



System Design Strategies

An ESRI White Paper • July 2002

Prepared by:

Dave Peters
Systems Integration

Environmental Systems Research Institute, Inc.
380 New York Street
Redlands, California 92373-8100

Copyright © 2002 Environmental Systems Research Institute, Inc.

All rights reserved.

Printed in the United States of America.

The information contained in this document is the exclusive property of Environmental Systems Research Institute, Inc. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by Environmental Systems Research Institute, Inc. All requests should be sent to Attention: Contracts Manager, Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373-8100, USA.

The information contained in this document is subject to change without notice.

U.S. GOVERNMENT RESTRICTED/LIMITED RIGHTS

Any software, documentation, and/or data delivered hereunder is subject to the terms of the License Agreement. In no event shall the U.S. Government acquire greater than RESTRICTED/LIMITED RIGHTS. At a minimum, use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in FAR §52.227-14 Alternates I, II, and III (JUN 1987); FAR §52.227-19 (JUN 1987) and/or FAR §12.211/12.212 (Commercial Technical Data/Computer Software); and DFARS §252.227-7015 (NOV 1995) (Technical Data) and/or DFARS §227.7202 (Computer Software), as applicable. Contractor/Manufacturer is Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373-8100, USA.

ESRI, ARC/INFO, ArcCAD, ArcView, *BusinessMAP*, MapObjects, PC ARC/INFO, SDE, and the ESRI globe logo are trademarks of Environmental Systems Research Institute, Inc., registered in the United States and certain other countries; registration is pending in the European Community. 3D Analyst, ADF, ARC COGO, the ARC COGO logo, ARC GRID, the ARC GRID logo, ArcInfo, the ArcInfo logo, the ARC/INFO logo, AML, ARC NETWORK, the ARC NETWORK logo, *ArcNews*, ARC TIN, the ARC TIN logo, ArcInfo LIBRARIAN, ArcInfo—Professional GIS, ArcInfo—The World's GIS, ArcAtlas, the ArcAtlas logo, the ArcCAD logo, the ArcCAD WorkBench logo, ArcCatalog, the ArcData logo, the ArcData Online logo, ARCEDIT, the ARCEDIT logo, ArcExplorer, the ArcExplorer logo, ArcExpress, the ArcExpress logo, ArcFM, the ArcFM logo, the ArcFM Viewer logo, ArcGIS, ArcIMS, the ArcIMS logo, ArcLogistics, the ArcLogistics Route logo, ArcMap, ArcObjects, ARCPLOT, the ARCPLOT logo, ArcPress, the ArcPress logo, the ArcPress for ArcView logo, ArcScan, the ArcScan logo, ArcScene, the ArcScene logo, ArcSchool, ArcSDE, the ArcSDE logo, the ArcSDE CAD Client logo, ArcSdl, ArcStorm, the ArcStorm logo, ArcSurvey, ArcToolbox, ArcTools, the ArcTools logo, ArcUSA, the ArcUSA logo, *ArcUser*, the ArcView GIS logo, the ArcView 3D Analyst logo, the ArcView Business Analyst logo, the ArcView Data Publisher logo, the ArcView Image Analysis logo, the ArcView Internet Map Server logo, the ArcView Network Analyst logo, the ArcView Spatial Analyst logo, the ArcView StreetMap logo, the ArcView StreetMap 2000 logo, the ArcView Tracking Analyst logo, ArcVoyager, ArcWorld, the ArcWorld logo, Atlas GIS, the Atlas GIS logo, AtlasWare, Avenue, the Avenue logo, the *BusinessMAP* logo, DAK, the DAK logo, DATABASE INTEGRATOR, DBI Kit, the Digital Chart of the World logo, the ESRI corporate logo, the ESRI PRESS logo, ESRI—Team GIS, ESRI—The GIS People, FormEdit, Geographic Design System, Geography Matters, GIS Day, the GIS Day logo, GIS by ESRI, GIS for Everyone, GISData Server, *InsiteMAP*, MapBeans, MapCafé, the MapCafé logo, the MapObjects logo, the MapObjects Internet Map Server logo, ModelBuilder, NetEngine, the NetEngine logo, the PC ARC/INFO logo, PC ARCEDIT, PC ARCPLOT, PC ARCSHELL, PC DATA CONVERSION, PC NETWORK, PC OVERLAY, PC STARTER KIT, PC TABLES, the Production Line Tool Set logo, *RouteMAP*, the *RouteMAP* logo, the *RouteMAP* IMS logo, Spatial Database Engine, the SDE logo, SML, StreetMap, TABLES, The World's Leading Desktop GIS, *Water Writes*, and Your Personal Geographic Information System are trademarks; and ArcData, ArcOpen, ArcQuest, *ArcWatch*, ArcWeb, Rent-a-Tech, @esri.com, and www.esri.com are service marks of Environmental Systems Research Institute, Inc.

The names of other companies and products herein are trademarks or registered trademarks of their respective trademark owners.

System Design Strategies

An ESRI White Paper

Contents	Page
1.0 System Design Process	1-1
2.0 ESRI Software Evolution	2-1
3.0 Network Communications	3-1
4.0 GIS Product Architecture	4-1
5.0 GIS User Needs	5-1
6.0 Sizing Fundamentals	6-1
7.0 Sizing Tools	7-1
8.0 System Implementation	8-1

Attachment A: System Architecture Design Services

A-1: Departmental System Architecture Design Consulting Services

A-2: Enterprise System Architecture Design Consulting Services

A-3: ArcIMS Architecture Design Consulting Services

Attachment B: System Architecture Design for GIS Class Description

Figures	Page
Figure 1-1: What Is System Architecture Design?	1-1
Figure 1-2: Why Is System Architecture Design Important?	1-2
Figure 1-3: System Design Process	1-3
Figure 1-4: Supporting Technology	1-4
Figure 1-5: System Design Support Efforts	1-5
Figure 2-1: GIS Enterprise Evolution	2-1
Figure 2-2: GIS Workstations/Services	2-2
Figure 2-3: Centralized Computing Environment	2-6
Figure 2-4: Distributed Computing Environment	2-7
Figure 2-5: Departmental GIS	2-8
Figure 2-6: Organizational GIS	2-9
Figure 2-7: Community GIS	2-10
Figure 2-8: System Architecture Options	2-11

Figures, continued	Page
Figure 2-9: Data Migration Road Map	2-15
Figure 2-10: Application Migration Road Map	2-17
Figure 3-1: GIS Applications Network Impact	3-1
Figure 3-2: Types of Networks	3-2
Figure 3-3: Communication Packet Structure	3-3
Figure 3-4: Network Transport Protocol	3-4
Figure 3-5: Client/Server Communications	3-4
Figure 3-6: Client/Server Performance	3-5
Figure 3-7: Shared Network Capacity	3-6
Figure 3-8: Typical 1-MB Map Display	3-7
Figure 3-9: Network Configuration Guidelines	3-8
Figure 3-10: Network Design Guidelines	3-9
Figure 3-11: Shared LAN	3-9
Figure 3-12: ArcIMS Network Performance	3-11
Figure 3-13: Data Download Performance	3-12
Figure 4-1: GIS Multi-tier Architecture	4-1
Figure 4-2: ESRI ArcGIS Software	4-2
Figure 4-3: ArcGIS Desktop Software	4-3
Figure 4-4: ESRI Software Environments	4-4
Figure 4-5: GIS for UNIX Product Architecture	4-5
Figure 4-6: GIS for Windows Product Architecture	4-6
Figure 4-7: Windows Terminal Server Product Architecture	4-8
Figure 4-8: ArcSDE Data Server Product Architecture	4-9
Figure 4-9: ArcIMS Web Services Product Architecture	4-10
Figure 4-10: ArcIMS Component Architecture	4-11
Figure 4-11: Single-tier Platform Configuration	4-12
Figure 4-12: Two-tier Platform Configuration	4-13
Figure 4-13: Three-tier Platform Configuration	4-14
Figure 4-14: All ArcIMS Components in DMZ	4-15
Figure 4-15: All ArcIMS Components in DMZ except Data Server	4-16
Figure 4-16: Web Server in DMZ, Remainder of Configuration on Secure Network	4-17
Figure 4-17: Web and Application Server in DMZ, Remainder on Secure Network	4-18
Figure 4-18: Multiple Web Server Configuration	4-19
Figure 4-19: Web Services with Proxy Server	4-20
Figure 4-20: All ArcIMS Components on Secure Network	4-21
Figure 4-21: ESRI Software Licensing	4-22
Figure 4-22: Ways to Store Spatial Data	4-24
Figure 4-23: Ways to Connect Spatial Data	4-25
Figure 4-24: Ways to Store Spatial Data (RAID)	4-26
Figure 4-25: Ways to Store Spatial Data (Additional RAID)	4-27
Figure 4-26: Ways to Backup Spatial Data	4-29

Figures, continued	Page
Figure 4-27: Ways to Move Spatial Data (Traditional)	4-31
Figure 4-28: Ways to Move Spatial Data (Database Transition)	4-32
Figure 4-29: Ways to Move Spatial Data (Database Replication)	4-33
Figure 4-30: Ways to Move Spatial Data (Disk-level Replication)	4-35
Figure 4-31: Ways to Move Spatial Data (Asynchronous)	4-36
Figure 4-32: New Ways We Can Access Spatial Data	4-37
Figure 5-1: Total GIS Environment	5-1
Figure 5-2: Application Needs Assessment	5-2
Figure 5-3: System Design Review	5-3
Figure 5-4: System Architecture Needs Assessment	5-4
Figure 5-5: Network Communications Overview	5-5
Figure 5-6: User Application Requirements Overview	5-6
Figure 5-7: Central Server with Workstation Clients	5-7
Figure 5-8: Distributed Data and Workstation Processing (WAN)	5-8
Figure 5-9: Choosing a System Configuration	5-19
Figure 5-10: Hardware Component Selection (Centralized Solution)	5-10
Figure 5-11: Hardware Component Selection (Distributed Solution)	5-11
Figure 6-1: Platform Performance Components	6-2
Figure 6-2: System Performance Profile	6-3
Figure 6-3: Dual-Processor SMP Performance	6-5
Figure 6-4: Quad-Processor SMP Performance	6-6
Figure 6-5: Eight-Way SMP Performance	6-7
Figure 6-6: Batch Processing Test Summary	6-8
Figure 6-7: Terminal Servers with File Server or Local Data	6-9
Figure 6-8: Terminal Servers with ArcSDE Data Source	6-10
Figure 6-9: Data Server Loads	6-11
Figure 6-10: ArcGIS Client/Server Network Traffic Analysis	6-12
Figure 6-11: ArcGIS File Server Network Throughput	6-13
Figure 6-12: ArcGIS File Server Network Performance	6-14
Figure 6-13: Data Server Performance Model	6-15
Figure 6-14: GIS Data Server Performance Model	6-16
Figure 6-15: ArcGIS Use Case Parameters	6-17
Figure 6-16: Terminal Server with File Server or Local Data	6-18
Figure 6-17: Terminal Server with ArcSDE Server	6-19
Figure 6-18: Terminal Server Performance Model	6-20
Figure 6-19: GIS Data Server Performance	6-21
Figure 6-20: Data Server Performance Summary	6-22
Figure 6-21: ArcSDE Performance Model Comparison	6-23
Figure 6-22: ArcIMS Platform Architecture	6-24
Figure 6-23: Map Services Performance Profile	6-25
Figure 6-24: ArcIMS Server Components	6-26

Figures, continued	Page
Figure 6-25: Multi-threaded ArcIMS Service Engines	6-27
Figure 6-26: ArcIMS Virtual Servers	6-28
Figure 6-27: Map Services Performance Profile	6-29
Figure 6-28: ArcIMS Performance Sizing Model	6-30
Figure 6-29: ArcIMS Server Loading Model	6-31
Figure 6-30: Optimized Service Agent Configuration	6-32
Figure 6-31: Web Server Configuration	6-33
Figure 6-32: Web Mapping Services Performance Profile	6-34
Figure 6-33: Web Mapping Services Performance Tuning Guidelines	6-35
Figure 6-34: Performance Testing Pitfalls	6-36
Figure 7-1: ArcInfo Platform Performance History	7-2
Figure 7-2: UNIX Platform Performance History	7-4
Figure 7-3: Theory of Relative Performance	7-5
Figure 7-4: How Do We Measure Change?	7-6
Figure 7-5: SPEC2000 Benchmark Suites	7-6
Figure 7-6: Platform Performance Chart Overview	7-8
Figure 7-7: Published SPEC Benchmark Results	7-9
Figure 7-8: ArcSDE Server Sizing Example	7-10
Figure 7-9: Platform Sizing Models	7-11
Figure 7-10: ArcInfo and ArcView Platform Recommendations	7-11
Figure 7-11: Workstation Platform Recommendations	7-12
Figure 7-12: Windows Terminal Server Sizing (File Server Data Source)	7-14
Figure 7-13: Windows Terminal Server Sizing (ArcSDE Data Source)	7-15
Figure 7-14: ArcIMS Spatial Server Sizing	7-16
Figure 7-15: ArcIMS Data Server Processing Loads	7-17
Figure 7-16: ArcIMS Map Service Time	7-18
Figure 7-17: Workgroup Data Server Sizing	7-20
Figure 7-18: Enterprise Data Server Sizing	7-21
Figure 8-1: System Architecture Deployment Strategy	8-1
Figure 8-2: Systems Integration Management	8-2

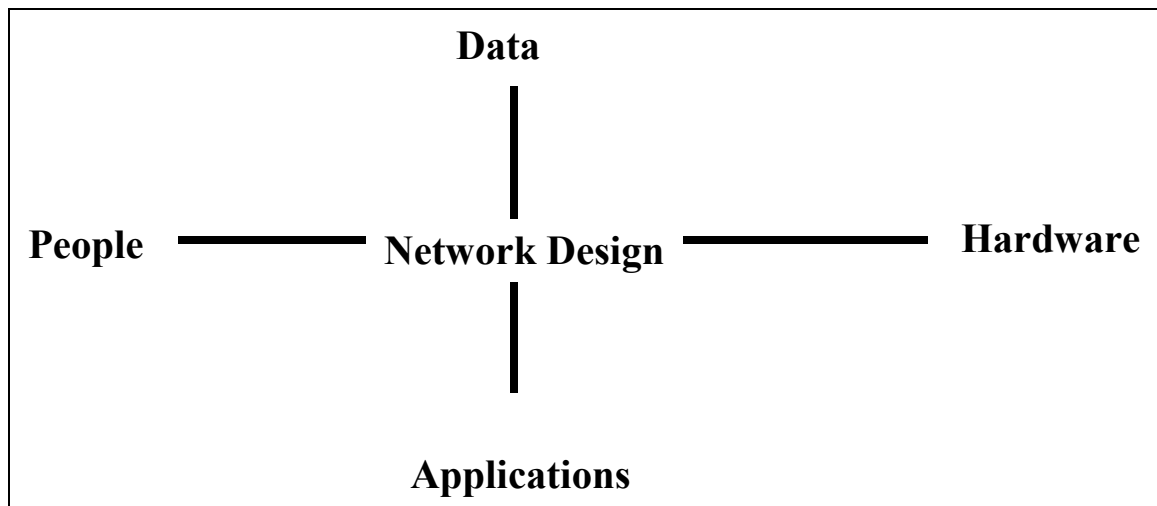
1.0 System Design Process

The purpose of this document is to share a system design strategy that promotes successful selection of geographic information system (GIS) enterprise architecture solutions. Guidelines include appropriate rationale and logic to support a system that will satisfy initial performance needs for most of our customers. Once the initial implementation is operational, the system environment can be further adjusted to fit specific user needs.

1.1 What Is System Architecture Design?

System architecture design is a process developed by ESRI to promote successful GIS implementations. This process supports existing infrastructure requirements and provides specific recommendations for hardware and network solutions based on existing and projected user needs. Application requirements, data resources, and people within an organization all are important in determining the optimum hardware solution. The ESRI system architecture design process provides specific hardware architecture and specifications based on identified user operational requirements.

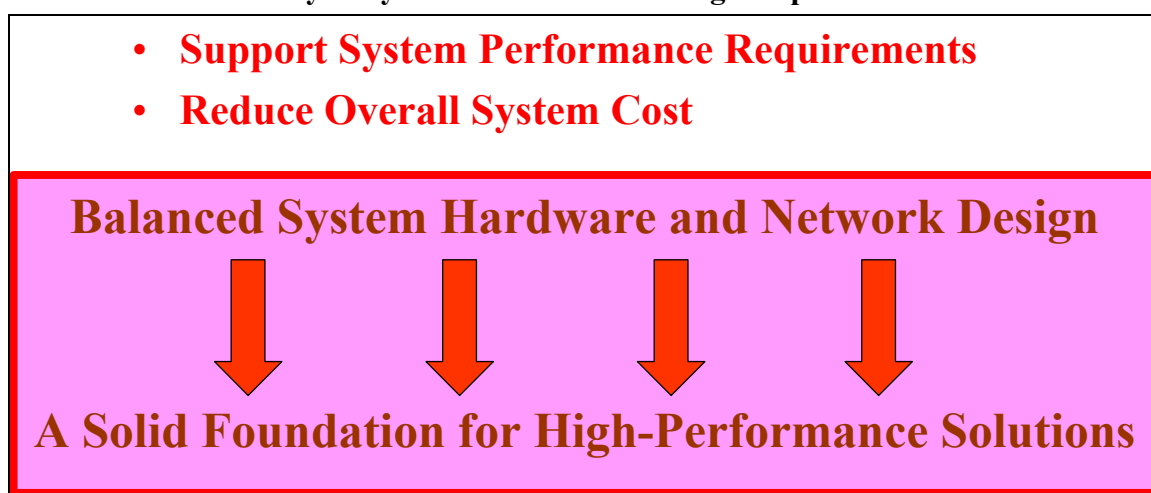
Figure 1-1
What Is System Architecture Design?



1.2 Why Is System Architecture Design Important?

A distributed computer environment must be designed properly to support user performance requirements. The weakest system component will limit performance. The system architecture design process develops specifications for a balanced system. Investment in hardware and network components based on a balanced system load model will provide the highest possible system performance at the lowest overall cost.

Figure 1-2
Why Is System Architecture Design Important?

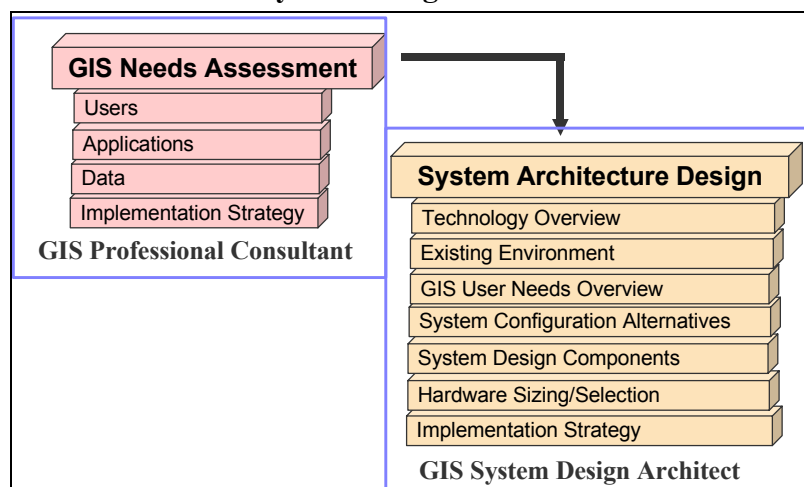


1.3 System Design Process

The GIS system design process includes a GIS needs assessment and a system architecture design review. The system architecture design review is based on requirements identified in the GIS needs assessment.

- **GIS Needs Assessment.** The GIS needs assessment includes a review of user work requirements and identifies where GIS applications can improve user productivity. This assessment identifies GIS application and data requirements, and an implementation strategy for supporting GIS user needs. A GIS professional consultant familiar with current GIS solutions and customer business practices can facilitate this planning effort.
- **System Architecture Design.** The system architecture design assessment is based on user requirements identified by the GIS needs assessment. The customer must have a clear understanding of their GIS application and data requirements before they are ready to develop system design specifications. System implementation strategies should identify hardware purchase requirements "just in time" to support user requirements.

Figure 1-3
System Design Process



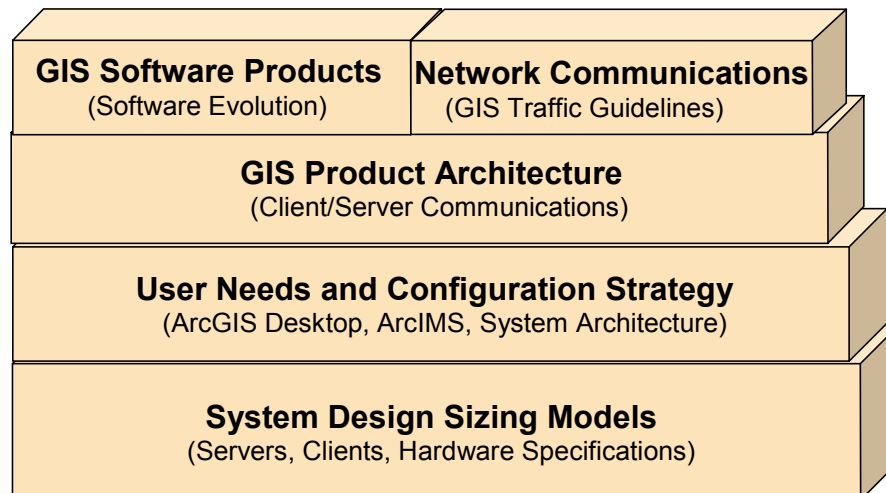
The ESRI system architecture design assessment begins with a technology exchange. The technology exchange provides a foundation for client support during the design process. Client participation is a key ingredient in the design process. The design process includes a review of the existing computer environment, GIS user requirements, and system design alternatives. The system design tools provided by ESRI translate user performance requirements to specific platform specifications. An integrated implementation strategy is developed to support GIS deployment milestones.

1.4 Supporting Technology

Distributed GIS solutions include integration of a variety of vendor products. Each product implements a component technology required to support the enterprise solution. Integration of this multi-vendor environment is made possible through voluntary compliance with generally accepted industry interface standards.

A general understanding of the primary supporting technologies associated with a GIS enterprise solution provides a foundation for supporting the system architecture design process. **Figure 1-4** identifies the key technologies supporting a distributed GIS environment.

Figure 1-4
Supporting Technology



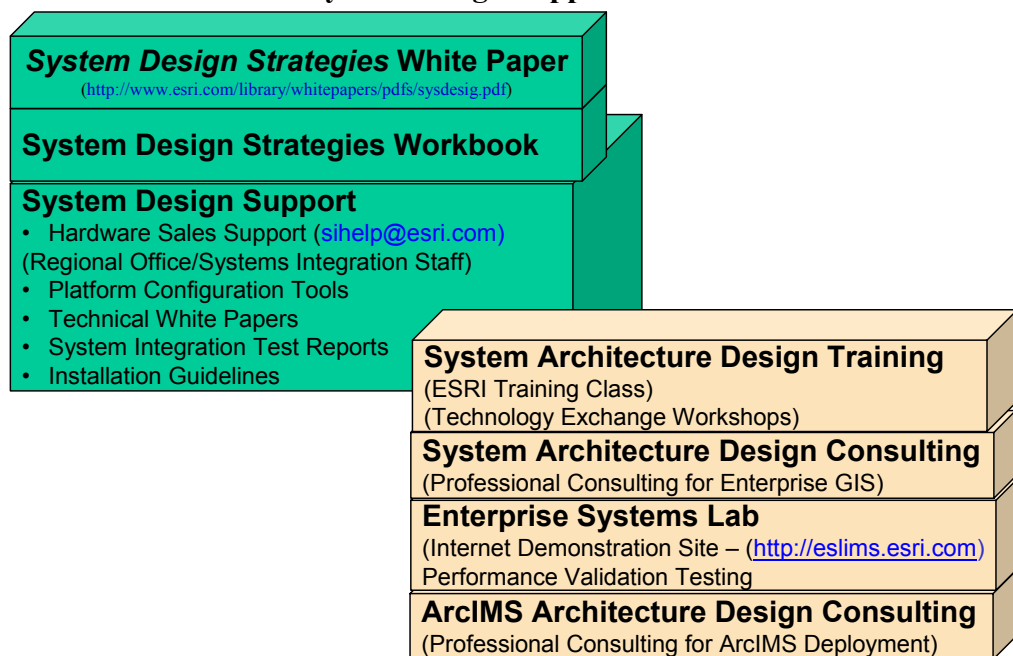
- **GIS Software Products.** A variety of GIS software solutions were developed by ESRI over the past 30 years. Each product solution was developed to support specific user requirements. **Section 2** describes how these solutions fit together to support the growing needs of the GIS user community.
- **Network Communications.** Network communications provide the connectivity for a powerful GIS enterprise solution. A fundamental understanding of network communications helps users develop better applications and more efficient shared data resources. **Section 3** provides an overview of network communication concepts and identifies GIS design standards for successful distributed GIS solutions.
- **GIS Product Architecture.** An enterprise GIS solution includes a variety of ESRI and third party vendor products. These products must interface with one another to support an integrated GIS solution. **Section 4** provides an overview of the communication protocols required to support distributed GIS solutions.
- **User Needs and Configuration Strategy.** The system architecture design process starts with results from the GIS user needs assessment, and provides a methodology to translate user performance requirements to appropriate GIS architecture and platform specifications. **Section 5** provides an overview of the user needs assessment and analysis required to develop an appropriate distributed GIS architecture design strategy.
- **System Design Sizing Models.** A fundamental contribution provided by the ESRI system architecture design consultant is the ability to translate user desktop application performance requirements to specific hardware and network specifications for an enterprise-wide design solution. **Section 6** discusses the technical assumptions supporting the ESRI sizing models, identifying fundamental performance relationships between the distributed computing platform components supporting the GIS hardware solution. **Section 7** identifies how to

apply these sizing models in the real world of rapidly changing hardware technologies, how to address system architecture design in the real world, including some simple sizing tools based on the performance sizing models for identifying required platform specifications.

1.5 System Design Support Efforts

The ESRI Systems Integration Department was established in November 1990 with the charter to promote successful GIS implementations. Several initiatives were developed over the years to simplify the design of successful GIS environments and reduce implementation risk. The following services are maintained to support the needs of the ESRI user community and promote effective enterprise GIS solutions.

Figure 1-5
System Design Support Efforts



- **System Design Strategies White Paper.** The *System Design Strategies* white paper defines the technical issues associated with design of an effective enterprise GIS solution and includes specific sizing guidelines to support hardware selection. ESRI customers, hardware consultants, and GIS technical staff can use this document as a guide for designing future distributed GIS solutions. The guidelines in this document are also useful in understanding existing hardware-related performance issues and identifying proper solutions. The document is updated twice each year and is accessible over the Internet at the following URL: <http://www.esri.com/library/whitepapers/pdfs/sysdesig.pdf>.
- **System Design Strategies Workbook.** The System Design Strategies workbook is a PowerPoint presentation used in a variety of ESRI systems integration support efforts. Changes are made on a project-by-project basis to reflect software and technology evolution. The workbook is used as a source document supporting ESRI system design consulting and training efforts.

- **System Design Support.** The ESRI Systems Integration Department provides on-line hardware marketing support through the sihelp@esri.com email alias. Simple questions related to hardware selection and sizing requirements can be addressed through this site. An Intranet Web site is maintained providing technical information for ESRI marketing staff. This information is replicated to the ESRI Partners and International Distributors Intranet site. This information includes simplified platform configuration tools, published Systems Integration test reports, and materials supporting our regional installation services staff. These materials are prepared and maintained for public release in support of the ESRI user community.
- **System Architecture Design Training.** A two-day System Architecture Design for GIS training class is available through the ESRI Learning Center. This class provides an in-depth review of the material presented in this white paper. A class brochure is included as **Attachment B**. A one-day System Architecture Design technical workshop is held the weekend prior to the annual ESRI User Conference providing a full technical overview of the same material presented in the training class.
- **System Architecture Design Consulting.** Professional system architecture design consulting services are available for ESRI customers. These services can be used to resolve performance problems with existing environments and develop infrastructure support strategies for future successful GIS implementations. These services provide IT professionals with the information they need to support GIS users within their organizations. A generic statement of work for these services is included as **Attachment A**.
- **Enterprise Systems Lab.** Our Enterprise Systems Lab supports test and evaluation of ESRI and third party products focused on system design and performance for enterprise GIS environments. The Lab maintains an Internet Web site, providing general public access to demos of ESRI products supported by Microsoft Windows Terminal Servers and Citrix MetaFrame technology. The site also includes performance demonstrations of ArcIMS map services and provides technical information on the supporting technologies. This site can be accessed over the Internet at the following URL: (eslims.esri.com). User name and password for accessing application demos included on this site can be obtained from your marketing representative or from the Systems Integration staff. This site demonstrates a very simple enterprise hardware solution for supporting remote GIS users over an Intranet, or the public Internet, with amazing performance.

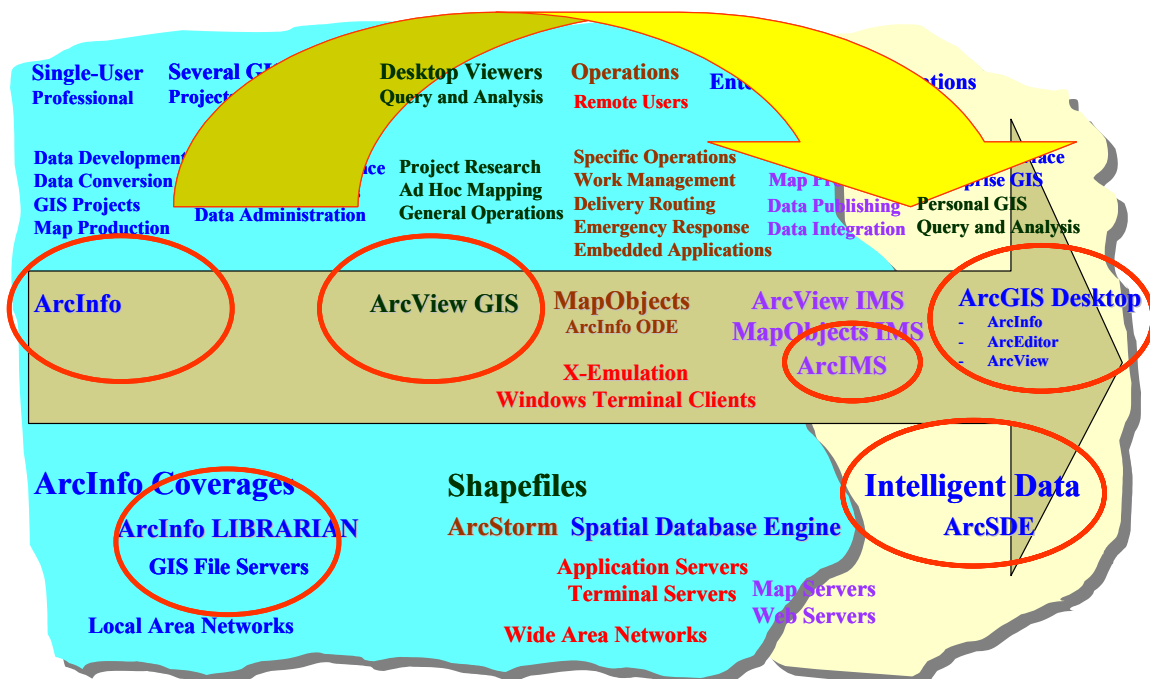
Several activities are included in the ESRI Annual International Users Conference in San Diego each summer. System Architecture Design technical workshops provide a technology overview and system design strategy refresher. System design architects are available at the Systems Integration island in the ESRI showcase area to provide customer design support during the conference. A System Implementation Papers track provides an opportunity for ESRI users to present their GIS implementation experiences.

2.0 ESRI Software Evolution

For more than 30 years, ESRI has continued to develop an evolving GIS software supporting functional requirements identified by the GIS user community. ESRI software developers leverage the latest computer hardware and software technology to produce the world's best GIS technology. ESRI aligns its resources to provide the best software and services in the GIS marketplace.

This section provides an overview of ESRI software and associated product technologies. Understanding the primary role for each member of the ESRI software family will help define current needs and provide a clear vision for migration to a successful enterprise GIS solution. **Figure 2-1** provides an overview of the ESRI software history and the associated third party technologies supporting effective GIS enterprise evolution.

Figure 2-1
GIS Enterprise Evolution

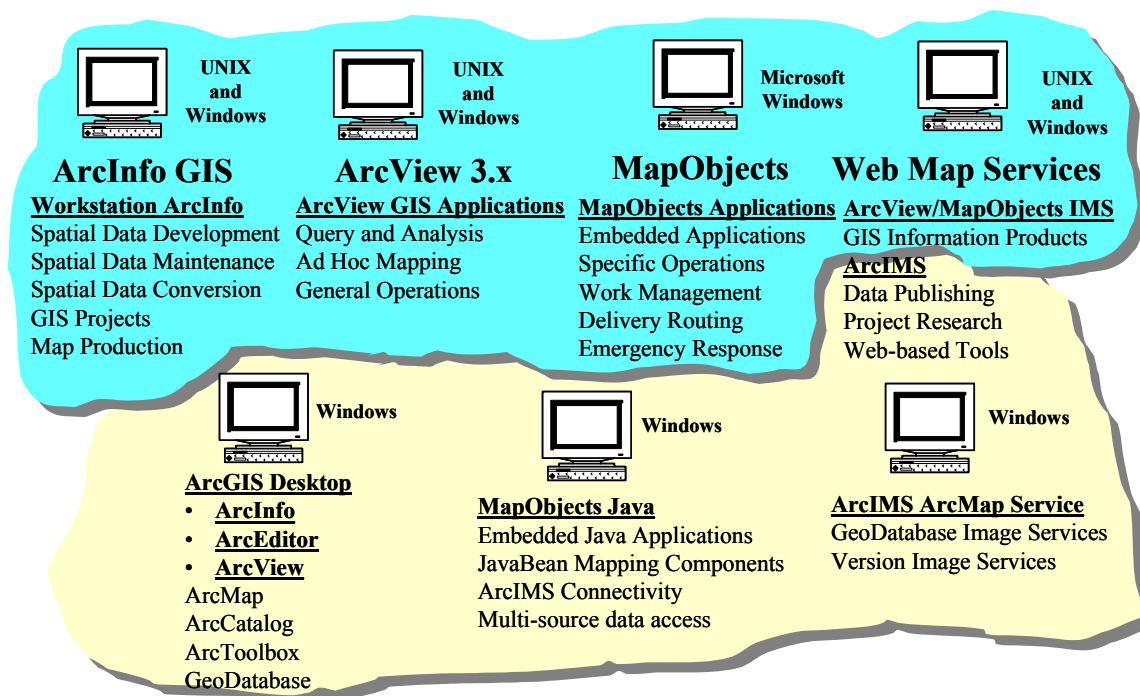


Many ESRI customers have developed effective enterprise solutions with the Workstation ArcInfo and ArcView GIS software provided in the 1990s. The new ArcGIS technology provides operational capabilities that were not available with the older technology. Many ESRI customers are in the process of migrating their data and applications from the old file-based technology to the current ArcGIS object-based geodatabase architecture. New customers are supporting enterprise GIS solutions directly from the ArcGIS architecture.

2.1 Desktop Workstation Environment

The traditional environment for GIS applications is on the user desktop. The primary ESRI desktop applications include ArcInfo, ArcView GIS, and custom applications developed using MapObjects software. ArcInfo 8 includes both Workstation ArcInfo and Desktop ArcInfo components. Desktop ArcEditor and ArcView were introduced with the Version 8.1 release. Web Services extends user access to GIS information products to every desktop.

Figure 2-2
GIS Workstations/Services



- **ArcGIS Software.** ArcGIS is a scalable family of software comprising a complete geographic information system, built on industry standards, that is rich in functionality and works out of the box. Organizations deploy the software of ArcGIS—ArcView, ArcEditor, ArcInfo, ArcSDE, and ArcIMS—in a configuration appropriate for their needs.

ArcGIS is used for the creation, management, integration, analysis, display, and dissemination of spatial data. Strong visualization, editing, and analysis, along with advanced data management, distinguish the ArcGIS software family as the leading GIS software.

ArcInfo License. ArcInfo is the complete GIS data creation, update, query, mapping, and analysis system. Professionals use ArcInfo for spatial data automation since it includes the most comprehensive collection of GIS tools available. As part of the ArcGIS software family, ArcInfo includes all the functionality of ArcView and ArcEditor and adds the advanced geoprocessing and data conversion capabilities that make it the de facto standard for GIS.

ArcEditor License. ArcEditor is a state-of-the-art GIS data visualization, query, and creation solution. Designed for the Windows desktop, ArcEditor can create and edit all ESRI-supported vector data formats including shapefiles, coverages, personal geodatabases, and multi-user geodatabases. As part of the ArcGIS software family, ArcEditor contains all the capabilities users find in ArcView while adding capabilities for managing geodatabase schema and advanced editing of geodatabases and coverages.

ArcView License. ArcView is the world's most popular desktop GIS and mapping software, with more than 500,000 copies in use worldwide. ArcView provides geographic data visualization, query, analysis, and integration capabilities along with the ability to create and edit geographic data.

ArcView is designed with an easy-to-use, Windows-like user interface and includes VBA for customization. ArcView consists of three desktop applications: ArcMap, ArcCatalog, and ArcToolbox. ArcMap provides data display, query, and analysis. ArcCatalog provides geographic and tabular data management, creation, and organization. ArcToolbox provides basic data conversion.

ArcSDE Database. ArcSDE is the tool that allows you to store and manage spatial data in your chosen DBMS. ArcSDE is open; it works with a variety of different databases—including Oracle, Informix, IBM DB2, and Microsoft SQL Server—that scale from work groups to large enterprise databases.

ArcSDE plays a fundamental role in a multi-user GIS. With ArcSDE, your ArcGIS software (ArcInfo, ArcEditor, ArcView, and ArcIMS) can work directly with spatial data managed in your DBMS. ArcSDE also works as an application server, delivering spatial data to many kinds of applications and serving spatial data across the Internet.

ArcIMS Web Services. ArcIMS software is the foundation for distributing GIS data and applications on the Internet. By providing a common platform for exchanging and sharing GIS resources, ArcIMS provides unique opportunities to leverage data from within the organization and to integrate information from other agencies. Key features of ArcIMS include data integration, standards-based communication, the Internet-enabling technology for ArcGIS, easy-to-use framework, a multi-tier architecture, support for a wide range of clients, highly scalable server architecture, and a wide range of GIS capabilities. ArcIMS supports Windows and UNIX platforms.

- **Workstation ArcInfo.** Workstation ArcInfo provides an exhaustive set of GIS tools representing over 30 years of software development effort, making this product the leading GIS professional software available today. Workstation ArcInfo includes a full set of tools to support spatial data development, maintenance, and conversion activities. For many users, this has been the primary software used by GIS professionals to support geographic studies, analysis, and map production. Workstation ArcInfo is supported on both UNIX and Windows workstations.
- **ArcView 3.x.** ArcView GIS 3.x is a user-friendly desktop GIS software, initially developed to support a growing population of query and analysis users within the ArcInfo community. ArcView GIS is able to directly read ArcInfo coverages and to store spatial views in a new

GIS shapefile data format. ArcView GIS quickly became very popular, significantly expanding the number of GIS users. It was not long before the number of ArcView GIS users outnumbered ArcInfo users. GIS professionals continue to use ArcInfo to create and maintain spatial data resources and to support high-technology spatial analysis and map products. ArcView GIS provides a simple GIS data access tool supporting general office query and analysis functions and producing simple ad hoc map products. A variety of ArcView GIS extensions have been developed since the initial ArcView GIS release to support the needs of the growing desktop user market, and ArcView GIS is now the leading desktop application throughout the GIS community. ArcView 3.x is supported on Microsoft Windows and UNIX platforms.

- **MapObjects.** MapObjects was developed using new Microsoft programming standards for Windows environments. ESRI took advantage of this new object-oriented environment with the release of MapObjects software. With MapObjects, developers can include map products within standard Microsoft application environments. This opened a new world of product opportunities and spawned rapid expansion of the ESRI developer community. MapObjects became a new standard for incorporating GIS in vertical market product solutions. MapObjects is able to read and display ArcInfo coverages, ArcInfo LIBRARIAN files, shapefiles, and Spatial Database Engine (ArcSDE) layers data sources. This is an optimum programming environment to support custom GIS Windows-based applications.

MapObject Java provides JavaBean mapping components to support Java programmers.

- **ArcInfo LIBRARIAN.** As the number of GIS users increase, software management tools are needed to support data maintenance operations. Users update features on a spatial coverage by copying the layer to memory on their workstation and making the appropriate changes, then replacing the updated layer on the GIS data server. Conflicts can occur when multiple users are responsible for updating data on the same layers.

ArcInfo includes a LIBRARIAN database management module that provides a controlled multiple-user data maintenance environment and supports more efficient management of spatial data resources. ArcInfo LIBRARIAN is a file-based storage solution that provides a continuous tile structure for all data layers on the GIS data server within a single database environment. Users are able to extract individual tiles within a coverage for editing, and the associated tiles within ArcInfo LIBRARIAN are “write locked” to avoid data changes during the update process. Once editing is complete, the updated tile is returned to the ArcInfo LIBRARIAN database and the associated tile is released for other users to update. ArcInfo LIBRARIAN provides an effective and simple way for several people to maintain and manage large continuous GIS data libraries.

- **ArcStorm Data Server.** Large enterprise GIS operations identified a need for enhanced transaction management of ArcInfo spatial data resources. ArcStorm software satisfies this need by providing feature-level transaction management, transaction history, and commit logic to support large GIS operational environments. ArcStorm includes several server processes that support transaction management and check-in/checkout processing. An ArcStorm weather service daemon manages the ArcStorm data files and generates additional processes as required for user transactions. ArcInfo provides functions to support maintenance of the ArcStorm libraries. ArcView GIS is able to directly read ArcStorm data

using standard NFS protocol (CIFS for Windows servers). Client workstation processes support all read operations.

- **Database Integration.** Spatial features within ArcInfo data coverages can be linked to other tabular attribute data throughout the organization. ArcInfo maintains a set of database integrator modules (application program interfaces [APIs]) that provide direct access to attribute data in Oracle, Informix, Sybase, and OpenIngres relational database management systems (RDBMSs). The Windows version of ArcInfo, ArcView GIS, and MapObjects are ODBC-compliant applications, supporting access to DBMS environments through standard ODBC driver connectivity. This connectivity has improved and expanded GIS access to additional ODBC-compliant database environments to include Microsoft SQL Server, IBM DB2, and a host of other data sources.

A variety of commercial gateways support online access to mainframe and AS400 legacy data business systems. Commercial replication services and administrative data transfers provide GIS user access to data from the more difficult proprietary data sources.

2.2 GIS Configuration Alternatives

GIS environments commonly begin with single-user workstations at a department level within the organization. Many organizations start with a single GIS manager, and evolve from a department level to an enterprise operation. This was common through the early 1990s, as many organizations worked to establish digital representation of their spatial data. Once these data are available, organizations expand their GIS operations to support enterprise business operations.

Spatial data are presented as layers of graphic lines, points, or polygons stored in a proprietary file format as GIS layers, similar to traditional Mylar sheets. These layers are overlaid by the client application to render a map image. Spatial data can be stored in proprietary file formats or as spatial feature types within a database table. Standard file formats include coverages and shapefiles. LIBRARIAN provides a tiled coverage format supporting data maintenance operations. ArcStorm includes a variety of server processes to manage database integrity with associated tabular attribute data sources.

Data can be shared between users in a variety of ways. Most organizations today have user workstations connected to local area networks (LAN) environments, and locate shared spatial data on dedicated server platforms. User applications connect to shared data sources to support GIS operations.

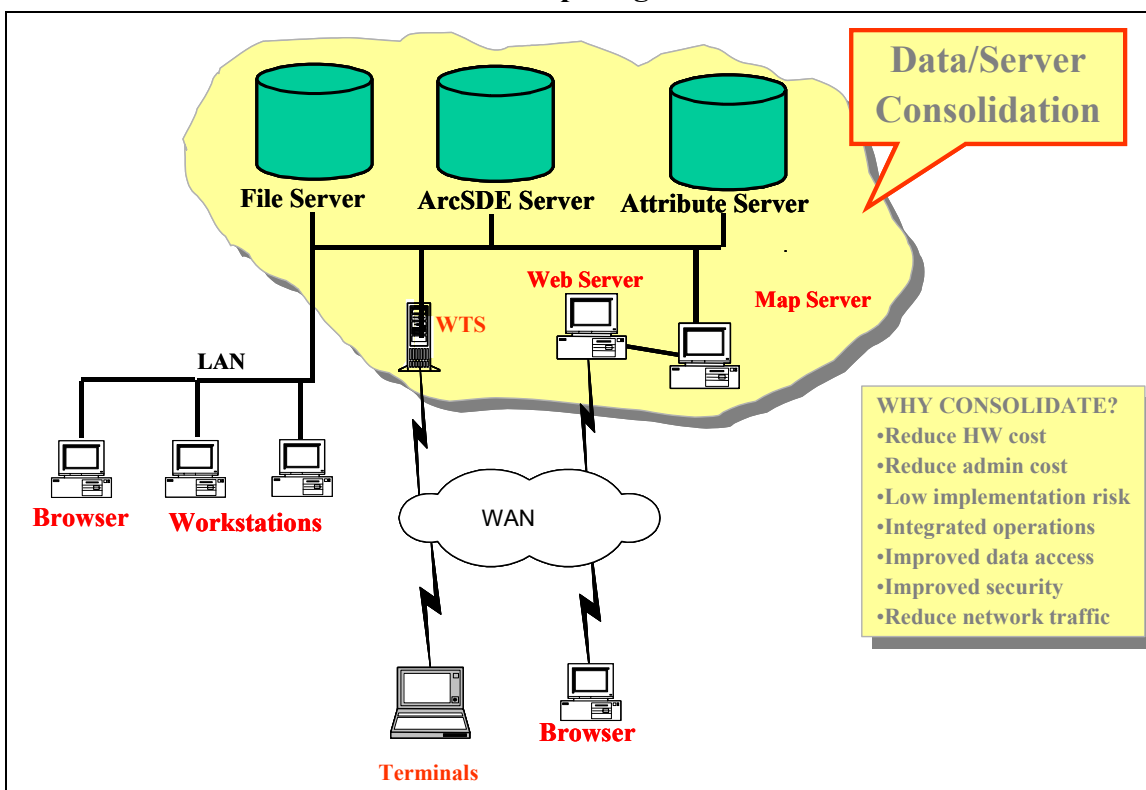
- **Centralized Data Configuration Alternative**

The most simple system architecture is supported by a central GIS database. A central database architecture supports one copy of the production database environment, minimizing administrative management requirements and ensuring data integrity. **Figure 2-3** provides an overview of a central data configuration architecture.

GIS desktop applications can be supported on user workstations located on the central LAN, each with access to central GIS data sources. Data sources can include GIS file servers, ArcSDE database servers, and related attribute data sources.

Remote user access to central data sources can be supported by central Windows Terminal Server farms, providing low bandwidth display and control of central application environments.

Figure 2-3
Centralized Computing Environment



Centralized application farms minimize administration requirements and simplify application deployment and support throughout the organization. Source data is retained within the central computer facility, improving security and simplifying backup requirements.

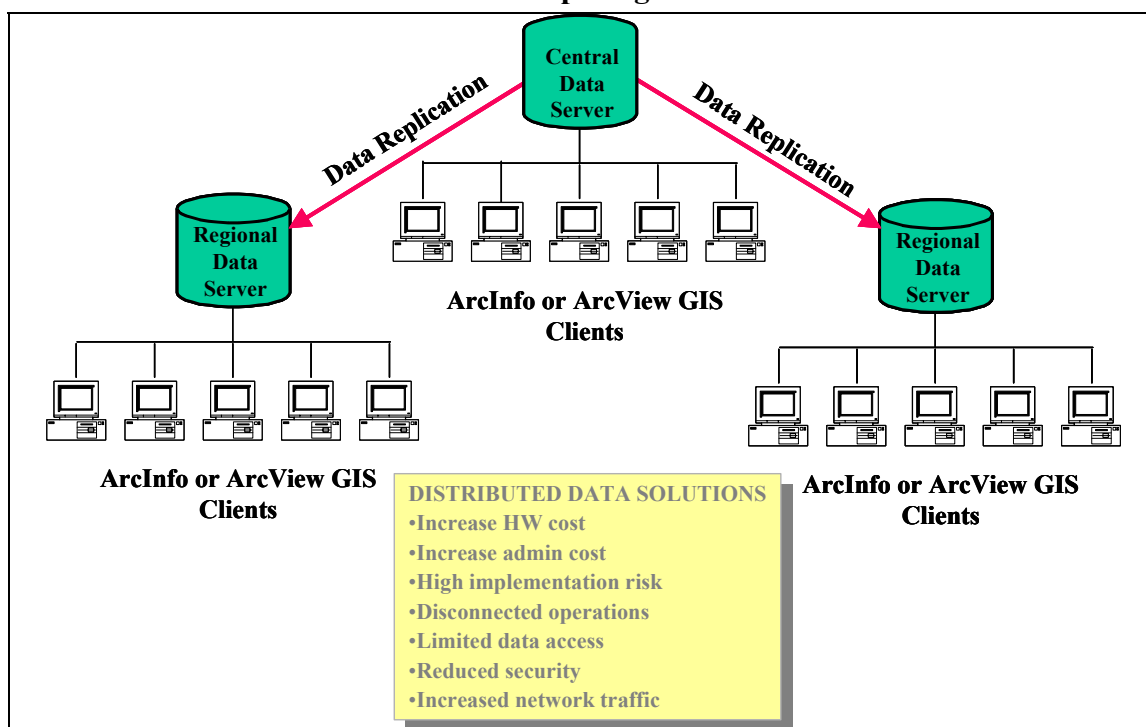
A variety of ArcIMS map services can support standard browser clients throughout the organization. Web mapping services support low bandwidth access to published GIS information products and services.

Distributed computing technology today can support consolidated architectures at a much lower risk and cost than similar distributed environments. For this reason, many organizations are in the process of consolidating their data and server resources. GIS can benefit from consolidation for many of the same reasons experienced by other enterprise business solutions. Centralized GIS architectures are generally easier to deploy, manage, and support than distributed architectures, and provide the same user performance and functionality.

■ Distributed Data Configuration Alternative

Distributed solutions are supported by replicated copies of the data at the remote locations, establishing local processing nodes that must be maintained consistent with the central database environment. Data integrity is critical in this type of environment, requiring controlled procedures with appropriate commit logic to ensure changes are replicated to the associated data servers.

Figure 2-4
Distributed Computing Environment



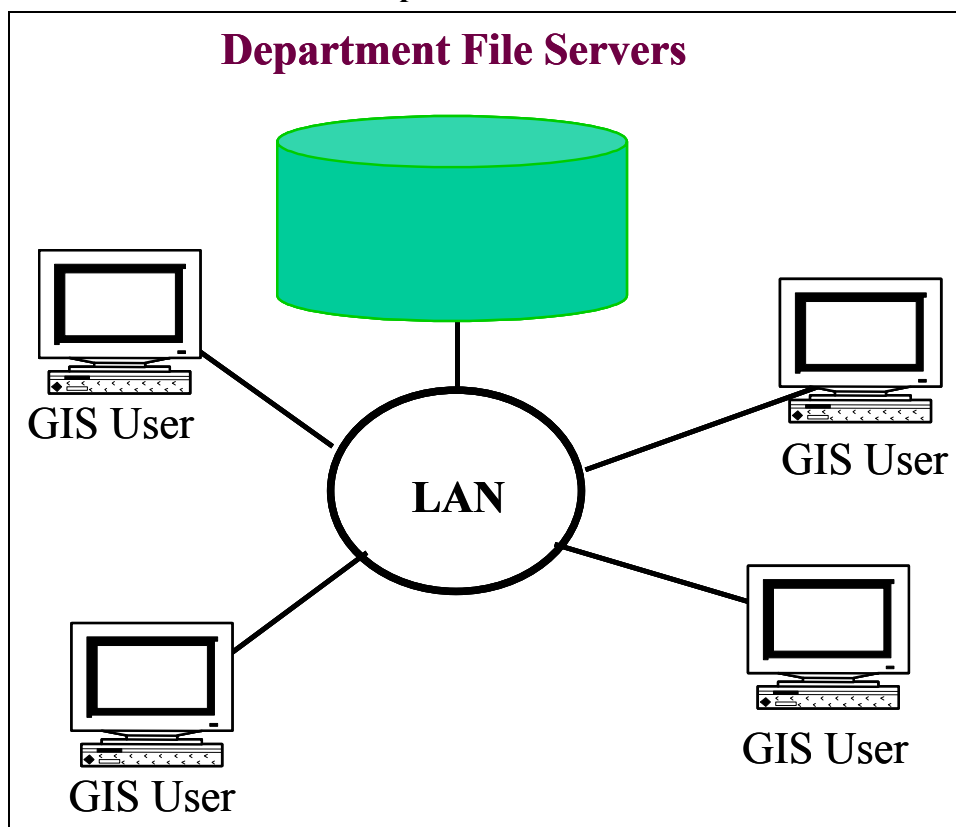
Distributed database environments generally increase initial system cost (more hardware and database software requirements) and require additional on-going system administration and system maintenance requirements. Distributed solutions are provided to support specific user needs, and generally increase system complexity, cost, and system deployment timelines.

In many cases, standard database solutions do not support replication of spatial data. GIS users with distributed database requirements must modify their data models and establish procedures to administratively support data replication. DBMS vendors have recognized the necessity to provide a spatial data replication option, and new releases are supporting some levels of spatial data replication. ESRI is also working on a solution to support geodatabase synchronization between multiple ArcSDE servers. The complexity of current geodatabase environments have complicated implementation of an automated efficient spatial replication solution.

2.3 Organizational GIS Evolution

GIS implementations grew in size and complexity throughout the 1990s. GIS started on the user desktop, and evolved to support GIS operations with department-level file servers. A majority of the GIS community is currently supported by department-level GIS architectures. **Figure 2-5** provides an overview of a department-level GIS architecture.

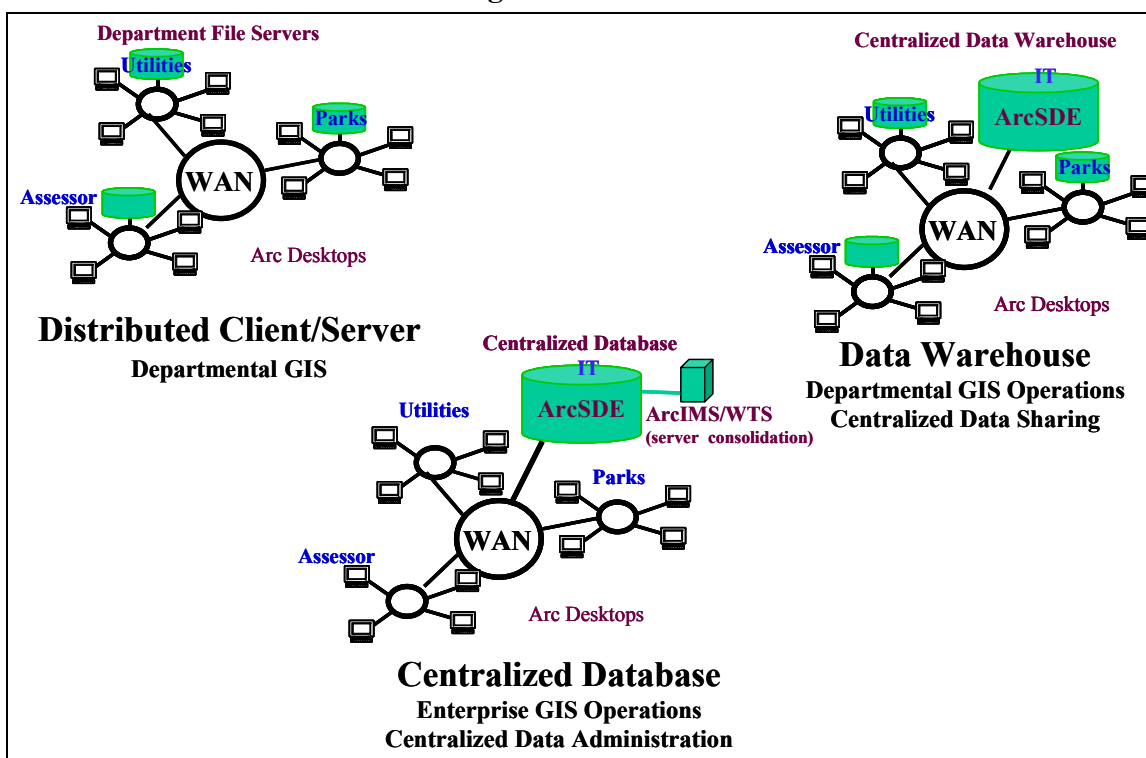
Figure 2-5
Departmental GIS



As GIS within a typical organization evolved over time, several departments would have local GIS operations which require data resources from other departments within the organization. The organization wide area network (WAN) became a way of sharing data between department servers. Data standardization and integrity issues would surface within the organization, since data were developed and managed from different department level sources.

The initial Spatial Database Engine (SDE) release in the mid 1990s supported enterprise data warehouse operations, and many organizations combined department GIS data resources on a central ArcSDE data warehouse. GIS staff were established within IT departments to integrate enterprise-wide GIS data resources within the data warehouse providing integrated data standards across the organization. The common data warehouse provided a reliable shared GIS data source for departments throughout the organization. This was a very common migration path for local government GIS operations. **Figure 2-6** provides an overview of organizational architecture alternatives.

**Figure 2-6
Organizational GIS**



The electric and gas industry started using GIS in the early 1990s to support management of their power distribution facilities. Most of these implementations were supported by a central database. Remote users were supported with terminal access to application compute servers (terminal servers) located in the central computer facility with the GIS database.

Many organizations today are migrating their department file-based GIS database environments to a single central enterprise ArcSDE database, and supporting terminal client access to these consolidated server environments. Departments retain responsibility for their data resources, updating and maintaining these data through terminal access to central ArcGIS applications. The central IT computer center supports general administration tasks, such as data backups, operating system upgrades, platform administration, etc. Users throughout the organization are provided browser access to published ArcIMS services over the Intranet. The complexity and sophistication of the ArcSDE geodatabase make central administration and support the most productive alternative for most organizations.

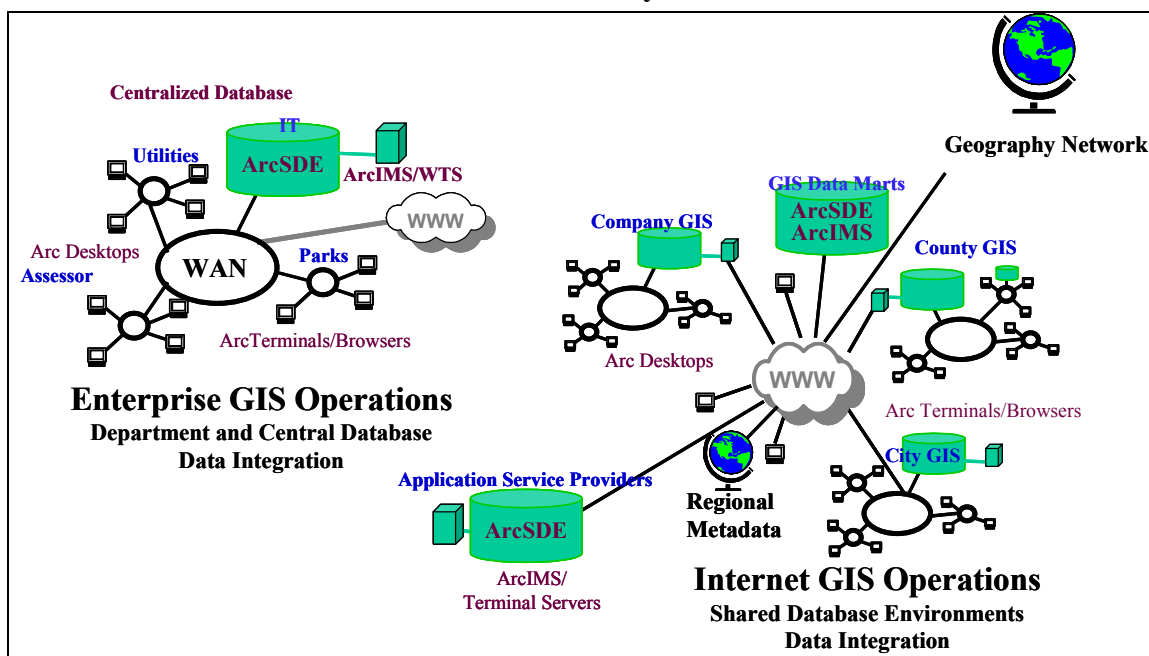
2.4 Community GIS Evolution

Year 2000 introduced a growing Internet awareness, demonstrating the tremendous value of sharing information between organizations and nations. Internet access was extended from the workplace to the home, rapidly expanding the user community. Communities and companies developed and deployed services to customers over the Internet. The Internet provided opportunities for organizations to share data and services between organizations. Users had access to data and services from a multitude of organizations through the Internet.

ESRI introduced the Geography Network, a metadata search engine that collected information about GIS data and provided direct Internet links between the customer and the data or service provider. ArcIMS provided a way for organizations throughout the world to share GIS data and services. The Geography Network provides a foundation to bring GIS data and services together, supporting a rapidly expanding infrastructure of worldwide communities sharing information about the world we all live in. Promotion of data standards and improved data collection technologies unlock enormous possibilities for GIS information products to help us better understand and improve our world.

GIS data resources are expanding exponentially. A few years ago, GIS data servers would seldom require a database over 25 to 50 Gigabytes in size. Today we are seeing a growing number of GIS implementations with ArcSDE data servers supporting database environments with over a Terabyte of GIS data.

Figure 2-7
Community GIS



State-level agencies are consolidating data to support municipalities and commercial activities throughout their states. National agencies are consolidating data to support their user requirements and sharing data between state and national communities. Community-level data marts are being established to consolidate GIS data resources and support Internet data sharing to organizations throughout county and state regional areas.

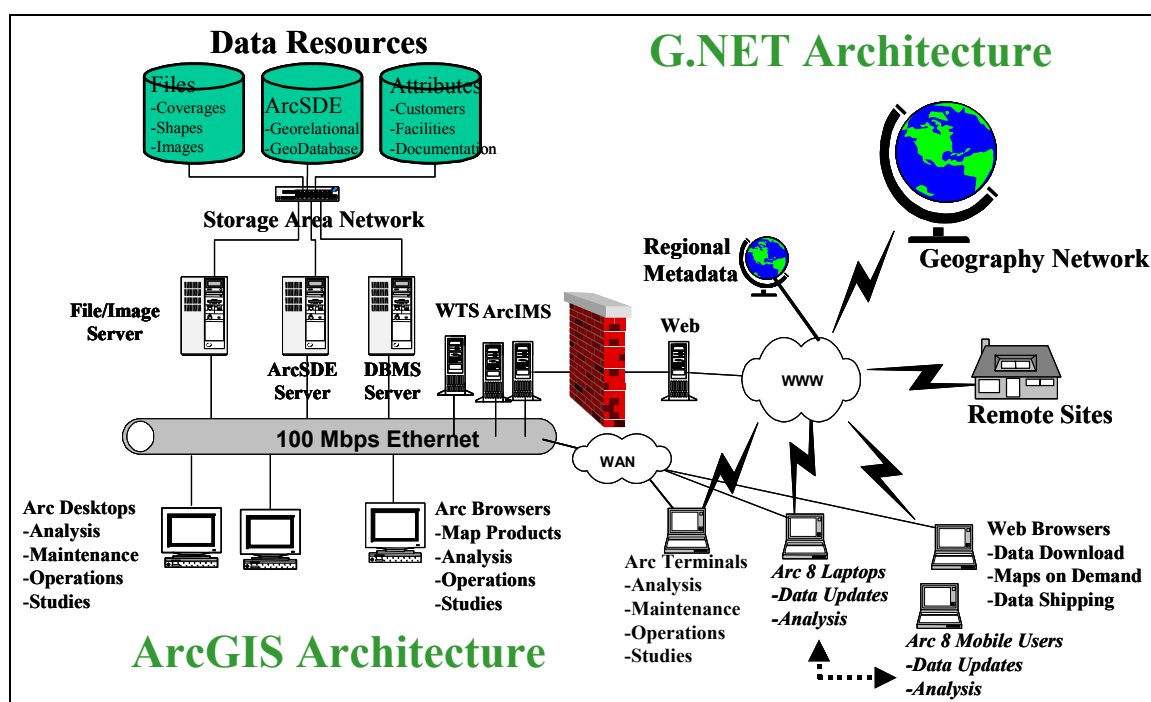
Many organizations are outsourcing their IT operations to commercial Internet Service Providers (ISPs). Application Service Providers (ASPs) support organizations with IT administration, providing opportunities for smaller organizations to take advantage of high-end GIS database and application solutions to support their business needs. State governments are hosting applications and data for smaller municipalities throughout their states so the smaller communities can take advantage of GIS technology in supporting their local operations.

Regional Geography Networks (G.Net) sites support sharing data throughout regional areas and within large state and federal agencies. The ArcIMS 4 software provides a metadata search engine that can be used by any organization to share their data and support their community operations. Cities can establish metadata sites to promote local commercial and public interests. States can consolidate metadata search engines for sharing data and services with municipalities throughout the state. Law enforcement can establish search engines to support national data sets. Businesses can establish metadata search engines to support distributed operational environments.

2.5 GIS Architecture Alternatives

Current GIS technology is available to support a rapidly expanding spectrum of GIS user needs. Solutions are supported by ESRI products integrated with a variety of vendor-enabling technologies.

Figure 2-8
System Architecture Options



Data storage and data management technologies are growing in importance, as organizations continue to develop and maintain larger volumes of GIS data. Direct connect storage solutions are evolving to Storage Area Networks (SANs) enhancing IT options for managing a large volume of data resources.

Supported data servers include file servers, ArcSDE servers, and attribute data servers. Desktop ArcGIS applications are supported on local workstation clients. These same applications can be supported for remote clients using Windows Terminal Servers. Windows Terminal Servers can also support local terminal clients.

ArcIMS provides published map servers to Web browser clients throughout the organization. ArcIMS can also provide public map services to clients across the Internet. ArcGIS clients are able to connect to ArcIMS as intelligent browser clients, enabling connection to unlimited data resources through the Internet Geography Network, as well as organization resources served through ArcIMS services. Users can access applications from home or from other locations. Mobile ArcGIS users can create red-line datasets in the field and submit them to a central ArcIMS site for final processing. ArcGIS desktop applications can include ArcIMS services as data sources with local ArcSDE and file servers, expanding desktop map production and analysis to include available Internet data sources. The “ArcGIS architecture” is supported by a combination of ArcSDE data sources, ArcIMS Web services, and ArcGIS desktop technology.

Implementation of the Geography Network metadata search engines, along with local metadata sites supported by ArcIMS 4 software, rapidly expand user access to data sources beyond the organization. This “G.Net architecture” expands traditional organizational GIS information resources to include Internet data sources providing a rich data environment to support growing GIS user needs.

A variety of design alternatives enable organizations to develop an enterprise GIS solution that best supports their user requirements.

2.6 ESRI ArcGIS Implementation

Many ESRI customers have been working with ESRI products and data formats for many years. As a result, there are a large number of custom applications and data sets that have been developed and maintained in various locations throughout the GIS community. This proliferation of data and applications has resulted in a variety of data and application environments.

The ArcGIS technology is particularly suitable for an enterprise GIS implementation and can help ease the transition from the existing “stovepipe” GIS environments into a true enterprise implementation. The geodatabase technology in particular will provide tangible benefits in accomplishing this transition. The ArcGIS technology is also suitable to those new users who are spatially enabling their existing tabular data resources to support implementation of enterprise GIS operations. Many spatial data resources are available to support organizations as they migrate their operations to take advantage of GIS technology.

Geodatabases come in two varieties - personal and multi-user¹. Personal geodatabases, implemented in Microsoft Access, are suitable for project-level GIS. Multi-user databases, deployed using ArcSDE require a DBMS such as Oracle, Microsoft SQL Server, Informix or DB2. Storing spatial and attribute data directly in a commercial database gives a geodatabase capabilities that are not available, or are more difficult to achieve, with other formats. Some of these benefits are listed below.

- A uniform repository for geographic data. All geographic data is centrally stored and managed in one database.
- Data entry and editing is more efficient. The use of subtypes, domains, and validation rules helps maintain database integrity and reduces database maintenance.

¹ This section has been derived from the article by Colin Childs, ArcUser July-Sept 2001

- Sets of features are continuous. Geodatabases can accommodate very large sets of features without tiles or other spatial partitions.
- Multi-user editing. ArcSDE geodatabases use a data management framework called versioning that lets multiple users access and edit features simultaneously and reconciles any conflicts.
- Feature-linked annotation. Geodatabase annotation can be linked to the feature that it describes. When the linked feature is moved or deleted, the related label is moved or deleted.
- Users work with more intuitive data objects. A properly designed geodatabase contains data objects that correspond to the user's model of data. Instead of generic points, lines, and areas, users work with objects of interest such as parcels, roads, and lakes.
- Using a geodatabase can be simple and straightforward. Geodatabases can be created, accessed, and managed through the standard menus and tools in ArcCatalog, ArcToolbox, and ArcMap. However, the geodatabase model supports intelligent features, rules, and relationships that advanced users can employ in complex GIS applications..

In addition to the benefits of the geodatabase technology, the ArcGIS desktop applications provide a Windows-based suite of GIS data query, analysis and management, and tools. These tools often provide comparable functionality to the applications that have been developed by ESRI users in the past using custom AML and Avenue applications. The commercial off-the-shelf (COTS) ArcGIS functionality can also be extended as needed using the COM-based ArcObjects technology. ArcIMS and ArcReader provide additional GIS capabilities for users who need browse and query access. Transitioning from existing data and applications to the ArcGIS technology will provide tangible benefits and efficiencies for the GIS community.

2.6.1 Data Transition Strategies

GIS data are traditionally stored, updated and managed at the local workgroup level. This results in duplication of many of the data themes and additional effort in reconciling the various datasets when aggregation of these data are required at higher levels within the organization. This aggregation becomes even more difficult when the schemas for the various workgroup datasets are inconsistent.

An important first step in resolving the difficulties with aggregating data for enterprise-level analysis is to define data standards to which all data managers are required to adhere. The standardization process, however, can be time-consuming and institutionally difficult because of differences in local business processes and geographies. Depending on the business area, the prospect for data standardization may be more or less challenging.

ArcGIS provides a number of tools that can help in the standardization and management of data. On the database design side, ArcCatalog and/or Visio's UML CASE tool can be used to define a geodatabase schema. The schema defines entities (feature and object classes) and their definition (field size and type, nullability, default values, domains, etc.), as well as relationships between entities. These schemas can be defined from scratch, from existing agency designs or developed based on the geodatabase template models provided by ESRI at:

(<http://arconline.esri.com/arconline/datamodels.cfm>).

Implementing the standardization can be accomplished in several ways. Three of the alternatives for implementing data standards are described below: (1) centralization of data storage and

access through terminal server technology, (2) disconnected editing and reconciliation of centrally managed data, and (3) periodic aggregation of local datasets adhering to enterprise data standards.

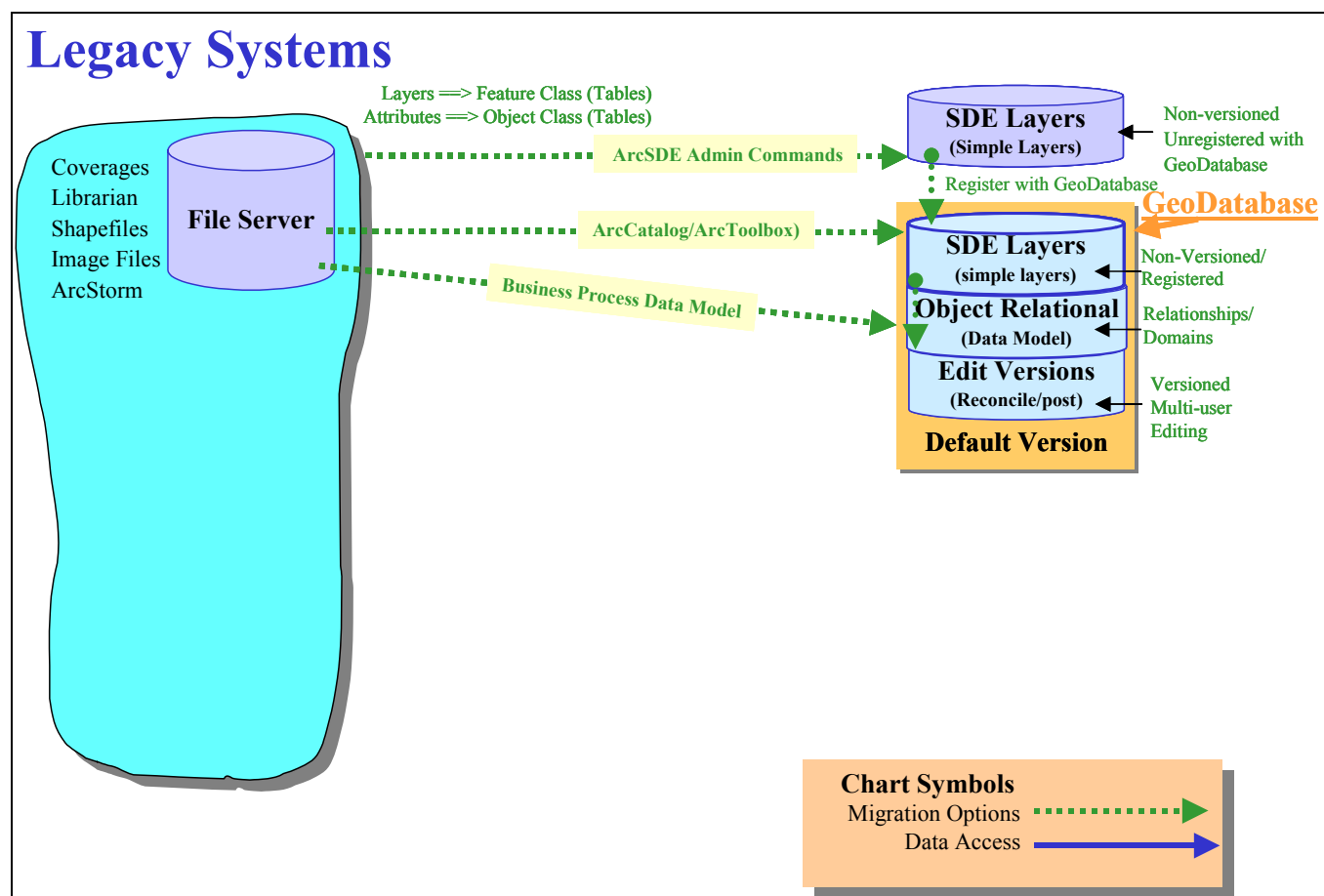
The simplest way to integrate data would be to enforce the schema definition so that all data imported into the ArcSDE database conforms to a single definition. This could be done through process or by centralizing the database and providing local access to the centralized database. The advantages to this approach include the centralization of data administration responsibilities and tight enforcement of the database schema.

An alternative to centralization is to define data standards for the essential elements of the schema, and allow local users to extend the “essential” model to include their added-value information. In this manner, the custodians responsible for aggregating local datasets would be able to map the required fields to the enterprise standard and ignore the additional information as desired. ArcCatalog provides an in-the-box graphical user interface (as well as programmable objects) for mapping the fields from one schema to another. ArcGIS layer files also provides tools for aliasing field names in the physical model to local names that may be more user-friendly. This can ease the difficulty of standardization when local nomenclature is different. The advantages to this approach include the flexibility for allowing local users to add additional fields into the schema for their local business needs.

Regardless of the alternative that is chosen, the migration of existing GIS data will be required at some level. Although ArcGIS supports the management, query and analysis of shapefiles and coverages, the content of the existing data would eventually need to be massaged to adhere to enterprise standards. In cases where the data are edited and maintained by multiple users, transitioning to a multi-user geodatabase in ArcSDE will provide the most robust model for managing data transactions.

Figure 2-9 below provides a data migration roadmap to convert legacy ArcInfo and ArcView data sets to both SDE simple layers (non-geodatabase) and into the geodatabase data model. It also shows how SDE simple layers can be integrated into the geodatabase data model and the different components and processes involved.

Figure 2-9
Data Migration Road Map



The options for migrating data from simpler to more robust data structures are as follows:

- Move file-based data (shapefiles, coverages, librarian tiles, etc.) to ArcSDE simple layers using the ArcSDE administration tools (command line). At this level, ArcSDE provides an improved data warehouse environment for an enterprise spatial data source.
- After creating ArcSDE layers, they may be registered with the geodatabase. Once a layer is registered with the geodatabase, the resulting *feature class* can be versioned to support multi-user data maintenance operations. Domains, relationship classes, and other object-relational behavior can be added to feature classes, further extending the geodatabase capabilities.

- Users can also move file-based data directly into the geodatabase using ArcCatalog or ArcToolbox. This migration path allows the user to create new feature classes or append updates to existing geodatabase feature classes. This is the preferred migration path once a geodatabase is established.

Legacy ArcInfo and ArcView clients will be able to view and query data in the default version of the new geodatabase. ArcIMS Image and Feature services also access the default version of the ArcSDE geodatabase.

2.6.2 Application Transition Strategies

In addition to the diversity of database schemas supported in the GIS community, there are also numerous custom applications that have been developed for managing and analyzing the data. While many of the different customer business processes may have unique aspects, the fundamental responsibilities for managing GIS data resources should be the same. As such, some of the applications required for managing GIS data resources could be standardized. Developing data standards is an important step in creating common GIS applications. Even so, the ArcGIS desktop tools provide a wealth of functionality for managing and analyzing data regardless of the data schema.

Legacy applications include ArcInfo custom applications, ArcView 3 custom applications, and ArcView 3 standard clients. Many of the functions developed in the custom applications are included in the ArcGIS standard clients or can be incorporated in the data models, which will reduce the requirements for custom application programming with the ArcGIS technology.

Technology upgrades include migration of clients to ArcIMS published map services, ArcGIS desktop client applications (ArcView 8, ArcEditor 8, or ArcInfo 8), or custom ArcGIS applications. The transition to ArcGIS from an existing operational environment will require a number of steps.

Training: Users should be trained in the capabilities and use of the COTS ArcGIS technology. Training a group of super-users or domain experts first will provide the background required for the next step in the transition: the application gap analysis.

Needs Assessment and Gap Analysis: A user application needs assessment and gap analysis should be completed to document functions required by each GIS user and to identify any functions that are not supported directly in the COTS ArcGIS desktop software. This analysis should identify those users that can migrate directly to ArcGIS and those users that must wait for custom ArcGIS applications. Custom applications can be built using any COM-compliant programming language (e.g. Visual Basic, C++) with the ArcObjects technology.

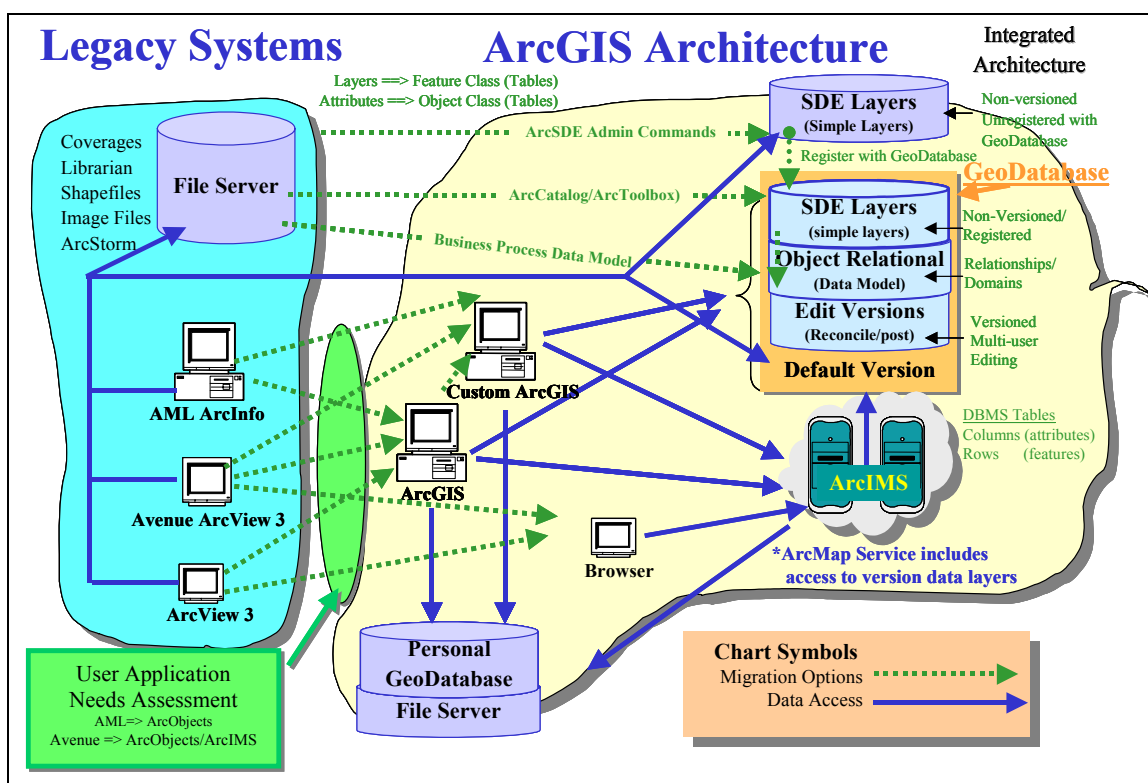
Power User Migration: Power users include GIS specialists who are trained in GIS analysis and technology. Many of their current tasks are undertaken on an ad-hoc basis and do not require custom applications. Many power users should be able to migrate directly to the standard ArcGIS desktop product without custom development. The ArcGIS desktop provides access to ArcIMS data sources and to versions and object relationships in the ArcSDE geodatabase.

Operational User Migration: Some customization may be required to support migration of standard desktop operations that currently are supported by custom ArcInfo applications. Many of the basic edit operations may be supported directly with the standard ArcEditor desktop.

There may also exist ArcView applications that can be supported with the ArcIMS technology for query and analysis.

Figure 2-10 below provides an application migration roadmap from legacy applications that use AML, Avenue, and COTS ArcView 3.x technology.

Figure 2-10
Application Migration Road Map



Note that the migration path from each of the existing platforms has a path from the existing implementation directly into ArcGIS without requiring any customization. It should also be noted that some ArcView 3 users might be able to perform their query and analysis functions through a thin browser client connected to an ArcIMS service. During the transition from the existing environment to the ArcGIS environment, there may be a need for continued use of existing AML applications. Workstation Arc/Info provides this capability, though the data accessed by existing AMLs would need to remain in coverage format until the existing code is updated to work with ArcSDE data or users become familiar with the ArcGIS Desktop tools.

3.0 Network Communications

Network communications provide the linkage that supports distributed computer processing. Network products establish a stable and dependable communication media for distributed data processing. A variety of communication protocols support distributed applications and data resources located at sites throughout an organization.

For several years, network technology was a relatively static environment while computer performance was increasing at an accelerating rate. Recent advances in communication technology support a dramatic shift in network solutions, introducing worldwide communications over the Internet and bringing information from millions of sources directly to the desktop in real time.

3.1 Desktop Workstation Environment

GIS applications rate among the heavy users of network traffic, along with document management and video conferencing. GIS technology provides a visual display environment to the user supporting very quick analysis of large amounts of graphic data. Access to distributed data sources for real-time display and analysis puts large demands on network communications. Data must be transported across the network to where the program is executed in order to display the information.

Figure 3-1
GIS Applications Network Impact

What GIS Does ...

- Graphic Data Representation (Maps)
- Large Quantity of Data Analysis
- Lots of Network Traffic

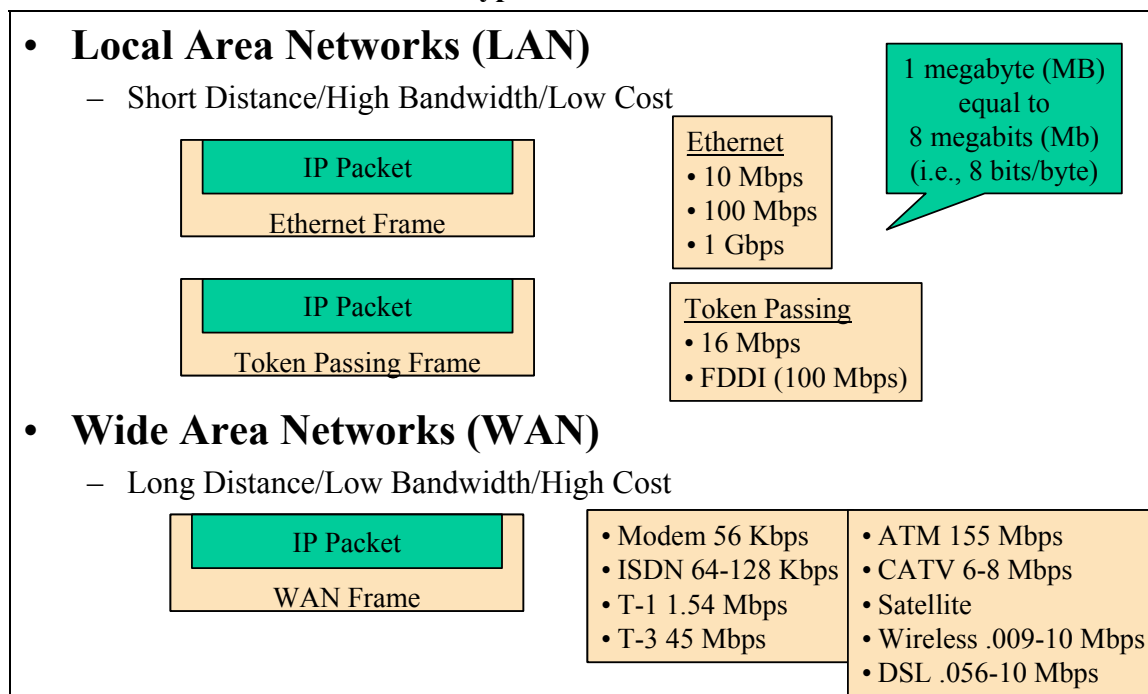
Data is a collection of digital computer information stored in media that has the capability to record and retain the data structure. These data are represented by little pieces of information called bits. Each bit takes up the same space on a storage or transmission media. For convenience, these little bits can be grouped into bytes of information with each byte containing eight bits. Data can be transported from one location to another within packets that protect the integrity of these data.

Data are typically transported from one storage location to another over copper wire or glass fiber physical networks. Other types of transport media include microwave, radio wave, and satellite digital transmissions. Each network protocol has limits to the volume of data they can transport based on the technology used to support the transmission.

Network transport solutions can be grouped into two general technology classes. These classes include local area networks (LANs) and wide area networks (WANs). The volume of data

(measured in bits) that can be transported per second represents the capacity of a specific network segment. This capacity is called network bandwidth and is typically measured in millions of bits (megabits) or billions of bits (gigabits) per second.

Figure 3-2
Types of Networks



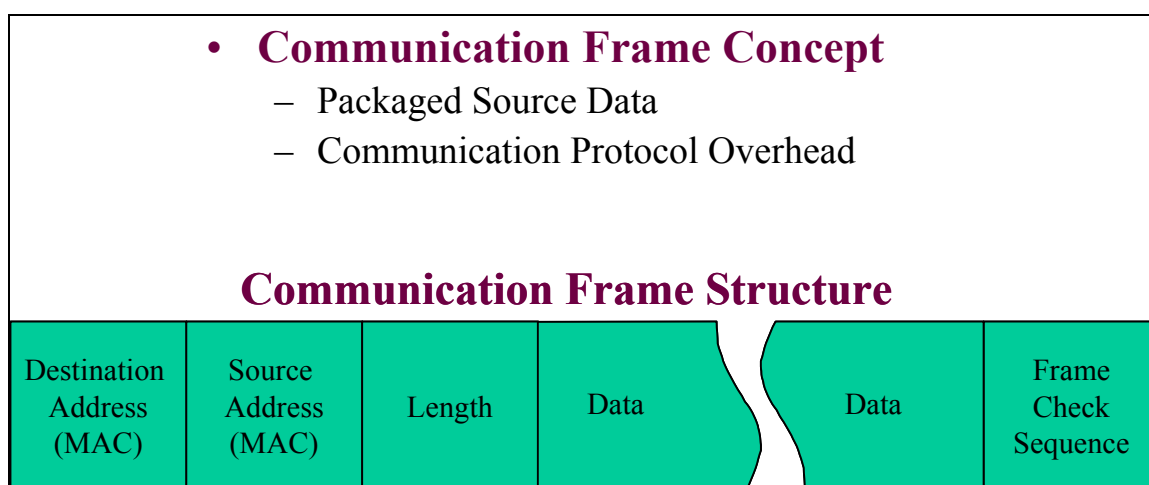
- **Local Area Networks (LANs).** Local area networks support high-bandwidth communication for short distances. This environment supports local operating environments (usually within a building or local campus environment). This is a data broadcast environment, which means only one user can send data on a single LAN segment at any one time. The cost for LAN environments is inexpensive relative to other hardware costs supporting the system environment.
- **Wide Area Networks (WANs).** Wide area networks support communication between remote locations. Technology supports much lower bandwidth than LAN environments, but transmission is possible over long distances. This is a data transport environment, which means data is packaged in a series of additional packets and transported as a stream of data along the transmission media. Cost for WAN connections are relatively high compared to LAN environments.
- **Data Units.** Data capacity is measured in terms of megabytes or gigabytes when stored on computer disk. Megabytes (MB) are abbreviated using a large “B” while megabits (Mb) are abbreviated using a small “b.” One must remember 1 MB = 8 Mb when converting data volume from disk storage to network traffic.

3.2 Client/Server Communication Concept

Applications move data over the network through proprietary client/server communication protocols. Communication processes located on the client and server platforms define the communication format and address information. Data are packaged within communication packets, which contain communication control information required to transport the data from its source client process to the destination server process.

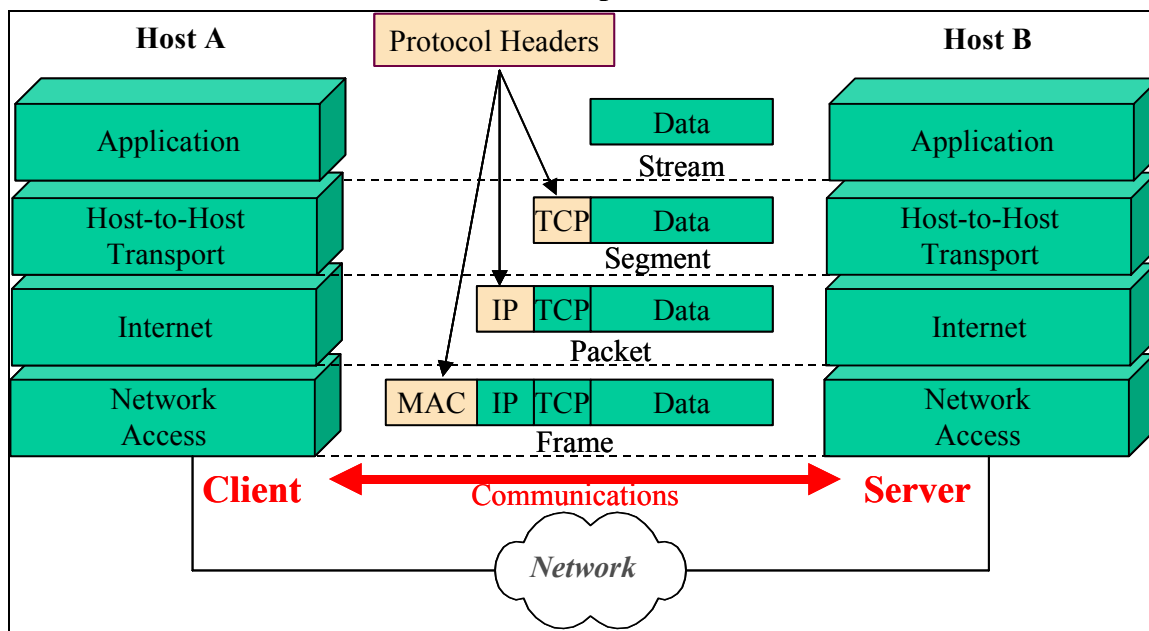
- **Communication Packet Structure.** The basic Internet packet structure includes destination and source addresses, and a series of control information in addition to the data structure. This information supports delivery of the packet across the network medium. The IP packet size will vary depending on the amount of data. The largest IP packet is around 65 Kilobytes (KB). Ethernet frames are limited to 1.5 KB. Data can be distributed across several packets to support a single data transfer.

Figure 3-3
Communication Packet Structure



- **Network Transport Protocol.** The communication packet is constructed at different layers during the transmission process. A data stream from the Host A application is sent through the protocol layers to establish a data frame for network transmission. The Transmission Control Protocol (TCP) header packages the data at the transport layer, the Internet Protocol (IP) header is added at the Internet layer, and the Media Access Control (MAC) address information is included at the physical network layer. The data frame is then transmitted across the network to Host B where the reverse process moves the data to the host application. A single data transfer can include several communications back and forth between the host applications.

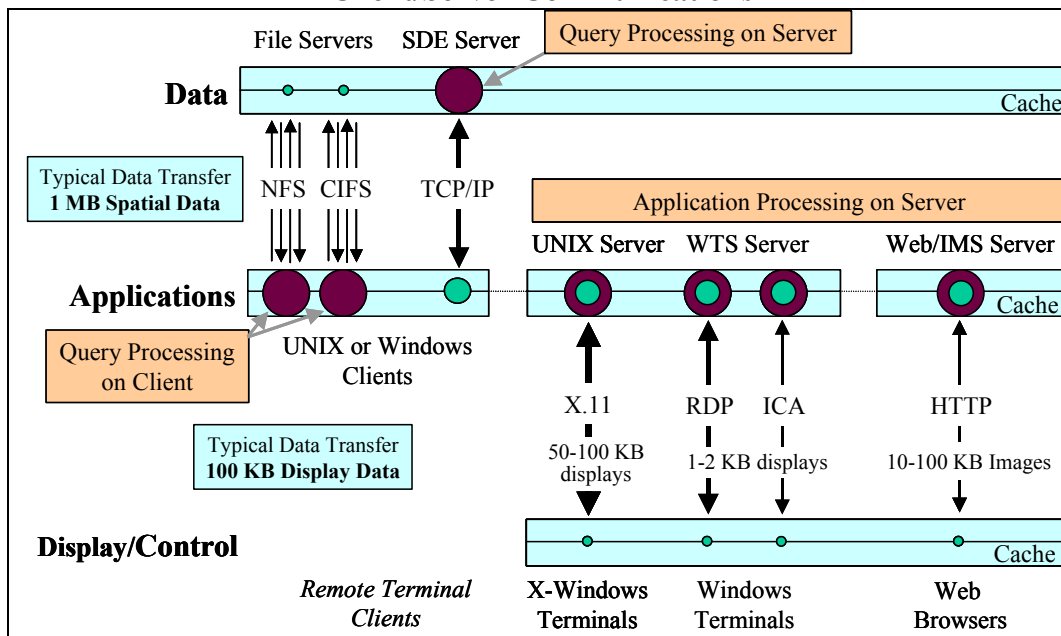
Figure 3-4
Network Transport Protocol



3.3 Client/Server Communications

Several client/server communication solutions are available to support network data transfer. Each solution includes a client and server component to support data delivery. The client process prepares the data for transmission and the server process delivers the data to the application environment. Primary protocols used by GIS solutions include the following.

Figure 3-5
Client/Server Communications



- **NFS (UNIX) and CIFS (Windows) Protocol.** Supports remote disk mounting enabling a client application to access data from a distributed server platform. All query intelligence is resident with the client application, directing access to data located on the server platform. Data must be transferred to the client application to support analysis and display.
- **ArcSDE API.** ArcSDE includes client and server communication components. The server component includes intelligence to support query processing. Data are compressed during transfer and uncompressed by the client application. Data must be transferred to the client application to support analysis and display.
- **X.11 Windows Protocol.** Supports remote terminal display and control of applications supported by a UNIX host application server. Supports transmission of display and control information to a terminal client.
- **ICA and RDP Protocol.** Supports remote terminal display and control of applications supported by a host Windows Terminal Server. Transmits display and control information to terminal client. Both ICA and RDP protocols compresses data for transmission.
- **HTTP Protocol.** Standard Web transmission protocol. Transaction-based environment, with product selection and display controlled by the browser client. Data is compressed for transmission.

3.4 Client/Server Network Performance

The data transfer volume and the network bandwidth can be used to establish expected user response times. A typical GIS application requires up to 1 MB of data to generate a new map display. A typical terminal environment requires 100 KB of data to support the display environment.

Figure 3-6
Client/Server Performance

<u>Client/Server Communications</u>	<u>Network Traffic Transport Time</u>			
	<u>56 Kbps</u>	<u>1.54 Mbps</u>	<u>10 Mbps</u>	<u>100 Mbps</u>
File Server to Workstation Client (NFS) •1 MB => 10 Mb + 40 Mb = 50 Mb	893 Sec.	32 Sec.	5 Sec.	0.5 Sec.
SDE Server to Workstation Client (SDE API) •1 MB => 10 Mb >> 5 Mb	89 Sec.	3.2 Sec.	0.5 Sec.	0.05 Sec.
UNIX Application Server to X-Terminal Client •100 KB => 1 Mb	18 Sec.	0.6 Sec.	0.1 Sec.	0.01 Sec.
Windows Terminal Server to Terminal Client (ICA) •100 KB => 1 Mb >> 20 Kb	0.3 Sec.	0.013 Sec.	0.002 Sec.	0.0002 Sec.
Web Server to Browser Client (HTTP) •50 KB => 500 Kb	9 Sec.	0.3 Sec.	0.05 Sec.	0.005 Sec.

Best Practices

The chart above takes the data transfer requirements in megabytes (MB), converts this to megabits (Mb) for transmission, includes any compression of these data performed by the communication protocol, and calculates the total volume of data in megabits (Mb) that must be transmitted (protocol overhead may be greater than what was used in this sample conversion). The transport time required to transfer these data are calculated for four standard bandwidth solutions. This simple illustration identifies the primary cause for remote client performance problems. Sufficient bandwidth is essential to support user performance requirements.

File server configurations support query from the client applications. When selecting data from a file (coverage or shapefile), the total file must be delivered to the client for processing. Data not required by the application is then rejected at the client location. This accounts for the considerable amount of network overhead experienced by these communications.

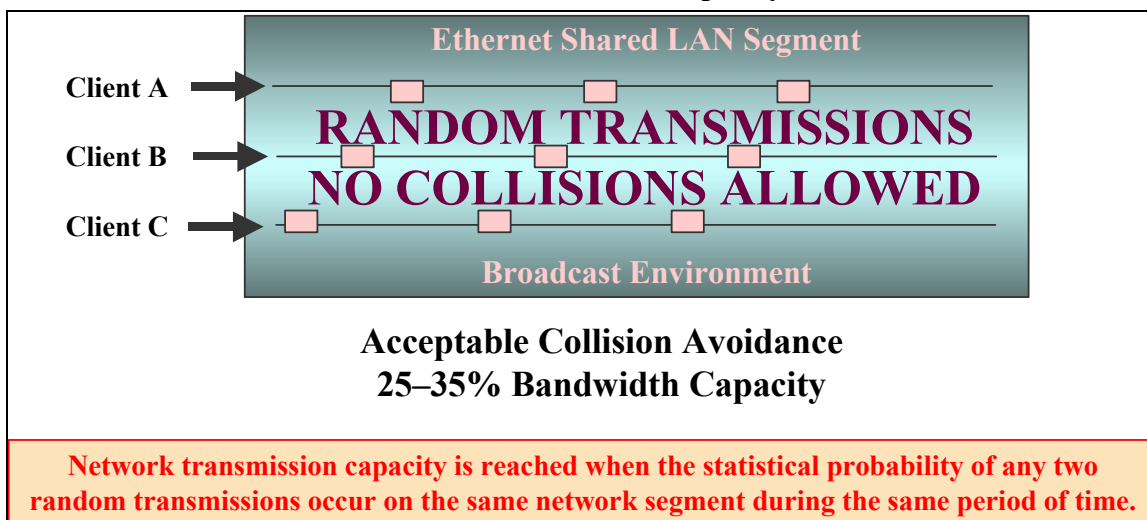
ArcSDE client/server configurations support query processing on the server platform. The query processing includes locating the requested data and filtering that data so only the specific data extent requested by the client is returned over the network. If the client limits requests to a small volume of data (i.e., point data or a single parcel in a parcel layer), the resulting data transfer can be very small and network transport performance would improve accordingly.

3.5 Shared Network Capacity

The total number of clients on a single network segment is a function of network traffic transport time (size of data transfer and network bandwidth) and the total number of concurrent clients. Only one client data frame can be transmitted over a shared network segment at one point in time. Multiple transmissions on the same Ethernet network segment will result in collisions, which will require another transmission to complete data frame delivery. Ethernet segments fail rapidly during saturation due to the rapidly increasing number of transmissions.

Figure 3-7

Shared Network Capacity



Token-passing environments operate somewhat differently. Concurrent transactions must wait for a token to support data frame delivery, resulting in transmission delays. Delays will increase as the network reaches saturation. Failure mode is much more graceful than Ethernet, supporting a higher acceptable bandwidth utilization.

3.6 Typical 1-MB Map Display

A GIS application may require 1 MB of spatial data, or up to 10 Mb of network traffic, to support a single map display. Applications can be tuned to prevent display of specific layers when the map extent exceeds defined thresholds. Only the appropriate data should be displayed for each map extent (i.e., it may not be appropriate to display individual parcel layers in a map showing the full extent of Houston, Texas). Proper tuning of the application can reduce network traffic and improve performance.

Figure 3-8
Typical 1-MB Map Display
(1:3,000 Scale (Feet), Average Features = 250)



3.7 Network Configuration Guidelines

Standard published guidelines are used for configuring network communication environments. These standards are application-specific and based on typical user environment needs. Communication environments are statistical in nature, since only a percentage of user processing time requires transmission over the network. Network data transfer time is a small fraction of the GIS users total response time (on properly configured networks). Network data transfer time can dominate response time when bandwidth is too small or when too many clients are on the same shared network segment.

Figure 3-9
Network Configuration Guidelines

- **Network Design Standards**
 - Current User Environment Needs
 - Addresses Statistical Application Use
 - Provides Basis for Initial System Design
- **Network Management**
 - Ongoing Traffic Management Task
 - Hardware, Application, and User Dependent
 - Strongly Affected by Work Environment and Changes in Computer Technology

The network must be designed to support peak traffic demands and is a function of the types of applications and user work patterns. Standard guidelines provide a place to start configuring a network environment. Once the network is operational, network traffic demands should be monitored and necessary adjustments made to support user requirements.

3.8 Shared Network Configuration Standards

Figure 3-9 provides recommended design guidelines for network environments. These guidelines establish a baseline for configuring distributed LAN environments. Five separate client/server environments are included for each network bandwidth. The number of recommended clients is based on experience with actual system implementations and does not represent worst-case environments. Networks should be configured with the flexibility to provide special support to power users whose data transfer needs exceed typical GIS user bandwidth requirements.

Figure 3-10
Network Design Guidelines

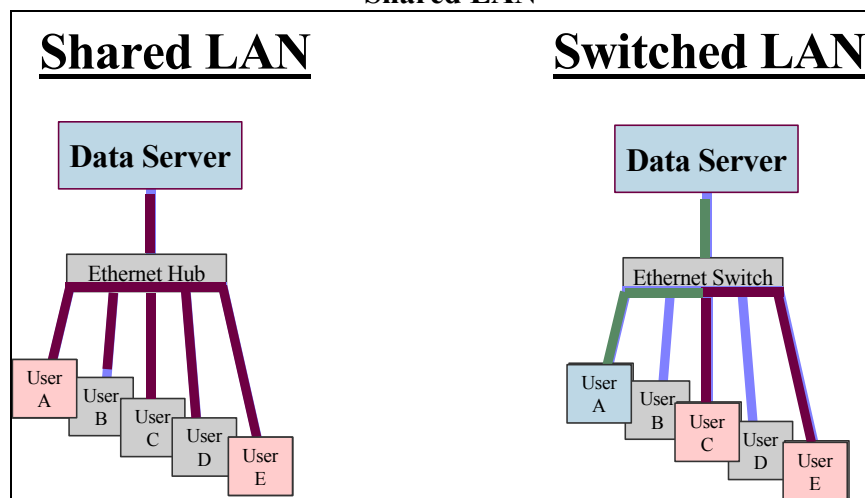
Local Area Networks Bandwidth	Concurrent Client Loads				
	File Servers	SDE Servers	X Emulation	Windows Terminals	Web Products
10 Mbps LAN	1-2	10-20	50-75	300-600	150-300
16 Mbps LAN	1-3	16-32	80-120	500-1000	250-500
100 Mbps LAN	10-20	100-200	500-750	3000-6000	1500-3000
1 Gbps LAN	100-200	1,000-2,000	5,000-7,500	30000-60000	15000-30000
Wide Area Networks Bandwidth	Concurrent Client Loads				
	File Servers	SDE Servers	X Emulation	Windows Terminals	Web Products
56 Kbps Modem	NR	NR	NR	2-4	1-2
128 Kbps ISDN	NR	NR	NR	5-10	2-4
256 Kbps DSL	NR	NR	1-2	10-20	5-10
1.54 Mbps T-1	NR	1-2	9-12	50-100	25-50
6.16 Mbps T-2	1-2	6-12	30-45	200-400	100-200
45 Mbps T-3	5-10	50-100	250-350	1500-3000	700-1500
155 Mbps ATM	15-30	150-300	850-1200	5000-10000	2500-5000

3.9 Local Area Network Component Configuration

Local area networks are generally constructed with wiring closets at different locations throughout the building (i.e., closet on each floor) with twisted-pair wiring from the wiring closet to each user desktop. Each user desktop computer is connected to a port on an Ethernet or token-ring hub located in the wiring closet. The hubs within the wiring closets are connected to the computer room over a glass-fiber or twisted-pair copper-wire backbone. The data servers located in the computer room are also connected to the network backbone. Network bandwidth capacity is a function of the computer network interface card and the port on the hub in the wiring closet.

There are two types of basic hubs supporting most network environments. The initial hub technology supported a shared network. This worked fine for sharing data between local area network clients. Distributed client/server processing configurations required much more bandwidth. A new switched hub technology was introduced to support the growing traffic demands.

Figure 3-11
Shared LAN



- **Shared Ethernet Hubs.** A broadcast protocol is used to support Ethernet communications. A standard Ethernet hub provides the backbone for the Ethernet LAN. Each user workstation and the data server is connected to a port in the Ethernet hub.

When User A communicates with the data server, the communication will broadcast over all LAN segments connected to the shared hub. Any other transmission attempted during this same period will result in a collision, delay for a random period of time, and then retransmit. As more users attempt to broadcast at the same time, collisions increase and the number of transmissions (and resulting collisions) grows exponentially until network traffic is saturated.

The maximum recommended communication rates for LAN environments are 25 to 35 percent utilization (depending on the total number of users). The probability of any two users initiating a transmission at the same time contributes to the total traffic supported by the network (higher utilization rates can be achieved with fewer users on the network).

Increasing bandwidth (100 Mbps versus 10 Mbps) can reduce data transport time (faster data transfer) and lower the probability of transmission collisions (each transmission spends less time on the network). A 100BaseT network can handle roughly ten times the number of clients as a 10BaseT network.

- **Switched Ethernet.** An Ethernet switched hub maps each user workstation address to the switch backplane, and transmissions are sent down the user LAN segment only when they are addressed to that user. The Ethernet backplane has a high bandwidth (gigabit range), so multiple transmissions can share access to the backplane, providing higher bandwidth to each user.

Uplinks to common network locations (central data server or network backbone) should include higher bandwidth. Several users can access the server virtually at the same time due to variation in network capacity. Example: A switched Ethernet network with five 10BaseT segments and one 100BaseT segment (for the data server) can support the cumulative equivalent of 50-Mbps bandwidth to the server.

Some Ethernet switches support single-user segments configured in duplex mode, which allows the segment to send and receive messages at the same time (each using one of the twisted-pair lines). The switch includes a buffer cache, which prevents collisions over the duplex connections. A duplex connection on a 100-Mbps port can provide up to 200-Mbps bandwidth (100 Mbps in each direction). Full duplex connections can support higher transmission loads, since they protect against collisions and retransmissions during peak traffic conditions.

Network technology has improved significantly over the past several years. Many switches today include support for a variety of network functions, including protocol conversion and communication routing. Introduction of switched gigabit Ethernet has enabled many organizations to replace expensive network backbones with switched gigabit Ethernet, with significant cost and performance benefits. Network components are becoming a commodity, with technology improving and costs going down. The Internet technology has significantly expanded the network communication marketplace, encouraging new and faster network technology.

3.10 Web Services Configuration Guidelines

Implementation of ArcIMS Web mapping services places additional demands on the network infrastructure. The amount of system impact is related to the complexity of the published mapping services. Map services with small (less than 10 KB) or a limited number of complex images will have little impact on network traffic. Large images (greater than 100 KB) can have significant impact on network performance.

Figure 3-12 provides an overview of network performance characteristics that should be considered when deploying a Web mapping solution. The top chart shows the maximum number of requests that can be supported over various WAN bandwidths based on average map image size. The bottom chart shows the optimum transmission time for the various size map images. Web information products should be designed to support user performance needs, which may be dominated by available bandwidth. Standard ArcIMS map services should produce images from 30 KB to 50 KB in size, in order to minimize network transport time to standard ArcIMS client browsers (50-KB image requires over 26 seconds of network transport time for 28-Kbps modem clients, limiting site capacity to a maximum of 7,400 requests per hour over a single T-1 Internet Service Provider connection). Users generally demand reasonable performance, or they will not be satisfied. Proper infrastructure bandwidth and careful map information product design are required to support high-performance Web solutions.

Figure 3-12
ArcIMS Network Performance

Wide Area Network Bandwidth	Peak ArcIMS Requests/Hour (based on Average Image Size)				
	10 KB	30 KB	50 KB	75 KB	100 KB
56 Kbps Modem	1,344	448	269	179	134
1.54 Mbps T-1	36,960	12,320	7,392	4,928	3,696
6.16 Mbps T-2	147,840	49,280	29,568	19,712	14,784
45 Mbps T-3	1,080,000	360,000	216,000	144,000	108,000
155 Mbps ATM	3,720,000	1,240,000	744,000	496,000	372,000
Note: 1 KB = 15 Kb HTTP traffic					
Wide Area Network Bandwidth	Image Transfer Time (sec) based on Average Image Size				
	10 KB	30 KB	50 KB	75 KB	100 KB
19 Kbps Modem	7.9	23.7	39.5	59.2	78.9
28 Kbps Modem	5.4	16.1	26.8	40.2	53.6
56 Kbps Modem	2.7	8.0	13.4	20.1	26.8
1.54 Mbps T-1	0.1	0.3	0.5	0.7	1.0
6.16 Mbps T-2	0.0	0.1	0.1	0.2	0.2
45 Mbps T-3	0.0	0.0	0.0	0.0	0.0
155 Mbps ATM	0.0	0.0	0.0	0.0	0.0

ArcIMS includes an Extract Service that will support data downloads to Web clients over the Internet. The ArcIMS extract service extracts data layers from ArcSDE based on identified extent, zips the data into a compressed file, and downloads the data to client. **Figure 3-13** identifies minimum download times based on available bandwidth and the size of compressed data packages. Data downloads should be restricted to protect Web site Internet service bandwidth. Data downloads can very easily dominate available bandwidth and impact performance to other Web mapping clients.

Figure 3-13
Data Download Performance

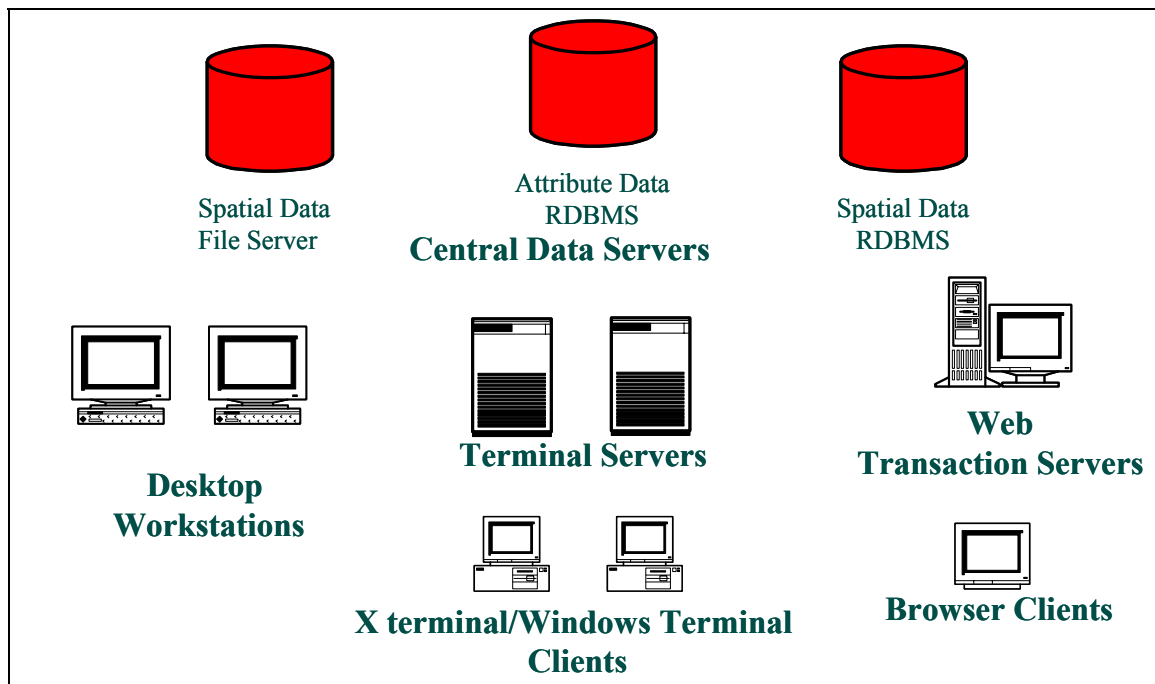
Wide Area Network Bandwidth	Peak FTP Downloads/Hour (based on Average File Size)				
	1 MB	5 MB	10 MB	20 MB	50 MB
56 Kbps Modem	17	3	2	1	0
1.54 Mbps T-1	462	92	46	23	9
6.16 Mbps T-2	1,848	370	185	92	37
45 Mbps T-3	13,500	2,700	1,350	675	270
155 Mbps ATM	46,500	9,300	4,650	2,325	930
Note: 1 KB = 12 Kb FTP traffic					
Wide Area Network Bandwidth	File Transfer Time (sec) based on Average File Size				
	1 MB	5 MB	10 MB	20 MB	50 MB
19 Kbps Modem	632	3,158	6,316	12,632	31,579
28 Kbps Modem	429	2,143	4,286	8,571	21,429
56 Kbps Modem	214	1,071	2,143	4,286	10,714
1.54 Mbps T-1	8	39	78	156	390
6.16 Mbps T-2	2	10	19	39	97
45 Mbps T-3	0	1	3	5	13
155 Mbps ATM	0	0	1	2	4

4.0 GIS Product Architecture

This section provides a basis for understanding components involved in distributed GIS applications. Understanding basic application architecture, relationships between commercial products and custom applications, and the component interfaces required to support GIS solutions provides a foundation for understanding distributed GIS design principles.

Enterprise-level GIS applications support a variety of users throughout an organization, all requiring access to shared common spatial and attribute data. System hardware and software environments for distributed GIS applications are supported by multi-tier client/server architecture. An overview of this architecture is presented in **Figure 4-1**.

Figure 4-1
GIS Multi-tier Architecture



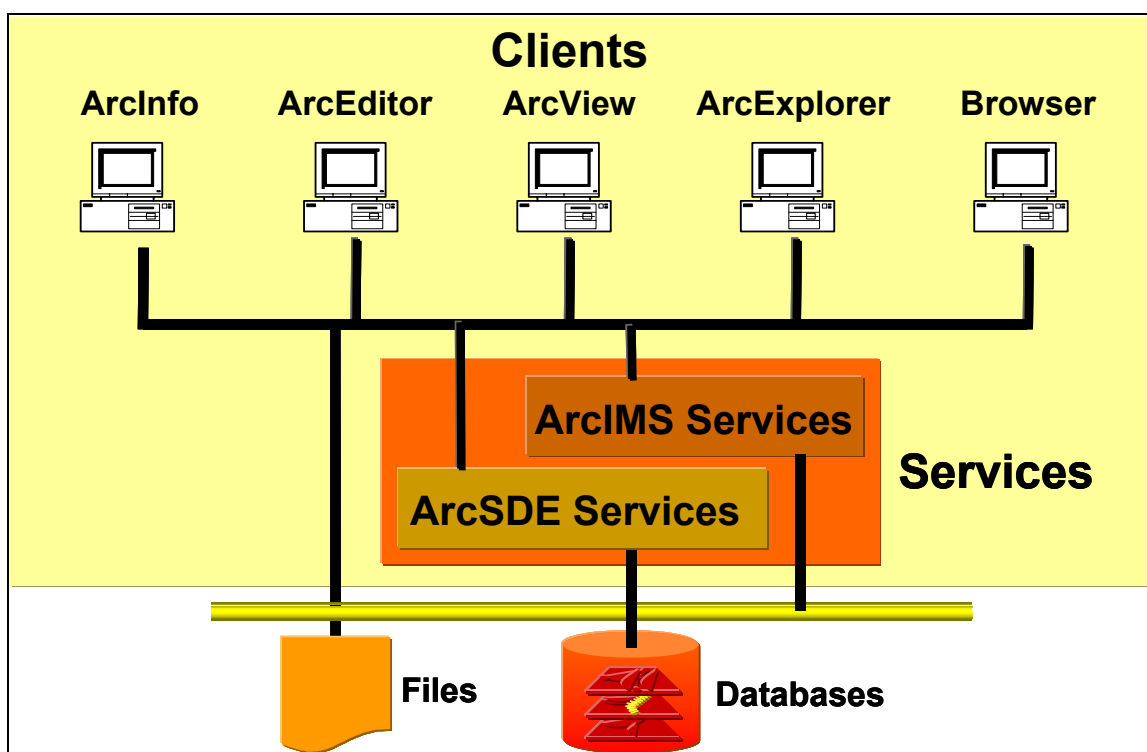
- **Central Data Servers.** Shared spatial and tabular database management systems provide central data repositories for shared geographic data. These database management systems can be located on separate data servers or on the same central server platform.
- **Application Compute Servers.** GIS applications are supported within the distributed configuration by hardware platforms that execute the GIS functions. In a centralized solution, application compute servers can be UNIX or Microsoft Windows platforms that provide host compute services to a number of GIS clients. These platforms include terminal servers and Web transaction servers (map servers). In smaller configurations, the application compute server and central data server may be on the same platform.
- **Desktop Workstations.** Display and control of application processes are provided by desktop workstations which, in many cases, are PCs running X-emulation software,

Windows terminal clients, or Web browser clients. In many GIS solutions, the client application server and desktop user workstation may be the same platform.

4.1 ArcGIS System Software Architecture

ArcGIS is a collective name representing the combined ESRI GIS technology. This technology includes a mix of ArcGIS desktop applications and ArcGIS server-based services. **Figure 4-2** provides an overview of the ArcGIS system environment.

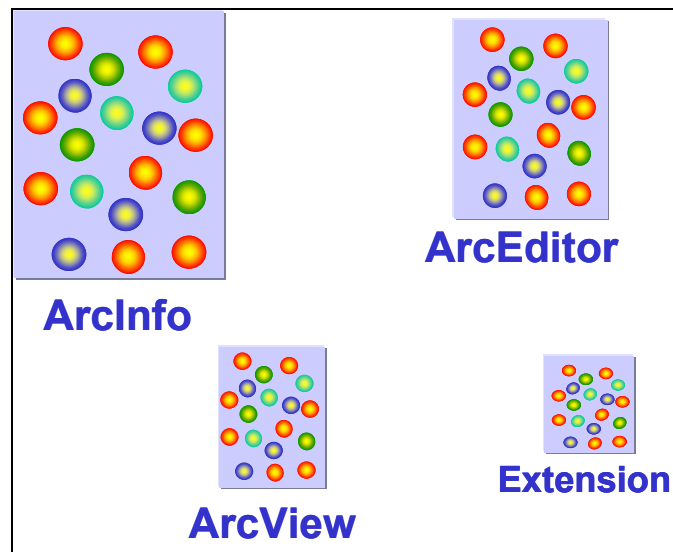
Figure 4-2
ESRI ArcGIS Software



4.2 ArcGIS Desktop Software Architecture

ArcGIS desktop applications include ArcInfo, ArcEditor, and ArcView licensing options. These applications are supported by a common set of ArcObjects developed using COM programming technology. ArcGIS desktop software is supported on the Microsoft Windows operating system. The ArcGIS desktop technology presents an object-oriented user-friendly interface along with a variety of GIS functionality. **Figure 4-3** provides an overview of the ArcGIS desktop software.

Figure 4-3
ArcGIS Desktop Software



An overview of ESRI commercial GIS software is provided in **Figure 4-4**. These include GIS applications (UNIX and Windows), data management solutions (GIS file server, ArcStorm, and ArcSDE), and remote access client solutions (Windows terminals and browsers).

Figure 4-4
ESRI Software Environments

- **GIS Applications**
 - GIS for UNIX Product Architecture
 - GIS for Windows Product Architecture
- **Data Management Solutions**
 - GIS File Server (coverages, shapefiles, LIBRARIAN, ArcStorm)
 - Spatial Database Engine (client/server data management)
- **Remote Access Client Solutions**
 - Windows Clients with Microsoft Terminal Servers
 - Browser Clients with ArcIMS

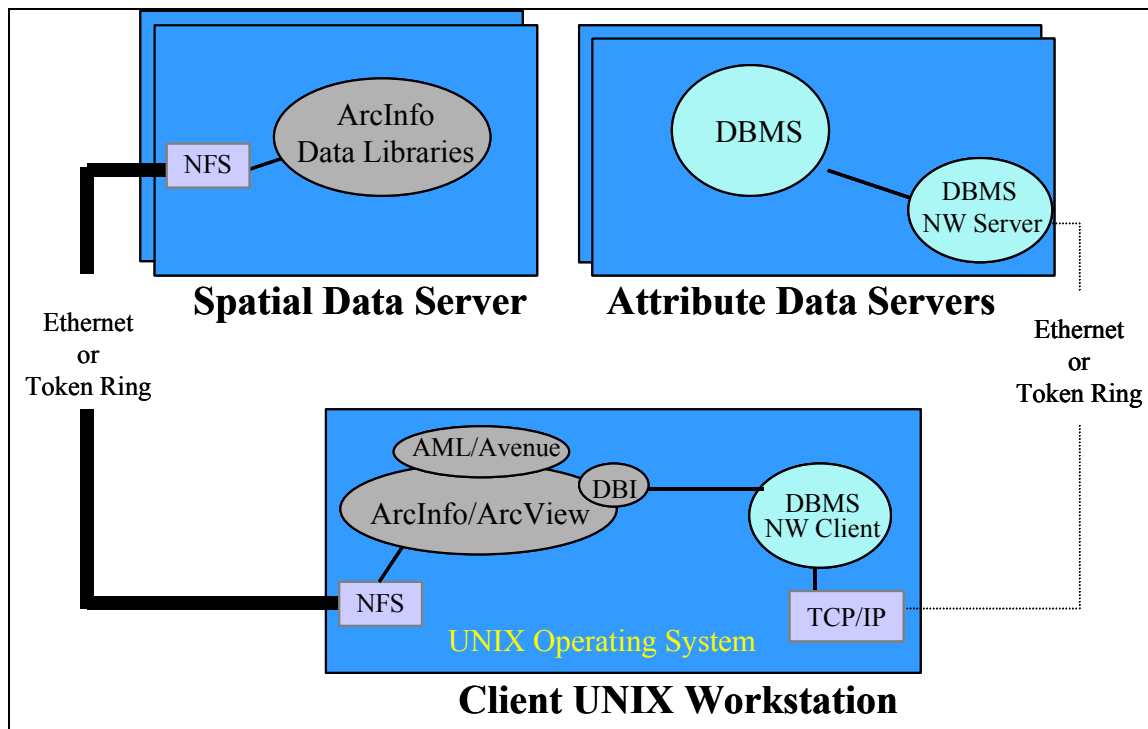
ESRI commercial software performance and associated system interface specifications provide a baseline for system design. Custom GIS application solutions require system resources to support the basic underlying commercial software functions.

GIS applications are supported by an open systems architecture. This architecture combines a variety of closely integrated commercial products to establish a fully supported system solution. Commercial software supports evolving communication interface standards to enable system integration with minimum customization. The importance of selecting well established software solutions cannot be over-emphasized, since all parts of the distributed configuration are critical in supporting overall system performance.

4.3 GIS for UNIX

Critical components supporting distributed ArcInfo for UNIX solutions are identified in Figure 4-5.

Figure 4-5
GIS for UNIX Product Architecture
 (UNIX File Server Spatial Data Source)



GIS application functions are executed within the UNIX workstation operating system environment. File-based spatial data management solutions (including spatial coverages, shapefiles, and ArcInfo LIBRARIAN) are supported entirely from the client workstation. The DBMS vendor software operates within the system environment provided by the attribute data server. Network file servers (NFS), provided with the UNIX operating system, mounts files located on the spatial data server to the client workstation, making them appear as if on local disks. Spatial data located on the GIS data server appear as local data to the GIS application.

A Database Integrator module is included with ArcInfo and ArcView GIS that translates communication from the GIS application to the DBMS client module located on the client workstation. DBMS client and server components handle network communication between the client workstation and the DBMS data server using appropriate network communication protocol. ArcInfo AML or other standard programming tools such as Visual Basic, Delphi, PowerBuilder, C, C++, Tcl/Tk, and/or Motif can be used to establish a custom user interface to ArcInfo application environments. Avenue macro language supports customized ArcView application environments.

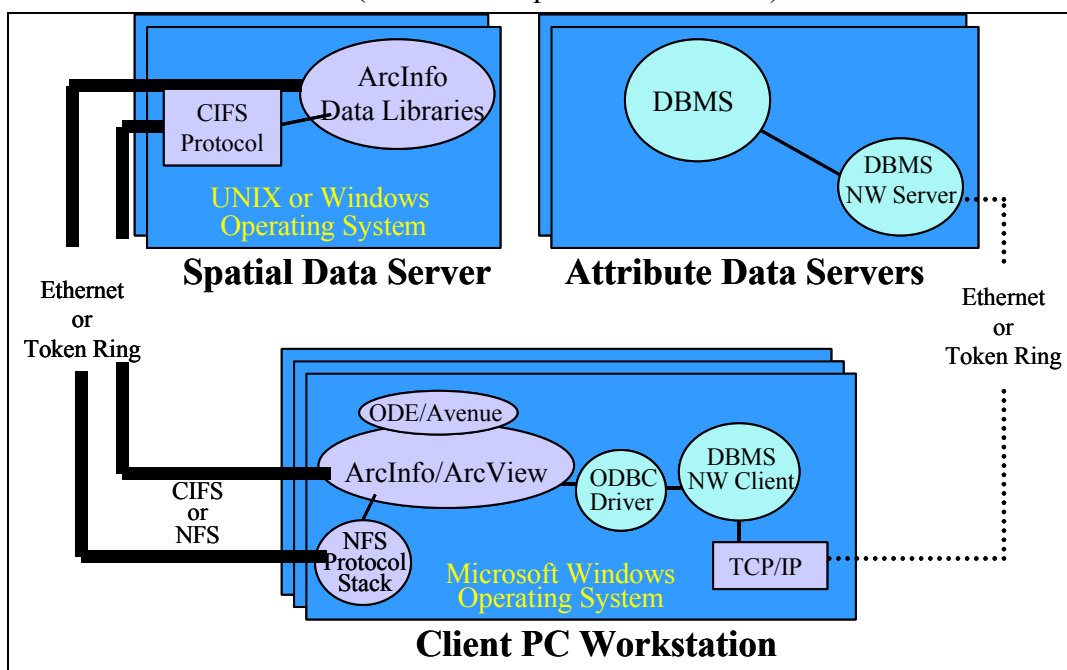
Communications between GIS workstation applications and spatial data located on a separate file server require a significant amount of network traffic overhead due to the connection-oriented protocol supporting NFS communications. Single-threaded application processes that function in memory on the client workstation execute program calls against the central database. Each line of code must receive a response over the network connection before the next line of code can be executed. Several transmissions are required to locate each source of required data. This type of communication generates lots of extra network traffic.

Communication between GIS workstation applications and remote attribute data requires very little network traffic. Queries to the central tabular DBMS are packaged in a single client request and sent as a message to the server for processing. The request is handled by a separate server process, which compiles the answer to the query and sends the response back to the client application for further processing. There are two reasons this is much more efficient than accessing spatial data—one is the use of message-oriented communication protocol (data processing is supported by a local server process); the other reason is the limited size of tabular data files, which are much smaller than spatial graphical files.

4.4 GIS for Windows

In 1996, ESRI ported the ArcInfo UNIX workstation application to the Windows platform, positioning ESRI to take advantage of the Microsoft Windows environment and the lower cost of PC hardware. ArcView GIS applications were also deployed on Windows workstations. Critical components supporting distributed GIS for Windows solutions are identified in **Figure 4-6**.

Figure 4-6
GIS for Windows Product Architecture
(File Server Spatial Data Source)



GIS application executables execute within the operating system environment provided by the Windows platform. Microsoft Common Internet File Services (CIFS) protocols included with Windows operating system support sharing of server disk resources to PC clients. Remote data servers appear as local disk drives to the client Windows workstation, and GIS applications access to spatial data is provided from the client workstation. The GIS Windows applications were developed to an Open Database Connectivity standard (ODBC-compliant) communication interface. The ODBC interface provides access to PC-based attribute data sources including Microsoft SQL Server and Microsoft Access. An ODBC driver must be obtained to complete the interface between GIS applications and each specific attribute database. AMLs developed in the UNIX environment will continue to perform with Workstation ArcInfo for Windows. Avenue scripts supporting custom ArcView GIS UNIX applications are supported by ArcView GIS for Windows software. Software are available to support conversion of UNIX system calls to Windows commands and may be required to support some custom code.

Communication between GIS applications and spatial data located on a separate data server (using CIFS protocol) requires network traffic overhead similar to UNIX NFS communications. Accessing remote data sources may take twice as long as accessing the same data on local disk, depending on the data storage technology.

Initially, Windows data servers supported department-level work groups. Server configurations capable of supporting enterprise-level GIS operations were not available. PC vendors (along with Intel) have made significant progress toward supporting enterprise server requirements. Windows multi-processor servers can be configured for up to eight processors. Microsoft provides a cluster fail-over solution for Windows operating systems that support server failover requirements.

Several solutions are available to support mixed Windows and UNIX environments. Software products are available to enable UNIX platforms to directly share disks to PC clients. Gateway products are available for Novell and Windows servers that support mapping of UNIX server disks to their supported PC work groups. PC client NFS products are available to support mounting of the UNIX server disk. In all these solutions, the server disk is mapped to the client workstation to allow ArcInfo access to the remote server data as if it were on a local PC drive.

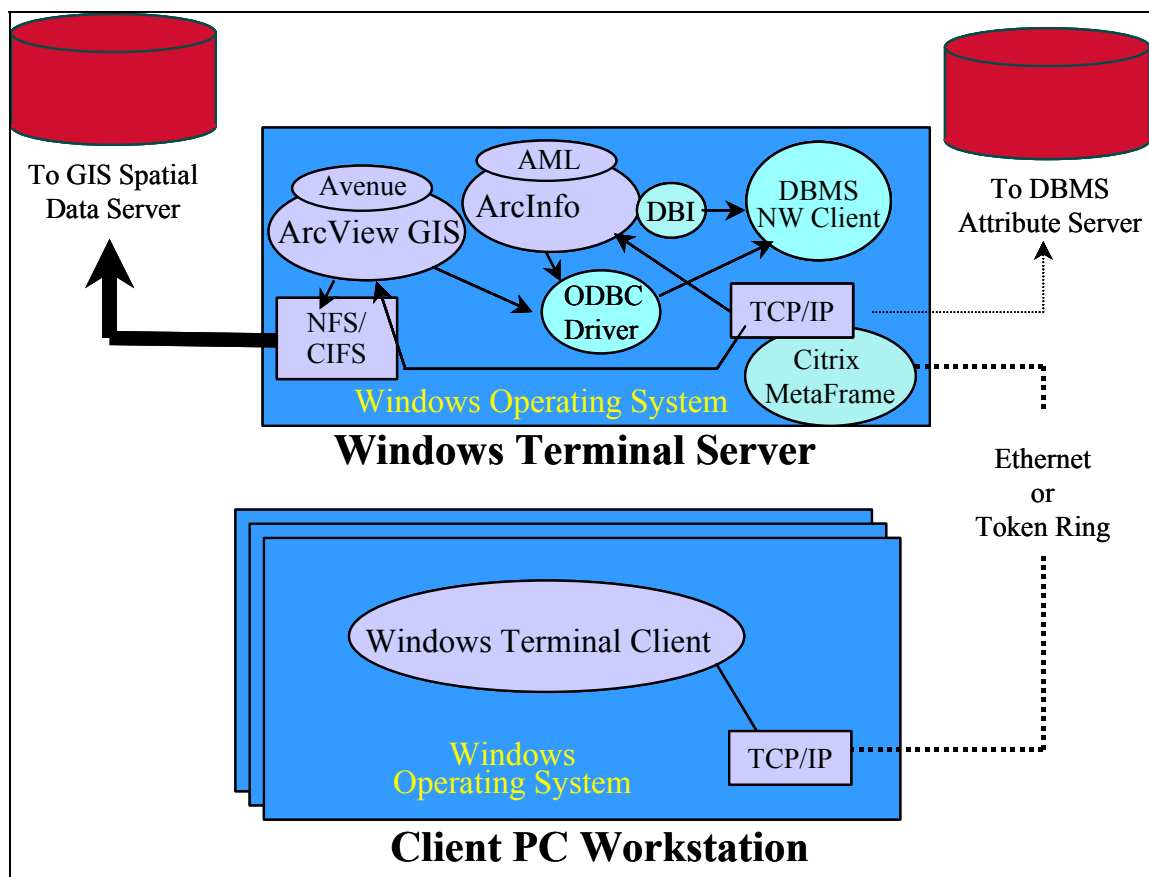
4.5 Microsoft Windows Terminal Server

The Microsoft Windows Terminal Server product establishes a multi-host environment on a Windows server. A Windows Terminal client provides display and control of applications executed on the Windows Terminal Server. Microsoft uses a standard remote desktop protocol (RDP) to support communication between the terminal server and client Windows platforms.

Citrix provides a MetaFrame extension product that provides a more efficient independent computing architecture (ICA) communication protocol to support communications between the terminal server and client Windows platform. The ICA protocol requires less than 28-Kbps bandwidth (rendering vector data information products) to support full Windows display and control of GIS applications supported on a Windows Terminal server. The MetaFrame product also includes client software for UNIX, Macintosh, and imbedded Web client applications.

Figure 4-7 provides an overview of the Windows Terminal Server communication architecture.

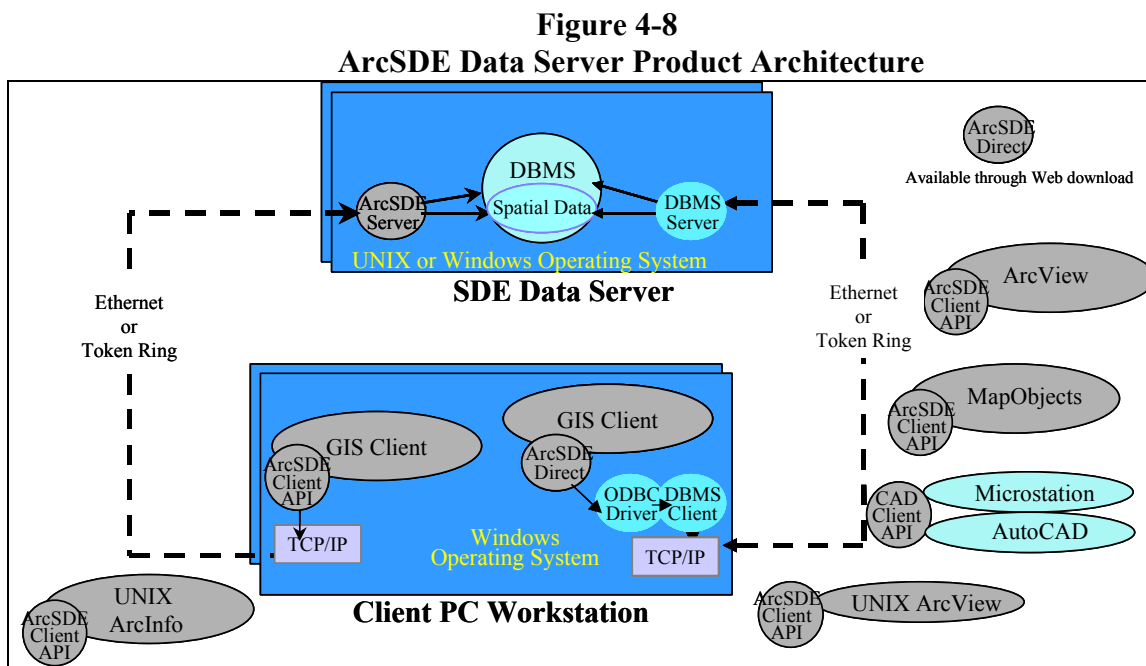
Figure 4-7
Windows Terminal Server Product Architecture



The Windows Terminal client communicates with the Windows Terminal Server through a compressed message-oriented communication protocol. The Windows environment display must be provided over the network to the client workstation, but this requires much less data than that required for spatial data processing by the terminal server or a workstation client. The terminal display traffic requirements are very small, supporting full server application performance over 28-Kbps modem dial-up connections (heavy image processing environments may require more bandwidth).

4.6 Spatial Database Engine (ArcSDE)

The Spatial Database Engine supports storage of spatial data within standard commercial DBMS applications. Integration of GIS spatial data with the enterprise-level DBMS environment provides a powerful enterprise-level data storage solution for large data stores and/or solutions supporting a high number of GIS clients. A variety of data performance and administration tools are provided with the associated DBMS solutions. Critical components associated with ArcSDE communication architecture are identified in **Figure 4-8**.



The ArcSDE solution combines spatial and attribute data into a single GIS database. Query processing is supported by the ArcSDE DBMS server. A transaction request is prepared by the GIS application client and sent to the ArcSDE server for processing. The ArcSDE server process completes the transaction query and returns the response back to the client for processing. There is no requirement in this configuration to map the data server disk to the GIS client. Application processes located on the ArcSDE data server process all spatial and attribute data requests received from the GIS client application.

ArcInfo, ArcView GIS, and MapObjects applications include ArcSDE client APIs that support direct access to central SDE server data. Also, both ArcSDE and ArcInfo support direct conversion of ArcInfo coverages to ArcSDE layers. An additional ArcSDE CAD client API is available to support storage of MicroStation and AutoCAD data in ArcSDE.

ArcSDE 8 supports the traditional georelational database and a new geodatabase-relational data model. This version also introduces the concept of a versioned database, and supports native data maintenance operations from the desktop ArcInfo 8 platform in the form of a versioned database.

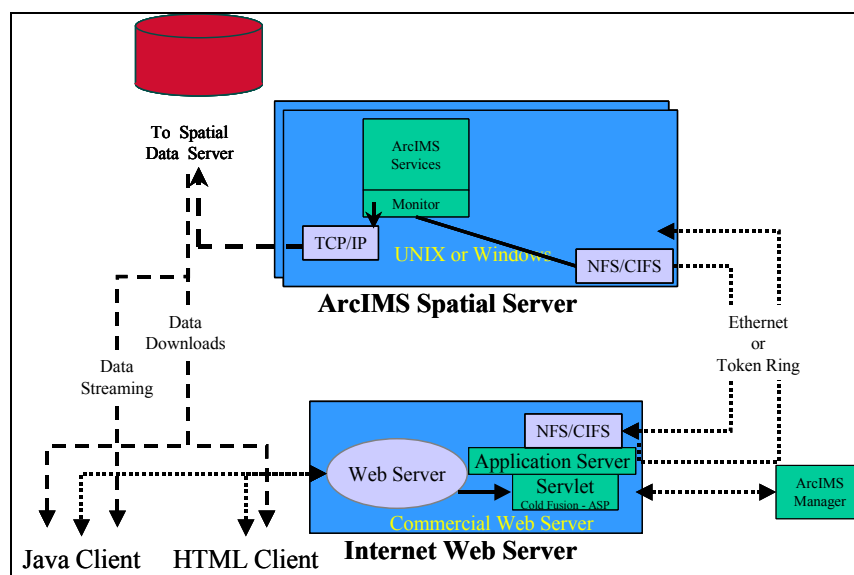
Standard DBMS client/server communications can be used with ArcSDE but are not required for ArcSDE queries. When used, they can provide standard access to attribute data through the DBMS client communication interface.

The ArcGIS 8.1 release includes a direct connect ArcSDE client access option for Microsoft SQL2000 and Oracle database solutions. The ArcSDE Direct option for Oracle will connect to the Oracle network client. The Oracle network client will communicate ArcSDE query calls to the server. The ArcSDE Direct option for SQL2000 will support direct communications to the server. Query processing will be supported by the DBMS functions on the server platform. Additional traffic will result between the server and client with this configuration, since the ArcSDE g-serv which filters the data for the client application will be supported on the client workstation. To upgrade existing ArcSDE client applications, the ArcSDE Direct Connect API is available for download. A mix of direct-connect and ArcSDE server-connect clients can be supported the same geodatabase environment.

4.7 Arc Internet Map Server (ArcIMS)

Arc Internet Map Server introduces a modern approach to serving map products over the Internet. ArcIMS has a Java-based application management environment that includes mapping services and map design tools to support a variety of Internet map services. Critical components associated with the ArcIMS communication architecture are identified in **Figure 4-9**.

Figure 4-9
ArcIMS Web Services Product Architecture



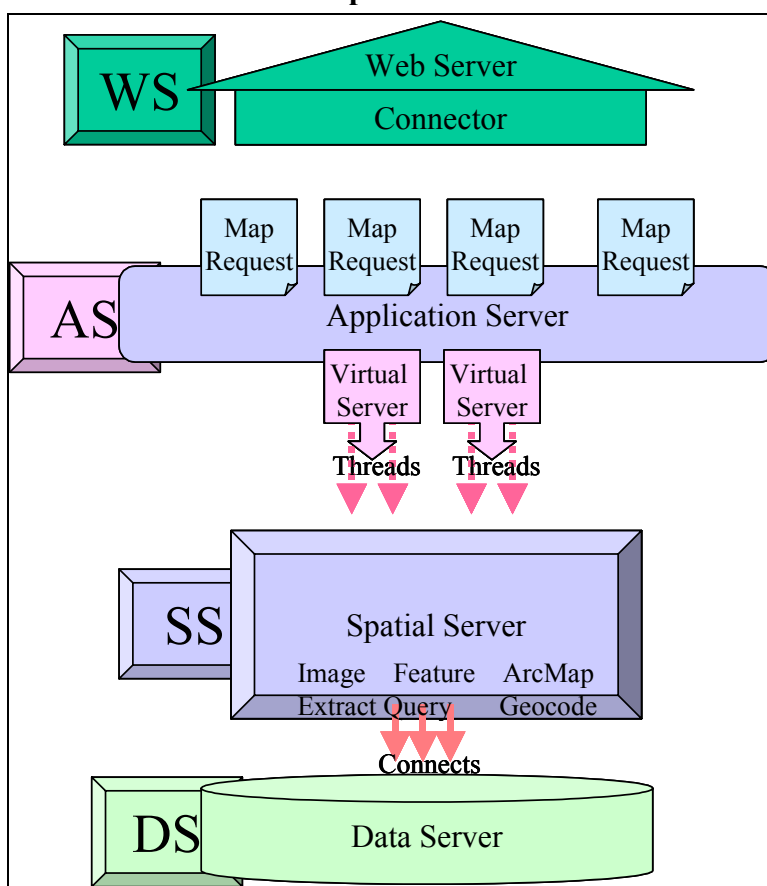
ArcIMS includes a Java Web servlet that supports ArcIMS, MapObjects, and ArcView map services. The standard product includes an ArcIMS manager and map development wizards that support design and authoring of most standard map products without special programming. ArcIMS also includes a Java client that supports standard Web products, data streaming, and data downloads. The Java client also supports integration of local vector data with map images obtained from the Web.

The map server platform supports the primary ArcIMS Web services. Map services are developed by the ArcIMS Manager design and authoring tools producing templates on the application server that manage the map service and performance as a service queue. Additional programming tools (Cold Fusion [CF] and Active Server Pages [ASP]) can be used on the Web server to enhance functionality of thin HTML clients.

4.7.1 ArcIMS Server Configuration Alternatives

There are several ways to configure an ArcIMS site. **Figure 4-10** provides an overview of the ArcIMS components and their locations.

Figure 4-10
ArcIMS Component Architecture



The location of the ArcIMS components and their configuration can directly impact the Web site capacity, reliability of the site, and overall output performance. The ArcIMS components can be divided into the following categories.

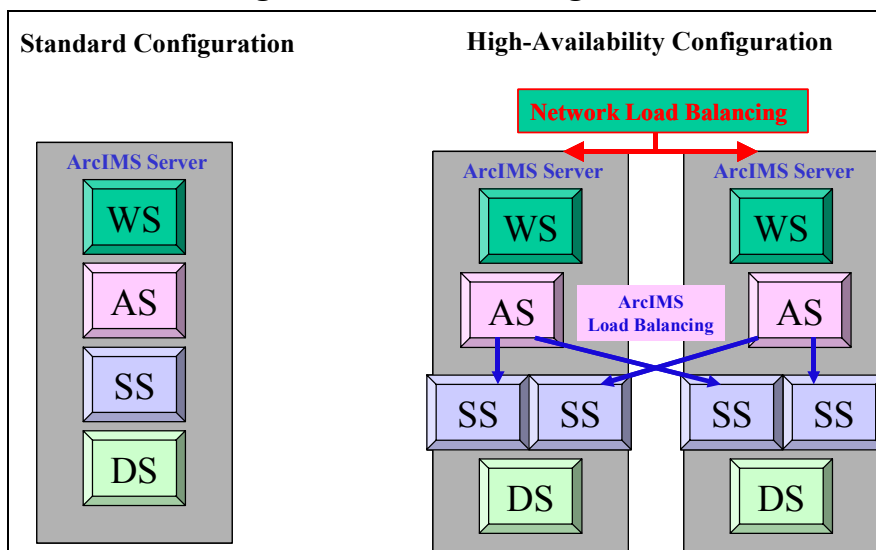
- **Web Server and Connectors (WS).** The Web server component supports communication between the ArcIMS map services and the Web client applications. The connectors on the Web server translate Web HTTP traffic to communication understood by the ArcIMS Web services. A very limited amount of processing is required to support this server.

- **ArcIMS Application Server (AS).** The ArcIMS application server component provides an application layer that supports the inbound map service request queues (virtual servers) and registered connects to the ArcIMS public service engines (Image, ArcMap, Feature, Extract). Inbound requests are routed to available service instances for processing. Roughly 10 percent of the ArcIMS processing load is required to support the application server functions.
- **Spatial Server (SS).** The ArcIMS spatial server includes the service engines (Image, Feature, Extract, Query, ArcMap and Geocode) that service the map requests. The spatial server also includes a metadata search engine. The spatial server scripts (engines) represent most of the ArcIMS processing requirements.
- **Data Server (DS).** The data server (GIS data source) is where the GIS data is stored. An ArcSDE data source supports query processing, roughly 20 percent of typical map service processing.

A minimum ArcIMS site can be configured with as few as one platform or as many as six platforms, depending on site requirements. Alternatives can be divided into single-tier, two-tier, and three-tier platform configurations. Simple configurations are easier to maintain and support. The more complex configurations support higher capacity sites and higher availability.

- **Single-tier Platform Configuration.** Figure 4-11 provides an overview of some example single-tier platform configurations. Single-tier configurations are supported with one or two platforms capable of supporting all the ArcIMS components.

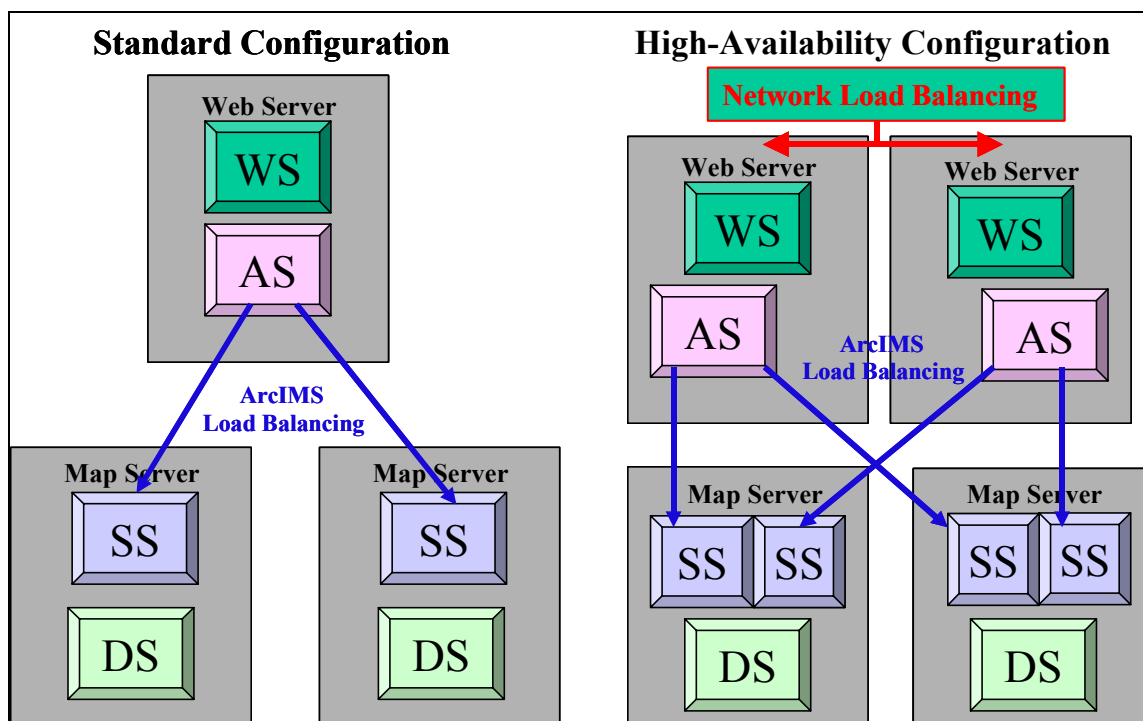
Figure 4-11
Single-tier Platform Configuration



- **Standard Configuration.** A complete ArcIMS site can be configured on a single hardware platform. This configuration would be appropriate for map service development, sites with limited number of service requests, and initial prototype development.

- **High-Availability Configuration.** Some ArcIMS operations require redundant server solutions, configured so the site remains operational in the event of a single platform failure. This configuration includes: (1) network load balancing to route the traffic to each of the servers during normal operations, and only to the active server if one of the servers fail; (2) ArcIMS load balancing to distribute spatial server processing load between the two platforms to avoid having requests back up on one server when extra processing resources are available on the other server; two spatial servers are required on each platform to support this configuration; and (3) each server will require a complete copy of the data.
- **Two-tier Platform Configuration.** Figure 4-12 provides an overview of some example two-tier platform configurations. The two-tier configuration includes a Web server and a map server, with the application server located on the Web server. Map server and data are on the same platform.

Figure 4-12
Two-tier Platform Configuration

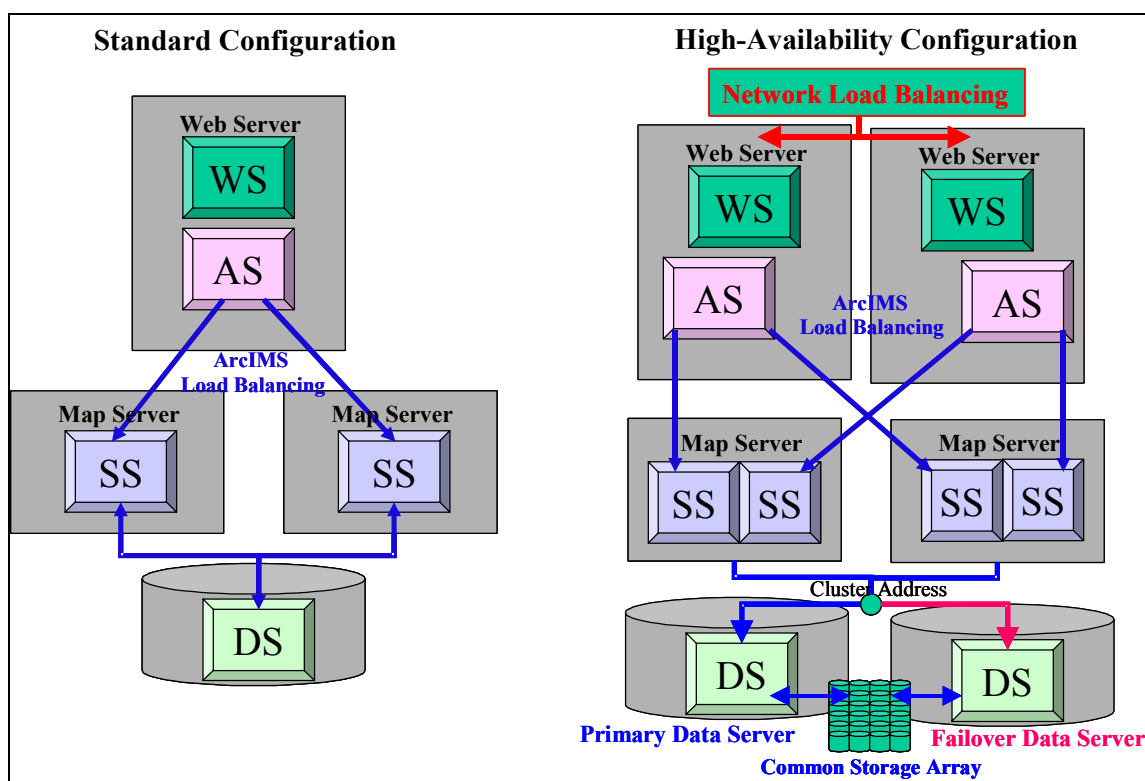


- **Standard Configuration.** The standard two-tier configuration includes a single Web server with a separate map server layer. The map server layer can be a single platform, or can be expanded to support several platforms, depending on the required site capacity. ArcIMS load balancing is provided by the application server.
- **High-Availability Configuration.** High-availability operations require redundant server solutions, configured so the site remains operational in the event of any single platform failure. This configuration includes: (1) network load balancing to route the traffic to each

of the Web servers during normal operations, and only to the active Web server if one of the servers fail; (2) ArcIMS load balancing to distribute spatial server processing load between the two map server platforms to avoid having requests back up on one server when extra processing resources are available on the other server; two spatial servers are required on each map server platform to support this configuration; and (3) each map server will require a complete copy of the data.

- **Three-tier Platform Configuration.** Figure 4-13 provides an overview of some example three-tier platform configurations. The three-tier configuration includes a Web server, a map server, and a data server tier with the application server located on the Web server.

Figure 4-13
Three-tier Platform Configuration



- **Standard Configuration.** The standard three-tier configuration includes a single Web server with a separate map server and data server tier. The map server tier can be a single platform or expanded to support several platforms, depending on the required site capacity. ArcIMS load balancing is provided by the application server. Both map servers use a common data source.
- **High-Availability Configuration.** This configuration includes two Web servers, two map servers, and two data servers. This solution includes: (1) Network load balancing is provided to route the traffic to each of the Web servers during normal operations, and only to the active Web server if one of the servers fail; (2) ArcIMS load balancing distributes the spatial server processing load between the two map server platforms to avoid having requests

back up on one server when extra processing resources are available on the other server; two spatial servers are required on each map server platform to support this configuration; and (3) two data servers are clustered and connected to a common storage array data source. The primary data server supports query services during normal operations, and the secondary data server takes over query services when the primary server fails.

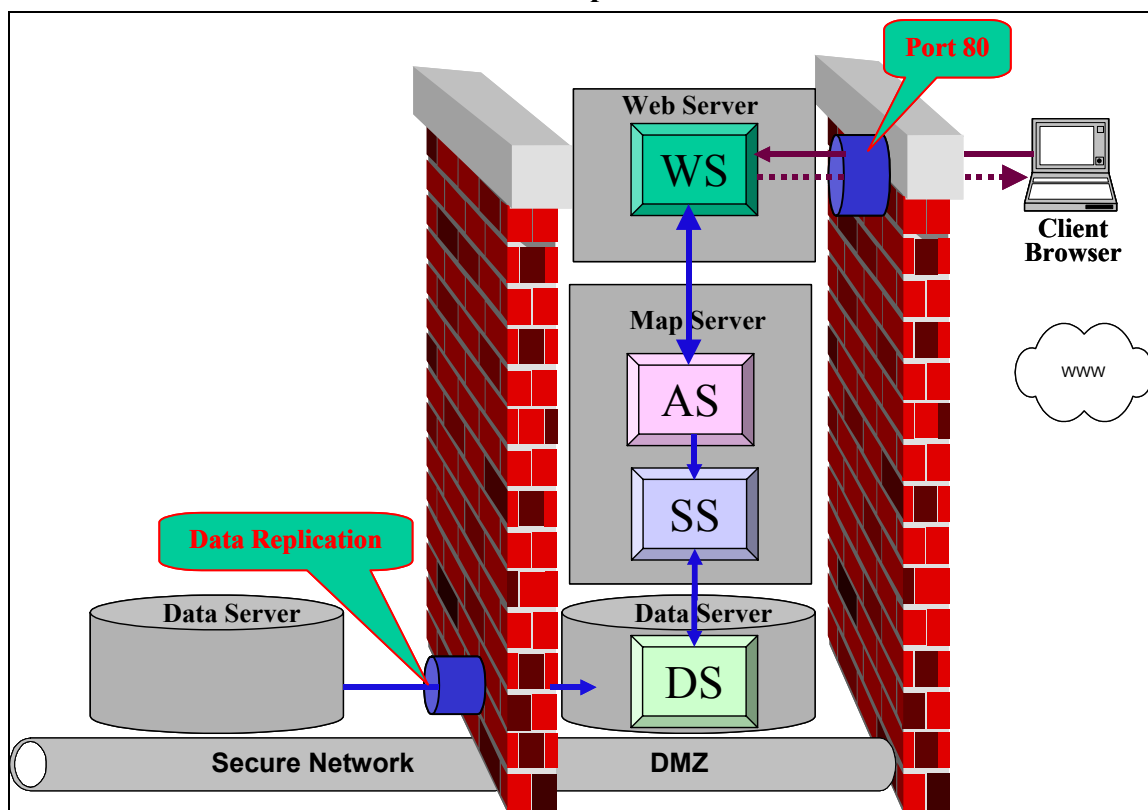
4.7.2 ArcIMS Firewall Configuration Alternatives

An ESRI white paper, *Security and ArcIMS*, addresses configuration options for secure ArcIMS environments. This paper is available on the ESRI Web site at the following URL:

<http://www.esri.com/library/whitepapers/pdfs/SecurityArcIMS.pdf>. Firewall configurations are provided to support various levels of security. A number of security options are identified here, based on the location of the ArcIMS software components.

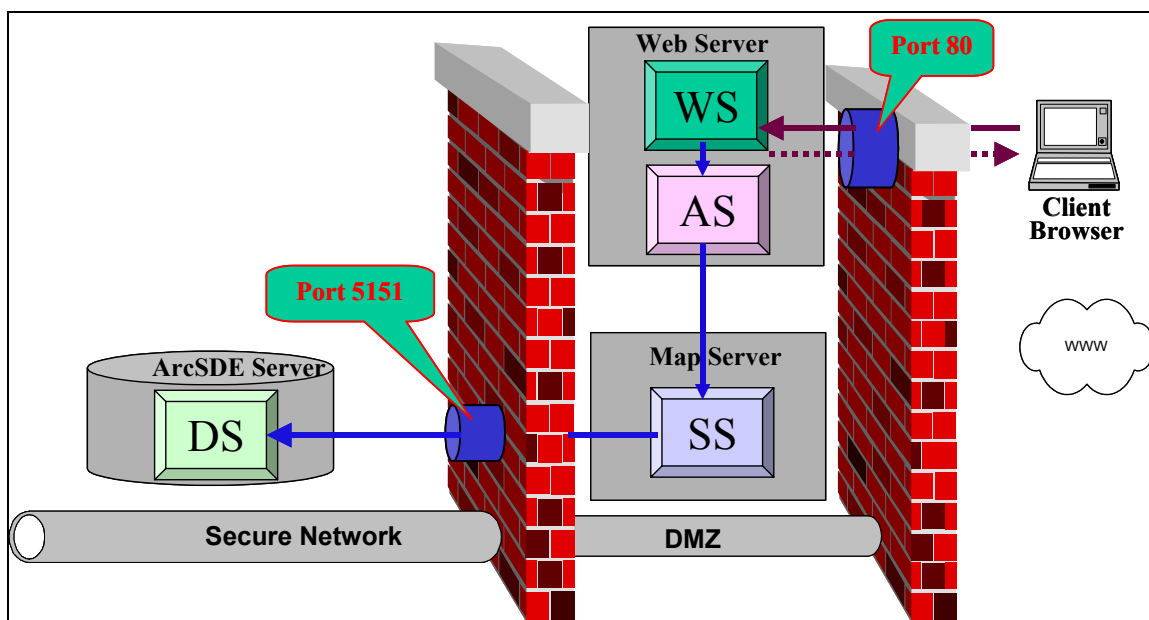
- **All ArcIMS Components in DMZ.** The most secure solution provides physical separation of the secure network from all of the ArcIMS software components. **Figure 4-14** shows the Web server, application server, spatial server and data server, are all located outside the secure network firewall within the DMZ. This configuration requires maintenance of duplicate copies of the GIS data. Data must be replicated from the internal GIS data server to the external data server supporting the ArcIMS services.

Figure 4-14
All ArcIMS Components in DMZ



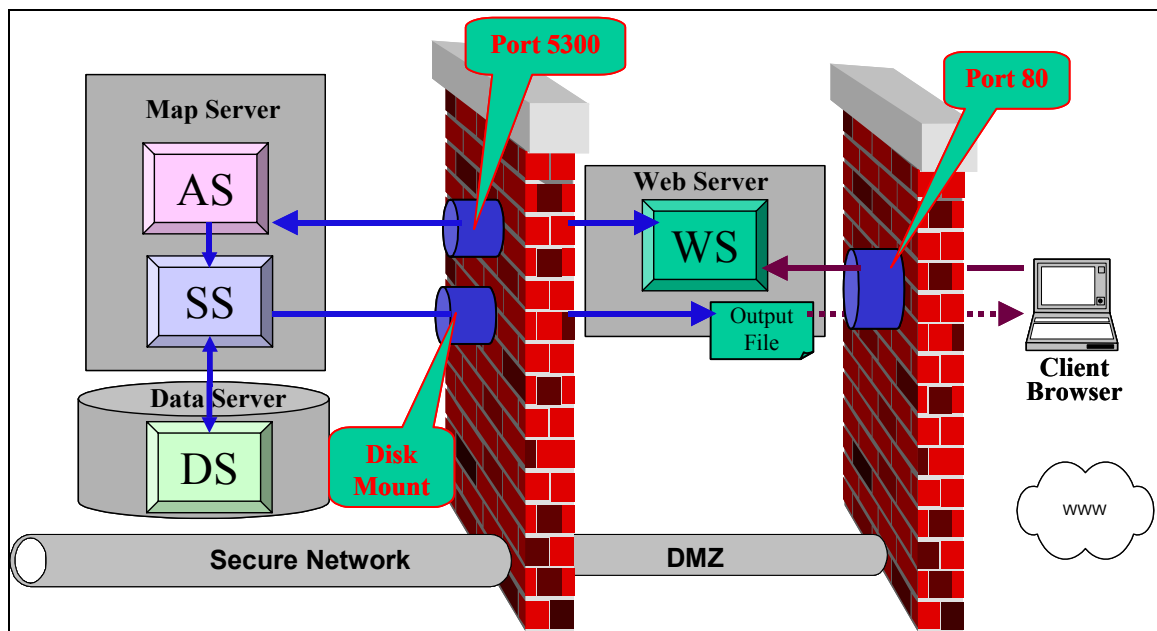
- **All ArcIMS Components in DMZ except Data Server.** Figure 4-15 shows the Web server, application server, and spatial server located in the DMZ, accessing the internal ArcSDE data server located on the secure network. Port 5151 access through the secure firewall will allow limited access to the ArcSDE DBMS data server. Security is provided through the ArcIMS services and the DBMS security. Firewall prevents traffic from any other source to access the internal secure network.

Figure 4-15
All ArcIMS Components in DMZ except Data Server



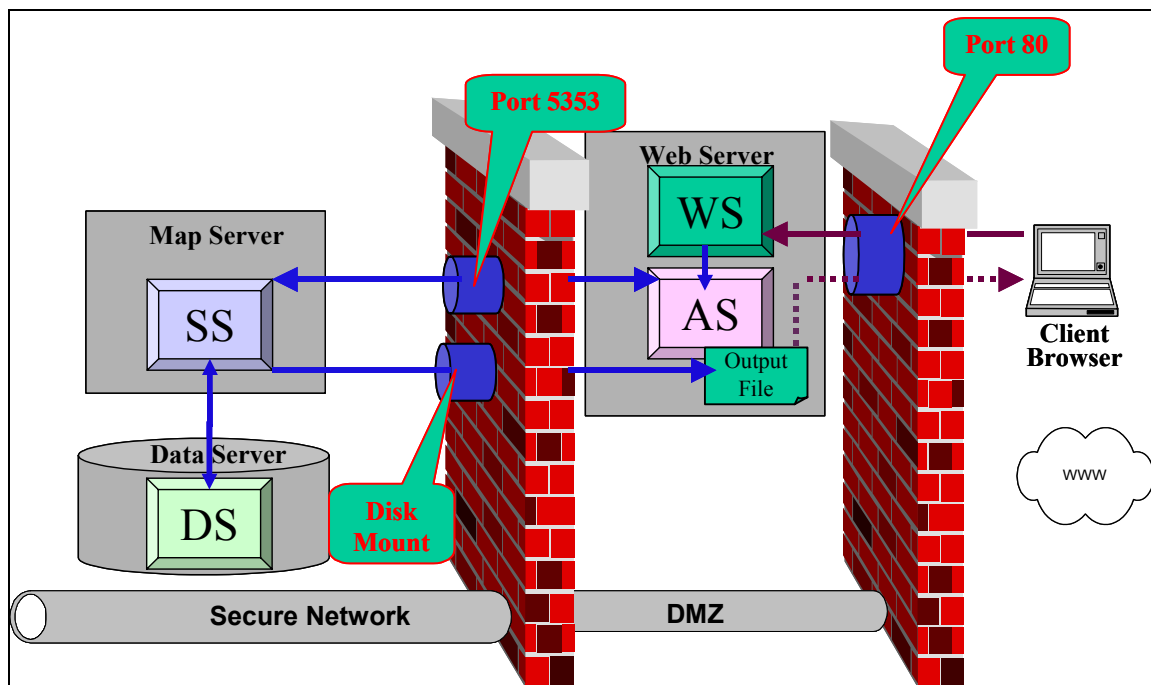
- **Web Server in the DMZ, Remainder of the ArcIMS Components on the Secure Network.** Figure 4-16 shows the Web server located in the DMZ, with the map server and data server located on the secure network. The application server and spatial server must be located on the internal network for this configuration to be acceptable. The output file, located on the Web server, must be shared to the map server. This disk mount will support one-way access from the map server through the firewall to the Web server.

Figure 4-16
Web Server in DMZ, Remainder of Configuration on Secure Network



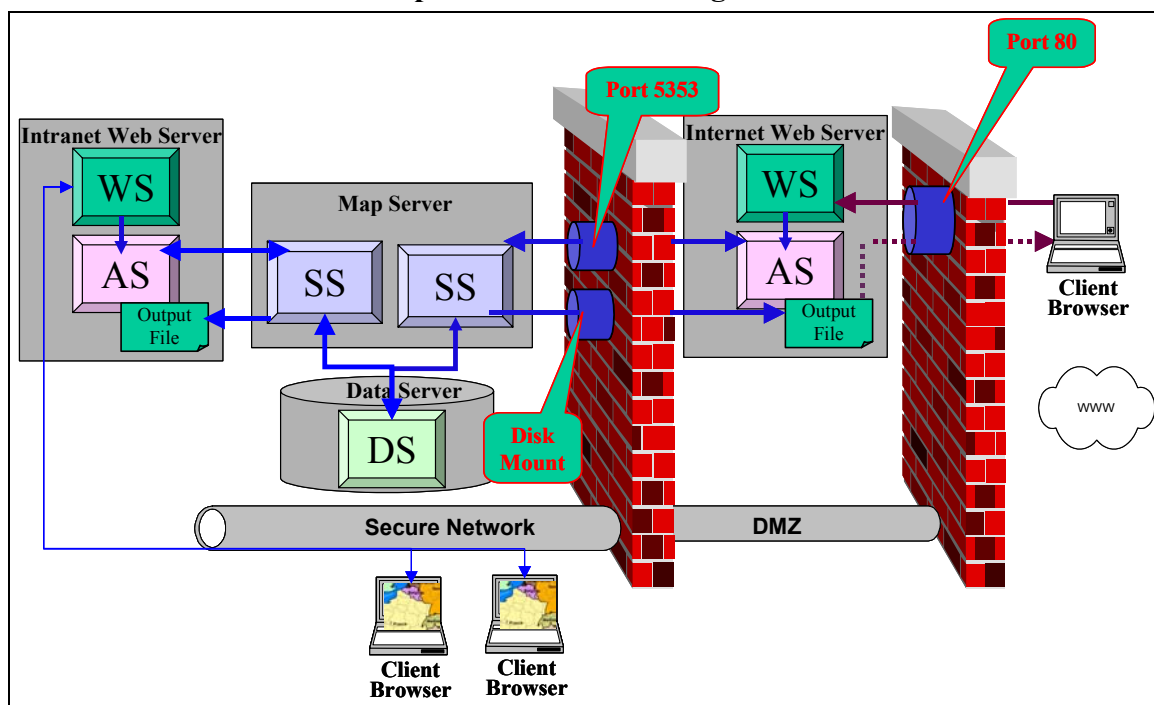
- **Web Server and Application Server in the DMZ, Remainder of the ArcIMS Components on the Secure Network.** Figure 4-17 shows the Web server and application server located in the DMZ, with the map server and data server located on the secure network. The output file, located on the Web server, must be shared to the map server. This disk mount will support one-way access from the map server through the firewall to the Web server.

Figure 4-17
Web and Application Server in DMZ, Remainder on Secure Network



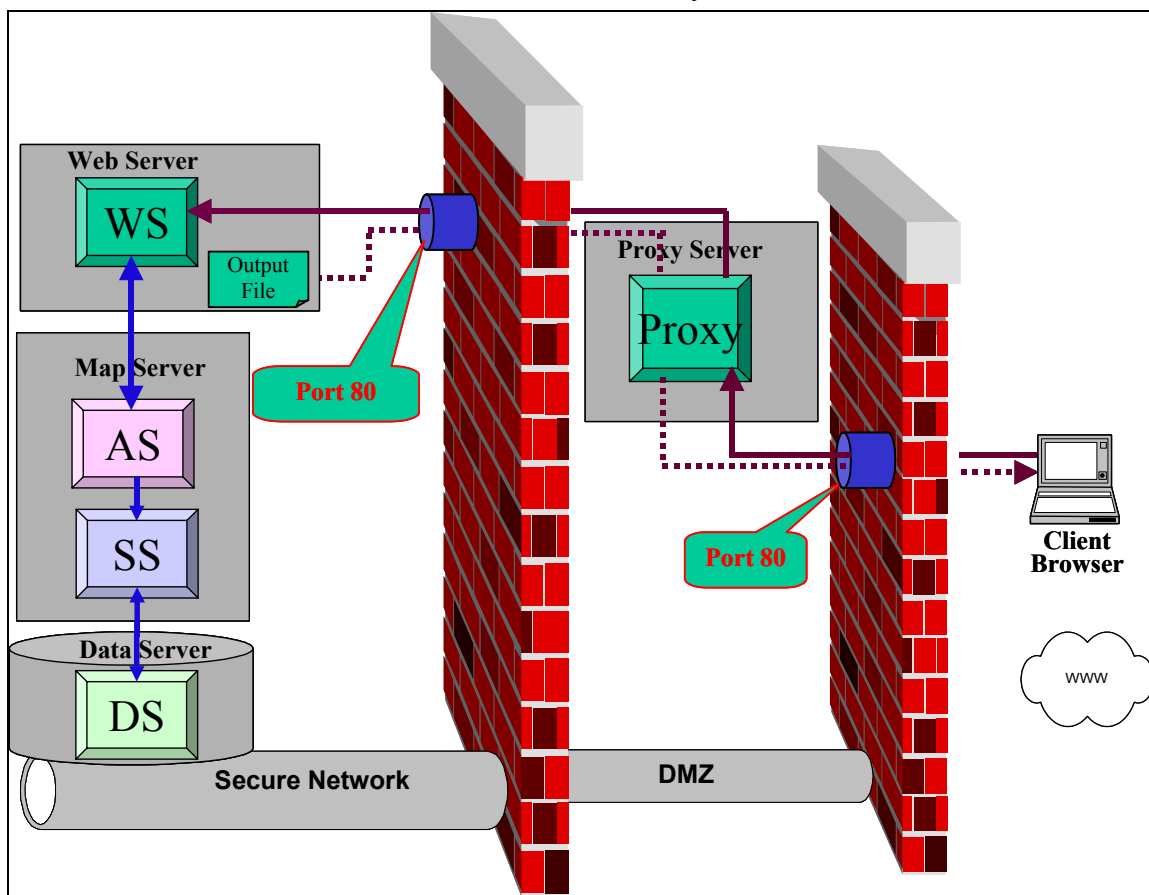
- **Multiple Web Server Configuration.** Figure 4-18 shows a multiple Web server solution, providing separate access security to Intranet and Internet browser clients. This hybrid solution provides shared use of a central map server compute environment while supporting separate user access security requirements. Separate map services can be deployed on the two Web servers, providing secure access to separate user communities from the same ArcIMS site.

Figure 4-18
Multiple Web Server Configuration



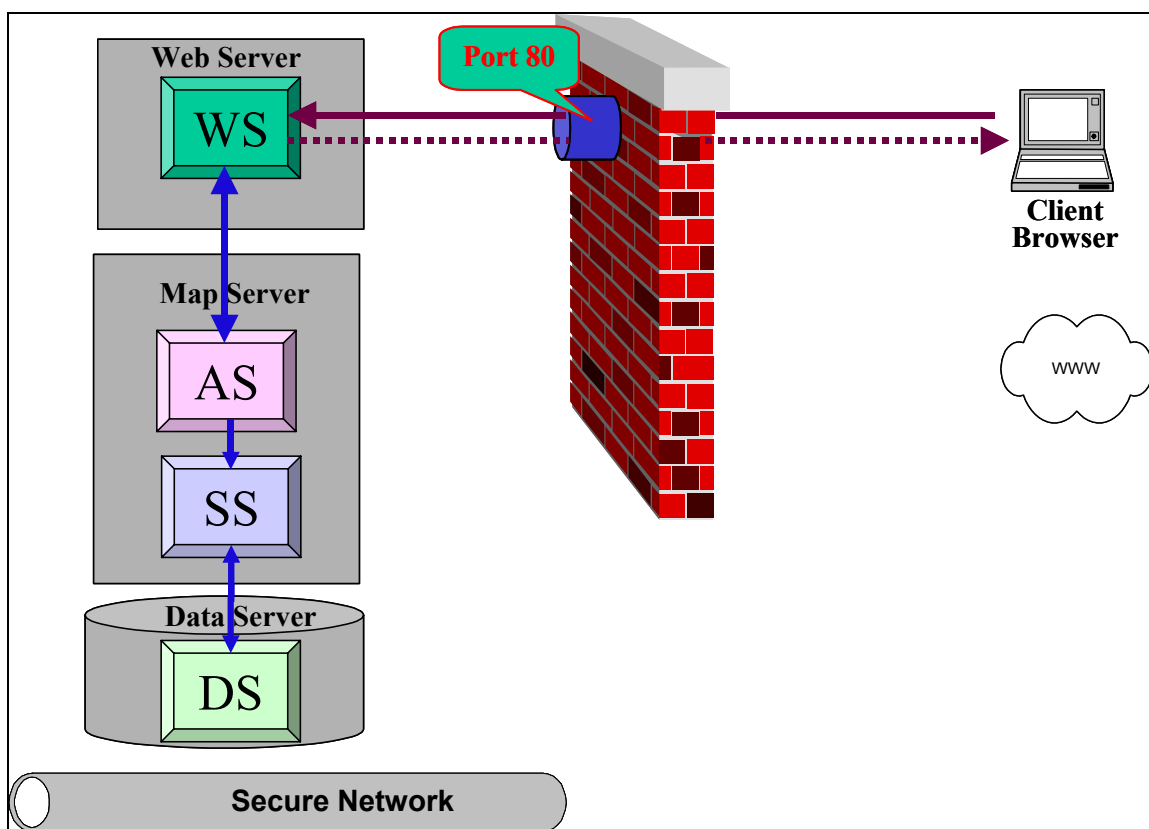
- **Web Services with Proxy Server.** Figure 4-19 shows interface with intranet Web Server configuration supported by a proxy server. This solution provides private network security through a proxy server, and supports the complete Web configuration on the private network. This configuration enables full management of the Web site on the private network.

Figure 4-19
Web Services with Proxy Server



- **All of the ArcIMS Components on the Secure Network.** Figure 4-20 shows the Web server, map server, and data server components all inside the firewall on the secure network. Port 80 must be open to allow HTTP traffic to pass through the firewall. Many organizations are not comfortable with this level of security.

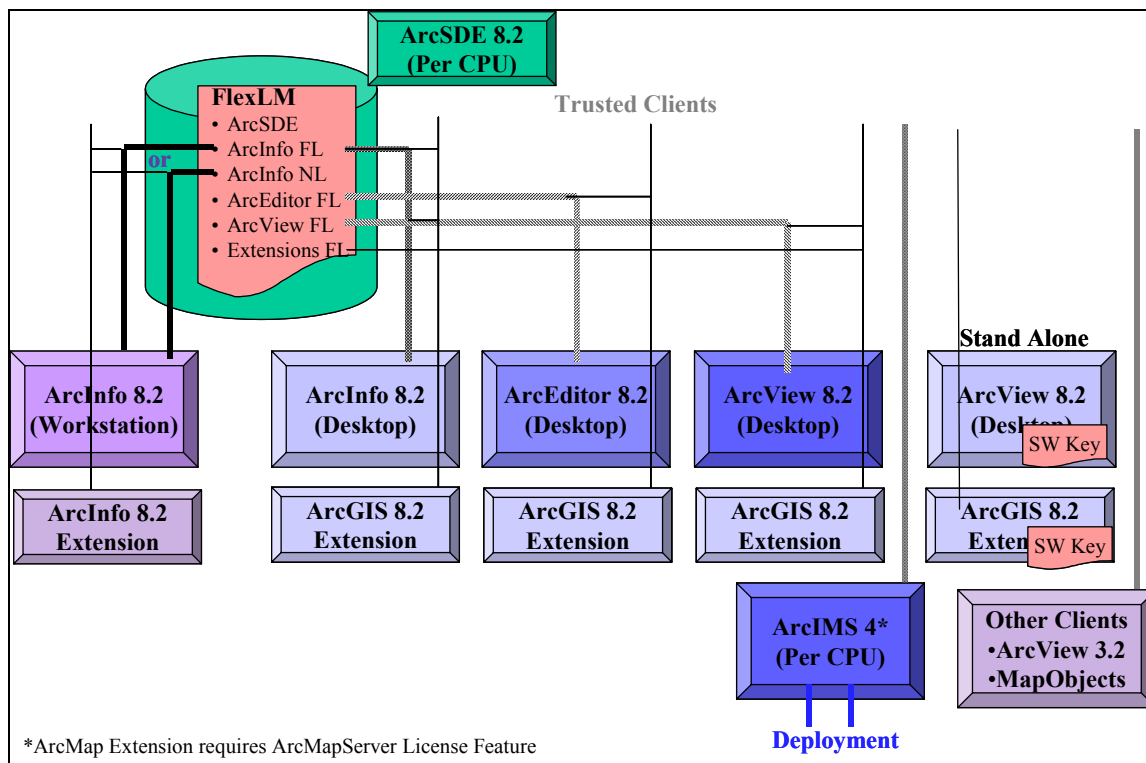
Figure 4-20
All ArcIMS Components on Secure Network



4.8 ESRI Software Licensing

Some changes were made in the ArcInfo, ArcView, and ArcSDE licensing with the deployment of the ArcGIS software (ArcInfo, ArcEditor, and ArcView) and their associated extensions.

Figure 4-21
ESRI Software Licensing



ArcView Standalone Desktop. ArcView desktop can be installed as a single-user license controlled by a software key. The license is issued for a specific platform and the license manager is downloaded over the Internet during installation. ArcGIS extensions for the single-user license will be supported in the same way. This license manager is not authorized for use on a Windows Terminal Server concurrent user licensing environment.

ArcInfo Workstation. Workstation ArcInfo licensing has not changed, and is controlled by a FlexLM license manager. Floating and node-lock UNIX licensing options are available, floating licenses are supported on the Windows platform.

A floating license authorizes a peak concurrent number of ArcInfo sessions configured on the network. Each concurrent user requires one license. Each distributed ArcInfo service (batch processing) will require one license.

A node-lock license supports terminal users running ArcInfo sessions on a UNIX application server. Licenses are bundled in groups of three or more users. Each concurrent user requires one license.

Licensing for ArcInfo Workstation extensions are configured in the FlexLM license manager. Each concurrent user requires one extension and one ArcInfo license for the extension software session.

ArcInfo Desktop. Desktop ArcInfo is controlled by the same FlexLM license manager as Workstation ArcInfo. ArcInfo desktop is supported on the Windows platform only, and will use the ArcInfo floating licenses. Each concurrent user requires one license.

ArcEditor Desktop. Desktop ArcEditor is controlled by the FlexLM license manager. Each concurrent user will require one floating license.

ArcView Desktop. A concurrent user license is available for Desktop ArcView. Each concurrent session is controlled by the FlexLM license manager. Each concurrent user will require one ArcView floating license.

ArcGIS Desktop Extensions. All of the ArcGIS Desktop extensions are supported with each of the ArcGIS core licenses. Each concurrent session will require the respective extension floating license. Each user session will also require one of the associated desktop (ArcInfo, ArcEditor, or ArcView) licenses. Each concurrent user requires one extension and one ArcGIS (ArcInfo, ArcEditor, or ArcView) license.

ArcSDE Licensing. ArcSDE licensing is based on number of CPUs on the associated DBMS server. There is no ArcSDE connection license required for ArcGIS and ArcIMS clients, and there is no control of the number of users accessing the server.

4.9 Data Management Technology Overview

Data management is a primary consideration when developing enterprise GIS architectures. Enterprise GIS normally benefits from efforts to consolidate agency GIS data resources. There are several reasons for supporting data consolidation. These reasons include improving user access to data resources, providing better data protection, and enhancing the quality of the data. Consolidation of IT support resources also reduces hardware cost and the overall cost of system administration.

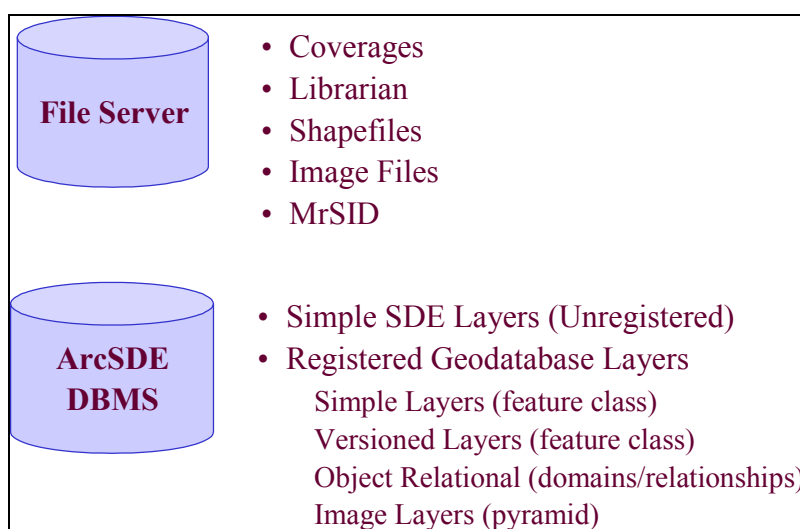
The simplest and most cost-effective way to manage data resources is to keep one copy of the data in a central data repository, and provide required user access to these data to support data maintenance and operational GIS query and analysis needs. This is not always practical, and many system solutions require that organizations maintain distributed copies of the data. There are significant compromises that may be required to support distributed data architectures. This chapter provides an overview of data management technology. Several basic data management tasks will be identified, along with the current state of the technology in supporting these tasks. These data management tasks include the following:

- Ways to store spatial data
- Ways to connect to spatial data
- Ways to protect spatial data
- Ways to backup spatial data
- Ways to move spatial data
- New ways to access spatial data

4.9.1 Ways to Store to Spatial Data

There are several different spatial data formats. These formats can be classified at a high level based on how we store and access data. Data can be stored in a native GIS format or imported into a data management system environment. The various standard spatial storage formats are identified in **Figure 4-22**.

Figure 4-22
Ways to Store Spatial Data



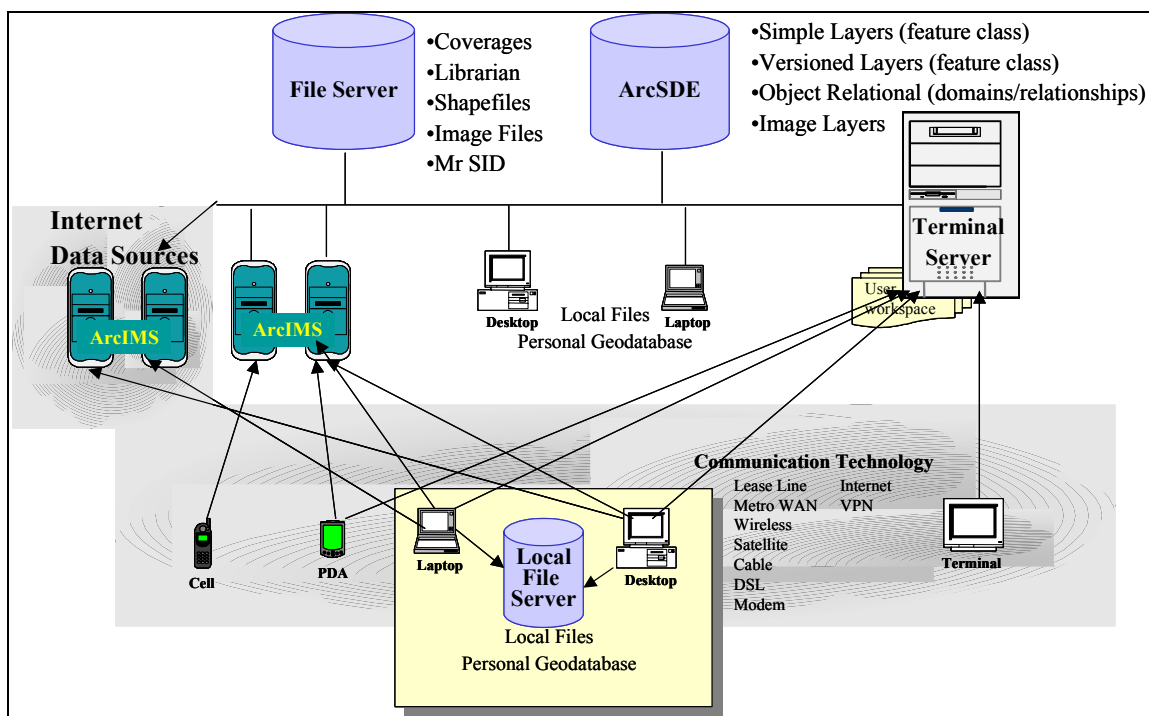
File Server Data Formats. File servers are used to store ArcInfo coverages, LIBRARIAN data files, shapefiles, image files, and images stored in MrSID compressed files. GIS applications support the query processing for file-based data sources. Network data sources are connected to the application through a disk mount, where the remote file system appears as a local application data source.

ArcSDE Server Data Formats. ArcSDE servers are used to store spatial vector data and image layers within a DBMS. ArcSDE supports spatial data stored in Oracle, SQL 2000, Informix, and DB2 DBMS environments. Spatial vector data can be stored as unregistered simple layers or as registered geodatabase layers. Geodatabase formats include simple layers, versioned layers, and object-relational data layers. ArcSDE also supports storage of image layers in an pyramid tiled structure in the DBMS.

4.9.2 Ways to Connect to Spatial Data

Growing acceptance and deployment of a broad range of communication standards and application interface devices has expanded user choices for accessing and displaying GIS information products. **Figure 4-23** shows a number of display devices that support ESRI application clients, and how these devices connect to spatial data products.

Figure 4-23
Ways to Connect to Spatial Data



The following types of connections are supported from these clients:

Local Area Network Connections. Data access within a local area network (LAN) is provided through application on desktop workstations, terminal servers, and ArcIMS servers. File server access is provided through standard operating system disk mounts. ArcSDE access is supported through the ArcSDE application program interface (API) or using an ODBC (open database connectivity standard) direct-connect interface to the DBMS client API.

Remote Data Connections. Standard recommended best practices restrict disk mount and direct DBMS connections to local area network (LAN) connections. ArcIMS Web mapping services and Windows Terminal Server client connections provide the most efficient data access over wide area network (WAN) or dial-up connections. ArcSDE connections are possible, but often result in unacceptable network traffic or performance impacts.

Remote Site Connections. Remote sites can access their respective local data servers, which may be replicated copies of central database environments. Users located at remote sites can also access ArcIMS data sources over the WAN, or use terminal client access to applications supported on central Windows Terminal Server environments.

Wireless Connections. Portable devices (PDAs) can support ArcIMS browser and Windows Terminal Server Citrix clients over wireless connections. ArcIMS sources can also provide simple mapping information products for cell phone displays.

4.9.3 Ways to Protect Spatial Data

Enterprise GIS environments depend heavily on GIS data to support a variety of critical business processes. Data is one of the most valuable resources of a GIS, and protecting these data is fundamental to supporting normal operations.

The primary data protection line of defense is provided by the storage solutions. Most storage vendors have standardized on RAID storage solutions for data protection. A brief overview of basic RAID storage technology is provided in the following figures.

Figure 4-24
Ways to Store Spatial Data
RAID (Redundant Array of Independent Disks)

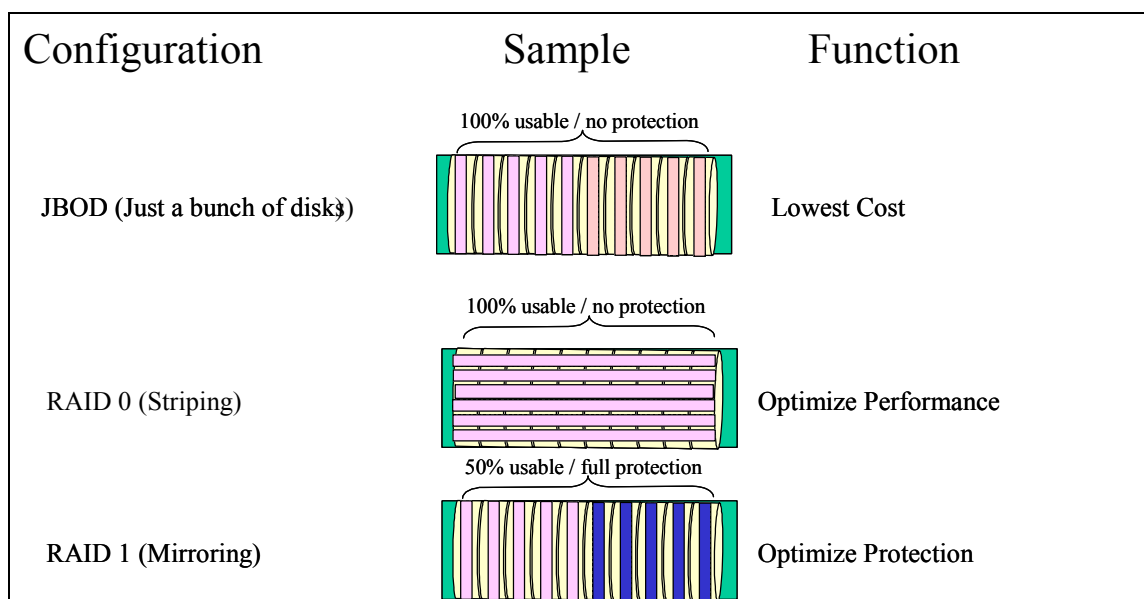


Figure 4-24 provides an overview of the basic RAID technology concepts. These include the following disk configurations:

JBOD (Just a bunch of disks). A disk volume with no RAID protection is referred to as a JBOD configuration, or just a bunch of disks. This represents a configuration of disks with no protection and no performance optimization.

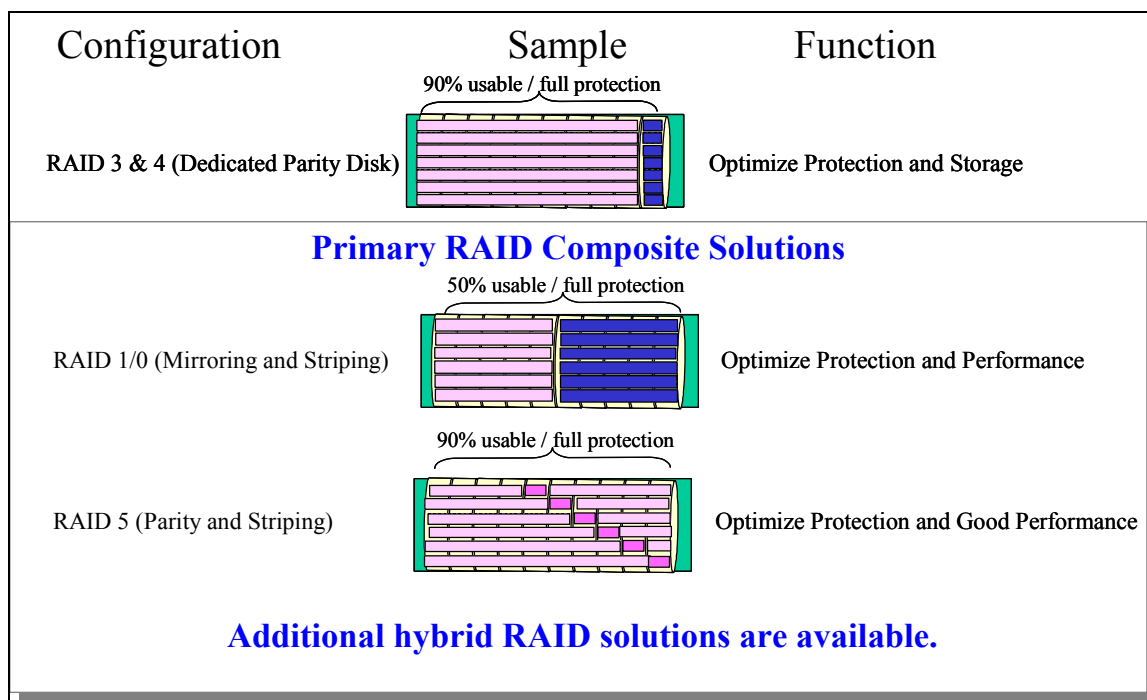
RAID 0. A disk volume in a RAID 0 configuration provides striping of data across several disks in the storage array. Striping supports parallel disk controller access to data across several disks reducing time required to find the requested data. Data is transferred to array cache once it is found on each disk. RAID 0 striping provides optimum data access performance with no data protection. 100 percent of the disk volume is available for data storage.

RAID 1. A disk volume in a RAID 1 configuration provides mirror copies of the data on disk pairs within the array. If one disk in a pair fails, data can be accessed from the remaining disk

copy. The failed disk can be replaced and data restored automatically from the mirror copy without bringing the storage array down for maintenance. RAID 1 provides optimum data protection with minimum performance gain. Available data storage is limited to 50 percent of the total disk volume, since a mirror disk copy is maintained for every data disk in the array.

Figure 4-25 provides an overview of the basic RAID 3 and 4 configuration options, and then provides an overview of the most popular composite RAID configuration strategies.

Figure 4-25
Ways to Store Spatial Data
(additional RAID configurations)



RAID 3 and 4. A disk volume in a RAID 3 or RAID 4 configuration supports striping of data across all disks in the array except for one parity disk. A parity bit is calculated for each data stripe and stored on the parity disk. If one of the disks fails, the parity bit can be used to recalculate and restore the missing data. RAID 3 provides good protection of the data and allows optimum use of the storage volume. All but one parity disk can be used for data storage, optimizing use of the available disk volume for data storage capacity.

There are some technical differences between RAID 3 and RAID 4 which, for our purposes, are at a level beyond our current discussion. Both of these storage configurations have potential performance disadvantages. The common parity disk must be accessed for each write, which can result in disk contention under heavy peak user loads. Performance may also suffer due to requirements to calculate and store the parity bit for each write. Write performance issues are normally resolved through array cache algorithms on most high-performance disk storage solutions.

4.9.4 Popular RAID Configurations

The following RAID configurations are the most common used to support ArcSDE storage solutions. These solutions represent RAID combinations that best support data protection and performance goals.

RAID 1/0. RAID1/ 0 is a composite solution including RAID 0 striping and RAID 1 mirroring. This is the optimum solution for high performance and data protection. This is also the highest cost solution. Available data storage is limited to 50 percent of the total disk volume, since a mirror disk copy is maintained for every data disk in the array.

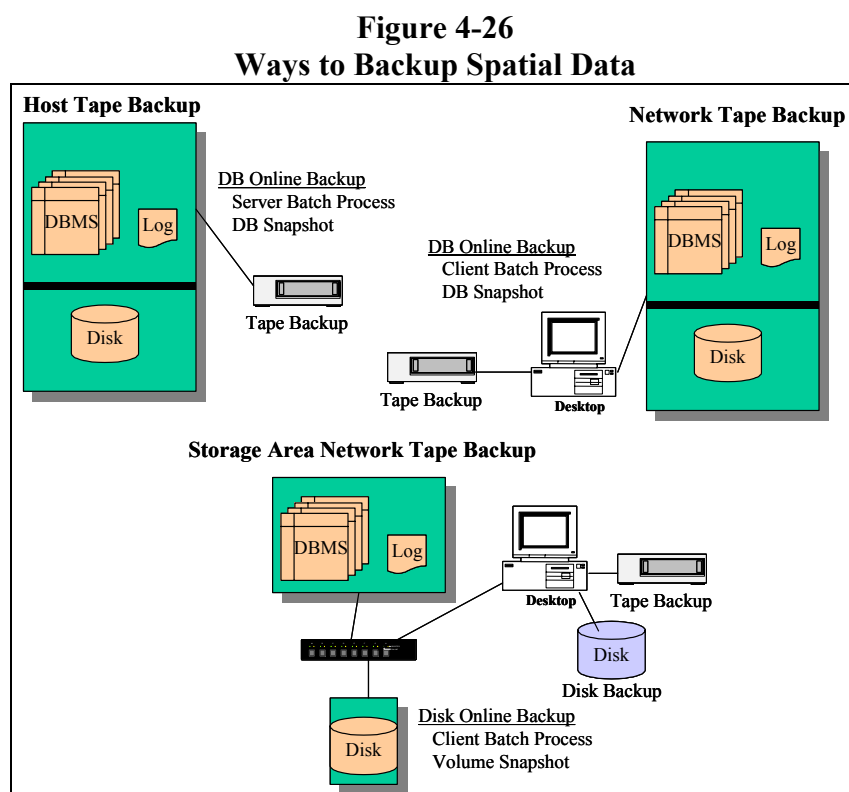
RAID 5. RAID 5 includes the striping and parity of the RAID 3 solution, and includes distribution of the parity volumes for each stripe across the array to avoid parity disk contention performance bottlenecks. This improved parity solution provides optimum disk utilization and near optimum performance, supporting disk storage on all but one parity disk volume.

Hybrid Solutions. Some vendors provide alternative proprietary RAID strategies to enhance their storage solution. New ways to store data on disk can improve performance and protection, and may simplify other data management needs. Each hybrid solution should be evaluated to determine if and how it may support specific data storage needs.

4.9.5 Ways to Backup Spatial Data

Data protection at the disk level will minimize the need for system recovery in the event of a single disk failure, but will not protect against a variety of other data failure scenarios. It is always important to keep a current backup copy of critical data resources at a safe known location away from the primary site.

Data backups typically provide the last line of defense for protecting our data investments. Careful planning and attention to storage backup procedures are important factors of a successful backup strategy. Data loss can result from many types of situations, with some of the most probable situations being administrative or user error. **Figure 4-26** provides an overview of the different ways to backup spatial data.



Host Tape Backup. Traditional server backup solutions use lower-cost tape storage for backup. Data must be converted to a tape storage format and stored in a linear tape medium. Backups can be a long drawn-out process taking considerable server processing resource (typically consume a CPU during the backup process) and requiring special data management for operational environments.

For database environments, these backups must occur based on a single point in time to maintain database continuity. Database vendors support on-line backup requirements by establishing a procedural snapshot of the database. A copy of the protected snapshot data are retained in a

snapshot table when changes are made to the database, supporting point-in-time backup of the database and potential database recovery back to the time of the snapshot.

Host processors can be used to support backup operations during off-peak hours. If backups are required during peak-use periods, backups can impact server performance.

Network Client Tape Backup. The traditional on-line backup can often be supported over the LAN with the primary batch backup process running on a separate client platform. DBMS snapshot may still be used to support point-in-time backups for on-line database environments. Client backup processes can contribute to potential network performance bottlenecks between the server and the client machine due to the high data transfer rates during the backup process.

Storage Area Network Client Tape Backup. Some backup solutions support direct disk storage access without impacting the Host DBMS server environment. Storage backup is performed over the SAN or through a separate fiber channel access to the disk array with batch process running on a separate client platform. A disk-level storage array snapshot is used to support point-in-time backups for on-line database environments. Host platform processing loads and LAN performance bottlenecks can be avoided with disk-level backup solutions.

Disk Copy Backups. The size of databases has increased dramatically in recent years, growing from 10s of Gigabytes to 100s of Gigabytes and, in many cases, Terabytes of data. Recovery of large databases from tape backups is very slow, taking days to recover large spatial database environments. At the same time, the cost of disk storage has decreased dramatically providing disk copy solutions for large database environments competitive in price to tape storage solutions. A copy of the database on local disk, or copy of these disks to a remote recovery site, can support immediate restart of the DBMS following a storage failure simply by restarting the DBMS with the backup disk copy.

4.9.6 Ways to Move Spatial Data

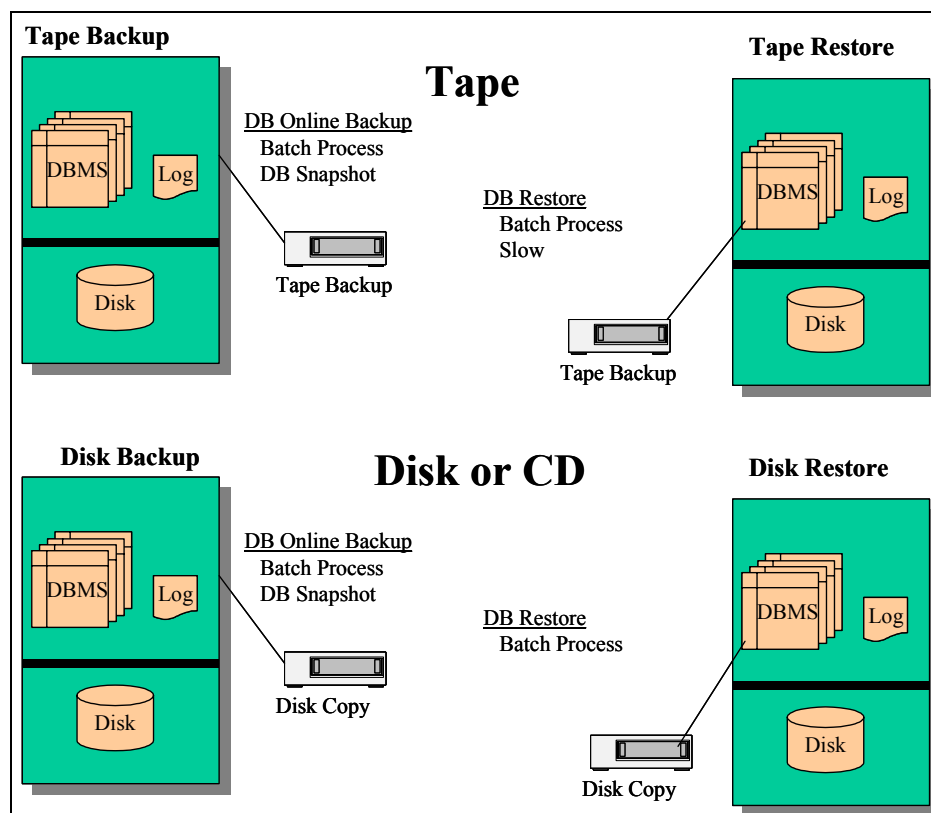
Many enterprise GIS solutions require continued maintenance of distributed copies of the GIS data resources, typically replicated from a central GIS data repository or enterprise database environment. Organizations with a single enterprise database solution still have a need to protect data resources in the event of an emergency such as fire, flood, accidents, or other natural disasters. Many organizations have recently reviewed their business practices, and updated their plans for business continuance in the event of a major loss of data resources. Events of this past year (loss of the New York Twin Towers on September 11, 2001) have emphasized the need and value of such plans, and increased interest and awareness of the need for this type of protection.

This section reviews the various ways organizations move spatial data. Traditional methods copy data on tape or disk and physically deliver these data to the remote site through standard transportation modes. Once at the remote site, data is reinstalled on the remote server environment. Technology has evolved to provide more efficient alternatives for maintaining distributed data sources. Understanding the available options and risks involved in moving data will be important in defining an optimum enterprise GIS architecture.

4.9.7 Traditional Data Transfer Methods

Figure 4-27 identifies traditional methods for moving a copy of data to a remote location.

Figure 4-27
Ways to Move Spatial Data
(Traditional Tape Backup/Disk Copy)



Traditional methods include backup and recovery of data using standard tape or disk transfer media. Moving data using these type of procedures are commonly called “sneaker net” providing a way to transfer data without the support of a physical network.

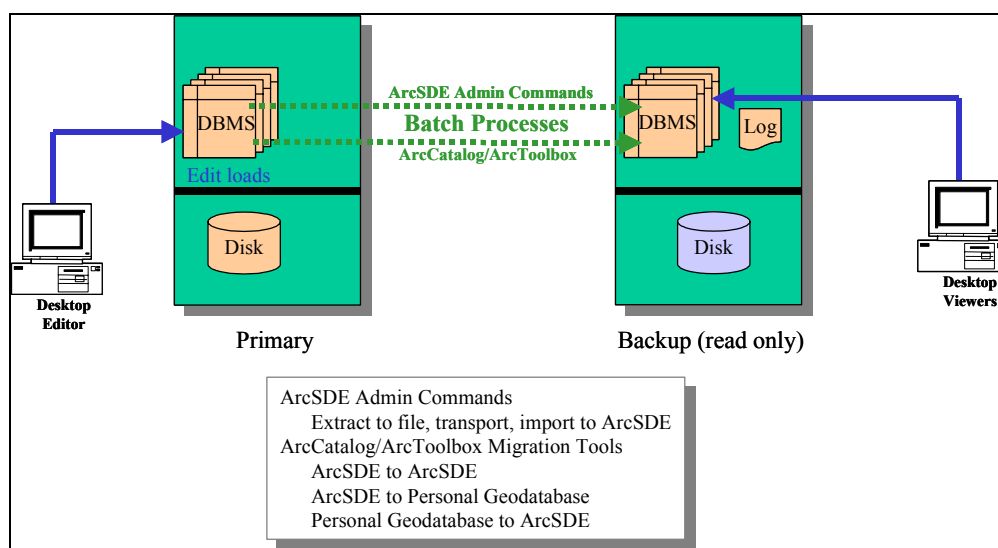
Tape Backup. Tape backup solutions can be used to move data to a separate server environment. Tape transfers are normally very slow. The reduced cost of disk storage has made disk copy a much more feasible option.

Disk Copy. A replicated copy of the database on disk storage can support rapid restore at a separate site. The database can be restarted with the new data copy and on-line with a very short recovery time.

4.9.8 ArcGIS Database Transition

Moving subsets of a single database cannot normally be supported with standard backup strategies. Data must be extracted from the primary database and imported into the remote database to support the data transfer. Database transition can be supported using standard ArcGIS export/import functions. These tools can be used as a method establishing and maintaining a copy of the database at a separate location. **Figure 4-28** identifies ways to move spatial data using ArcGIS data transition functions.

Figure 4-28
Ways to Move Spatial Data
(Database Transition)



ArcSDE Admin Commands. Batch process can be used with ArcSDE admin commands to support export and import of an ArcSDE database. Moving data using these commands are most practical when completely replacing the data layers. These commands are not optimum solutions when transferring data to a complex ArcSDE geodatabase environment.

ArcCatalog/ArcToolbox Commands. ArcCatalog supports migration of data between ArcSDE geodatabase environments, extracts to a personal geodatabase, and import from a personal geodatabase to an ArcSDE environment.

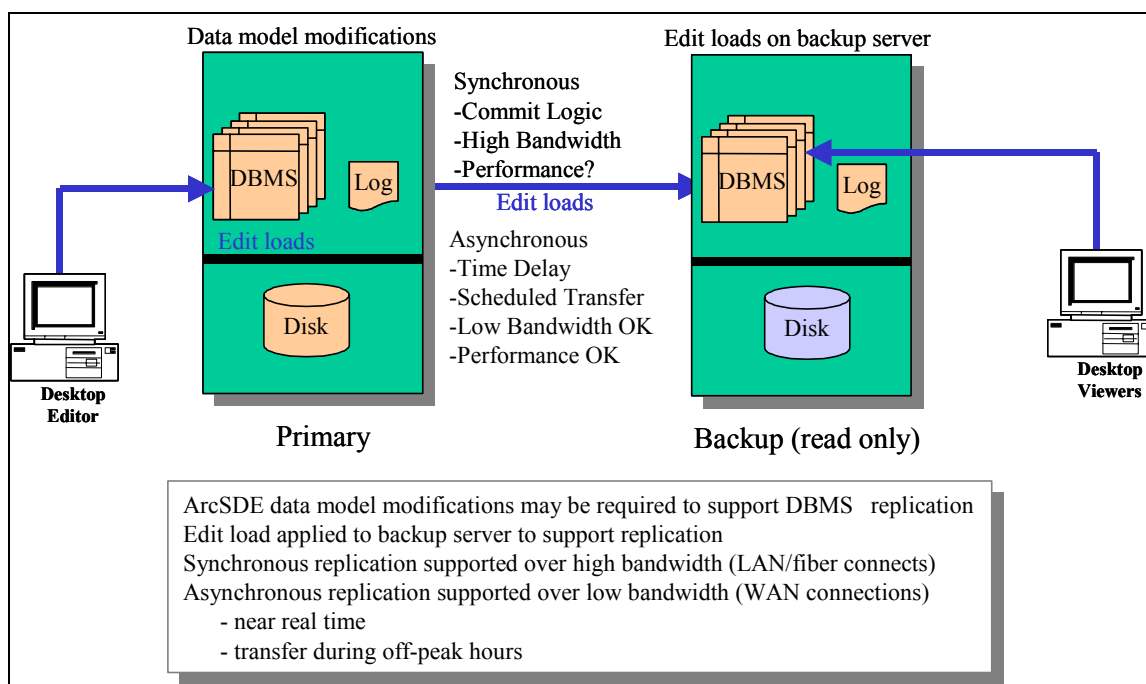
4.9.9 Database Replication

Database replication of spatial data using standard DBMS replication solutions has recently been supported for spatial data types within the database. Replication solutions have traditionally been very complex, and very few organizations have successfully implemented automatic replication solutions for ArcSDE spatial data.

Customers have experienced a variety of technical challenges when configuring DBMS spatial data replication solutions. ArcSDE data model modifications may be required to support DBMS replication solutions. Edit loads will be applied to both server environments, contributing to potential performance or server sizing impacts. Data changes must be transmitted over network connections between the two servers, causing potential communication bottlenecks. These challenges must be overcome to support a successful DBMS replication solution.

Customers have indicated that DBMS replication solutions can work but require a considerable amount of patience and implementation risk. Acceptable solutions may be available by some DBMS vendors to support replication to a read-only backup database server. Dual master server configuration strategies significantly increase the complexity of an already complex replication solution. **Figure 4-29** presents the different ways to move spatial data using database replication.

Figure 4-29
Ways to Move Spatial Data
(Database Replication)



Synchronous Replication. Real-time replication requires commitment of data transfer to the replicated server before releasing the client application on the primary server. Edit operations with this configuration would normally result in performance delays due to the typical heavy

volume of spatial data transfers and the required client interaction times. High-bandwidth fiber connectivity (1000-Mbps bandwidth) is recommended between the primary server and the replicated backup server to minimize performance delays.

Asynchronous Replication. Near real-time database replication strategies decouple the primary server from the data transfer transaction to the secondary server environment. Asynchronous replication can be supported over WAN connections, since the slow transmission times are isolated from primary server performance. Data transfers (updates) can be delayed to off-peak periods if WAN bandwidth limitations dictate, supporting periodic updates of the secondary server environment at a frequency supporting operational requirements.

4.9.10 Disk-level Replication (Synchronous)

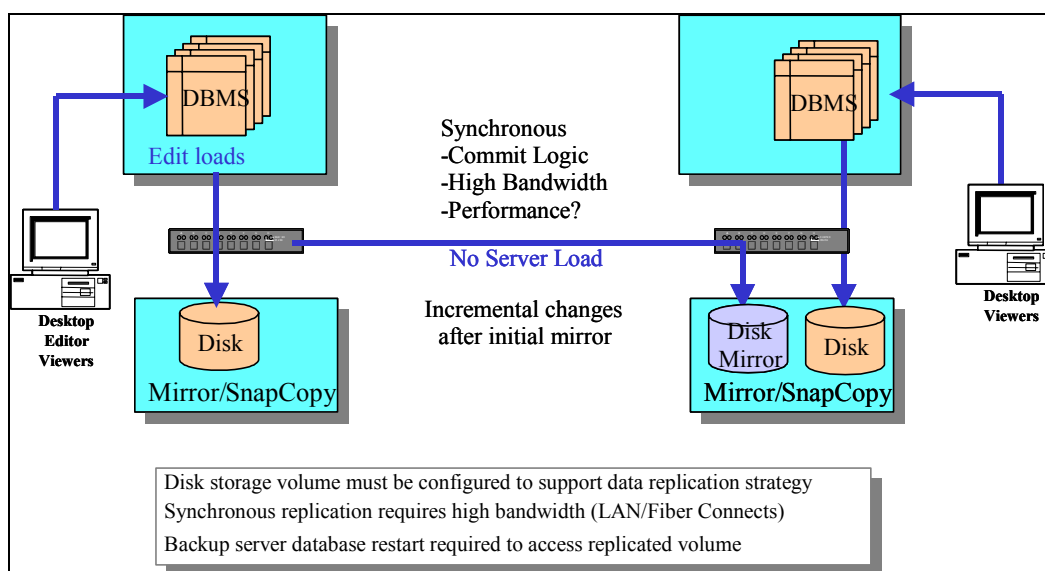
Disk-level replication is a well-established technology, supporting global replication of critical data for many types of industry solutions. Spatial data is stored on disk sectors very similar to any other data storage and, as such, does not require special attention beyond what might be required for other data types. Disk volume configurations (data location on disk and what volumes are transferred to the remote site) may be critical to ensure database integrity. Mirror copies are refreshed based on point-in-time snapshot functions supported by the storage vendor solution.

Disk-level replication provides transfer of block-level data changes on disk to a mirror disk volume located at a remote location. Transfer can be supported with active on-line transactions with minimum impact on DBMS server performance capacity. Secondary DBMS application must be restarted to refresh DBMS cache and processing environment to the point-in-time of the replicated disk volume.

Synchronous Replication. Real-time replication requires commitment of data transfer to the replicated storage array before releasing the DBMS application on the primary server. High-bandwidth fiber connectivity (1000-Mbps bandwidth) is recommended between the primary server and the replicated backup server to avoid performance delays.

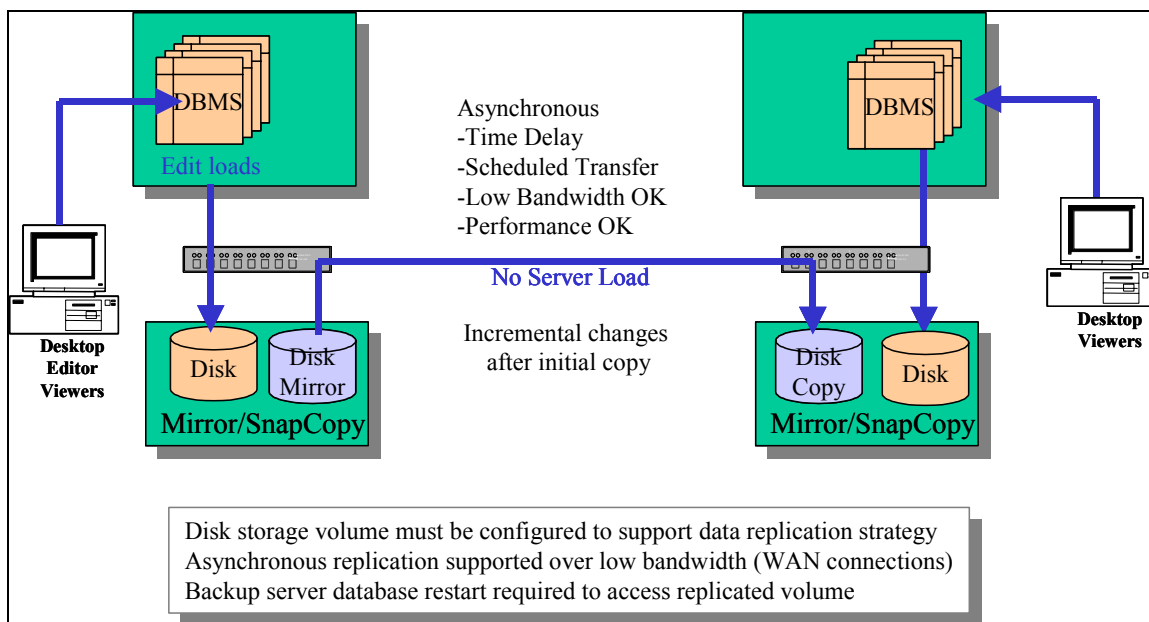
Figure 4-30 presents the different ways to move spatial data using synchronous disk-level replication.

Figure 4-30
Ways to Move Spatial Data
(Disk-level Replication)



Asynchronous Replication. Near real-time disk-level replication strategies decouple the primary disk array from the commit transaction of changes to the secondary storage array environment. Asynchronous replication can be supported over WAN connections, since the slow transmission times are isolated from primary DBMS server performance. Disk block changes can be stored and data transfers delayed to off-peak periods if WAN bandwidth limitations dictate, supporting periodic updates of the secondary disk storage volumes to meet operational requirements. **Figure 4-31** presents the different ways to move spatial data using asynchronous disk-level replication.

Figure 4-31
Ways to Move Spatial Data
(Disk-level Replication - Asynchronous)

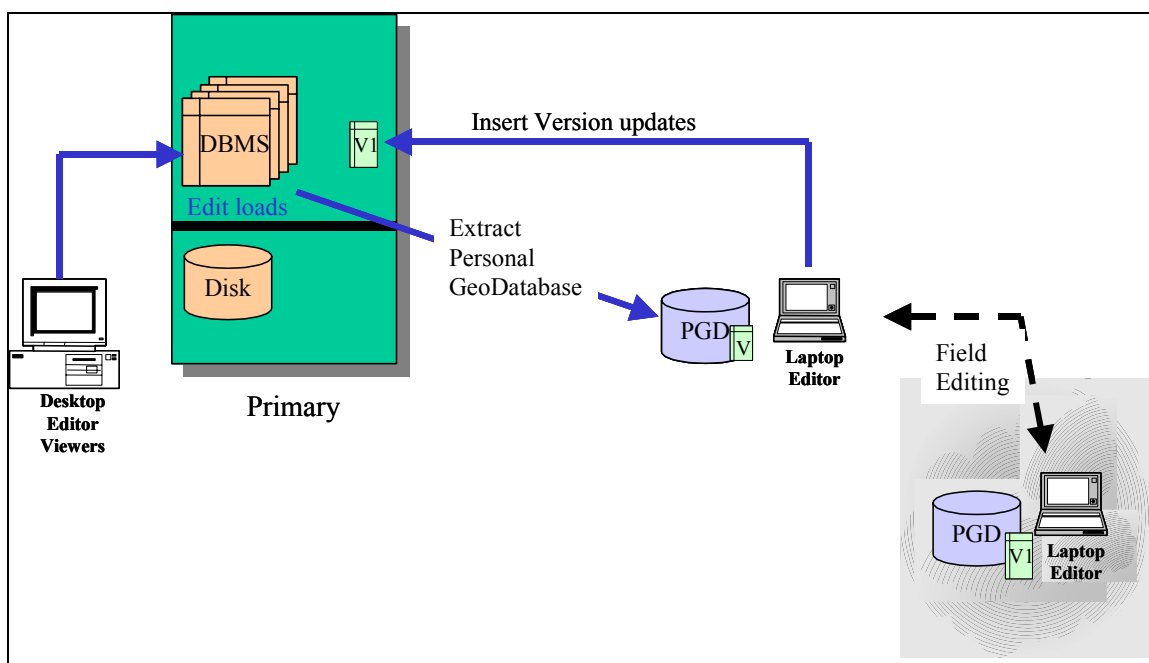


4.9.11 New Ways to Access Spatial Data

ESRI is supporting a new disconnected editing solution with the future ArcSDE 8.3+ release expected this fall. The primary objective of this solution will support a registered geodatabase version extract to a personal geodatabase for disconnected editing purposes. The version adds/deletes will be collected by the disconnected editor, and on reconnecting to the network, can be uploaded to the central ArcSDE database as a version update.

This new technology may expand the ways ArcSDE will support remote site operations. We expect this technology may support more efficient synchronization of central ArcSDE database environments with remote DBMS servers and loosely connected editor environments. **Figure 4-32** presents the different new future ways to access spatial data using disconnected editing, available with ArcGIS 8.3+.

Figure 4-32
New Ways We Can Access Spatial Data
(Future Disconnected Editing – ArcGIS 8.3+)



4.9.12 Data Management Overview

Support for distributed database solutions has traditionally introduced high-risk operations, with potential for data corruption and use of stale data sources in supporting GIS operations. There are organizations that support successful distributed solutions. Their success is based on careful planning and detailed attention to their administrative processes that support the distributed data sites. The more successful solutions support central consolidated database solutions with effective remote user access and support. Whether centralized or distributed, the success of enterprise GIS solutions will depend heavily on the administrative team that keeps the system operational and provides an architecture solutions that supports user access needs.

5.0 GIS User Needs

System architecture design provides a hardware and network solution that supports the performance and communication needs of GIS application users. Distributed computer environments are commonplace throughout the world, providing applications that support users where they work throughout the organization.

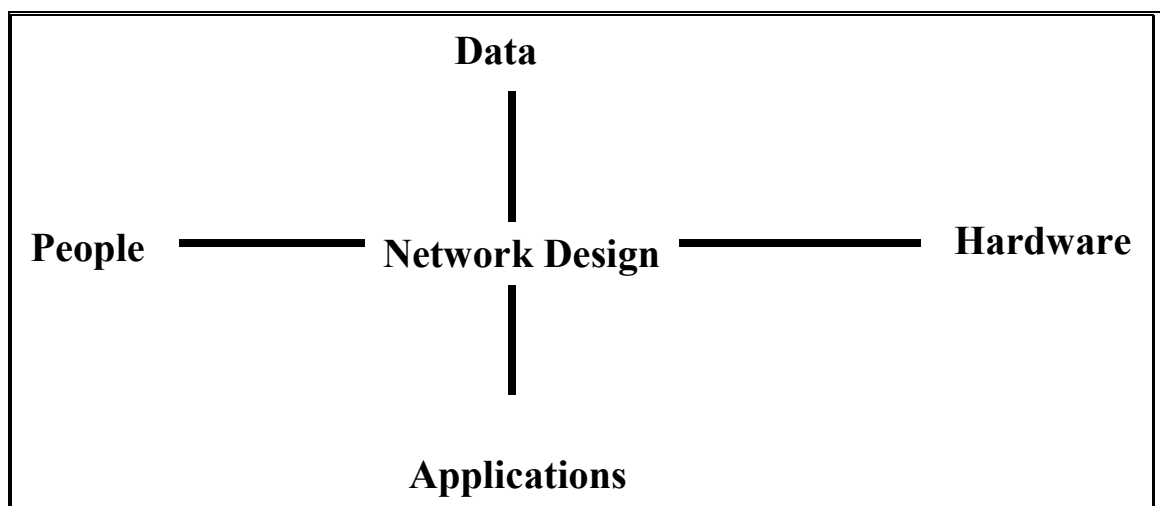
GIS applications provide new and powerful ways to evaluate information. Application tools allow users to visually analyze large volumes of data. Access, retrieval, and display of these data place significant demands on the associated distributed computer environment.

The GIS software supported by the distributed computer environment will establish primary requirements for the system design. Some user software requirements have a small impact on system design, and others require significant support from system resources. Understanding how GIS will be used in the workplace is important in order to develop a successful design solution.

5.1 System Design Prerequisites

Many factors contribute to a successful enterprise GIS solution. A GIS solution must have the right people, data resources, applications, and hardware to be successful. The type and nature of the GIS solution must match the needs and skills of the organization.

Figure 5-1
Total GIS Environment



Some basic objectives must be satisfied before an organization is ready to develop an enterprise system design strategy. These objectives include establishing a set of organizational goals for GIS development and identifying application needs and data requirements.

Approved organizational goals provide a foundation for getting resources necessary to support implementation. Implementation of an enterprise GIS solution will require top management support and must include a commitment from a number of departments throughout the organization. GIS goals should include objectives that define what changes the organization

would like to see over the next three- to five-year period. These implementation objectives establish baseline requirements for the system design review.

5.2 GIS Application Needs

ESRI marketing staff provide a wide variety of support to customers who need help identifying solutions for their GIS needs. The technical marketing staff understand what GIS applications will best support user needs. They concentrate their efforts on understanding their customer's and how ESRI applications can best support specific user requirements.

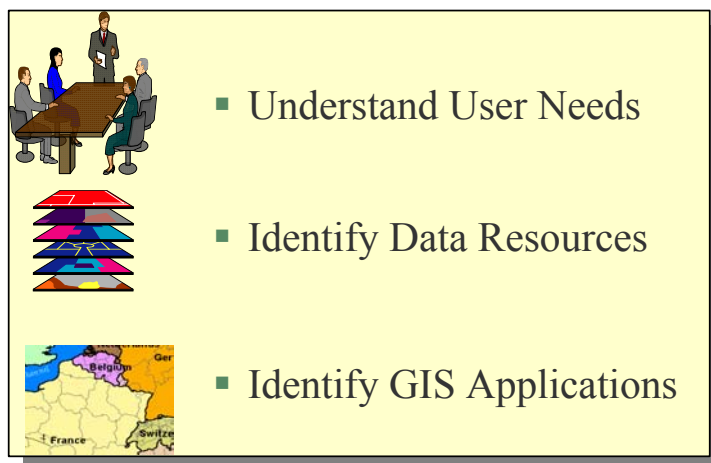
Larger enterprise solutions may require support from GIS application consultants. A variety of consultants are available to support users in defining their GIS software needs. ESRI has an experienced staff of consultants who support users in defining the appropriate software to support their particular operational requirements.

ESRI software provides applications that support users in a broad range of vertical market disciplines. ESRI has technical vertical marketing staff trained and focused on supporting GIS use in the work environment, and ESRI software solutions provide the tools required to support your operational needs.

GIS application needs and data requirements should be identified before evaluating hardware requirements. An application needs assessment can be completed by internal staff or by a professional GIS consultant, and it should be documented and approved by senior management.

Application needs assessments typically include interviewing department staff throughout the organization, identifying GIS application needs and data requirements, and evaluating the benefits for implementing GIS solutions. The assessment will typically identify requirements for supporting the applications and data development and provide a proposed implementation schedule.

Figure 5-2
Application Needs Assessment



Once the application needs are understood, a GIS system design consultant can work with the user community to identify the location and number of users who will need access to the application and establish the frequency with which they will be using the system. This is

generally accomplished by direct discussions between the system design consultant and selected user representatives.

5.3 System Architecture Design Review

People skills and experience in maintaining distributed computer system solutions are important considerations when selecting a system design. Maintenance of the distributed computer environment will be a critical consideration in selecting appropriate vendor solutions.

Experience and training in maintaining specific computer environments may identify a particular design solution as the best fit for your organization. **Figure 5-3** provides an overview of issues to address during a system design review.

- **Platform and Network Environments.** The design consultant should review the current vendor platforms and network environments. These are the environments that are currently maintained by the organization. Hardware experience, maintenance relationships, and staff training represent a considerable amount of investment for any organization. Proposed GIS design solutions should take advantage of corporate experience gained from working with the established platform and network environment.

Figure 5-3
System Design Review



- **Hardware Policies and Standards.** Organizations develop policies and standards that support their hardware investment decisions. Understanding management preferences and associated vendor relationships will provide insight into a design solution that can be supported best by the organization.

- **Operational Constraints and Priorities.** Understanding the type of operations supported by the GIS solution will identify requirements for fault tolerance, security, application performance, and the type of client/server architecture that would be appropriate to support these operations.
- **System Administration Experience.** The skills and experience of the systems support staff provide a foundation for supporting the final design solution. Understanding network administration and hardware support experience, in conjunction with support staff preference for future administration strategies, will help guide the consultant to compatible hardware vendor solutions.
- **Financial Considerations.** The final solution must be affordable. An organization will not implement a solution that is beyond their available financial resources. With system design, cost is a function of performance and reliability. If cost is an issue, the system design must facilitate a compromise between user application performance, system reliability, and cost. The design consultant must identify a hardware solution that provides optimum performance and reliability within identified budget constraints.

Current technology supports distribution of GIS solutions to clients throughout an enterprise environment, but there are limitations that apply to any distributed computer system design. It is important to clearly understand real GIS user needs and discuss alternative options for meeting those needs with your systems support staff in order to identify the most cost-effective solution.

5.4 GIS User Needs Assessment

There are a few basic user requirements that must be understood to support an effective system design solution. These requirements include identifying where the GIS users are located in relation to the associated data resources (site locations), what network communications are available to connect user sites with the GIS data sources, and what types of users are located at these locations. **Figure 5-4** provides an overview of the system architecture needs assessment.

Figure 5-4
System Architecture Needs Assessment



5.4.1 GIS User Locations

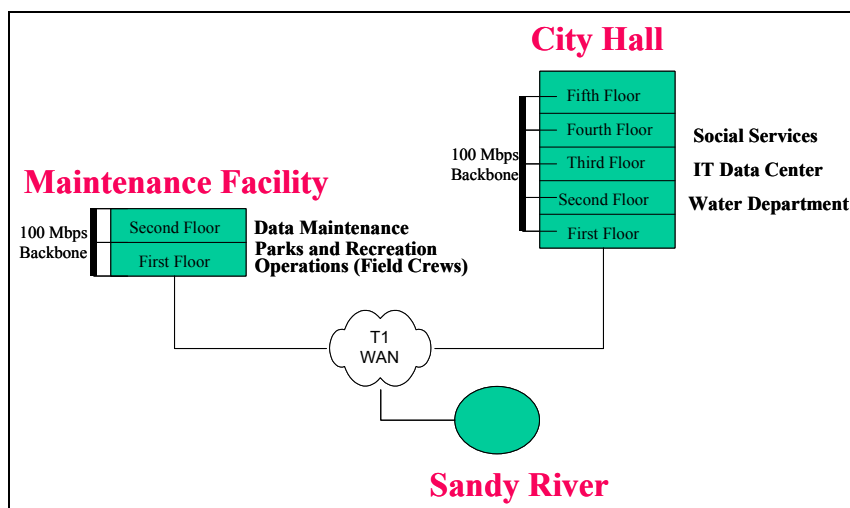
All locations within the organization that require access to GIS applications must be identified. The system design for each location will be developed to support the number of GIS users at that location.

5.4.2 Network Communications

The network communications between the different locations must be identified to support the system design assessment. Network bandwidth will identify communication constraints that must be supported in the system design solution.

Figure 5-5 provides a sample format for identifying locations throughout a typical organization. This drawing provides an overview of the facility locations and the associated network environment to be addressed in the system design study.

Figure 5-5
Network Communications Overview



5.4.3 GIS User Types

The types of GIS users can be divided into two basic categories. The ArcGIS desktop users will require desktop applications for GIS processing. The remaining users can be supported by ArcIMS map services.

- **ArcGIS Desktop.** Includes Desktop GIS specialists supporting general spatial query and analysis studies, simple map production, and general-purpose query and analysis operations, including all ArcEditor and ArcView clients. GIS applications for custom business solutions and other ArcView 3 GIS clients that support specific business needs should also be included in this category.

- **ArcIMS Map Services.** ESRI Internet Map Services provide transaction-based map products to intranet and Internet browser clients. ArcIMS supports transaction-based Web services for users requiring access to standard map products. ArcIMS can also be used as a data source for ArcGIS desktop clients.

Figure 5-6 provides an overview of a sample organization's user requirements. The total number of concurrent users will provide a basis for establishing server platform specifications and network bandwidth requirements. Each department manager should carefully review their GIS needs and identify the number of concurrent application users required to support their work environment. Peak concurrent use provides a foundation for software licensing requirements and estimating peak system infrastructure processing loads.

Figure 5-6
User Application Requirements Overview

User Requirements	December 1, 2002			December 1, 2003		
	ArcGIS		ArcIMS	ArcGIS		ArcIMS
	Total	Peak	Req/hr	Total	Peak	Req/hr
City Hall						
IT Data Center	2	2		2	2	
Social Services	5	2		20	10	
Water Department	10	2		40	25	
Subtotal City Hall	17	6		62	37	
Maintenance Facility						
Landbase Data Maintenance	40	35		70	55	
Parks and Recreation	12	3		33	12	
Operations (Field Crews)	10	5				
Subtotal Maintenance Fac	62	43		103	67	
Remote Locations						
Sandy River	10	5		10	10	
Subtotal Remote Buildings	10	5		10	10	
CITY TOTALS	89	54	2,000	175	114	4,000

5.5 System Configuration Alternatives

Several GIS configuration alternatives are available to support user processing requirements. The optimum configuration strategy can be identified once organizational and user requirements are understood.

5.5.1 Central Computing Architecture

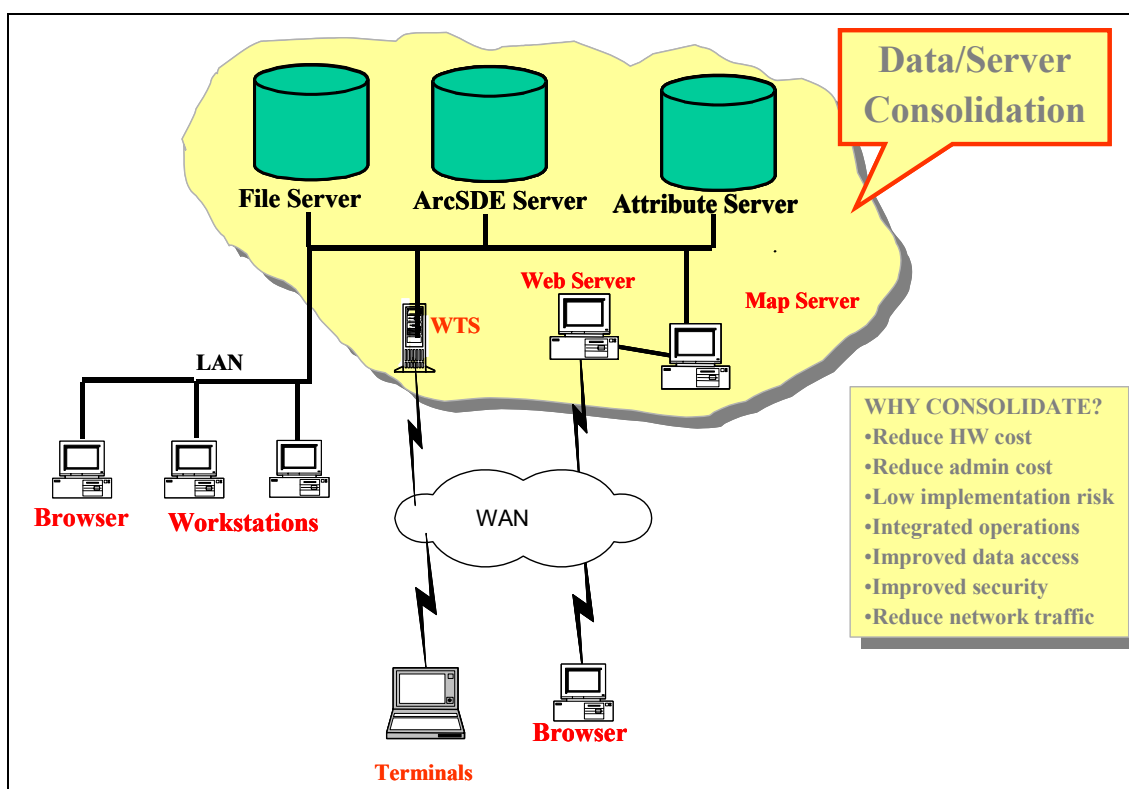
A central computing architecture can support ArcGIS desktop and ArcIMS map service clients throughout the enterprise network environment.

- ArcGIS desktop applications can be deployed in a distributed local area network (LAN) environment taking full advantage of standard network protocols. Applications are executed

in supported local workstation operating system environments. Direct access to distributed GIS data is provided using standard network disk-sharing protocols.

- ArcGIS desktop applications can be installed on central application server platforms, hosting multiple remote terminal-based clients. Applications running on the application server are displayed and controlled by users located at remote desktop display terminals. Applications and GIS data servers can be located in the same controlled central computer environment.
- ArcIMS Web services support thin client GIS users and provide a data source for ArcGIS desktop clients. This environment allows a single application process to provide simultaneous support to a large number of concurrent GIS users. The ESRI Internet Map Server ArcIMS components are installed with a standard Web server that directs inbound map requests to waiting map service agents located on the map server. Each request is serviced within seconds and returned to browser clients throughout the enterprise. **Figure 5-7** provides an overview of ESRI centralized architecture solutions.

Figure 5-7
Central Server with Workstation Clients



Server and data consolidation is an effective way to reduce overall GIS implementation and support cost. There are several advantages of supporting GIS operations from a central processing environment:

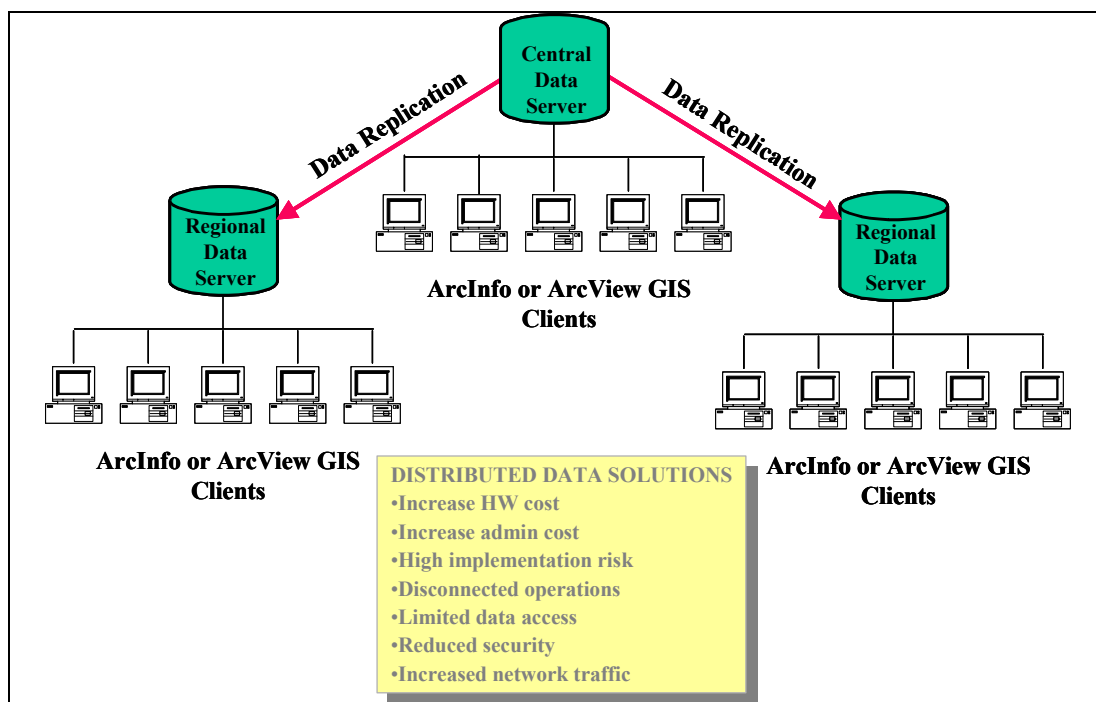
- **Reduce Hardware Cost.** Client workstation upgrades are much more expensive than supporting heavy applications on a central terminal server.

- **Reduce Administration Cost.** Applications and user workspaces are much easier to support from a centralized processing environment.
- **Low Implementation Risk.** Central solutions are much easier to deploy than distributed architectures. Once the system is installed and operational in the computer center, it can be deployed throughout the enterprise by simply upgrading the user group profiles.
- **Integrated Operations.** Data integration is much simpler to support with all data sources located in the computer center on the same network as the client applications.
- **Improved Data Access.** High bandwidth within the computer center reduces network contention and improves application performance.
- **Improved Security.** Applications and data access are restricted to the central computer processing environment.
- **Reduced Network Traffic.** Heavy traffic is isolated on the computer center network, enterprise traffic is limited to terminal client display information and plotting traffic.

5.5.2 Distributed Data and Workstation Processing (WAN)

This solution provides GIS data resources to each site location. It also includes applications running on the client desktop, with network access to distributed data sources at each location. This is a typical solution when users are located at a reasonable number of distributed sites and WAN communications are not reliable enough to support centralized operational requirements.

Figure 5-8
Distributed Data and Workstation Processing (WAN)



A central data server maintains a composite copy of the GIS data libraries. Required data is replicated from the central data server to regional data servers to support data currency requirements. Data updates from the regions are replicated back to the central data server for final processing.

Distributing data requires additional servers and disk storage. Distributed solutions require multiple copies of the data. As the size of data grows, the cost for supporting distributed solutions will increase. Distributed solutions continue to present both implementation risk and follow-on maintenance challenges, and clearly result in higher cost environments. A clear justification should be validated before recommending a distributed architecture environment.

5.6 Choosing a System Configuration

The best solution for a given organization will depend on the distribution of the user community and the type of operational data in use. User requirements determine the number of machines necessary to support the operational environment, the amount of memory required to support the applications, and the amount of disk needed to support the system solution.

Figure 5-9
Choosing a System Configuration

- **Best Solution Leverages Current IT Investments**
- **User Needs Set Foundation for Hardware Selection**
- **Platform Selection Based on User Needs**
 - Types of user applications
 - User locations and network communications
 - Location of GIS data resources
 - Peak user loads on system
 - Peak processing loads on selected hardware components
- **Contributing Factors to Final Design**
 - Existing hardware and support experience
 - Organizational system design policies/preferences
 - Operational constraints and priorities
 - Future growth plans/available budget

A basic ingredient to understanding and applying results of the user needs assessment is identification of the type of users on the system and the workstation performance needed to support these functions. Information required includes number of users on the system, the percent of time each will be using their GIS application, size of the user directories (workspace), size and type of other applications on the system, and user performance requirements. In addition, where data files will be located on the system, how users will access these data, and how much disk space will be needed to store the data are considerations. Also, it is important to understand the facility layout and available network communications and to evaluate the environment for potential performance bottlenecks. Other factors include accounting for existing

equipment, organizational policies, preference toward available system design strategies, future growth plans, and available budget.

5.7 Hardware Component Selection

Once a configuration strategy is identified, the specific server platforms required to support the solution can be identified.

Figure 5-10
Hardware Component Selection
Centralized Solution

User Requirements	December 1, 2002			December 1, 2003		
	ArcGIS		ArcIMS	ArcGIS		ArcIMS
	Total	Peak	Req/hr	Total	Peak	Req/hr
City Hall						
IT Data Center	2	2		2	2	
Social Services	5	2		20	10	
Water Department	10	2		40	25	
Subtotal City Hall	17	6		62	37	
Maintenance Facility						
Landbase Data Maintenance	40	35		70	55	
Parks and Recreation	12	3		33	12	
Operations (Field Crews)	10	5				
Subtotal Maintenance Fac	62	43		103	67	
Remote Locations						
Sandy River	10	5		10	10	
Subtotal Remote Buildings	10	5		10	10	
CITY TOTALS	89	54	2,000	175	114	4,000
Server Platforms	December 1, 2002			December 1, 2003		
	ArcGIS		ArcIMS	ArcGIS		ArcIMS
	Total	Peak	Load	Total	Peak	Load
City Hall Data Center						
ArcSDE GeoDatabase		54			114	
Windows Terminal Servers		48			77	
ArcIMS Server	2,000		Requests per hour	4,000		Requests per hour

- Centralized Solution.** The first step is to identify required platform components to support user requirements and comply with infrastructure constraints. In this centralized solution, we selected an ArcSDE geodatabase server, a Windows Terminal Server, and an ArcIMS server to be located at the City Hall IT Data Center. The central ArcSDE database will support all shared data resources, the Windows Terminal Server will provide remote user access to applications located at the central computer facility (this will comply with WAN bandwidth restrictions), and the ArcIMS server will support Web mapping services from the central database.

The next step is to translate peak user loads identified at the site and department levels to the associated platforms that support their application and data processing. The result of

this analysis identifies peak user loads for each of the selected hardware components. **Section 7** will take these user loads and generate performance specifications for each of the selected hardware components.

Figure 5-11
Hardware Component Selection
Distributed Solution

User Requirements	December 1, 2002			December 1, 2003		
	ArcGIS		ArcIMS	ArcGIS		ArcIMS
	Total	Peak	Req/hr	Total	Peak	Req/hr
City Hall						
IT Data Center	2	2		2	2	
Social Services	5	2		20	10	
Water Department	10	2		40	25	
Subtotal City Hall	17	6		62	37	
Maintenance Facility						
Landbase Data Maintenance	40	35		70	55	
Parks and Recreation	12	3		33	12	
Operations (Field Crews)	10	5				
Subtotal Maintenance Fac	62	43		103	67	
Remote Locations						
Sandy River	10	5		10	10	
Subtotal Remote Buildings	10	5		10	10	
CITY TOTALS	89	54	2,000	175	114	4,000

Server Platforms	December 1, 2002			December 1, 2003		
	ArcGIS		ArcIMS	ArcGIS		ArcIMS
	Total	Peak	Load	Total	Peak	Load
City Hall Data Center						
ArcSDE GeoDatabase		11			47	
Windows Terminal Servers		5			10	
ArcIMS Server	2,000		Requests per hour	4,000		Requests per hour
Server Platforms						
ArcSDE GeoDatabase		43			67	

- Distributed Solution.** The first step is to identify required platform components to support user requirements and comply with infrastructure constraints. In this example, the GIS maintenance users were all located at one of the remote locations and insisted they have a local data server. A data server was also required at the central City Hall Computer Center to support other users throughout the City. Data updates will be processed at the Maintenance Facility database with weekly updates replicated to the central City Hall database to support the remaining City users.

In this alternative distributed solution, we selected an ArcSDE geodatabase server, a Windows Terminal Server, and an ArcIMS server to be located at the City Hall Data Center. We identified a separate ArcSDE geodatabase server that would be located at the

Maintenance Facility. The central ArcSDE database will support all shared data resources except those located at the Maintenance Facility, the Windows Terminal Server will support remote user access for the Sandy River location to applications located at the Central Computer facility (this will comply with WAN bandwidth restrictions), and the central ArcIMS server will support Web mapping services from the central City Hall database.

The next step is to translate peak user loads identified at the site and department levels to the associated platforms that support their application and data processing. All users at the Maintenance Facility will use the Maintenance Facility ArcSDE database. The result of this analysis provides peak user loads on the components we selected to support the enterprise GIS operations. **Section 7** will take these user loads and generate performance specifications for each of the selected hardware components.

Sizing models introduced in **Sections 6 and 7** will be used to identify candidate hardware platform configurations that support the peak user loads identified above.

6.0 Sizing Fundamentals

Computer platforms must be configured properly to support system performance requirements. There are many factors that contribute to overall system performance. Enterprise GIS solutions include distributed processing environments, where user performance can be the product of contributions from several hardware platform environments. Many of these platform resources are shared with other users.

ESRI has implemented distributed GIS solutions since the late 1980s. System performance models have been used by ESRI consultants to support distributed computing hardware solutions since 1992. These same performance models have been used to identify potential performance problems with existing computing environments. This section will present the basis for these models. A fundamental understanding of these models will help users better understand their computing environment.

The models in this section will be presented with the assumption that all platforms have the same processing capacity. This is a simplified approach to presenting and understanding the performance models. The following section (Sizing Tools) will identify how these models are applied in the real world, where platform performance is not the same and is rapidly changing.

6.1 System Performance Profile

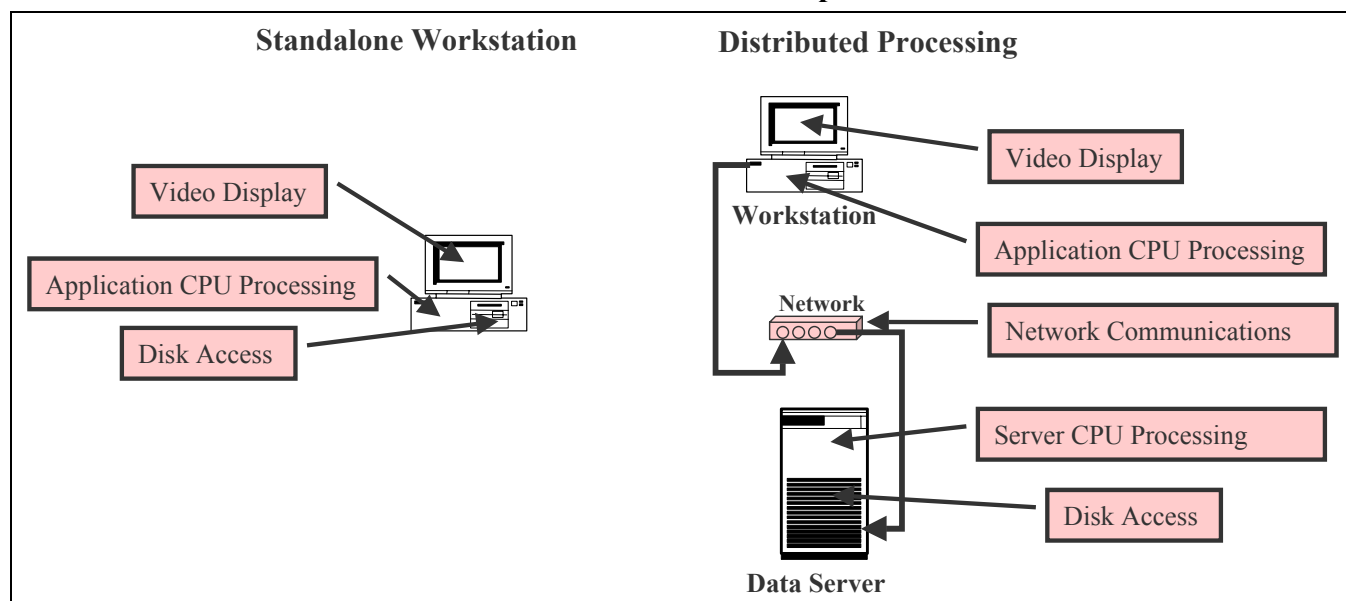
Computer platforms are supported by several component technologies. Each component technology contributes to the overall computer performance. Hardware vendors build computers with the appropriate component resources to optimize overall platform performance.

In much the same way, distributed computing solutions (enterprise computing environments) are supported by several hardware platforms that contribute to overall system performance. Each hardware component contributes to the overall system performance. Hardware platforms supporting a computing environment must be carefully selected to support optimum overall system performance.

The primary objective of the system design process is to provide the highest level of user performance for the available system hardware investment. Each hardware component must be selected with sufficient performance to support processing needs. Current technology can limit system design alternatives. Understanding distributed processing loads at each hardware component level provides a foundation for establishing an optimum system solution.

Figure 6-1 provides a simplified overview of the components in a standalone workstation and a distributed processing configuration. Each component participates sequentially in the overall program execution.

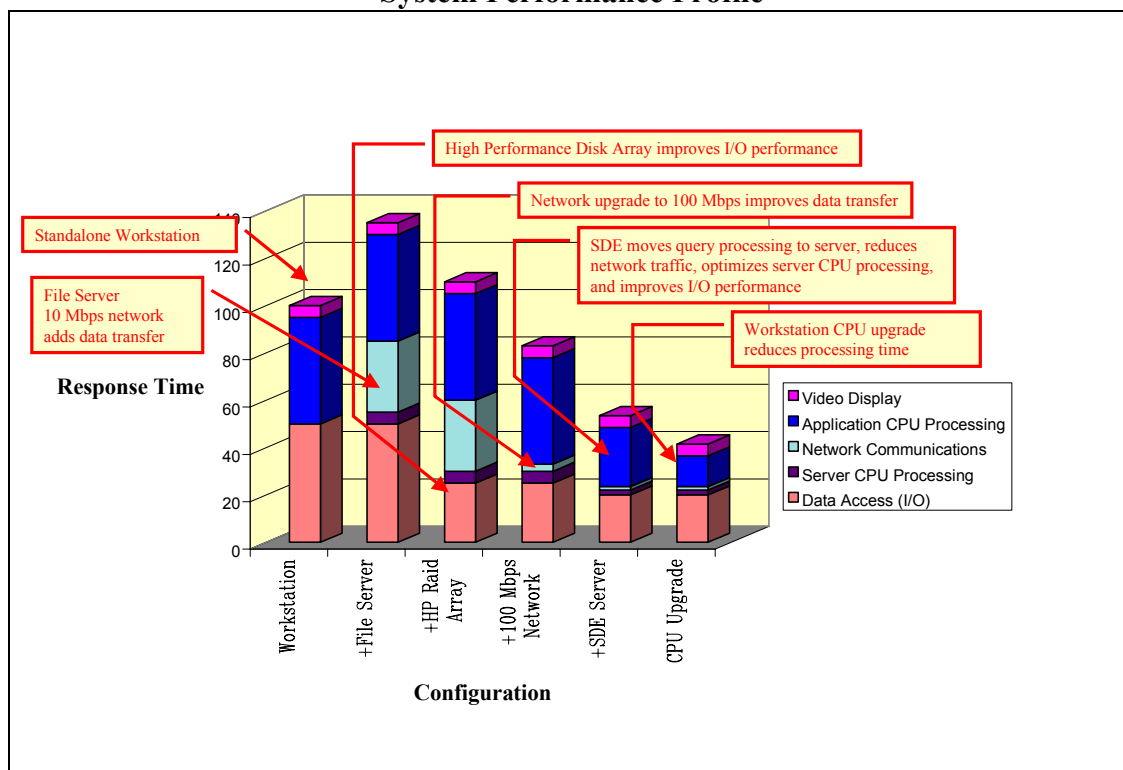
Figure 6-1
Platform Performance Components



The total response time of a particular application query will be a collection of the responses from each of these components. A computer vendor optimizes the component configuration within the workstation to support the fastest computer response to an application query. An IT/Systems department has the responsibility to optimize the organization's hardware and network component investments to provide the optimum system-level response at the user desktop. System performance can directly contribute to user productivity.

Figure 6-2 identifies performance gains as a result of a series of hardware investments. These investments contribute to the relative performance experienced at the user desktop.

Figure 6-2
System Performance Profile



The first column represents the performance of a standalone workstation performing a typical GIS operation (requesting display of a new map extent on the user screen). Experience shows that GIS applications tend to be both compute intensive and input/output (I/O) intensive. A standalone workstation spends roughly the same amount of time on data access and compute processing. A relatively small amount of remaining time is required to send the resulting map product to the video display.

The second column shows the system performance profile when accessing the data from a file server in place of local disk. This distributed solution includes additional server CPU processing and network data transfer. These additional system components extend the overall response time to include the additional system loads. Accessing data from a file server over a 10-Mbps network can reduce performance by roughly 30 percent.

The third column shows the result of upgrading the JBOD (just a bunch of disks) configuration on the file server to a high-performance RAID storage solution. High-performance RAID storage solutions can improve disk access performance by as much as 50 percent.

The fourth column shows the effect of increasing network bandwidth from 10-Mbps Ethernet to 100-Mbps Ethernet, a factor of 10 reduction in data transfer time. Users experience improved performance over having data on local workstation disk with this configuration.

The fifth column shows the result of moving the spatial data to an ArcSDE server. The ArcSDE server solution will improve performance in several areas. The ArcSDE server technology relocates query processing traditionally supported by the client application to the server platform. This reduces client CPU processing requirements by roughly 50 percent. Spatial data is compressed by 30 to 70 percent on the ArcSDE server, reducing network traffic by an additional 50 percent. The ArcSDE server also filters the requested data layer so only the requested map layer extent is sent over the network to the client, further reducing network traffic. The query processing performed by the ArcSDE server, using DBMS query indexing, data cache and search functions, will also reduce the processing load on the server to less than half that used to support the traditional client query processes. Moving spatial data to ArcSDE can significantly improve overall system performance in a distributed computing solution.

The final column shows the effect of upgrading the workstation CPU to one with twice the performance, reducing the workstation CPU processing time by 50 percent.

Hardware component investments contribute directly to user productivity, and the overall productivity of the organization. Computer technology is changing very rapidly, and the product of this change is higher performance and improved productivity at the user desktop. Organizations need to budget for this change, and make wise investments in their infrastructure portfolio to maintain high productivity in the workplace. A smart investment strategy pays very large dividends in supporting GIS operations.

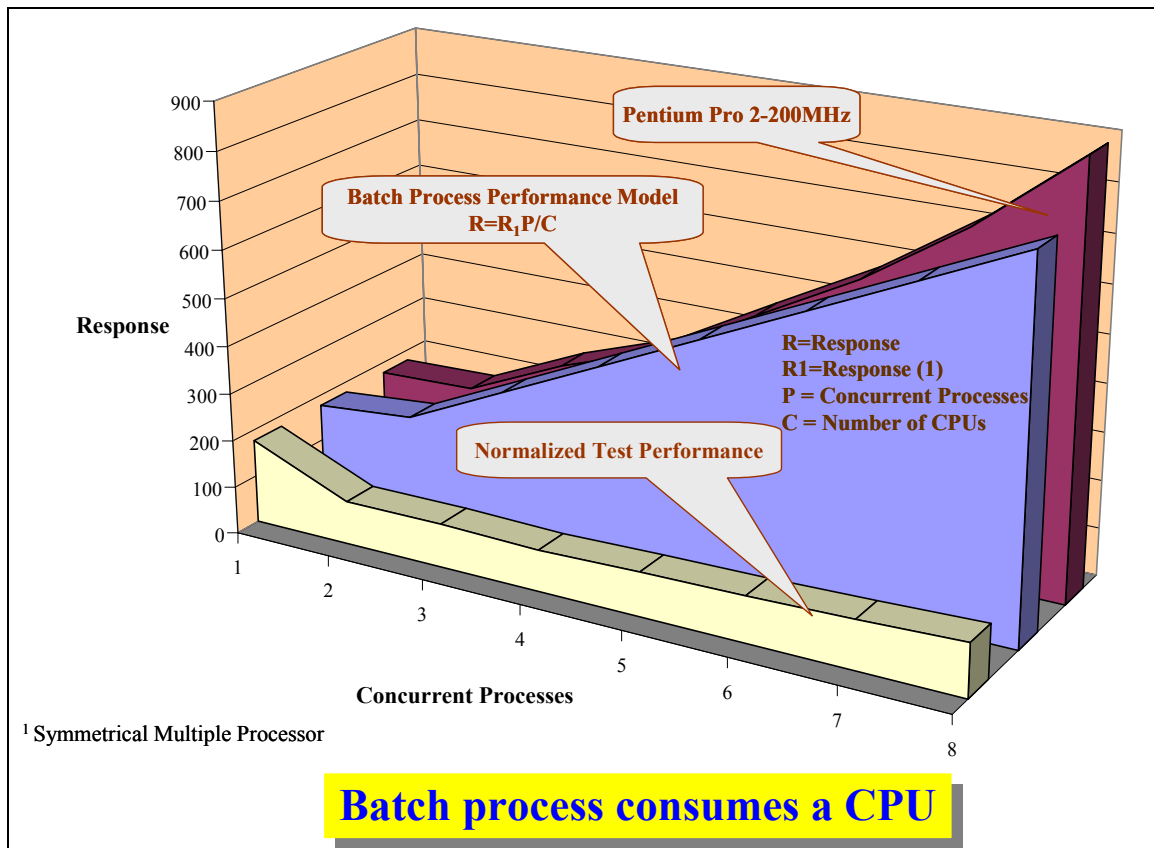
6.2 System Performance Testing

ArcInfo performance testing with the Microsoft Windows Terminal Server 4.0 Edition was conducted at the Data General development labs in Westborough, Massachusetts, in July 1998. This was an update of ArcInfo performance testing completed by ESRI in 1993. This performance testing provides a foundation for the sizing models used for configuring distributed GIS computing environments. The objectives of these models are to support proper selection and configuration of system hardware components for distributed GIS computing.

The ESRI system performance models are based on an understanding of how computer platforms respond to an increasing number of concurrent ArcInfo process loads. An ArcInfo performance benchmark is used to evaluate platform response to increasing user loads. Excessive memory was configured on each server platform (2-4 GB) to avoid paging and swapping of executables during the testing (recommended physical memory requirements are established by separate tests that measure memory allocated to support each application process).

Figure 6-3 provides a summary of the ArcInfo benchmark results from testing a dual-processor Pentium Pro 200 Windows Terminal Server.

Figure 6-3
Dual-Processor SMP Performance



An individual test was completed for each concurrent process configuration (1 through 8). The third-row graphic plots the average response time measured for each of the concurrent process test runs. The first-row graphic shows the rate at which the server platform was processing the ArcInfo instructions. The center row shows a plot of the sizing model for concurrent ArcInfo batch processing. The results of this test validate the ArcInfo design model, and demonstrate good Windows Terminal Server scaling performance with this platform configuration.

Figure 6-4 provides a summary of the ArcInfo benchmark results from testing a quad-processor Pentium Pro 200 Windows Terminal Server. An individual test was completed for each concurrent process configuration (1 through 16). The results of this test validate the ArcInfo design model, and demonstrate good Windows Terminal Server scaling performance with this platform configuration.

Figure 6-4
Quad-Processor SMP Performance

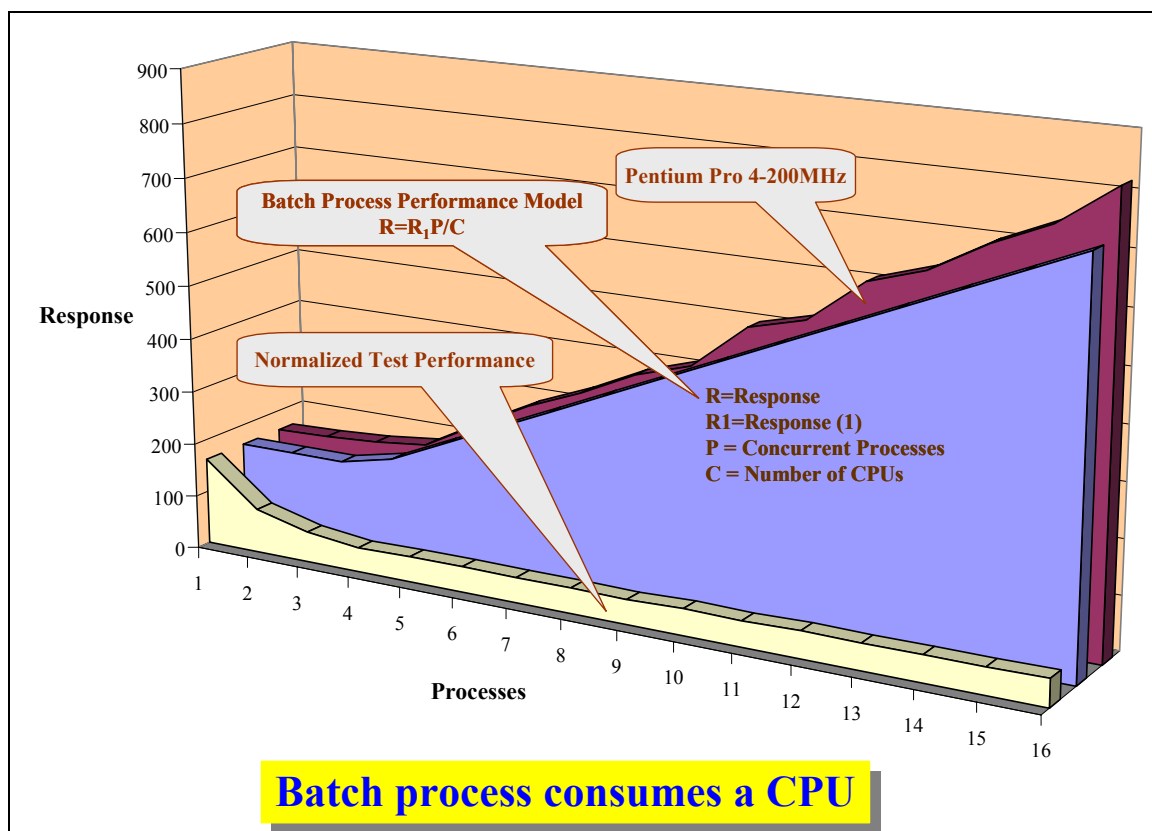
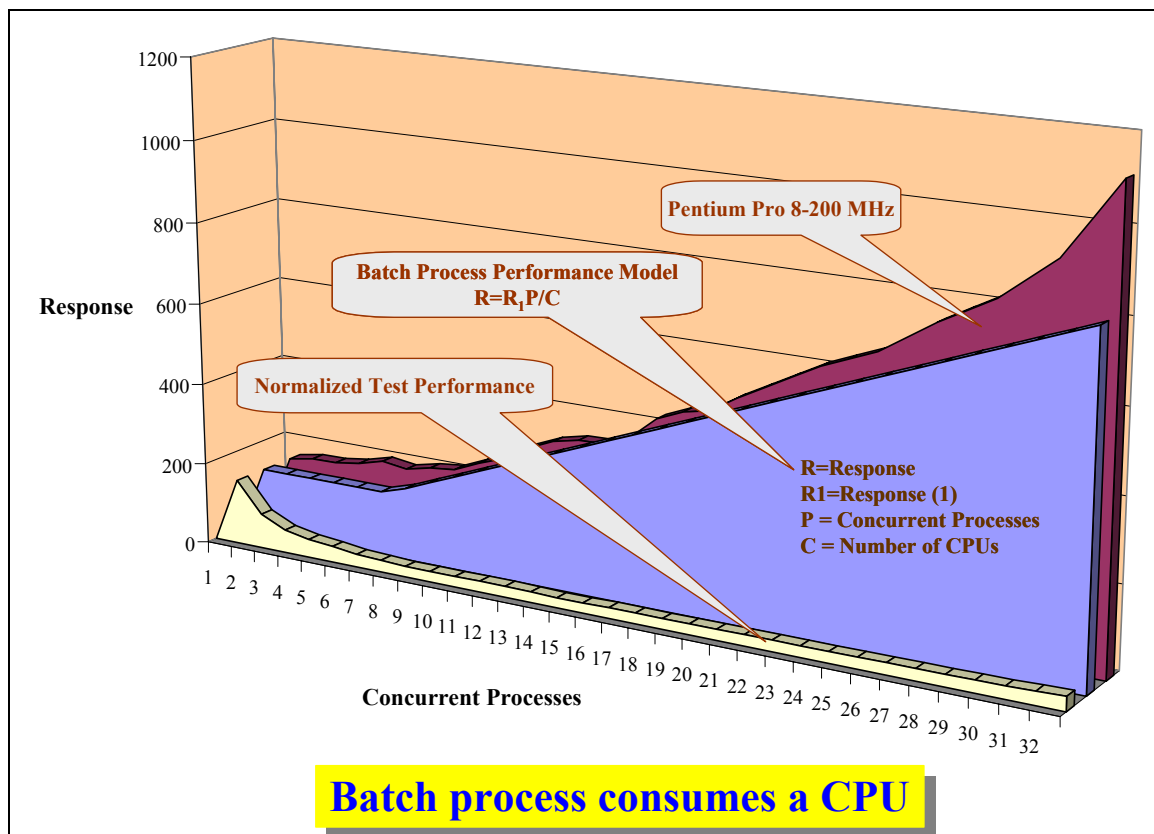


Figure 6-5 provides a summary of the ArcInfo benchmark results from an eight-processor Pentium Pro 200 Windows Terminal Server. An individual test was completed for each concurrent process configuration (1 through 32). The results of this test validate the ArcInfo design model, and demonstrate good Windows Terminal Server scaling performance with this platform configuration.

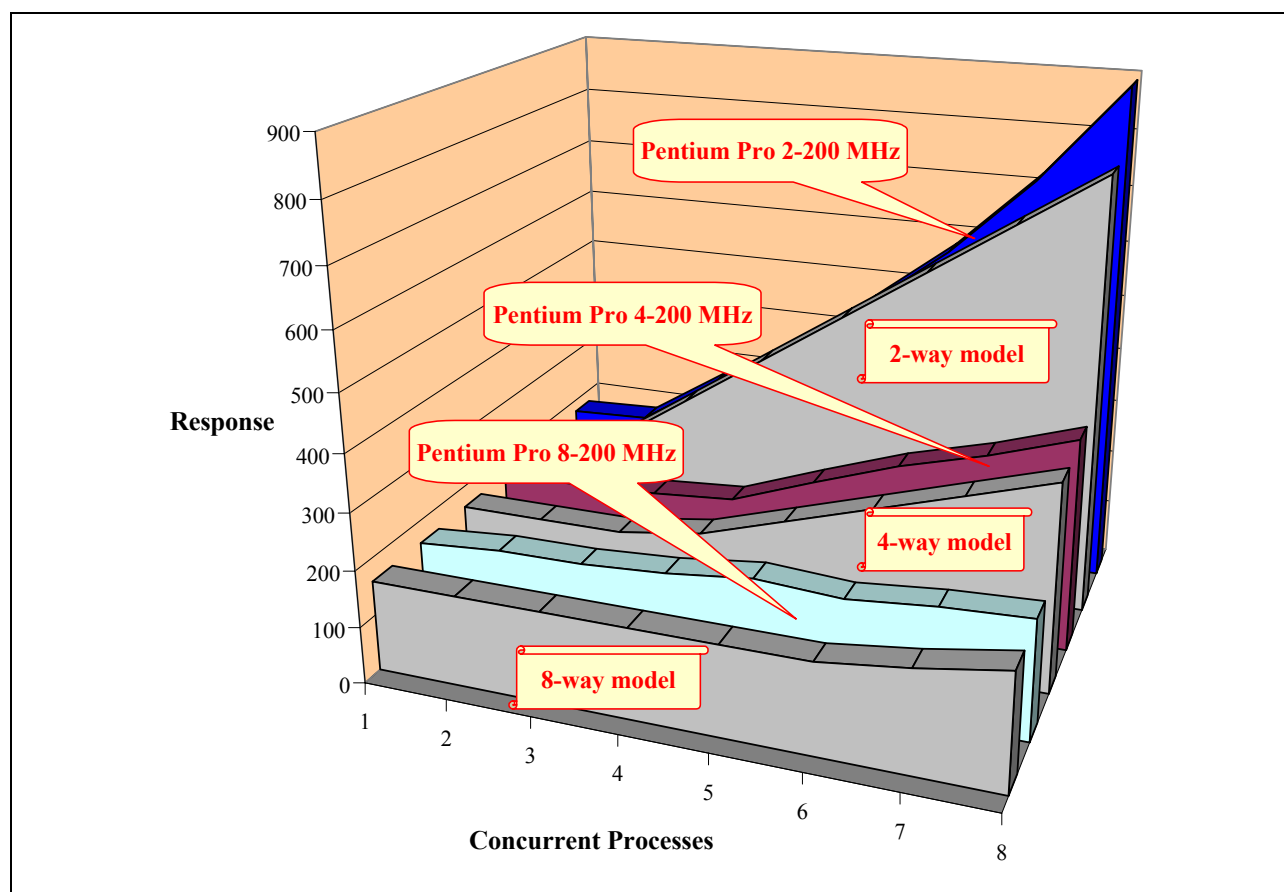
Figure 6-5
Eight-Way SMP Performance



Each benchmark test series was extended to evaluate platform support for four batch processes per CPU. The ArcInfo sizing model is seldom used to identify performance expectations beyond two concurrent batch processes per CPU. The sizing models perform very well at the lower range of these test series.

Figure 6-6 provides an overview of the Windows Terminal Server batch process testing. This same test series was conducted on a dual-processor UNIX platform (Sun Ultra 60). The performance scaling results (shape of the performance curve) for the UNIX application server testing was very similar to what we measured in the Windows server tests, suggesting performance scaling for Windows platforms through eight processors is similar to what we see with UNIX platforms.

Figure 6-6
Batch Processing Test Summary



6.3 Client/Server Models

Additional ArcInfo benchmark testing was conducted to evaluate processing performance with data located on a separate file server and on an ArcSDE server. The results of these tests provide a foundation for understanding distributed computing loads and developing our client/server computing performance-sizing models.

6.3.1 Batch Processing Performance

Figure 6-7 provides an overview of batch processing performance for symmetrical multiple processor (SMP) application server platforms (1- through 4-CPU configurations) accessing data located on a separate file server. Query processing is supported by the application server with data located on local disk or on a remote file server. Each batch process consumed the resources of a single CPU. The SMP operating system effectively distributes available processing loads across all CPUs. Performance remains the same until the number of concurrent batch processes exceeds the available number of CPUs. Performance slows at a relatively constant rate when the number of concurrent processes increases beyond the available CPUs.

Figure 6-7
Terminal Servers with File Server or Local Data

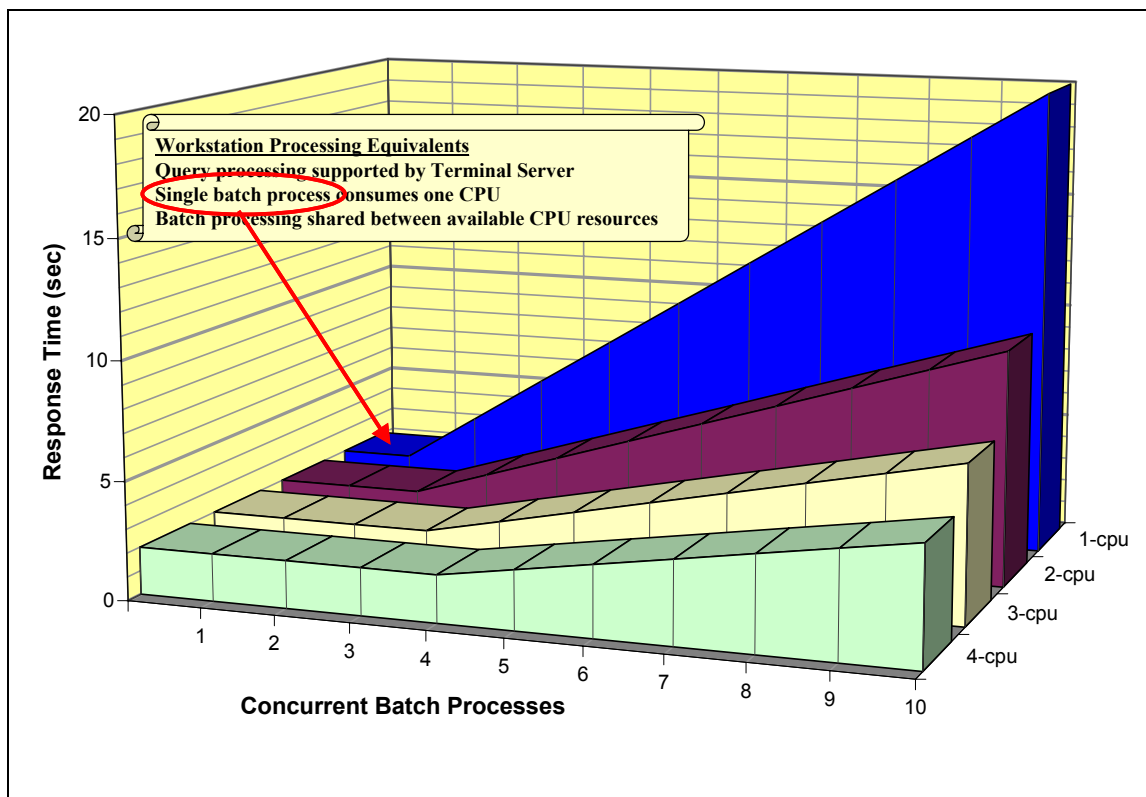
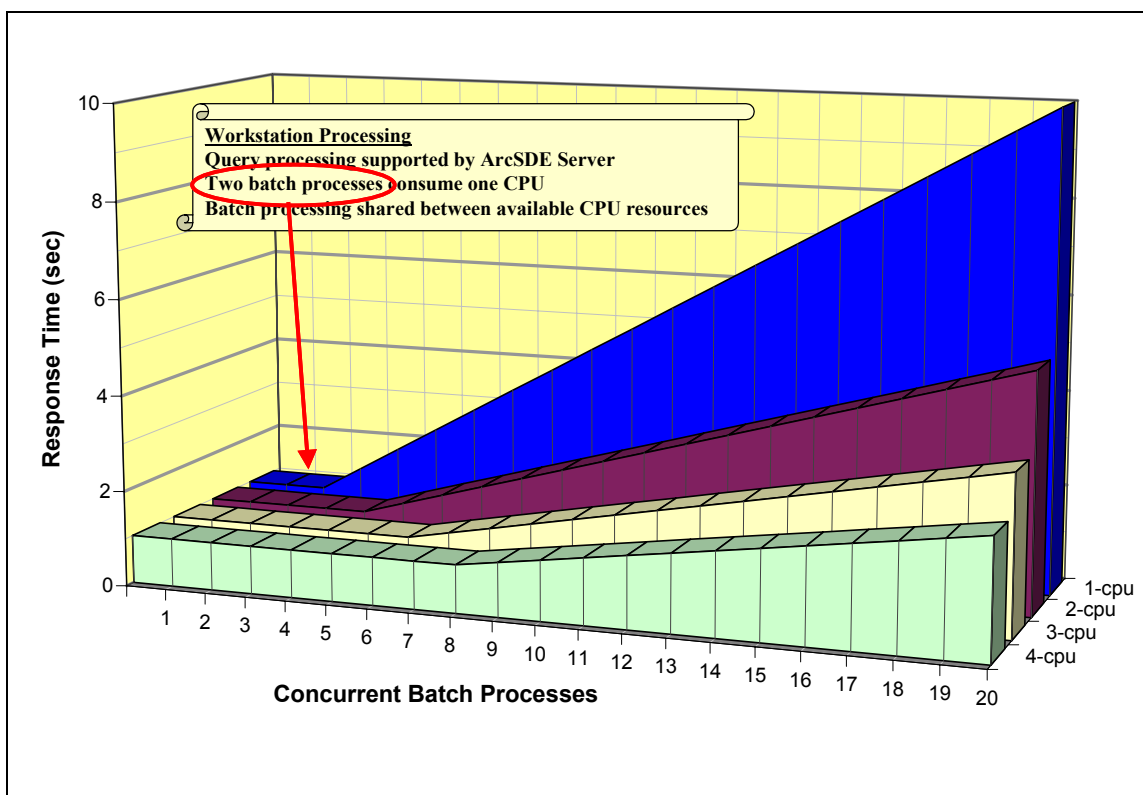


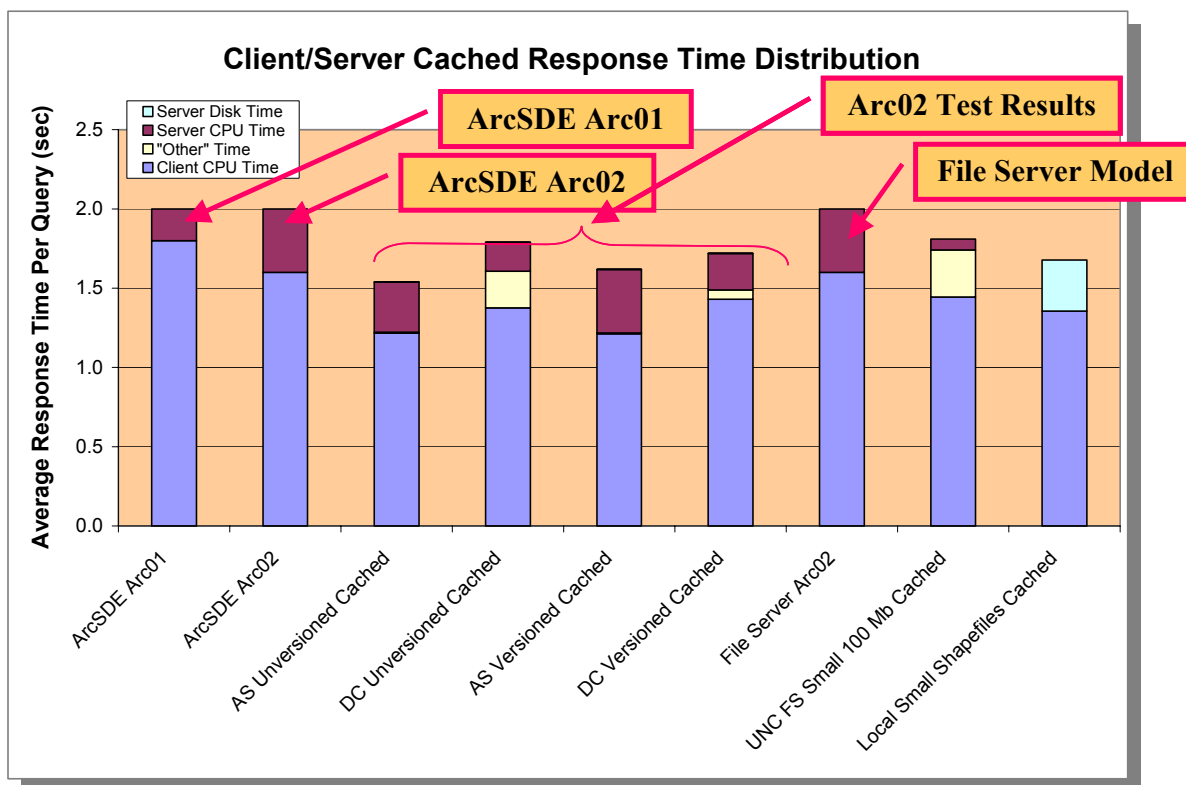
Figure 6-8 provides an overview of batch processing performance for SMP application server platforms (1- through 4-CPU configurations) accessing data located on an ArcSDE server. Query processing is supported by the ArcSDE server. Two batch processes are supported by a single CPU. The SMP operating system effectively distributes available processing loads across all CPUs. Performance remains the same until the number of concurrent batch processes exceeds twice the available number of CPUs. Performance slows at a relatively constant rate when number of concurrent processes increase beyond twice the available CPUs.

Figure 6-8
Terminal Servers with ArcSDE Data Source



ESRI ArcGIS 8.2 performance validation testing results are shown in **Figure 6-9**. The test results show roughly 20 percent of the distributed processing load (query processing) supported by the ArcSDE server, and roughly 80 percent of the processing load (map rendering) supported by the ArcGIS desktop client. The 2002 (Arc02) ArcSDE performance sizing model is adjusted to reflect these loads. The file server model has not changed, and will continue to be based on 20 percent of the total processing load.

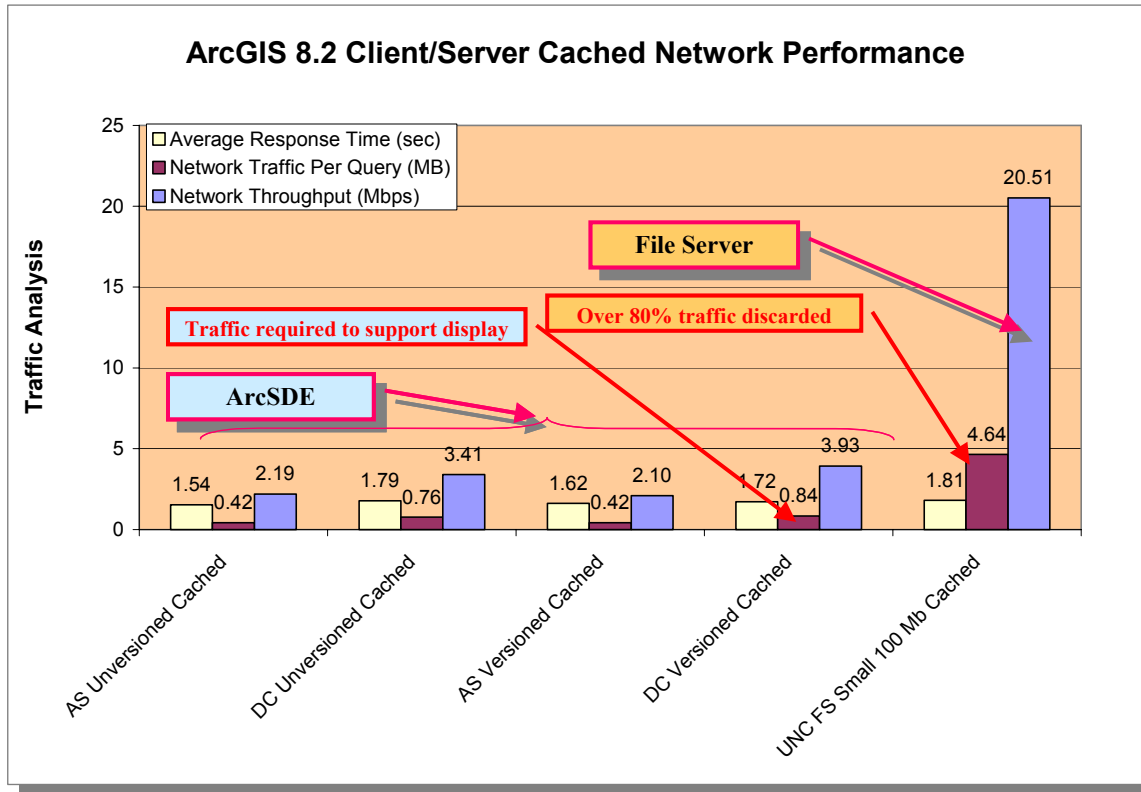
Figure 6-9
Data Server Loads



The foundation for the file server model was based on several years of experience configuring distributed ArcInfo enterprise environments. File servers tend to show less CPU processing intensity, although quickly start to exhibit performance problems if these sizing guidelines are not followed.

Figure 6-10 provides a summary of ArcSDE and file server test results, showing average response time, network traffic per query, and average network throughput measured during ESRI ArcGIS 8.2 performance benchmark testing. This chart highlights the performance advantages of an ArcSDE data source. Over 80 percent of the data transferred from the file server to the client workstation are discarded, and are not required to support the map display.

Figure 6-10
ArcGIS Client/Server Network Traffic Analysis



The baseline Arc02 performance test dataset used for the previous ArcSDE/file server performance comparison is a subset of a larger San Diego dataset. **Figure 6-11** compares client/server performance sensitivity to larger database environments, by running the exact same benchmark test using the complete San Diego database as a file data source. The file data source tests showed a significant increase in network throughput. This increase is expected, since the larger shape files must be transferred to the client workstation to support the query processing. The ArcSDE test results were the same as the previous chart, and are not effected by the size of the database.

Figure 6-11
ArcGIS File Server Network Throughput

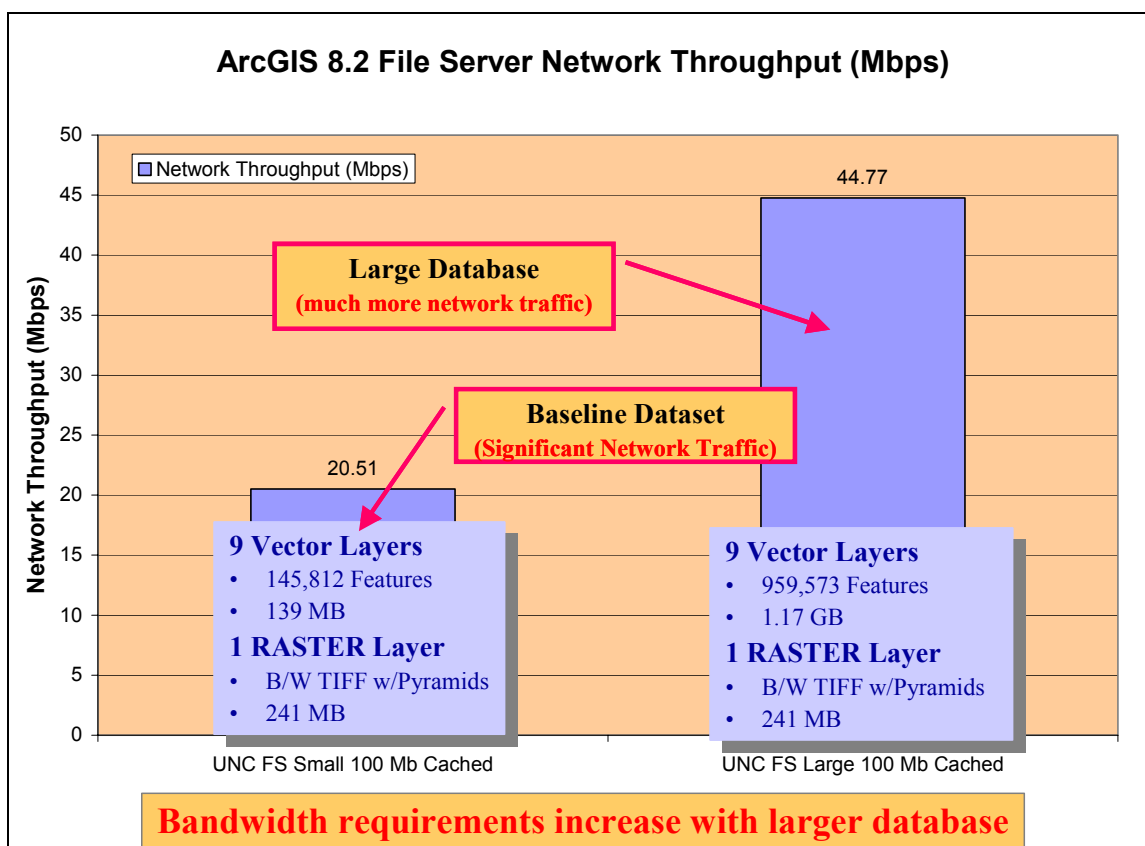


Figure 6-12 provides a summary of the average response time for the same benchmark test when using a file data source. Response time with the file data source is much slower due to the larger shapefiles. In comparison, very little difference in performance was measured with the ArcSDE data source testing.

Figure 6-12
ArcGIS File Server Network Performance

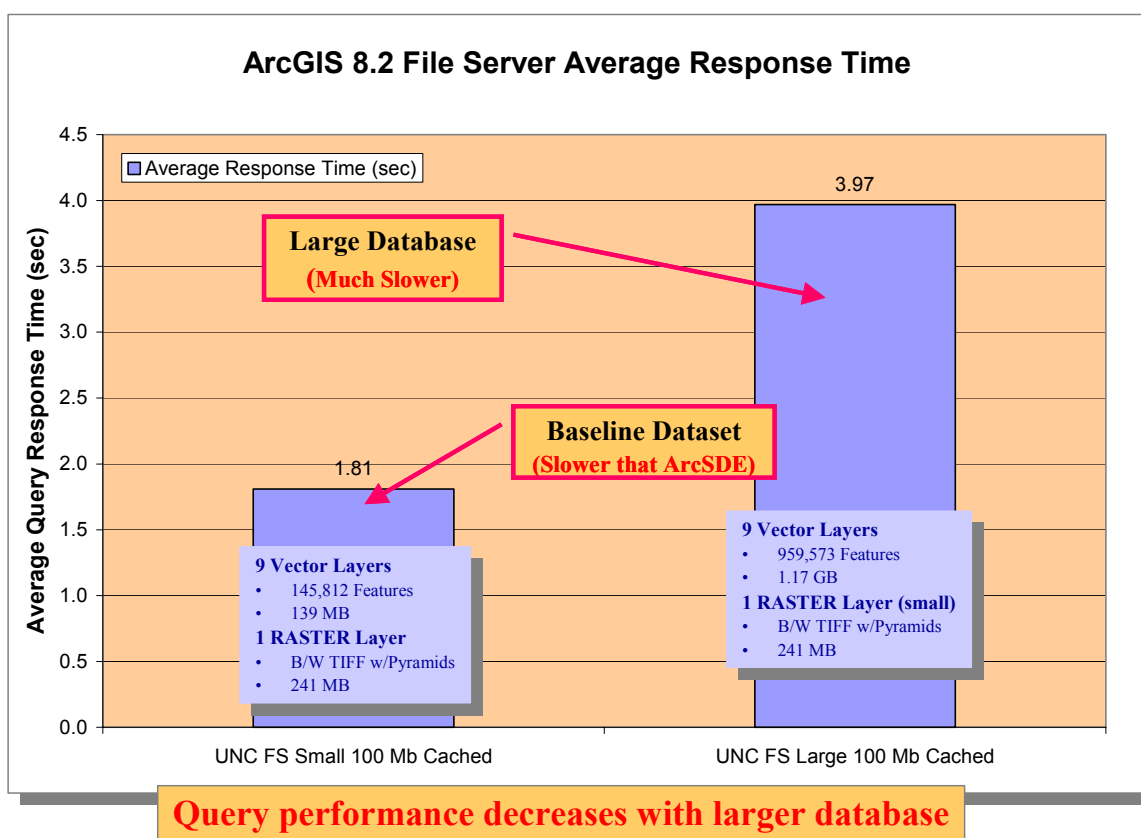
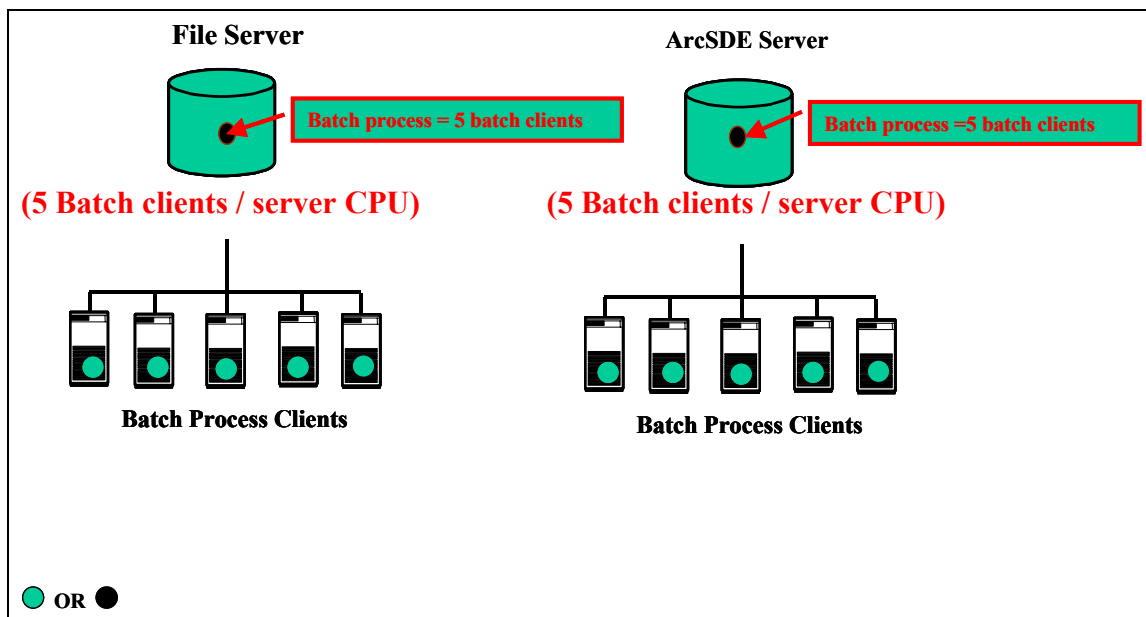


Figure 6-13 provides an overview of the data server sizing requirements for concurrent batch processing. Each diagram shows a balanced distributed platform configuration, with each of the server and workstation platforms having one CPU (all CPUs are the same). A single-CPU data server can support five batch client workstations (same model for file and ArcSDE data servers).

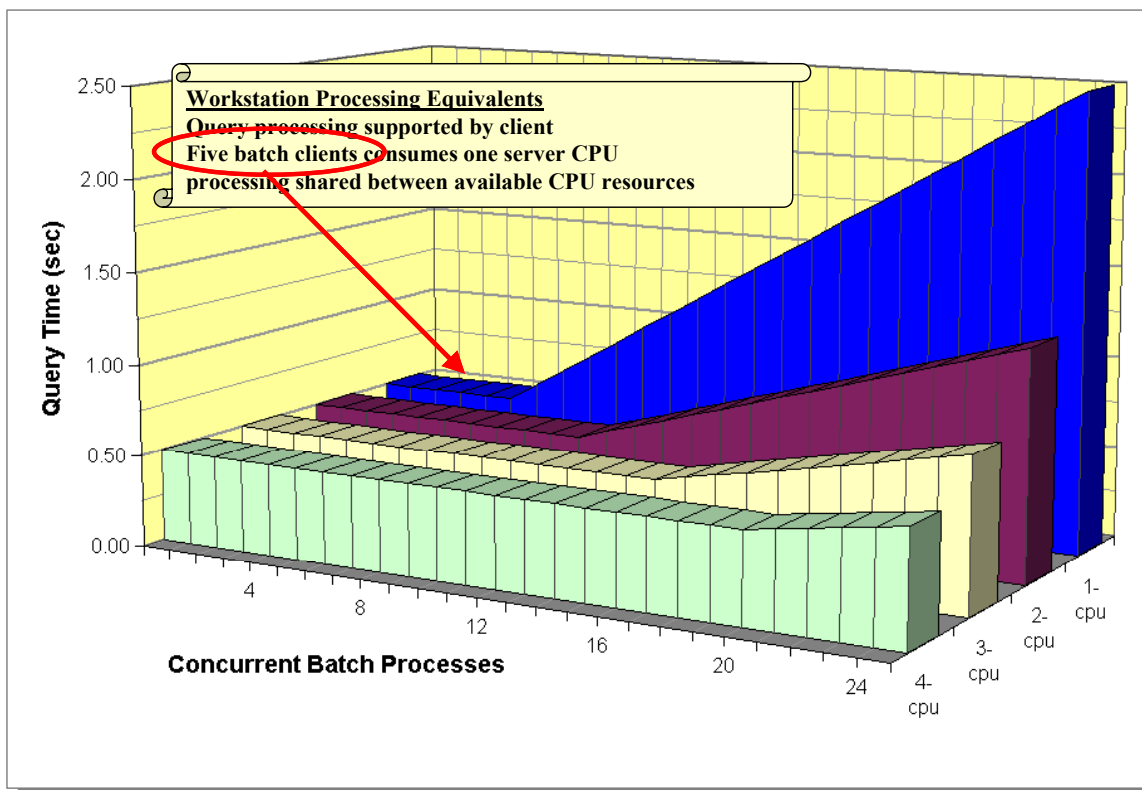
As a corollary, a batch process executed on the data server would be equivalent to five batch clients (either condition consumes a server CPU).

Figure 6-13
Data Server Performance Models



The GIS data server performance model is provided in **Figure 6-14**. The model shows a single data server CPU supports up to five (5) concurrent batch clients for each server CPU. Additional clients supported by the server would result in reduced query performance. This performance profile clearly applies to compute intensive ArcSDE server environments, where a variety of testing and real-life implementations has demonstrated the CPU-bound nature of the ArcSDE server processing.

Figure 6-14
GIS Data Server Performance Model



6.3.2 ArcGIS Desktop Terminal Server Performance

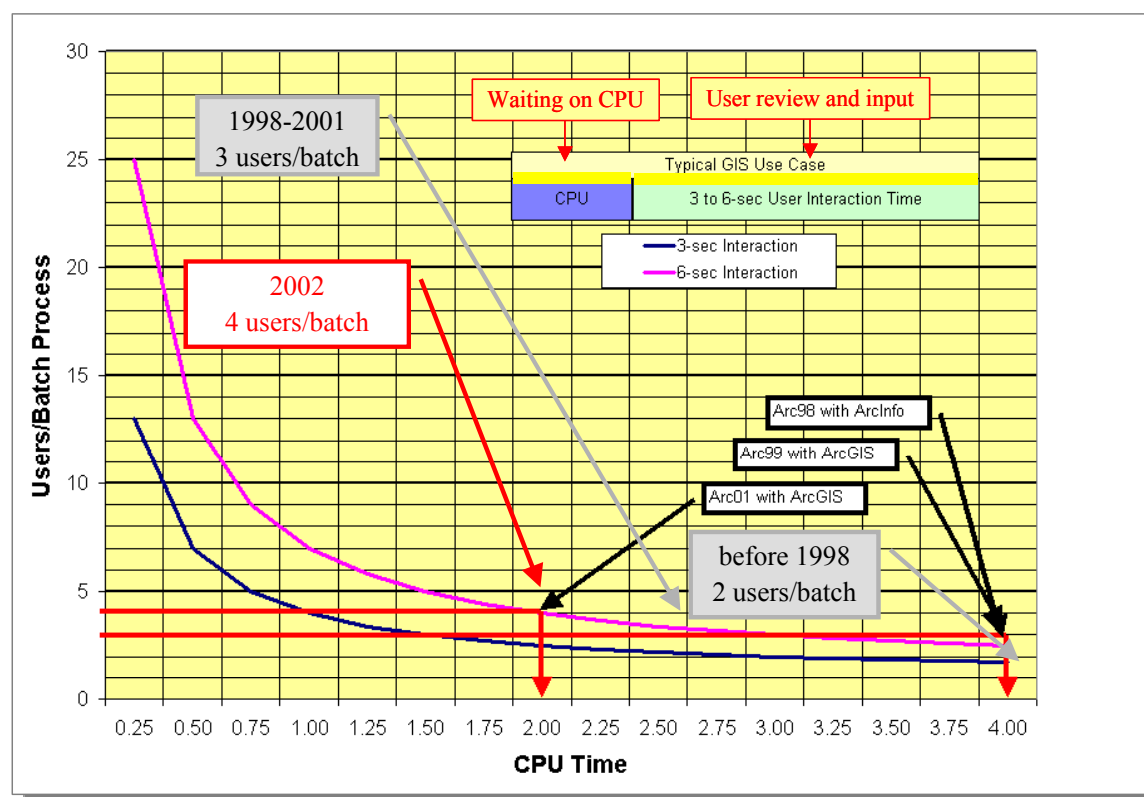
ArcGIS desktop users include a variety of professional analyst, data maintenance personnel, and business query and analysis specialists that use GIS to manage their data processing requirements. GIS project efforts can take minutes, hours, and sometimes several days to complete, including review of very large volumes of data resources during the project sessions.

User requirements can be used as a foundation for identifying the number of terminal sessions supported by the server during a peak processing period. The definition of a concurrent user (desktop use case) is based on comparison of real-life user loads on the system during live benchmark testing and reviewing real live operational GIS customer experiences.

Figure 6-15 provides an overview of a typical GIS use case, demonstrating how user performance (user experience) has changed as workstation performance has improved over the past 10 years. The standard use case applied to the Arc02 performance models assumes a typical

client display can be rendered within roughly 2 seconds, and the average user interaction time (time to review the display and enter data) is roughly 6 seconds. This results in an average user cycle time (between screen displays) of roughly 8 seconds. These assumptions establish the basis for translating our benchmark test results (batch process models) to sizing models for real interactive user environments. For the Arc02 performance models, users wait on CPU processing about 25 percent of the time. With these assumptions, a single ArcGIS desktop batch process provides the equivalent processing load of four concurrent ArcGIS desktop clients.

Figure 6-15
ArcGIS Use Case Parameters



The amount of time it takes a typical user to view a new display and enter appropriate data for the next display (user interaction time) is the fundamental difference between batch processing (no user interaction) and a real user work environment. GIS processing is compute intensive, resulting in users routinely waiting on computer processing during normal operations. The sizing models used prior to 1999 (Arc95 through Arc98) assumed ArcInfo users had to wait about 50 percent of their time for computer processing (terminal servers could support two users per CPU). Computing performance has increased over ten times of what was available five years ago. GIS operations have become more productive during this same period. The Arc99, Arc00, and Arc01 performance models assume GIS concurrent users wait on CPU processing 33 percent of the time, while the remaining time is spent on data review and input. The Arc02 model assumes users wait on CPU processing 25 percent of the time.

Platform performance has improved every year, and this technology improvement trend is expected to continue for the foreseeable future. As platform processing performance improves, the time GIS users wait on CPU processing will decrease and operational productivity may improve. At some point, which may already be the case, user interaction time will reach a minimum required response time based on user performance limitations. As the CPU processing time continues to decrease faster than the user interaction time, the relationship between batch processes and real interactive users will continue to improve.

Figure 6-16 provides an overview of the current terminal server performance model for a local disk or file server data source. Each CPU can support four (4) ArcInfo users, thus performance will stay roughly the same until the total number of users exceed four (4) times the number of CPUs. Performance slows at a relatively constant rate when the number of concurrent users increases beyond four (4) times the number of available CPUs.

Figure 6-16
Terminal Server with File Server or Local Data

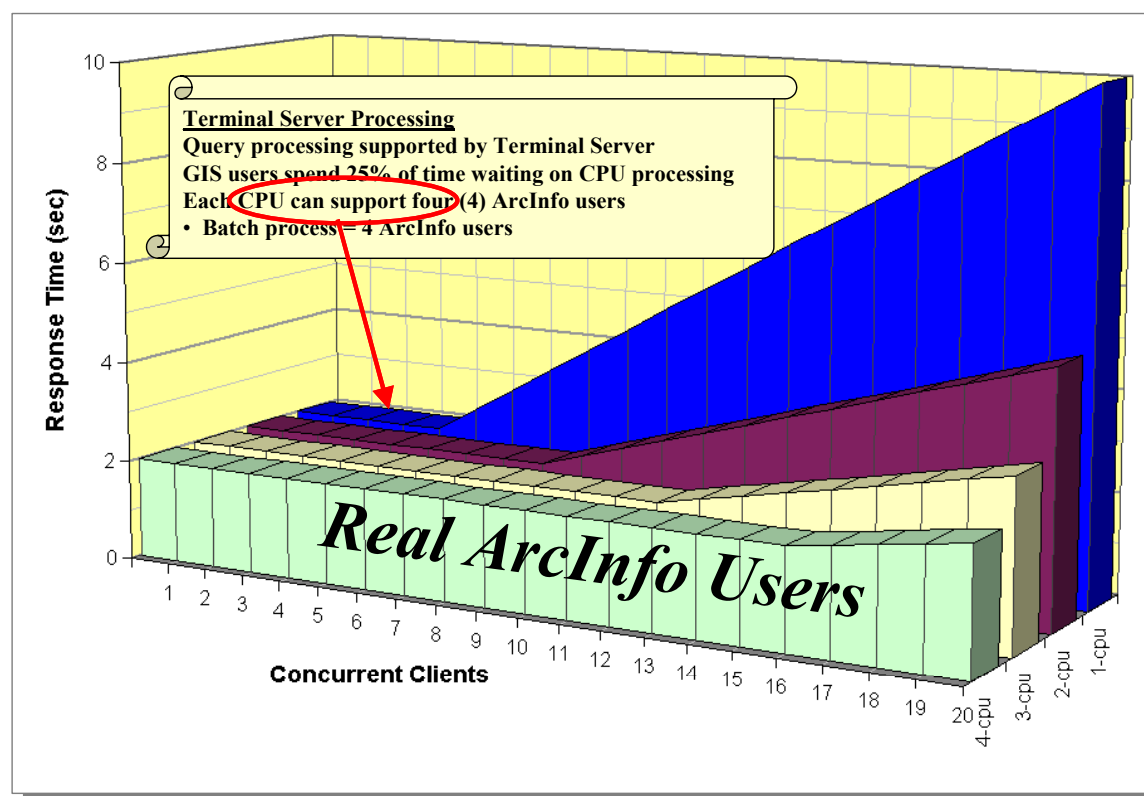


Figure 6-17 provides an overview of the current terminal server performance model with data on a remote ArcSDE server. The ArcSDE server supports query processing, reducing each user load on the terminal server by 50 percent. Each CPU can support eight (8) ArcInfo users, and performance will stay roughly the same until the total number of users exceeds eight (8) times the number of CPUs. Performance slows at a relatively constant rate when the number of concurrent users increases beyond eight (8) times the number of available CPUs.

Figure 6-17
Terminal Server with ArcSDE Server

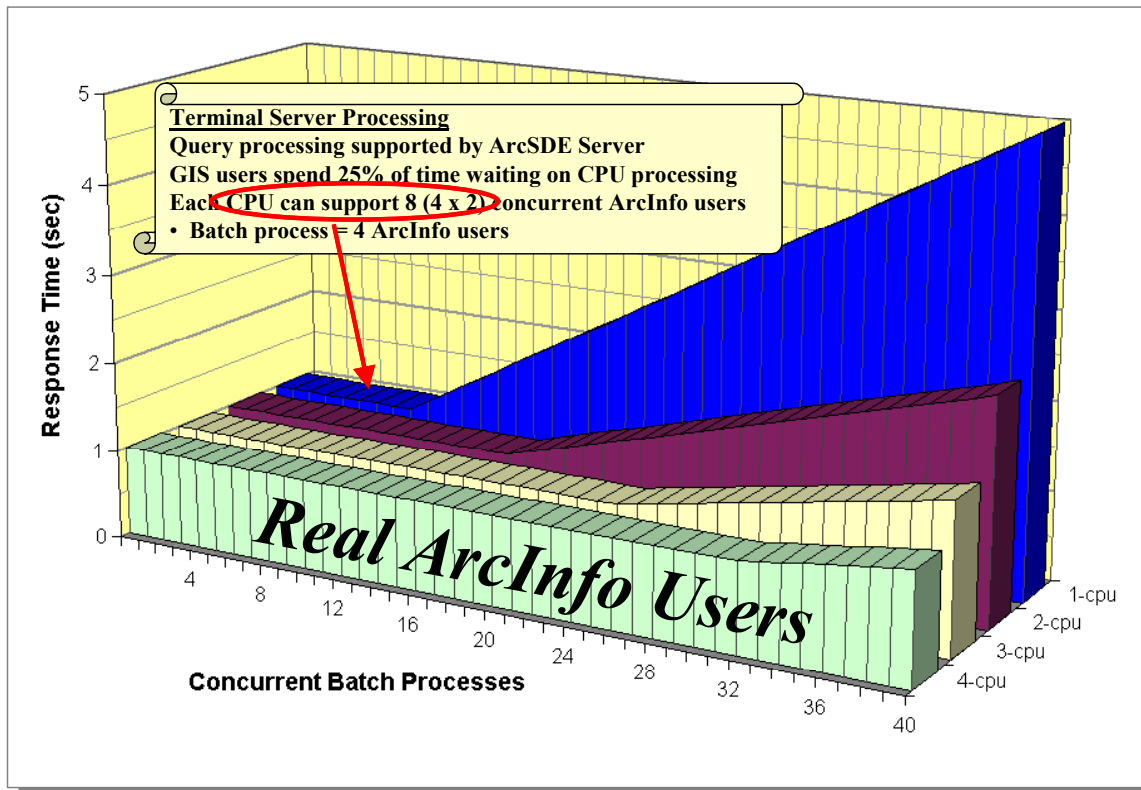
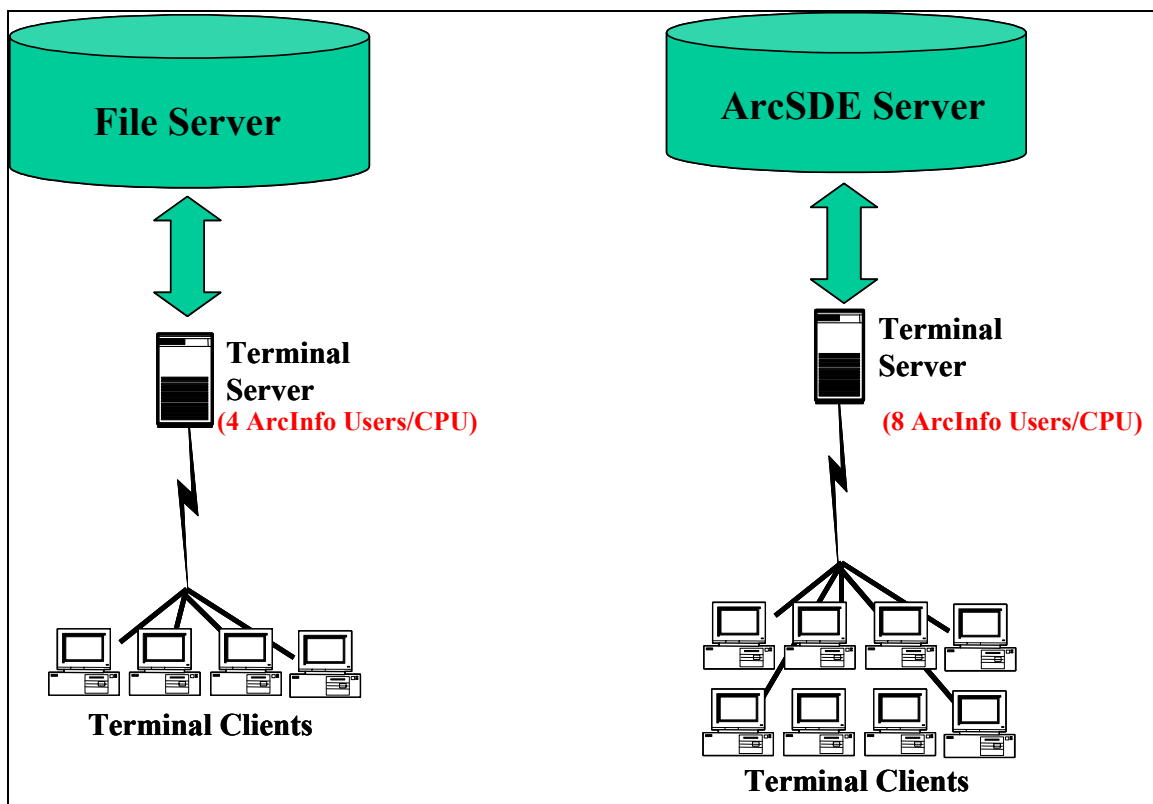


Figure 6-18 provides an overview of the terminal server sizing models based on configuring a system with the same single-CPU servers. Four (4) ArcInfo users can be supported per CPU when accessing data on local disk or a separate file server. Eight (8) ArcInfo users can be supported per CPU when accessing data on an ArcSDE server.

Figure 6-18
Terminal Server Performance Model



6.3.3 Data Server Performance

Data server client performance is an extension of the models introduced with batch processing. There are two primary types of GIS data servers. File servers support simple query processing from the client workstation. ArcSDE servers incorporate high-performance geodatabase solutions that support query processing on the data server platform. The Arc02 data server sizing models have been simplified based on ArcGIS 8.2 performance validation test results, providing a single model that will address both architectures.

The GIS file server model has supported effective system sizing requirements for GIS workgroup environments since the early 1990s. Customers that have followed these design guidelines have been able to support their server performance requirements. File server performance is directly related to the size of the files supporting the client query. Smaller files (spatial files with an extent relatively close to the size of the query display) will perform very well from a file server, and very little performance gain will be achieved with migrating these data to an ArcSDE server environment. On the other hand, larger data files (spatial files much

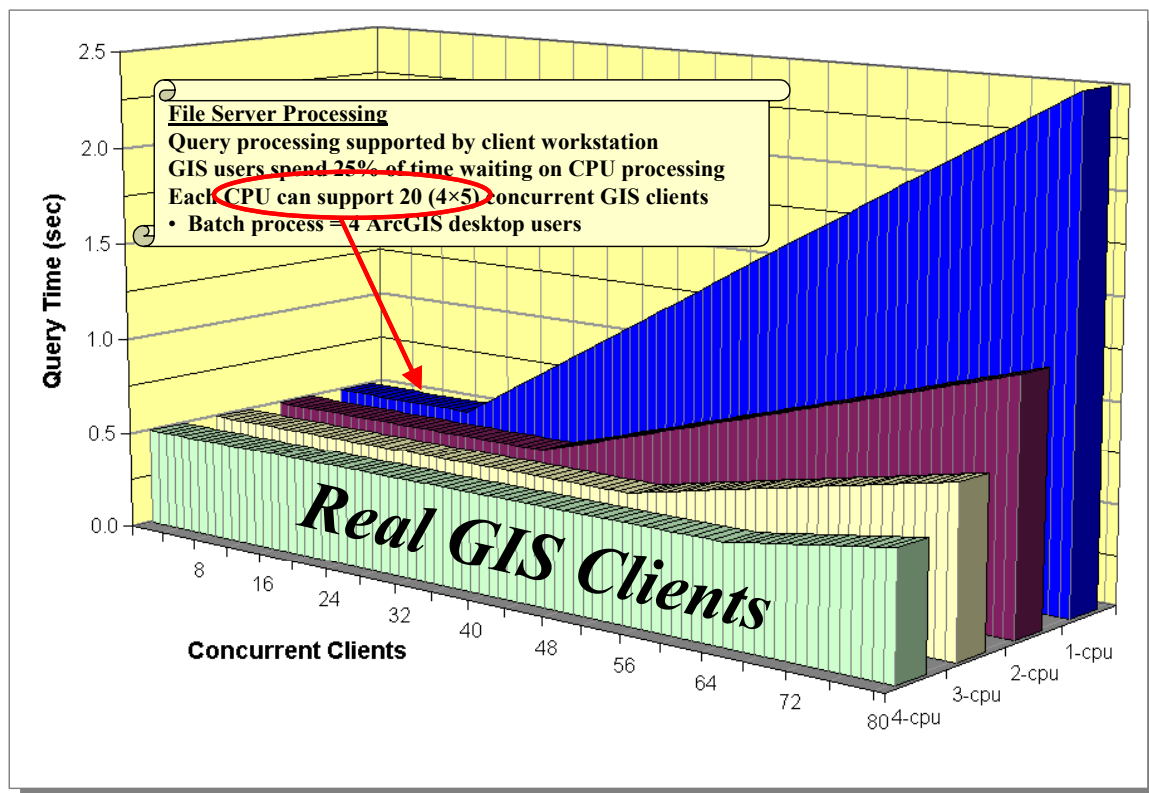
larger than the average user query display) will see a significant performance benefit when migrating these data to an ArcSDE server environment.

The ArcSDE data server provides the most scalable and highest performance GIS data source. Server query processing takes advantage of high-performance DBMS query functions and optimized data caching, reducing the overall server processing load and improving data access performance. ArcSDE is a logical choice for larger enterprise data sources and operational system environments.

The Arc02 performance model supports four (4) users per batch process, and each GIS data server CPU can support five batch processes. The GIS data server performance model will support 20 (5 x 4) concurrent ArcGIS desktop “real user” clients per CPU. Server query performance will remain roughly the same until data server clients exceed 20 times the number of available CPUs. Performance slows at a relatively constant rate when the total number of clients increases beyond 20 times the number of available CPUs.

Figure 6-19 provides an overview of the current GIS data server (ArcSDE and file server data source) performance sizing models.

Figure 6-19
GIS Data Server Performance

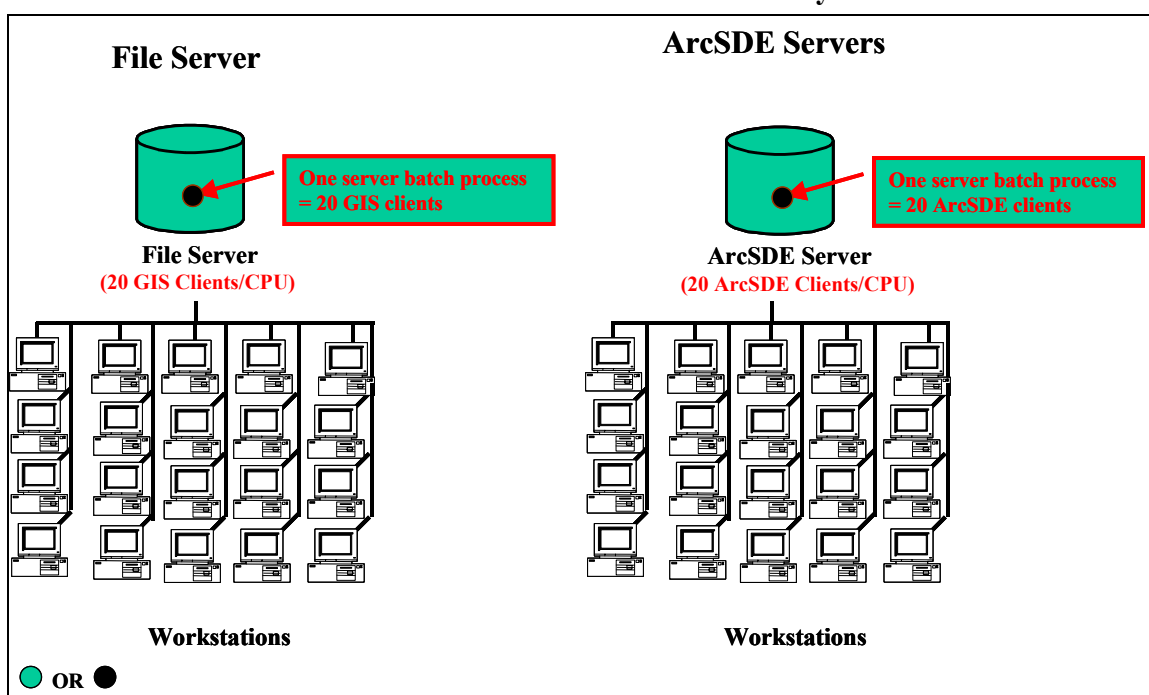


ArcSDE geodatabase performance is somewhat dependant on the complexity of the data model. The object-relational geodatabase provides a rich environment for establishing data standards and spatial feature dependencies. Relationships between geodatabase tables can generate

additional server loads during edit and query operations. Simple data models will perform better than the more complex geodatabase environments. Also, versioned database environments require more server load than non-versioned feature data sets. Several additional tables are included in a versioned query, which results in increased processing load on the server. ArcSDE performance tuning can be very effective, and performance considerations should be an integral part of database planning and administration.

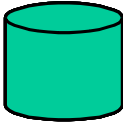
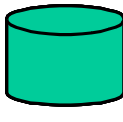
Figure 6-20 provides an overview of the data server sizing models based on configuring a system with the same single-CPU servers. A single-CPU GIS data server can support 20 GIS clients (a batch process located on the data server will take the place of 20 clients). The data model is the same for both GIS file server and ArcSDE server architectures.

Figure 6-20
Data Server Performance Summary



Several ESRI customers have established their current system architecture based on the previous year's platform sizing models. As technology and user performance expectations change, these models are reviewed and adjusted to support future design efforts. The ArcSDE sizing model represents the most critical and most significant adjustment from last years models. **Figure 6-21** provides a comparison of the Arc01 and Arc02 ArcSDE performance models.

Figure 6-21
ArcSDE Performance Model Comparison
(Arc01 and Arc02)

		Arc01 Model (30 GIS Viewers/CPU, 1 editor = 2 viewers)	Arc02 Model (20 ArcSDE Clients/CPU)
Peak Users		 ArcSDE Server	 ArcSDE Server
Editors	Viewers		
150	50	$350 / 30 = 11.6 \text{ CPU}$	$200 \div 20 = 10 \text{ CPU}$
100	100	$300 / 30 = 10 \text{ CPU}$	$200 \div 20 = 10 \text{ CPU}$
50	150	$225 / 30 =$	$200 \div 20 = 10 \text{ CPU}$

The Arc01 ArcSDE sizing model supported 30 concurrent GIS viewer users per CPU, and ArcGIS editor clients were doubled in the analysis to account for the heavier geodatabase processing loads with the more complex environments. Efforts this past year to reduce the editor loads through increasing edit cache and introducing disconnected editing solutions have supported adjustment to a more generic Arc02 performance model. The Arc02 sizing model supports 20 concurrent ArcGIS clients, with editors and viewers treated equally. The result of this model adjustment would favor sites with a high percentage of ArcGIS desktop editors, and provide the same platform recommendations for customers that have an equal mix of ArcGIS desktop editors and viewers. Customers with more ArcGIS desktop viewers than editors would require slightly higher server performance capacity when using the new Arc02 sizing models.

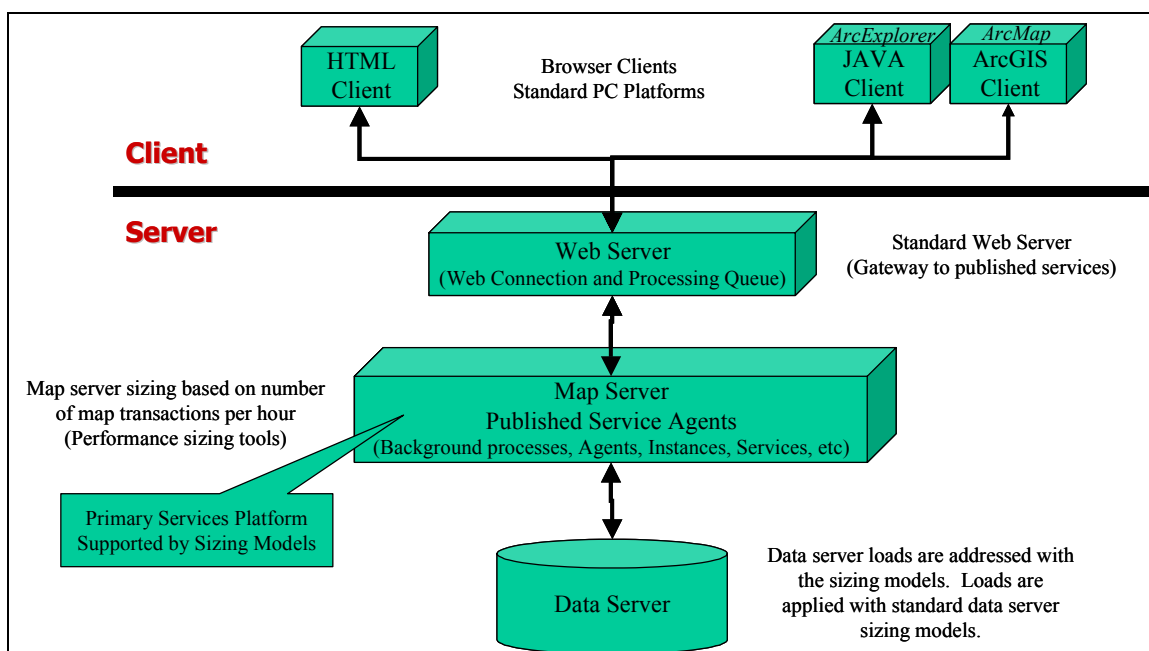
6.4 ArcIMS Web Map Services Models

ArcIMS Web services provide a transaction-based computing environment. Service agents (ArcIMS spatial servers) are established on map server platforms to support browser client requests for map services (i.e., create a map product). These map services are published (connected to the Web server) to support browser client Web service requests. Each map service can be considered a published map template that is processed by ArcIMS spatial server batch processes to support service requests from a browser client. Each client transaction generates a new map product.

The Web services architecture is a departure from traditional application environments, where user sessions are supported by dedicated executables (programs) running on workstations or application servers. Web services support random client requests (transactions), supporting a large number of users with a single service engine (background process).

Figure 6-22 provides an overview of the typical platform components supporting an ArcIMS Web site. Clients can be simple HTML-based browsers or more sophisticated client applications with access to the Web server. ArcIMS provides a Java client (ArcExplorer) that can be supported by a client browser to support an enriched client application environment. ArcIMS servers can also support basic HTML client browsers. ArcGIS clients (ArcInfo, ArcEditor, and ArcView 8.1) can access ArcIMS services as an intelligent browser client.

Figure 6-22
ArcIMS Platform Architecture



Several server components work together to support map services. The primary Web services components include the Web server, map server, and data server.

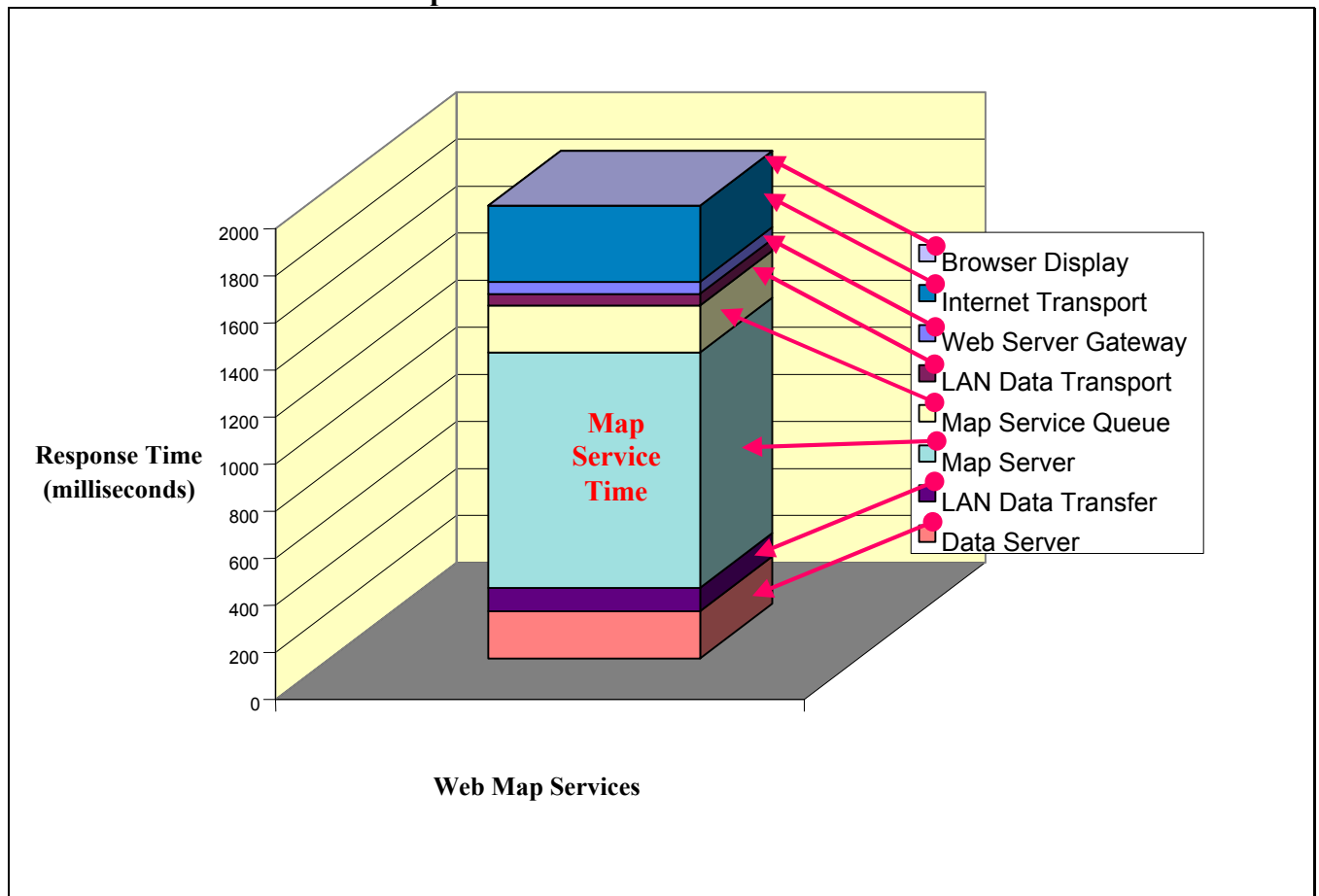
The Web server connects the browser clients with the ArcIMS map services. The specific ArcIMS Java servlet connectors and map service processing queues are typically installed on this platform.

The map server supports most of the ArcIMS processing load for the Web site. The map server includes the executables that service the map request (create the map product). For larger sites, this is typically a dedicated server or cluster of servers supporting the spatial server executables.

Data can be installed on the map server, or accessed from a separate GIS data server. A separate server is normally used for larger sites with multiple map server platforms.

Figure 6-23 provides an sample performance profile representing the life of a typical map service request. This graphic provides an overview of the platform and network loads associated with processing a Web transaction and delivering the HTML page and graphic image to the client browser.

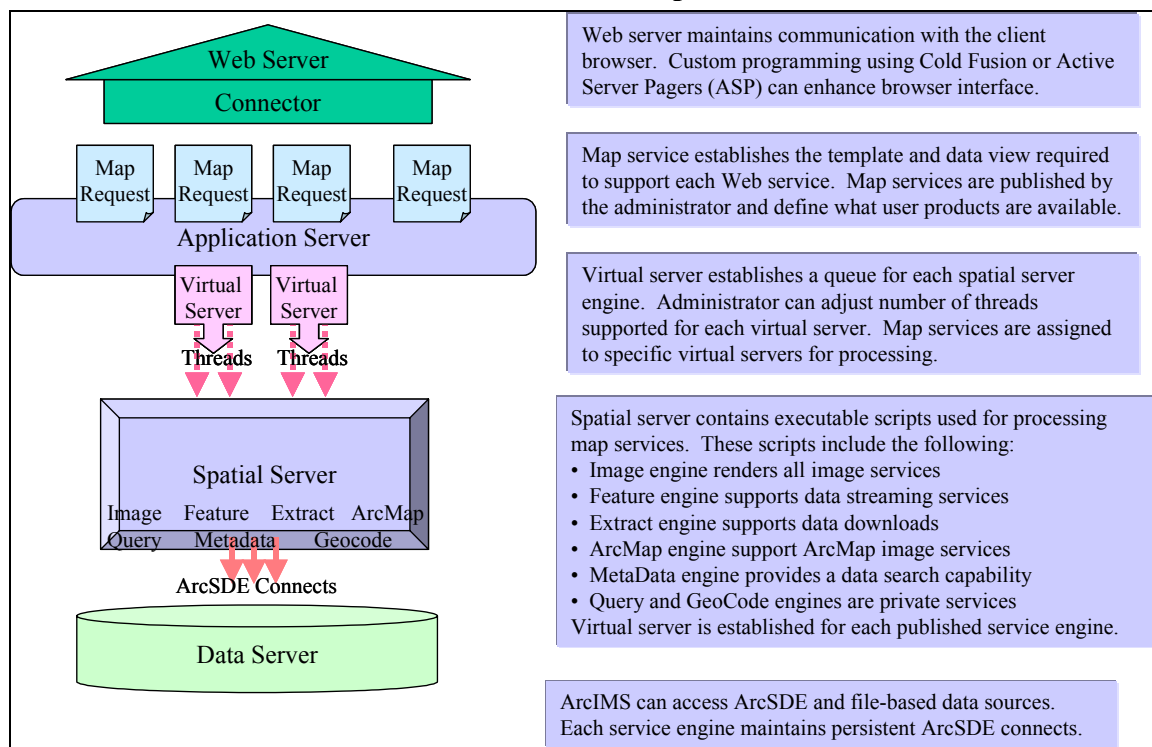
Figure 6-23
Map Services Performance Profile



6.4.1 ArcIMS Server Components

The ArcIMS software components provide a variety of configuration alternatives that support the Web services environment. **Figure 6-24** identifies the ArcIMS software components and where they are supported in the system platform architecture.

Figure 6-24
ArcIMS Server Components



The ArcIMS server is supported by the following functional components:

- **Web Server.** The Web server manages the interface and translation between the external network client communications and the ArcXML communications supporting the ArcIMS services. Web server components include Java connectors that translate HTTP XML traffic to ArcXML service requests.
- **Application Server.** The ArcIMS application server component establishes a physical location for the virtual servers. A virtual server is established for each ArcIMS service engine supported by the site environment (Image Virtual Service, Feature Virtual Service, etc.). The virtual server is registered with all the ArcIMS service engine threads supporting the designated service, and provides a waiting queue for inbound service requests and assigns these requests to available service threads for processing. Each inbound map service is assigned to the appropriate virtual server for processing, and waits in the queue until an agent thread is available to service the request.
- **Spatial Server.** The ArcIMS spatial server, or map server, provides the “container” for the ArcIMS service engines that process the inbound map service requests.

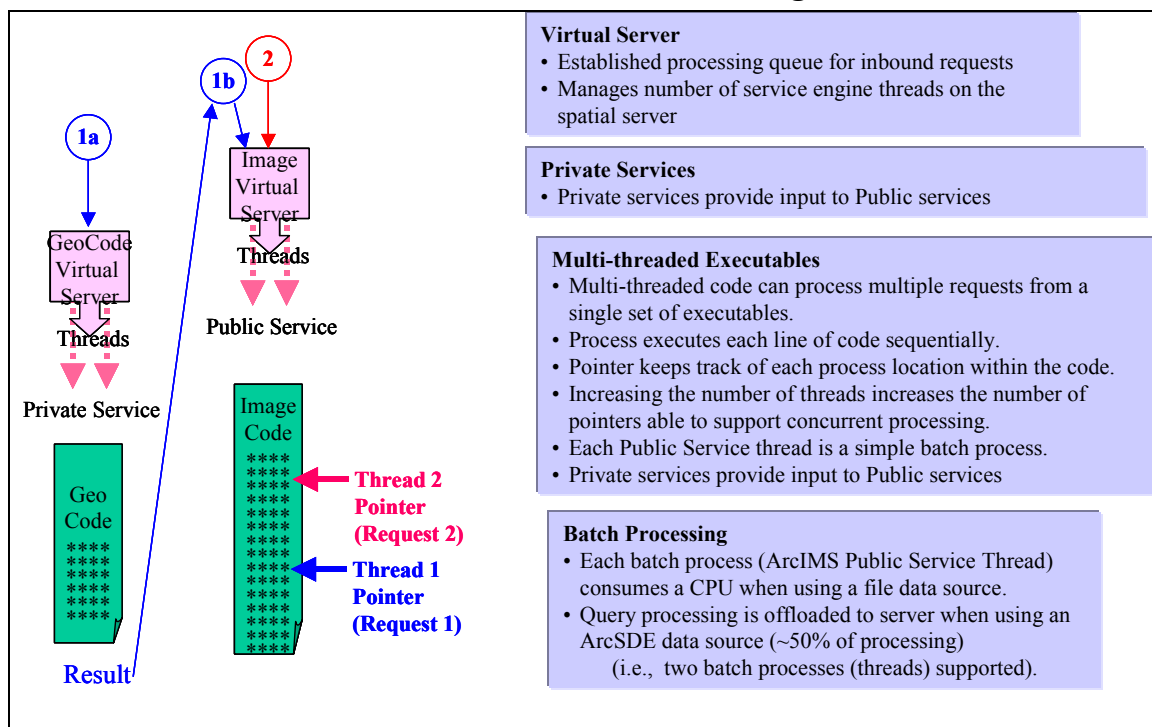
- **Data Server.** The ArcIMS service engines can access local shapefile, image, and ArcSDE data sources. The service engines support query processing for local shapefile and image data sources. The ArcSDE DBMS supports query processing for ArcSDE data sources. An ArcSDE for Coverages interface is provided to support access to ArcInfo coverage files.

ArcIMS is a transaction-based service, configured to support requested services as they arrive at the site. Each request is processed using preconfigured background scripts with persistent connections to the associated data source. During the service time, these scripts consume CPU resources the same as any other batch process. These scripts are idle when not processing requests. During peak processing loads, requests are held in a processing queue and assigned for processing as each script completes delivery of the previous request.

6.4.2 ArcIMS Multi-threaded Service Engines

The ArcIMS service engines located on the spatial server are compiled as multi-threaded executables. This means each executable can be configured to process multiple requests with a single set of code. This type of configuration simplifies site tuning and reduces the amount of executables supported by platform memory. **Figure 6-25** provides a depiction of how multi-threaded executables can process concurrent multiple requests.

Figure 6-25
Multi-threaded ArcIMS Service Engines



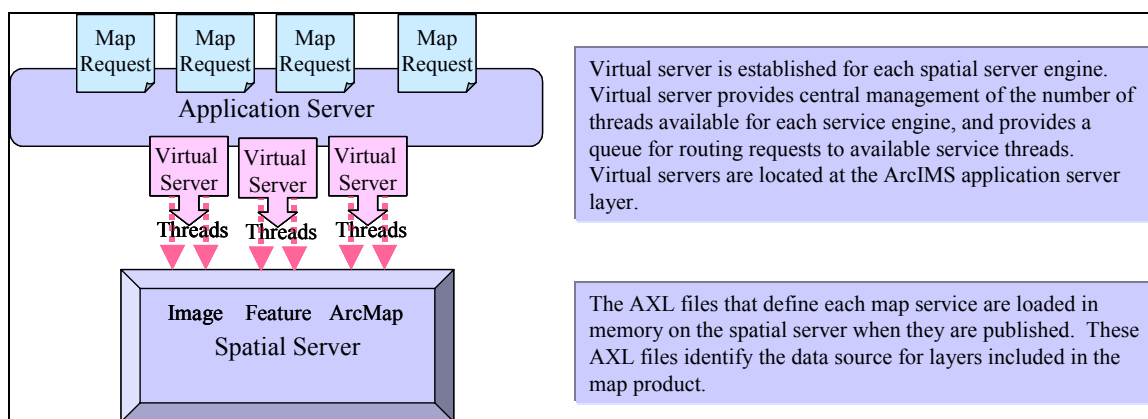
Each request is tracked through the code using a pointer. Each line of code is executed as the pointer selects that line of instruction. Each request will execute only one instruction at a time, same as any other batch process and, as such, cannot take advantage of more than one CPU at a

time. If two threads are assigned to a single server, that service can take advantage of up to two CPUs at the same time since it will be processing instructions for both pointers at the same time.

6.4.3 Virtual Services and the Processing Queue

When configuring the ArcIMS server, a virtual service is identified for each active service engine. Map services are then published relative to an associated virtual server. Requests for a specific map service are routed to the associated virtual server for processing. The virtual server is registered to associated spatial server service threads that will be used for processing the service request. Requests are held in the virtual server queue until an associated thread is available, at which time the service request is assigned to the thread for processing. Once the processing is complete, the next waiting service request can be assigned to that same thread. A single thread can process only one request at a time. **Figure 6-26** shows three virtual servers, each assigned two threads. One virtual server supports an image service engine, another virtual server supports a feature service engine, and the third virtual server supports the ArcMap service engine. The virtual servers are located with the ArcIMS application server component.

Figure 6-26
ArcIMS Virtual Servers



6.4.4 Published Map Services

Map services are authored and published using the ArcIMS Manager administration tools. Wizards are included to support publishing of ArcIMS map services. A large variety of simple map services can be configured out-of-the box using simple authoring tools. The map services are simple instruction sets (identifying layers and layout for the map service in an AXL file). These AXL files are assigned to a virtual server and appropriate network connections are preconfigured to support deployment of these service instructions. Once the map services are deployed, they can be requested by users through the Web site.

6.4.5 ArcIMS Performance Sizing Model

The ArcInfo batch process sizing model is used to support ArcIMS map server sizing. The maximum capacity of the ArcIMS spatial server is achieved with 100 percent map server CPU

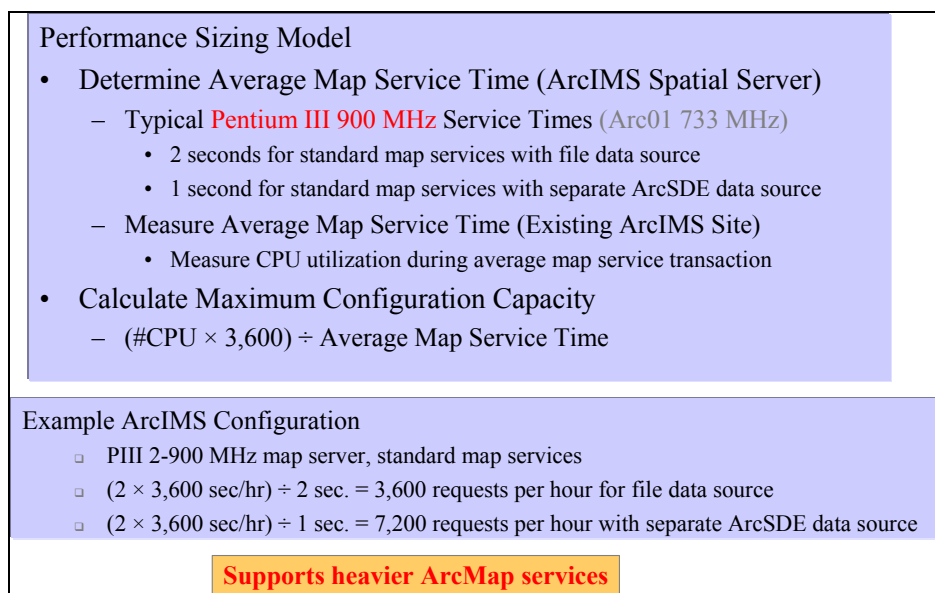
utilization. Each map service engine thread performs as a batch process, and consumes a CPU when servicing a request using a file data source (thread performs both query and rendering functions). Two service threads are required to consume a CPU when using an ArcSDE data source, same as with previous batch process models.

Figure 6-27 provides alternative options for estimating the peak transaction capacity of an ArcIMS map server. The peak requests per hour supported by the ArcIMS spatial server can be determined by dividing the number of seconds per hour by the average map service time in seconds.

ArcIMS spatial servers with ArcSDE data sources offload the query processing to the ArcSDE data server, reducing the service time on the map server by roughly 50 percent. ArcIMS map servers with a separate ArcSDE data source can generally support twice as many transactions per hour as the same map server with a file data source.

During ArcIMS peak loads, the average request wait time (time in the virtual server queue) is roughly four times longer than the average map service time (these estimates assume random arrival times; estimate is based on standard queuing theory). Total average wait plus service time is five times the average service CPU processing time during calculated peak system loads (100 percent CPU utilization on map server).

Figure 6-27
Map Services Performance Profile

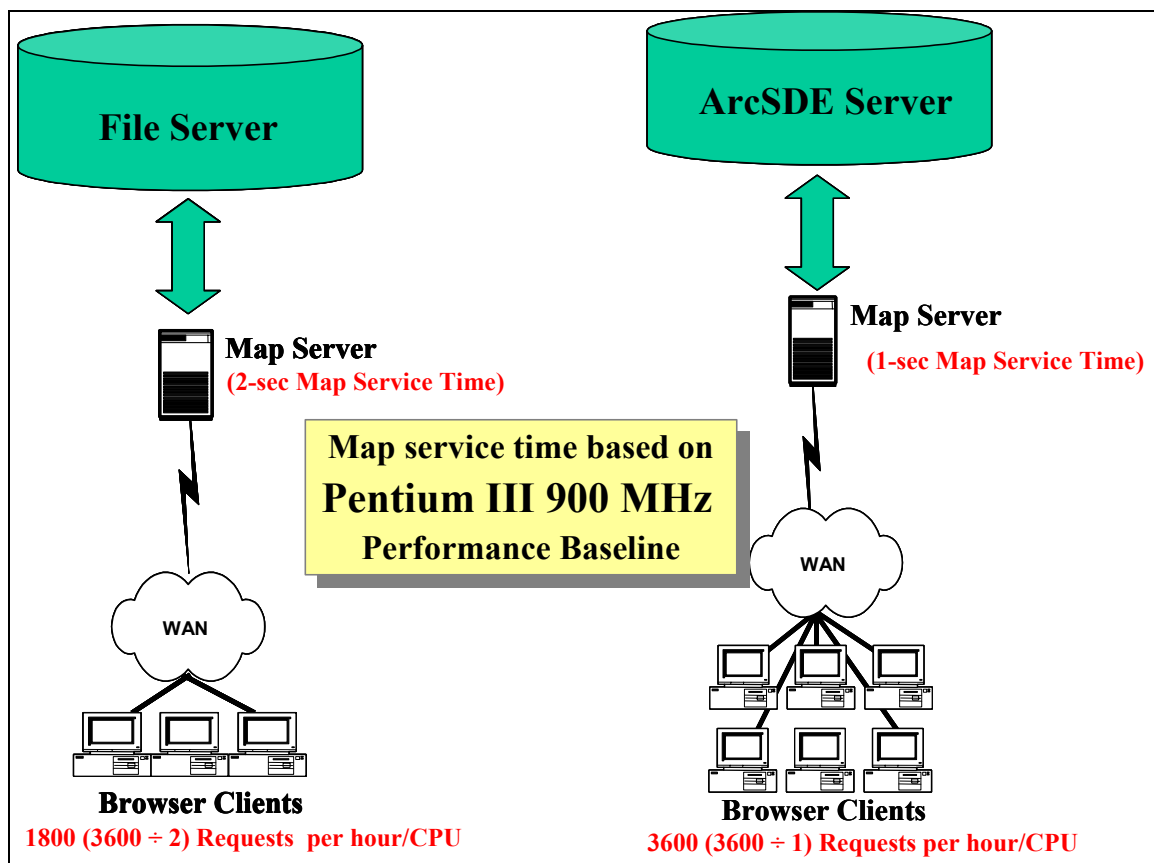


In many cases it is not practical to measure published ArcIMS map service times, since the specific map services have not been developed or published for testing. In these situations, hardware selection must be based on established performance standards. ArcIMS map server sizing models, based on a performance review of standard published map services, are provided to support users in selecting an appropriate hardware solution.

Figure 6-28 provides an overview of ArcIMS map server performance sizing models. A single-CPU Pentium III 900-MHz map server can support 1,800 transactions per hour with a file data

source and 3,600 transactions per hour with an ArcSDE data source. These transaction rates are based on average map service times of 2 seconds and 1 second, respectively.

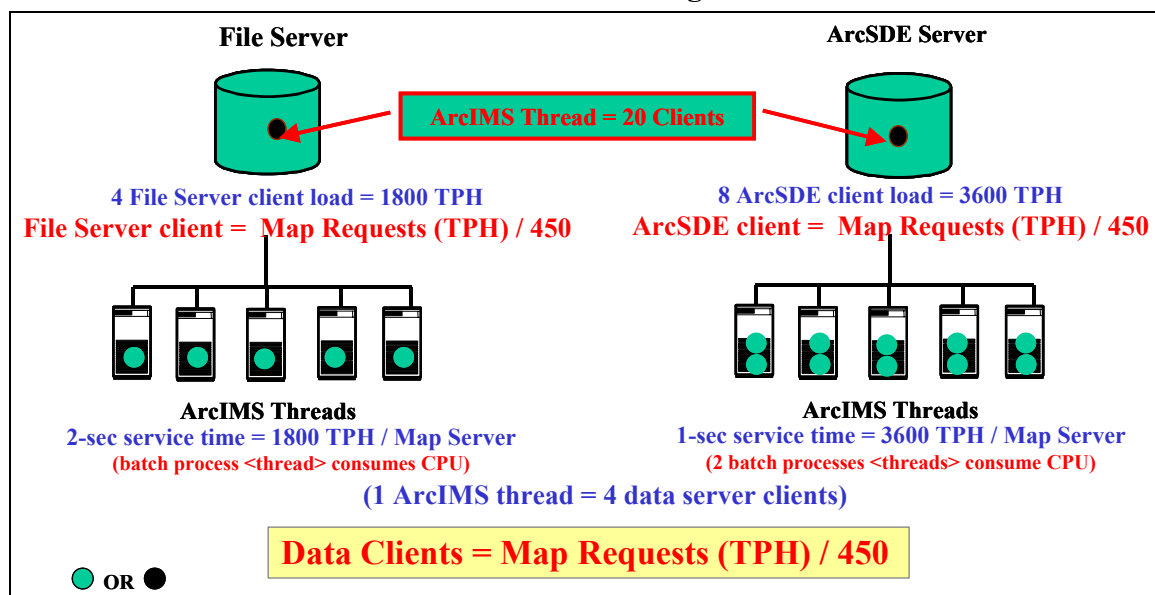
Figure 6-28
ArcIMS Performance Sizing Model



6.4.6 ArcIMS Data Server Loading Model

During peak ArcIMS loads, the map server CPUs are performing at 100 percent utilization. This is the same peak client platform performance experienced when supporting batch processes. For this reason, the ArcInfo batch process server load model can be used to identify ArcIMS peak data server loads. Peak data server loads are roughly 20 percent of the map server loads (1 data server CPU can support 5 map server CPUs). **Figure 6-29** provides an overview of the ArcIMS server loading model.

Figure 6-29
ArcIMS Server Loading Model



During peak transaction loads, an ArcIMS thread is equivalent to an ArcGIS desktop batch process. Based on our Arc02 performance models, a batch process is equivalent to four concurrent ArcGIS clients.

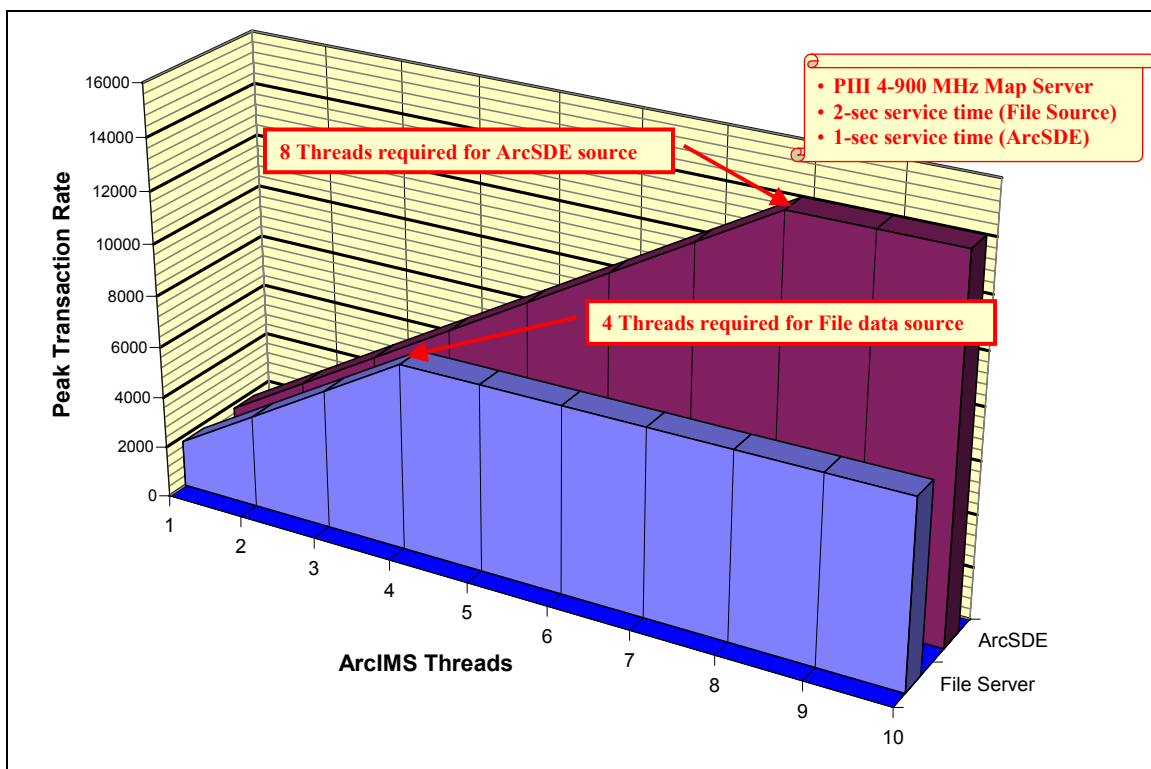
File Server Data Source. An ArcIMS map server with a file data source will reach peak capacity with a single ArcIMS thread (all primary processing supported by the ArcIMS service). Map service time is roughly 2 seconds, which supports a peak transaction rate of 1,800 map services (transactions) per hour. The generated file server load by the single ArcIMS service thread is equivalent to four (4) concurrent ArcGIS clients, which renders 1,800 map services per hour. A single ArcGIS client load would be equivalent to a peak ArcIMS load of 450 transactions per hour.

ArcSDE Data Source. An ArcIMS map server with an ArcSDE data source will require two (2) ArcIMS threads to reach peak platform capacity (query processing is supported by the separate ArcSDE server platform). Map server CPU processing time is roughly 1-second per request, which supports a peak transaction rate of 3,600 map services (transactions) per hour. The generated ArcSDE processing load from the two (2) ArcIMS service threads is equivalent to eight (8) concurrent ArcGIS clients, which renders 3,600 map services per hour. A single ArcGIS client load would be equivalent to a peak ArcIMS load of 450 transactions per hour.

The equivalent data client load generated from separate map servers can be estimated by dividing the projected ArcIMS peak transaction rate by 450 transactions per hour. If the ArcIMS spatial server were installed on the data server, the peak load per ArcIMS thread would consume a server CPU, which would be equivalent to 20 concurrent ArcGIS clients.

Figure 6-30 addresses the importance of configuring the proper number of ArcIMS threads on the map server to support peak transaction loads. The chart shows a Pentium III 4-CPU 900-MHz platform as the map server, and shows the peak transaction rate as a function of the number of ArcIMS threads. A minimum configuration of four (4) ArcIMS threads is required to reach full map server capacity with a file data source, and eight (8) ArcIMS threads with an ArcSDE data source.

Figure 6-30
Optimized Service Agent Configuration
Number of Threads Matter

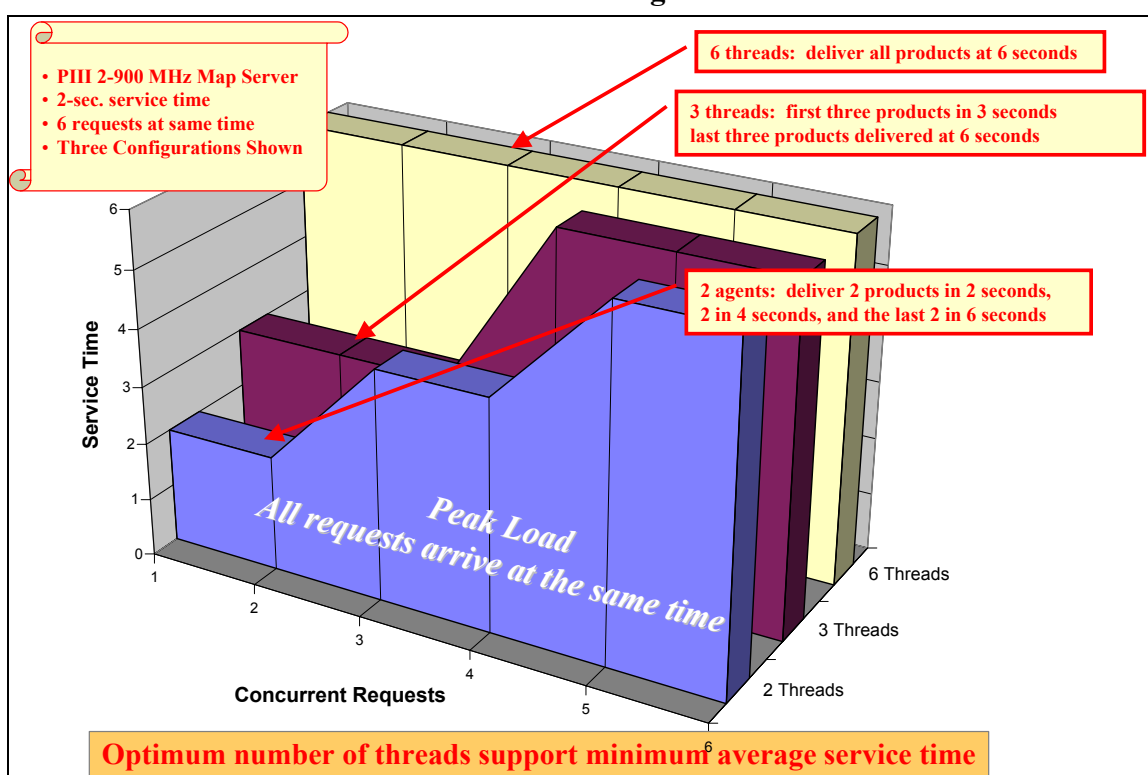


During ArcIMS operations, the virtual server queue does not release a request for processing once it receives the output from the ArcIMS spatial server thread. There can be a slight delay from the time the map server completes service processing and starts processing the next request. This delay is a result of network transfer time between the map server and the Web server. During ArcIMS tuning, it may be necessary to add an additional thread to achieve maximum map server output capacity. An additional thread will support an additional process on the map server to compensate for this delay.

Figure 6-31 addresses the importance of configuring the proper number of ArcIMS threads. This chart shows a Pentium III 900-MHz map server with two CPUs. Three different ArcIMS thread configurations are displayed on the chart (two threads, three threads, and six threads). Corresponding client response times for each of the six map requests are provided for each ArcIMS thread configuration, assuming all six requests arrive at the Web site at the same time.

- **Six-thread Configuration.** All six map requests are assigned by the virtual server to the six available spatial service threads. All requests are processed by the two available map server CPUs (service time for all six requests is 6 seconds).

Figure 6-31
Web Server Configuration



- **Three-thread Configuration.** Six requests arrive at the Web site. Three are assigned to the three available threads for processing, and three are held in the service queue waiting for available threads. The first three requests are processed in 3 seconds. The next three requests are processed in the next 3 seconds. Total service and wait time (query response time) for the first three requests is 3 seconds, and 6 seconds for the second three requests.
- **Two-thread Configuration.** Six requests arrive at the Web site. Two are assigned to the two available threads for processing and four are held in the service queue waiting for available threads. The first two requests are processed in 2 seconds, followed by the second two requests which are processed in the next 2 seconds, followed by the third two requests which are processed in the final 2 seconds. Total service and wait time for the first two

requests is 2 seconds, the second two requests 4 seconds, and the third two requests 6 seconds.

This simple analysis of the service sequence of arriving requests demonstrates the penalty for configuring too many ArcIMS threads. This analysis also shows the advantage of configuring the proper number of threads, supporting minimum average map service delays. The following guidelines are provided for configuring service agents for optimum Web performance.

- **Map Servers with File Data Source.** For each map service, configure one ArcIMS public service thread for each map server CPU.
- **Map Servers with ArcSDE Data Source.** For each map service, configure two ArcIMS public service threads for each map server CPU.

Include one additional thread to take advantage of map server CPU resources not being used during the slight delay between completing a request and starting on the next request (network transfer delay between map server and application server).

6.4.7 ArcIMS Server Performance Tuning

There are variables within the ArcIMS architecture that can be measured and modified to improve site performance. **Figure 6-32** identifies the associated performance measurements and configuration variables available in tuning an ArcIMS configuration.

Figure 6-32
Web Mapping Services Performance Profile

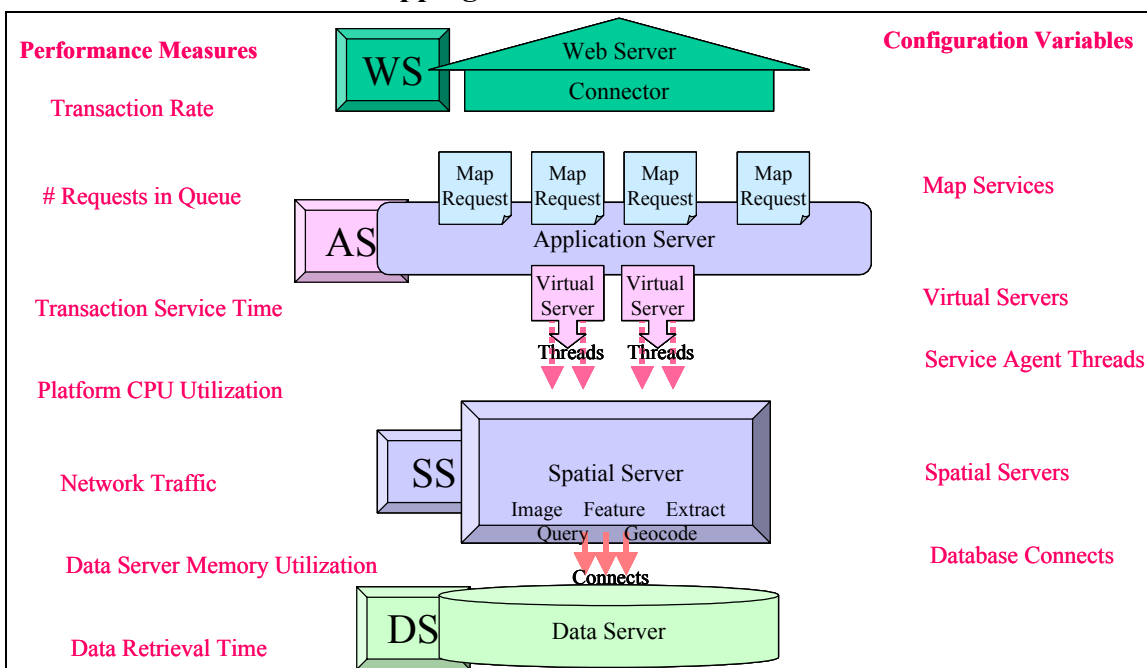


Figure 6-33 identifies things you can measure, and how you can modify ArcIMS map services or component variables to optimize site performance.

Figure 6-33
Web Mapping Services Performance Tuning Guidelines

Performance Measure	Tuning Options
Transaction Rate	Transaction rate identifies the number of requests supported by the site configuration. Peak transaction rate is the maximum capacity of the site to support incoming requests. Peak capacity can be increased by reducing the time required to generate each map service (simpler information products) or by increasing the number of CPUs (more processing power). Sufficient service agent threads should be included to take full advantage of the available CPUs (usually 2-3 threads per CPU is sufficient). Once CPUs are fully utilized, additional threads will not improve site capacity.
# Requests in Queue	Each virtual server acts as a processing queue for inbound requests. Requests are held in the queue until there is a service thread available to process the request. If a request arrives and the Virtual Server queue is full, the browser will receive a server busy error. Queue depth can be increased to avoid rejecting browser requests.
Transaction Service Time	CPU time required to process the published service, once it has been assigned to a service agent for processing. Long service times can significantly reduce site capacity, and should be avoided if possible. Simple map services (light data and minimum number of layers) can significantly improve site capacity.
CPU Utilization	Sufficient threads should be configured on the public service agents to support maximum CPU utilization. Peak site capacity is reached with CPU utilization reaches a peak level (close to 100% utilization). Increasing service threads beyond this point without increasing number of CPUs will increase average client response time, and is not recommended.
Network Traffic	Sufficient network bandwidth must be available to support information product transport to the client browser. Network bottlenecks can introduce serious client response delays. Bandwidth utilization can be improved by publishing simple map services, keeping image size from 30 KB to 50 KB, and ensuring sufficient bandwidth to support peak transaction rates.
Data Server Memory	Sufficient physical memory must be available to support all processing and adequate caching for optimum performance. Memory utilization should be checked once the system is configured to insure more physical memory exists than what is being used to support the production configuration.
Data Retrieval Time	CPU processing time on the ArcSDE server. Query time can be optimized by proper indexing and tuning of the ArcSDE database.

6.4 Performance Testing Pitfalls

For many of us, early in our childhood our science teacher introduced us to something called the scientific method. Few of us understood or remember the fundamental concepts that were

introduced in those early science fair projects. The scientific method provides a foundation for all scientific disciplines, and is very important.

Computer technology does not hold an exemption to the scientific method principals. Too many organizations try to understand system performance through testing, without a fundamental understanding of the system variables represented in the test. This practice can very easily result in misinterpretation and false conclusions.

The scientific method stresses the importance of modeling our environment, understanding what we are trying to prove and the variables contributing to our analysis before we test. Once we think we understand how our system works, we can test our hypothesis and further improve our understanding.

In summary, we are not ready to test until we think we understand what the result of the testing will be (this is our hypothesis). Testing is conducted to prove our hypothesis. If the testing fails, we go back to our models and try to develop a more accurate hypothesis. This iterative process is necessary to understand and quantify the performance of complex distributed system environments.

Figure 6-34 identifies some common pitfalls to watch out for when conducting or reviewing system performance test results. Conclusions drawn from simple benchmark testing seldom render a clear prediction of operational system performance.

Figure 6-34
Performance Testing Pitfalls

- **False Sense of Security**
 - Tendency to accept test results over analysis
 - Problem 1: Test results are a function of input parameters often not understood
 - Problem 2: System bottlenecks in testing can generate false conclusions
- **Simulated Load Testing May Not Represent Real World**
 - Load generation seldom represents actual user environments
 - Relationships between load generation and real world are seldom understood
 - Several system configuration variables can contribute to test anomalies
- **Performance Testing Best Practices**
 - Model system components and response parameters
 - Validate models based on real-world user loads testing
 - Predict test results from models (hypothesis) before conducting testing
 - Evaluate test results against models and original hypothesis
 - Update models and hypothesis, repeat testing until reaching consensus

“TEST ONLY WHEN YOU THINK YOU KNOW THE ANSWER”
“TESTING ONLY CONFIRMS WHAT YOU ALREADY KNOW”
“TESTING DOES NOT TEACH YOU WHAT YOU DON’T KNOW”

6.5 Conclusion

The sizing models presented in this section provide a foundation for understanding system performance issues. The models were presented throughout this section assuming CPU resources were all the same. In the real world, platform performance is not the same. Hardware vendors release new platforms every three to six months as faster CPUs enter the market and new technology is introduced to improve processing performance. The next section will address how these models can be applied to support system design in a world where each platform configuration has its own performance capacity.

7.0 Sizing Tools

The real world we live in today is experiencing the benefits of rapidly changing technology. Technology advancements are directly impacting our individual productivity—the way we all deal with information and contribute to our environment. Our ability to manage this change and take advantage of its benefits can contribute to our success in business and in our personal life.

Section 6 (Sizing Fundamentals) provided an overview of the system configuration performance models assuming all hardware platforms were the same. This section will identify how we can use these same models to make proper hardware selections when all hardware platforms are different.

The system design process identifies the optimum system configuration strategy, selects the appropriate hardware components, and provides platform specifications that will support user performance requirements. **Section 5** presented a methodology for identifying an enterprise system configuration strategy that would support specific user requirements. This methodology provides peak user loads for a list of selected platform components required to support the system solution. This section will provide practical tools that can be used to convert peak user platform loads to specific platform specifications.

7.1 *Performance Baseline Selection*

To develop a system design, it is necessary to identify user performance needs. User performance requirements are represented by the workstation platforms selected by users to support computing needs. Application and data servers must be configured to support user desktop performance requirements.

User performance expectations have changed significantly over the past several years. This change in performance needs is encouraged primarily by faster platform performance and lower hardware costs. This change directly contributes to increased user productivity.

Figure 7-1 identifies the change in single-user ArcInfo platform requirements since the ArcInfo 7.0.2 release in September 1994.

Figure 7-1
ArcInfo Platform Performance History

- **ArcInfo 7.0.2 (Sept. 1994)**
 - Sun SPARCstation 10 Model 40, 32 MB Memory
- **ArcInfo 7.0.4 (Feb. 1996)**
 - Sun SPARCstation 20 Model 71, 64 MB Memory
- **ArcInfo 7.1.1 (Feb. 1997)**
 - Pentium Pro 200, 64 MB Memory
- **ArcInfo 7.2.1 (April 1998)**
 - Pentium II 300, 128 MB Memory
- **ArcInfo 8 (July 1999)**
 - Pentium III 500, 128 MB Memory
- **ArcInfo 8.0.2 (May 2000)**
 - Pentium III 733, 256 MB Memory
- **ArcInfo 8.1 (February 2001)**
 - Pentium III 900, 256 MB Memory
- **ArcInfo 8.2 (May 2002)**
 - Pentium 4 1500, 512 MB Memory

The Sun SPARCstation 10 Model 40 was a powerful single-user workstation in 1994. This platform was competitive with similar price/performance options provided by other UNIX vendors during this same period.

In February 1996, ESRI released ArcInfo 7.0.4. Sun released their SPARCstation 20 models, and the SPARCstation 20 Model 71 was a common selection as a single-user workstation. The SPARC 20 Model 71 was roughly 2.5 times faster than the SPARC 10 Model 40.

In February 1997, ESRI released ArcInfo 7.1.1. The first Windows ArcInfo product was supported on the Intel Pentium Pro 200. The Pentium 200 platforms provided standard workstation performance for ArcInfo users during 1997. The Pentium 200 performed more than four times faster than the SPARC 10 Model 40.

In April 1998, ESRI released ArcInfo 7.2.1. Intel was selling the Pentium II-300 platform, which provided the standard workstation performance baseline for ArcInfo users during 1998. The Pentium II-300 platforms performed more than six times faster than the SPARC 10 Model 40.

ArcInfo application performance improved throughout this same period as the code was optimized with new ArcInfo 7 incremental releases. A script developed with ArcInfo 7.0.2 in September 1994 would run faster using the ArcInfo 7.2.1 release, both running on the same platform. The change in the ArcInfo performance baseline was not a software-driven requirement, but rather a change in user performance expectations brought on by the faster and less expensive platform technology.

In July 1999, ESRI demonstrated the first release of ArcInfo 8. The Intel Pentium III-500 platform was the popular ArcInfo workstation candidate. The Pentium III-500 was over 11 times faster than the SPARC 10 Model 40.

In May 2000, ESRI released ArcInfo 8.0.2. The Pentium III-733 was selected as the performance baseline supporting the summer 2000 deployments. The Pentium III 733-MHz platform is over 22 times faster than the SPARC 10 Model 40.

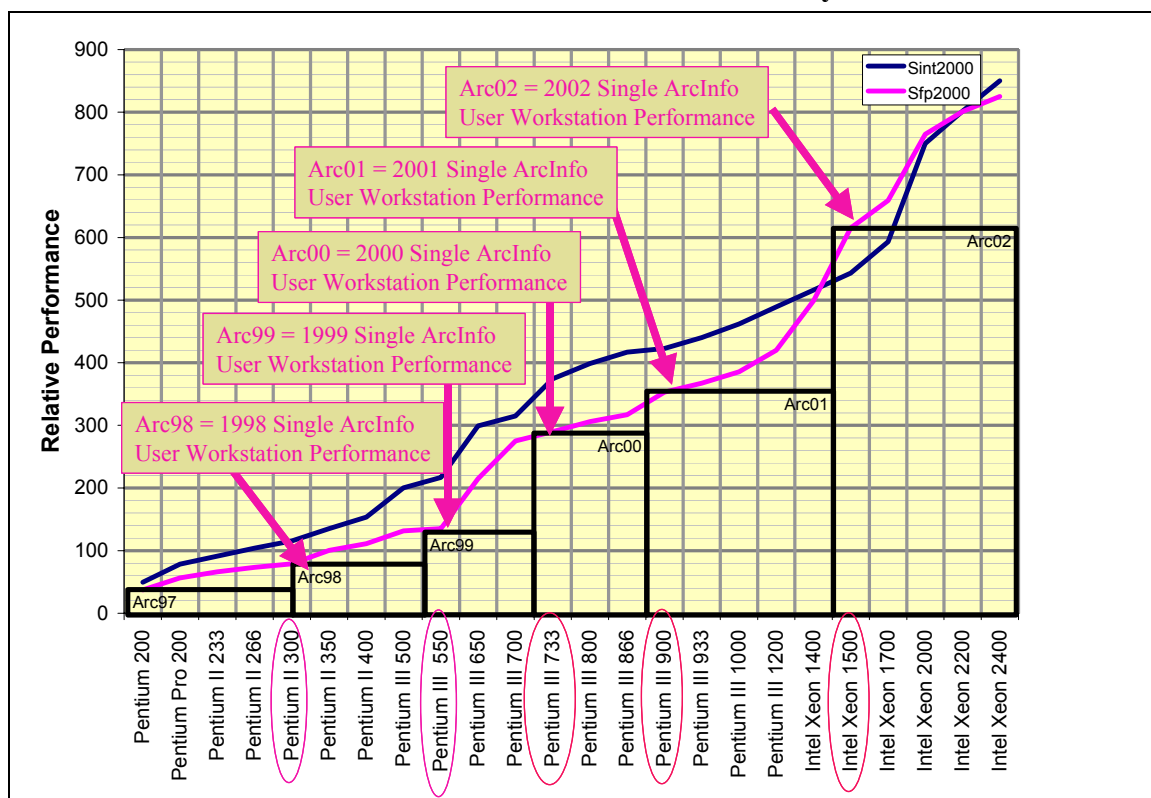
Modest performance improvements continued into 2001. The slowdowns in performance increases were contributed to development of the next-generation processor technology. The Pentium III 900-MHz platform is selected as the performance baseline supporting the ArcInfo 8.1 release.

Intel release of the Intel Xeon MP 1500-MHz server provides a 2002 performance boost for Intel-based server platforms. Intel server technology has been lagging behind workstations, which are reaching the Intel Xeon 2400-MHz performance thresholds by mid-year. IBM and Sun UNIX server platforms are keeping pace with the Intel workstation performance levels. The Intel Xeon MP 1500-MHz platform was selected as the performance baseline for the ArcInfo 8.2 release.

Platform performance improvements have a direct impact on infrastructure investment strategies. Most organizations today find it essential to reinvest in hardware on a three- to five-year cycle to stay productive.

Figure 7-2 provides a graphic overview of Intel workstation performance over the past ten years. Vendor-published commercial benchmark results establish the relative performance represented on this chart.

Figure 7-2
UNIX Platform Performance History



The boxes along the bottom of the chart represent ArcInfo workstation user performance requirements. Each box is labeled to represent the corresponding year (Arc98 represents the 1998 performance baseline, Arc99 the 1999 performance baseline, and so forth, through Arc02 which represents the year 2002 performance baseline). These performance baselines agree with the user platforms identified on **Figure 7-1**, and represent user performance requirements during that time period. You can see from the chart that ArcInfo platform performance is more than 15 times faster today than five years ago.

7.2 How Do We Handle Change?

Hardware platforms used to configure an enterprise GIS environment may include many configurations and different models. Each configuration must support user performance needs for the system to perform properly. Platform performance is not represented by the number of CPUs or the CPU speed (MHz) alone. There are many components within a specific platform configuration that contribute to overall performance including the design and model of the CPU, the speed of the system bus, and the installed operating system.

Benchmark performance of a specific platform model and configuration is a better measure of total performance capacity than the number of CPUs. A common set of benchmarks can be used to measure the relative performance of two different hardware platform configurations. If you know the relative performance of two platform configurations, you can identify the corresponding relative performance capacity of those two platforms.

Figure 7-3 provides a definition of the theory of relative performance that can be used along with the sizing models in **Section 6** to quantify platform specification requirements based on user needs.

Figure 7-3
Theory of Relative Performance

$$\begin{array}{c} \textbf{Theory of Relative Performance} \\ \textbf{The relative performance of two servers is} \\ \textbf{directly proportional to their compute capacity} \\ \\ \frac{\textbf{Performance of Server A}}{\textbf{Performance of Server B}} = \frac{\textbf{Clients of Server A}}{\textbf{Clients of Server B}} \end{array}$$

The theory of relative performance identifies the compute capacity of a new server based on its relative performance to a known server configuration. As an example, if we know an existing server can support 100 clients and that the new server has twice the performance as the existing server, we then can expect the new server will support 200 clients.

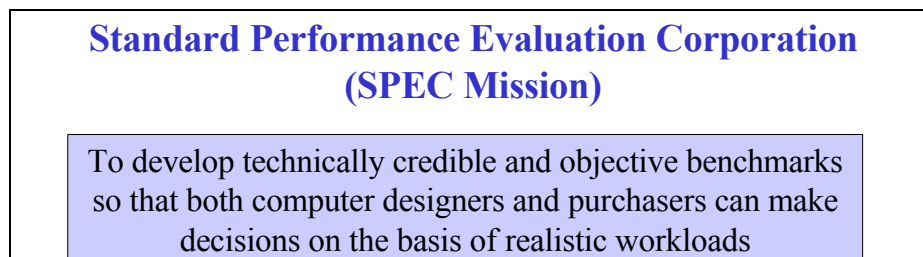
7.3 *How Do We Measure Change?*

It is a challenge to find a reliable and objective set of benchmarks that can be used to establish relative performance between vendor platforms. It is desirable to find a set of benchmarks that are simple, consistent, and reliable over time (we would like to know the performance of newer workstations relative to older models that exist in the current inventory). We would also like every vendor platform to be tested against these benchmarks, and these benchmarks published by the vendor and made available on the Web for customer reference. We would also like these benchmarks to be neutral in nature, and not to have their validity challenged by the vendors.

The Standard Performance Evaluation Corporation (SPEC) was formed during the 1980s to identify and create objective sets of application-oriented tests that can serve as common reference points and be used to evaluate performance across different workstation platforms.

Figure 7-4 provides the current SPEC charter. ESRI system design consultants have used the SPEC benchmark suite since 1992 as an independent source to identify relative performance of supported hardware platforms. The SPEC relative performance measures have been very helpful in representing relative platform performance, and have been used with ESRI sizing models to identify appropriate platform solutions to support GIS user performance needs.

Figure 7-4
How Do We Measure Change?



SPEC announced the release of the SPEC2000 benchmark suites in early 2000. **Figure 7-5** provides an overview of the SPEC2000 benchmark suite. SPEC2000 includes two sets (or suites) of benchmarks: CINT2000 for compute-intensive integer performance and CFP2000 for compute-intensive floating-point performance. The two suites provide component-level benchmarks that measure the performance of the computer's processor, memory architecture, and compiler. SPEC benchmarks are selected from existing application and benchmark source code running across multiple platforms. Each benchmark is tested on different platforms to obtain fair performance results across competitive hardware and software systems.

Figure 7-5
SPEC2000 Benchmark Suites

- **SPEC2000 Comprises Two Suites of Benchmarks**
 - **CINT2000: Compute-Intensive Integer**
 - Twelve (12) CPU-intensive integer benchmarks (C and C++ Language)
 - Conservative (SPECint_base2000, SPECint_rate_base2000)
 - Aggressive (SPECint2000, SPECint_rate2000)
 - **CFP2000: Compute-Intensive Floating Point**
 - Fourteen (14) CPU-intensive floating-point benchmarks (FORTRAN 77 & 90 and C Languages)
 - Conservative (SPECfp_base2000, SPECfp_rate_base2000)
 - Aggressive (SPECfp2000, SPECfp_rate2000)
- **Sun Ultra 10 300-MHz Reference Platform**

The CINT2000 suite, written in C and C++ languages, contains twelve CPU-intensive integer benchmarks. CINT2000 is used to measure and calculate the geometric mean of twelve normalized ratios (one for each integer benchmark) when compiled with aggressive (SPECint2000) and conservative (SPECint_base2000) optimization for each benchmark, and the geometric mean of twelve normalized throughput ratios when compiled with aggressive (SPECint_rate2000) and conservative (SPECint_rate_base2000) optimization for each benchmark.

The CFP2000 suite, written in FORTRAN 77 & 90 and C languages, contains fourteen CPU-intensive floating-point benchmarks. CFP2000 is used to measure and calculate the geometric mean of fourteen normalized ratios (one for each floating point benchmark) when compiled with aggressive (SPECfp2000) and conservative (SPECfp_base2000) optimization for each benchmark, and the geometric mean of fourteen normalized throughput ratios when compiled with aggressive (SPECfp_rate2000) and conservative (SPECfp_rate_base2000) optimization for each benchmark.

7.4 *Hardware Platform Selection*

Several factors should be considered when selecting a platform vendor. The two most visible and obvious criteria are platform performance and purchase price. These criteria are critical to system design and acceptance but do not address other hidden factors that contribute to total system cost. Today's leading hardware platform vendors are very competitive in both technical performance and pricing. Purchase of a GIS platform environment may extend over several months or possibly years, and platform performance leadership will change several times between leading hardware manufacturers during this period.

System supportability may be a deciding factor. In many cases, companies invest heavily in training and staff infrastructure in support of a particular hardware vendor solution. Hiring, training, and maintaining staff to support the appropriate computer environment is an important factor contributing to system success.

Company relations with vendor sales and support personnel may be a deciding factor in choosing the appropriate vendor. Vendor support is not the same throughout the country, and these personal relationships can make a difference in how well the system will be supported.

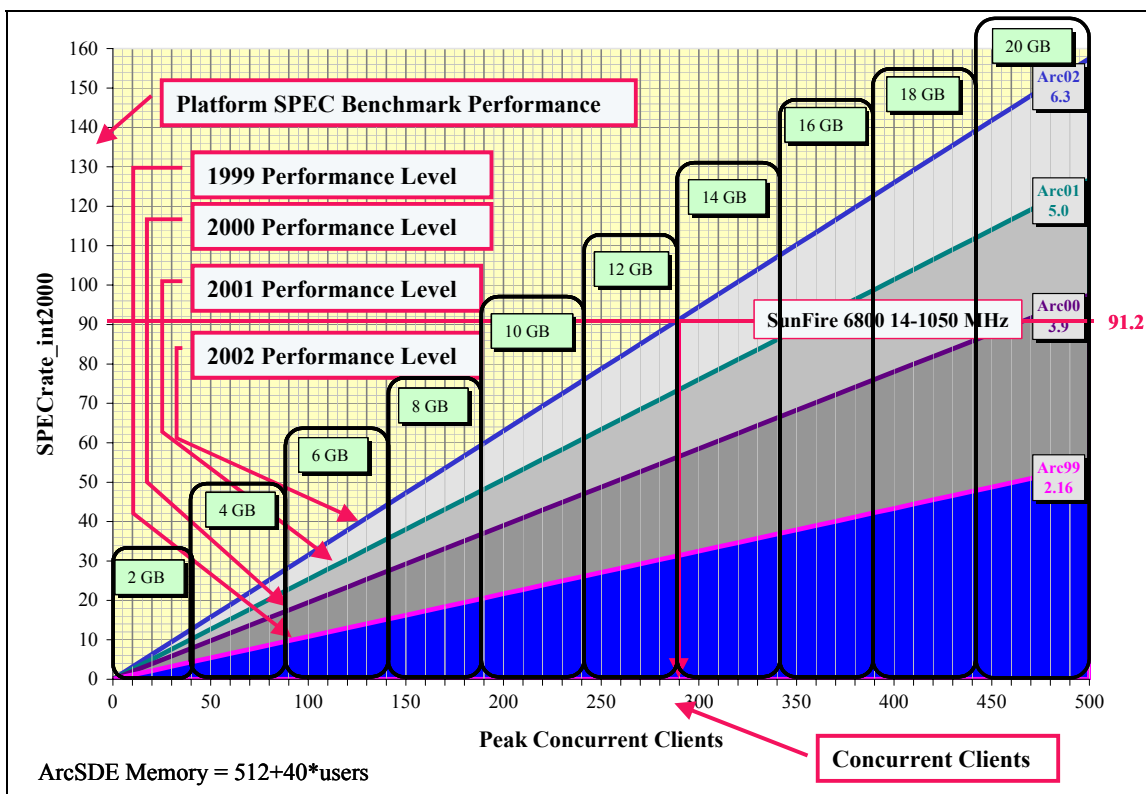
Total life-cycle cost and performance of the system should be the primary considerations in selecting a vendor solution. This section will identify a methodology for establishing vendor configurations with common performance expectations. System solutions with common performance expectations provide a basis for life-cycle cost comparisons.

Once a vendor is selected, a review of the available candidate hardware platforms, along with their associated SPEC benchmark ratings, will provide a basis for satisfying configuration needs. The following information is helpful in evaluating candidate platforms for configuring a system solution.

7.4.1 Platform Sizing Model Format

Figure 7-6 introduces a methodology for presenting the performance models for each of the ESRI product solutions. Understanding how to read and apply this chart will provide a foundation for using the performance models in the rest of this section.

Figure 7-6
Platform Performance Chart Overview



The left side of this chart represents the platform performance based on published SPEC benchmarks for each selected platform configuration. Application compute platforms will use the SPECrates_fp2000 benchmark results, while data server platforms will use the SPECrates_int2000 benchmark results.

This chart provides a plot of performance levels for 1999, 2000, 2001, and 2002 represented by diagonal lines across the chart. These lines are labeled Arc99, Arc00, Arc01, and Arc02 to represent performance baselines for the respective years. The published single processor benchmark values are identified on each performance line.

The concurrent clients are represented along the X-axis of the chart. Memory recommendations are a function of the concurrent client connects and are represented on the performance chart. The formula used for establishing memory requirements is identified on the bottom left corner of the chart.

A specific platform configuration is represented on this chart by a horizontal line (SunFire 6800 14-CPU 1050-MHz platform is a fourteen-CPU configuration with a SPECint_rate2000 performance benchmark of 91.2). The intersection of the platform performance with the respective baseline performance line identifies the memory requirements and number of clients that can be supported at that performance level by the identified platform configuration. (The example shows 290 clients can be supported at the Arc02 performance level, 14-GB memory would be required to support this number of clients). The same chart can be used to identify configuration requirements for each of the other performance levels (Arc99, 00, 01, and 02) following the same procedure. (The same SunFire 6800 14-CPU 1050-MHz configuration can support close to 500 concurrent clients at the Arc00 performance level, 20 GB of physical memory would be required to support this number of client connects).

7.4.2 Where Can I Find the Performance Benchmarks?

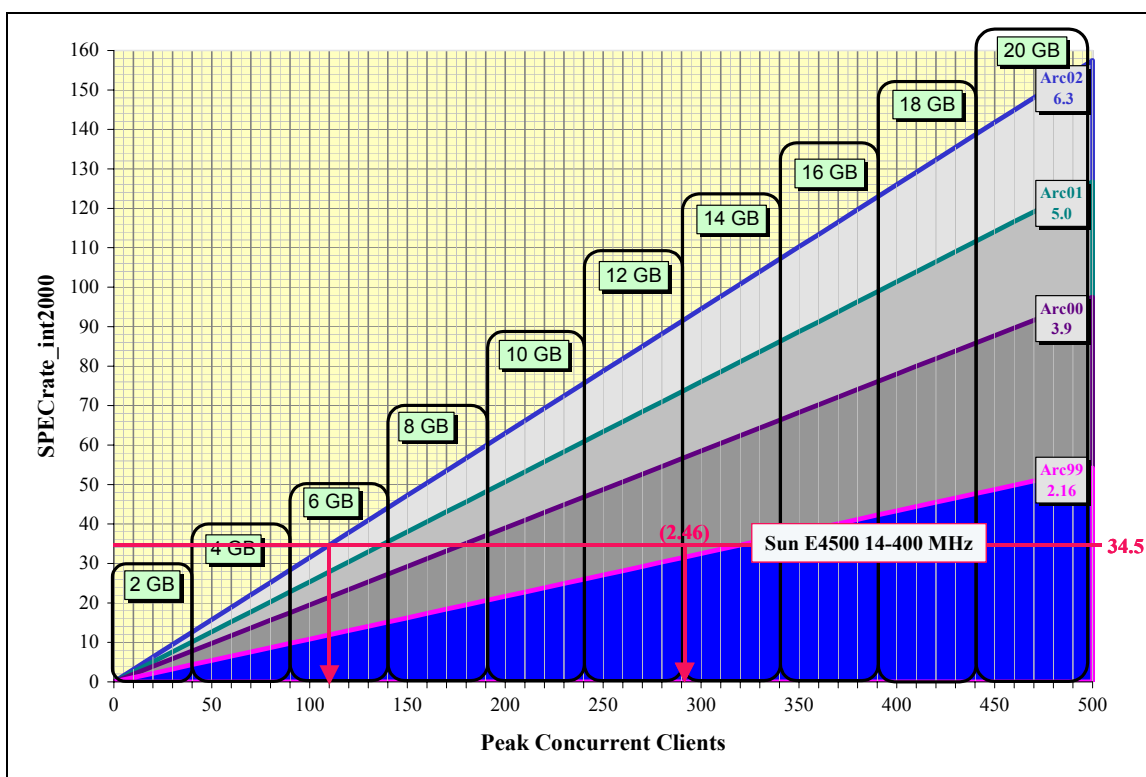
Figure 7-7 provides a summary of Sun Enterprise 3500 and 4500 server configurations with 400-MHz processors. This summary was developed from a search function provided on the SPEC Web site, with the results identified on the chart. This is a listing of the SPECint_rate2000 benchmarks published by Sun Microsystems for these platform configurations.

Figure 7-7
Published SPEC Benchmark Results

SPEC2000 Benchmarks (http://www.specbench.org/osg/cpu2000/)					
Company	System	# CPU	Processors (CPU)		Result
Sun Microsystems	Sun Enterprise 3500/4500	1	400 MHz	UltraSPARC-II	2.46
Sun Microsystems	Sun Enterprise 3500/4500	2	400 MHz	UltraSPARC-II	4.86
Sun Microsystems	Sun Enterprise 3500/4500	4	400 MHz	UltraSPARC-II	9.69
Sun Microsystems	Sun Enterprise 3500/4500	8	400 MHz	UltraSPARC-II	19.5
Sun Microsystems	Sun Enterprise 4500	14	400 MHz	UltraSPARC-II	34.5

The Sun Enterprise 3500 with fourteen 400-MHz processors has a published SPECint_rate2000 benchmark result of 34.5. I will also note on the chart that a single-CPU configuration of this same platform has a benchmark result of 2.46. I can plot the selected platform configuration benchmark value on the sample chart in **Figure 7-8**, and use this chart as a sizing tool to identify how many GIS viewers I can support if I use this platform for an ArcSDE data server.

Figure 7-8
ArcSDE Server Sizing Example



The chart above shows the selected Sun E4500 platform configuration intersecting the Arc02 performance line at 120 concurrent users. I can also see from the chart that the maximum performance provided by this platform (single-CPU performance of 2.46) is slightly faster than the Arc99 performance line (this is an older platform, roughly 3 times slower than current Arc02 technology). For older technology, it is not possible to support current levels of performance. The performance capacity of this platform can be as high as 290 concurrent clients at the Arc99 performance levels, if the server is configured with 14-GB memory.

7.5 Performance Sizing Charts

The rest of this section will provide individual performance sizing charts for each of the ESRI software solutions identified in **Figure 7-9**, and include sample platforms plotted on the chart for reference purposes.

Figure 7-9
Platform Sizing Models

- **Workstations (ArcInfo, GIS Viewers, Terminal Clients)**
- **Application Compute Servers**
- **IMS Map Servers**
- **GIS File Servers**
- **ArcSDE Servers**

Workstations and application compute servers (Windows Terminal Servers) support application processing. GIS application processing is compute-intensive with a preference for floating-point calculations. Application processing is required to support a variety of GIS analysis tasks (examples include spatial analysis, routing, address matching, address geocoding, map product, data maintenance, and data conversion). Application processing is required to render the GIS information product provided on the user display. Application processing supports roughly 50 percent of typical GIS operations. The remaining 50 percent of the resources are required for data query. Application processing and data query is supported by the application layer (user workstation, application server, or IMS map server) when accessing local or remote GIS file systems. Query processing is offloaded to the data server when using an ArcSDE data source.

7.5.1 GIS Desktop Platforms

Figure 7-10 provides recommended configurations for the GIS desktop platforms.

Figure 7-10
ArcInfo and ArcView Platform Recommendations

Application	Platform	Memory	SPECfp2000
ArcGIS 8.1	Intel Xeon 1500+	512	630
ArcView 3 GIS	Pentium III 900+	256	340
Terminal/Browser Clients	Pentium 200+	64	50+

There are two categories of GIS user workstations. A high-performance configuration is required to support Workstation ArcInfo 8 applications and ArcGIS operations. This includes the ArcGIS desktop applications (ArcInfo, ArcEditor, and ArcView 8.1). A standard performance configuration is required to support ArcView 3 GIS query and analysis applications. Standard Windows Office desktop environments can support terminal and browser clients.

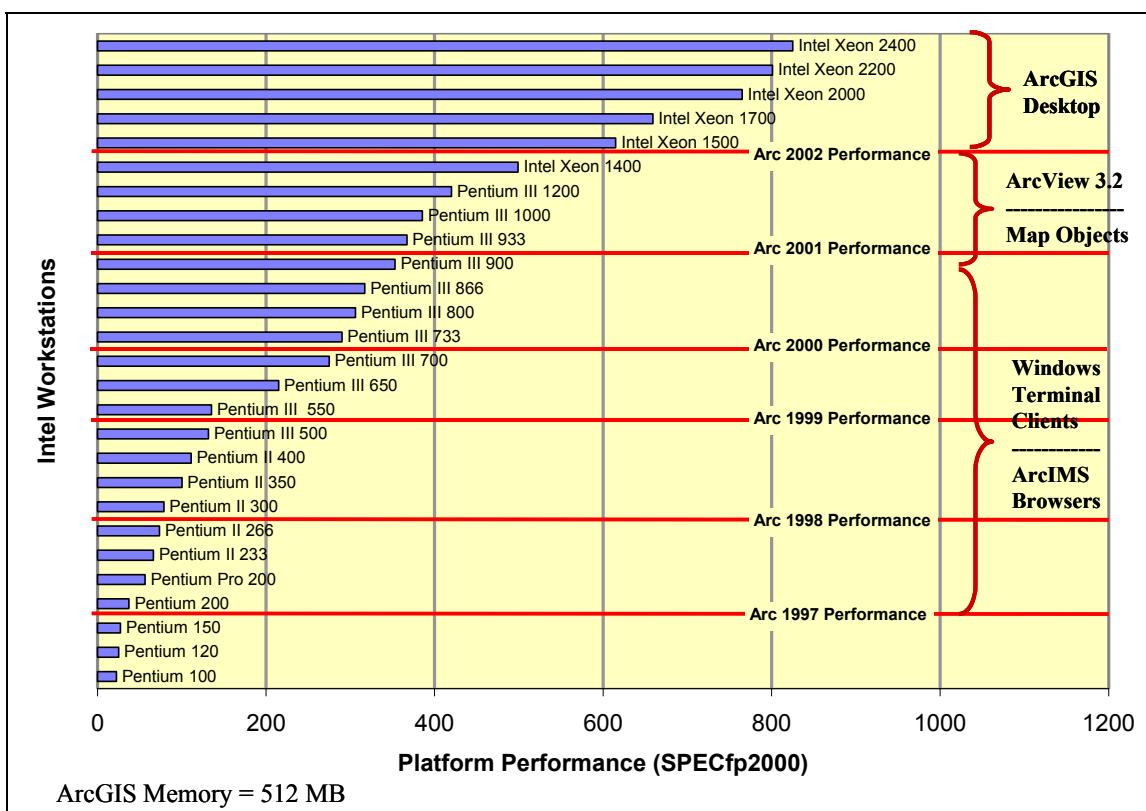
ArcInfo Workstation and ArcGIS Specifications. ArcInfo and ArcGIS desktop platform requirements include a Pentium 4-class processor (current processor is Intel Xeon -1500+) with 512-MB physical memory and 20-GB SCSI hard disk (40-GB disk would provide additional local storage for large project files). Minimum 17-inch display (20-inch preferred for heavy users). Video display card should include 32-MB VRAM supporting minimum of 1280 x 1024

resolution and full color (additional VRAM will provide higher resolution and better display performance for 3-D applications). Platform should include a CD ROM, 1.44-MB floppy disk, and 10/100-Mbps Ethernet controller. Include Windows operating system.

ArcView GIS 3.x Windows Workstation Specifications. ArcView 3.x for Windows platform requirements include a Pentium III-class processor (current processor is Pentium III-900) with 256-MB memory (512-MB preferred) and 20-GB SCSI hard disk (40-GB disk would provide additional local storage for large project files. Minimum 17-inch display (20-inch preferred for heavy user). Video display card should include 32-MB VRAM supporting minimum of 1280 x 1024 resolution and full color (additional VRAM will provide higher resolution and better display performance for 3-D applications). Platform should include a CD ROM, 1.44-MB floppy disk, and Ethernet controller. Include Windows operating system.

Figure 7-11 provides a performance overview for current Windows platform environments. Recommended performance thresholds are identified for ArcInfo workstations, GIS viewing workstations (ArcView GIS and MapObjects applications), and Windows Terminal clients (this would include Web browser clients).

Figure 7-11
Workstation Platform Recommendations



7.5.2 Application Compute Servers

The application compute server supports execution of the primary GIS applications. This platform can be used as a user workstation as well as a compute platform for a number of users with display and control supported by terminal clients.

The total number of concurrent users supported by a specific application compute server platform can be calculated by dividing the published SPECrate_fp2000 performance rating by the associated SPECrate_fp2000 performance baseline platform rating and multiplying by a factor of four. This identifies the number of users who can be supported by each candidate platform.

Once the number of users is identified for each candidate platform, memory requirements can be established based on software specifications. For ArcGIS desktop sessions, physical memory requirements are roughly 512 MB for the first user and 256 MB for each additional user.

Windows Terminal Servers. There are two charts provided to support configuration recommendations for Windows Terminal Servers. **Figure 7-12** provides general guidelines for configuring Microsoft Windows Terminal Servers when accessing a file data source. As a general guideline, application servers with a file data source will maintain performance levels with three to four ArcInfo users for each CPU. Additional users supported by the same application server will result in reduced application performance.

Figure 7-12
Windows Terminal Server Sizing
(File Server Data Source)

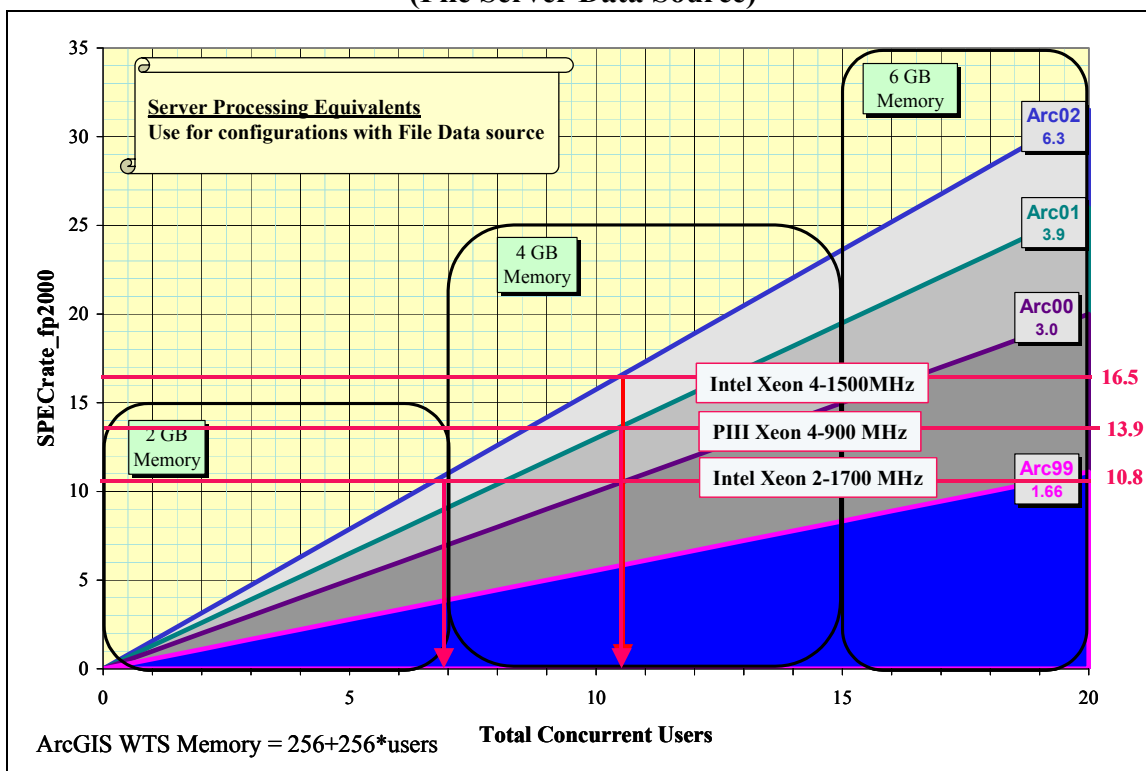
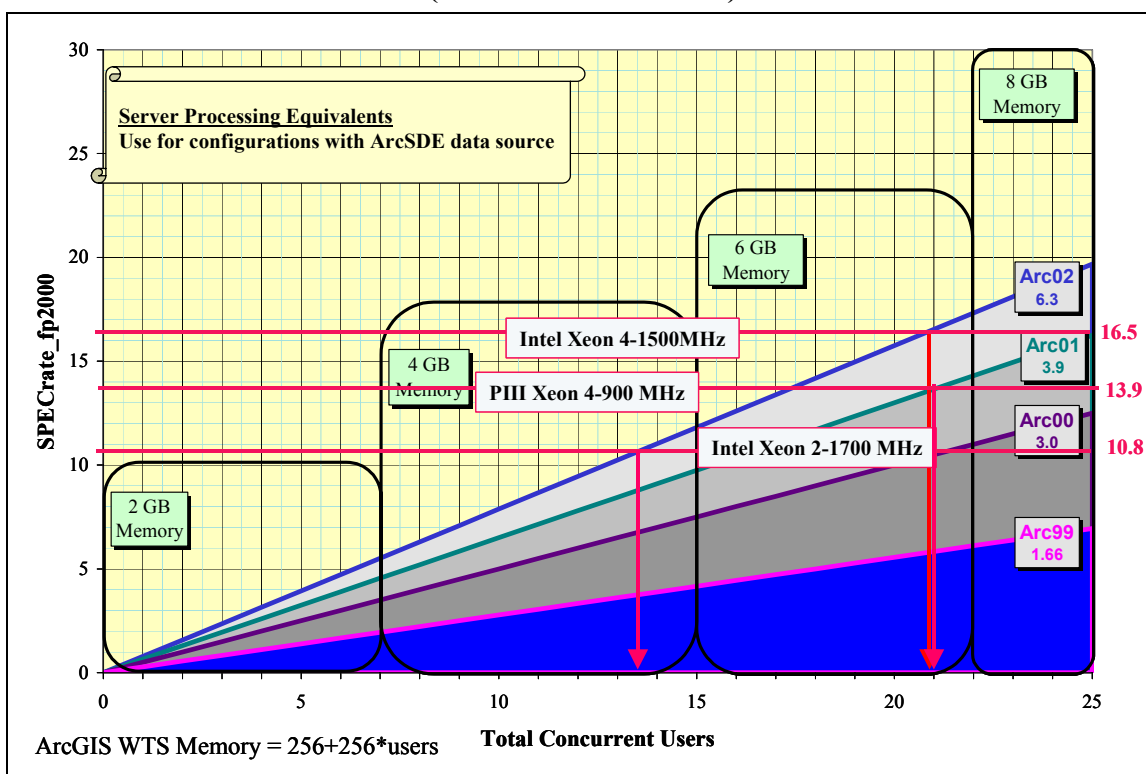


Figure 7-13 provides recommendations for Windows Terminal Servers when accessing a separate ArcSDE data source. Application servers with a separate ArcSDE data source will off-load query processing to the ArcSDE server, and as a result can maintain performance levels with up to seven or eight ArcInfo users for each CPU. Additional users supported by the same application server will result in reduced application performance.

These charts can be used as a quick reference for selecting the appropriate platform CPU and memory requirements. Memory and performance requirements are provided as a function of peak concurrent users. Platform should be configured with the appropriate number of CPUs required to support peak user performance needs, since users will generally be disappointed with a performance loss even if this loss is predicted in the system design.

Figure 7-13
Windows Terminal Server Sizing
(ArcSDE Data Source)



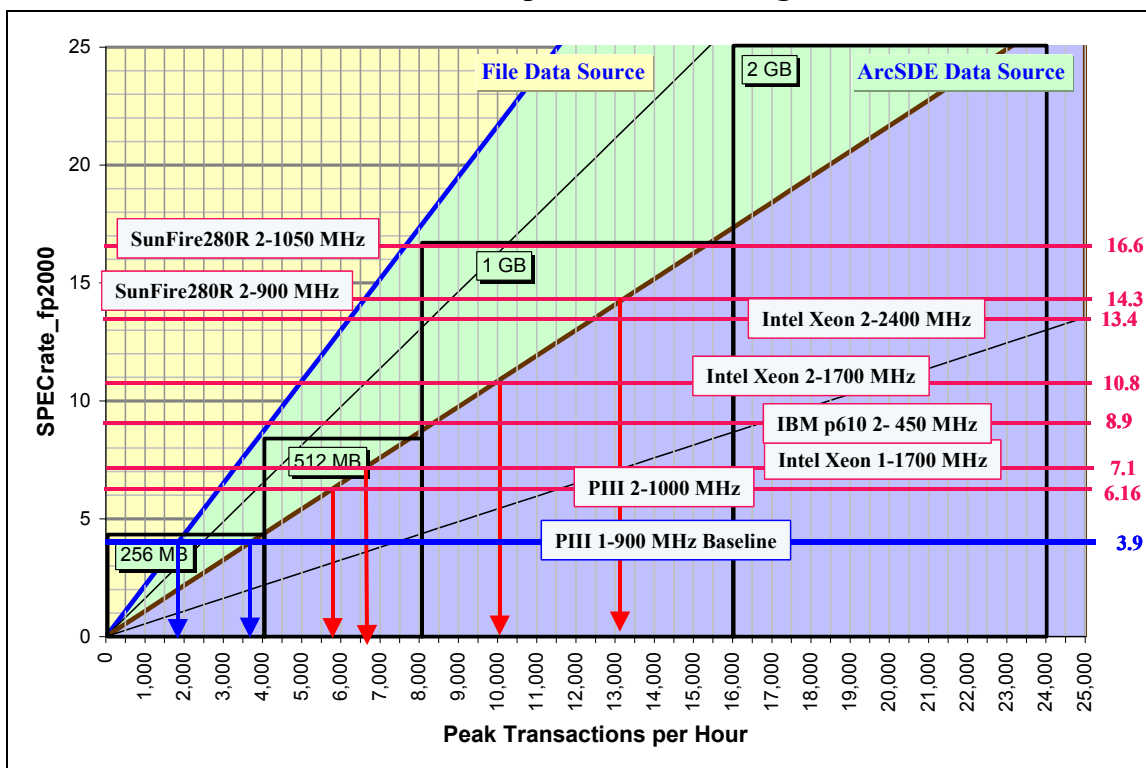
UNIX Application Servers. ESRI performance sizing model for the UNIX application servers is identical to the Windows Terminal Server environment. ArcInfo Workstation and ArcView 3.2 UNIX application servers can use these same charts, with the appropriate UNIX platform, to identify appropriate configuration recommendations.

7.5.3 Internet Web Services

ESRI Web services provide the technology to publish map products to browser clients throughout the internal enterprise environment or over the public Internet. The Web site includes several components that support the Web services. These components can be installed on a single platform, or configured on separate platforms for optimum performance and high availability. A complete discussion of the ArcIMS architecture and sizing models were presented in **Section 6**.

Figure 7-14 provides sizing recommendations for ArcIMS spatial servers (map servers).

Figure 7-14
ArcIMS Spatial Server Sizing



Sample platforms are included on the performance chart to demonstrate current capabilities. Map server platforms are represented by a horizontal line on the chart (based on their SPECrates_fp2000 performance benchmark). The file data source line represents 2-second service rate for the Pentium III 900-MHz baseline platform, and is recommended for use when configuring with a file data source or heavy map service loads. The ArcSDE data source line represents 1-second service rate for the Pentium III 900-MHz baseline platform, and is recommended for use when configuring standard services with an ArcSDE data source or light map service loads. The estimated peak transactions per hour are provided on the horizontal axis, and are identified at the platform intersection with the appropriate data source line.

Map server transaction rates can be estimated for service times not represented on the chart. Average service time is inversely proportional to the peak transaction rate (i.e., a 4-second service time would support half the peak transaction rate of the 2-second service time).

Figure 7-15 identifies ArcIMS data server loads. This chart is required when the source data is located on a separate data server platform. Enter the chart from the left side with the required peak transaction rate (request rate). The equivalent data server load is provided on the horizontal axis, and is identified by the intersection of the peak transaction rate line with line on the chart.

Figure 7-15
ArcIMS Data Server Processing Loads

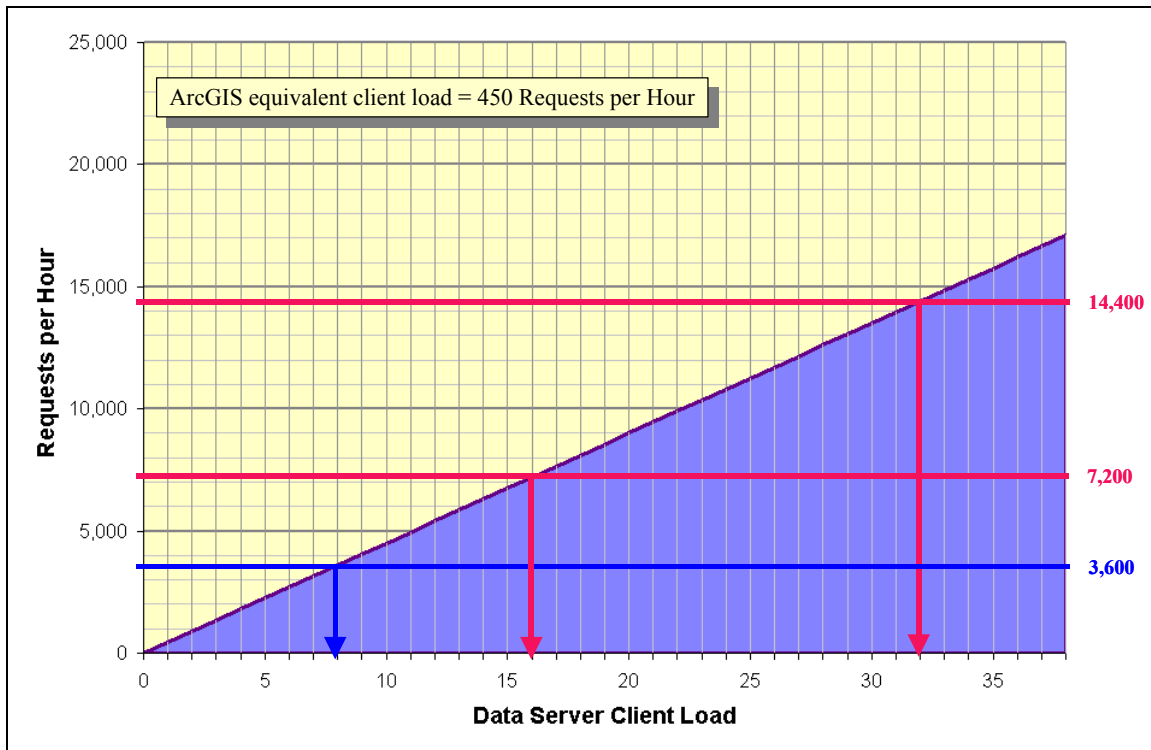
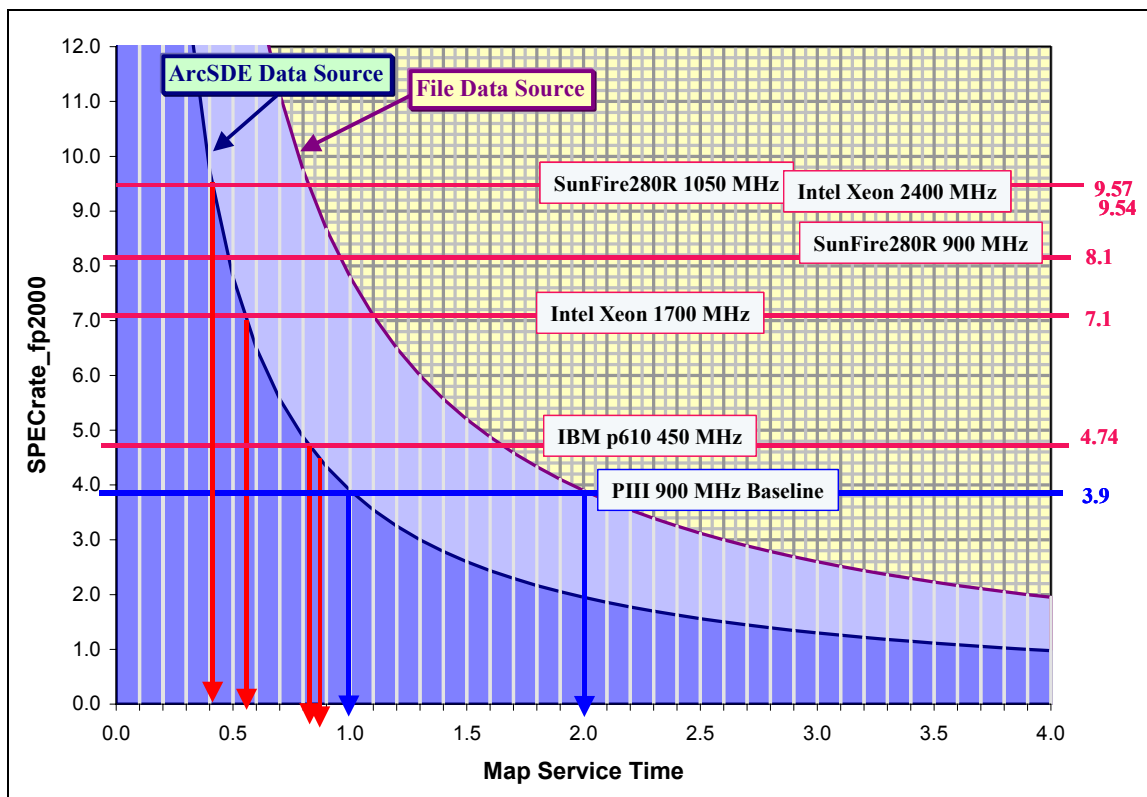


Figure 7-16 can be used to translate map service rates measured on one platform to expected service rates on another platform, based on published relative performance benchmarks. For example, a map service rate of 0.55 seconds on an Intel Xeon 1700 MHz platform would require roughly 1.0 seconds on a Pentium III 900 MHz platform, or 0.4 seconds on a SunFire 280R 1050-MHz platform.

Figure 7-16
ArcIMS Map Service Time



7.5.4 GIS Data Server Sizing (ArcSDE and File Servers)

Data resources typically represent the largest and most valuable asset supporting GIS operations. GIS applications provide tools to support management and analysis of GIS data. The configuration and location of GIS data resources is a primary consideration for any GIS system design.

There are several data formats and storage strategies that can be used to support GIS data resources. Data can be stored in user directories on local workstations or application servers for individual GIS projects or studies. Shared GIS data resources are normally located on a GIS file server or ArcSDE DBMS server. Enterprise data server solutions are normally supported by an ArcSDE DBMS server. Web services can be supported by data replicated on the ArcIMS spatial server or by a central file server or ArcSDE DBMS server.

The data location will impact system performance. Data located on a data server must be transmitted across the network to the application to support processing (minimum of 100-Mbps bandwidth is recommended between the application processing platform and the GIS data source). Larger bandwidth capacity may be required when using a file data source. The GIS application will support query processing when data is stored locally or on a file server. The ArcSDE server will support query processing for data located in the ArcSDE DBMS.

Data server sizing charts are provided for workgroup and enterprise environments. Both of these charts are based on the same sizing model, and can be used for selecting file server or ArcSDE data servers. There is no additional performance penalty for application access to data on local disk (this is supported by the application server and workstation performance charts).

The GIS data server provides an environment for GIS spatial data and associated server processes. The GIS data server must be configured to support the maximum number of concurrent clients. Any additional processes running on the server platform must also be accounted for in the sizing analysis.

The total number of concurrent GIS clients that can be supported by a specific GIS data server platform can be calculated by dividing the published SPECrate_int2000 performance rating by the associated performance baseline platform rating and multiplying by a factor of 20. This identifies the maximum number of users supported by the recommended candidate platform configuration.

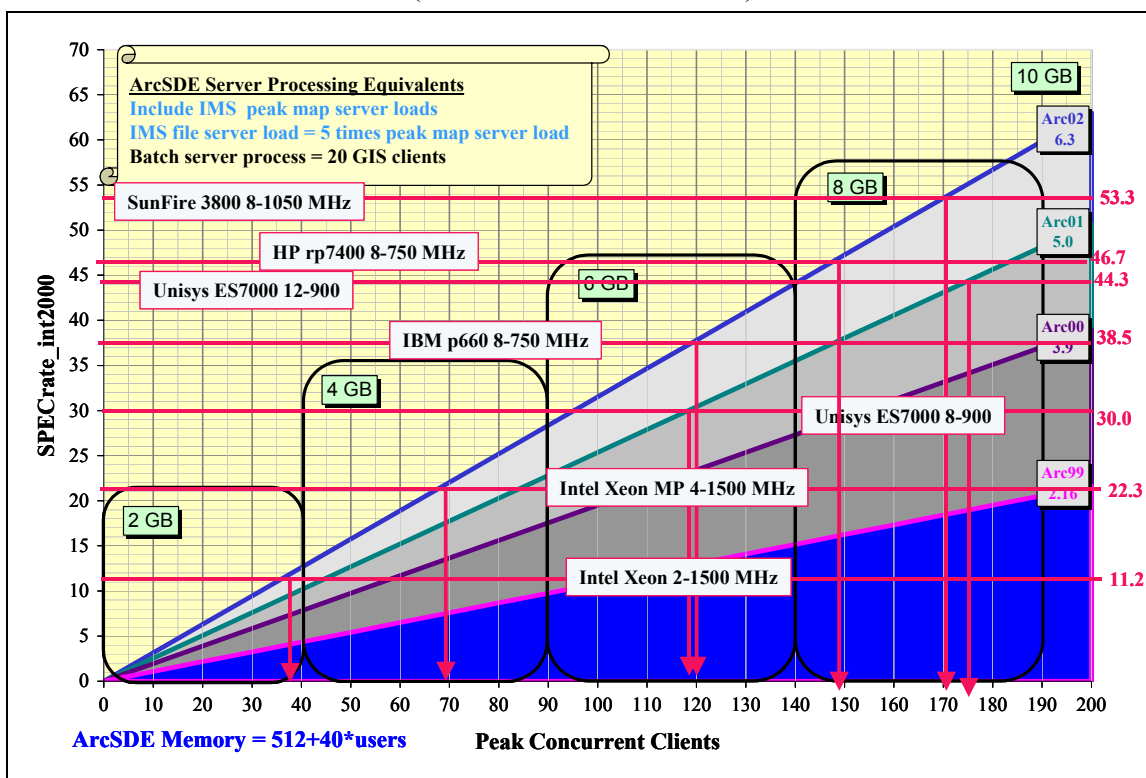
In some configurations, the GIS data server will also support additional application or server processes. These additional processes must be included in determining the total compute load on the server. Additional client application processes (ArcInfo or ArcView GIS) running on the server take roughly the same CPU resources as five data server clients. Any batch processes running on the server will consume a CPU, which is equivalent to the resources required to support 20 data server clients. These adjustments can be made to estimate the total number of equivalent concurrent GIS clients.

Standard spatial data file server solutions may include ArcInfo coverages, shapefiles, ArcInfo LIBRARIAN, and ArcStorm data management solutions. For ArcStorm, an ASuser build process is executed on the server during data check-in and checkout operations. This is a batch process that uses the equivalent CPU resources of 20 data server clients. A reasonable estimate of the concurrent ASuser processes should be made or this may dominate server CPU calculations. ArcStorm check-in and checkout operations are normally a small fraction of the editing process, which implies concurrent checkout sessions would be a small percentage of the total concurrent ArcInfo editor clients. If necessary, check-in or checkout efforts can be completed during off-hours when resources are available. In heavy editing environments, a separate data server may be required for view-only users.

The appropriate memory components of the additional processes identified above must be identified and included in the total calculation. This provides a minimum recommendation for physical memory requirements. The data server should be configured to optimize performance of the client workstations and should not be a performance bottleneck in the configuration. Performance should be monitored during implementation, and memory or CPU resources increased if they are not adequate to meet user performance needs.

Memory and performance guidelines for dedicated GIS data server sizing are displayed in **Figure 7-17**. These charts can be used as a quick reference for sizing GIS data servers.

Figure 7-17
Workgroup Data Server Sizing
(ArcSDE and File Servers)

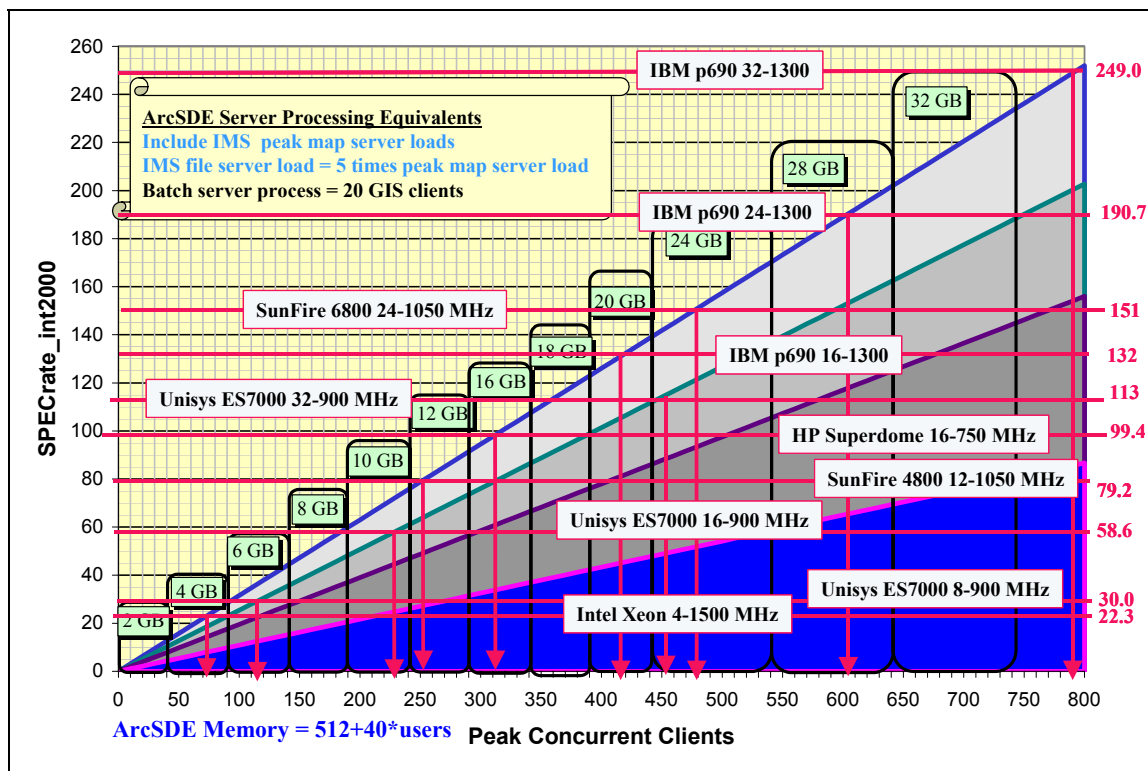


The ArcSDE data server provides client/server data management integrated with standard commercial DBMS product solutions. ArcSDE provides GIS transaction management on the server platform and supports network communication with the GIS client application. The ArcSDE server should be configured to support the peak number of concurrent ArcSDE clients. Any additional processes running on the server platform should also be accounted for in the sizing analysis. This would include any DBMS clients in addition to the GIS users.

The SDE solution is supported on a variety of DBMS platforms including Oracle, Informix (incorporated in the Informix Spatial Datablade), DB2 (incorporated with the DB2 Spatial Extender), and Microsoft SQL Server. Communications between client and server are supported by the ArcSDE application programming interface (API). Both spatial and attribute data are stored in the DBMS. Standard DBMS administration and performance tuning features apply to both the spatial and attribute data.

Memory requirements are identified to support standard DBMS and ArcSDE server processing and should be increased to support planned data caching needs. Increasing memory cache can significantly improve server performance. These charts can be used as a quick reference for sizing GIS data servers.

Figure 7-18
Enterprise Data Server Sizing
 (ArcSDE and File Servers)



8.0 System Implementation

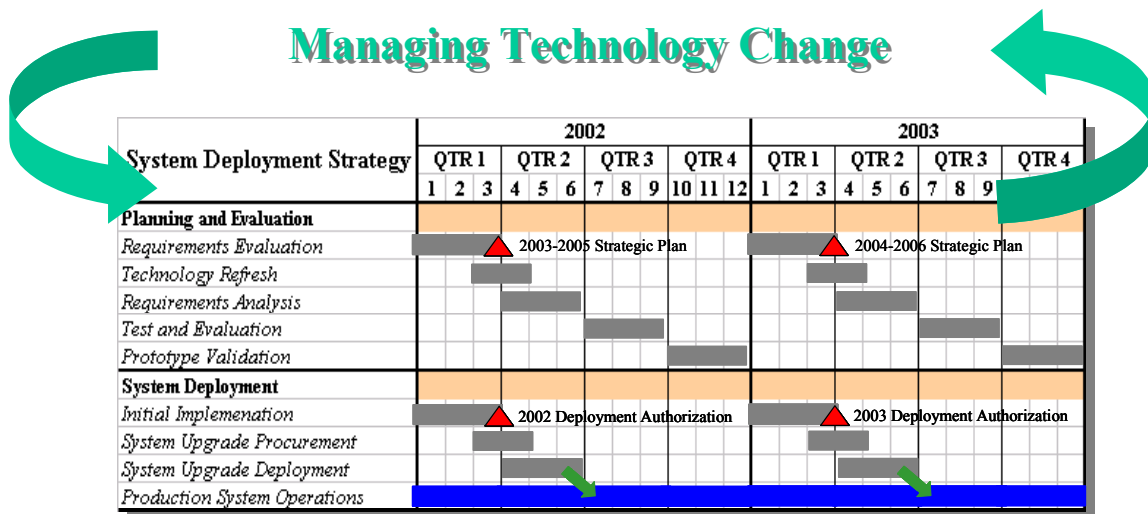
Implementation guidelines are provided to support initial hardware and software selection and establish the opportunity to identify problems early and provide solutions that meet the schedule.

8.1 System Architecture Deployment Strategy

Enterprise GIS operations require a combination of strategic planning and a continued investment in technology. Technology is changing very rapidly, and organizations that fail to manage this change fall behind in productivity and operational cost management. Managing technology change is a major IT challenge.

Enterprise operations should include a periodic cycle of coordinated system deployment. The planning and technology evaluation should occur one periodic cycle ahead of technology deployment, and these efforts should be coordinated to support operational deployment needs. **Figure 8-1** identifies a conceptual system architecture planning and deployment strategy for technology change management.

Figure 8-1
System Architecture Deployment Strategy



- **Planning and Evaluation.** Planning activities should be established on a periodic cycle, coordinated to support the organizations operational and budget planning needs. Strategic plans should be updated to support a multi-year deployment strategy, and published on an established periodic basis (normally on an annual cycle).

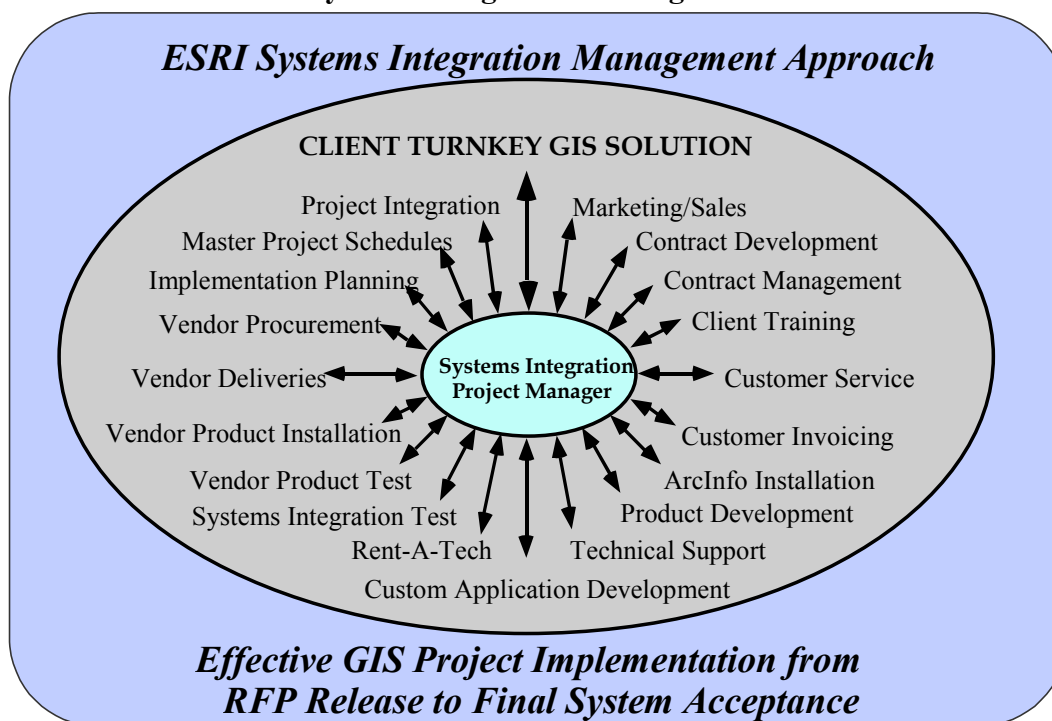
The planning and evaluation process should include a requirements evaluation (strategic plan update), technology refresh (training and research), requirements analysis (process and requirements review), test and evaluation (evaluate new technology alternatives), and prototype validation (pilot test programs). Efforts should be planned on a schedule to support the annual system deployment upgrade cycle.

- **System Deployment.** Operational system upgrades should be planned on a periodic cycle, scheduled to implement validated operational enhancements from the planning and evaluation program. System deployment phases should include initial implementation (implementing changes in an operational test environment) to support deployment authorization. The program should include planned schedules for new technology procurement and deployment on a perioding schedule (normally on an annual cycle). All production system upgrades should be planned and scheduled with full support for on-going operations.

8.2 Systems Integration Management

Basic management guidelines promote success, and experience indicates that this level of supervision is valuable to support successful implementation. **Figure 8-2** represents ESRI's commitment to a total integrated systems management approach.

Figure 8-2
Systems Integration Management



- **Systems Integration Focal Point.** A single management-level focal point should be assigned with responsibility for overall project success. The focal-point person should have authority and responsibility for total project success.
- **Project Team.** A project team should be established with each functional project area represented.
- **Joint Configuration Review.** A review of the configuration should be completed by the project team to identify and resolve implementation problems, allowing time to make

adjustments before installation. This review should be completed prior to any major procurement effort.

- **Implementation Schedule.** Once the project is initiated, delivery and installation schedules should be confirmed. The goal is to identify implementation problems early, avoid unnecessary costs, and support implementation requirements.
- **System Testing.** The installation schedule should support incremental test and acceptance. Install each deliverable on the system, and test it to identify and correct any problems with integration as early as possible. Install ArcInfo last and use it as a tool for final verification that all components of the system function properly.

8.3 Conclusion

Successful implementation depends on a good solid design, appropriate hardware and software product selection, successful systems integration, and careful incremental evaluation during installation. A phased approach to implementation reduces project risk and promotes success, providing the opportunity for early success and flexibility to incorporate new technology at low risk prior to final system delivery. Guidelines are available to support a successful system design, even for large complex systems. Final purchase decisions are influenced by both operational requirements and budget limitations, introducing unique challenges for system design.

Attachment A

System Architecture Design Consulting Services

ESRI provides three levels of system architecture design services to meet specific customer planning needs. Differences between these services are to address the complexity of the design effort. Each of these services result in a complete system architecture deployment strategy to support customer-defined GIS user implementation objectives.

The system design process used by ESRI Systems Integration has been validated by a large number of successful customer implementations over the past 10 years. The ESRI system design methodology is constantly evaluated and updated to provide our customers with the highest return on their planning investment.

A-1: Departmental System Architecture Design Consulting Services

These services are most appropriate for smaller organizations or workgroups that need technical support in establishing a specific system architecture strategy for supporting their GIS deployment needs. GIS users and IT staff should be represented in the system design process. Consultant will work with the customer design team to establish a system architecture strategic plan, with specific hardware platform recommendations and deployment objectives to support GIS deployment objectives. A current user needs assessment and clear understanding of user deployment objectives is a prerequisite for this effort, and will be required as an input to the system architecture design process. Consultant will provide a user needs template to be completed prior to the design effort.

A-2: Enterprise System Architecture Design Consulting Services

These services are most appropriate for larger organizations with several departments or workgroups that need technical support in establishing a specific system architecture strategy for supporting their enterprise GIS deployment needs. All departments should be represented in the system design process. Services include a System Architecture Design for GIS on-site training session and a customer task to consolidate GIS user needs deployment information to support the enterprise design process. User needs templates will be provided following the training class and completed prior to the system design review. Consultant will work with the customer design team to establish a system architecture strategic plan, with specific hardware platform recommendations and deployment objectives to support enterprise GIS deployment objectives. A current user needs assessment and clear understanding of user deployment objectives is a prerequisite for this effort, and will be required as an input to the system architecture design process.

A-3: ArcIMS Architecture Design Consulting Services

These services are most appropriate for organizations that need technical support in establishing a specific system architecture strategy for supporting their ArcIMS deployment needs. Consultant will work with the customer design team to establish a system architecture strategy, with specific hardware platform recommendations and deployment objectives to support ArcIMS deployment objectives. A current business needs assessment and clear understanding of Web services deployment objectives is a prerequisite for this effort, and will be required as an input to the system architecture design process.

A-1: Departmental System Architecture Design Consulting Services

Scope of Effort

The CUSTOMER is seeking recommendations that will identify their current distributed computer hardware and network communication needs and provide recommendations for implementing a department-wide GIS. CUSTOMER should have a current GIS user needs assessment and have a defined GIS implementation strategy as a prerequisite for supporting this system architecture design consulting effort.

The intent of this proposal is to outline the type of support activities that ESRI can provide. The primary goal is to work with CUSTOMER staff to review existing system architecture and identify system design strategies that will include recommendations for supporting planned GIS application needs.

Activity 1.0: System Architecture Design Review

Date: TBD
Duration: Two (2) days
Location: CUSTOMER Site

This activity provides up to 32 hours of consulting support, including preparation and travel time. The purpose of this activity is for the ESRI Systems Integration consultant to visit the CUSTOMER site and meet with appropriate staff to assess the following:

- Current system design and facility layout
- Current software license structure
- Current user application needs
- Future systems design needs
- Hardware procurement requirements
- Implementation support needs

The on-site effort will follow this general format:

General Technology Transfer

Meet with the CUSTOMER GIS project manager and an Information Systems (IS) representative for general introductions and project overview. This will be a working meeting to get acquainted and establish a foundation for systems design and integration program support. Discussions will include a general overview of the current distributed network environment and computer resources.

Meet with GIS representatives to review system design methodology and the GIS user environment. Attendees should include the GIS project manager, the appropriate system administration representatives (hardware and networking), and representatives from each of the user groups as appropriate. This meeting will establish a common foundation to support the system design effort. The following items will be discussed:

- GIS system design methodology (morning session)
- User needs working session (afternoon session)

Specific topics include GIS product architecture and system design methodology, GIS user needs (identify total and concurrent users, quantify application requirements, identify data sources, and evaluate performance needs), and data storage strategies.

System Design Alternatives

Meet with GIS staff to identify components of the computer system environment. Meetings should be scheduled with appropriate staff to review the following distributed system components:

- UNIX and/or NT compute environments
- User desktop environment/workgroups
- Network communications
- Access to legacy systems
- Implementation strategies

Meetings with GIS staff to develop a complete system design approach. Discussions will identify options appropriate for the CUSTOMER GIS system solution consistent with established distributed computer system expansion plans. Specific objectives will be identified for the system architecture design strategy documentation, including actions to address any design issues that are not resolved during these meetings.

Activity 2: System Architecture Design Summary

Date: Following the on-site design review
Duration: Fifty (50) hours
Location: ESRI Headquarters - Redlands, California

This activity provides up to 50 hours of consulting support. The purpose of this activity is for the ESRI Systems Integration consultant to develop a system architecture design document based on the Activity-1 on-site visit to address the following:

- Overview of GIS product architecture
- Overview of user needs, GIS applications, GIS users, computer system design options, recommended computer system design solution, GIS data libraries, licensing requirements, and remote applications.
- System performance sizing, System performance model, SPEC benchmark suites, platform vendor selection, performance baseline selection, network sizing guidelines.
- Specific system design recommendations (immediate and long term)

Draft and final system architecture design documents will be provided to complete this activity. The draft document will be delivered approximately four (4) weeks following the on-site visit.

A-2: Enterprise System Architecture Design Consulting Services

Scope of Effort

The CUSTOMER is seeking recommendations that will identify their current distributed computer hardware and network communication needs and provide recommendations for implementing an enterprise-wide GIS. CUSTOMER should have a current GIS user needs assessment and have a defined GIS implementation strategy as a prerequisite for supporting this system architecture design consulting effort.

The intent of this proposal is to outline the type of support activities that ESRI can provide. The primary goal is to work with CUSTOMER staff to review existing system architecture and identify system design strategies that will include recommendations for supporting planned GIS application needs.

Activity 1: Systems Architecture Design Training

Date: TBD
Duration: Two days
Location: CUSTOMER Site

Two-day on-site training class for up to 20 CUSTOMER staff. Attendees include both systems administrators and key GIS staff chosen to represent the various CUSTOMER departments throughout the enterprise system architecture design process. Training will provide a complete overview of the system architecture design process, identifying work that must be completed by the CUSTOMER to support the system design process.

Activity 2: CUSTOMER User Needs Review and Template Updates

Date: TBD
Duration: Four weeks
Location: CUSTOMER Site

CUSTOMER staff complete the user needs assessment, and fill-out and submit form templates [to ESRI's Systems Integration group] describing their computing environment (i.e., total and peak user loads, user locations, network bandwidth between user sites, existing hardware and infrastructure standards, etc.). Templates will be provided during the class and due prior to the system architecture design review meetings.

Activity 3: System Architecture Design Review

Date: TBD
Duration: Two (2) days
Location: CUSTOMER Site

This activity provides up to 32 hours of consulting support, including preparation and travel time. The purpose of this activity is for the ESRI Systems Integration consultant to visit the CUSTOMER site and meet with appropriate staff to assess the following:

- Current system design and facility layout
- Current software license structure
- Current user application needs
- Future systems design needs
- Hardware procurement requirements
- Implementation support needs

The on-site effort will follow this general format:

General Technology Transfer

Meet with the CUSTOMER GIS project manager and an Information Systems (IS) representative for general introductions and project overview. This will be a working meeting to get acquainted and establish a foundation for systems design and integration program support. Discussions will include a general overview of the current distributed network environment and computer resources.

Meet with GIS representatives to review system design methodology and the GIS user environment. Attendees should include the GIS project manager, the appropriate system administration representatives (hardware and networking), and representatives from each of the user groups as appropriate. This meeting will establish a common foundation to support the system design effort. The following items will be discussed:

- GIS system design methodology (morning session)
- User needs working session (afternoon session)

Specific topics include GIS product architecture and system design methodology, GIS user needs (identify total and concurrent users, quantify application requirements, identify data sources, and evaluate performance needs), and data storage strategies.

System Design Alternatives

Meet with GIS staff to identify components of the computer system environment. Meetings should be scheduled with appropriate staff to review the following distributed system components:

- UNIX and/or NT compute environments
- User desktop environment/workgroups
- Network communications
- Access to legacy systems
- Implementation strategies

Meetings with GIS staff to develop a complete system design approach. Discussions will identify options appropriate for the CUSTOMER GIS system solution consistent with established distributed computer system expansion plans. Specific objectives will be identified for the system architecture design strategy documentation, including actions to address any design issues that are not resolved during these meetings.

Activity 4: System Architecture Design Summary

Date: Following the on-site design review
Duration: Fifty (50) hours
Location: ESRI Headquarters - Redlands, California

This activity provides up to 50 hours of consulting support. The purpose of this activity is for the ESRI Systems Integration consultant to develop a system architecture design document based on the Activity-1 on-site visit to address the following:

- Overview of GIS product architecture
- Overview of user needs, GIS applications, GIS users, computer system design options, recommended computer system design solution, GIS data libraries, licensing requirements, and remote applications.
- System performance sizing, system performance model, SPEC benchmark suites, platform vendor selection, performance baseline selection, network sizing guidelines.
- Specific system design recommendations (immediate and long term)

Draft and final system architecture design documents will be provided to complete this activity. The draft document will be delivered approximately four (4) weeks following the on-site visit. It is anticipated that one ESRI technical designer will support this activity.

Activity 5: System Design On-site Presentation

Date: Following delivery of the system architecture design document
Duration: One day
Location: CUSTOMER Site

The purpose of this task is for the ESRI Systems Integration consultant to present the results of the design study and system architecture design recommendations to select CUSTOMER staff.

A-3: ArcIMS Architecture Design Consulting Services

Scope of Effort

CUSTOMER is seeking recommendations that will review their current computer hardware and network communication needs and provide recommendations for the best utilization of a new ArcIMS system and the associated data storage requirements.

The intent of this proposal is to outline the type of support activities that ESRI can provide. The primary goal is to work with the CUSTOMER staff to develop and implement system design strategies that will meet requirements for an ArcIMS implementation. Services will address ArcIMS platform alternatives and capacity planning, as well as alternative data storage strategies (file or ArcSDE DBMS-based data sources).

Activity 1.0—System Architecture Design Review

Date: TBD
Duration: Two (2) days
Location: CUSTOMER Site

This activity provides up to 32 hours of consulting support, including preparation and travel time. The purpose of this activity is for ESRI to visit the client site and meet with appropriate staff to assess:

- Current system design and facility layout
- Current deployment objectives
- Future hardware needs
- Hardware procurement requirements
- Implementation support needs

The on-site effort will follow this general format:

General Technology Transfer

Meet with the GIS project manager and an information system representative for general introductions and project overview. This will be a working meeting to get acquainted and establish a foundation for system design and integration program objectives. Discussion will include a general overview of the current distributed network environment and computer resources.

Meet with user ArcIMS administrator to review system design methodology and user environments. Attendees should include the GIS project manager, the appropriate system administration representatives (hardware and networking), and the ArcIMS administrator. This meeting will establish a common foundation to support the system design effort. The following items will be discussed:

- ArcIMS technology and configuration strategies
- ArcIMS deployment objectives
- Specific topics include the ArcIMS product architecture and system design methodology
- GIS user needs (identify peak target transaction rates) and data storage strategies
- UNIX and/or Windows NT compute environments
- Network communications
- Implementation strategies

Meetings will address a complete ArcIMS design approach. Specific objectives will be identified for the ArcIMS architecture design documentation including actions to address any design issues that are not resolved during the meetings.

It is anticipated that one ESRI technical designer will support this activity.

Activity 2.0—ArcIMS Architecture Design Document

Date: Following the on-site design review
Duration: Three (3) days (Document delivery two weeks following on-site visit)
Location: ESRI Headquarters - Redlands, California

This activity provides up to 25 hours of consulting support. The purpose of this activity is to develop an ArcIMS architecture design document based on the site visit to address the following:

- Overview of ArcIMS product architecture
- Network bandwidth considerations
- System deployment objectives
- System performance sizing
- Specific system design recommendations (immediate and long term)
- System implementation guidelines

Draft and final system architecture design documents will be provided to complete this activity.

Attachment B

System Architecture Design for GIS Class Description

J-6017

System Architecture Design for GIS

Overview

This two-day course introduces a proven system architecture design methodology for developing successful GIS design and implementation techniques. This methodology was developed and tested during years of successful ESRI system design consulting efforts. The objective of this course is to share this methodology to help users improve performance of existing and future GIS environments. The system design models and configuration guidelines discussed provide class participants with a proven path to successful GIS solutions. Lectures and hands-on exercises help those responsible for GIS system architecture decisions select a system design that will support GIS user performance requirements.

Audience

System Architecture Design for GIS will appeal to those in charge of developing and maintaining hardware or software systems designs and to those in the business of supporting software or application development and technical marketing for system design, testing, and configuration of client solutions. It also provides an excellent conceptual framework for anyone in the position of supporting and securing GIS hardware or software solutions. Senior architecture consultants will benefit from the GIS design methodology presented, while GIS managers will come away with a better understanding of system architecture and hardware selection criteria.

Goals

- Understand the relationships that are available to support successful GIS solutions
- Learn where ESRI software solutions fit in an evolving GIS environment
- Learn how to integrate ESRI software in a distributed enterprise environment
- Learn how to provide high-performance remote user access requirements
- Learn how to integrate GIS applications with legacy data sources
- Understand the prerequisites and recommendations that must be completed before selecting hardware solutions
- Learn how to identify user locations and existing network communications

- Learn how to summarize user requirements to support system architecture design
- Identify system components that contribute to application performance
- Learn how to select an optimum enterprise design solution
- Learn how to identify relative performance of Windows and UNIX platforms
- Apply practical sizing models for selecting central application and data servers
- Learn practical design guidelines for network communications
- Understand the process for conducting a system design review and distributed GIS hardware solution

Topics covered

- System design strategies
- GIS software solutions
- Network communications
- GIS product architecture
- GIS user needs
- System sizing fundamentals
- System sizing tools
- System implementation

Prerequisites and recommendations

Registrants should have an interest in understanding GIS product architecture and how today's computer technologies can support successful GIS solutions. It may also be helpful to review the white paper titled *System Design Strategies* at www.esri.com/library/whitepapers/pdfs/sysdesig.pdf.

Cost: \$800 (Two days)

Classes can be taught onsite at a client's facility (up to 12 participants) for a fixed price of \$5,500.* Special pricing is available to the United States federal government and to qualifying educational institutions, libraries, and museums. Contact the ESRI Learning Center for eligibility requirements. Prices are subject to change without notice.

*For on-site training of fewer than three days, add \$500 to the total charge.

J-6017

Course Outline System Architecture Design for GIS

DAY ONE

System design process

- What is system architecture design
- Why is system design important
- System design process overview
- Supporting technologies overview
- System design support efforts

GIS software solutions

- Overview
- GIS workstations
- Local area networks
- Remote access clients
- Web map products
- Distributed data solutions
- GIS migration strategies
- ArcInfo license management
- GIS high-availability solutions
- *Review questions*

Network communications

- GIS network impact
- Types of networks
- Client/Server communications
- Client/Server performance
- Network configuration guidelines
- Network technology overview
- *Network sizing exercise*

GIS product architecture

- GIS multi-tier architecture
- GIS applications
 - *ArcInfo, ArcView GIS, MapObjects*
- Data management solutions
 - *GIS file servers, ArcStorm, Spatial Database Engine*
- Remote Access Solutions
 - *UNIX application servers, Windows Terminal Servers, Internet Map Servers*
- *Review questions*

GIS user needs

- Total GIS environment
- Application needs assessment
- System architecture review
- User needs assessment

- WAN communications overview
- User application requirements
- Configuration strategies
- Selecting a system solution
- Identifying platform loads
- *Project planning exercise*

DAY TWO

System sizing fundamentals

- System performance profile
- Performance testing
- Client/Server models
 - *Batch processing performance*
 - *Terminal server performance*
 - *Data server performance*
 - *Web services*

System sizing tools

- GIS performance history
- How do we handle change?
- SPEC benchmark suite
- Platform performance charts
 - *GIS workstations*
 - *Application servers*
 - *Internet Map Servers*
 - *File servers*
 - *SDE servers*
- Platform vendor selection
- *Platform sizing exercise*

System implementation

- Phased implementation
- Project management
- Installation schedule
- Lessons learned

Putting it all together

System design exercise

- System design review
- User needs assessment
- System architecture solution
- Platform sizing evaluation
- Platform vendor selection
- Network design specifications

