Anatomy of an archiving project

Contents

Executive summary

Business critical ERP, CRM and other transaction processing applications represent an ongoing corporate investment. These applications empower new business initiatives, support daily operations and drive revenue generation. They also collect and accumulate a wealth of business information – a valuable asset that must be managed and preserved throughout its lifecycle.

Companies worldwide are affirming the value of database archiving as a best practice for managing data growth, completing faster application upgrades and supporting data retention compliance initiatives. In fact, enterprise data management that includes database archiving as a best practice can deliver business value and a greater return on investment from your mission-critical business applications.

So, what are the basics of an archiving project? What are the benefits? What factors should sites consider to successfully implement a full-lifecycle data management strategy? As with any major project, up-front planning drives success.

This white paper explains how the seven basic principles for managing enterprise application data can help your organization:

• Establish effective policies for full-lifecycle enterprise data management to control data growth and lower storage costs.

• Meet service level goals to ensure the timely completion of key business processes for mission-critical applications.

• Support data retention compliance initiatives and mitigate risk for audits and e-discovery requests.

• Implement scalable archiving strategies that easily adapt to your ongoing business requirements.

The IBM® Optim™ Data Growth Solution is a consistent, scalable solution that includes comprehensive capabilities for managing enterprise application data across applications, databases, operating systems and hardware platforms. Learn how you can align the management of your enterprise application data with your business objectives to improve application service levels, lower costs and mitigate risk.

What is enterprise data management?

Your enterprise applications simply must deliver measurable business value. So how can you capitalize on your investment? The answer rests in your ability to align continuous control of your application data with your business objectives through the power of enterprise data management.

More companies are realizing the benefits of a full-lifecycle enterprise data management strategy that includes database archiving to derive more business value and provide more options for:

- Managing continued application data growth and its impact on service levels, operational costs and risks, as well as storage requirements.

- Aligning application performance and service levels to meet operational and business objectives, for timely reporting and processing, and ensuring the delivery of information to decision makers when and how they need it.

- Supporting data retention compliance requirements and effective storage strategies that keep historical transaction records accessible until legal retention periods expire.

- Improving application portfolio management to decommission redundant applications and simplify the IT infrastructure.

Why archive?

According to a 2006 study conducted by the Enterprise Strategy Group (ESG), database archiving has helped companies mitigate the harmful impact of rapid data growth.[1] Archiving strategies, whether applied to ERP, CRM, custom or legacy applications, are shown to provide benefits to both the IT organization and the lines of business. Clearly, archiving has "crossed the chasm" to become an operational best practice.

The research also reveals that enterprises are leveraging the deployment of archiving to address issues beyond data growth. For example, companies are applying archiving strategies to improve the application upgrade process. Archiving historical application data prior to an upgrade reduces the amount of data to convert and substantially curtails downtime.

In short, enterprise data management and database archiving capabilities deliver additional benefits that support IT and business objectives. However, before you can get started, consider the following capabilities as essential for the success of any archiving project:

- **Archiving complete business objects.** Focus on ensuring the integrity of the archived data in its complete business context. For example, archiving a complete "paycheck" would include time and labor transaction details, as well as the employee master data needed to fulfill reporting requirements without ties to production data.

- **Supporting full-lifecycle archiving.** Focus on managing archives and retention periods cost-effectively and consistently over the full lifespan of your data. Capabilities for saving archives on a variety of storage media allow you to future-proof methods for managing data, based on its business value and access requirements. Keep business records accessible until legal retention periods expire and archives can be deleted.

Managing your archiving project

The process of planning and sustaining an effective database archiving project must include some basic project management considerations. The following guidelines will be helpful:

• Involve all stakeholders to align the business and legal requirements with the technology infrastructure to meet them. Establish clear lines of accountability and individual responsibility. Ensure that IT, business units and compliance professionals work together.

• Establish common objectives for promoting archiving best practices within your organization. It is important to ensure that business users are appropriately involved with IT and informed about how their information will be managed and how their business requirements for data access will be met.

• Monitor, review and update documented archiving policies and procedures. Continue to improve archive processes to support your ongoing business objectives for providing appropriate application service levels and supporting retention compliance requirements and reducing costs.

Seven basic principles for archiving application data

Every archiving project begins with asking pertinent questions, as you consider a variety of options for managing enterprise application data. After you determine which approach best suits your organization's requirements, you can evaluate solution alternatives that support that approach.

Field experience has shown that there are seven basic principles for archiving and managing enterprise application data:

Table 1. Basic principles summary

Principle	Description
Assess	Determine which applications are most in need of archiving, grouping them into categories based on your business requirements.
Classify	Document business rules and data retention policies to govern active, inactive and compliance-managed data.
Archive	Segregate historical business objects or transaction records from current activity. Safely move them to a secure archive.
Store	Store archived historical records securely and cost-effectively, according to the evolving business value.
Access	Apply service levels that provide authorized business users with access to the historical records they need, when and how they need them.
Tune	Monitor operations to ensure that archive operations continue to support desired service levels and access requirements.
Dispose	Prevent information assets from becoming information liabilities by deleting historical records after they are no longer required for compliance or business purposes.

Each of these basic principles and the questions you should consider for designing your archiving project are discussed in more detail in the remainder of this paper.

Principle 1: Assess your application portfolio

Accumulating current transactions and retaining historical transactions in the same database causes the volume of application data to increase exponentially. As a first step in assessing your portfolio, determine which applications are demonstrating

the symptoms of rapid data growth: delayed report processing, slow response for ad hoc queries and slow backup and recovery. Are these problems more prevalent for some applications than others? When transaction processing slows down, how is your business affected?

Consider also your future plans for application deployment. Where might you save money by consolidating similar applications or decommissioning systems that no longer add value? What applications and modules need to be upgraded, and when? All of these issues will impact your data management strategy. With these ideas in mind, you can group your enterprise applications according to their unique parameters. Suggested categories are described in Table 2.

Table 2. Assess your applications

Mission critical

Applications that are critical to your daily business operations. An example might be a company's process application for manufacturing consumer goods. Without this application, the company simply cannot produce or ship its products – or generate revenue.

Business critical

Applications that are important to your daily business operations, but can be unavailable for short periods of time. Some examples might include billing, accounts payable and internal reporting systems.

Candidates for retirement

Redundant, obsolete or de-supported "legacy" applications that are no longer actively in use. Not only do these applications add clutter to your IT infrastructure, but they also drain valuable resources and increase operational risks.

When designing your archiving project, keep in mind that each application category will have its own requirements. As an example, imagine a company called Acme Manufacturing, which has identified both mission-critical applications and retirement targets that may benefit from archiving.

Table 3. Application assessment example

Application	Assessment	Module	Business object
Financials	Mission Critical	General	Ledger Journals
Inventory Management	Retirement (Sunset)	Payments	Payment Receipts

Acme's goal for its mission-critical Financials application is to improve service levels and speed transaction processing. For its legacy Inventory Management system, now targeted for retirement, the goals are different. In this case, Acme wants to maximize storage savings, ensure the immutability of the data and enable prompt response to audit or e-discovery requests.

Principle 2: Classify business objects

Business objects, such as journals, ledgers, invoices, payment receipts and purchase orders, represent the basic building blocks of application processing. Classification schemes can be based on any criteria. However, as a simple example, you can classify a business object based on its business value or how often it needs to be accessed. By classifying these objects, you can begin to define the rules for managing them at different stages in the information lifecycle. Consider the following questions:

• What are the post-archive use cases? In other words, who needs access to archives for what purpose, and how fast do they need it?

• Do access requirements change as the archives age?

An example of a data classification scheme, based on business value, is shown in the Table 4.

Table 4. Classify business objects

Critical

Used in key business processes. Required to support minimum acceptable work levels in the event of a disaster. Must be retained for legal reasons.

Vital

Used in routine business processes. Represents a substantial investment of company resources that may be difficult to recoup. May not be required immediately in the event of a disaster. May be considered company confidential or secret.

Essential

Used in routine business operations for which there are alternate sources available. Can be reconstructed fairly easily in case of loss.

Non-critical

Can be reconstructed easily with minimal cost. Duplicates existing data that has minimal security requirements.

Again, archive deployment requirements will vary across applications, business objects and use cases. Continuing with our example, Acme's classification scheme might be planned as shown in Table 5.

Table 5. Data classification example

Application/ module	Business object/ classification	Production	Online archive	Offline archive	Delete
Financials/ General Ledger	Journals/Critical	Years 1 to 2	Years 3 to 5	Years 6 to 10	Yes, after Year 10
Inventory Management/ Payments	Payment Receipts/Vital	N / A	N / A	Years 0 to 7	Yes, after Year 7

Acme plans to archive historical General Ledger transactions in the Financials application. For the first two years, records remain in production. In years three to five, these records move to an online archive, where Finance Department analysts will use them for trend analysis and reporting. In years six to ten, the records move to an offline archive to address retention compliance requirements.

Should Acme receive an e-discovery request, the Controller can query the archives to generate information reports. If necessary, she can even retrieve records from the archive and restore them to an application environment. After year ten, the legal retention period expires, and Acme deletes the records. The approach differs for the Inventory Management system. Here, Acme will archive the Payment records and maintain them in an offline environment for seven years, until deletion.

Principle 3: Archive application data

Archiving is the process of segmenting business objects (application transaction records) by age and status, and then safely moving them into a secure archive. Consider the following questions:

- Does archive processing capture complete business objects, including both transaction detail and master or reference data?

- Does archive processing perform the appropriate functional condition checks?

Archive complete business objects. Conceptually, the complete business object represents a historical "point-in-time" snapshot of a business transaction. This historical reference snapshot must include both transaction details and related master information. For example, the business object "purchase orders" would contain line items (the transaction details), as well as the corresponding vendor name, customer name and credit card number (master and reference data).

After capturing the complete business object, the archive process should also perform the appropriate functional condition checks. These condition checks identify which specific records in a defined group are safe and appropriate to archive. For example, an invoice should not necessarily be archived simply because it is two years old. Before moving to the archive, the invoice must first be fully paid and posted. You should validate that the archive process applies all the necessary condition checks.

Archiving complete business objects ensures that all the related details, master and reference data remain together. When the complete business object is captured, your archives serve as an intact, accurate, standalone repository of transaction history. You can query this repository to respond to customer inquiries or e-discovery requests, without the need to reference information stored in separate schemas.

In contrast, when a history database is simply split off from the production database, it contains only the transaction details. Storing master and reference data separately from transaction details introduces unnecessary risk into your archiving project. Remembering the purchase order example, you risk creating orphans for request-for-quote, contracts, shipments and returns-to-vendor, or any other detail records, where parent records have been changed or deleted.

Acme Manufacturing will validate that its selected archive solution captures the complete Financials and Inventory Management business objects, leaving no orphan records behind.

Principle 4: Store archives throughout the full data lifecycle

Storage decisions play a major role in project planning, and archiving offers the potential for substantial storage savings. Focus your analysis on full-lifecycle archive capabilities so you can easily manage archives and retention periods cost-effectively for three years, seven years and ten or more years into the future. Consider the following questions:

- In what format should we store our archives? Can we compress the archives to maximize storage savings? Can we index archives for faster retrieval?

- How many storage tiers and what types of storage devices should be deployed? NAS or SAN? WORM devices? Optical disk? Tape?

- What business objects should be placed on which storage devices, and for what periods of time? How do we manage archives throughout the full lifecycle?

Alternatives for archive formats. Determine whether to store archives in a database, in a file or perhaps in different formats at different points in time. Each alternative offers benefits. For example, when you begin an archiving project, your staff can probably manage the archive database easily. However, as time progresses, the archive database can balloon to the size of the original production environment, forcing you to archive that database as well. Any potential storage savings are erased, and project management efforts quickly become burdensome.

In contrast, compressed file formats require a smaller footprint, so you can maximize storage savings. Compressed files can also be indexed, enabling rapid retrieval of archived data. Most importantly, managing archives in a file format offers the broadest range of access methods *over the life of the archive*. You can use a variety of industry standard access methods, like ODBC/JDBC, XML and SQL, and reporting tools, like Business Objects™, Cognos® or even Microsoft® Excel®. Leveraging this access path, you "future proof" your archives because you retain access long after the originating application has been upgraded, changed or even retired.

Tiered storage options enable greater cost savings. Consider storing archived data based on its business value and access requirements. Over time, the value of a business transaction naturally declines unless needed to respond to an immediate request. However, as long as transaction records remain within their legal retention period, the business value will never be reduced to zero. Accordingly, Acme must ensure access to its archives throughout the retention period. If the company should receive an audit inquiry, it will require flexible access alternatives to provide a prompt and accurate response.

When Acme Manufacturing first books a sales order, its highest priority is to fulfill that order. Once the order is shipped and payment collected, Acme will shift its resources to fulfilling the next order, and so on. Yet, without an archiving strategy, Acme must use the same expensive storage resources for all orders – regardless of priority.

Archiving enables Acme Manufacturing to deploy a tiered storage strategy and assign the appropriate level of resources to each transaction, based on priority, age, status or other parameters. Tiered storage strategies can be managed manually or using an integrated storage controller, such as IBM Tivoli® or Symantec™, to improve control and administration throughout the lifecycle. An example of full-lifecycle management is shown in Table 6.

Table 6. Tiered storage strategy

Application/module/business object	Storage tier deployment
Financials/General Ledger/ Ledgers and Journals	Online – Years 1 to 2 Nearline – Years 3 to 5; Offline – Years 6 to 10; Delete after Year 10
Inventory Management/Payments/ Payment Receipts	Offline – Years 0 to 7; Delete after Year 7

Acme plans to manage current Financials data for two years in the production environment, on the fastest storage device available. After two years, historical transactions are archived and stored in a less expensive, nearline storage environment, where they remain until the end of year five. In year six, the records shift out of the nearline archive and onto a secure WORM ("Write Once, Read Many") device. They are deleted upon expiry after year ten. Since the Inventory Management application is being retired, it is only necessary to retain the archives offline until the retention period expires after year seven.

Principle 5: Access the archives

Early in the project planning process, you determined your post-archive use cases, that is, who needs access to archived data, what they do with it and so forth. These use cases become the primary driver for selecting the optimal access method. Consider the following questions:

• What access methods are available to us? What are the costs and benefits of each?

• What Service Level Agreements (SLAs) are appropriate for accessing the various archived business objects? Do these SLAs vary over time?

• Would we ever want to retrieve transactions from an archive and reload them into a temporary environment?

Alternatives for accessing archives. Typical alternatives for accessing archived data include native (application-based) access and application-independent access. Some archive solutions provide only one access method or the other, but few provide both.

You need to weigh the advantages of each method for meeting your specific access requirements. For example, native access allows you to interact with archived data through the original application interface. This method allows users to access the information they need, using familiar formats and processes. However, if you plan on retiring an application and still want to retain access to your archives, application-independent access may be the better alternative.

Application-independent access provides the most flexible range of access alternatives over the life of the archive. This approach enables authorized users to interact with archives using industry standard methods, such as ODBC/JDBC, XML tools, and report writers, such as Cognos and Business Objects. Most importantly,

application-independent access allows you to future-proof your data. This approach makes it possible to access historical transaction records long after an application has been retired.

Apply distinct access and service levels. While government regulations require that certain records be retained for specific periods of time, sub-second retrieval is not necessarily a requirement. In fact, archiving makes it possible to apply discrete service levels for different use cases and business objects. Accordingly, Acme Manufacturing will provide more resources and faster service for mission-critical business activities. However, where priorities are less urgent, the company can reduce resource allocation and still meet user needs.

Incorporating both distinct service levels and optimal access methods for each scenario, the company's enterprise archiving deployment is shown in Table 7.

Table 7. Access your archives

Application/module/ business object	Storage tier deployment	Archive access method	SLA
Financials/GL/Ledgers and Journals	Online – Years 1 to 2	Native	On Demand
Financials/GL/Ledgers and Journals	Nearline – Years 3 to 5; Offline – Years 6 to 10; Delete after Year 10	Application Independent	Standard reports on demand Ad-hoc queries, based on access capabilities of storage hardware
Inventory Management/ Payments/ Payment Receipts	Offline – Years 0 to 7; Delete after Year 7	Application Independent	Standard reports on demand; Ad-hoc queries, based on access capabilities of storage hardware

Principle 6: Tune archive operations

Tuning involves monitoring operations to ensure that archiving processes operate efficiently and continue to support service level goals. Consider the following questions:

- How quickly does data continue to accumulate? How often, and when, should we archive?

- Should our archive processes be automated or manual? Online or offline?

- Does archive administration have a native application look and feel? Does processing incorporate our site-specific application customizations?

- Do our routine archive processes operate according to our expectations? How well is database archiving supporting our business objectives?

Determine the optimal frequency for archive processing. Periodically monitoring data growth in the application database is the best way to determine the optimal frequency for archive processing. Your goal is to maintain response times and application availability to support your defined SLAs. Continued rapid data growth may signal a need to increase the frequency of archiving or to expand the scope to incorporate additional business objects.

Databases are architected to perform at high speed when you add or update data, not necessarily when you delete it. The process of deleting data is "expensive," meaning that it creates overhead for the database. Therefore, it may sometimes be necessary to evaluate indexing and partitioning strategies and tune the SQL to make the delete process run more efficiently. With these points in mind, Acme's IT group will periodically review the effectiveness of their archiving operations, looking for ways to refine their deployment for even greater business value.

Evaluate archive administration and controls. Ensuring archiving administration, control and job separation are important considerations. You should verify that the archive process incorporates and respects your site-specific customizations. For example, the solution should offer a clean visual editing environment where custom tables can be added easily. As validation, preview the table traversals, including the custom tables, before executing the archive process. Verify also that the records to be archived are complete and accurate.

Give thought to an archiving solution that allows manual processing, but also provides integration into job schedulers for automated processing. Keep other operational requirements in mind as well. Do available windows allow enough time to take your application offline for archiving? Or, would you benefit from the flexibility of online processing alternatives? Acme will use a third party batch job scheduler to run archive processes. General Ledger data will be archived annually.

Principle 7: Dispose of archived data

In a business climate conditioned to strict retention regulations, the concept of data disposal seems counterintuitive. Both business executives and IT managers hesitate to delete application records for fear of legal repercussion. However, it is not only expensive to "retain everything forever," it is also risky. Any data that exists can become a target for discovery. As a prudent and cost-effective alternative, organizations are mapping retention policies to defined business objects, enabling an orderly disposal. Consider the following questions:

• Who decides the retention policies for each business object?

• How should we dispose of archived data after the retention period expires? Can the process be automated?

• What audit trails will prove that data has been deleted according to our policies?

Promote cross-functional ownership. Typically, the business units own the data and set the data management policies, while IT owns the infrastructure and controls data management processes. Accordingly, business managers are responsible for defining who can touch the data and what they can do with it. IT must implement a technology infrastructure that supports these business policies. Promoting cross-functional ownership for data management, archiving, storage and retention policies is perhaps the greatest indicator of project success because all groups have a vested interest in a positive outcome.

In practice, organizations build cross-functional teams to define data management and archiving policies. So that all relevant issues can be considered, these teams will commonly include line-of-business representatives, application owners, DBAs, storage administrators, legal counsel, risk managers and other subject-matter experts. Business managers can define use cases, legal counsel can map retention requirements to deletion policies, and IT can validate that the archive process meets the defined requirements.

Plan and practice an orderly disposal. After all stakeholders have signed off on the archiving and data retention policies, IT can develop a plan to implement those policies. Consider solutions that generate notification reports, identifying which archives are nearing expiry.

You may want to initiate the delete process manually at first, until the practice of deleting expired data becomes more comfortable. Also consider a solution that allows you to verify the data targeted for deletion before running the delete process. Later on, you may want the option to progress to automated deletion upon expiry. Finally, ensure that your solution provides an adequate audit trail, so that you can verify compliance to your stated deletion policies.

To conclude our example, Acme's CIO consulted with both General Counsel and the CFO to determine how and when to delete expired data. The group was most concerned about General Ledger transactions, which are subject to legislative requirements. Based on the General Counsel's input, Acme will delete its General Ledger archives upon expiry; that is, after 10 years. Because the Inventory management application was not subject to the same retention requirements, the deletion policy was based on the expected possibility of a business inquiry, and a seven-year deletion period was considered appropriate.

Archiving: the time to begin is now

Effective enterprise data management strategies that include database archiving capabilities provide companies with a full-lifecycle approach for managing enterprise application data. The basic principles for managing an archiving project presented in this paper can help you initiate and plan an archiving strategy that meets your organization's requirements.

Database archiving is a recognized best practice for managing data growth that offers significant advantages. Only the IBM Optim Data Growth Solution provides proven, full-lifecycle capabilities for archiving and managing application data, with the broadest range of implementation options to ensure your success.

Now is the time to exploit the power of enterprise data management to realize measurable business value across your enterprise:

- Align application performance with business processes and profit from superior performance and availability.

- Simplify database administration, ensure business continuity and speed disaster recovery.

- Streamline application and database upgrades and reduce resource requirements for key IT operations.

- Automate data retention to support compliance initiatives and respond quickly and accurately to audit and discovery requests.

- Leverage existing investments in applications, databases and storage and eliminate IT budget variances.

Optim supports all leading enterprise databases and operating systems, including IBM DB2®, Oracle®, Sybase®, Microsoft® SQL Server®, IBM Informix®, IBM IMS™, IBM VSAM®, Microsoft Windows®, UNIX®, Linux® and IBM z/OS®. And it supports the key ERP and CRM applications in use today – Oracle® E-Business Suite, PeopleSoft® Enterprise, JD Edwards® EnterpriseOne, Siebel® and Amdocs® CRM, as well as your custom and packaged applications.

Optim provides the flexibility to manage large volumes of data over long periods of time, allowing you to deploy whatever business policies you determine are appropriate for managing your valuable application data. Our enterprise data management experts and our partners work with your project team to define business objectives and processes for managing enterprise application data in your environment. After implementation, you can count on continued optimization and technical support.

About IBM Optim

IBM® Optim™ enterprise data management solutions focus on critical business issues, such as data growth management, data privacy compliance, test data management, e-discovery, application upgrades, migrations and retirements. Optim aligns application data management with business objectives to help optimize performance, mitigate risk and control costs, while delivering capabilities that scale across enterprise applications, databases and platforms. Today, Optim helps companies across industries worldwide capitalize on the business value of their enterprise applications and databases, with the power to manage enterprise application data through every stage of its lifecycle.

For more information

To learn more about IBM Optim enterprise data management solutions, contact your IBM sales representative or visit: **www.optimsolution.com**

IBM