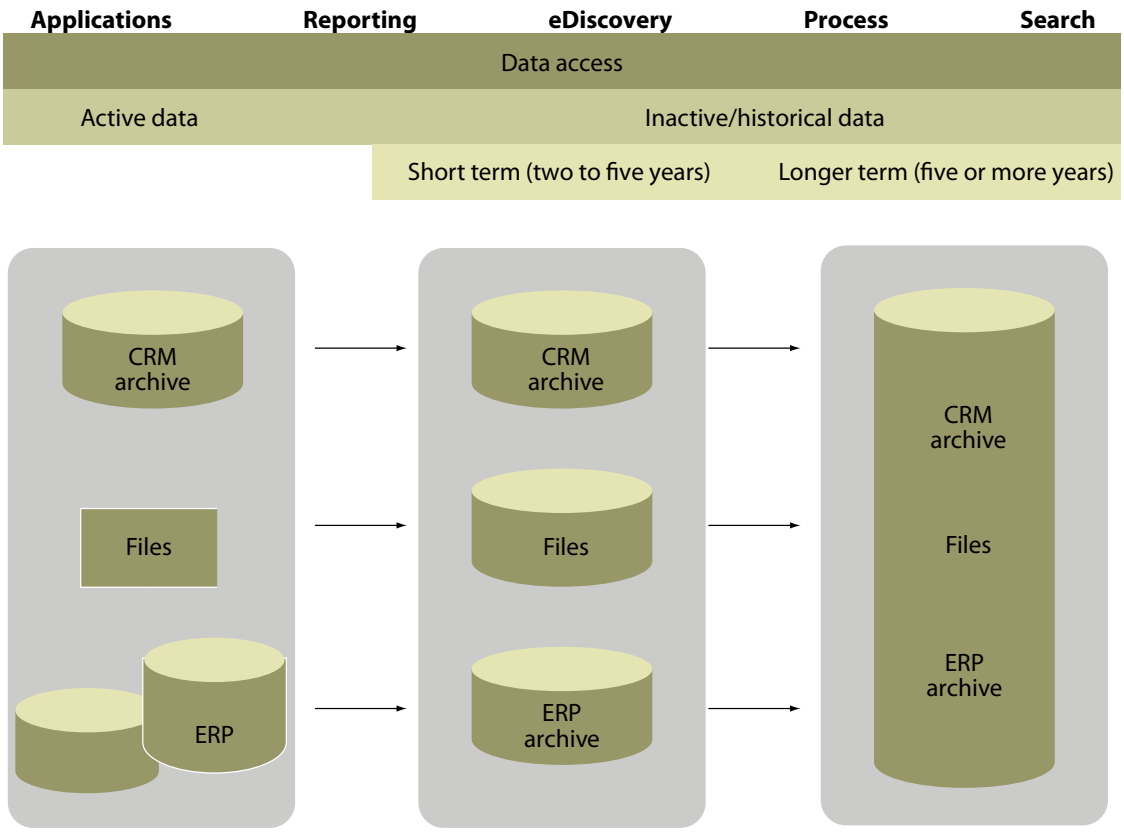
58169 Source: Forrester Research, Inc.

BUILDING A SUCCESSFUL DATA ARCHIVING STRATEGY REQUIRES CAREFUL PLANNING

Your data archiving strategy is a plan to ensure that inactive data is stored, managed, secured, accessed, and destroyed according to the policies you define. Building a data archiving strategy requires the participation of several stakeholders, including data architects, enterprise architects, database administrators (DBAs), developers, information security and infrastructure professionals, business analysts, and business users. Some organizations are not clear on what inactive data they must retain and what they may archive or purge. As a result, these organizations avoid archiving for fear of failing to meet compliance requirements. In addition, too many organizations manage storage reactively as opposed to proactively, hiding the cost implications of failing to act. Once established, an archiving strategy helps clearly define what data you must retain, what you can archive or purge, and how users should access archived data.

Your data archiving strategy should focus on archiving inactive data after two years, moving it to lower-cost tiered storage (see Figure 5). After retaining that archived data for another period, typically three years, consider moving it to longer-term offline storage such as tapes — though your policies may differ depending on the need to keep data online versus offline.

Figure 5 Data Archiving Architecture



58169 Source: Forrester Research, Inc.

The 10 Essential Components Of Building An Unbreakable Data Archiving Strategy

To build an effective data archiving strategy, you must understand not only the data but also the business requirements and compliance mandates. Forrester defines 10 essential components for building an effective data archiving strategy:

1. **Data discovery.** An essential requirement for any data archiving strategy is having a deep understanding of source applications' data and metadata. This involves discovering data relationships and determining inactive data. Without deep knowledge of this data and metadata, archiving can fail or even cause data corruption. Leverage third-party tools to automate the process of discovering inactive data.

2. **Data retention.** Determine how long you need to keep archived data, as this factor strongly influences your strategy's design. A strong understanding of compliance regulations affecting your organization is a critical prerequisite to any data archiving initiative. This involves understanding what data needs to be stored for how long. For example, one must retain public company financial records for at least seven years, federal email for 10 years, and hospital medical records for at least three years. And don't forget that compliance regulations often differ by country or region and might prevent some data from going outside of the country or region, such as private information from citizens of the EU.

3. **People.** A successful data archiving initiative requires collaboration from several key stakeholders, including data architects, enterprise architects, developers, DBAs, information security and infrastructure professionals, business analysts, and key business users. In addition, ensure that the data archiving project team meets regularly to discuss process, technology, and implementation approaches.

4. **Process.** The archiving process should be well defined — including defining the process for data discovery, access, movement, searching, reporting, and purging of inactive data — before you implement it. Focus on the *what*, *how*, *who*, and *when* of data archiving: Your strategy must address how the archiving process will work, what data you will extract, where you will move it, and how long you will keep it. Archive data on a regular basis to meet your storage goals and improve your data access consistency.

5. **Technology.** One key decision you must make is whether to build or buy an archiving solution. Building a solution often takes from six to nine months to support only some of your applications, and this approach requires a development team and ongoing maintenance. In addition, building a solution brings a higher risk of data corruption, especially for complex applications. Buying a third-party archiving solution offers many benefits, including greater ease of use, support for very large applications, higher performance, predefined compliance reports, integration with existing applications for seamless data access, process automation, and enterprisewide scale to support hundreds or thousands of applications and databases.

6. **Data access.** It's best to provide several options to access archived data, especially where many business users need to gain access to the data. This includes providing access to the archived data using production applications and predefined reports. Before defining your data access approach, consider all business users and their requirements to access the archived data.

7. **Data repository format.** It's usually best to store archived data in the same type of DBMS as the production system, which enables it to integrate as well as be used for recovery if necessary. For example, if you have SAP running on an Oracle database for your enterprise resource planning (ERP) application, use an Oracle database as your archival data repository. Other archival repository options include flat files and XML databases. Regardless of the format, make all archived data read-only to ensure accuracy, especially to support eDiscovery compliance requirements.

8. **Archived data storage.** A critical component of any data archiving strategy is the choice of storage media. Key decisions include whether archives from multiple applications will reside in a single archived data repository, multiple repositories on the same server, or even different servers. Also consider whether you want to store data using cloud options such as Amazon EC2 or other public clouds; if so, you must determine how users will access the archived data in the cloud.

9. **Data governance.** This important component ensures the archived data's protection from misuse and guarantees that only authorized personnel can access it. Treat data governance of inactive data as seriously as you would treat governance of active data. Assign a data custodian or information security professional to ensure that you will manage and access archived data in a manner consistent with your compliance and business policies.

10. **Cost.** Organizations often archive to reduce operating expense (OPEX), save on storage and servers, and shrink staffing requirements. When building an archiving strategy, understand the cost breakdown on production systems as well as how best to save money on an archived system. Consider lower-cost storage, data compression, and lower-cost servers for an archived system. Also, don't just consider the acquisition cost of the archived solution; take into account the overall total cost of ownership (TCO). When using a third-party solution, most organizations typically achieve payback within six months.

THIRD-PARTY ARCHIVING SOLUTIONS ARE ESSENTIAL FOR AN ENTERPRISEWIDE STRATEGY

Although archiving may seem a simple process of moving and purging data in production systems, it can require significant effort and resources, depending on the scale, and it can also necessitate changes to your culture, policies, and processes.

Leading Third-Party Vendors Deliver Reliable And Scalable Archiving Solutions

HP, IBM, and Informatica lead the database archiving market, with strong growth and good features and functionality to support most database archiving initiatives. Open Text and Solix continue to enhance their solutions to focus on enterprisewide capabilities and broader support for applications.

In addition, Oracle and SAP offer basic application archiving capabilities, focusing mainly on their own packaged applications. Key data archiving solutions include:

- **HP Database Archiving.** HP acquired OuterBay five years ago to expand on its data management capabilities, including support for database and application archiving. HP Database Archiving supports packaged and custom applications across mainly distributed platforms such as UNIX and Windows. HP offers a tighter integration of the archiving solution with HP's storage and hierarchical storage management (HSM) solutions. In addition to database and application archiving, HP also offers file, email, and messaging archiving solutions. Although HP has not been as aggressive as IBM and Informatica in the data archiving space, it offers a viable platform to support most customer requirements. Its large customer base includes 7-Eleven, Agilent, Aramark, BellSouth, Home Depot, McGraw-Hill, MetLife, Motorola, NCR, O2, and Sony.

- **IBM InfoSphere Optim.** IBM's acquisition of Princeton Softech three years ago made a significant contribution to IBM's data management offering. Today, IBM continues to lead the industry with the most comprehensive data archiving solution and the largest installed base. In addition, it has the broadest support for packaged applications and DBMSes, including strong support for mainframe-based databases. IBM's customer list includes American Airlines, Bank of America, Blue Cross Blue Shield of Illinois, Boeing, BT, Honda, DirectTV, Johnson & Johnson, Prudential, and Toshiba. In addition to archiving, IBM offers a suite of integrated data management products to support subsetting, test data privacy, test data management, application retirement, data migrations, and data governance. IBM's customers spoke highly of the Optim solution's reliability and strong performance.

- **Informatica Data Archive.** Until its acquisition of Applimation in 2009, Informatica did not offer a data archiving product. The acquisition has helped Informatica expand its information management capabilities and expand its market significantly in various vertical industries. Informatica Data Archive mainly supports structured data archiving, although Informatica does partner with EMC to archive unstructured data such as emails and messaging. Besides archiving, Informatica offers support for test data privacy and subsetting. In addition, earlier in 2010, Informatica announced support for archiving to the cloud via integration with Amazon EC2 and other public clouds. Its large customer base includes ADP, Aetna, AT&T, Bank of America, Canon, Citigroup, Dell, EMC, GE, IKON Office Solutions, Michelin, and United States Postal Service. Informatica's customers like its ease of use and high-performance capabilities.

- **OpenText Data Archiving for SAP Solutions.** Open Text was founded in 1991 and is a publicly traded company with 2010 fiscal year revenues of $912 million. SAP certifies and sells the database archiving solution OpenText Data Archiving for SAP Solutions, which archives a complete range of business-critical documents and transactional data for SAP. In addition to SAP applications, OpenText offers support for archiving files, unstructured data, content, emails, and messages. OpenText Data Archiving for SAP Solutions has customers including Delhaize

Group, Husky Energy, Northrop Grumman, Owens Corning, Smurfit-Stone, TXU Energy, Volvo Aero, and World Kitchen.

- **Solix.** Solix is the only remaining independent vendor of database- and application-level archiving, and it has a good track record supporting enterprise accounts. Founded in 2001, Solix supports DB2, Informix, MySQL, Oracle, and SQL Server databases and applications such as Baan, JD Edwards, Oracle E-Business Suite, PeopleSoft, SAP, Siebel, and custom applications. Besides archiving, it offers support for test data management, test data privacy, and application retirement. In addition, earlier in 2010, Solix announced the availability of Solix ExAPPS, the industry's first application retirement appliance to simplify and reduce the cost of legacy applications. Solix offers a viable archiving solution and is a likely target of acquisition in the coming years, most likely by EMC, NetApp, or Oracle. Solix customers include Forbes Marshall, Helen of Troy, Korea Telecom, Lakshmi Machine Works, *Reader's Digest,* Rediff.com, Traco, Tibco Software, and University of Pennsylvania.

Consider More Than Just Basic Archiving Features When Choosing A Third-Party Solution

When choosing an archiving solution, consider:

- **Data integrity.** Ensure that the archiving solution maintains data integrity, taking into consideration entities and relationships, primary and foreign keys, and other constraints. Without data integrity, data can get corrupted or lead to inconsistent results.

- **Application knowledge.** A key strength of archiving solutions is the ability to support leading packaged applications, such as JD Edwards, Oracle E-Business Suite (EBS), PeopleSoft, and SAP, out of the box. These solutions automatically understand the data model and automate the archiving process with minimal effort. However, if the application has been customized, the data architect would have to be involved to help with customizing the archiving data model to reflect those changes. Look for solutions that can support most of your packaged applications running on mainframe or distributed systems.

- **Performance.** Some solutions are better than others are when it comes to moving and purging millions and billions of records. If you plan to run archiving regularly and have a set window for data movement in order to minimize impact to production activities, benchmark the solution to ensure that it can move and purge data within that window.

- **Ease of use.** This criterion is critical when implementing an enterprisewide data archiving strategy that spans many applications and databases. A point-and-click user interface with predefined configuration can improve data architect and DBA productivity. Look at solutions that require no or minimal training and support hundreds or thousands of databases from the same console.

- **Enterprisewide deployment.** An archiving solution should be able to support hundreds or thousands of databases and applications, ensuring automation for the complete archiving process. Look for solutions that require minimal effort when supporting an enterprisewide strategy.

- **DBMS support.** Although most leading archiving solutions support the leading DBMSes, ensure that your choice has credible references, especially on mainframe platforms such as z/OS and VMS if these are important to you. Look at solutions that can support all of your DBMSes, even if currently you are not using archiving for all of them.

- **Data access.** A key differentiator among the leading archiving solutions is the ability to access the archived data using existing applications and simplified data access using SQL and common reporting tools such as IBM Cognos and SAP BusinessObjects. Look for solutions that offer a range of options for data access, including support for the data access standards that matter to you, such as JDBC, Microsoft OLE DB, ODBC, etc.

- **Data repository flexibility.** Although most organizations use databases to store archived data, you should also consider XML and flat-file options. In addition, ensure that you can store the data repository in any leading DBMS of choice, such as DB2, Oracle, SQL Server, or Sybase.

- **Repository security.** Ensure that the solution offers strong security measures, such as auditing, monitoring, encryption, and access controls to the archived data repository. The solution should be able to control the repository and report any suspicious access or changes.

- **Role-based access.** DBAs do not implement data archiving solutions; they involve several other roles, such as data architects, information security professionals, and enterprise architects. Ensure that the solution offers role-based access supporting the various roles you plan to assign to data archiving deployment.

- **Compliance-driven search and reporting.** A key reason why enterprises archive data is to support various compliance requirements such as SOX, PCI DSS, GLBA, and state laws. Ensure that the solution can perform compliance-driven searches and generate reports with minimal effort.

- **Data compression.** Although archived data is often stored on lower-cost storage, ensure that the solution can compress data to conserve space.

- **Integration with HSM/storage solutions.** Ensure that the solution can integrate with HSM/storage solutions to offer control over data movement, storage, and retention.

- **Automation.** A higher degree of automation helps minimize human error and offers scale across hundreds or thousands of databases and applications.

- **Flexible frequency.** Ensure that the solution can deliver a flexible frequency ranging from daily, weekly, or monthly to custom intervals.

BEST PRACTICES IN DATABASE ARCHIVING

Although archiving may seem a simple process of moving and purging data in production systems, it often requires changes to culture, policies, and processes. Key best practices include:

- **Don't archive everything.** Archiving everything is a waste of storage, resources, and effort. Why archive 10,000 tables when only 1,000 are relevant? Understand the data model, and determine which tables and rows are relevant in the archival process.
- **Remember that backups are not archives.** A data backup is solely for recovery of data; it does not deliver a platform required to support inactive data across its various stages of archiving. Unlike database backups that work at the table and database level, archives work at the record level, identifying those that the system should retain, archive, or store offline.
- **Test recovery of archived data.** In rare circumstances, it may be necessary to restore archived data to production because of data corruption issues or other integration reasons. Be prepared to test the recovery of archived data, ensuring that the solution cleanly moves data back to production without any integrity issues.
- **Use lower-cost storage for archived data.** You defeat the purpose of archiving if you move production data onto an archived system that also contains high-speed, expensive storage and high-end servers. Consider lower-cost storage devices and disks, such as Serial Advanced Technology Attachment (SATA) drives or a hosted cloud platform such as Amazon EC2.
- **Keep an eye on data integrity.** Forrester has encountered several large implementations that purged data incorrectly as part of the archival project. After days and months of purging, this created data integrity issues that were a challenge to repair.
- **Be proactive rather than reactive in archiving.** Don't archive just because the database size increased dramatically or auditors plan to audit your environment in the coming weeks. Rather, implement archiving proactively as part of normal processes, running your archiving process at scheduled intervals.
- **Monitor the archival process regularly, and set it to alert administrators to any errors.** Data archiving processes can be very sensitive to changes in schema structures, database and application upgrades, changes to configuration and system patches, and changes in network speeds. Check for any errors, and keep a count of records archived and purged.
- **Avoid archiving too frequently.** Schedule archiving weekly, monthly, or quarterly, but don't do it daily or hourly, as archiving this frequently can adversely affect production performance and data consistency.

- **Consider using data compression in production and archival repositories.** Data compression in databases is a mature feature that can dramatically reduce the storage requirement by 50% or more. Based on customer feedback, Forrester finds that compression does not often adversely affect performance and in many cases actually improved performance by providing the ability to fetch more rows in a single disk I/O block.

- **Use archiving tools, especially with large and complex applications.** Archiving tools can help enormously by automating the archiving process, ensuring data integrity, and moving and purging data, especially when dealing with large or complex custom or packaged applications.

DATA ARCHIVING IS POISED FOR STRONG GROWTH AND MORE INNOVATION

In Forrester's November 2010 Global Database Management Systems Online Survey, 33% of respondents stated that they were currently pursuing data archiving, with another 18% looking to do archiving in 2011 — a growth of more than 50% year over year. From a technology perspective, future data archiving solutions promise to deliver even more automation, ease of use, role-based tooling, and higher performance. Over the next three to four years, Forrester expects:

- **Tighter integration with private and public clouds.** Although some vendors such as IBM and Informatica already offer cloud archiving solutions, more integration lies ahead. We are likely to see stronger integration of inactive data across on-premise and cloud platforms to enable users to search, access, and move such data.

- **A common solution that supports structured, unstructured, and semistructured data.** One should not have to distinguish between these different data types for compliance or search. Although currently no archiving solution provides comprehensive support for all data types including structured, unstructured, and semistructured, we are likely to see that change in the coming years as vendors such as EMC, HP, IBM, and Informatica ramp up their efforts to deliver a common and unified solution.

- **Role-based tooling.** Archiving involves multiple people for planning, implementation, and support, including data architects, DBAs, developers, enterprise architects, and business analysts. We are likely to see more-powerful role-based tools supporting archiving projects, allowing each role to interact with the solution differently based on its requirements.

- **Greater automation.** Although most of the leading archiving solutions offer automation, the future promises even more automation to support self-managing, self-securing, and self-integrating capabilities.

- **Tighter integration with storage and other media.** Through the process of archiving, inactive data moves across different types of storage media, such as SAN, network attached storage (NAS), and direct attached storage (DAS) systems. In the near term, we will see archiving

solutions taking advantage of vendor-specific storage systems, offering greater reliability, performance, and availability.

- **Application vendors offering more comprehensive support.** Even though SAP and Oracle offer a basic data archiving capability, they do not tightly integrate archiving with the DBMS and underlying storage subsystems. In the future, application vendors are likely to offer more-comprehensive archiving solutions that will seamlessly access archived data and offer tighter integration with storage solutions.

- **Data security vendors integrating with archiving solution providers.** As adoption of data archiving grows, data security vendors such as ArcSight, IBM, LogLogic, and RSA will integrate their security solutions to support compliance and eDiscovery initiatives for inactive data.

- **Information-as-a-service (IaaS) vendors integrating with archiving solutions.** Although today's IaaS implementations are mostly focusing on active data, we believe that in the future, being able to access inactive data will become an essential requirement, especially to support compliance initiatives and predictive analytics.[1]

- **A common repository that serves more applications and databases.** Today, most implementations store archived data from different applications in separate data repositories, mainly to offer easier management and access. In the near future, we are likely to see solutions that will support common data repositories and integration across hundreds of applications, creating an information fabric layer for inactive data.

RECOMMENDATIONS

AN ENTERPRISEWIDE DATA ARCHIVING STRATEGY HAS BECOME CRITICAL FOR ALL

It's never too late to start an archiving initiative to help control data management cost, meet compliance requirements, and improve operational efficiencies. However, a successful data archiving strategy requires careful planning regarding what, how, where, and when to archive. As part of their archiving strategy, architects should:

- **Identify potential large databases.** Any OLTP production database that contains more than a terabyte of data or data warehouse that contains more than five terabytes of data is a good candidate for archiving. In addition, these databases should have data growth of at least 20% annually for data that will be required for long-term compliance retention purposes.
- **Understand the data.** If you don't understand the data model or data, you are running a high risk of potentially corrupting data, as the archiving process purges production data. Understand what data is active and what's not as well as the best approach to archive data that's more than two years old.

- **Develop an economic justification.** Understanding archiving's long-term benefits will help justify short-term investment. Although long-term savings can be considerable, demonstrating some short-term positive impact will be critical in securing funding support for your archiving initiative. Understand the hardware and software cost on production systems as well as how archiving can help reduce it over time.
- **Get management and business owners to buy in.** Without support from management and business owners, an archiving initiative is likely to fail. Business owners and information security personnel are the best people to help determine when to retain data, what data to retain, and how long the organization should retain it.
- **Develop the plan.** Without a proper plan that involves the appropriate stakeholders, any archiving project is likely to run into issues. Develop a plan that includes architectural considerations, various approaches, policies, data access, and data governance of archived data — and test the process before deploying it into production.
- **Roll out the strategy across various applications.** Don't hold back on just one or two applications; focus on an enterprisewide archiving strategy across hundreds of databases and applications.
- **Use a cloud platform to lower cost.** Cloud platforms offer an attractive cost-value proposition that you should leverage wherever possible, but bear in mind appropriate data security practices. When storing data in a public cloud, ensure that it is fully encrypted and protected at all times.
- **Use third-party vendor solutions — which are necessary for any archiving initiative.** Although an organization could build a data archiving solution by itself, third-party vendors such as HP, IBM, Informatica, OpenText, and Solix deliver a strong value proposition that ensures reliability and security, minimizes the risk of data corruption, and lowers costs through automation and data compression. Although archiving solutions require a capital expenditure (CAPEX), they often save 20% or more in hardware and software cost over time and tend to provide payback in less than six months.

SUPPLEMENTAL MATERIAL

Methodology

Forrester's November 2010 Global Database Management Systems Online Survey was fielded to 194 database management professionals. For quality assurance, participants are required to provide contact information and answer basic questions about their firms' revenue and budgets.

Forrester fielded the survey from September to November 2010. Respondent incentives included a summary of the survey results and a copy of the final report.

Exact sample sizes are provided in this report on a question-by-question basis. Panels are not guaranteed to be representative of the population. Unless otherwise noted, statistical data is intended to be used for descriptive and not inferential purposes.

Companies Interviewed For This Document

BP
EMC
HP
IBM
Informatica
Oracle

ENDNOTES

[1] Forrester's information fabric vision of enterprise information virtualization has evolved as enterprises have implemented more parts of the vision and as vendors have extended information-as-a service products to deliver broader capabilities. Although a complete information fabric implementation is still rare, many large enterprises already have some elements and a number of information services deployed or are planning deployment over the next two to three years. See the April 9, 2007, "Information Fabric 2.0: Enterprise Information Virtualization Gets Real" report.

FORRESTER®

Making Leaders Successful Every Day

Headquarters

Forrester Research, Inc.
400 Technology Square
Cambridge, MA 02139 USA
Tel: +1 617.613.6000
Fax: +1 617.613.5000
Email: forrester@forrester.com
Nasdaq symbol: FORR
www.forrester.com

Research and Sales Offices

Forrester has research centers and sales offices in more than 27 cities internationally, including Amsterdam; Cambridge, Mass.; Dallas; Dubai; Foster City, Calif.; Frankfurt; London; Madrid; Sydney; Tel Aviv; and Toronto.

For a complete list of worldwide locations visit www.forrester.com/about.

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or clientsupport@forrester.com.
We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Research, Inc. (Nasdaq: FORR) is an independent research company that provides pragmatic and forward-thinking advice to global leaders in business and technology. Forrester works with professionals in 19 key roles at major companies providing proprietary research, customer insight, consulting, events, and peer-to-peer executive programs. For more than 27 years, Forrester has been making IT, marketing, and technology industry leaders successful every day. For more information, visit www.forrester.com.

FORRESTER®