



# Hitachi's New Generation Active Archive Platform

## *Separating Physical Storage from Logical Archiving Services*

### Executive Summary

---

Historically, archivists and document managers have struggled to demonstrate substantial technology payback, in large part due to the unreliability and expense of managing microfilm and DVD jukeboxes. Recently, the advent of new disk storage technologies has meant that more reliable and cost effective solutions can be provided using standard IT components. However, the IT industry continues to struggle to fully understand archiving requirements.

The vision of “Information Lifecycle Management” (ILM) is to ensure that all documents are archived automatically, can be accessed and deleted as required, and are placed on the most sensible medium, given specific performance requirements. Although the IT industry has delivered some standalone systems that have solved certain ILM problems, users have found in practice that those have been limited in scope, inflexible, and difficult to integrate.

ITCentrix believes the approach taken by Hitachi, while less grand, is much more practical and likely to succeed. It focuses on providing archiving services independent of storage and allows applications to easily use these services to exploit information.

This paper explores in depth the merits of this new approach and provides a case example comparing digital archiving solutions with traditional methods of archiving using jukebox technologies. Moreover, the study compares Hitachi's Active Archive Platform with competitive best-of-breed digital solutions in the marketplace today.

Our research focuses on an 11TB hospital example and demonstrates that Hitachi's approach delivers a net present value increase of nearly \$3.7 million, relative to jukebox technologies. Moreover, in this example Hitachi's new platform has the potential to deliver more than twice the value of the traditional digital archiving systems in the marketplace today.

ITCentrix strongly encourages archivists, document managers, independent software developers (ISVs), and IT executives to evaluate this innovative approach to active archiving.

## TABLE OF CONTENTS

Executive Summary .....	1
Introduction .....	3
Archiving Fundamentals .....	4
Hitachi Content Archive Platform Architecture.....	6
Case Study .....	8
Conclusions and Recommendations.....	11

---

*Disclaimer: The software, tools and methods used to create this analysis and the data contained herein do not have any express or implied warranty. ITCentrix provides sophisticated business analytics intended to provide as complete and current an assessment as commercially reasonable based on the inputs provided to its tools. The inputs for this analysis are believed to be accurate and current as of June 2006. There are many factors that can alter actual customer investment results, including staff experience, staff turnover, technology changes, cost overruns, competitive responses, and changes in organizational structure and strategy. Each of these, as well as other factors, can impact the performance of an investment. ITCentrix therefore makes no express or implied warranties or representations with respect to the information contained within this report. In addition, ITCentrix makes no representation regarding the accuracy or completeness of information provided in this report. Users of this information will indemnify and hold harmless ITCentrix for any error or omission.*

# Introduction

Figure 1 details the architecture that Hitachi has developed. Hitachi provides both the storage hardware and storage infrastructure architecture layers, while the archiving services layer is powered by Archivas. This platform can be exploited by both ISVs and in-house applications.

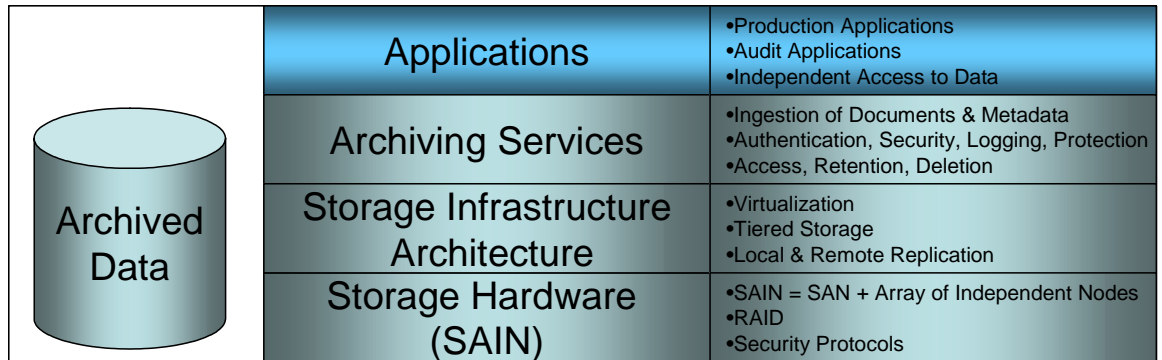


Figure 1 - Hitachi Content Archive Platform (HCAP)

The advantages of this architecture are as follows:

- It separates the storage hardware and allows storage to be added and managed independently of other layers. This allows the optimization of storage cost and functionality within each layer.
- It separates storage services, such as virtualization, tiered storage, and replication from the hardware. This allows the usage of best-of-breed storage management software and function to be utilized and enhanced independently of other layers.
- The archiving services are powered by Archivas, allowing Hitachi to provide the services an archivist or document manager would expect, such as ingestion of documents, authentication, protection, security, logging, access control, retention control, and deletion.
- Hitachi is sending a clear message that it is not competing in the application layer. Hitachi is focused on making the platform the easiest to use by application developers. The interface with the data is clear. The metadata and data can be directly available to an application, which means new applications look across all data and are not constrained to a “stovepipe” view.

The case study in this paper illustrates that the HCAP solution is much more cost effective than jukeboxes, with a breakeven of eight months and a Net Present Value (NPV) of \$3.6 million. Figure 2 gives a summary of the business case for 11 terabytes of HCAP archiving storage compared to a jukebox and traditional solutions.

<b>Comparison of Installed Jukebox vs. Digital Archiving</b>	<b>Traditional Digital Archiving System</b>	<b>Hitachi Content Archive Platform (HCAP)</b>	<b>Future HCAP</b>
Yearly Cost Savings	\$137,940	\$168,896	\$194,007
Yearly Incremental Benefits	\$598,000	\$1,349,725	\$1,703,975
NPV (3 years)	\$1,754,094	\$3,682,205	\$4,620,044
IRR (3 years)	434%	643%	724%
Breakeven (months)	9	8	8
Annual ROI (3 years)	672%	1586%	2040%

Figure 2 - Summary of Case Study Financials

Figure 2 shows the case for comparing an installed jukebox system with 1) a best-of-breed traditional digital archiving system, 2) Hitachi's content archive platform, and 3) enhancements to HCAP to allow the integration of TagmaStore, Hitachi's virtual controller platform. The chart shows a delta case, meaning the values represent the change from the existing installation.

Beyond straight cost savings, the case study illustrates incremental business benefits, such as improved availability of data, improved user productivity, and significant risk mitigation from unavailability of data or noncompliance, which are increasingly important and more expensive due to fines and penalties.

ITCentrix believes the architecture developed by Hitachi conceptually provides the best archiving platform available in the industry. It simplifies the development of applications and provides optimum performance. Because data is held independent of the applications, archive data can use the same infrastructure resources as any other production data.

The Hitachi content archive software is powered by Archivas, and HDS must demonstrate that they can deliver an aggressive road map to implement the vision for this platform. If they succeed, they will achieve economies of scale across all archiving applications and become a low-cost provider of archiving storage and archiving services. The Hitachi content archive platform could then compete to become the platform of choice for the development and management of archiving applications.

## Archiving Fundamentals

Archivists and document managers care about the following aspects of a physical document:

- Provenance – Where did it come from? Can the source be trusted? Has it been altered?
- Chain of custody – Who had the document and when?
- Control over disposition – When can or should the document be destroyed? What is the best way to ensure that it is not destroyed before that time?
- Accession – Who can have and who has had access to the document?

Translating these requirements into system terms, the Hitachi/Archivas solution provides the following:

- Document object immutability: protection from corruption and tampering.

This is achieved by computing a digital signature for each object and periodically recalculating the digital signature to ensure integrity of the document over time.

- Document object retention: when the retention period is set, there should be no possible way to delete the document before the retention period has expired.

This is achieved by internally managing each document as an object, with metadata that drives or enforces the protection policy. There is one single authoritative record, although many copies can be kept. When the document is “shredded,” there is no ability to retrieve the document or copies of the document.

- Protection: multiple copies of a document can be stored within the system. These copies can be distributed across multiple sites.
- Access and search mechanisms are available on both the archived and metadata, allowing records to be found and extracted.
- A comprehensive audit trail offers full transparency.

Audit trail is not provided now, but is on the short-term road map.

One additional requirement from an IT standpoint is that data and metadata should be available to other applications. For example, in a hospital, the PACS system manages the imaging data. However, if a doctor wanted to review a patient’s cardiology and other medical records, the system should have the ability to search across all archived data. The Hitachi content archive platform achieves this by storing the metadata and data separately while maintaining links to standard files that can be searched for and retrieved independent of the application that originally archived it. This allows separation of the data from the application, which enables better archiving and increased ability to reuse the data.

Figure 3 shows the functional components of HCAP, which is based on a globally accepted archiving standard (ISO 14721:2002). The preservation planning sets the parameters for retention and preservation. The administration function determines the configuration of the system and controls who is allowed to access the system.

The producers are usually applications that create the documents. HCAP then ingests the documents and metadata and stores these as an archive object. Consumers (users or applications) of the data in the archive can then access the data. Often additional documents are created during this process, which are in turn ingested by the system.

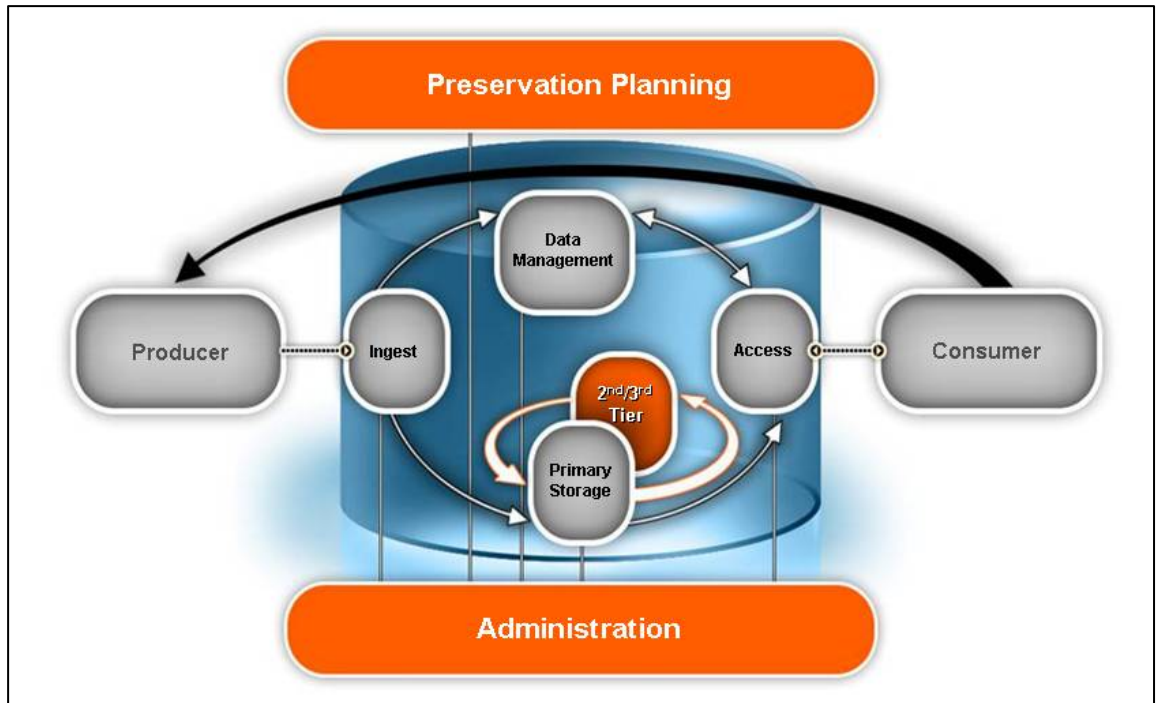


Figure 3 - Description of the Functions of the Hitachi Content Archive Platform (Source Hitachi Data Systems, 2006)

## Hitachi Content Archive Platform Architecture

### RAIN vs. SAIN

Most stand-alone archiving solutions utilize a redundant array of independent nodes (RAIN) architecture.

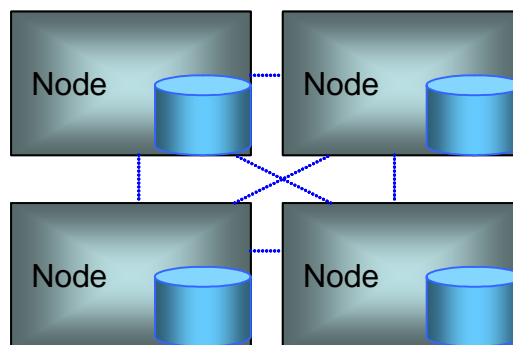
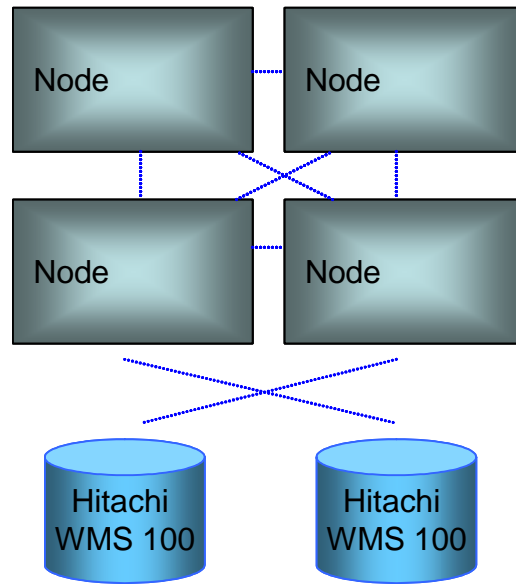


Figure 4 - RAIN Architecture

Figure 4 illustrates the RAIN architecture, with nodes based on industry-standard processors and storage connected together in an array by software. The advantage of this approach is that the hardware is inexpensive and can be easily distributed over multiple sites for data protection. The disadvantage is that any scaling requires processor and storage resources, and software has to provide all management services required, even if established solutions are available in the marketplace.

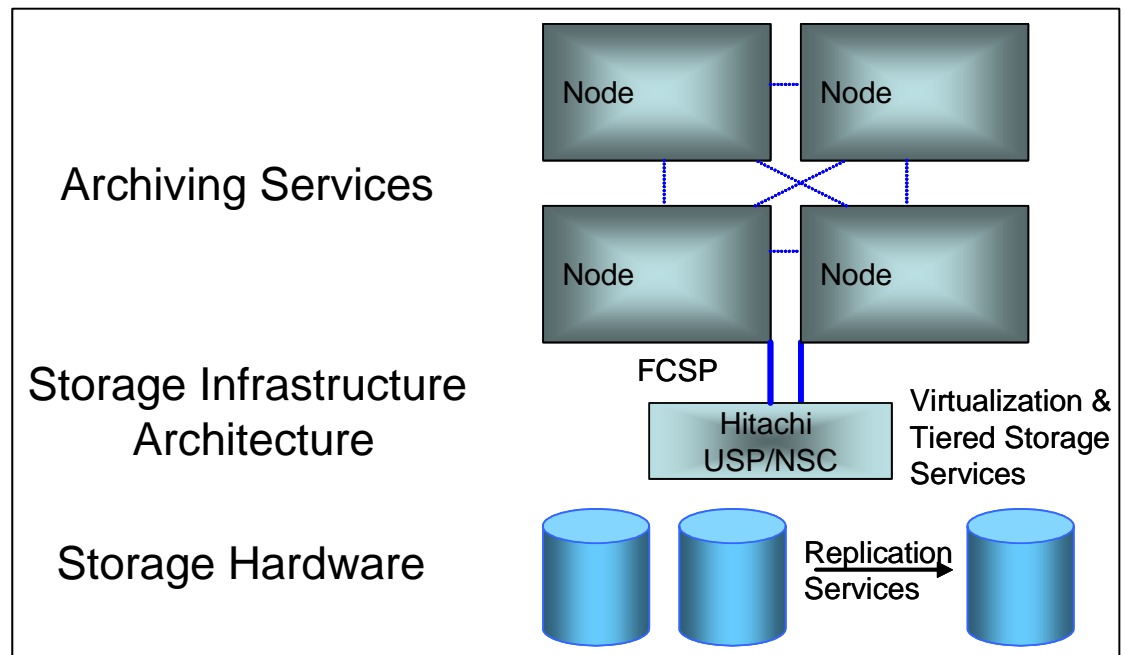
The current Archivas solution is implemented with RAIN architecture. The partnership between Hitachi and Archivas has developed a SAIN architecture that combines the benefits of advanced archive software and storage technologies. SAIN stands for SAN + Array of Independent Nodes.

Figure 5 shows the first implementation of the Hitachi SAIN solution. This separates out the storage from the nodes, using serial ATA (SATA) disk drives.



**Figure 5 - SAIN Architecture Stage 1**

Figure 6 shows the next implementation of the SAIN architecture. It will leverage the fibre channel security protocol (FCSP), part of the ANSI T11 FCSP draft standard, which provides robust authentication in a SAN environment, as well as data integrity for both host-to-switch and switch-to-switch communication. Full ratification is expected later in 2006. This will allow for the volumes to be entirely part of the SAN, connected through the Hitachi TagmaStore Universal Storage Platform (USP) and TagmaStore Network Storage Controller (NSC). This will provide a full set of services to the archived storage, including virtualization, tiered storage, WORM devices, and local and remote replications.



**Figure 6 – Full SAIN Architecture Implementation**

This architecture allows the complete separation of the storage hardware layer, the storage infrastructure layer, and the archiving services layer. The storage hardware layer can take advantage of different and new storage technologies (Hitachi or non-Hitachi) as performance and budget requirements dictate. The storage infrastructure layer can be enhanced independently to add new services. The archiving services layer can focus on archiving functionality and not worry about storage. If successfully implemented, this SAIN architecture should accelerate future functionality and reduce the cost of development.

## Case Study

### A Practical Implementation

The case study is from a general hospital with revenues of about \$1 billion. It was storing cardiology records and imaging records on DVD jukebox technology. The requirement was to keep general records for seven years and pediatric records for as long as 21 years.

Concern about the long-term ability to read the DVD media and the difficulty of identifying unreadable data were the main issues with the system. In addition, the availability of the system was only 97 percent. The hospital information security office was troubled that if the unavailability of data led to an adverse outcome, the hospital would be found liable. Adverse outcomes are settled at an average of \$8 million.

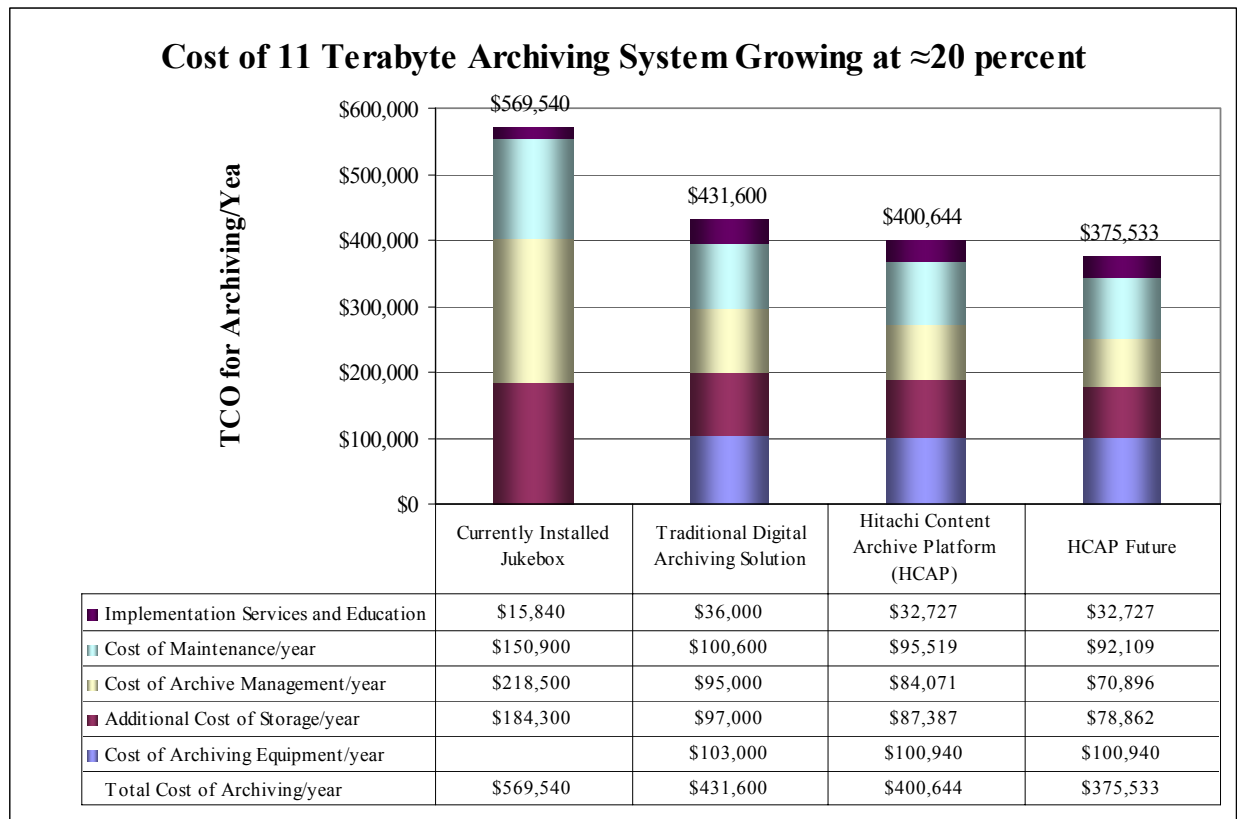
The requirements for the replacement system for archiving the cardiology and radiology records were defined as follows:

- Data available (and shown to be readable) for as long as 21 years
- Multiple copies to be held at two sites within the campus
- Tamperproof, so documents can be used in any court proceedings
- Prove to any legal audit that records can be accessed and have not been altered
- Provide an availability of at least 99.5 percent, in line with general systems availability

- Ability to work with the current storage system and applications
- Ability to scale, to hold all records, and extend to other patient records (sonograms, EKG strips, etc.) and systems (e.g., e-mail), if required
- Ability to delete files if required

Tape and optical solutions were evaluated, but were not options that improved the situation significantly. The IT department therefore evaluated disk-based archive solutions. The Archivas solution was selected because it met all criteria. In addition, the HCAP solution was unique in the following areas:

- It was flexible enough to work with block sizes of the current virtualization system installed (avoided changing infrastructure).
- It could read the DICOM header of the imaging records and include it in the metadata.
- It was able to search the data provided by all the applications as a single source. Although not implemented yet, this was seen as a significant advantage in future systems development.



**Figure 7 - Archiving Total Cost of Ownership (TCO)**

ITCentrix looked at the total cost of ownership (TCO) for four alternatives:

- The existing jukebox system
- Traditional archiving solutions
- HCAP solution

- Future HCAP solution including projected functionality

The results of the analysis are detailed in the chart in Figure 7. The most significant differences were the cost of additional storage and the cost of maintaining the jukebox system.

Yearly Costs and Benefits of providing 11 Terabytes Archiving Growing at ≈20 percent						
	Archiving Topology				Delta Business Case	
	Currently Installed Jukebox	Traditional Digital Archiving Solution	Hitachi Content Archive Platform (HCAP)	HCAP Future	HCAP vs. Traditional	HCAP Future vs. Traditional
Productivity from Increased Availability (200 Users, 10% Usage)	\$0	\$29,250	\$30,225	\$30,225	\$975	\$975
Improvement in Productivity from Improved Access to data (500 Users, 10% Usage)	\$0	\$243,750	\$975,000	\$1,316,250	\$731,250	\$341,250
Expected Avoidance of Claims from Adverse Outcomes	\$0	\$182,000	\$188,500	\$188,500	\$6,500	\$0
Avoidance of Fines from Non-Compliance	\$0	\$143,000	\$156,000	\$169,000	\$13,000	\$13,000
<b>Total Benefit/year</b>	<b>\$0</b>	<b>\$598,000</b>	<b>\$1,349,725</b>	<b>\$1,703,975</b>	<b>\$751,725</b>	<b>\$355,225</b>
Cost of Archiving Equipment/year		\$103,000	\$100,940	\$100,940	(\$2,060)	\$0
Additional Cost of Storage/year	\$184,300	\$97,000	\$87,387	\$78,862	(\$9,613)	(\$8,526)
Cost of Archive Management/year	\$218,500	\$95,000	\$84,071	\$70,896	(\$10,929)	(\$13,175)
Cost of Maintenance/year	\$150,900	\$100,600	\$95,519	\$92,109	(\$5,081)	(\$3,410)
Implementation Services and Education	\$15,840	\$36,000	\$32,727	\$32,727	(\$3,273)	\$0
<b>Total Cost of Archiving/year</b>	<b>\$569,540</b>	<b>\$431,600</b>	<b>\$400,644</b>	<b>\$375,533</b>	<b>\$ (30,956)</b>	<b>\$ (56,067)</b>
NPV (3 years)		\$1,754,094	\$3,682,205	\$4,620,044	\$1,928,237	\$2,866,178
IRR (3 years)		434%	643%	724%	∞	∞
Breakeven (Months)		9	8	8	-	-
Annual ROI (3 years)		672%	1586%	2040%	8102%	9305%

**Figure 8 – Costs and Benefits of 11 Terabyte Archiving System Growing at ≈20 percent**

The total initial solution was installed at a cost of about \$300,000 (\$100,000/year). There were significant savings from media and reduced staff to manage the system. Additional business benefits to the hospital were identified as the following:

- The improvement in productivity from hospital staff from significantly improved availability
- The mitigation of risk of data not being available and not being compliant with government regulations
- The ability to extend the system quickly and easily to other applications
- The ability to search seamlessly across all data

Figure 8 shows the benefits in detail. One of the most important factors is improved access to data and the ability to search and link to data across different applications. The hospital was interested in providing physicians automatic retrieval of all other data about a patient when they retrieve, for example, an image. This would encourage usage of the data to improve treatment and avoid adverse outcomes.

Overall, the current HCAP shows a net present value (NPV) of \$3.7 million compared with continuing to run the system on jukeboxes. The breakeven is eight months, and the IRR is 643 percent. The delta case for HCAP against traditional digital archiving systems shows an NPV of \$1.9 million.

In evaluating the system, the hospital information security officer said it met all their criteria for the archiving system and “worked as advertised.”

## Conclusions and Recommendations

---

The Hitachi content archive platform has the potential to provide the first open architecture in the archiving space. This will allow the independence of storage platform, the ability to leverage existing storage management investments, and the ability to access data independent of other applications. If delivered as specified, Hitachi Data Systems will be providing a unique contribution to the industry.

ITCentrix believes the architecture will fit extremely well into the market of emerging archiving solutions. By being application agnostic and focusing on providing the active archive infrastructure, Hitachi Data Systems and its suppliers can achieve economies of scale across the whole base of archiving solutions.

Users will eventually choose solutions based on the functionality of applications and the ability to integrate these applications with other systems and infrastructure. ISVs will make their decision of which platforms to support based on confidence in the platform, general availability to users, and leverage with existing offerings. ISVs will also want to ensure that platform vendors are not competing with them. If Hitachi Data Systems can focus on delivering the infrastructure and archiving services, meet its aggressive road map goals, and be responsive to the requirements of ISVs, the Hitachi content archive platform will become a significant player in the rapidly growing archiving market space.

### **ITCentrix**

1661 Worcester Road  
Framingham, MA 01701  
Tel: 508-620-6336  
www.itcentrix.com

## About ITCentrix

---

ITCentrix, which serves the needs of IT audiences, is a division of Barometrix Software Corporation. Barometrix is a privately held software company founded in 1999 by leading business and technology professionals. The company develops enterprise decision management and collaboration software designed to increase company performance through improved resource allocation and management.

The company's software products are used at several hundred organizations in North America, Europe, the Asia/Pacific region, and in emerging countries, to focus investments on returning optimal business value.

ITCentrix / 1661 Worcester Road / Framingham, MA 01701 / Tel: 508-620-6336