



# Content Storage for a New Millennium

By **Derek Gascon**, VP Marketing, Caringo  
March 2008

---

"[Storage] systems need to be massively scalable, provide predictable and reliable performance, and come at an affordable price point. With the potential to store billions of files and objects, they must also be easy to manage across multiple systems or geographies. In addition, they must interface with Web-based protocols and have replication capability."

Storage Magazine  
November 2007

## Introduction

The reality of today is that we live in a largely digital society across all aspects of our lives. Everything from documents and graphics to audio, video and images are being created, distributed and stored digitally. It is estimated that the amount of digital information created will grow from 281 Exabytes (281 billion GB) in 2007 to nearly 1,800 Exabytes in 2011<sup>1</sup>. This explosion in new digital content is stretching traditional file storage environments beyond their breaking points causing them to become increasingly complex, costly and brittle.

Storage as usual will not suffice in this new reality and customers are demanding a simpler, more affordable, scalable and reliable infrastructure for digital content.

CASTor™ is innovative storage software that enables customers to implement high-performance clustered storage at any scale and take advantage of inexpensive commodity servers. Based on content addressed storage architecture (CAS), CASTor creates a flat address space providing a single storage environment for billions of content objects (files) and up to Petabytes of capacity. The simplicity inherent in CASTor's cluster design further reduces overall operating costs with self-managing, self-healing and self-configuring features. The software-only design implementation allows greater choice for hardware and enables customers to seamlessly add capacity as needed and upgrade as more advanced systems are introduced.

## Content, Complexity and Cost

Businesses have been generating increasing amounts of data for decades, but just recently have encountered the seemingly uncontrollable production of unstructured data. Such data content or files includes office documents, graphics, images, video, engineering diagrams, geospatial maps and more. Not only is more digital content being created, regulations and internal governance policies are requiring file retention for longer periods of time. Existing solutions in the market, such as tiered storage, designed to reduce costs actually increase complexity and, over the long-term, increase operating costs as well.

Companies have implemented document, records and image management applications to maintain greater control and oversight for critical business information. It is not uncommon for one company to have several content management systems and repositories in their environment that complicates managing these storage repositories. According to a survey by Forrester Research 78% of companies have more than one content repository and 43% have six or more. These represent disparate silos of information that need to be controlled to address regulatory and corporate governance mandates. To effectively accomplish this organizations are looking at ways to simplify their infrastructure and consolidate content into a common storage environment.

### Number of content management systems

**78%** of companies have more than one content repository

**43%** of companies have six or more

Forrester Research

One of the largest costs associated with the complexity of storage and its management is related to the amount of administration and management overhead required to keep it in production. Reducing cost and complexity is the top IT priority for companies, which was cited by 59% of companies in a recent survey and second is the reduction in cost and time required for provisioning (43%)<sup>2</sup>. The combination of the volume of new digital content that needs to be managed and retained for years to decades with the complexity of today's storage solutions is a significant challenge for organizations.

It is estimated that 20% of the information generated by organizations is subject to compliance rules and standards<sup>3</sup>, which today would be in the neighborhood of 60 Exabytes (60 billion Gigabytes). Retention of content (business records) for a specified time along with other requirements are things file systems just cannot address. With the exception of setting a basic retention period, file systems do not provide the capability to validate the authenticity of content, manage lifecycle policies, or support custom metadata.

The concept of information lifecycle management (ILM) has risen in importance as a process to meet these rules and retain content for the long-term. The current model for ILM is fraught with added complexity as individual, point products are needed to categorize content, migrate it between tiers of storage and ensure its integrity for the long term. ILM is discussed as a means of simplifying the management of critical business information and that may be the case, however the infrastructure requires a complex set of independent technologies all working cohesively together that's just not yet the reality. In theory it is argued that storing data on the right tier of storage based on its value and access characteristics will save money although a lot of cost goes into building this type of environment.

### Top IT Priorities<sup>3</sup>

#### Reduce cost & complexity:

- File data consolidation
- File migration
- File virtualization

### Top IT Trends<sup>3</sup>

#### Long-term content retention:

- Information Lifecycle Management
- Management of Unstructured Data

Not only are organizations creating more digital content, but consumers now generate personal content in a volume that well surpasses it. Approximately 70% of it is consumer generated and 85% of that will be handled in some way by the commercial sector<sup>1</sup>. Many of the new Web 2.0 companies are offering consumers destinations for social networking along with capacity for storing personal content, e.g. images, video and blogs. Others are providing online content storage for consumers to backup their growing library of content as well as simply store and share photos, music and video. For these types of companies storage is a critical component of their infrastructure and they require more cost effective, scalable and reliable technology for serving active content in order to support their business models. Traditional file storage and even newer clustered file system products are unable to meet the demands of this new customer segment.

### **A New Storage Model Is Required**

Traditional file or content storage technologies are not meeting the requirements and pain companies have identified such as scalability, time and resource consuming administration of file systems and storage provisioning as well as expense. Some newer solutions offered to mitigate these issues treat the symptoms but don't necessarily address the root cause. With storage being such a major cost of the IT infrastructure in terms of both capital and human resources there is a strong business imperative to move to a new model. The economic and financial incentive to operations is evident and simplification will naturally translate into lower costs.

## **Behind the File Wall**

### **The File System**

File systems were invented some 30 years ago to provide an interface for end users and applications to store file data (non-database data) on block storage at a time when the volume of digital content stored today was unforeseen. For the most part file systems have not kept pace in addressing the capacity demanded. There are limitations on the number of files that can be stored, the size of files as well as the amount of capacity they can support. Today's file system imposes constraints when facing the need to store hundreds of millions to billions of files, scale from a few to hundreds of Terabytes to Petabytes and handle large multi-Gigabyte video and geospatial files.

Approximately 80% of data inside an organization is residing on file servers in formats as diverse as documents, spreadsheets, graphics, images, video, seismic maps, etc. and expectations are that file-based storage will increase by 50% to 120% a year<sup>4</sup>. As these file systems grow so too does the complexity in managing them. Maintaining file servers and file systems is a daily job for system administrators. They often need to migrate users and replicate data, which is so time consuming they often over-provision the file servers to forestall migrations as long as possible resulting in under utilizing storage. File serving has become mission critical yet companies are hitting a wall in terms of how efficient and effective they are versus what they need them to be.

### **Reality Check**

"It's clear that Web 2.0 features can benefit companies in vertical markets ranging from manufacturing to health care, retail to executive recruitment. Much of the current [storage] infrastructure wasn't designed to handle Web 2.0 application requirements at a price point that enables a company to deliver a profitable service."

"It's vital for IT leaders to understand and start evaluating their specific business requirements to determine the appropriate infrastructure. That will require deviating from traditional solutions and approaches. A shift in mindset, perhaps even in culture, may also be needed."<sup>5</sup>

Storage Magazine,  
November 2007

## **The NAS Appliance**

In order to reach beyond underlying file system limitations an appliance approach was taken to consolidate a small number of individual file servers through Network Attached Storage (NAS). Essentially, this is a network access node for applications and end users to a pool of shared storage. While it is one approach to consolidating file serving it doesn't fundamentally change the infrastructure for storage other than consolidating the hardware. If the NAS appliance runs out of resources in terms of network bandwidth or storage capacity, the only solution is to add a second appliance, which means segmenting and migrating users and data. This translates into more cost and more effort.

Administrators must still manage file systems, provision storage to ensure users and applications have sufficient space as well as back up the file systems. NAS appliances also don't scale beyond a single system, meaning that free space on one appliance cannot be used by another appliance. Such systems are delivered with software on proprietary hardware and are also relatively expensive.

## **The Clustered NAS Appliance**

Within the last several years clustered file systems or clustered NAS appliances were developed and introduced to mitigate some of the above constraints. These vendors deliver their own operating system (OS) and have proprietary file systems that often support their own version of CIFS and NFS. One major benefit of a clustered NAS is the ability to have shared write access to the data. There are some specific customer requirements where such a solution makes sense such as in video rendering. In order to have simultaneous write access to the same file a locking mechanism is required and the cluster's performance is heavily dependent on its granularity. For environments with small numbers of large files, e.g. video, this is not a performance problem. However, in environments that have a large number of small files or a mix (traditional file serving) write performance drops significantly because the cluster is so busy with locking on writes. This is the primary reason these solutions are largely targeted at the high-end of the market such as in oil and gas and broadcast companies that can afford the higher price.

## **The Global Name Space**

Technologies in this area provide another variation on distributed file systems by building a directory of physical file locations and cross referencing them to logical file share names for access by users and applications. The users or applications don't need to know where their file is located, but simply refer to the global name and are then routed to the file. These solutions don't solve the fundamental problem, but provide a practical mechanism to ease the customers' current pain. System administrators still need to manage and support all the underlying file systems, which remains time and resource intensive. Essentially it provides file server consolidation and relies on the global name space to direct requests to data that is scattered around other file systems on the network. These solutions are often delivered as an appliance, come at a fairly high cost and have not yet proven scalability at the enterprise level.

## **Archives Locked and Loaded**

Since 2002 government regulations have been introduced worldwide concerning the retention and preservation of business records in digital form often referred to as electronically stored information (ESI). These encompass everything from e-mail, instant messages, documents and spreadsheets to scanned images, medical images and call detail records representing a significant amount of storage. A new product category was introduced to provide protected archiving on disk and address regulatory compliance for retention, integrity and authenticity.

The majority of archive solutions are based on first or second-generation content addressed storage (CAS) architectures that meet basic archiving requirements.

Today all the major storage vendors have an archive storage offering in their portfolio that are delivered as an additional tier of storage. These are proprietary, packaged solutions that include software, servers and storage capacity in an appliance model. Current archive storage products have high switching costs associated with them that appeal to these vendors, but create lock-in for customers. Vendors describe this product attribute as “sticky.” Once the customers’ content is archived, it is very difficult to move it to a new hardware platform. Functionality is also being bolted onto these solutions to create greater “stickiness” such as index/search, categorization and content management that belong at the application layer, not storage.

## A New Millennium Requires a Fresh Approach

With file systems showing their age and limitations in handling the explosive growth of file data or digital content in this online era, companies are seeking a fresh, innovative approach to content storage. Caringo understands this and designed its CAStor content storage software to meet customer demands. In evaluating the current state of the CAS market it was recognized that there was much more customer value to be provided by a new, 3<sup>rd</sup> generation CAS product. Customers are being locked-into expensive, proprietary hardware platforms, are not getting the benefit of price declines in commodity server technology or the performance necessary for active content.

When Paul Carpentier, Caringo co-founder and CTO, and Jan Van Riel, Caringo VP of Advanced Technology, invented CAS in the 1990’s they were working on peer-to-peer networking concepts and recognized the need for a simple mechanism that allowed for the unique identification of a piece of content, regardless of where it was located. CAStor is the result of his innovations since then and represents a highly advanced storage infrastructure that delivers the performance needed for active content serving as well as the retention and preservation features of an archive. Unlike other CAS solutions available today, CAStor delivers an affordable content storage infrastructure with the performance, scalability and reliability needed across various customer segments.

### Elegance Delivered Through Simplicity

CAStor is software enabling the implementation of a clustered storage environment that effectively handles content of all types other than database data. The software will run on any x86 platform eliminating the lock-in to proprietary hardware products. This allows the customer to choose the commodity server hardware that best meets their budget, performance, power and price requirements. Installing CAStor is as easy as plugging in a USB thumb drive and booting a server with it. In the time it takes the server to boot you have created a CAStor content storage node with as much capacity as is internally available on that server.

The cluster’s storage capacity can continue to be expanded by adding new server nodes with the CAStor software. The new node(s) automatically join the cluster, the available storage is added to overall capacity and it begins to accept new I/O service requests. Adding new capacity requires no provisioning or configuration, which reduces administrative and management overhead unlike what is needed for traditional file systems. This can all be done without having to take the

#### Customer Advantage

**Simplicity.** Simple installation and low-cost commodity server hardware combined with lower administrative overhead translates to a total cost of ownership without rival.

system out of service so there is no interruption to applications or availability of content. Because of CAStor's cluster architecture it is self-managing, self-balancing and self-healing providing additional reductions in administrative resources.

### Tipping the Scale

Given the projections estimated for the demand for content storage over the next several years and beyond, CAStor's ability to scale in capacity, numbers of objects and performance makes it the ideal solution. A CAStor cluster can start with as little as 1 Terabyte (TB) of capacity and scale to 100s and even 1000s of Terabytes of storage in a single cluster. It provides the ability to start small as the business demands and scale in capacity increments according to your growth pattern. It can be thought of as just-in-time storage or the ultimate in thin provisioning. Plug in a new server node with 1TB, 3TB or 10TB, boot and start using the new storage immediately.

CAStor presents a huge, single flat address space that can store billions of objects in one cluster. File systems hit a wall at high volumes such that a new file system needs to be implemented, provisioned and presented to the user or the application. This is not the case with CAStor as it can easily extend beyond these limitations. Traditional file systems need to manage multiple blocks for each file. Imagine the complexity and constraints this imposes on a traditional file system when there are millions of files and it needs to manage multiple millions of block addresses. CAStor simply needs to manage a single unique identifier (UUID) that is assigned to each piece of content when stored.

Performance is a critical aspect of content storage that is highly and continuously considered by Caringo engineering. Earlier CAS implementations had come to be known as having slow performance and only being useful for archived content that is rarely touched. CAStor breaks down this point of view by providing a storage cluster with near perfect scaling in throughput. Not only does capacity increase with the addition of new server nodes, so too does the cluster's performance improve for I/O as the new processing power is added to the pool. The performance characteristics of CAStor supports the massive number of reads and writes for small files as well as the bits or bytes per second demanded by large files such as videos or medical images. This benefits customers as diverse as those that need to archive large volumes of e-mails on a daily basis to Web 2.0 oriented applications/services that are serving up active content to online communities.

### Storage You Can Rely on

There are several dimensions to reliability that CAStor delivers to ensure the on-going protection and availability of content. When a content object is stored in CAStor it creates one or more exact replicas depending on the policy set by an administrator or application and each is identified by the same UUID and stored on different nodes in the cluster. This ensures continuous data availability in the event a disk or node was to fail as a second replica can be read from a different node that stores the replica. The self-healing nature of the cluster automatically initiates a recovery mode if such an event occurs to create new replicas of the content on the remaining operational nodes. This is accomplished without interrupting applications or data availability and the failed node or disk can be replaced when convenient for the customer.

#### Customer Advantages

**Massive Scalability.** Seamlessly scale from small cluster implementations to massive size according to business requirements.

**High-Performance.** Realize high throughput for small and large files as performance scales with each node added to the cluster.

CASor clusters are also self-balancing in terms of processing I/O requests. Its internal, market-based bidding mechanism ensures that the optimal node to service a given request is selected thus keeping performance at a high level. The cluster is totally symmetrical, which means that each node is independent and provides an entry point into the cluster. This allows the workload to be evenly distributed across all nodes in the cluster and avoid hot-spots that are prevalent in NAS and clustered file systems. A self-managing/balancing clustered storage environment means greater stability and reliability without constant administration and tuning. CASor provides a browser-based administrative console that is the same whether you're managing three nodes in a cluster or hundreds.

Reliability also extends to the individual content objects in the CASor repository and assurances that its integrity remains intact. CASor uses a hash algorithm to generate a digest for each unique content object it stores, which it uses to validate the integrity of the content. A background health processor continuously runs to check the hash digest stored in the metadata associated with each object to make sure there has not been any corruption on disk. If CASor discovers an object has become corrupted, CASor will automatically generate a new replica from the other clean one to replace it.

Additionally, CASor provides robust replication capabilities over local and wide area connections for disaster recovery (DR) and business continuance (BC). It can replicate at LAN speeds to sub-clusters co-located, in a separate room or separate building in a campus environment. Wide area replication allows content to be mirrored or replicated over distance to a disaster recovery site. Replication of CASor clusters can be done in one-to-one, many-to-one, or one-to-many configurations.

## Built to Last

CASor was designed and built with the long-term availability of content, storage and archiving in mind. The ability to access and retrieve content is a key business consideration as it represents corporate assets and intellectual property that must be available months, years, even decades into the future. Much of this is also governed by regulatory mandates for retention of authentic digital records for a prescribed period of time. Caringo understands these key customer requirements and is why it focused on supporting open standards and commodity hardware technology, which is augmented by inherent CAS features that meet compliance mandates for content.

## Anytime, Anywhere Access

Access to content is absolutely critical and is the reason CASor leverages the standard HTTP 1.1 protocol as its primary interface. This protocol has been around for years and will continue to be supported well into the future. CASor's Simple Content Storage Protocol (SCSP) is a subset of HTTP 1.1 that is very easy to implement and similar to interfacing with an Apache web server. Companies in the Web 2.0 segment find this appealing since they are delivering services via the web or offering software as a service over the Internet and HTTP is the common language in that domain.

### Customer Advantages

**Self-healing.** CASor is self-healing that delivers operational reliability and continuous data availability.

**Self-managing.** And self-balancing keep CASor clusters running at optimal performance while assuring content integrity.

**Robust Replication.** Replicate content locally or to geographically dispersed locations for disaster recovery (DR) and business continuance (BC) purposes.

Caringo has also developed the CASTor FSG component (File System Gateway) that runs on SCSP and presents the CIFS and NFS protocols allowing legacy applications to utilize a traditional file system interface to CASTor without having to modify code. Running this along with CASTor essentially provides the most scalable file system available in the market.

### **Non-Stick Storage Hardware**

Storage is a fundamental infrastructure component and a major consideration for IT organizations is the longevity of the media itself as are the costs associated with refreshing hardware technology as it evolves. This is especially true for organizations evaluating content archive storage for retention and compliance that needs to retain records for years or decades. CASTor is software-only content storage that is agnostic to the underlying hardware (any x86 architecture) allowing customers to use the server technology they choose as well as non-disruptively upgrade it when new hardware is introduced.

With a CASTor cluster x86 platforms can be intermixed in terms of hardware configuration and vendor, e.g. Dell, HP or IBM. When a new server is introduced to the market you can simply boot it as a new node into the cluster and retire the aging node. The content on the retired node is replicated within the cluster and when that is complete it can simply be removed. There is no disruption to applications or content availability and this can be done one node at a time or in multiple node increments. Essentially, CASTor abstracts the content from the underlying storage media freeing the customer from lock-in and enabling a continuous hardware lifecycle.

### **Industrial Strength Archive**

Archiving is an area where CAS cut its teeth and CASTor hits on all cylinders here as well as advancing the state of the art. The integrity and authenticity of content are integral parts of an archive especially where compliance is concerned. Establishing authenticity of a specific content object is achieved by running it through a hash algorithm, such as MD-5, which generates a 128-bit digest that is used as a digital signature. Unlike previous CAS generations that use this digest as the address for the content, CASTor creates a UUID (128-bit random number) for the address and stores the digest separately in the content metadata. This allows CASTor to transparently upgrade the hash algorithm (patent pending) before the one that was originally used becomes compromised such as what occurred with MD-5 and SHA-1.

#### **Customer Advantages**

**Open Interface.** Open standard for interfacing to CASTor ensures content is accessible well into the future.

**Hardware Agnostic.** Freedom of choice in hardware eliminates lock-in and enables seamless technology refresh and longevity.

**Content Integrity.** Archive for compliance with assured content integrity, retention, authenticity and protection in WORM storage.

Another key to archiving is the capability to include metadata elements that enable the retention and lifecycle of content to be managed within the storage environment. CASTor allows applications or administrators to define retention periods, number of replicas and deletion policy in the metadata. The metadata values ensure that the content is retained for the required period, that a certain number of replicas are stored so there's always a backup and that deletion is permitted when the retention period expires. In addition, CASTor provides a metadata element known as a *LifePoint* in which the complete lifecycle of a particular content object can be defined and is managed by the *HealthProcessor*. Another compelling feature of CASTor is the ability for applications and/or customers to define custom metadata elements that can be used to uniquely manage content.

## Conclusion

The demands of digital content, the volume being created today and what is expected in the years ahead exceed the capabilities of traditional file system solutions. Customers are encumbered by complexity and cost inherent in these technologies and need a better solution to better manage their infrastructure and support their business. Emerging companies in the Web 2.0 sector that provide community oriented offerings also need affordable robust storage infrastructure for massive amounts of consumer-generated content. CASstor clustered storage software for active and archive content delivers the features and functionality customers have expressed a need for. It's simple to implement, scale and manage. It eliminates hardware lock-in leveraging affordable, commodity server hardware instead of expensive proprietary appliances. It separates content from the underlying hardware so that the CASstor repository can evolve as the server technology does without impact to operations or content availability. It provides the retention, preservation and protection required that allow organizations to meet archive and regulatory compliance mandates.

CASstor is content storage simplified.

## About Caringo

With CASstor content storage software, Caringo has re-imagined content addressed storage (CAS) from the ground up. Its flagship CASstor product is third-generation technology that leverages standard commodity hardware that dramatically changes the economics of corporate content storage.

For more info, visit [www.caringo.com](http://www.caringo.com).

### Footnotes

- 1 The Diverse and Exploding Digital Universe, IDC, Christopher Chute et al, March 2008
- 2 Brocade Customer Survey Results, September 2007
- 3 The Expanding Digital Universe, IDC, David Reinsel et al, March 2007
- 4 The Economic Impact of File Virtualization, IDC, Richard Villars, May 2007
- 5 Web 2.0 Storage: Challenges and Choices, Bob Laliberte, Storage Magazine, November 2007