Kazeon™

Left unmanaged, information can put your company at risk

Ten questions to ask
as you assess your risk

What's on your network?

Which files are sensitive?
What's regulated?
What's not being used?

What should be saved?
For how long?
What should be deleted?
When?

Who can see what?
What did they do?

Most of your company's information is unstructured

In the last 10 years, the information stored on corporate networks has changed—a lot. Networks now store radically more information (hundreds of terabytes for some large businesses) in hundreds of file types (from text to video, from spreadsheets to MP3s). The information is growing at 50 to 75% per year and is increasingly distributed. In fact, over 80% of what's on the network is stored outside databases—it's unstructured.

All this unstructured information is increasingly important to business. It's a primary tool for communications and an important resource for making decisions. It's also increasingly regulated (e.g., SOX, HIPAA, SEC 17a-4).

Meanwhile, storage infrastructure has not kept up with changes in content. Sure, networks are bigger and faster—but not smarter.

95% of all unstructured information is unmanaged

The problem is: Unstructured information is essentially invisible. The sheer volume of information hampers finding the things you need. The hundreds of file formats obscure your content. In some cases, you can't even tell what's on your network.

And if you can't see your information, you can't manage it.

Unmanaged information is unmanaged risk

The risk of not managing your company's unstructured information can be huge. Bad decisions. Unknown decisions. Untimely responses. Privacy and security breaches. Failure to comply with policies and regulations.

The consequences can be serious and far-reaching. Business losses. Expensive lawsuits. Regulatory fines. Even criminal prosecution.

Kazeon™ has solutions. We give you visibility and control over your unstructured information—so you can reduce your risk.

Information visibility and control means reduced risk

Managing unstructured information

In order to keep pace with the changes in corporate information—and to meet the needs of corporate executives, compliance officers, and lawyers—today's storage administrators need tools that provide visibility and control over unstructured information.

That's why Kazeon is introducing a new generation of systems for managing unstructured information. They bring together advances in search, database, and storage management technology to create a new class of content-aware solutions. They provide a common view (regardless of original format or file type) and consistent control over information.

Kazeon's solutions use both file content and metadata to create an abstract for each of your unstructured files. These file abstracts create an application independent view of your unstructured information. They provide a basis for consistently classifying documents and applying governance policies. They are also the basis for advanced searching and reporting functions that will help you find information and audit your storage infrastructure.

That's the essence of Unstructured Information Management—classifying, managing, and retrieving documents. Unstructured Information Management enables a number of new applications: Storage Search and Classification; Compliance Archiving; Information Lifecycle Management.

Each of these new applications increases information visibility and control—and reduces risk.

With Kazeon, you will understand the business value of your files, make sure they are managed correctly, and be able to find what you need, when you need it. What's more: You will also increase service levels, lower costs, and maximize the use of your storage and administrative resources.

Introducing the Kazeon Information Server

A complete solution in a scalable appliance

The Kazeon Information Server is the first complete solution for unstructured information management. It's well-integrated, scalable, flexible, and efficient.

The Kazeon Information Server is remarkably easy to deploy. It comes in a plug-and-play appliance package. It integrates seamlessly with the surrounding environment. And Kazeon also supports standard file system interfaces, ties into existing storage and security systems, and provides powerful management capabilities—without changing user behavior or inserting itself into the production environment.

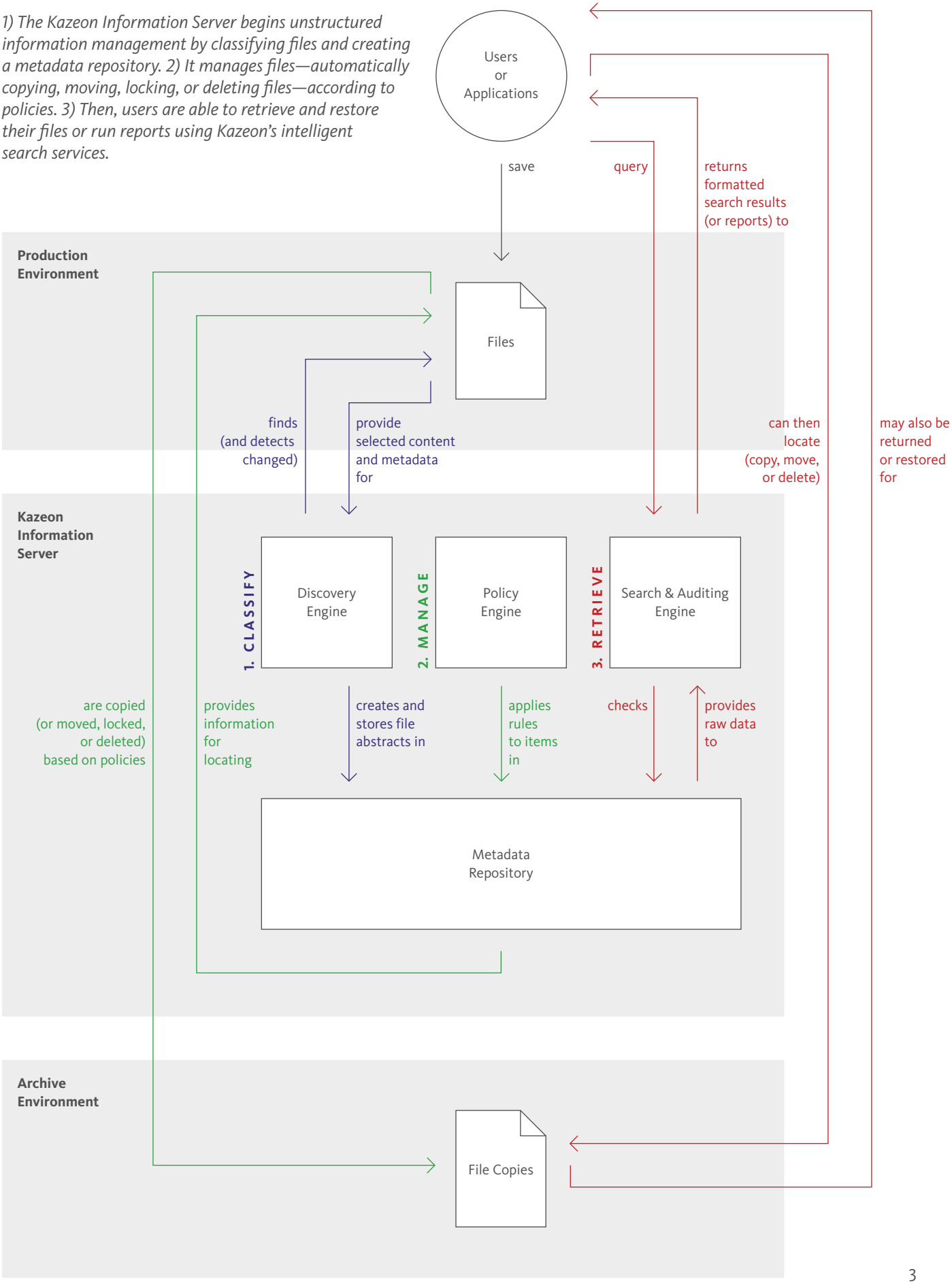
We designed the Kazeon Information Server to scale. A single Kazeon Information Server can manage millions of files. You can add servers to increase performance. Several servers can be clustered together into a single system capable of managing more than 100 million files. Multiple Kazeon systems deployed across the enterprise can provide integrated access to billions of files.

The Kazeon Information Server offers a flexible interface—three in all. For end-users, it has an easy-to-use graphical user interface (GUI). For advanced users, it also has a powerful command-line interface (CLI). For integration with applications, it has a well-documented application programming interface (API).

The Kazeon Information Server provides powerful classification, management, and search features without straining your resources. In most cases, it requires less than 20% storage overhead. This cost-effective approach to managing unstructured information puts Kazeon in a class by itself.

Logical view of the Kazeon Information Server

1) The Kazeon Information Server begins unstructured information management by classifying files and creating a metadata repository. 2) It manages files—automatically copying, moving, locking, or deleting files—according to policies. 3) Then, users are able to retrieve and restore their files or run reports using Kazeon's intelligent search services.



Three steps to information management
1. Classify 2. Manage

Creating visibility

Kazeon creates structure from unstructured files in a two-step process. First, Kazeon searches the network for every file and then creates a unique fingerprint for each file. This fingerprint is generated from the file's unique content and can be used to identify duplicate files or detect corruption of a file.

Kazeon can open files and extract content, such as dates, telephone numbers, or customer IDs. By analyzing file content, Kazeon uncovers hidden structure in unstructured information.

In addition to file content, Kazeon also collects metadata from the file system, such as size, name, owner, path, and format. Kazeon recognizes hundreds of standard file formats, including PDF, media file, and Microsoft Office. File formats are identified using file content, not file extensions. So, for example, MP3 files with a .dat extension would still be recognized as MP3 files.

The file fingerprint, file content, and file metadata are gathered together into a single abstract that is stored in a metadata repository. Additional data or tags can be added to a file abstract through the Kazeon user interface or the API. For example, when a customer account closes, files with that customer account number could be marked as closed. Files can also be automatically tagged using predefined rules. For example, PDF files containing the text "datasheet" can be categorized with the word "Marketing."

Kazeon's reporting services surface information from the metadata repository, offering insights into enterprise file assets and storage use throughout a network. Reports can be run based on any metadata or content field collected during file classification. These reports uncover duplicated files, show a breakdown of files by file type, or detail file use by users or departments, supporting effective planning and management. Information from the metadata repository provides the building blocks for enterprise governance policies.

Automating policy-based control

File management policies are created using the policy engine, which defines rules and then ties them to file abstracts. Policies can be set for almost any action, such as moving a file to a specific storage device based on age or file content. For example, a policy can be set to move files tagged as "Customer" to a WORM device and retain them for 3 years. Or, marketing documents that haven't been accessed in 90 days can be automatically moved to low-cost archive storage.

Kazeon's management services handle the copying and movement of files from production storage into an archive. Files are archived in two ways. First, predefined policies can automatically move and copy files. Second, files can be selected for movement and copying from search results or reports.

Kazeon manages and controls where the file is moved when it is archived. Files can be moved to a specific type of storage, such as WORM, enabling consistent and effective use of storage devices and resources.

Auditability is built into the Kazeon environment. Once a file is archived, Kazeon tracks it with an audit trail, saving information on the file's movement over the network or within an archive. Along with the original information about a file, such as its owner, file name, and file path, Kazeon tracks file access patterns. For selected files, it is possible to track when a file was created, how often it was accessed, and when it was modified. In some cases, Kazeon can even track who performed these actions.

3. Retrieve

Finding what you need when you need it

Most data management solutions focus on getting data onto disk. They provide little help in making sure you can find and retrieve it. An effective unstructured information management system can get data out as quickly and easily as it is put in. Robust and flexible search is critical for any system managing very large numbers of unstructured files.

At Kazeon, we have focused our efforts on applying search on a very large scale and integrating it throughout the system. Our engineers have collectively logged several man-years on the problem of search. From this work, we have built a robust and flexible tool for searching very large document archives.

Kazeon's powerful search and indexing technology is the basis for many of the system's services, from data classification to content-driven polices and information retrieval. It is also tightly integrated with the policy engine so you control what users see in search results. By default, the Kazeon Information Server will only show users search results for files for which they have access privileges. And, you can further control results by defining policies beyond the file system controls. For instance, you can grant a compliance auditor rights to see search results for all documents managed by Kazeon without granting full read permissions on your file servers. You could also use policies to prevent end users from seeing search results that contain social security numbers.

The search and auditing engine is not just a passive information resource. It is designed to be an active management tool. Any result from a search or auditing query can be managed directly from the Kazeon interface. Files can be tagged with additional metadata, moved to another location, or even deleted directly from the search and auditing results screen.

Optimizing search to solve business problems

Kazeon focuses search technology on solving broad unstructured information management problems in the enterprise. We refer to this new type of search as storage search. Storage search mirrors the information diversity of enterprise search and the scalability of web search products. But, it more deeply integrates with existing enterprise infrastructure and is tailored for information management tasks such as data classification and information discovery. By focusing on these areas, Kazeon has transformed search into an efficient and powerful data management tool.

Kazeon designed its search technology specifically for crawling and indexing large shared file systems and enterprise data repositories. Kazeon optimizes the crawling process and includes a broad range of metadata within the search index. Kazeon recognizes hundreds of standard file types such as Word, PowerPoint, Excel, and PDF. It can also be easily tailored to extract the unique information types created by every enterprise, for example, invoices, trade records, or contracts. The result is a fast, efficient indexing process tailored to the unique needs of each enterprise.

Traditional search technologies focus on finding "the best" result for a given query. This approach can create an index that is twice the size of the original data set. In contrast, storage search focuses on finding a complete result set for queries such as "find all files created between these two dates" or "find every document that contains this customer account number." This focus enables Kazeon to optimize resources for maximum efficiency. As a result, our index is a fraction of the size of the original data set.

We provide a range of options for tuning required resources. For example, using the staged search option requires a permanent index that is less than 10% of the size of the original data set. This permanent index is used to create temporary full-text indices that can support deep searching of large archives. Files are easily searchable within normal discovery or audit windows—but with an order of magnitude reduction in storage resources.

Practical applications—Kazeon at work

Storage Search and Classification

Companies with large amounts of unstructured files need tools to understand what those files are and what information they contain. Kazeon provides visibility for all enterprise file assets, taking a full inventory of every file on servers and storage devices. Armed with a complete picture of a file system infrastructure, companies have better control over the use of storage media, as well as support for litigation discovery and regulation compliance.

Once connected to a network, Kazeon takes an inventory of all enterprise file assets. This data is saved in the metadata repository. Reports run on the metadata repository uncover duplicate data, policy violations, and storage use by individuals and departments. Kazeon assigns a unique fingerprint to every file based on each file's unique content. The system identifies duplications by finding files that have that same fingerprint. Kazeon also identifies file type based on file content, not file extensions. If users are violating policies by saving MP3 files using different extensions, Kazeon will still find the files.

In addition to enabling effective file management, Kazeon also supports a full range of search options for finding files on a network. Unlike traditional intranet search solutions, Kazeon's search technology is fine-tuned to work with very large amounts of files. In addition, the search feature uses a familiar user interface making it possible to offer search to individual users. For example, lawyers can run their own discovery searches or compliance officers can keep track of file and file content use.

Compliance Archiving

Corporations need solutions to comply with the requirements of today's complex regulatory environment. Regulated data has diverse retention requirements which affect how files are stored and when they should be purged. For example, organizations must be able to prove that certain types of information, such as broker communications or patient records, have not been tampered with and are retained for a specific time period. For financial records, they are required to capture every saved version of a file on read-only media. If a company does undergo an investigation, audit, or litigation, it must be able to quickly retrieve and produce relevant information.

Kazeon supports the full range of archiving services required to manage corporate governance policies, litigation, and regulatory compliance. A comprehensive file archiving solution can be deployed by pairing the Kazeon Information Server with a WORM storage device. A company's compliance policies can be configured using the Kazeon policy engine. These policies define which files should be moved to WORM storage and how long they should be saved there. For example, policies might state that Excel spreadsheets from the Finance department should be saved for 3 years and Word files from Legal should be stored for 10 years.

Kazeon finds files that require protection and regulation based on the configured policies, copying them from network storage or servers to a WORM device and creating a fully searchable metadata repository of all the files in the archive. Because each file has a unique fingerprint based on the file's unique content, Kazeon can immediately identify if a file's contents have been modified.

The serious consequences of violating government regulations makes it imperative for organizations to find a viable option to traditional methods of storing and archiving. The Kazeon solution enables consistent, auditable control of files and fast information retrieval in a cost-effective manner.

Information Lifecycle Management (ILM)

In many production environments, 50 to 80% of stored files are not in active use. With a Kazeon system, these inactive production files are moved onto a fully searchable and manageable archive. Decreasing the amount of data in the production environment saves money and improves disaster recovery service levels.

Kazeon builds file system archives using near-line storage subsystems in a tiered storage environment. Kazeon classifies files in a designated production environment and then moves specified files to archive media. Files are archived based on file content and file properties, which ensures that critical documents remain in the production environment. Files in the archive are fully searchable by content, not just file name or file path.

With a file system archive, organizations can quickly prune their production storage and significantly reduce backup costs while improving success rates. Most production environments are backed up weekly. Decreasing the number of files in the production environment reduces capacity needs for weekly backups. If 60% of your production data can be moved into an archive, you can reduce your backup requirements by half. And, you can lower the costs of capacity expansion by using lower-cost, near-line storage options.

Archive and backup: What's the difference?

File archiving is frequently equated with tape backup. In fact, archiving and backup have two distinct—and complementary—functions within an enterprise. Backup systems manage files for disaster recovery and business continuity; they are optimized for fast and reliable copying and restoration of files. Archiving systems manage files for retention, which requires specific knowledge about files and their content, as well as tools for quick and easy file retrieval.

With the advent of cheaper and faster disk-based storage, many backup systems are now more efficient, but they are still designed for long-term storage and disaster recovery. They do not provide the intelligent file management critical for unstructured information management.

Archiving systems manage both files and file content based on business value. Business value determines how long to keep a file, how to protect it, how to secure it, and what media type to use for storing it. To find a file, archiving systems provide services for searching by both file metadata and content.

Archiving and backup systems are both needed for effective storage management. In an optimized storage and file management system, active files are stored in the production environment and inactive files are archived and stored in near-line disk-based storage or WORM devices. Backup is optimized for both production and archive classes of storage. The end result is streamlined storage operations and higher service levels.

About Kazeon

Founded in 2003, Kazeon develops application-independent information management services that provide consistent information visibility and control across diverse unstructured data sets, such as user documents, application files, and email. The company's founders and executive team are experts in leveraging search, database, and storage technologies to improve the manageability of large, distributed information infrastructures. They have applied their experience and knowledge to create products that integrate seamlessly with existing infrastructure, scale to support billions of objects, and efficiently address the challenges of data classification, governance, and discovery.

Kazeon's first product is the Kazeon Information Server, an appliance for classifying, managing, and retrieving data contained within large data sets. The product is currently in beta test with select companies and is expected to be released in early 2005. Kazeon is based in Mountain View, California, and is privately funded by top-tier venture firms Redpoint Ventures, Clearstone Venture Partners, and Goldman Sachs.

Kazeon
1161 San Antonio Road
Mountain View, CA 94043

650 641 8100 main
650 641 8195 fax
www.kazeon.com

March 16, 2005