

## The eDiscoveryJournal Report:

# The Importance of Storage in eDiscovery

January, 2011

By:

Barry Murphy

## CONTENTS

<b>Executive Summary .....</b>	<b>3</b>
<b>E-discovery matters to organizations today .....</b>	<b>4</b>
<b>Storage – Not Just Where Data Lives .....</b>	<b>5</b>
<b>Processing Performance And Scalability .....</b>	<b>6</b>
<b>Information Governance Optimization.....</b>	<b>7</b>
<b>Supporting Chain-of-Custody Management .....</b>	<b>8</b>
<b>What You Should Consider When Evaluating Storage For E-discovery.....</b>	<b>9</b>
<b>Conclusion.....</b>	<b>11</b>

### Disclaimer:

eDiscoveryJournal is not a law firm. All expressed opinions and content are provided for general educational purposes only and are not specific legal advice, even if the author is a practicing attorney. Neither eDiscoveryJournal nor the information contained herein should be used as a substitute for competent legal advice from a licensed professional attorney in your state.

eDiscoveryJournal believes reasonable efforts have been made to ensure the accuracy of all eDiscoveryJournal original content. Content may include inaccuracies or typographical errors and may be changed or updated without notice. All eDiscoveryJournal original content is provided “AS IS” and while we endeavor to keep the information up to date and correct, we make no representations or warranties of any kind, express or implied, about the fitness for a particular purpose, completeness, accuracy, reliability, suitability, or availability with respect to the information, products, services, or related graphics for any specific purpose. Any reliance you place on such information is therefore strictly at your own risk.

In no event will eDiscoveryJournal or any of its contributors be liable for any direct, indirect, punitive, incidental, special, or consequential damages or damages for loss of profits, revenue, data, down time, or use, arising out of or in any way connected with the use of the document or performance of any services, whether based on contract, tort, negligence, strict liability or otherwise.

## **EXECUTIVE SUMMARY**

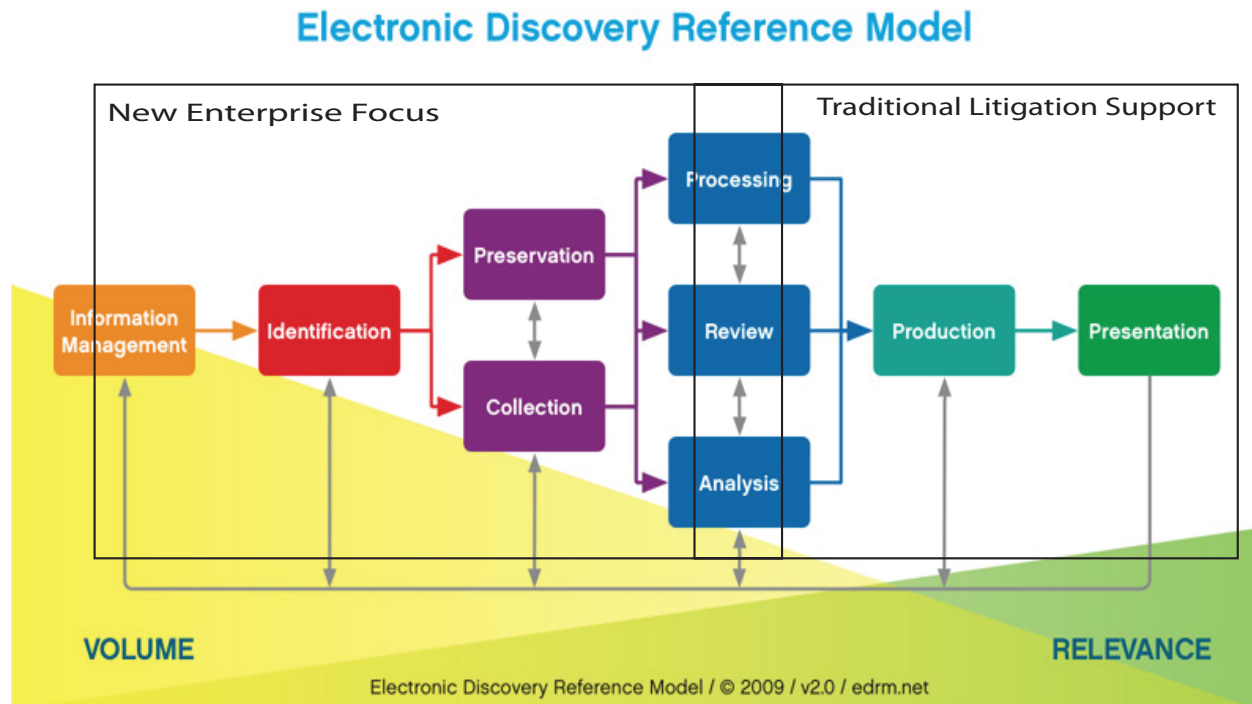
While you may not know it, storage plays a huge role in all aspects of eDiscovery. It's easy to take storage for granted and think of it as the commodity on which data resides, but that would drastically underestimate the importance of storage to optimizing eDiscovery programs. Storage plays a critical role in ensuring fast and accurate processing of potentially huge volumes of electronically stored information (ESI), building a scalable eDiscovery infrastructure that supports diverse file types and sizes, and providing a foundation for managing the chain-of-custody of ESI within the eDiscovery process. This document explores the role of storage in eDiscovery and gives IT, legal, and litigation support professionals alike the critical factors to consider when working storage requirements into short- and long-term eDiscovery projects.

## **E-DISCOVERY MATTERS TO ORGANIZATIONS TODAY**

As corporations develop litigation response and information governance infrastructures, there is a desire to convert what has traditionally been a reactive, outsourced process to one that can be supported by a proactive eDiscovery infrastructure. Rising eDiscovery costs and sanctions for non-compliance with federal rules force all organizations to take cost containment and risk mitigation measures. For corporations, this means having the ability to collect massive amounts of information and gain insight into it quickly in order to make informed decisions about how to handle any given matter. The clear trend is for corporations to build internal capabilities for the early phases of the EDRM model. As the Figure 1 below shows, many corporations extend from the left side of the model out to the right, creating the capability to conduct early case assessment (ECA) in order to make more informed legal decisions efficiently. Law firms and litigation support service providers continue to offer more valuable solutions like advanced analytics and build out scalable data centers to support corporations downstream in the eDiscovery process.

Figure 1 - New Trends in eDiscovery

- Early Review of Potentially Relevant Records
- Some implementations inside the firewall - "ECA in-place"
- Potentially substantial savings in time and money for enterprise and law firm



Recent data show that only about half of US corporations have repeatable eDiscovery processes and programs. As they build out eDiscovery capabilities, it's very important to understand the critical role that storage plays as a pillar of eDiscovery infrastructure.

## STORAGE – NOT JUST WHERE DATA LIVES

Storage infrastructure is one of the pillars of eDiscovery, whether it is conducted in the data centers of corporations, law firms, or EDD service providers. Specific applications such as identification, collection, and preservation software, early case assessment (ECA), and review management applications are important and offer feature and functions to make eDiscovery more efficient and defensible. Such applications, though, depend on the right storage for peak performance. And, while applications are typically mixed and matched as needed, storage

infrastructure cannot be swapped out without considerable effort and expense. Thus, it is extremely important to choose the right storage foundation to support your eDiscovery activities and applications. Some applications are very demanding of storage infrastructure; if the storage foundation is inadequate, the whole eDiscovery process can be hobbled. It's easy to think of storage as simply the media where digital information sits. But, storage is so much more than that – it's the power behind an optimized eDiscovery infrastructure.

## **Processing Performance And Scalability**

The performance and scalability of eDiscovery applications – important to corporate legal, law firm, IT, and litigation support professionals alike – is highly dependent on the underlying storage platform. If the storage system can't keep up with the demands imposed on it by a high-volume processing, both the speed and the quality of the final output will suffer. Data throughput performance will be stressed, for example, in enterprise-scale operations like eDiscovery that utilize tens or hundreds of clients hosted on modern multi-threaded, multi-core CPUs. Whereas most systems require a relatively steady state of storage and computing power, an eDiscovery infrastructure must support fast, unexpected bursts of computing needs to churn through massive collections of information. And, with eDiscovery, there is no way to forecast when these resources will be necessary. Instead, it is prudent to be sure the eDiscovery infrastructure is built for high-performance from the ground up.

ESI collections often include a number of large (~300GB or greater) source files, such as logical evidence files (LEF) accumulated from disk imaging. Converting these files to smaller target files requires a storage platform with capacity for high IOPS (input/output per second), as well as the capacity to perform many multiple concurrent read/write operations. Processing the millions of small files that are extracted from containers like LEFs creates a storage IOPS challenge since the ratio of the file protocol handling to the size of the data is so significant.

When application vendors tout high processing performance numbers, a major part of the equation depends upon the storage used in the testing. It's important to realize that you might not get the same performance if you just use existing storage that you have in your data center.

Scalability is also highly dependent on storage. Organizations that conduct collection, preservation, analysis, and review on a case-by-case basis and tend to have very small matters might be able to get away with using an application built for limited scale. However, for the enterprises, law firms, and service providers conducting large-scale eDiscovery operations, scalability is a must. With eDiscovery, storage requirements for a given matter are not known until the identification and collection phases are complete; at that point, there is no time for a typical information technology infrastructure purchase cycle. The right infrastructure – with the ability to scale and process data at the right speeds – is a critical upfront investment if there is going to be a chance to optimize both the effectiveness and efficiency of eDiscovery.

## Information Governance Optimization

In addition to being a critical factor in maintaining high performance for eDiscovery operations, storage is also an important component of a solid information governance infrastructure. Such an infrastructure truly matters for organizations looking to better manage information before litigation strikes. The right storage platform ensures minimal interruption to the IT environment by:

- *Completing backups on time.* Backing up information is a vital data protection activity and especially important when dealing with preservation repositories. A storage environment that can't complete backups in a timely manner will compromise eDiscovery operations and ultimately result in higher costs. In addition to fast backups, a good storage platform will support fast application restores so that eDiscovery solutions maintain good uptime.
- *Allowing for efficient management of storage capacity.* High performance and scalability are important for eDiscovery operations, but that does not mean that total cost of ownership (TCO) is not a consideration. A good storage platform can be high-performance while still allowing for cost management via tiered storage capabilities and low resource utilization.

- *Allowing for storage provisioning.* Storage provisioning is the process of assigning storage space in order to optimize performance of the platform. One reason we've seen clients turn to network attached storage (NAS) platforms versus storage area networks (SAN) for eDiscovery infrastructures is the ability to more readily virtualize storage, simplifying data management and making storage provisioning much easier.
- *Supporting deployment of effective multi-pathing solutions.* Multi-pathing is the ability of a system to use more than one read/write path to a storage device. It is critical to high availability solutions that provides fault tolerance against single point of failure in hardware components. Multi-pathing provides load balancing of I/O traffic, thereby improving system and application performance. In order for eDiscovery operations to function at the level necessary in today's environment, the storage platform must support multi-pathing.

When considering the storage platform for eDiscovery operations, be sure to factor in ease-of-use in terms of data center management. eDiscovery requires high-performance in the data center while maintaining reasonable costs. One client we spoke with considered implementing a storage area network (SAN) as its storage platform, but found that SANs require specialized, dedicated resources in order to manage. Those kinds of costs can be prohibitive. This client found Network attached storage (NAS) solutions easier to manage because they behave just like file systems. The client's existing IT resources could manage the NAS solution – meaning no incremental management costs. Deploying the right NAS solution allowed this client the performance necessary to support eDiscovery operations at a low total cost of ownership (TCO).

## Supporting Chain-of-Custody Management

In addition to providing much of the necessary horsepower for processing digital information, storage software is critical in maintaining defensible collection and preservation. While eDiscovery applications run mechanisms such as hashing and provide reports to prove that ESI has not been altered during eDiscovery operations, it is the storage software that must support combining collection of data sets from different platforms and support the migration of data from those platforms to a specific matter set. For example, some storage devices have file allocation

tables that are FAT 16 whereas others are FAT 32. Your storage infrastructure must support collecting data when there are potential differences in file system fields/properties from the various sources of data in a way that allows you to show the defensibility of the collection, preservation, and management of the data. eDiscovery applications are ultimately responsible for managing chain-of-custody, but it is important to be aware of the differences in the storage platforms of various ESI sources and use the collection software to track operating system (OS) fields before they are altered. What is important is that the actual storage be homogeneous, so that there are no secondary/tertiary changes after the initial collection. Storage administrators need to be aware of potential issues and provide expertise to support the legal team in proving that data is not altered.

## WHAT YOU SHOULD CONSIDER WHEN EVALUATING STORAGE FOR E-DISCOVERY

Clearly, the storage used in your eDiscovery operations can be the difference between moderate efficiency gains and drastic cost reduction and risk mitigation. Many corporations have existing storage systems in place and assume those systems can handle the load that eDiscovery operations will demand. While in some cases that may be true, it is a better practice to have a dedicated storage platform to support eDiscovery. Not every organization will need the same eDiscovery storage platform, but there are some best practices that all should consider:

- **Understand the amount of eDiscovery operations that will take place in your data center.** Storage requirements depend upon how much of eDiscovery will be managed in your data center. For corporations, service providers, and law firms that will be processing large amounts of data, storage is a critical element for success. When it comes to storage, there are many options. Some will opt for storage area networks (SAN) which can require more resources to manage, while others will want to use network attached storage (NAS), which can be administered and managed with considerably less effort. Don't just think about near-term eDiscovery; rather forecast the long-range activity and capacity needs in order to determine performance requirements.

- **Ensure centralized, historical tracking of ownership and access rights for shares.**  
Data will be sitting on multiple file shares – trying to manage the ownership and access rights for each share is hard to do in a distributed manner. It's important to have that centralized management capability and centralized reporting of any changes to the shares.
- **Create a documented change management process that keeps legal in the loop.**  
eDiscovery is a process that needs to be managed. Your storage infrastructure must support that process. As business is dynamic and doesn't stop while you evaluate a storage platform, be sure to document all changes and inform your legal department of any potential impact that changes may have. Legal should have some say in decisions – or at least validate that any changes won't be show-stoppers for eDiscovery.
- **Make sure any information on legal hold can remain on legal hold with full metadata and context intact on the new storage.** One of the worst things that can happen in eDiscovery is spoliation, where data is altered somewhere in the process of collection, preservation, review, or production. It is important to recognize the differences between storage platforms of collected data sets and ensure that your eDiscovery storage platform be able to accommodate all the diverse collections.
- **Plan for end-of-life, both of data and storage.** How does data get defensibly disposed of? Does legal have a sign off? How does data get migrated from EOL storage? How do you actually destroy drives so that the data is not recoverable or sold to a recycling company with confidential information?
- **Make chain-of-custody management an integral part of your storage strategy.**  
Understand the potential differences in file system fields/properties before migrating content under preservation. While the eDiscovery applications you use will provide reports and methodologies such as hashing to manage chain-of-custody, it is important that your eDiscovery storage platform be homogeneous, so that there are no secondary/tertiary changes after the initial collection.

## CONCLUSION

eDiscovery presents challenges to every single organization, large or small, public or private. In the last several years, solutions have emerged that allow organizations to cut costs, mitigate risks, and better manage eDiscovery as a process. In order to harness the power of these solutions, though, it's important to plan for the long-term and deploy those solutions on the right storage infrastructure. Without the right storage foundation, many eDiscovery applications will struggle to process data quickly enough to meet the tight timeframes required by the FRCPs and to scale to handle the amount of data that organizations create and store today. Poor planning will lead to applications that fail and only exacerbate the problem. The chart below summarizes the important factors to consider when planning a storage infrastructure to support eDiscovery.

Consideration	Explanation
eDiscovery operations in both the near- and long-term	Storage requirements depend upon how much of eDiscovery will be managed in your data center. For corporations, service providers, and law firms that will be processing large amounts of data, storage is a critical element for success. When it comes to storage, there are many options. Some will opt for storage area networks (SAN) which can require more resources to manage, while others will want to use network attached storage (NAS), which can be administered and managed with considerably less effort. Don't just think about near-term eDiscovery; rather forecast the long-range activity and capacity needs in order to determine performance requirements. How does data get defensibly disposed of? Does legal have a sign off? How does data get migrated from EOL storage? How do you actually destroy drives so that the data is not recoverable or sold to a recycling company with confidential information?
Security and access rights	Data will be sitting on multiple file shares – trying to manage the ownership and access rights for each share is hard to do in a distributed manner. It's important to have that centralized management capability and centralized reporting of any changes to the shares.

Consideration	Explanation
Legal Sign-off	eDiscovery is a process that needs to be managed. Your storage infrastructure must support that process. As business is dynamic and doesn't stop while you evaluate a storage platform, be sure to document all changes and inform your legal department of any potential impact that changes may have. Legal should have some say in decisions – or at least validate that any changes won't be show-stoppers for eDiscovery.
Minimize potential for spoliation	One of the worst things that can happen in eDiscovery is spoliation, where data is altered somewhere in the process of collection, preservation, review, or production. It is important to recognize the differences between storage platforms of collected data sets and ensure that your eDiscovery storage platform be able to accommodate all the diverse collections. Understand the potential differences in file system fields/properties before migrating content under preservation. While the eDiscovery applications you use will provide reports and methodologies such as hashing to manage chain-of-custody, it is important that your eDiscovery storage platform be homogeneous, so that there are no secondary/tertiary changes after the initial collection.