

EMC and ESRI Solutions for the GIS Community

Image Storage Architecture

July 2004

Copyright © 2004 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, Symmetrix, Celerra, CLARiiON, CLARAlert, DG, E-Infostructure, HighRoad, Navisphere, PowerPath, ResourcePak, SnapView/IP, SRDF, VisualSAN, where information lives, and The EMC Effect are registered trademarks and EMC Automated Networked Storage, EMC ControlCenter, EMC Developers Program, EMC Enterprise Storage, EMC Enterprise Storage Network, EMC OnCourse, EMC Proven, EMC Snap, Access Logix, AutoAdvice, Automated Resource Manager, AutoSwap, AVALONidm, C-Clip, CacheStorm, Celerra Replicator, Centera, CentraStar, CLARevent, Connectrix, CopyCross, CopyPoint, CrosStor, Direct Matrix, Direct Matrix Architecture, EDM, E-Lab, Enginuity, FarPoint, FLARE, GeoSpan, InfoMover, MirrorView, NetWin, OnAlert, OpenScale, Powerlink, PowerVolume, RepliCare, SafeLine, SAN Architect, SAN Copy, SAN Manager, SDMS, SnapSure, SnapView, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix DMX, TimeFinder, Universal Data Tone, and VisualSRM are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

Part Number H1116.1

Table of Contents

Overview	3
What Is Image Storage Architecture?.....	3
Why Is ISA so important?	3
Increased storage requirements	3
Expanding data resources.....	3
Online requirements	3
New business processes.....	3
IT evolution	4
Increasing storage costs	4
Complex storage management	4
GIS Database Technology Overview.....	5
ISA Technology Overview.....	6
ISA Deployment Strategies.....	9
ISA General Cost Comparison.....	14
Overview of Technology Alternatives	14
More Information	14

Overview

GIS data resources have expanded rapidly in recent years. These resources are supported by a proliferation of new data collection technology, in conjunction with expanding initiatives, to promote sharing and distributed use of geographic information services. Requirements for online data resources have grown to capitalize on new Web-based geographic mapping services technology. And, Federal requirements for integrated information management and real-time data access are placing growing demands on connected storage solutions.

This white paper discusses the work ESRI and EMC have done to evaluate available storage technology alternatives and identify viable storage strategies that address the growing need for online data access. It explains Image Storage Architecture, a unique cost-effective storage architecture created to support large volumes of image and historical data in an enterprise GIS production environment, and its value to today's ESRI users.

What Is Image Storage Architecture?

ISA is a family of storage solutions designed around the specific needs of the ESRI user community. This family of storage solutions was jointly developed by EMC and ESRI to support the growing needs of the GIS community. These packaged, integrated, and tested solutions mitigate risk and support turnkey deployment. Turnkey solutions include hardware, software, and implementation services.

The ISA platform environment supports a "buy-as-you-need" deployment strategy, where customers pay for only what they need at the time they need it. The modular architecture supports future expansion requirements with minimum loss of investment. The ISA can contribute to over 50 percent reduction in customer storage costs, with very little reduction in user performance and system reliability.

Why Is ISA so important?

Increased storage requirements

Most geographic database environments in the 1990s needed production database environments that required up to 100 GB of storage, supported primarily by vector data resources (points, polygons, lines, etc.). Most GIS operations today, however, require over a terabyte of storage that supports a mix of vector and image data resources.

Expanding data resources

Traditional geographic vector database environments requiring 10-20 GB of data in the 1990s have grown to over 100 GB. These production database environments are being supported by a proliferation of additional digital image data. Image data resources have grown from 100-200 GB file-based systems in the 1990s to several terabytes of data in current operation storage environments.

Online requirements

Internet-based Web mapping services offer an alternative source for geographic information and spatial data resources. Web mapping services provide user access to enterprise GIS data resources supporting standard business operations. Enterprise servers provide online access to very large image data resources supporting a variety of core business operations.

New business processes

Corporate business process evolution is demanding online access to historical data resources. Reduced dependence on paper information repositories is placing high demands on improved data archival and historical digital document integrity. Increasing interest in historical geographic research and evaluation of

spatial data resources over time places increasing interest and business demands on maintaining historical spatial vector and image data resources.

IT evolution

Infrastructure requirements have grown from supporting departmental GIS operations to organizational, community level, state, and international sharing of GIS data resources. Central data repositories are growing in number, along with an exponential growth in geographic data storage capacity requirements.

Increasing storage costs

The cost of storage is outpacing available operational budgets. Storage capacity requirements are rapidly expanding, much faster than what can be offset by the reduction in storage disk cost. The cost of storage is becoming the predominant technology investment, matching or exceeding the cost of enterprise server platforms and operational software technology in supporting most enterprise operations. The operational capacity and associated cost for data storage are growing at an alarming rate.

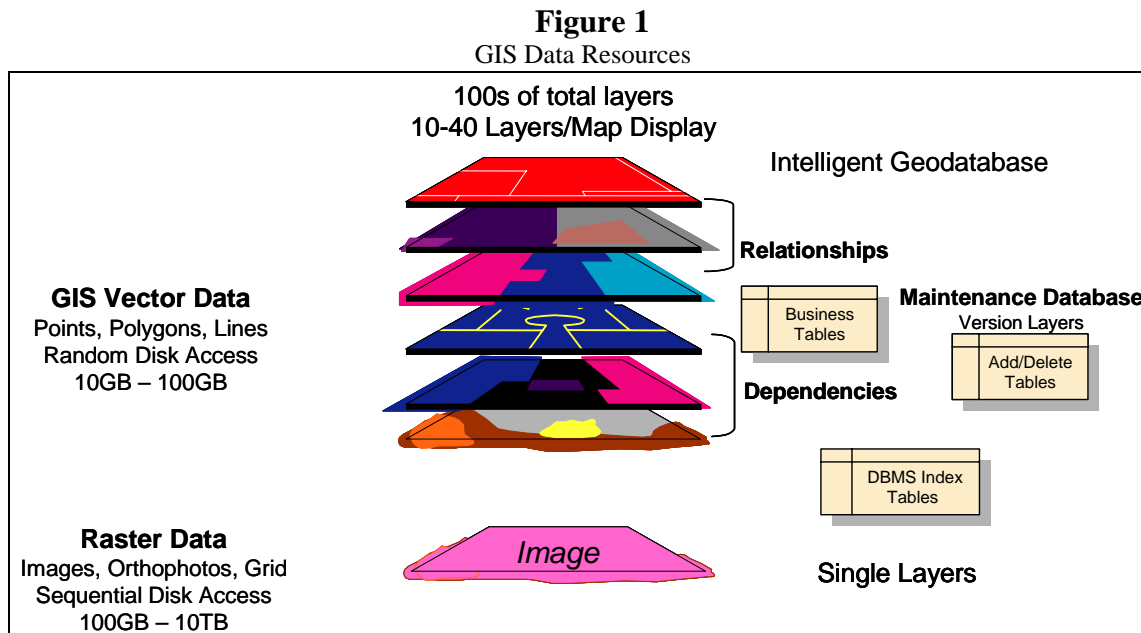
- CY 2000: \$5,000 (50 GB database at \$100 per GB)
- CY 2002: \$30,000 (1 TB database at \$30 per GB)
- CY 2003: \$250,000 (10 TB database at \$25 per GB)
- CY 2004: \$2,000,000 (100 TB database at \$20 per GB)

Complex storage management

Skyrocketing storage hardware capacity requirements present a growing level of complexity in managing these data resources. Optical systems have proven unreliable, driving cost of ownership beyond limits. These volumes of data resources cannot be effectively retrieved or referenced reliably from tape silos. The decreasing cost of disk has made tape and optical storage environments much less desirable and cost effective.

GIS Database Technology Overview

Geographic information products provide a map display of a variety of data resources. These data resources include GIS vector data, business attribute data, and a wide range of image data. An overview of the types of GIS data resources is provided in Figure 1.



ESRI production database environments are supported by an intelligent object-relational data model, which incorporates relationships and dependencies between spatial data features stored in the database. This data model establishes a feature-based object-relational geodatabase that supports geographic analysis and display functions executed by the ESRI ArcGIS desktop applications. This geodatabase environment provides the business framework to maintain and support enterprise business applications. The geodatabase business data can be presented on top of spatially registered (georeferenced) raster data (images) to support a composite map display.

Geographic raster data can be stored as files or in an ArcSDE database environment. Storing images in an ArcSDE database improves query performance and quality management of the image resources.

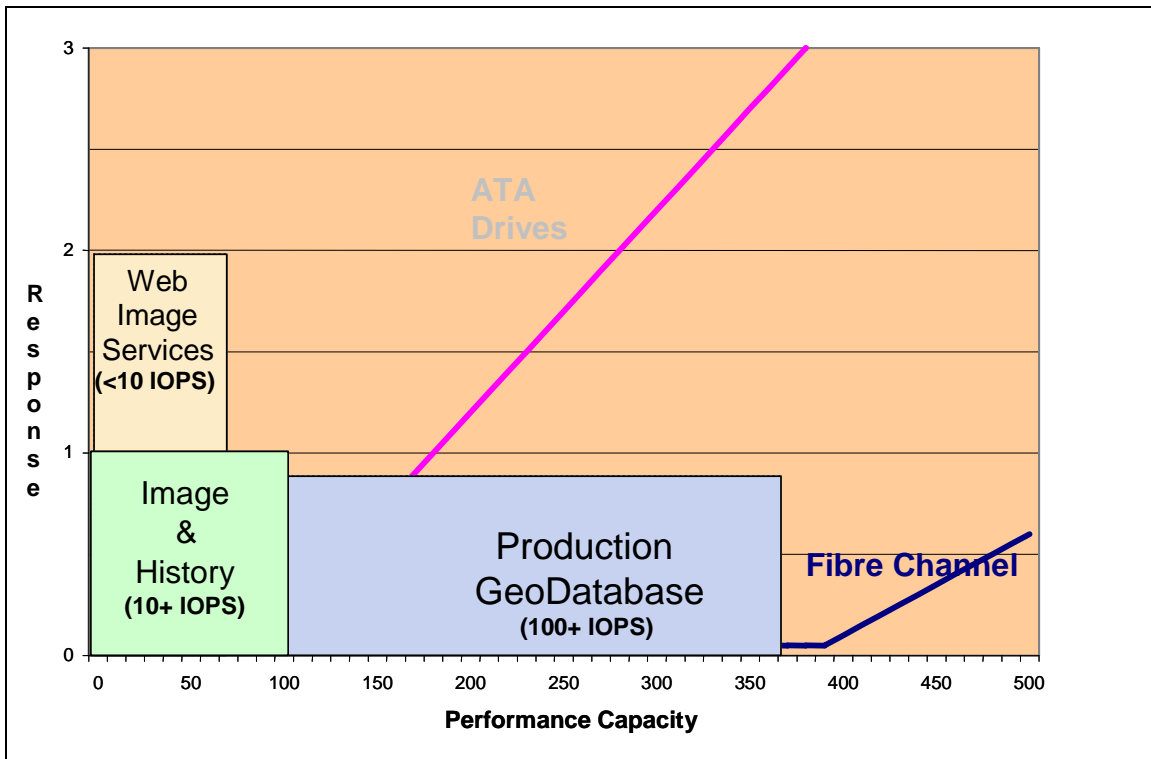
GIS data resources are maintained within a central ArcSDE database environment. ESRI ArcGIS desktop applications and ArcIMS Web Services are supported by network access to the ArcSDE server. A typical ArcGIS desktop map display is supported by a range of 10-40 individual vector data layers normally overlaid on top of one raster image layer. Similar map displays are generated by ArcIMS Web Services. Each layer display is supported by a series of database queries to the ArcSDE server, and supports DBMS disk I/O requests to the storage environment. A typical 1-sec map display may generate over 100 database queries and render over a megabyte of spatial data extracted from the ArcSDE database storage environment. These queries place heavy I/O processing demands on the disk storage subsystem. The majority of these I/O transactions are focused on accessing the production geodatabase (GIS vector and tabular business data). The image data queries are less than 10 percent of the overall query transactions and are normally stored sequentially on the disk storage media.

ISA Technology Overview

The Image Storage Architecture (ISA) provides an optimum mix of high-performance Fibre Channel disk storage technology with high-capacity, lower cost ATA disk storage as a standard GIS production storage architecture. Both disk technologies are integrated in a high-performance, highly available, fault-tolerant hardware RAID storage environment supported by best-of-breed storage management technology. A variety of storage strategies are supported with current vendor technology to cover a wide range of customer storage capacity needs.

The ISA takes advantage of dramatic differences between GIS disk performance demands for vector and image data resources in supporting standard map display queries (see Figure 2). A very large number of vector data queries are required to support a single map display, while a relatively small number of image queries support the same map display transaction. Vector and tabular data are supported on high-performance/capacity Fibre Channel disk storage, while image and historical data are supported on higher density lower cost ATA drive technology. The relatively high-capacity storage requirements for GIS image data result in a significant reduction in overall ISA storage cost.

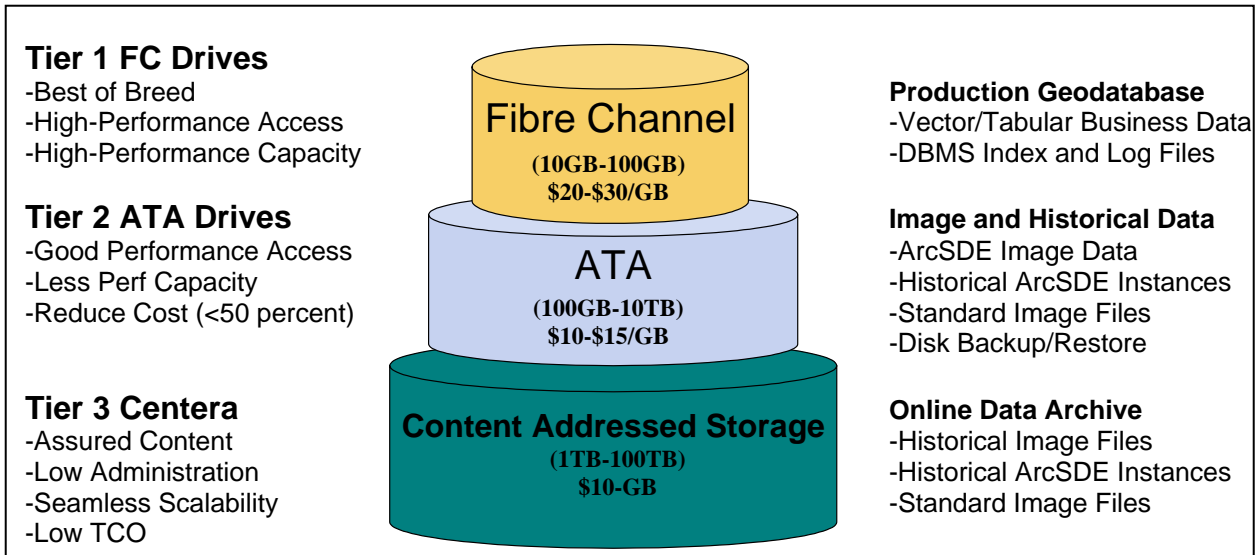
Figure 2
ISA Disk Access Performance



High-performance Fibre Channel storage is important to avoid disk contention when a very high quantity of disk queries must be supported from the storage volumes. Fibre Channel technology provides the highest level of disk access performance and capacity available in the storage marketplace. This technology is important to support an ArcSDE production database, particularly with peak concurrent client loads in access of 50 clients. The total number of disks must be sufficient to avoid I/O performance bottlenecks. Most Web mapping services fall below the equivalent load of 50 concurrent desktop clients (peak map requests of 75,000 TPH would generate an equivalent load of 50 concurrent desktop clients).

The primary ISA production environment is a two-tier architecture (see Figure 3). A tray of standard Fibre Channel high-performance/capacity disk drives supports the first tier. The second tier is supported by a variable number of trays with standard ATA high-capacity disk drives. The optional third tier consists of an API and EMC Centera™ to operate as an online image data repository.

Figure 3
ISA Data Placement Strategy



The first storage tray supports the ArcSDE production geodatabase. This includes both vector and tabular business tables and database index and log files. Standard GIS storage recommendations should be followed to keep index and log files separate from the vector and tabular business tables. A single storage tray can support up to 15 Fibre Channel disks, with disk volume selections of 32 GB, 73 GB, or 146 GB, depending on customer data capacity requirements. The first storage tray supports over 2 TB of disk capacity (when using 146 GB drives), providing plenty of capacity to support ESRI customer current-period production geodatabase environments.

The remaining storage trays are used to support ArcSDE image data, historical online ArcSDE geodatabase instances, and standard image files. This includes both vector and tabular business tables and database index and log files. Index and log tables for this data can be supported on the Fibre Channel tray for improved access performance. Currently, a single storage tray can support up to 15 ATA disks, each currently supporting 320GB of storage capacity. The first ATA storage tray supports over 4.8 TB of disk capacity, which is ideal for most entry-level image storage environments. ISA disk storage solutions can support up to 16 trays in a single rack-mounted storage environment, supporting over 74 TB of GIS data resources from a single high-performance disk controller. Several of these systems can be supported in an enterprise storage environment to support online storage of over a petabyte of online image data resources.

Included in the ESRI packages are EMC CLARiiON® Fibre Channel storage systems. EMC CLARiiON CX storage provides scalability, the highest availability, bullet-proof reliability, and data integrity with a broad range of powerful, easy-to-use management tools. The CLARiiON platform supports multiple server environments, including Windows, Linux, and UNIX. CLARiiON also supports multiple networked storage environments, including SAN, NAS, and direct attached. EMC is the first major storage vendor to implement ATA technology deployed in standard storage array configurations. CLARiiON with ATA drive technology is less expensive than fibre or SCSI drive components and is higher-performing and more reliable than tape or optical storage media.

An optional third tier that includes EMC Centera™ content addressed storage is also available for archival of image data. When image data must be archived, the traditional methods of tape and optical storage devices have proven to be costly to maintain and manage. The ISA third tier takes advantage of the Centera with its many features and benefits and provides a secure repository for static image files. Centera meets the unique requirements of fixed content—unchanging digital assets that are retained for active reference and long-term value. This non-rewriteable, non-erasable system provides online access with assured content authenticity and seamless scalability. Centera also self monitors to detect and repair faults. It helps ensure information accessibility by automatically reconfiguring itself and replicating objects as necessary. Nirvana SRB MAPServer and LizardTech ExpressServer provide standalone NFS or CIFS access to the Centera image storage repository.

A policy-based intelligent data manager, Nirvana SRB MAPServer software enables the movement of vector data, spatial imagery, and satellite image data from the CLARiiON storage to and from the Centera system—where archived content lives—and provides users immediate access to this critical information. The Centera archive solution is specifically designed to meet the unique requirements of fixed content, supporting a broad range of standard image file data formats. For additional information on SRB MAPServer, go to <http://www.nirvanastorage.com>. To read the ESRI testing report with SRB MAPServer and Centera, go to http://www.esri.com/systemsint/pdf/srb_centera_poc_testing.pdf.

Lizardtech's ExpressServer 4.6 provides a high-speed access and retrieval API for MrSID image file formats. Lizardtech also provides an encoder to convert most file formats into the MrSID format. Files generated with LizardTech ExpressServer 4.6 can be stored in EMC Centera for assured content authenticity and can be efficiently distributed to users across the hall or around the world. Efficient *pixel-stream* technology makes it possible for existing Web servers to provide instant viewing of high-resolution aerial photographs, satellite imagery, and other digital content on both wireless and broadband devices. The integrated solution also supports many other common file formats and provides users with a fast, efficient, and cost-effective alternative for long-term storage and content distribution without sacrificing quality, resolution, or legibility. ExpressServer 4.6 integration with Centera and ESRI image solutions bridges the gap between massive offline storage of geospatial imagery and end-user access. For additional information on ExpressServer 4.6, go to <http://www.lizardtech.com/solutions/emc/>. To read the ESRI testing report with ExpressServer and Centera go to http://www.esri.com/systemsint/pdf/express_server_centera_poc_testing.pdf.

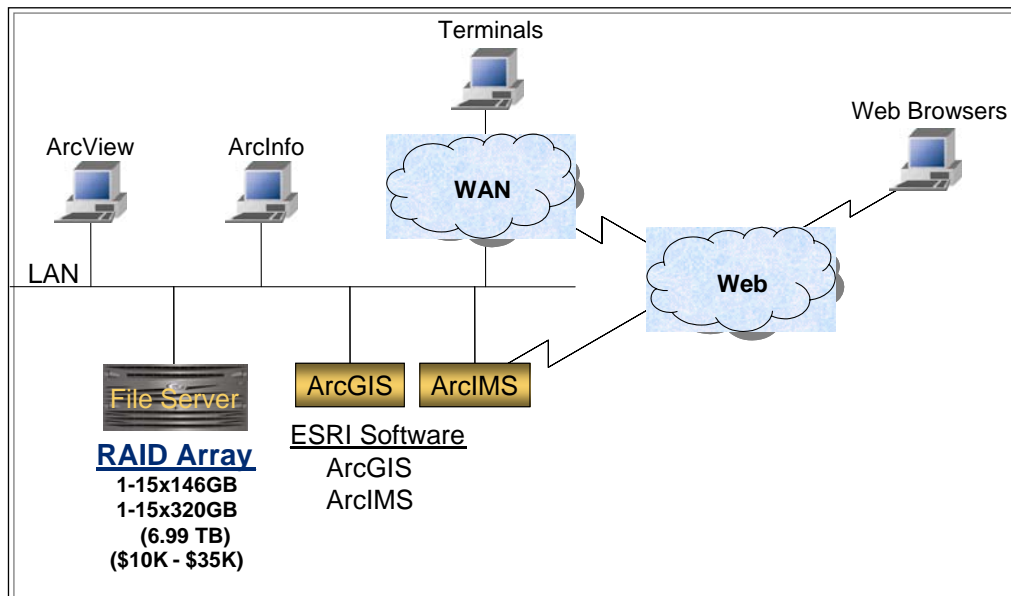
ISA Deployment Strategies

The ISA can be deployed to support a variety of ESRI customer production environments. These environments range from entry level to the largest ESRI customer environments. Typical implementation strategies are highlighted in the following sample architecture drawings.

- Entry Level ISA Departmental/Regional File Server Environment

Some of the smaller entry-level GIS operations are supported by GIS file data sources. These environments are typically department-level GIS operations and have limited IT administration resources. This solution does not support an ArcSDE database environment.

Figure 4
Entry-Level ISA File-Based Departmental/Regional Server Environment

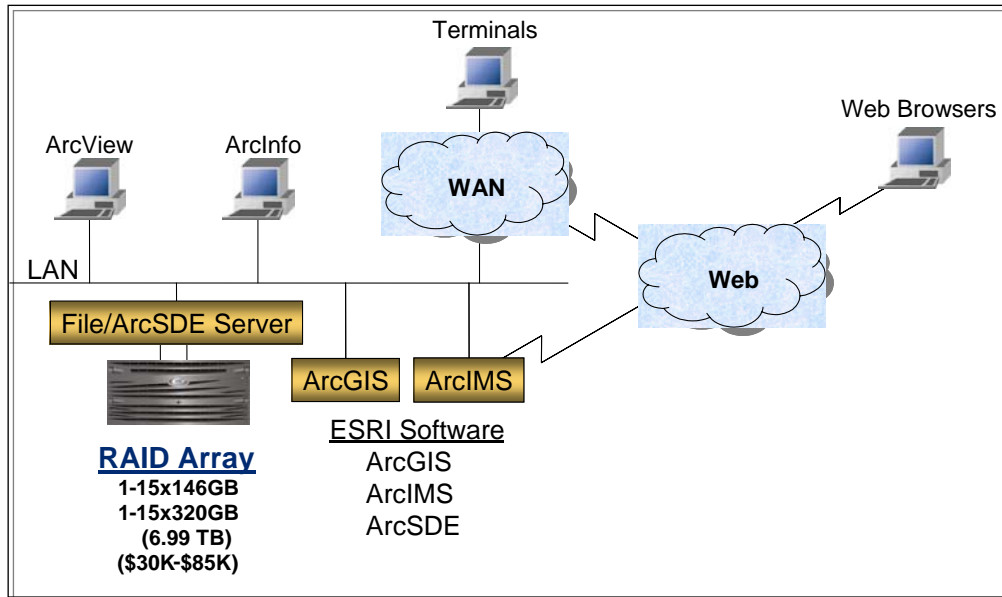


The GIS file server is supported by a network attached storage appliance. The first storage tray supports the primary GIS vector layers and the second tray supports the image data. A single rack-mounted system supports up to 6.99 TB of total disk storage capacity.

- Entry-Level ISA ArcSDE Enterprise Environment

Many of the smaller entry-level GIS operations are supported by ArcSDE geodatabase production data and a variety of file data sources. These environments can vary from department-level GIS operations to small community or commercial business operations. This environment will require a DBMS database administrator and typically is supported by a production IT administration team.

Figure 5
Entry-Level ISA ArcSDE Enterprise Environment

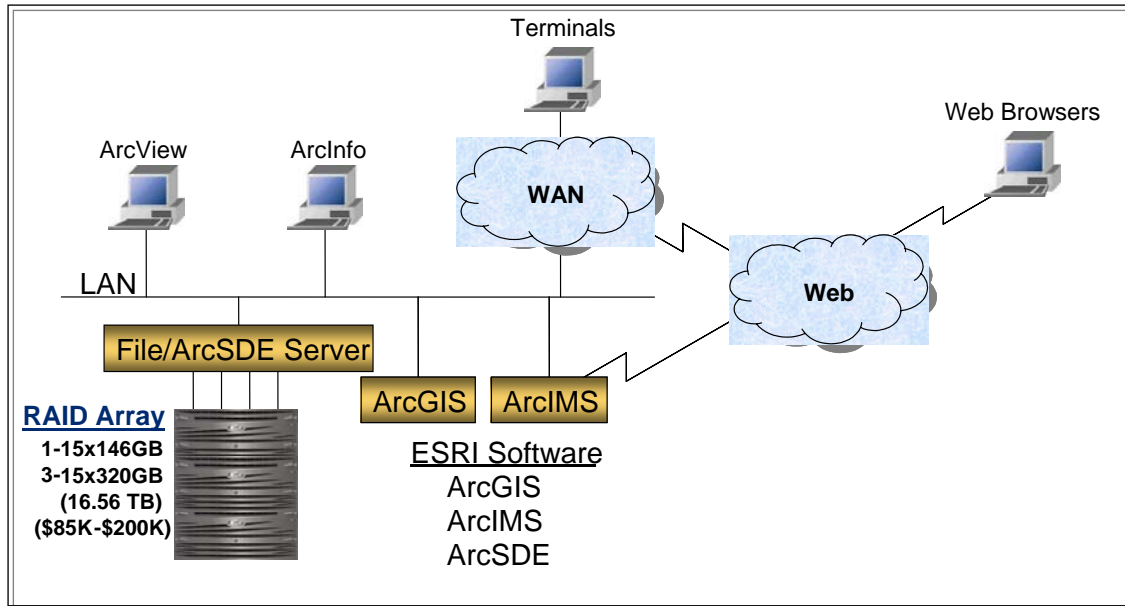


A standard ArcSDE server supports the database and file server environment. The first storage tray supports the primary ArcSDE production geodatabase and database index and log tables. The second tray supports the image and history data sources. A single rack-mounted system supports up to 6.99 TB of total disk storage capacity.

- Standard ISA ArcSDE Enterprise Environment

Most GIS production operations can be supported by this storage environment. This environment supports a full ArcSDE production geodatabase and up to 14 TB of image data.

Figure 6
Standard ISA ArcSDE Enterprise Environment

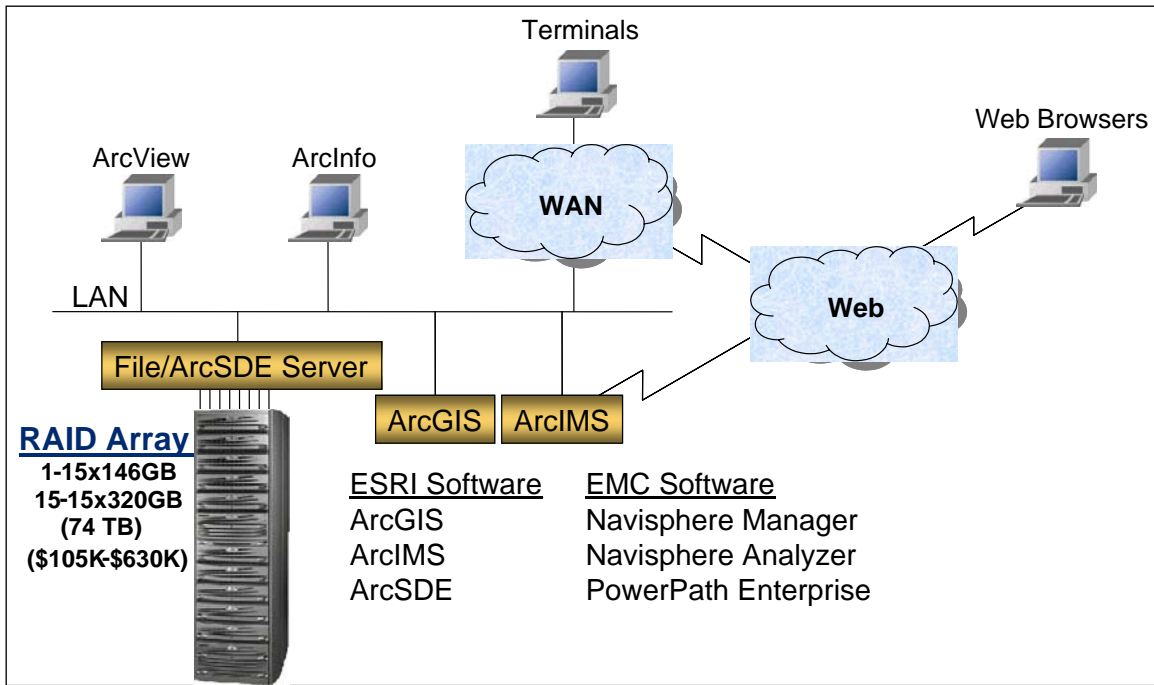


A standard ArcSDE server supports the database and file server environment. The first storage tray supports the primary ArcSDE production geodatabase and database index and log tables. The second tray supports the image and history data sources. A single rack-mounted system supports up to 16.56 TB of total disk storage capacity.

- High-Capacity ISA ArcSDE Enterprise Environment

The largest GIS production operations can be supported by this storage environment, which supports a full ArcSDE production geodatabase and up to 72 TB of image data. Storage requirements can be expanded up to 1 petabyte of online data by integrating several of these rack-mounted storage solutions behind a storage area network environment.

Figure 7
High-Capacity ISA Enterprise Environment

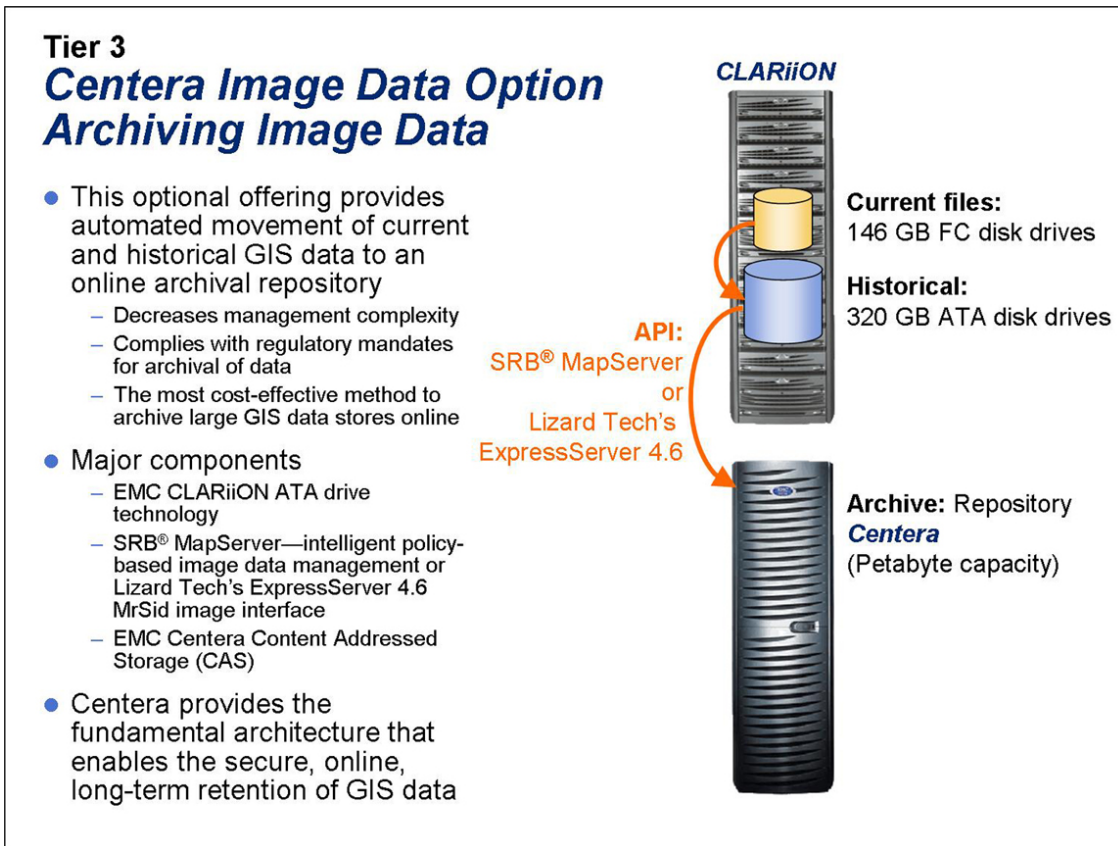


A standard ArcSDE server supports the database and file server environment. The first storage tray supports the primary ArcSDE production geodatabase and database index and log tables. The remaining storage trays support the image and history data sources. A single rack-mounted system supports up to 74 TB of total disk storage capacity.

- Centera Image Data Archiving Option

The optional third tier with Centera may be configured with any of the above storage configurations to provide the customer with the maximum flexibility for Image Storage in GIS environments. See Figure 8 for typical ISA configurations with the Centera option.

Figure 8
Centera Image Data Archiving Environment



ISA General Cost Comparison

Significant cost savings can be realized from the Image Storage Architecture. A comparison of market storage prices for standard high-performance, best-of-breed storage solutions is provided in Figure 8.

Figure 9
Market Storage Price Comparison

Standard FC Storage			ISA Storage		
Disk (GB)	Market	\$/GB	Disk (GB)	Market	\$/GB
4,300	\$ 49,980	\$11.62	5,000	\$ 35,000	\$ 7.00
4,300	\$ 98,980	\$23.02	5,000	\$ 84,000	\$16.80
8,760	\$247,940	\$28.30	13,000	\$203,000	\$15.62
35,000	\$854,700	\$24.42	59,000	\$630,000	\$10.68

This pricing comparison shows over \$165,000 savings over standard Fibre Channel storage when using the ISA with 10 TB of data.

Overview of Technology Alternatives

The Image Storage Architecture (ISA) provides a cost-effective design alternative for supporting large volumes of image and historical ArcSDE production instances in an enterprise GIS production environment. Figure 9 provides an overview of the technology alternatives included in the EMC/ESRI ISA design study.

Figure 10
ISA Design Alternatives Overview

Solution Storage Technology	Figure 4	Figure 5	Figure 6	Figure 7
	ISA-1 NAS	ISA-2 DAS	ISA-3 DAS/SAN	ISA-4 DAS/SAN
Min. Capacity	700 GB	700 GB	2 TB	4 TB
Max. Capacity	6.99 TB	6.99 TB	16.56 TB	74 TB
ESRI Software	ArcGIS ArcIMS	ArcGIS ArcIMS ArcSDE	ArcGIS ArcIMS ArcSDE	ArcGIS ArcIMS ArcSDE

More Information

For more information about the ISA solution, please call your ESRI sales representative or go to www.esri.com. To find out more about EMC products and solutions, visit the website at www.emc.com.