



January / February 2008

E-Discovery and Enterprise Data Management – What You Need to Know

By Eric Offenberg
IBM Optim Product Marketing Manager

Brave new world of e-discovery

New products, new technologies, new legislation — and new liabilities! With the advent of stronger laws that focus on data retention and accountability, such as Sarbanes-Oxley, SEC-17a, Basel II and others, companies across industries face a much higher risk for audits and potential lawsuits. Along with increased risk, the associated costs can multiply into millions of dollars.

A key factor in satisfying audit requirements and defending lawsuits, or at a minimum, avoiding costly penalties, is having the capability to retrieve the appropriate information and supporting evidence to meet audit, legal and business requirements. However, with the rise of the “new data center” and “paperless” application environments, almost 95 percent of all business information is created and stored electronically, hence the evolution of electronic discovery (“e-discovery”).

According to Forrester Research, e-discovery is the process of collecting, preparing, reviewing, formatting and producing enterprise content that may exist in either paper or digital form in response to internal, litigation, or regulator requests.¹ Examples include files, databases, email and any other electronically stored information (ESI) that could be relevant evidence in a lawsuit.

While companies recognize that managing business-critical information electronically offers many advantages, the practice also poses many IT challenges. IT organizations need technology with proven capabilities for managing and locating the specific information in order to deliver a fast response to any e-discovery or audit request. Without an effective enterprise data management strategy and early IT involvement to plan for data retention and audit challenges, the way your company responds to an e-discovery request can mean the difference between success and failure.

¹ Barry Murphy and Robert Markham, “eDiscovery Bursts Onto the Scene,” Forrester Research, March 1, 2006.

Why prepare for e-discovery?

The potential for litigation is on the rise, not only in the U.S., but also globally.² The associated legal costs and penalties can be prohibitive. In addition, because most cases settle before going to trial, the discovery phase is often the most critical and expensive part of litigation. For example, it takes time and resources to identify and collect the necessary information.

First, you must identify the necessary data, which may be managed in different application databases operating on different platforms across your enterprise. Hopefully, this data has been maintained in a way that makes it easy to access. Secondly, you must ensure that the integrity of the data and metadata is preserved so that it can be retrieved, accurately and unaltered in its business context. Next, reviewing the data ensures that you can separate the information that must be produced from the data that is privileged or irrelevant and must be withheld. Finally, you must produce the data in a transferrable medium.

The costs can multiply quickly, as the metrics from a recent Gartner Research Report indicate:³

Task	Time
Identify appropriate data	\$200 / hour
Preserve the data	\$100 to \$300 / hour
Collect the data	\$200 to \$300 / hour
Review the data	\$120 to \$350 / hour
Produce the data	\$1000 to \$2,100 / GB

Based on the cost per hour or per GB of data, it is easy to calculate the expense of the discovery phase into the hundreds of thousands of dollars.

² “Fulbright Launches its Third Annual Litigation Trends Findings – U.S.,” Press Release, October 10, 2006.

³ Debra Logan, “Mapping Technology and Vendors to the Electronic Discovery Reference Model,” Gartner Research, ID Number: G00153110, November 9, 2007.

All things considered, even if your organization has never faced an audit or e-discovery request, it simply makes good business sense to be prepared. First and foremost are the Federal Rules of Civil Procedure (FRCP), which went into effect on December 1, 2006. The new rules govern lawsuits filed in Federal Court and outline the rights and obligations of all organizations, public or private.

The FRCP rules hold companies accountable for taking "reasonable steps" to implement processes and technology to safeguard data. These rules recommend that companies implement procedures for capturing and preserving information that relates directly to an actual or pending lawsuit. However, the caveat is that keeping more data than is required makes that data available for e-discovery.⁴ In addition, current or pending litigation warrants a "litigation hold." This means that companies are required to monitor procedures to ensure that relevant data is not altered or destroyed.

Effective data retention policies are critical for compliance. The good news is that the technology is available to enforce effective retention policies and support e-discovery initiatives, but there are still many challenges.

What data management challenges impact e-discovery?

A recent Gartner Research Report notes that, "Organizations are experiencing an increase in the complex demands for the management and retention of historical information, driven by evolving compliance and discovery requirements. The volume of electronically stored content continues to increase dramatically, along with the diversity of content types that need to be managed . . . Organizations will need to intertwine their content governance and retention policies with information archiving technologies to meet their commitments effectively."⁵

The report recognizes that "Few organizations have archiving strategies for both their structured files and unstructured content that exists in transaction-oriented applications. Rather than defining a content lifecycle management plan that would include archiving, the active data store is typically allowed to grow, putting a strain on the ability of IT to support business continuity, as well as backup and recovery for these critical applications." Gartner goes on to note, "Structured data that is typically stored in databases will require special database archiving solutions."⁶

Why? Because the ERP, CRM and custom applications that drive your business operations collect large volumes of data that must be managed, maintained and secured. Application data or "structured data" poses additional challenges for e-discovery. This is particularly true when the data is stored in one or more complex relational databases.

For example, if data is referenced from two different databases, it must be archived together, while maintaining the data relationships. A capable database archiving solution must manage complex relational data and keep that data referentially intact. When the data model is straightforward, creating referentially intact subsets of data is relatively simple. However, the databases powering most of today's enterprise applications rely on highly complex data models.

To support e-discovery requirements, the ideal database archiving solution must capture data at the business object level of detail and preserve the referential integrity of the data. 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.

⁴ Alan Brill and Michele C.S. Lange, "The New E-discovery Rules: What You Don't Know Can Definitely Hurt You and Your Company," Computerworld – Storage, December 11, 2006.

⁵ Kenneth Chin and Carolyn DiCenzo, "Key Issues for Information Archiving and Retention," 2007, Gartner Research, ID Number: G00146211, March 9, 2007, p. 2.

⁶ Ibid. p. 3.

How does enterprise data management support e-discovery initiatives?

An effective enterprise data management strategy that includes comprehensive database archiving capabilities allows you to manage application data throughout its lifecycle. By archiving historical data, based on its business value, and implementing tiered storage strategies, you can optimize storage utilization and defer the cost of high-speed resources, while providing authorized data access for e-discovery and other reporting requirements. Archived data remains accessible no matter where it is stored. Some of the advantages are explained in the following paragraphs.

Managing data growth

As a recognized best practice, routine database archiving keeps application databases at a manageable size, reducing capacity requirements and lowering operational costs. Capabilities for segregating historical from current data enable companies to manage continued growth, while keeping the data accessible for reporting purposes, audits or e-discovery requests.

Accessing and retrieving data

Typically, e-discovery requests have tight deadlines. Archive indexing improves search accuracy, which is critical for locating needed data. Additional audit capabilities for monitoring who has access and when archived data was last retrieved provide additional levels of control. Enterprise data management options permit authorized access to archived data. In addition to native access and application independent access, an open data management access facility ensures fast data retrieval, while mitigating risk.

Protecting data privacy

Corporations must protect all sensitive information, whether it resides in a production system, a development database or anywhere else within organizational boundaries. Companies need the capability to block out or mask sensitive data that may not be required as part of an e-discovery request. This process is called "redaction". By implementing capabilities that allow you to mask or transform sensitive information, you have the data management strategies you need for redaction, while preserving the integrity of your data.

Defining data retention policies

According to Forrester Research, "There is a compelling case today for retention management — lower eDiscovery costs and more efficient compliance processes. Organizations that do not put initiatives in place now not only will lose out on today's business benefits but also will lose the opportunity to manage information better overall."⁷

First you must determine which applications manage data that would most likely be required for e-discovery and group them into categories for archiving based on your business requirements. Next, classifying the data makes it easy to define and document business rules and data retention policies to govern active, inactive and compliance-managed data. You can prevent information assets from becoming information liabilities by deleting historical records after they are no longer required for compliance or business purposes.

Differentiating between backups and archives

Consistent backup processing ensures that in the event of a disaster or inadvertent data loss, operational data can be recovered quickly from the last backup point. Successful recovery depends on how often and how well backup procedures are performed. However, because backups are often overwritten, this procedure is not designed to preserve data for the long term and would not support your e-discovery requirements.

In contrast, archives are designed for long-term data retention. Ongoing database archiving allows you to manage and maintain historical data and associated metadata for reporting and compliance purposes. Your archived data remains accessible on demand to support your e-discovery efforts. As an additional benefit, implementing routine database archiving can significantly reduce backup and recovery times. Keeping databases at a manageable size allows you to maintain acceptable levels of performance and availability.

⁷ Barry Murphy, "Abysmal: The State Of Retention Management, Forrester Research, July 17, 2007, p. 10.

Storing data throughout its lifecycle

Companies need a variety of cost-effective storage alternatives and capabilities for managing data in an immutable format for e-discovery. Tiered storage strategies enable managing data based on its business value and access requirements, until retention periods expire. For instance, it makes sense to manage active data in a high-performance storage environment for rapid transaction throughput. Reporting data can be relocated to mid-tier storage, so you control costs, while still meeting service requirements.

Finally, reference data can be maintained in an immutable format on a secure WORM (Write Once, Read Many) device, where defined policies govern its access and ultimate disposal. The flexibility of enterprise data management allows IT organizations to leverage the best price/performance platform for storing enterprise data throughout its lifecycle.

Retiring applications

Periodically reviewing your application portfolio helps you balance the cost of operating and maintaining your applications against the business value they provide. Your objectives may be to simplify your IT infrastructure by eliminating redundancy, migrating applications to more cost-effective platforms or reducing dependence on systems that are no longer supported by a vendor.

An enterprise data management strategy that includes database archiving ensures access to application data, even after the originating application is taken out of service. Sites can satisfy an e-discovery request by using industry standard methods to query and report on archived business records.

What should you look for in an Enterprise Data Management Solution?

The ideal enterprise data management solution should provide a consistent technology and methodology that can be scaled across diverse applications, databases, operating systems and hardware platforms. Find a solution that provides proven capabilities for addressing the critical IT issues that accompany e-discovery and audit requests.

Archive and retrieve complete business objects and metadata

Focus on archiving capabilities that ensure the integrity of the data in its complete business context. The ideal enterprise data management solution must capture complete business objects and the associated metadata. You should also have capabilities that allow you to selectively retrieve archived transaction records without having to restore, and capabilities for restoring archived data to production or to an alternate target environment.

Integrated archiving and indexing capabilities can promote faster searches across archives. The ideal solution should allow you to establish one or more indexes for each archive to reduce the time it takes to retrieve the precise data you need.

Support a variety of storage alternatives

Determine whether to store archives in a database, in a file or perhaps in different formats and storage media 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*.

Ensure long-term data access options

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 and report writers. 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.

Satisfy data retention and litigation hold requirements

Next, focus on capabilities for 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. Look for a solution that offers a litigation hold feature. You can keep business records accessible until legal retention periods expire and archives can be deleted either manually or by means of automated disposal of expired transaction records.

Conclusion

Preparing for e-discovery simply makes good business sense. Enterprises must recognize that there is business value in organizing their information and enterprise application data. Managing enterprise application data for e-discovery poses many IT challenges. To comply with FRCP rules, companies must take some “reasonable steps” to implement processes and technology to support e-discovery requirements.

Implementing a proven enterprise data management solution, such as IBM® Optim™, provides comprehensive data management and archiving capabilities that enable organizations to manage, store, access and retrieve application data to support e-discovery requirements. As a recognized best practice, database archiving also offers other benefits in the form of improved application performance and availability, while reducing risk and lowering operating costs.

###



© Copyright IBM Corporation 2008. All rights reserved. IBM, the IBM logo, and Optim are trademarks or registered trademarks of the IBM Corporation in the United States, other countries or both. All other company or product names are trademarks or registered trademarks of their respective owners.