Environmental Systems Research Institute, Inc., 380 New York St., Redlands, CA 92373-8100 USA • TEL 909-793-2853 • FAX 909-307-3014

ESRI Systems Integration Technical Brief

Image Storage Architecture

Prepared by:

Dave Peters
ESRI Systems Integration Department
Redlands, California

January 5, 2004

Image Storage Architecture

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 the 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.

ESRI has teamed with EMC to evaluate available storage technology alternatives and to identify viable storage strategies that address the growing need for online data access. As a result of this research, a unique cost-effective storage architecture, Image Storage Architecture (ISA), was created to support large volumes of image and historical data in an enterprise GIS production environment.

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 more than a 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 to20 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- to 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-, state-, and international-level 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 is 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)

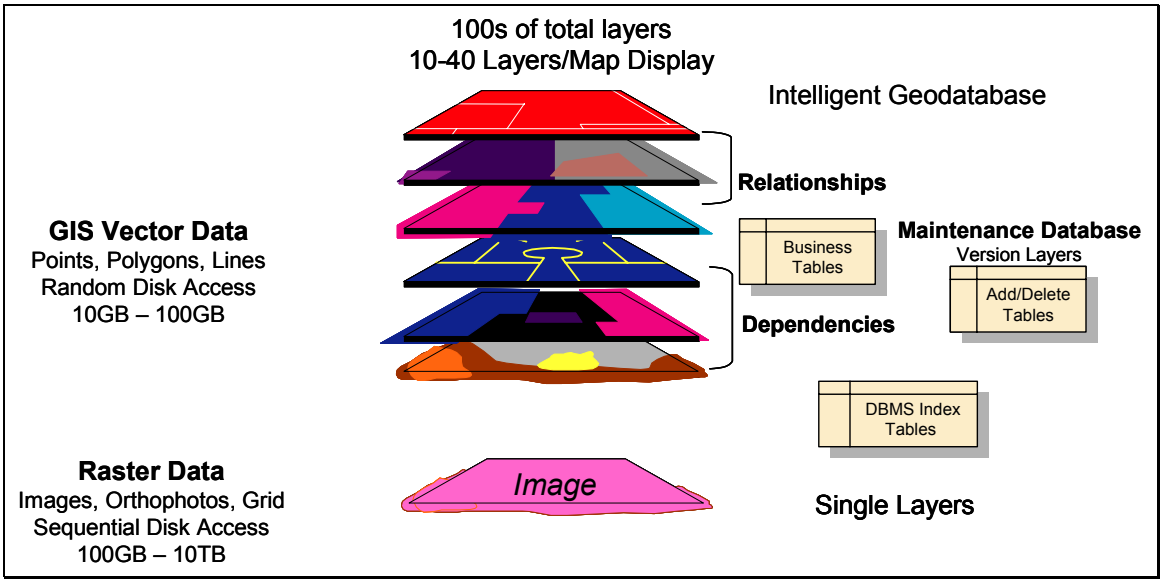
Complex storage management

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.

Figure 1
GIS Data Resources



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 to 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-second 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).

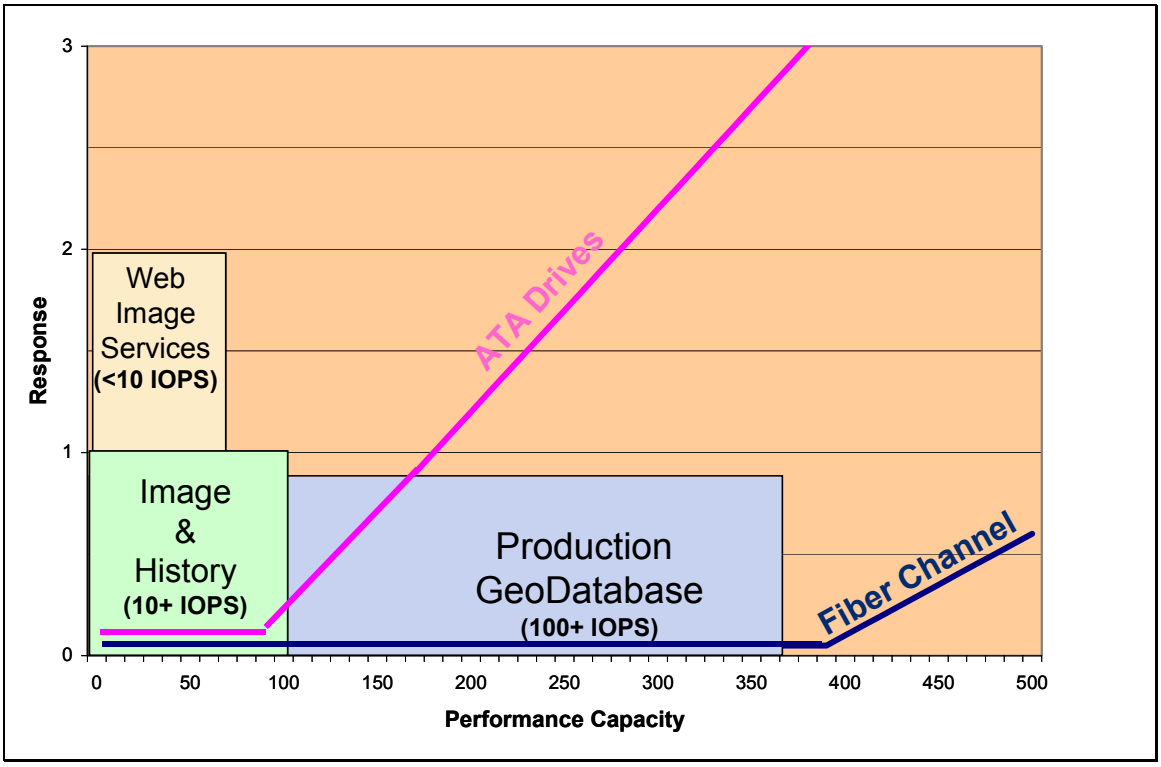
The image data queries are less than 10 percent of the overall query transactions and normally are stored sequentially on the disk storage media.

ISA Technology Overview

The Image Storage Architecture (ISA) provides an optimum mix of high-performance fiber-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 fiber-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.

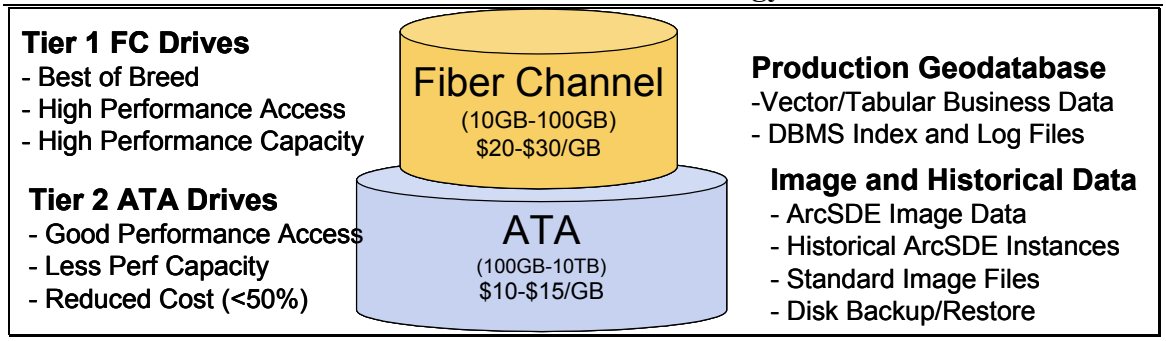
Figure 2
ISA Disk Access Performance



High-performance fiber-channel storage is important to avoid disk contention when a very high quantity of disk queries must be supported from the storage volumes. Fiber-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 fiber-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.

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 fiber-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 fiber-channel tray for improved access performance. Currently, a single storage tray can support up to 15 ATA disks, each supporting 250 GB of storage capacity. The first ATA storage tray supports over 3.7 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 57 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.

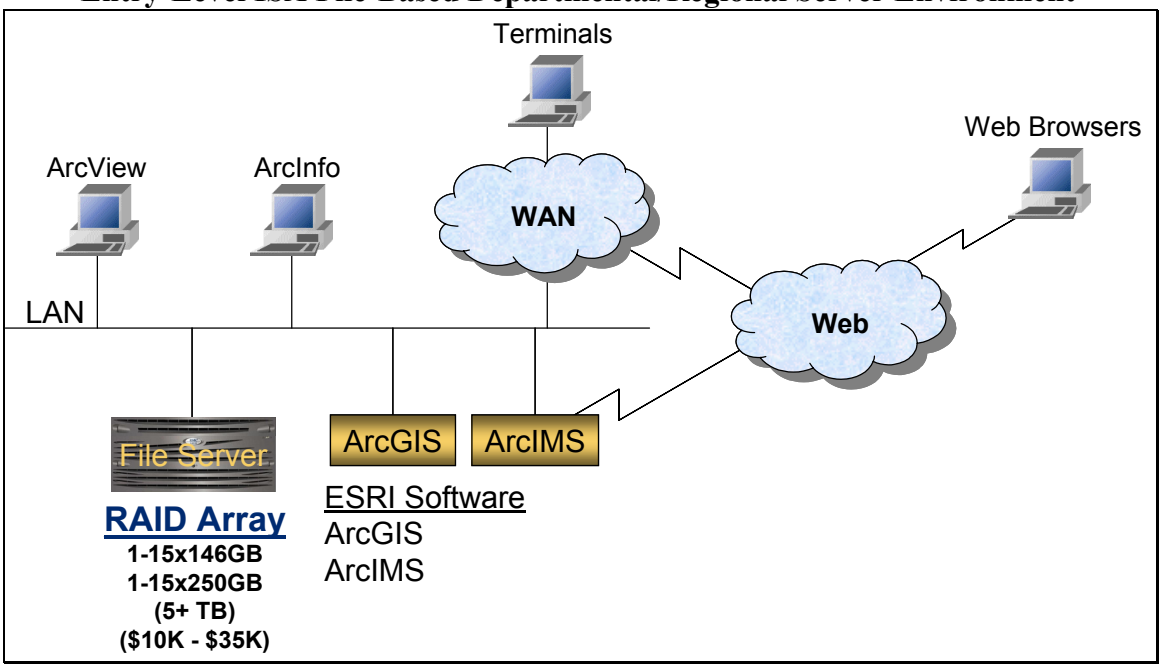
ISA Deployment Strategies

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

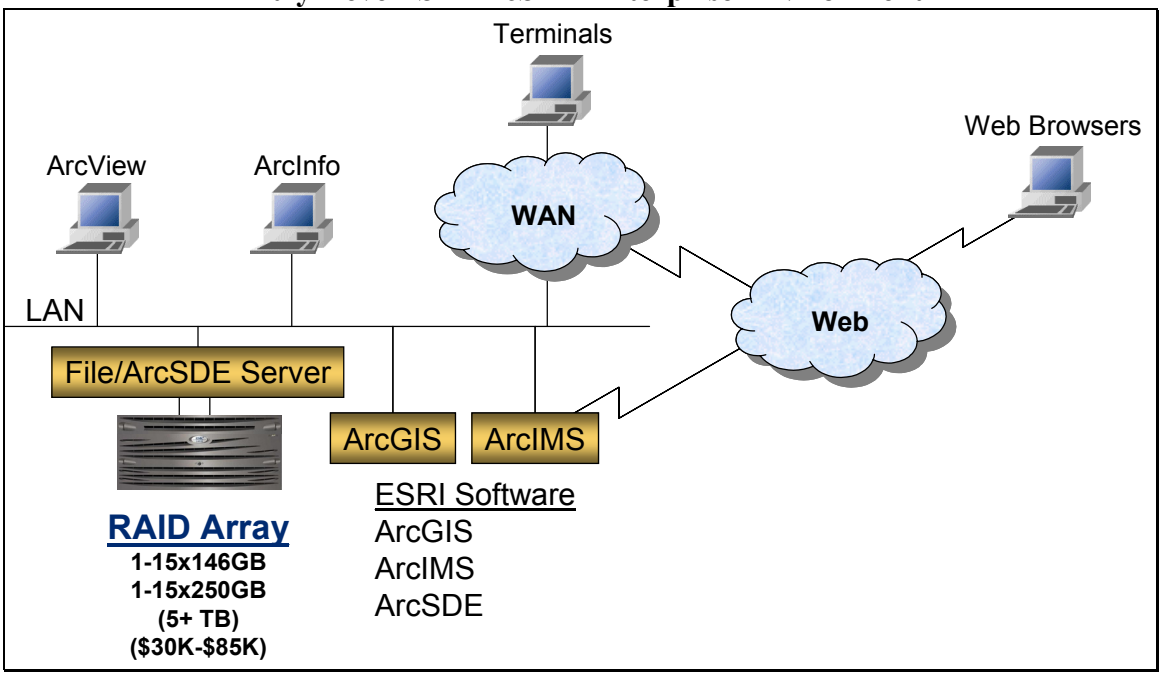


The GIS file server is supported by a network attached storage appliance. The first storage tray supports the primary GIS vector layers and the second tray supports the image data. A single rack-mounted system supports up to 5 TB of total disk storage capacity.

- 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 is typically supported by a production IT administration team.

Figure 5
Entry-Level ISA ArcSDE Enterprise Environment

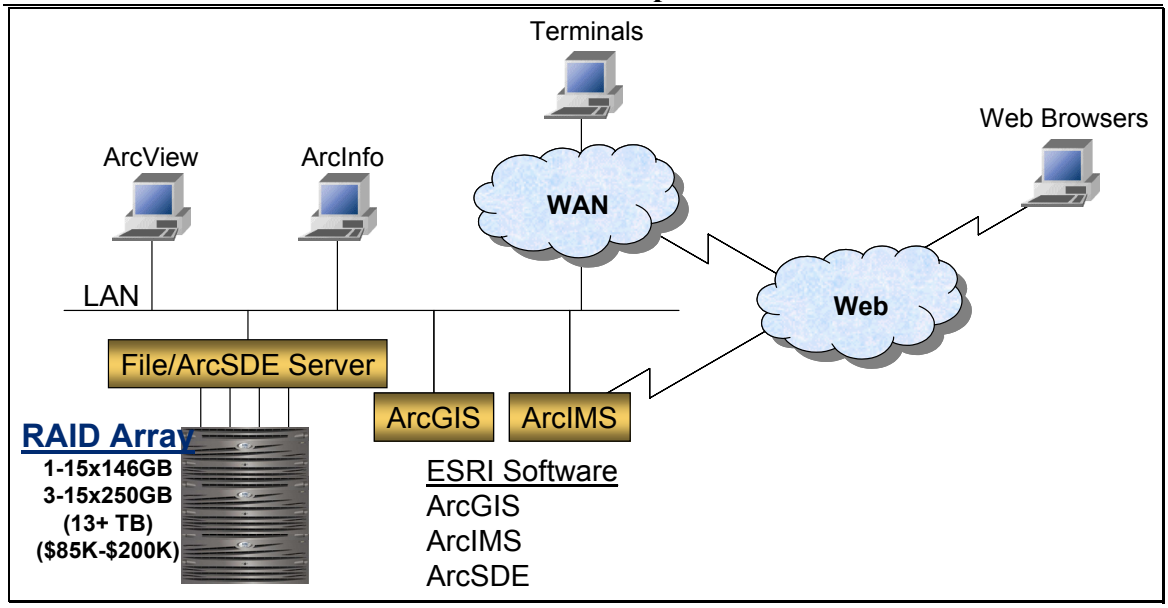


A standard ArcSDE server supports the database and file server environment. The first storage tray supports the primary ArcSDE production geodatabase and database index and log tables. The second tray supports the image and history data sources. A single rack-mounted system supports up to 5 TB of total 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 10 TB of image data.

Figure 6
Standard ISA ArcSDE Enterprise Environment

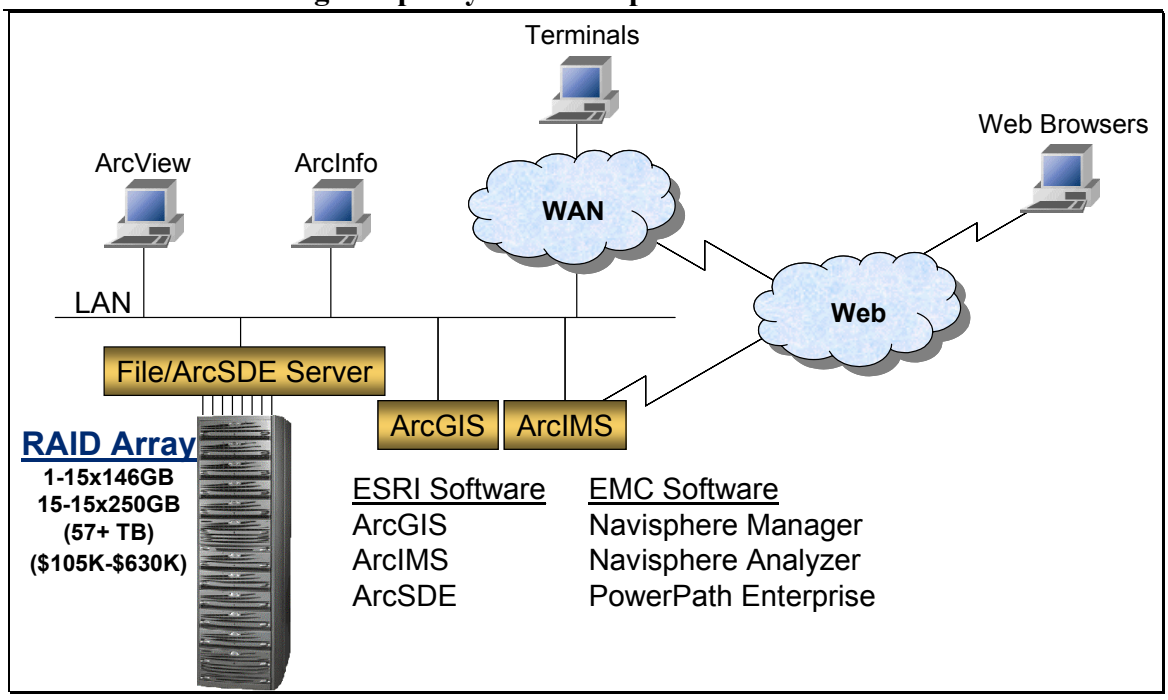


A standard ArcSDE server supports the database and file server environment. The first storage tray supports the primary ArcSDE production geodatabase and database index and log tables. The second tray supports the image and history data sources. A single rack-mounted system supports up to 13 TB of total 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 50 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 57 TB of total disk storage capacity.

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 8
Market Storage Price Comparison

Standard FC Storage			ISSA 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 fiber-channel storage when using the ISA with 10 TB of data.

Summary

The 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 Image Storage Architecture design study.

Figure 9
ISA Design Alternatives Overview

Solution Storage Technology	ISSA-1 NAS	ISSA-2 DAS	ISSA-3 DAS/SAN	ISSA-4 DAS/SAN
Min. Capacity	700 GB	700 GB	2 TB	4 TB
Max. Capacity	5 TB	5 TB	18 TB	57 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 contact us at sihelp@esri.com. To find out more about EMC products and solutions, visit their Website at www.emc.com.