

Centera Foundations

Upon completion of this module, you will be able to:

- Define Content Addressed Storage (CAS)
- Define the terms associated with CAS data flow
- Describe how data is processed in a CAS environment
- List the benefits of using Centera to store fixed content

Economic changes in communications and storage have made it necessary to move fixed data content to a network-accessible format. EMC's Centera is the first of its kind platform, designed and optimized specifically to deal with the characteristics of fixed data content. Huge forces are driving customers to manage that content on-line. For example, new regulatory requirements compel healthcare providers and insurance companies to provide access to medical records. Centera's RAIN (Redundant Array of Independent Nodes) architecture and use of C-Clip technology for content addressing allow organizations to meet these needs.

The objectives for this module are shown here. Please take a moment to read them.

Centera Foundations

CENTERA AND CAS – A DEFINITION

In this first section, we will briefly define what Content Addressed Storage (CAS) is and discuss the hardware platform that supports it, namely, Centera.

EMC Education Services

EMC²
where information lives[®]

Traditional Storage vs Content Addressed Storage

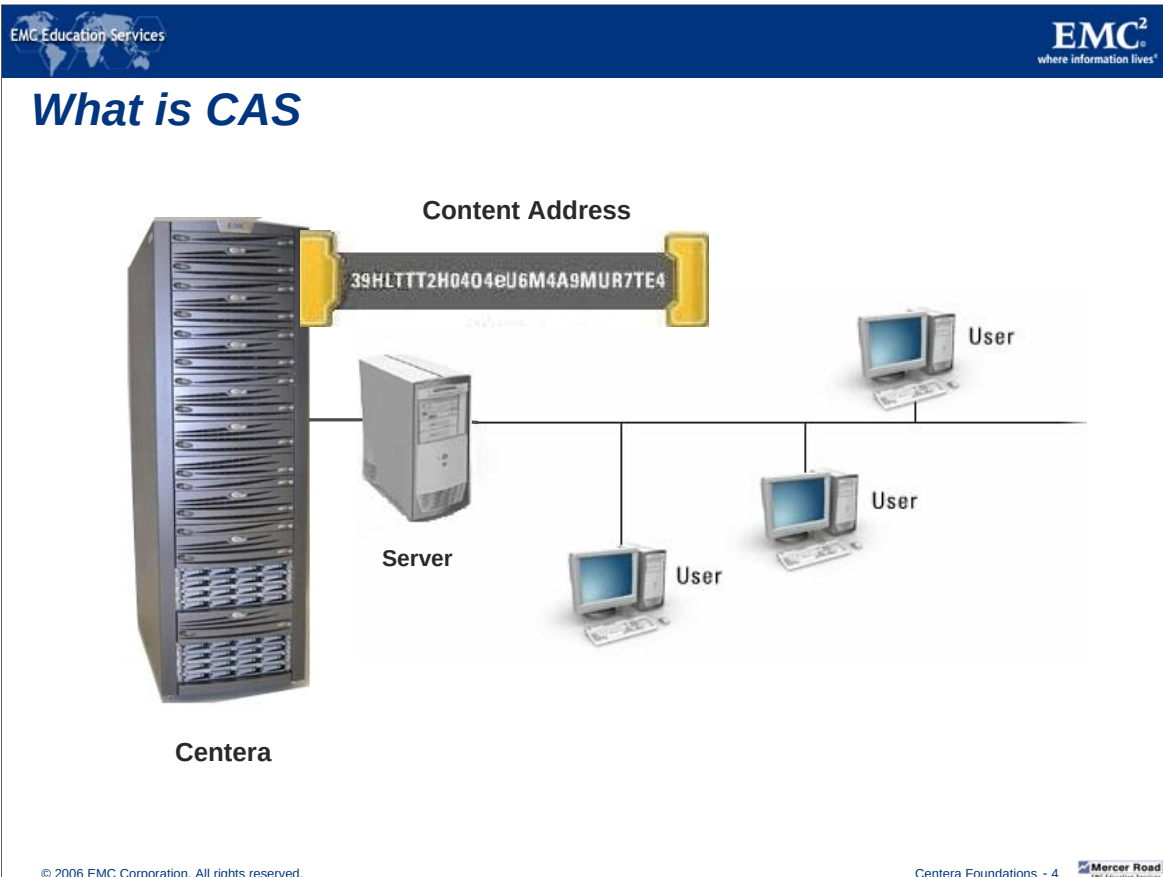
	SAN Storage Area Networks	NAS Network-Attached Storage	CAS Content Addressed Storage
Type of transport	Fibre Channel IP (emerging)	IP	IP
Type of data	Block	File	Object, fixed content
Key requirement	Deterministic performance	Multi-protocol Sharing	Longevity, integrity assurance
Typical applications	OLTP, data warehousing, ERP	Software and product development, file server consolidation	Content Management, Archive
<i>Information Lifecycle</i>	<i>Content is created and actively shared</i>		<i>Content is fixed and preserved</i>

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 3

Mercer Road

Traditional disk storage systems use block or file access schemes that are well suited to transaction oriented, update intensive, data storage solutions. In a fixed content environment, it becomes a challenge to manage the logistics of data placement and capacity scaling, while also assuring authenticity of the content over its lifetime.

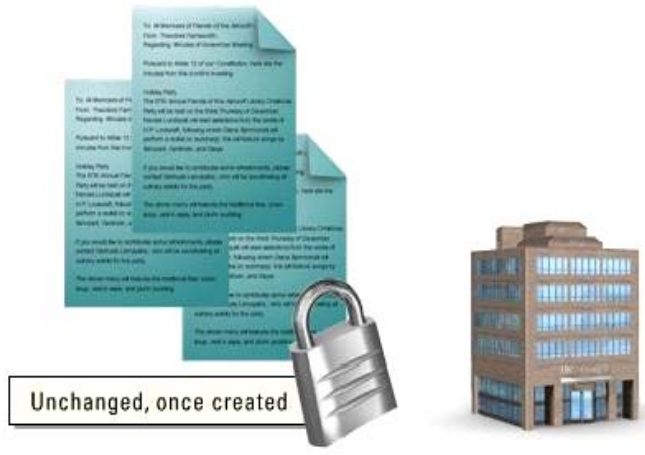


CAS (Content Address Storage) is a new category of storage designed for the secure online storage and retrieval of fixed content. Rather than access a data object by its file name at a physical location, a CAS device uses a Content Address to store and retrieve the object, where the address of the object (e.g. a file) is created from the unique content of that file. The Content Address is a globally unique identifier, generated by a hashing algorithm. Content Address or “content addressing” will be discussed more thoroughly throughout this module.

The CAS market cuts across multiple vertical industry segments. In each of these market segments, content must be preserved intact for years, if not decades. This kind of content has often ended up on tape or optical disk where, if the data can still be accessed, retrieval may be so delayed that the usefulness of the information is often negated.



What is Fixed Content



Unchanged, once created

Data in its final form

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 5



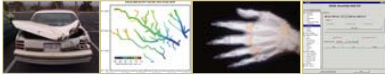
Fixed content refers to any informational object retained for future reference and business value, including electronic documents and many types of newly digitized information. Unlike transactions or files, it is typically unchanged once created. If you think about the lifecycle of information, it ultimately all leads to fixed content. Content, like email, clinical trial data, CAD/CAM drawings, or electronic documents, may begin as transactional or collaborative work but ultimately becomes fixed content. It is at this point that its value comes from expanded use and not its ability to change.

Fixed content is often contained in large, long-lifetime objects. The quantity is constantly expanding. Regulatory, auditing, and consumer access needs prevent changes to the information. Frequent and fast retrieval is often required, and there are typically many users in many locations. Online availability significantly increases the business value of archived reference information.



Examples of Fixed Content

Unchanging Data Objects With Long-Term Value

X-rays	MP3s	
MRIs	Movies	
CAT scans	Recordings	Genomic data
Blueprints	Transcripts	Proteomic data
Insurance photos	Professional photos	Clinical trial results
Contracts	Consumer photos	Biometric data
	Educational videos	Lab notebooks
	Surveillance videos	Backups
Newspapers	Seismic data	Historical documents
Periodicals	Astronomic data	Presentations
Books	Spreadsheets	Monthly reports
Marketing collateral	Graphics	Video conferences
Check images	Source code	Audio conferences
White papers		News clips
CAD/CAM originals		Sports videos
E-mail and attachments	Training materials	Government records
	Manuals	

© 2006 EMC Corporation. All rights reserved.

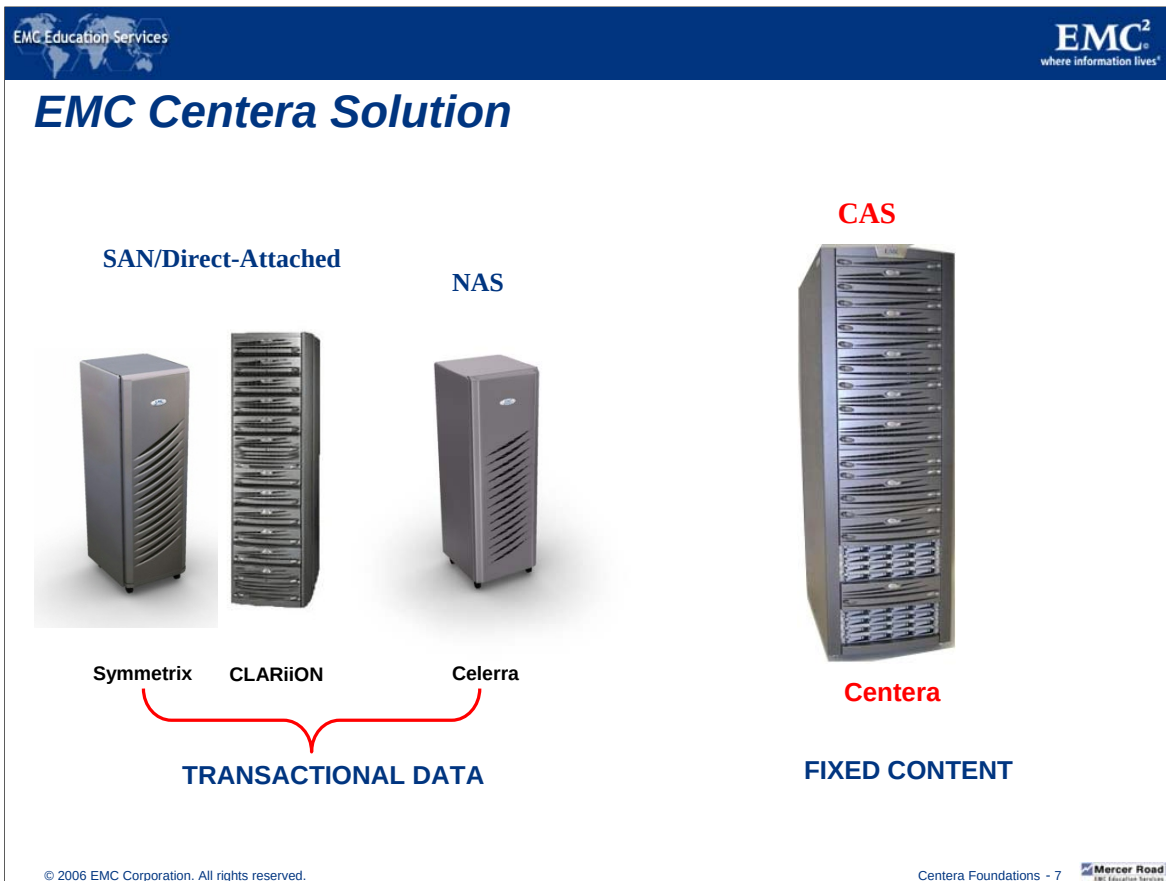
Centera Foundations - 6 

Traditionally, the majority of fixed content has been stored on tape or optical technologies. While these technologies can store this content, neither these, nor traditional magnetic disk solutions, were built to handle the very unique requirements for storing final form content. CAS can:

- Provide online access with assured content authenticity
- Efficiently store the content by eliminating the storage of duplicate content
- Scale easily and seamlessly to hundreds of terabytes
- Provide low administration costs by having self-configuring, self-healing, and self-managing functionality

When looking for optimum solutions for fixed content, tape and optical solutions are inadequate. They are too slow, there have been too many in-technology changes that have resulted in lost or unusable content, and reliability is questionable (a tape concern), as is the industry's commitment to the technologies (a point specific to optical).

Common storage alternatives have not been designed with the storage management capabilities found in Centera. These typically do not scale beyond a few terabytes (and/or individual devices) before the operational complexities (and costs) become a significant barrier. For example, if an application requires more storage than fits within a single volume or physical storage device, management complexity increases significantly. Not only is the application challenged by the expanding filesystem hierarchy, but the storage manager is faced with time-consuming reallocation and data relocation, not to mention the complexities of replicating information to multiple sites for purposes of sharing or disaster recovery.



Requirements to store and retrieve fixed content items are much different and, in many ways, more taxing, than those of transactional data (which is traditionally handled by SAN and NAS storage solutions). Organizations need new ways to manage these increasing amounts of reference information, which is typically unchanged once it is created, and may need to remain online for many years due to regulatory or consumer access requirements. Centera fulfills these requirements by providing faster record retrieval than traditional backups, single instance storage, guaranteed content authenticity, and self-healing, as well as numerous industry regulatory standards.

EMC offers networked storage solutions for every business need: SAN for business and technical applications requiring optimized transaction performance; NAS for high-availability file sharing and collaboration; and CAS for storage and retrieval of fixed content. Whether you need SAN, NAS, CAS, or a combination, only EMC can deliver and integrate all three to work together seamlessly in your environment.

Centera Models

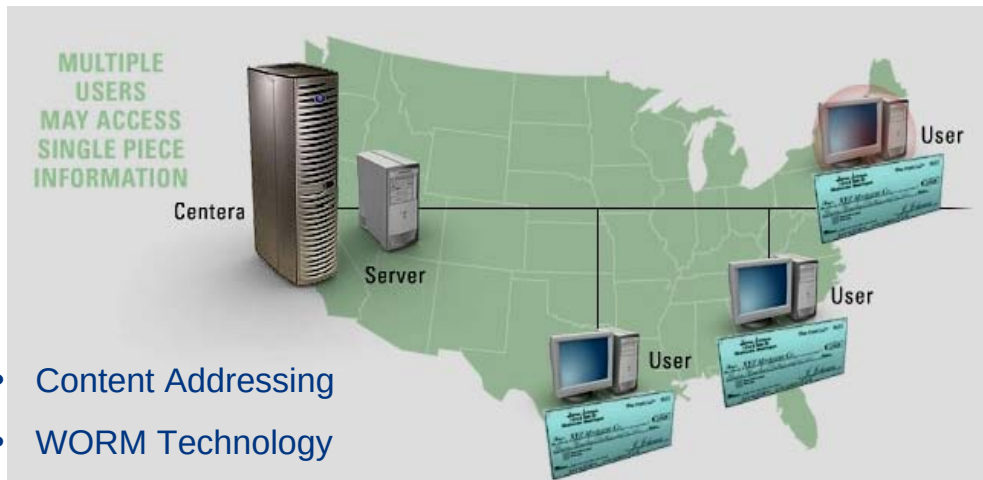
- **Centera Basic**
 - Provides all functionality without enforcement of retention periods
- **Centera Governance Edition**
 - Process-centric on the lifecycle of electronic records and enabling policies and technologies
 - Restricts the retention and deletion of data but does not conform to SEC regulations
 - Suitable for most regulations
- **Centera Compliance Edition Plus (CE+)**
 - Designed for the strictest of regulation requirements, specifically SEC 17a-4
 - Restricts the retention and deletion of data according to SEC regulations

The Centera is offered in 3 different models, dependent on the needs of the individual customer. These needs are based on the data being stored and the stringency of the customers regulatory needs. The Centera Basic model is the least restricted of the models and does not provide any enforcement of retention periods or data shredding. The Centera Governance Edition is suitable for most regulatory needs and does restrict retention and deletion of data. Centera Compliance Edition Plus is the most secure model and follows the strictest regulation requirements demanded by the SEC.

Later in the presentation, we will give a few examples of the regulations, and how Centera features meet and fulfill their requirements.

Centera Advantages –

Faster record retrieval than other methods of storing fixed content



- Content Addressing
- WORM Technology
- GUID
- Single Instance Storage

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 9 

Centera employs a unique storage/retrieval method called "content addressing", which is performed through the use of C-Clip technology. Centera offers a vast number of advantages such as WORM (Write Once Read Many) technology, GUID (Globally Unique Identifier), and one instance of data for multiple clients (Single Instance Storage).

WORM technology is non-editable data storage. Any changes made to the data will create a new object on the Centera, leaving the original object intact.

GUID is a pseudo random number used to uniquely identify the data stored on the Centera through the Content Address.

Single Instance Storage is the ability to better utilize storage resources by only storing a single copy of data but allowing it to be referenced multiple users. This is discussed in more detail later in the presentation.

Centera Foundations

CONTENT ADDRESSED STORAGE – TERMINOLOGY

As with any new technology innovation, there will always be a number of new concepts and names for those concepts. In this section, we will briefly describe the new objects within CAS.



CAS Terminology



API



BLOB



Metadata

- **Application Programming Interface (API)**
 - A set of function calls that enables communication between applications or between an application and an operating system
- **BLOB**
 - The Distinct Bit Sequence (DBS) of user data represents the actual content of a file and is independent of the filename and physical location
- **Metadata**
 - Metadata or "data about data" describes the content, quality, condition, and other characteristics of data

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 11



As previously mentioned, Centera uses Content Addressing to store and retrieve data. To follow the sequence of data from a Client to the Centera, new terminology must be defined.

Application Programming Interface (API)

A set of function calls that enables communication between applications or between an application and an operating system.

BLOB

The Distinct Bit Sequence (DBS) of user data. The DBS represents the actual content of a file and is independent of the filename and physical location.

Metadata

Metadata, or "data about data", describe the content, quality, condition, and other characteristics of data.



Review of CAS Terminology (continued)



C-Clip



Content Address



CDF



C-Clip ID

- **C-Clip**
 - A package containing the user's data and associated metadata
- **Content Address (CA)**
 - An identifier that uniquely addresses the content of a file and not its location. Unlike location-based addresses, content addresses are inherently stable and, once calculated, they never change and always refer to the same content
- **C-Clip Descriptor File (CDF)**
 - The additional XML file that the system creates when making a C-Clip. This file includes the content addresses for all referenced BLOBs and associated metadata
- **C-Clip ID**
 - The content address that the system returns to the client. It is also referred to as a C-Clip handle and C-Clip reference

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 12 

C-Clip

A package containing the user's data and associated metadata.

C-Clip ID

The content address that the system returns to the client. It is also referred to as a C-Clip handle and C-Clip reference.

C-Clip Descriptor File (CDF)

The additional XML file that the system creates when making a C-Clip. This file includes the content addresses for all referenced BLOBs and associated metadata.

Content Address (CA)

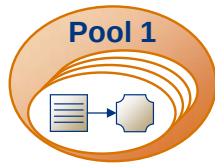
An identifier that uniquely addresses the content of a file and not its location. Unlike location-based addresses, content addresses are inherently stable and, once calculated, never change and always refer to the same content.

Review of CAS Terminology (continued)



- **Access Profile**

- Access Profiles are used by access applications to authenticate to a Cluster, and by Clusters to authenticate themselves to each other for replication or restore connections.



- **Virtual Pools**

- Virtual Pools will allow a single logical cluster to be broken up into arbitrary number of smaller logical clusters.

Access Profile

Access Profiles are used by access applications to authenticate to a Cluster, and by Clusters to authenticate themselves to each other for replication or restore connections.

Virtual Pools

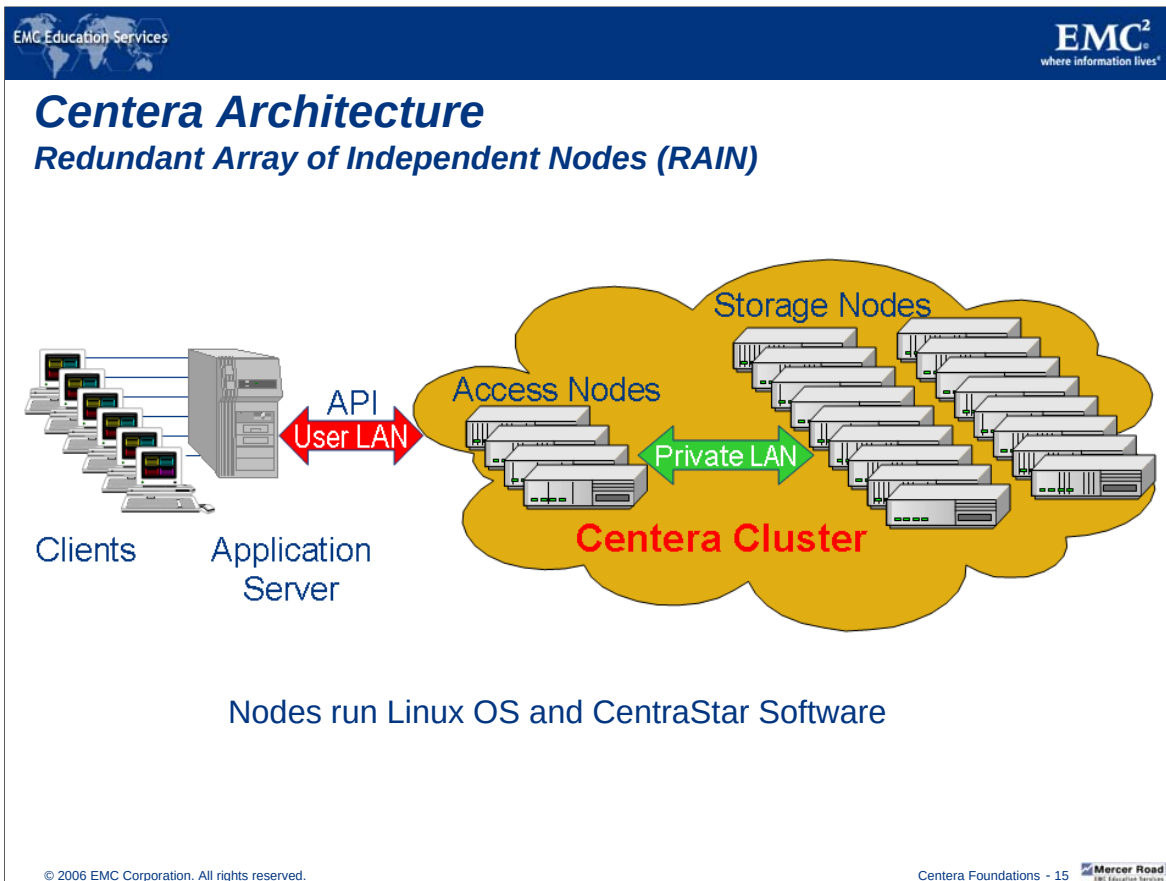
Virtual Pools will allow a single logical cluster to be broken up into arbitrary number of smaller logical clusters.

CENTERA ARCHITECTURE



Centera Foundations - 14

Centera Foundations - 14 



The architecture of the Centera system, based on Redundant Array of Independent Nodes (RAIN), is designed to be highly scalable and hold petabytes of content. Each node has its own Linux OS and CentraStar microcode and utilizes a distributed workload. The nodes are connected together to form a cluster. CentraStar is the application (microcode) that runs the Centera and all its features.

Client applications write to the Centera using the Centera API. Application store and retrieval requests are sent to the Access Node via “public” IP connections. The Access Node uses a unique Content Address to locate the requested information from the Storage Nodes over a private internal LAN, and supplies the necessary information back to the client through the API. The Storage Nodes are responsible for the long-term storage and protection of the BLOBs and CDFs.

Centera Generation 4 Configuration

- 4 Node entry system



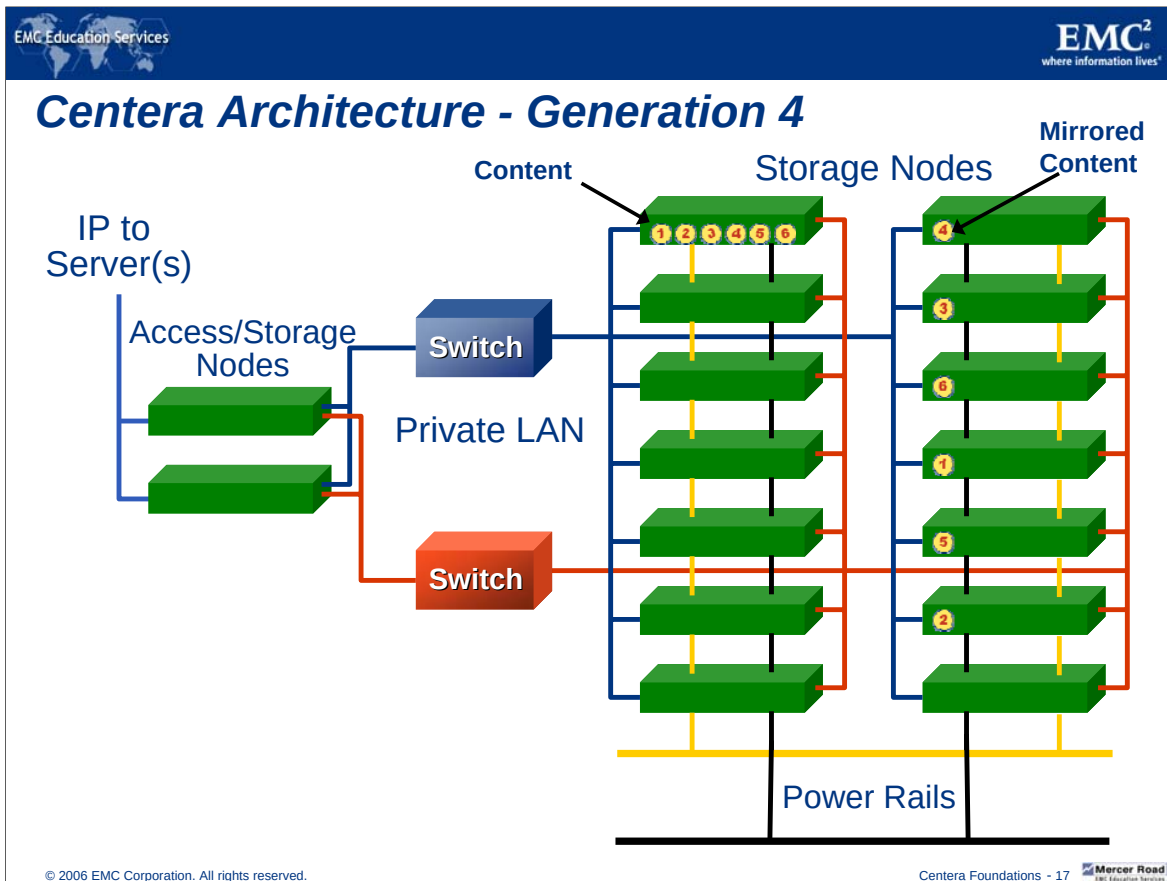
1 Cube = 16 nodes + 2 internal switches

1 Cabinet = 2 Cubes



While the Centera cabinet may contain 32 nodes, the minimum configuration can contain as few as four Generation 4 (Gen4) nodes and 2 switches, and is expandable in increments of 4 nodes. A cube is configured with a maximum of 16 nodes and 2 internal switches. So a fully configured cabinet will be comprised of 2 cubes. The Gen4 hardware is also available as an unbundled solution. It can be installed in an approved non-EMC cabinet.

Each Gen4 node contains a built in automatic transfer switches (ATS) to allow them to connect to two power sources. Each power outlet on a node will connect to a separate power supply. This has eliminated the needs for both the external ATS required for previous generations as well as connecting each node to a different power rail.



Each storage node contains more than 1TB of usable capacity, and both access and storage roles can reside on the same node. So a node may act as both an access node as well as a storage node. Also the nodes now have built in optical access capabilities where, before, a separate adapter was required.

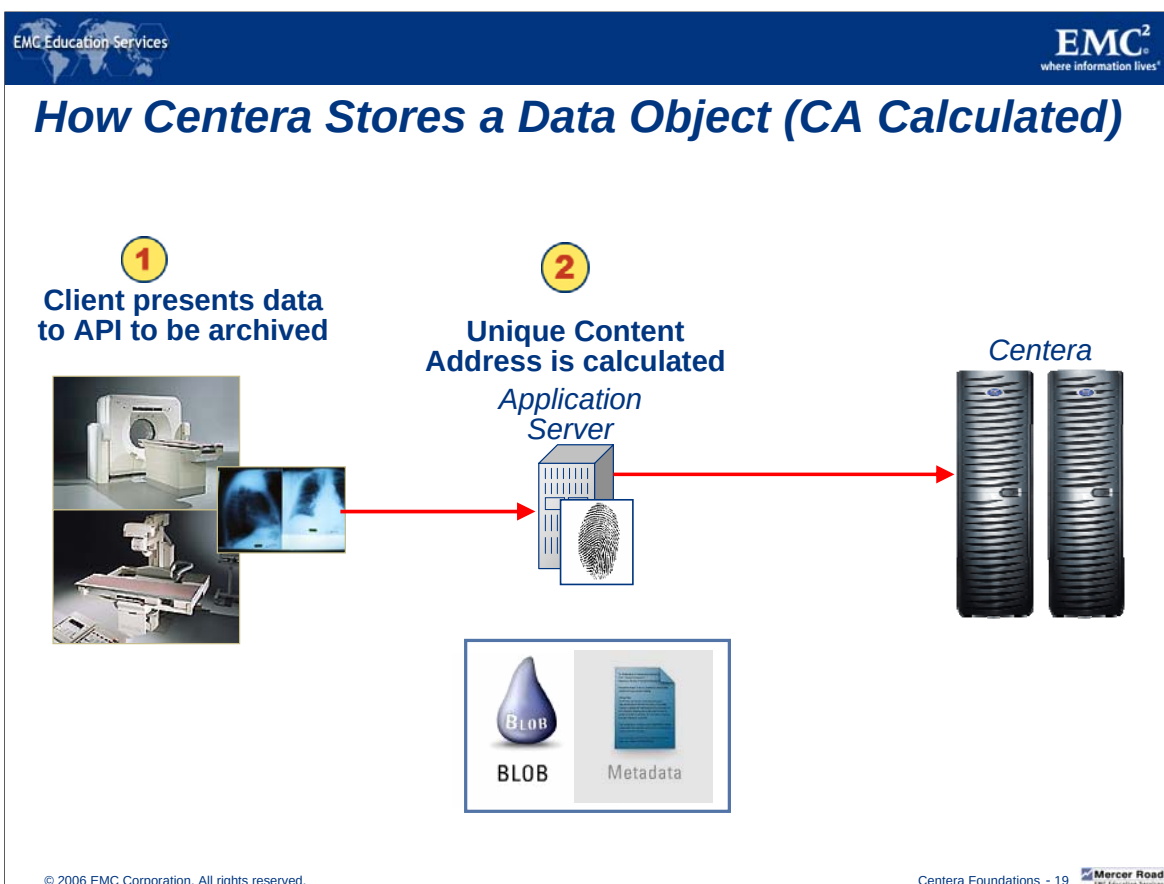
The internal switches are 24 port 2GB switches that provide communications for up to 16 nodes within the private LAN. The Centera cabinet may contain up to 32 nodes, with a minimum configuration containing as few as 4 nodes. Several cabinets can be connected to form a Centera cluster. Applications may support multiple clusters within a Centera domain. Root switches are used for connecting 3 or more cubes together into a single cluster. Each cabinet has from 4 TB to 23 TB or more of usable protected capacity, depending on whether they have Content Protection Parity (CPP) or Content Protection Mirrored (CPM) set.

Note: This slide shows a CPM configuration with the content mirrored to other nodes.

Centera Foundations

THEORY OF OPERATION

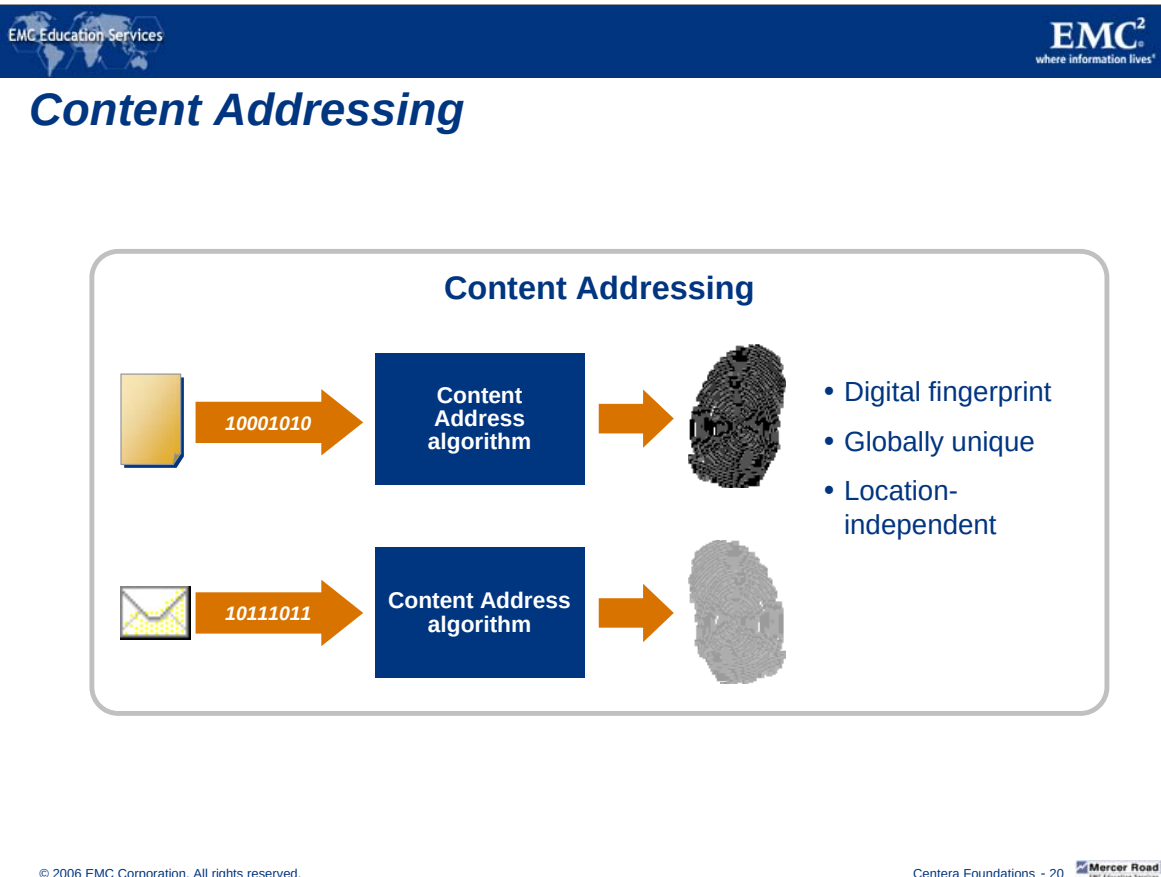
Having discussed the hardware aspect of the solution, we will now briefly discuss the theory of operation behind CAS.



The Centera API facilitates access from an application server to the Centera cluster. Applications that typically interface with Centera are applications that manage the storage and retrieval of fixed digital content such as X-rays, check images, scanned mortgage contracts, and more. This content is fixed in the sense that once it has been stored, it does not change. This type of content is stored according to the WORM principle: “written once and read many”.

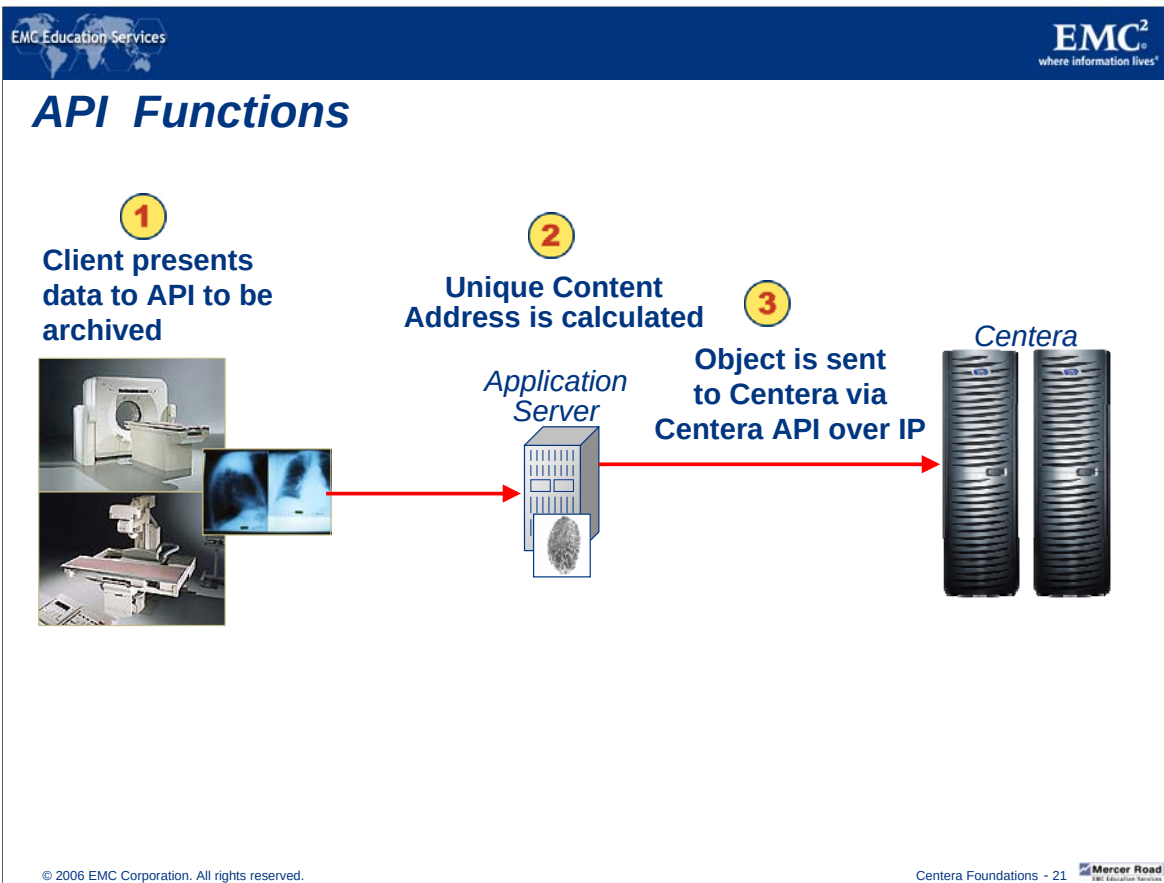
End users input their data to content management applications that interface with the Centera system via the Applications Program Interface (API). The API is part of the Centera SDK (Software Development Kit).

The API separates the actual data (BLOB) from the metadata and the Content Address (CA) is calculated from the object's binary representation.



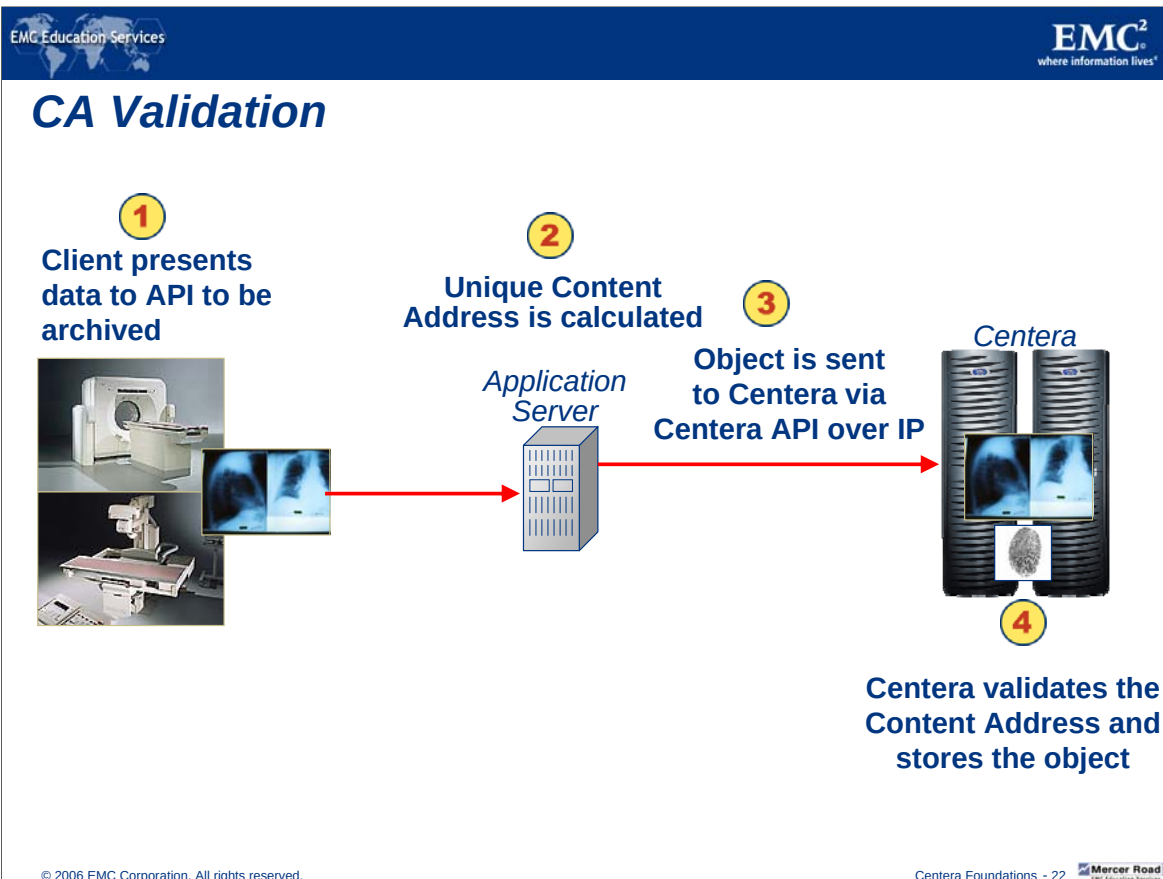
Centera's Content Address is a derived address. It is the result of a hashing algorithm run across the binary representation of the object. It takes into account all aspects of the content, even file type. And what it returns to a user's application is a content fingerprint unique to that content.

A unique number is calculated by the hash algorithm from the sequence of bits that constitutes the content of a file. If a single byte changes in the file, then any resulting calculation will be different. This fingerprint is now used as the Content Address for the data that is to be stored on the Centera. When viewed, this unique number will be displayed in a 27 or 53 character format, depending on the type of storage strategy (Storage Strategy Capacity or Performance) and collision avoidance chosen by the customer.

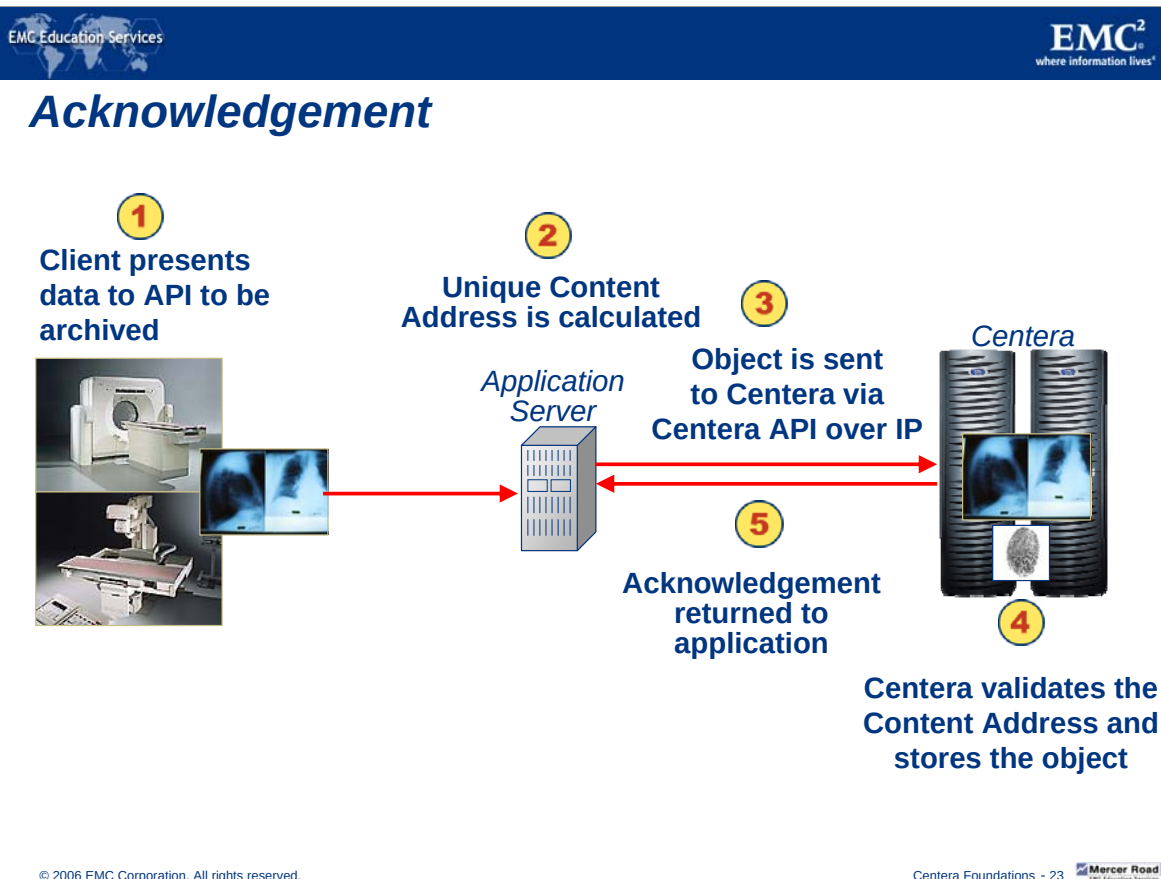


The content address and metadata of the object, such as its file name and creation date, are then inserted into an XML file known as the C-Clip Descriptor File (CDF) and transferred to the Centera. The combination of the BLOB and the CDF is referred to as the C-Clip, which is stored on the Centera.

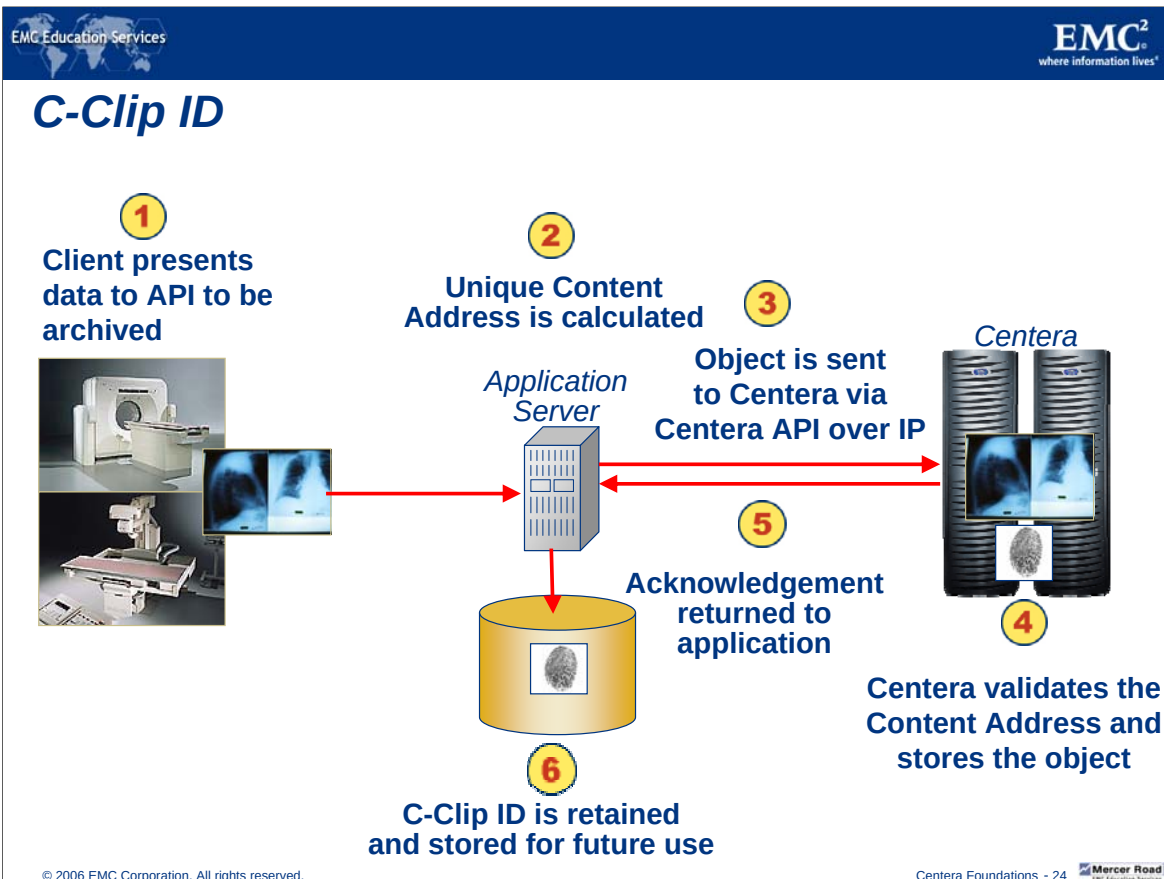
NOTE: XML refers to Extensible Markup Language. It allows the flexible development of user defined document types.



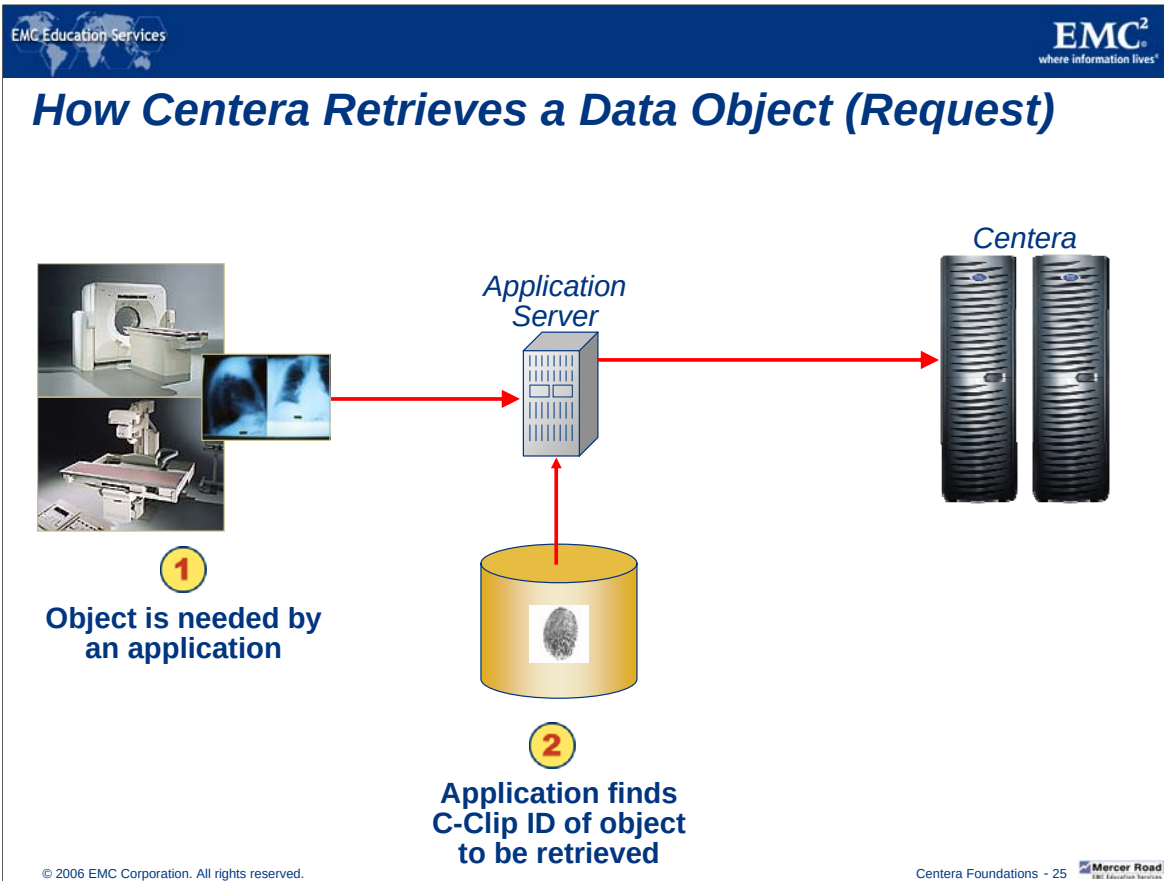
Centera recalculates the object's Content Address as a validation step and stores the object. This is to ensure that the content of the object has not changed. If the data has been modified, then a new CA will be generated, and the object will be stored separately as its own BLOB.



An acknowledgment is only returned to the API once a mirrored copy of the C-Clip Descriptor File (CDF), and a protected copy of the BLOB, have been safely stored in the Centera repository. Once a data object is stored in the Centera repository, the API is given a C-Clip ID (also called a C-Clip handle or C-Clip reference).



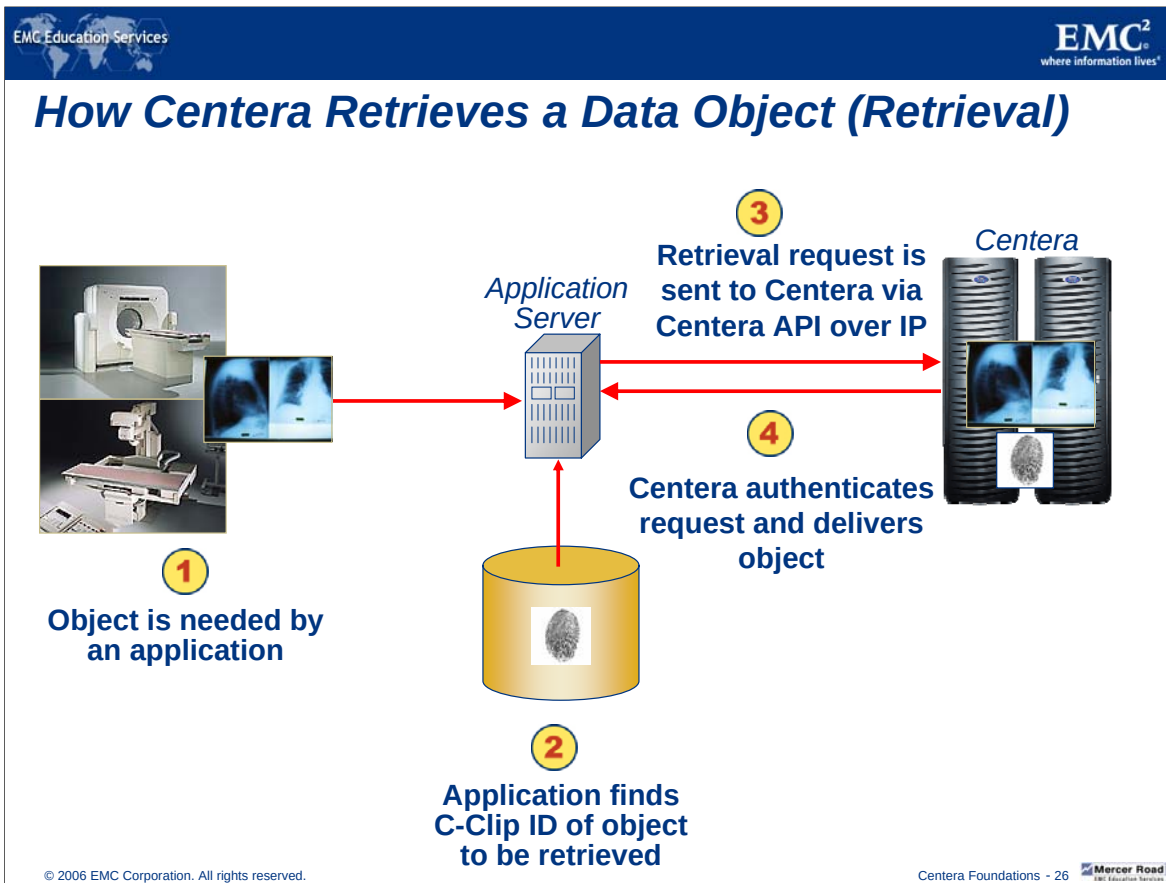
The C-Clip ID is a content address of the CDF, which contains the CA of the actual data on the Centera. Using the C-Clip ID, the application can read the data back from the Centera at any time. There is no centralized directory in the Centera, no pathnames or URLs are used. Where the data is stored on the Centera is transparent to the application.



Here is the process on how the Centera retrieves a data object:

Step 1. An object is required by a user or an application.

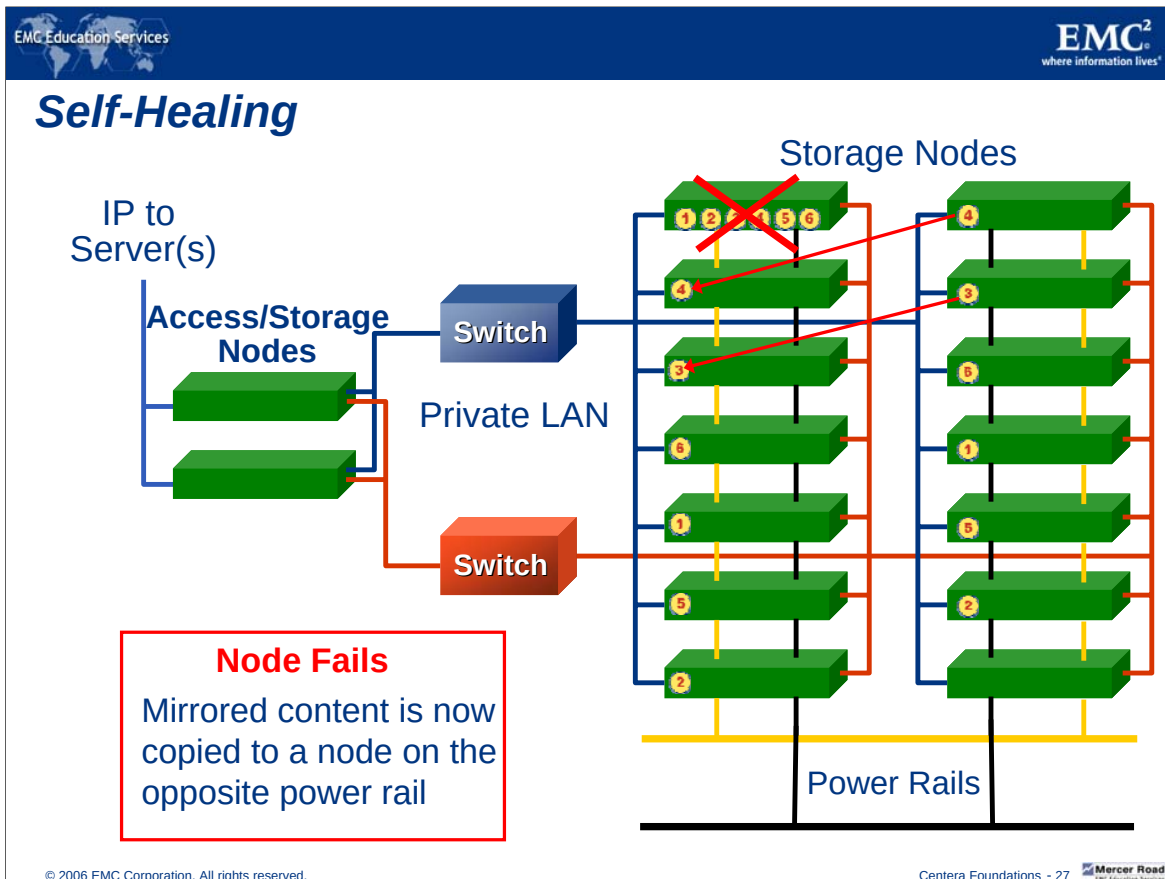
Step 2. The application queries the local table of C-Clip IDs and locates the C-Clip ID for the needed object.



Step 3. Using the Centera API, a Retrieval request is sent, along with the C-Clip ID, to the Centera.

Step 4. Centera delivers the requested information to the application which, in turn, delivers it to the client.

The Client does not need to know where the data is physically located, as it is the Centera that determines where the data is stored. It only requires the unique address that is used to identify the data so that it can be retrieved.



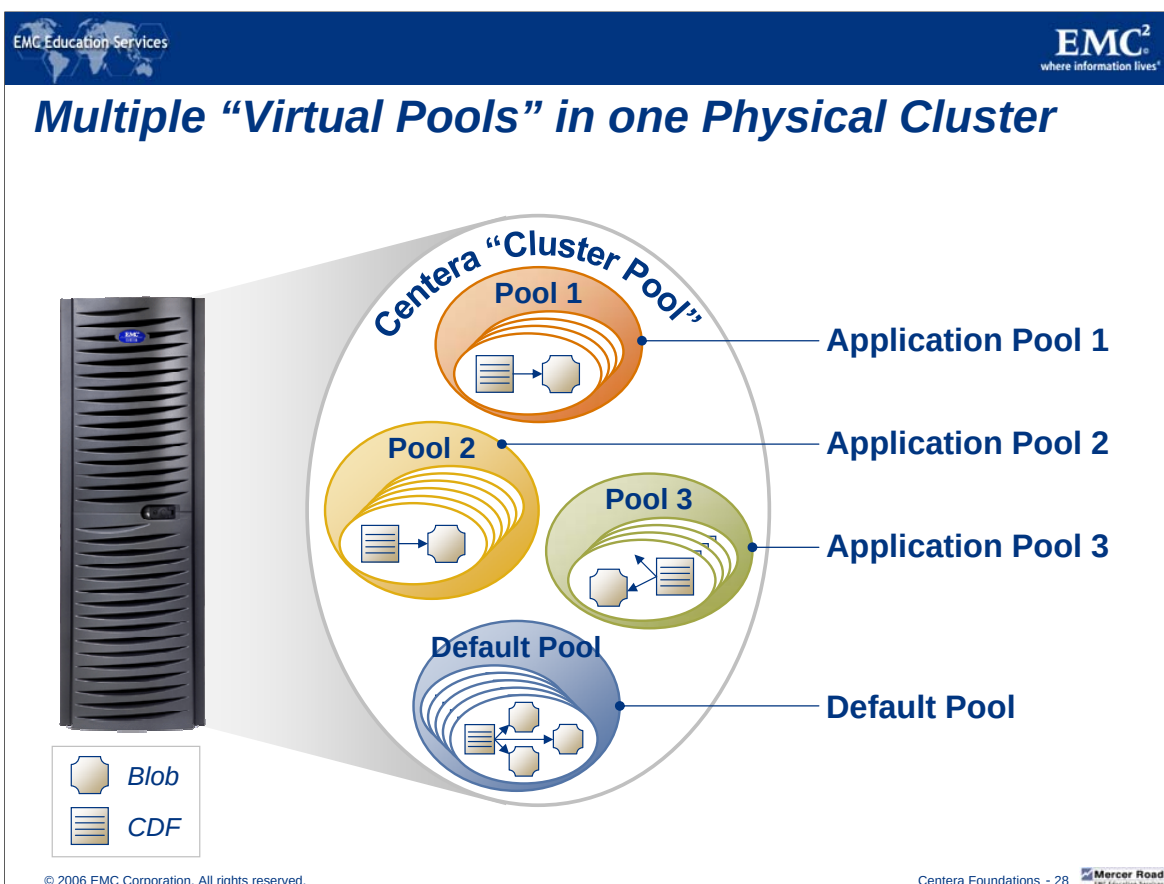
The Centera is a self-configuring, self-healing, and self-managing solution. It offers different methods of content protection. The example in the above slide shows Content Protection Mirrored.

CPM (Content Protection Mirrored) is where every data object is mirrored. There are 2 copies of every piece of data sent to the Centera that will reside on different nodes. If a node or disk should fail, the Centera software would automatically broadcast to the node with the mirrored copy to regenerate another copy to a different node so that there will always be 2 copies available.

In CPP (Content Protection Parity), the data is fragmented into segments, with a parity segment. Each segment is on a different node, similar to a file type RAID. If a node or disk should fail, then the other nodes would recreate the missing segment and put it on a different node. Either protection scheme provides total protection against failure through the Centera's unique self-healing functions.

Centera performs continuous Content Integrity Checking:

- Centera constantly validates the integrity of its data objects and structure
- Centera does ongoing background data scrubbing
 - If the data is not protected, Centera will automatically ensure the content is protected
- Centera does constant authenticity checking to prevent data corruption
- Automated Garbage Collection



The Centera uses Virtual Pools to store the data. Clips can be written to a specific pool where access can be granted to some applications but not others. With Virtual Pools, a Cluster can have multiple pools (up to 98 application pools). Access to the Virtual Pool is granted to the Access Profile, which is used by the applications to gain access to the clips contained in a particular pool. This provides better security, flexibility, and control of the data on the Centera.

Replication, Restore, Seek, and Chargeback can now also be performed on a Virtual Pool basis.

Virtual Pools, coupled with Centera's existing capabilities for searching storage-based metadata, allow for management of an archive as broad or as granular as desired.

EMC Education Services

EMC²
where information lives[®]

Centera Foundations

CENTERA TOOLS



© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 29 

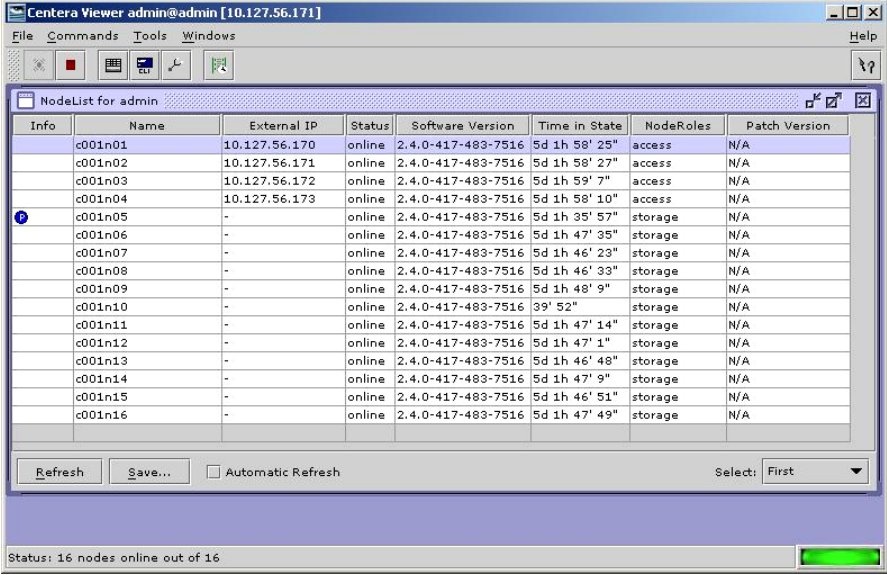
Some tools that are available for administration of the solution will be briefly discussed here.

EMC Education Services

EMC²
where information lives[®]

Centera Tools

Example of Centera Viewer



The screenshot shows the Centera Viewer application window titled "Centera Viewer admin@admin [10.127.56.171]". The window has a menu bar with "File", "Commands", "Tools", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations and a search icon. The main content area displays a table titled "NodeList for admin". The table has columns: Info, Name, External IP, Status, Software Version, Time in State, NodeRoles, and Patch Version. The table lists 16 nodes, all with a status of "online". The first four nodes (c001n01 to c001n04) have external IP addresses and are assigned the "access" role. The remaining 12 nodes (c001n05 to c001n16) have a "-" for external IP and are assigned the "storage" role. At the bottom of the window, there is a status bar that reads "Status: 16 nodes online out of 16" and a green progress indicator.

Info	Name	External IP	Status	Software Version	Time in State	NodeRoles	Patch Version
	c001n01	10.127.56.170	online	2.4.0-417-483-7516	5d 1h 58' 25"	access	N/A
	c001n02	10.127.56.171	online	2.4.0-417-483-7516	5d 1h 58' 27"	access	N/A
	c001n03	10.127.56.172	online	2.4.0-417-483-7516	5d 1h 59' 7"	access	N/A
	c001n04	10.127.56.173	online	2.4.0-417-483-7516	5d 1h 58' 10"	access	N/A
	c001n05	-	online	2.4.0-417-483-7516	5d 1h 35' 57"	storage	N/A
	c001n06	-	online	2.4.0-417-483-7516	5d 1h 47' 35"	storage	N/A
	c001n07	-	online	2.4.0-417-483-7516	5d 1h 46' 23"	storage	N/A
	c001n08	-	online	2.4.0-417-483-7516	5d 1h 46' 33"	storage	N/A
	c001n09	-	online	2.4.0-417-483-7516	5d 1h 48' 9"	storage	N/A
	c001n10	-	online	2.4.0-417-483-7516	39' 52"	storage	N/A
	c001n11	-	online	2.4.0-417-483-7516	5d 1h 47' 14"	storage	N/A
	c001n12	-	online	2.4.0-417-483-7516	5d 1h 47' 1"	storage	N/A
	c001n13	-	online	2.4.0-417-483-7516	5d 1h 46' 48"	storage	N/A
	c001n14	-	online	2.4.0-417-483-7516	5d 1h 47' 9"	storage	N/A
	c001n15	-	online	2.4.0-417-483-7516	5d 1h 46' 51"	storage	N/A
	c001n16	-	online	2.4.0-417-483-7516	5d 1h 47' 49"	storage	N/A

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 30

Mercer Road
EMC Education Services

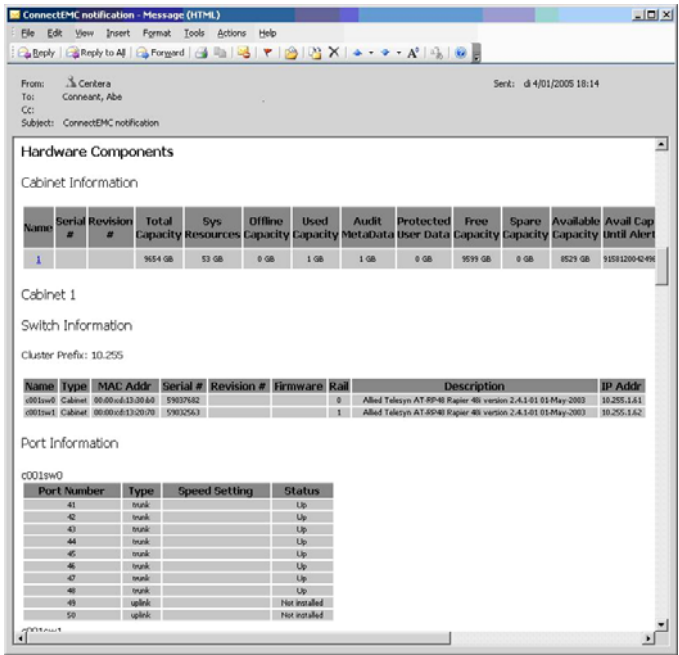
System Administrators do not need to worry about volume creation/management, file system structure/maintenance, or data location, as the Centera controls all of these functions. They do need to be able to monitor the Centera's capacity and performance, and EMC personnel/partners need to maintain the Centera and to enable and disable its features as needed.

A group of tools is provided, both to the customers as well as the service personnel. The most commonly used tool is the Centera Viewer, a GUI (shown above) loaded onto a Windows PC with network access to the Centera that provides a simple means of displaying capacity utilization and operational performance of the Centera. It also enables the system administrator to change any site-specific information such as the public network information, as well as end-user contact information. It is also a tool most commonly used by EMC personnel and partners to troubleshoot failures and for upgrading the CentraStar code. There is also a Command Line Interface (CLI) that can be used with the GUI or as a standalone tool in a UNIX environment.

Other tools available are Centera Monitor, which allows the customer to monitor a CE+ Centera, and Simple Network Management Protocol (SNMP) is used to alert an enterprise network management system to any faults that might occur within a Centera. Centera can also be monitored with the EMC ControlCenter v5.2 or higher through sensor agents built into the CentraStar code..

EMC Education Services
EMC²
where information lives[®]

Centera Health Reporting



Example of a
Health Report in
HTML email
format

© 2006 EMC Corporation. All rights reserved.
Centera Foundations - 31

The System Health Report is an automatic email message that a Centera cluster periodically sends to the EMC Customer Support Center with a list of predefined recipients. It reports on the current status of the Centera cluster. The report is sent to the EMC Customer Support Center in an XML format. EMC is then able to monitor the Centera cluster remotely and detect any hardware or software problems.

When it is sent to other recipients, it is converted into an HTML format for easy reading of its data (as shown in the above slide).



Centera Foundations

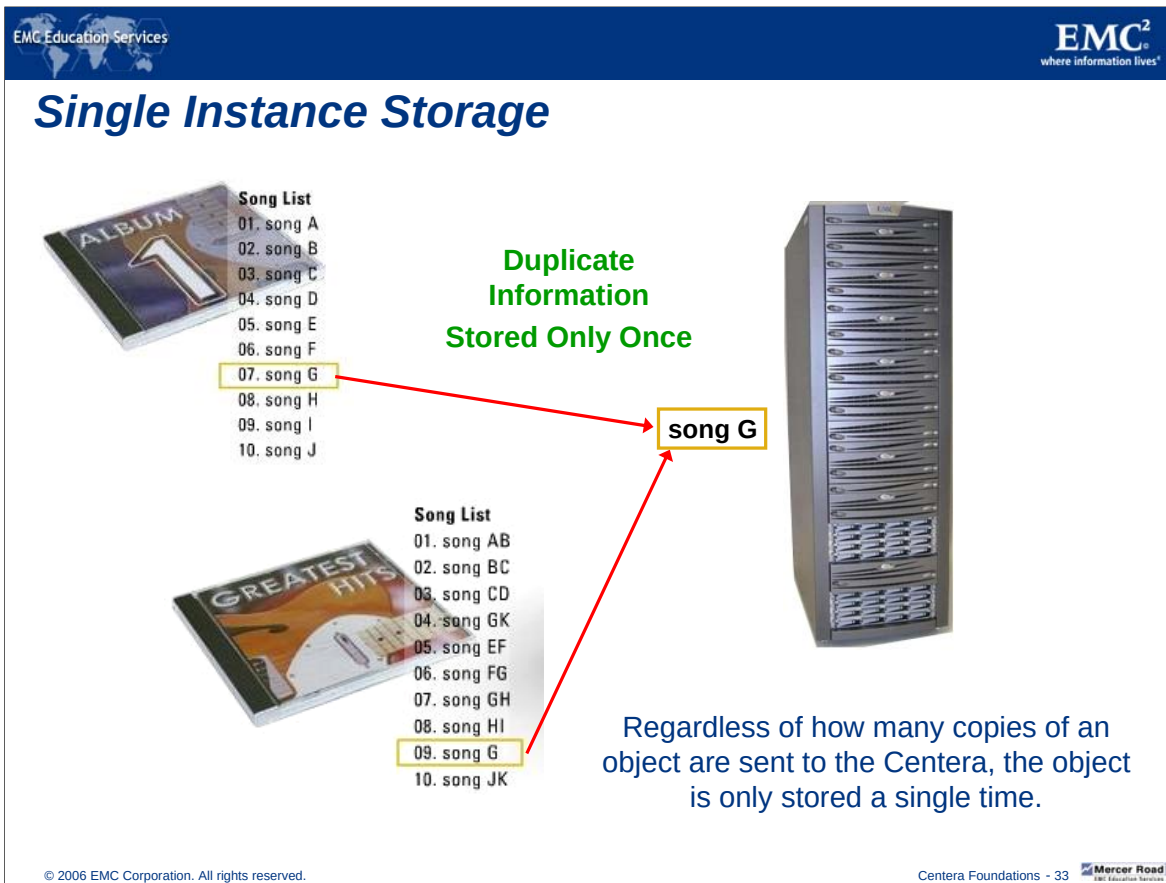
BUSINESS JUSTIFICATION



© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 32 

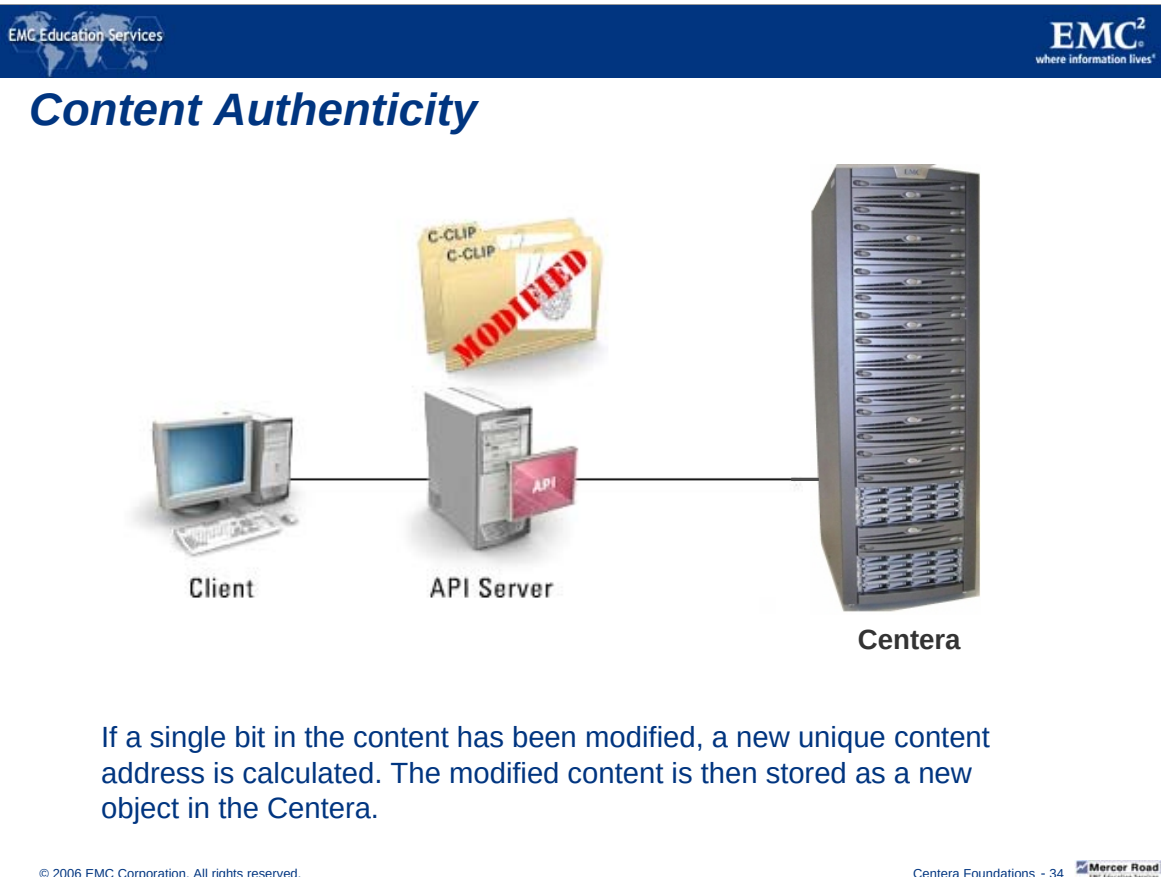
How the Centera and the concept of Content Addressed Storage assist businesses will be discussed in this section.



To address both the business need of minimizing data growth and the IT need to consolidate data, this concept of single instance storage is one of the key factors in the deployment of a Centera solution.

Rather than accessing a data object by its file name at a physical location, a CAS device uses a handle that is derived from each object's unique binary representation to store and retrieve the object. This is accomplished using breakthrough C-Clip technology, where subsequent access of the data object is made by simply giving the handle that uniquely identifies the object back to the repository. The data object is then returned. Content addressing greatly simplifies the storage resource management tasks, especially when handling hundreds of terabytes of static objects.

Also, this content-derived address is unique to ensure that only one protected (mirror or RAID) copy of the content is stored (single instance storage), no matter how many times applications store the same information. This significantly reduces the total number of copies of information stored, and is a key factor in lowering the cost of storing and managing content.



The Content Address is a digital fingerprint for the content and it never changes.

Because a Content Address is globally unique, it ensures data mobility. Content can move as needed without concern on the part of the user. Applications present Centera with a Content Address and get the specific content in return.

In addition, the Content Address assures content authenticity. A change to content generates a new copy of that content with a new, unique Content Address.

Identical objects are only stored once, dramatically improving storage efficiency by eliminating redundant copies of content. An object's Content Address is derived from the content itself. This results in no more than one instance of identical content stored in a cluster.

EMC Education Services

EMC²
where information lives[®]

Distributed Content for Business Continuity and Disaster Recovery

Pools and Replication

- Replicate selected Virtual Pools

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 35

Mercer Road
EMC Education Services

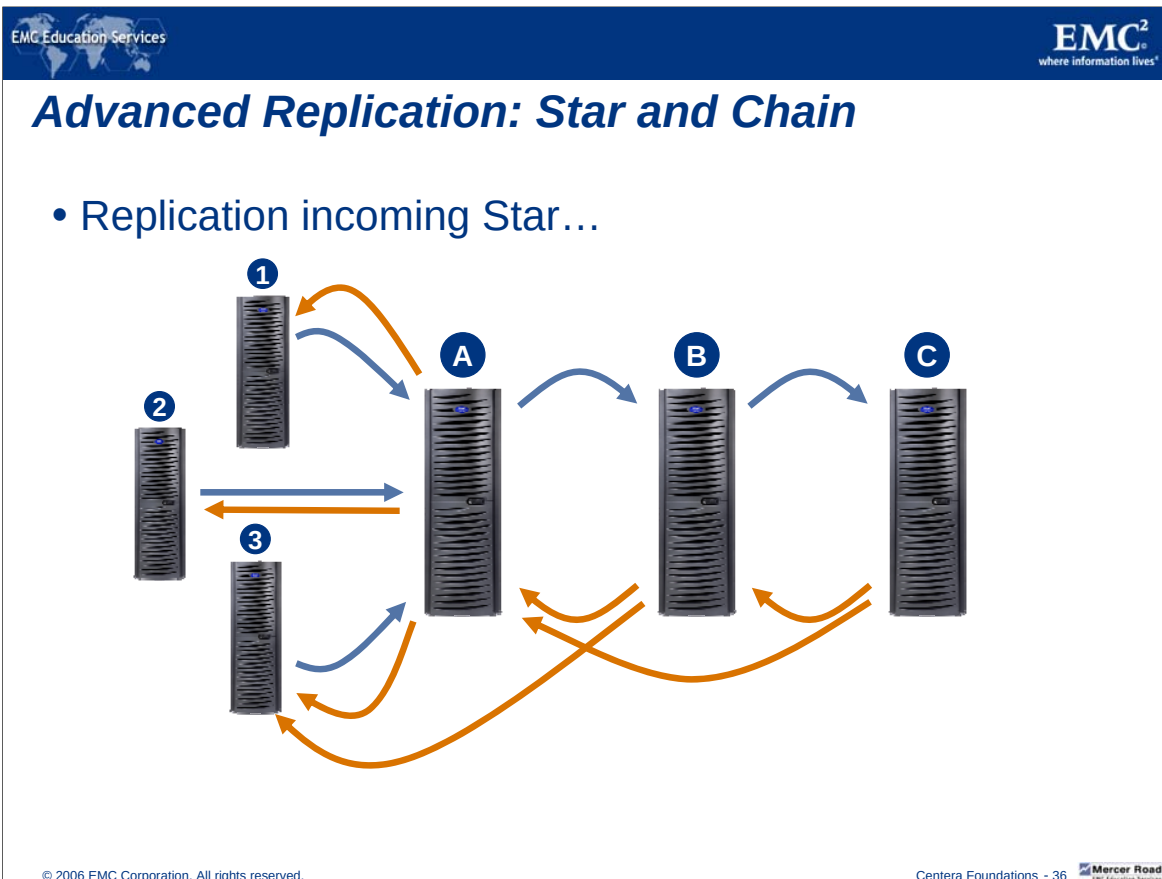
Available as a layered product is Centera's remote-replication software, Centera Remote Replication, which replicates content from a local repository to a remote repository.

When an object is initially stored in the local Centera, the object will be asynchronously and automatically replicated to the remote site over a wide-area network (WAN), which means the content will be stored both locally and remotely. If a piece of requested content cannot be found on the primary Centera via the Centera API, it will be sought on the remote Centera.

Centera replication can be implemented in three ways:

- Configured unidirectional
- Bidirectional
- Replicate Pools

New in Centera is Quiescent Mode, which enables Administrators to halt replication during busy times, then turn it back on during off-peak times.



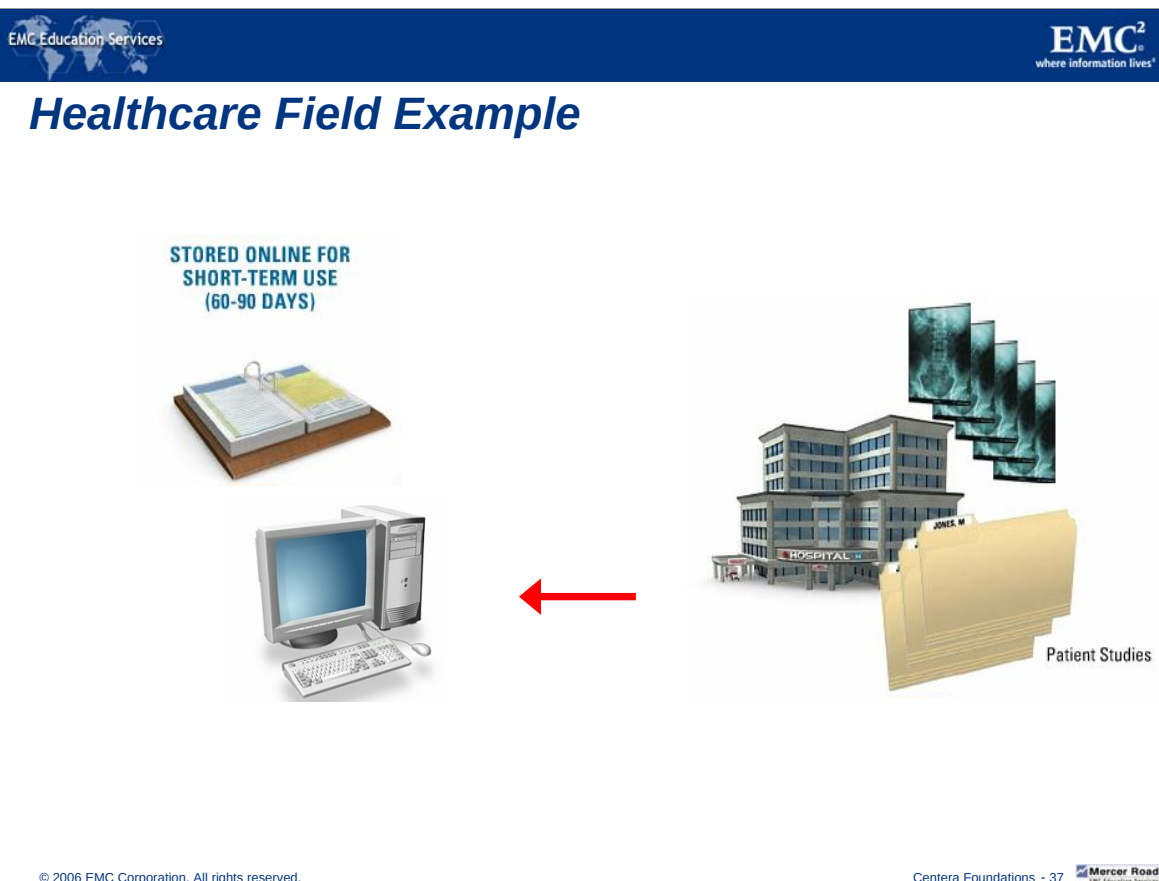
Centera has added advanced replication topologies to the existing unidirectional and bidirectional topologies:

- Star—allows up to three clusters to replicate to a single central Centera
- Chain—Replication to a third Centera (e.g., A to B to C) to provide additional redundancy

Long recognized as the industry-leading storage solution for fixed content and application-specific active archiving, Centera's Virtual Pooling and advanced replication topologies meet the needs of a new set of fixed-content users—specifically, those people and functions concerned about their organizations' "total" archiving needs and management. Whether it is the CIO, Legal Counsel/Department, Corporate Records Manager, or Archivist (a title found in Government entities), these functions are chartered with implementing a flexible organization-wide strategy to standardize policy management, address eDiscovery, simplify management, and constantly improve content protection.

Note:

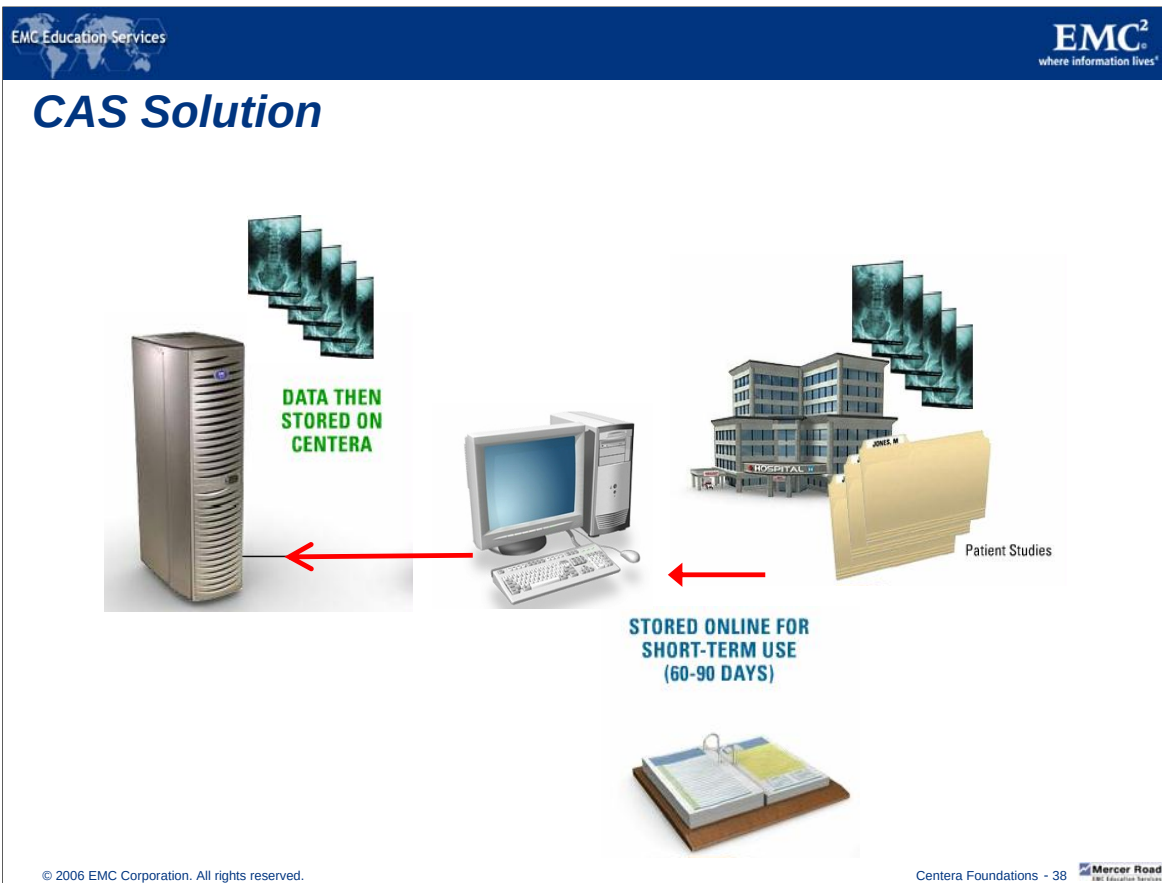
- Star-topology replication is not bi-directional. This would create an outgoing star-replication configuration that is not supported.
- Individual Virtual Pools within the same Centera cluster must replicate to the same target cluster.



Over 400 million patient studies were completed in the U.S. in 2000. Each study was composed of one or a series of images, which range in size from about 15 MB for standard digital X-rays to over 1GB for oncology studies. As X-rays are created in the radiology department or hospital, they are stored online for immediate use by attending physicians for a period of 60-90 days.

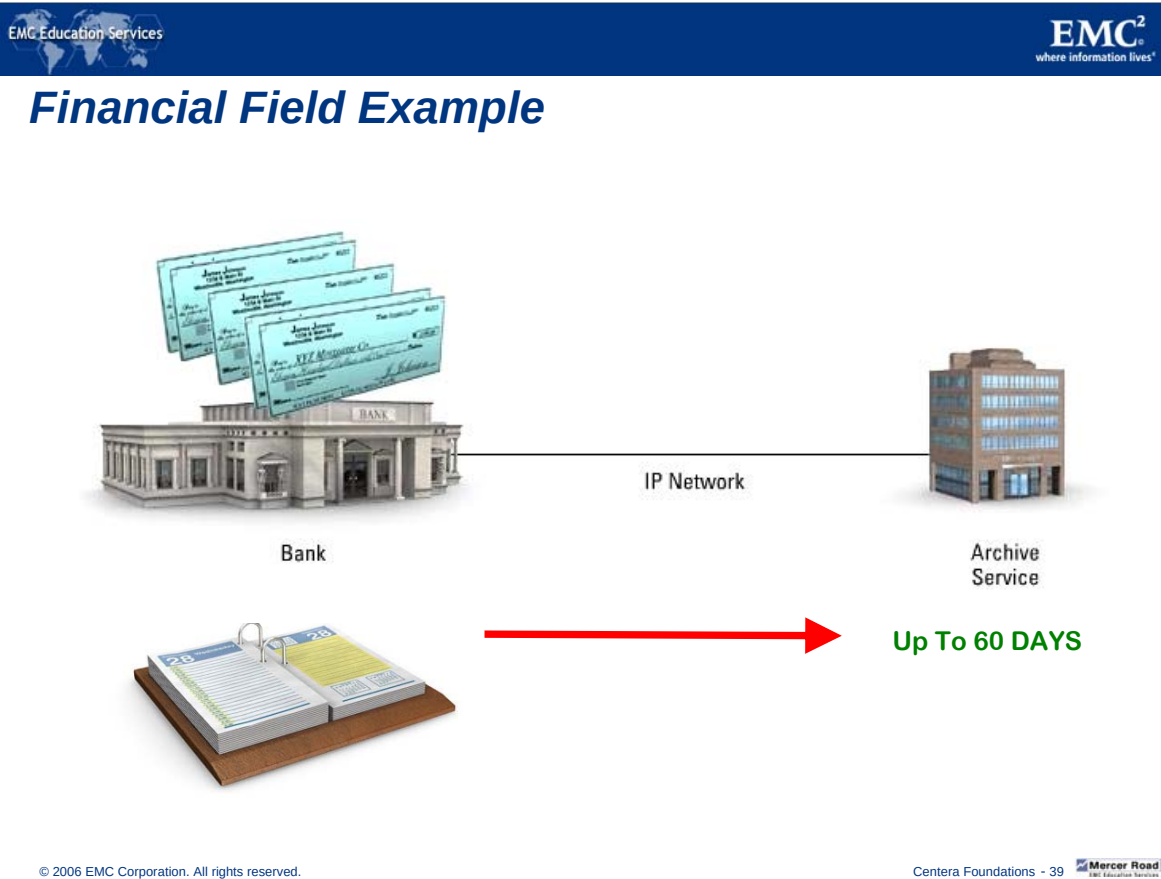
At the point where the patients are cured or discharged, the access needs for their particular X-ray drop off dramatically. However, HIPAA* requirements stipulate that these studies must be kept in their full, glossy image formats for a minimum of 7 years.

*HIPAA: Health Insurance Portability and Accountability Act of 1996

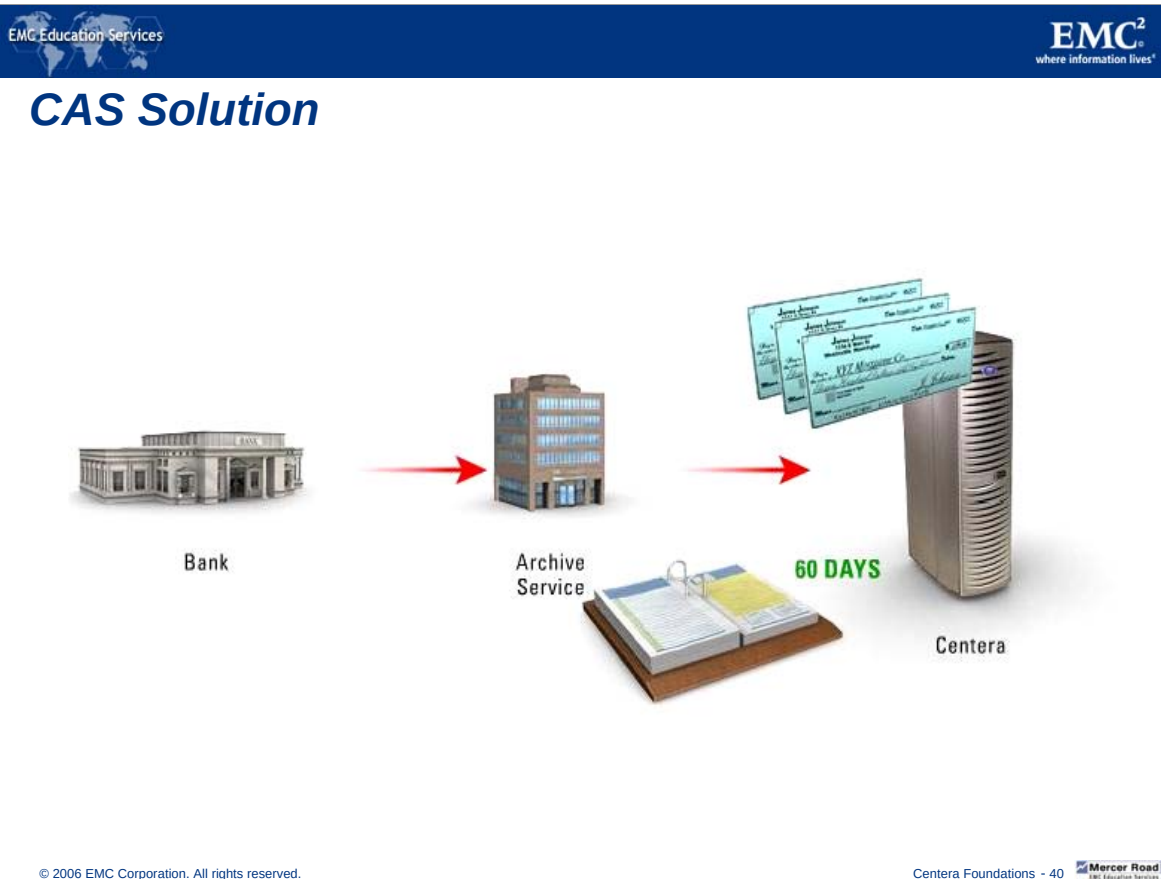


Beyond 90 days, hospitals may back up images to tape or send them to an offsite archive service for long-term retention. The cost of restoring or retrieving an image when in long-term storage could be 5-10 times more expensive than leaving the image online. Long-term storage can also involve extended recovery times of hours or days.

Medical image solution providers offer hospitals the capability of viewing medical studies, such as X-rays online, with sufficient response times and resolution to allow rapid assessment of patient situations. Centera is the optimum target storage device to facilitate long-term storage and immediate access of medical images online within a hospital or clinician's office.

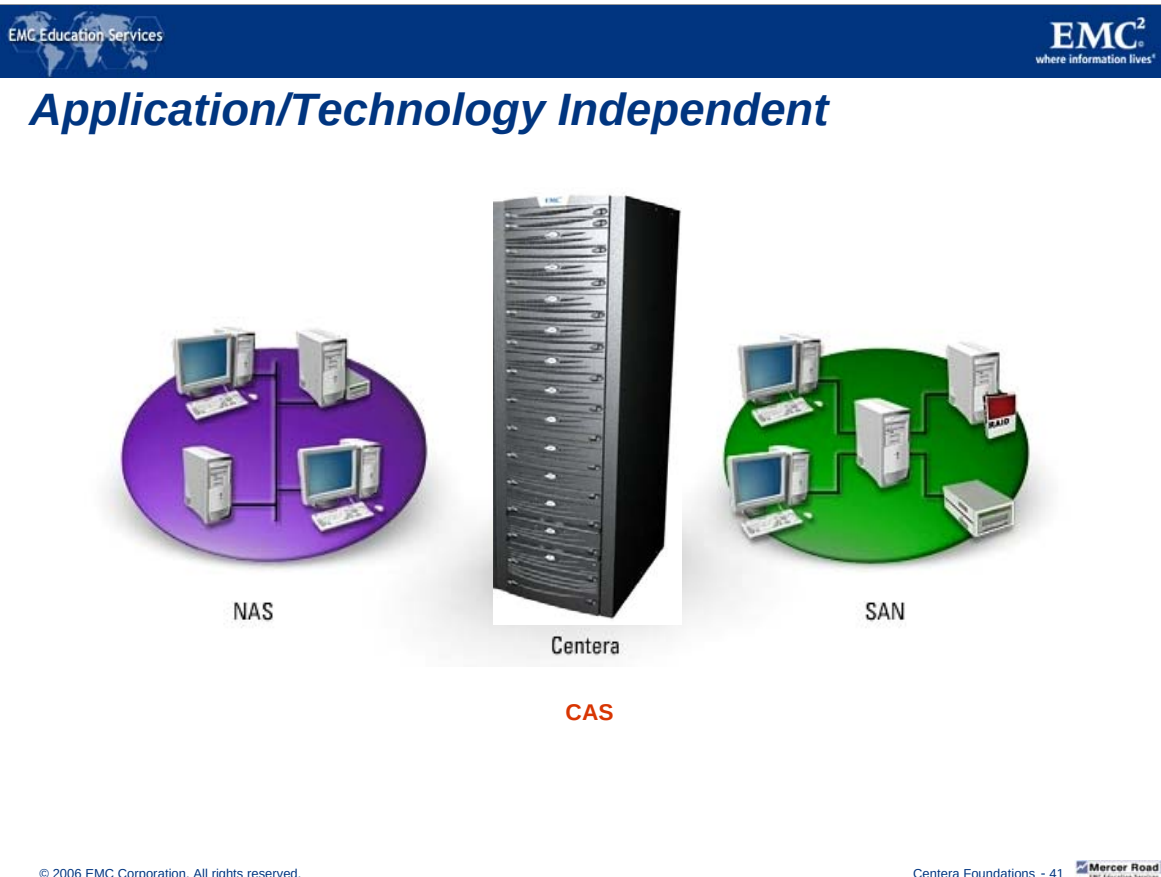


Check images, each with a capacity of about 25 KB, are created at the bank and sent to archive services over a standard IP network. A check imaging service provider may process 50-90 million check images a month. Typically, check images are actively processed in transactional systems for about 5 days.

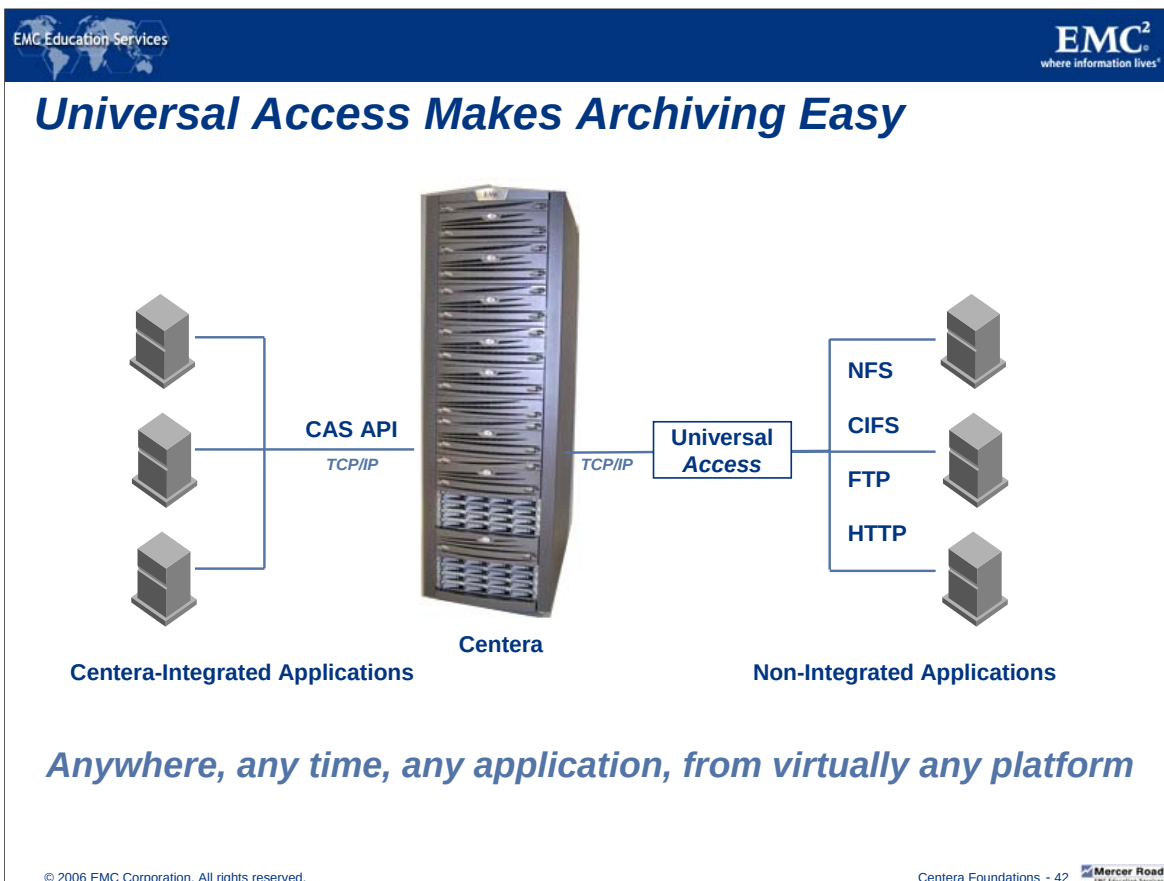


For the next 60 days, check images may be requested by regional banks or individual consumers for verification purposes, at a rate of about ½ percent of the total check pool (250,000-450,000 requests for check look-ups). Beyond 60 days, access requirements drop dramatically to as few as 1 for every 10,000 checks. In this case, the check images would be stored on Centera, starting at day 60 and held there indefinitely. A typical check image archive can approach 100 TB.

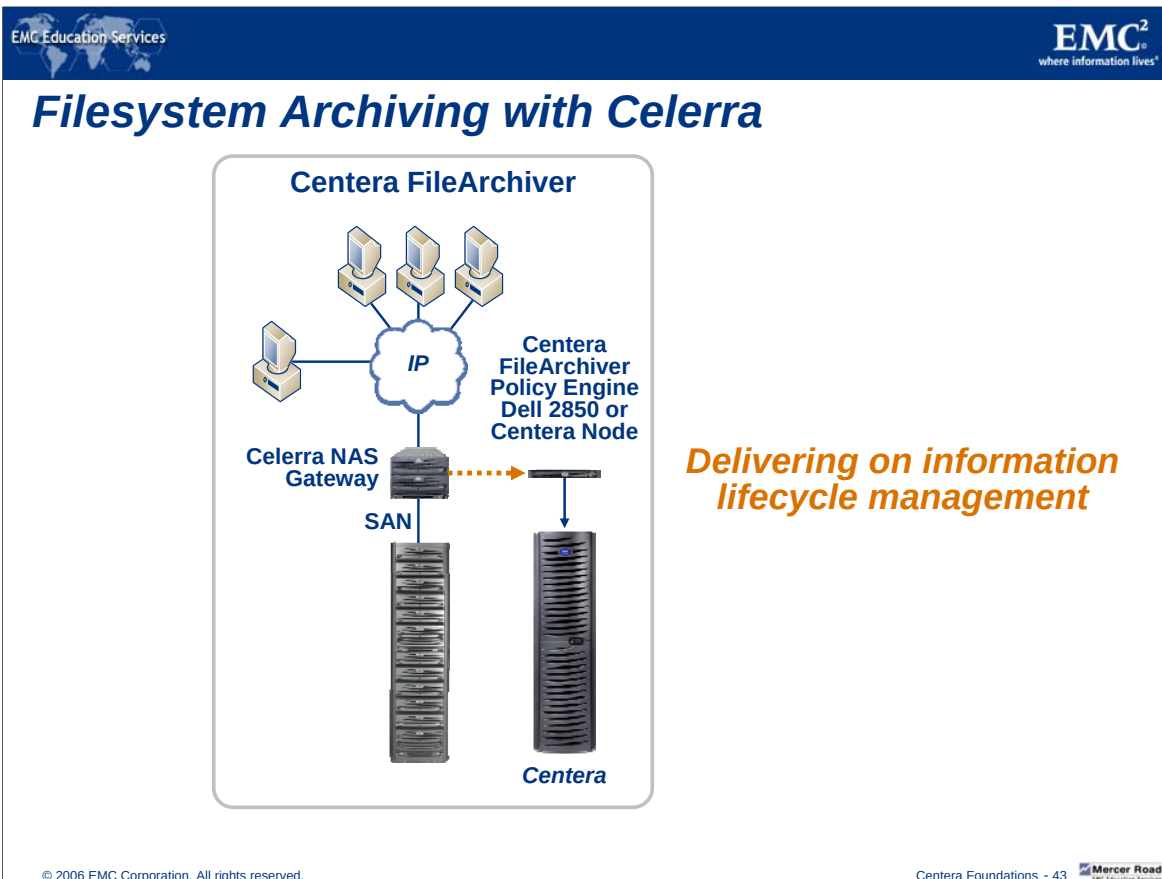
Check imaging is one of many financial service applications requiring the content storage facilities of Centera. Customer transactions initiated by e-mail, contracts, and security transaction records also need to be kept online for as long as 30 years.



Centera is easily installed and offers non-disruptive serviceability. The content addressed storage (CAS) technology allows for unique data representation and content authentication. In some applications, multiple customers concurrently accessing content may cause bottlenecks and access delays when using conventional solutions. Although these solutions may appear less expensive due to lower initial acquisition costs, in the long run, they cost more as they require increased focus on manual content management, such as movement to tapes and conversions to new formats. Centera alleviates the need to manage vast amounts of separated networked components and multiple file systems, caused by using low-end NAS or SAN alternatives to tape. Finally, Centera can be configured to maintain a replica of the fixed content at a remote site, eliminating the possibility of a site disaster destroying all copies of information.



EMC Centera™ Universal Access is a networked software appliance that enables enterprise applications to work with Centera using industry-standard file system protocols such as NFS (for UNIX and Linux applications) and CIFS (for Windows applications), plus Internet protocols such as FTP and HTTP. With the Centera Universal Access, any enterprise application that can mount a network drive or use FTP and HTTP can immediately capture many of Centera's unique benefits such as long-term retention and assured, long-term content integrity. From home-grown applications to nonintegrated versions of applications, Centera Universal Access makes it possible to utilize Centera in customer environments with no change to the existing application, greatly simplifying and accelerating deployment.

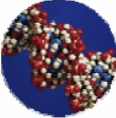


A key component of an information lifecycle management (ILM) strategy is the ability to move data between storage tiers based on its value over time. Businesses need to move data between storage tiers for a variety of reasons including, cost, protection, or compliance. The ability to create flexible automated policies to map data movement between tiers of storage against business requirements is a significant step forward in provide ILM functionality. Centera FileArchiver offers the benefits of enterprise-grade, policy-based archiving while complementing the existing infrastructure.

Centera FileArchiver, a policy engine integrated to the Celerra FileMover API, provides transparent file-level archiving between Celerra NAS and Centera CAS. It addresses the business need to manage information growth without storing large quantities of duplicate information



Centera Meets Regulatory Standards

	<u>Centera</u>
 • Financial Services – SEC Rule 17a-4	✓ Compliance Edition Plus, data shredding
 • Healthcare – HIPAA	✓ Content Addressing, mirroring, replication, data shredding
 • Life Sciences – 21 CFR Part 11	✓ Content authenticity, integrity checking
 • Government – DoD 5015.2	✓ Content Addressing, authenticity, data shredding

© 2006 EMC Corporation. All rights reserved.

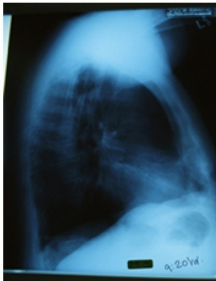
Centera Foundations - 44 

In all the regulated areas, the key elements of compliance are people, processes, and technologies. The mix and technology implications vary by regulation, and at our last count there were over 4,000 regulations across industries dealing with records authentication and retention. Here are only four of the many regulations that Centera addresses:

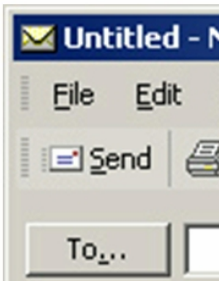
- SEC Rule 17a-4: The storage media requirements written into SEC (Security Exchange Commission) regulation 17a-4, viewed by many as one of the most stringent, if not the most stringent, regulated environments for information authenticity, retention, and protection. In Centera, data shredding and Compliance Edition Plus (CE+) handles these requirements with retention periods and data deletion restrictions as well as content authenticity.
- HIPAA: Centera complements the physical and administrative safeguards required by HIPAA, enhancing the protection and security of electronic patient information with content addressing, mirroring, replication and lifecycle integrity
- 21 CFR Part 11: Centera's digital fingerprinting and integrity checking of content throughout its retention provide functionality to detect altered, or compromised records beyond the grasp of any application, and is superior to that of conventional storage technology. With online access, Centera allows accurate and ready retrieval of all regulated records as required within Part 11.
- DoD 5015.2: Centera delivers additional levels of protection and security with Centera's Content Addressing, time/date stamping, and lifecycle integrity checking. Its Data Shredding feature exceeds the requirements of 5015.2 by ensuring privacy and eliminating liability.



A Few Integrated Applications and Partners



**Medical Imaging
Life Sciences**
GE Medical
Fuji Medical
AGFA/Mitra
Scientific SW
Amicas
Mediface
Merge



E-mail Archiving
KVS
Legato
Educom
IXOS
Entelagent
Assentor/Illumin
Tumbleweed
NICE Systems



**Enterprise
Content
Document
Management**
Documentum
FileNet
Tower Tech
daa Systemhaus
System Ware
Alliance
Windream

Plus many more partners for the following areas:

- Document / Check Imaging*
- HSM/FS Gateways*
- E-Learning*
- Oil and Gas*
- Audio Archiving*
- Backup/Archiving /Workflow*
- Media and Entertainment*

© 2006 EMC Corporation. All rights reserved.

Centera Foundations - 45 

Centera has hundreds of partners involved in its Centera Partner Program, taking advantage of Centera's free and open API. A small sampling of these partners is shown above.

Many of their integrated applications are already being used with the Centera by customers. Centera supports most OS platforms including Win2K/XP, Win2003, Linux, HP, AIX, IRIX, Solaris, and z/OS mainframe.

Module Summary

Key points covered in this training:

- Content Addressed Storage (CAS) is a new category of online disk storage designed specifically for fixed content, data that is in its final form
- Centera employs a unique storage/retrieval method called “content addressing” which is performed through the use of C-Clip technology
- Centera offers a number of benefits over that of other long-term backup solution such as:
 - Faster record retrieval
 - Single instance storage
 - Guaranteed content authenticity
 - Self-Healing
 - Meets regulatory standards

Key points covered in this module are shown here. Please take a moment to review them.