



# EMC Centera and the zSeries Mainframe: An Enterprise-Class Active Archiving Solution

Enterprise Computing &amp; Networking

by Stephanie Balaouras and Andy Efstathiou

May 2005

## Executive Summary

The mainframe is one of the most important computing platforms for mission-critical business applications. Mainframes are highly reliable and scalable. They can support a significant number of users and accommodate mixed workloads, high transaction volumes and large amounts of data. For these reasons, mainframes often run the business applications of Global 2000 companies, particularly in the finance, banking and insurance industries, and other compute-intensive applications. In recent years, there has been a resurgence of mainframe deployments due to the ability to host heterogeneous operating systems and applications as well as a return to centralized IT management.

Across mainframe and open systems, archiving is a critical data management function for several reasons, including:

- Reducing the costs of maintaining data on expensive, high-performance storage
- Achieving regulatory compliance
- Deriving business intelligence from historical data and transactions

Until recently, tape has been the preferred storage solution for mainframe archives because of its cost-effectiveness and long archive life. However, regulatory compliance and business intelligence mean that enterprises need to consider storage solutions that rapidly retrieve, aggregate and present information from an archive. When the information is available online, businesses can improve customer self-service, reduce costs of regulatory compliance and even turn the archive into a revenue generator. A storage solution needs to provide online access, ensure authenticity of the stored data, be highly scalable, but remain less expensive to acquire and manage than high-performance storage systems (see exhibit 1).

The EMC Centera achieves those goals. The EMC Centera is purpose-built for storing fixed or archived content and it:

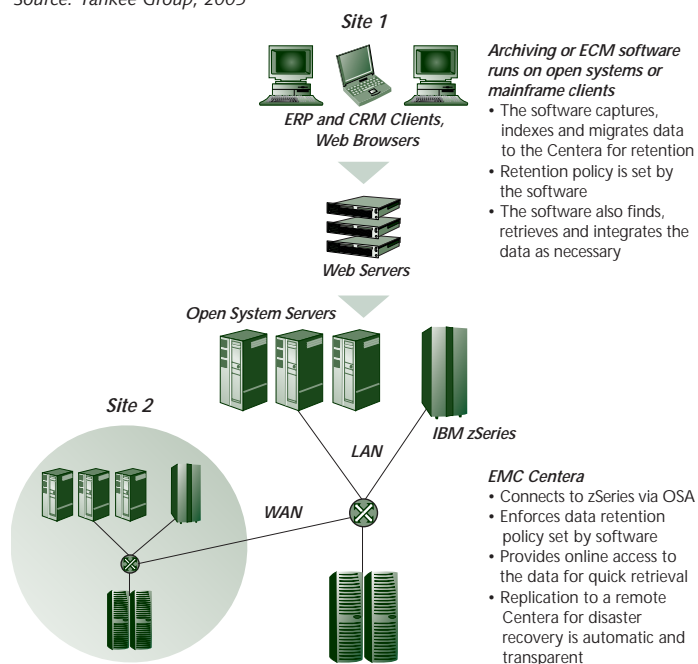
- Uses low-cost drive options and interfaces
- Is highly scalable
- Provides a mechanism to ensure long-term data authenticity and integrity
- Significantly reduces the complexity of managing a large-scale archive
- Has a native interface to the market leader in mainframe servers—the IBM zSeries

In this report, Yankee Group examines the value of EMC Centera as the primary archive for an IBM zSeries environment.

### Exhibit 1

#### EMC Centera Active Archiving Solution for the IBM zSeries

Source: Yankee Group, 2005



## Table of Contents

I. Introduction . . . . .	2
II. Mainframe Archiving Market Overview . . . . .	3
III. Product Background . . . . .	4
IBM zSeries . . . . .	4
EMC Centera . . . . .	5
IV. Solution Benefits . . . . .	6
V. Conclusions . . . . .	7
Recommendations for Enterprises . . . . .	7

## I. Introduction

The mainframe environment served by IBM's zSeries has grown and changed with the times. A data explosion in mainframes, coupled with changing user requirements for using the data, necessitates a re-evaluation of information management and storage solutions. This report identifies key information management and storage pain points in mainframe environments and assesses how well the EMC Centera addresses these pain points relative to other storage systems and storage mediums (such as tape and optical disk) for IBM's zSeries servers. We evaluate the following criteria:

- **Data authenticity and integrity:** The ability to ensure the integrity of the data for the purposes of regulatory compliance or corporate policy. This requires a write-once-read-many (WORM) capability or equivalent.
- **Data accessibility:** To meet regulatory requirements, legal discovery and corporate policy for data accessibility, access to data must be fast.
- **Scalability:** The ability to scale a single managed archive repository to multi-terabytes without significant complexity.
- **Archive life:** The expected archive life of the storage medium should meet or exceed the data retention requirements of the government regulation or corporate policy, or the business will need to implement a tiered-storage strategy.

- **Management simplicity:** The archive should be simple and require the same amount of labor resources to configure, manage and troubleshoot as it scales.
- **Data protection capabilities:** The data in the archive should be mirrored or parity protected and either be portable (as in the case of tape) or have data mirroring/replication capabilities for disaster recovery (as in the case of disk storage).
- **Partner ecosystem:** The archive should have product integration with the following independent software vendors: backup vendors, archive vendors and enterprise content management vendors.
- **Environmental:** How much floor space does the storage solution occupy in the data center? Are there special cooling and electrical requirements?

This report also identifies other opportunities afforded by the Centera solution such as online access to mainframe information, improved customer and partner service, and potential revenue-generating opportunities.

## II. Mainframe Archiving Market Overview

Mainframe servers such as IBM's zSeries are at the heart of mission-critical applications at global enterprises. Data management is crucial for this environment and includes the following functions and tasks:

backup/restore, data replication, archiving, hierarchical storage management (HSM) and data mobility. In the Yankee Group *2004 U.S. Enterprise Storage Survey* of enterprise businesses, backup and data retention (archiving) ranked as the top priorities. Archiving has become a priority for businesses because of several important trends:

- **Regulatory compliance:** Compliance is driving an increased demand for archiving solutions because the effectiveness of compliance initiatives can only be measured and tested by retrieving and analyzing archived data. Key performance differentiators for archiving solutions that are optimized for compliance are data search, speed of retrieval and security of stored data.
- **Digital content explosion:** The digitization of content such as reports, statements, office documents, e-mail, check and document images, medical information (patient records and medical imaging), media (audio, video and music files), drug discovery data and software source code has created enormous capacity requirements for reference data.
- **Business intelligence:** Enterprises are trying to improve business performance through a better understanding of business processes and constituent behavior. As adoption of analytical solutions has increased, the need to retain ever-larger amounts of increasingly granular data for analysis has grown.

Historically, mainframes have been considered highly proprietary and new deployments were either stagnant or diminishing in favor of open system, distributed servers. This limited the number of secondary storage options that were developed for the mainframe market. In fact, until recent years, secondary storage options for mainframe servers were limited to proprietary tape offerings from IBM (3590 and 3490E tape formats) and StorageTek (9840 and 9940 tape formats). These tape formats are offered at a premium versus tape formats for the open systems market such as LTO and DLT/SDLT offered by numerous vendors.

Recent Yankee Group research, primary interviews and surveys suggest that mainframe server environments are characterized by the following:

- Large data archives with significant expected growth rates. About 55% of mainframe environments will have between 1 TB and 30 TB of archived data; 45% will have more than 30 TB. About 55% will experience archive growth rates of between 1% and 20%; 45% will experience more than 20%.
- Archive manageability, reliability and availability are overwhelming issues.
- Tape is the most widely used storage medium for archiving and HSM, but disk utilization is sizable and growing. Approximately 55% of respondents utilize tape, 26% utilize disk and 19% utilize optical. In the coming year, 44% will invest in disk, 33% in tape and 22% in optical.
- Most businesses regard disk as the best for ease-of-use and data access, optical disk as the best for archival life, and tape as the best for total cost of ownership despite the significant price reductions in disk with the advent of low-cost disk options such as ATA. The tape library ecosystem requires a significant number of labor resources to manage and occupies a large physical space in the data center. Consider:
  - On average, as many as 12 individuals are needed to manage the tape archive. As the archive grows, one additional full-time labor resource per year will likely be needed.
  - The number of tape cartridges in the environment is about 1,800 and growing by about 5% per year.
  - Tape archives currently consume about 60 floor tiles and will likely require an increase of 5% to 10% per year.

### III. Product Background

This section briefly describes the product offerings of two market leaders: the IBM zSeries mainframe server and the EMC Centera, a purpose-built storage system for storing fixed or archived data.

#### IBM zSeries

IBM's zSeries is the dominant mainframe in the marketplace with clear thought leadership in terms of development. Until recently, many industry veterans were anticipating the demise of the mainframe. Investments in the past 5 years by IBM have not only kept the zSeries relevant, but have increased its appeal and relevance in the marketplace.

IBM has made key improvements to the zSeries:

- **Increased storage options:** Primary storage options for the zSeries over ESCON or Fibre Channel over ESCON (FICON) are available from all the leading storage vendors.
- **Improved management software:** IBM has improved the software stack around the mainframe to enhance manageability. For example, in the virtualization stack, z/VM 5.1 was introduced in September 2004 with a price that is half of V4.
- **Lower entry point pricing:** IBM introduced the new z890 in April 2004, providing a new, lower entry point for mainframe customers.

Key application workloads on the mainframe are mission-critical, transaction-intensive applications (see Exhibit 2). Because of these workloads, the mainframe produces large amounts of data that must be stored and periodically retrieved. For example, in banking the deposit system processes many transactions to each account. All of these transactions must be stored and readily accessible in the event there is a question regarding the final balance. Many different constituencies, such as customers, regulators, counterparties and internal auditors, have an interest in reconstructing transaction histories. Most mainframe applications show the same characteristic of requiring accurate recovery of transaction detail for extended periods of time.

This requirement necessitates fixed content storage that is easily searchable and easily recoverable. The increasing levels of regulation, litigation and transaction feature capture are driving increased demand for storage that is secure, rapidly searchable and retrievable.

#### Exhibit 2

##### Key Mainframe Application Workloads Drive Growth in Data Archiving

Source: Yankee Group, 2005

Vertical	Key Mainframe Application Workloads	Archiving Growth Drivers
<b>Financial Services</b>	Transaction processing (loans, deposits, securities trading, accounting), financial accounting, regulatory reporting and billing	Compliance (NASD/SEC), risk management, online trading, customer billing applications, CRM, financial accounting applications and decision support
<b>Manufacturing</b>	ERP, SCM and financial reporting	Compliance (export regulations/Sarbanes-Oxley Act), supply chain management, online inventory control, financial accounting applications and decision support
<b>Transportation</b>	SCM, financial reporting, shipping and inventory management	Compliance (safety), shipping/inventory applications, financial accounting applications and decision support
<b>Government</b>	Database management and legacy applications (e.g., judicial, regulatory, taxation, military and intelligence)	Compliance (US Homeland Security) and data warehousing
<b>Telco</b>	BOSS, billing, financial reporting and decision support	Compliance (FCC), CRM, online billing applications, financial accounting applications and decision support/data warehousing
<b>Retail</b>	SCM, financial reporting and inventory management	Financial accounting and decision support
<b>Oil and Gas</b>	Exploration, inventory management, SCM and financial reporting	Compliance, online inventory applications and financial accounting applications
<b>Biotechnology</b>	Research, billing, SCM, ERP and financial reporting	Compliance (FDA/HIPAA), clinical research and decision support
<b>Utilities</b>	Billing, SCM and financial reporting	Compliance, financial accounting and online billing
<b>Healthcare</b>	Billing, patient records processing and financial accounting	Compliance (HIPAA), billing and customer service applications, patient records and financial accounting applications

## EMC Centera

The EMC Centera content addressed storage (CAS) system is a bundled software and disk platform that targets the storage requirements of fixed content (storing hundreds of terabytes of unchanging content for long periods and having that content readily available online).

The Centera storage system includes a redundant array of independent nodes (RAIN) interconnected via a private TCP/IP local-area network (LAN) in a cabinet. Each node executes an instance of the Centera software and has processing power and storage capacity. Each node acts as either an access node or storage node. Access nodes are the conduits between the applications and the storage nodes. Applications, such as content management applications, are written or customized to integrate with the Centera API. Applications send requests to store and retrieve objects over the TCP/IP LAN to the Centera access nodes.

- **Integrity and long-term retention:** Centera ensures the authenticity of the object to be stored. When an application sends a request to Centera to store an object, the Centera software calculates a content address for that object based on its attributes and then stores the object in the Centera repository. The content address is globally unique and acts like a digital fingerprint for the object. This ensures that the object is stored uniquely and is not modified. A change to the object generates a new copy of the object with a new unique content address. Authenticity is crucial for business requirements but also for compliance requirements for SEC 17a-4, HIPAA, Sarbanes-Oxley Act and others.
- **Scalability:** Content addressing reduces storage complexity. Because Centera determines the physical placement of objects, system administrators don't have to specify and manage the underlying file or logical unit number structure. Additional nodes can be added as capacity is needed and Centera self-configures the additional capacity. Each cabinet can accommodate up to 32 nodes and 8 cabinets can be clustered together. An application can access multiple Centera clusters (up to 1 petabyte) through the Centera API.

- **Data protection:** Centera offers two options for local content protection: content mirroring protection (CMP) and content parity protection (CPP). The user selects the most appropriate option based on the level of protection required. CMP mirrors the object on two different storage nodes within the Centera cluster. Centera ensures that the object is mirrored on the nodes that are the least busy, have available capacity and are on different power rails. CMP uses 50% of the raw storage capacity. CPP separates an object into six separate data segments plus one parity segment and stores the segments across seven different storage nodes. If there is a device failure, CPP can recreate the object using the remaining segments. CPP uses 75% of the raw storage capacity.
- **Disaster recovery:** Centera offers functionality to support disaster recovery. When an object is stored in a local Centera, the object can automatically be mirrored asynchronously to a remote Centera over the wide-area network (WAN). The process is transparent to users once it is implemented and read failover is automated.
- **Manageability:** Centera requires minimal manual administration. As discussed in the preceding paragraph, Centera auto-detects and self-configures capacity. It is also self-managing and self-healing. Centera monitors and auto-detects any failures of any component (drive, node, switch, power rail, etc.) and will map around the issue. It monitors and auto-detects software faults and restores any processes. It constantly verifies content integrity using multiple schemes and techniques.
- **Online accessibility:** Unlike tape or optical storage media, Centera provides fast, online access to data via the TCP/IP LAN.
- **Integration:** Centera provides an API so that independent software developers (such as database archiving and enterprise content management software vendors) can integrate their solutions with Centera.
- **zSeries mainframe support and offerings:** Centera offers several connectivity options and integrated partner offerings for the zSeries mainframe server:

- **Connection through an emulation gateway:** zSeries mainframes can connect using ESCON or FICON to the IP-based Centera using an emulation gateway such as Bus-Tech's Mainframe Appliance for Storage (MAS). MAS connects to the zSeries using two or three ESCON or two FICON channels and to Centera using a GigE connection. It emulates up to 64 IBM 3480/3490 tape drives and is OS and application independent. With this approach, businesses don't have to change any of their existing processes or procedures to integrate Centera into the environment.
- **Native connection through Centera API:** Centera provides a native API to the IBM zSeries server. Centera communicates directly over the TCP/IP LAN to the zSeries through the IBM Open Systems Adapter (OSA). OSA provides direct connection for the zSeries to high-speed routers, servers and clients on the LAN. Thus, no external hardware communication is required for integration.
- **Partner ecosystem:** Centera has partner integration with several mainframe solution providers: Mobius, Systemware, BMC Software, Computer Associates and RSD for output management/enterprise content management; Bus-Tech, Intercom and Luminex for emulation gateways; Princeton Softech for database archiving.

#### IV. Solution Benefits

In Section I of this report, Yankee Group outlined key criteria for a mainframe archive. The EMC Centera satisfies all these criteria. Compared to other common options for mainframe archiving, such as tape and optical disk, it has several advantages that can dramatically reduce complexity, provide fast access to data and improve the total cost of ownership:

- **Simplicity and scalability:** Up to a petabyte of storage capacity can be managed in a single Centera instance; and as capacity is added, it is self-configured. In addition, due to its location-based addressing, Centera requires little active storage management (there is no requirement for a storage administrator to determine the physical layout of data on the storage). Its self-healing

capabilities reduce the need for ongoing monitoring of content integrity and availability. This is a significant reduction in management complexity and labor costs relative to the need to manage potentially thousands of tapes (both onsite and offsite) and several libraries. It's estimated that a single administrator can manage several hundred terabytes of Centera capacity.

- **Online access to data:** Centera provides online access to archived data and when integrated with solutions from database archiving and enterprise content management (ECM) software vendors. These joint solutions provide rapid recovery of data for regulatory compliance and legal discovery as well as presentment of the data to other applications in support of decision support systems and customer service applications. In addition, these solutions can help businesses implement broader information lifecycle management strategies. Partners include Mobius, Systemware, BMC Software, Computer Associates and RSD, and Princeton Softech for database archiving.
- **zSeries support offerings:** Centera can integrate into the zSeries environment with TCP/IP connections by using IBM's OSA hardware or with ESCON or FICON connections using an emulation gateway. The choice is ultimately up to the user. The ability to integrate natively with the zSeries is a distinct advantage over other low-cost or purpose-built disk storage systems. For businesses utilizing tape in an archive or enterprise content management solution, the emulation gateway allows the business to substitute Centera without disrupting existing software configurations, processes and procedures.
- **Advanced data protection and disaster recovery:** Centera offers the ability to mirror or parity archived data automatically for local content protection from device failures. For disaster recovery, Centera can automatically mirror data from one Centera to another remote Centera. This eliminates the need to duplicate tapes (duplication is a time-consuming batch process) or to manually transport tapes to a disaster recovery location. It not only provides a higher level of data availability, but also saves a significant amount of time and cost. In addition, it eliminates the potential of lost, stolen or damaged tapes. For example, in February 2005, a large financial institution lost several tapes containing

sensitive customer data of federal employees while transporting the tapes to a disaster recovery location.

- **Heterogeneous host support:** Centera can simultaneously support open systems and zSeries hosts. This enables businesses to use the Centera as the single archive repository for applications across the enterprise. This is in contrast to tape, where open systems and mainframes require different tape formats. Therefore, separate cartridges, drives and libraries are required in support of each.

## V. Conclusions

**M**ainframe customers are struggling to store, manage and access terabytes of archived data effectively to meet business and regulatory demands. In the past, tape was the only cost-effective storage medium available to store such huge amounts of data. However, data stored on tape cannot be accessed fast enough to give business and end users online access to critical information. In addition, as tape scales, so does its management complexity—thus negating many of the cost savings in the initial acquisition price. Mainframe customers require a storage solution that provides online access but is still cost-effective for terabytes of data, requires minimal active storage management and can ensure the authentication of the data.

Mainframe environments are typically processing- and transaction-intensive and mission-critical. As a result, mainframe managers are risk-averse when considering changes in their environment. Mainframes have typically addressed the archiving challenge by storing data on tape. Tape is non-intrusive to the operating environment but costly to manage and difficult to access, even with virtual tape solutions. Mainframe managers need to actively consider alternatives to tape.

EMC Centera's content addressing system scales to petabytes and requires minimal storage management administration because it needs no device, volume or file system management and ensures data authenticity. Centera's cost-effectiveness enables companies to replace tape and provide online access to data that previously had been inaccessible on tape. In addition to helping meet regulatory

compliance, Centera's integration with enterprise content management vendors and online access to archived data improves customer service and creates new revenue-generating opportunities.

## Recommendations for Enterprises

- **Mainframe administrators need to consider the benefits of disk-based storage.** Customers with existing tape libraries need to carefully compare the total cost of ownership (TCO) of continuing their current tape solution with the TCO of converting to a disk-based storage system. Low-cost disk, scalability, manageability and the advanced data protection options of a disk-based storage system such as the EMC Centera can be a more cost-effective alternative. It's also important to understand the value of having data online and readily available for compliance and legal audits, as well as to enhance business intelligence applications or improve customer service.
- **Consider disk-based storage before upgrading tape systems to the latest generation.** Tape system vendors introduce new generations of cartridges, drives and libraries every 1.5 to 2 years. Generally, new generations are read/write-compatible with the previous generation and only read-compatible with the previous two generations. Typically during this cycle, a vendor will stop offering the previous product generation, essentially forcing businesses to upgrade. It's important to understand the long-term viability of your current generation of tape technology. If considering an upgrade of your technology, this represents another opportunity to consider a disk-based solution.
- **Develop an information lifecycle management strategy.** Archiving is just one component of an overall information lifecycle management strategy. Understanding the value of corporate data and determining effective ways to manage it will be increasingly strategic for businesses. This means working closely with content management and storage vendors that can manage all elements of the information lifecycle as well as establishing content value. Yankee Group believes information lifecycle management will be a core issue for IT executives in the next 2 years.

## Yankee Group

Yankee Group has research and sales staff located in North America, Europe, the Middle East, Africa, Latin America and Asia-Pacific. For more information, please contact one of the sales offices listed below.

### Corporate Headquarters

31 St. James Avenue  
**BOSTON, MASSACHUSETTS 02116-4114**  
T 617.956.5000  
F 617.956.5005  
info@yankeegroup.com

### EMEA

55 Russell Square  
**LONDON WC1B 4HP**  
**UNITED KINGDOM**  
T 44.20.7307.1050  
F 44.20.7323.3747  
euroinfo@yankeegroup.com

### North America

260 Terence Matthews Crescent, Suite 200  
**KANATA, ONTARIO, CANADA K2M 2C7**  
T 613.591.0087  
F 613.591.0035  
canadainfo@yankeegroup.com

### Decision Services

Yankee Group Decision Service annual memberships offer clients access to research and one-to-one expert guidance.

Decision Services represent our best value for clients. The services help our members understand industry, regulatory, competitive and market-demand influences, as well as opportunities and risks to their current strategies.

Membership includes an invaluable in-person strategy session with Yankee Group analysts, direct access to a team of analysts, DecisionNotes<sup>sm</sup> and regular Online Decision Forums on relevant topics.

We offer Decision Services on almost 30 selected topics in Telecommunications; Wireless/Mobile Communications; Consumers, Media & Entertainment; and Information Technology Hardware, Software & Services.

### Decision Instruments

Yankee Group offers a full portfolio of technology and market forecasts, trackers, surveys, and total cost of ownership (TCO), return on investment (ROI), selection and migration tools. Decision Instruments provide our clients the data required to compare, evaluate or justify strategic and tactical decisions—a hands-on perspective of yesterday, today and tomorrow—shaped and delivered through original research, in-depth market knowledge and the unparalleled insight of a Yankee Group analyst.

#### Trackers

Trackers enable accurate, up-to-date tactical comparison and strategic analysis of industry-specific metrics. This detailed and highly segmented tool provides discrete proprietary and performance data, as well as blended metrics interpreted and normalized by Yankee Group analysts.

#### Surveys

Surveys take the pulse of current attitudes, preferences and practices across the marketplace, including supply, delivery and demand. These powerful tools enable clients to understand their target customers, technology demand and shifting market dynamics.

#### Forecasts

Forecasts provide a basis for sound business planning. These market indicators are a distillation of continuing Yankee Group research, interpreted by our analysts and delivered from the pragmatic stance our clients have trusted for decades.

### Signature Events

Yankee Group's Signature Events provide a real-time opportunity to connect with the technologies, companies and visionaries that are transforming Telecommunications; Wireless/Mobile Communications; Consumers, Media & Entertainment; and Information Technology Hardware, Software & Services.

Our exclusive interactive forums are the ideal setting for Yankee Group analysts and other industry leaders to discuss and define the future of conversable technologies, business models and strategies.

### Consulting Services

Yankee Group's integrated model blends quantitative research, qualitative analysis and consulting. This approach maximizes the value of our solution and the return on our clients' consulting investment.

Each consulting project defines and follows research objectives, methodology, desired deliverables and project schedule. Many Yankee Group clients combine Decision Service memberships with a custom-consulting project, enabling them to augment our ongoing research with proprietary studies.

Thousands of clients across the globe have engaged Yankee Group for consulting services to hone their corporate strategies and maximize overall return.

### www.yankeegroup.com

Yankee Group believes the statements contained in this publication are based on accurate and reliable information. However, because our information is provided from various sources, including third parties, we cannot warrant that this publication is complete and error-free. Yankee Group disclaims all implied warranties, including, without limitation, warranties of merchantability or fitness for a particular purpose. Yankee Group shall have no liability for any direct, incidental, special, or consequential damages or lost profits. This publication was prepared by Yankee Group for use by our clients.