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

ESRI Systems Integration Technical Brief

Enterprise GIS Storage Technologies – Exploring the Differences Between DAS, SAN, and NAS

Technical Brief #2
May 14, 2003

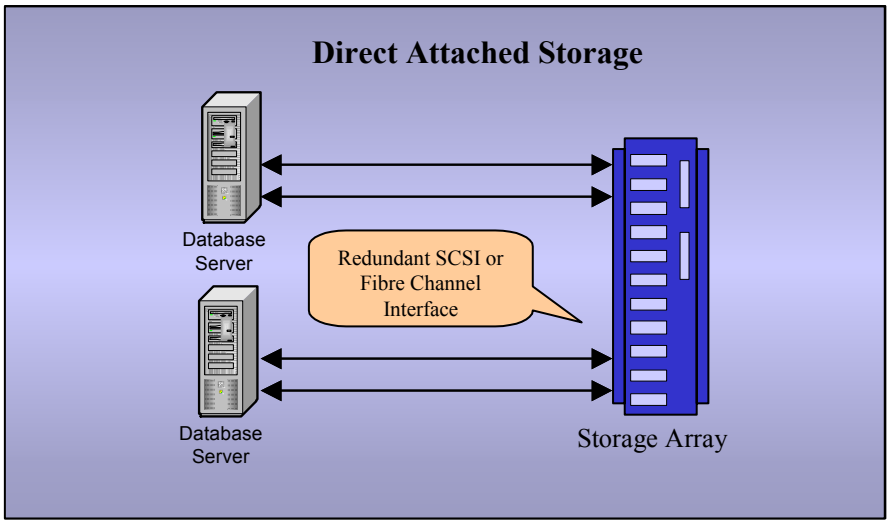
Introduction

Several storage technologies exist to support modern enterprise GIS systems. This technical brief will briefly explain the three primary technologies: Direct Attached Storage (DAS), Storage Area Networks (SAN), and Network Attached Storage (NAS). Internet SCSI (iSCSI) is an up and coming alternative to traditional SAN fabrics and is also discussed.

Direct Attached Storage (DAS)

Direct Attached Storage (DAS) involves the direct connection of servers to a storage subsystem without the use of a back-end storage network. This connection is usually through a SCSI or Fibre Channel interface. Because of cabling and storage design constraints, the number of servers that can connect to the storage is typically limited to four. Therefore, the servers along with the storage array are usually configured as a single rack solution. Multiple host cable paths to the storage array are encouraged for high availability requirements and various RAID solutions are used to protect data at the disk level. Clustering is also supported with shared DAS storage where the clustering software determines which server can see a particular logical disk at any one time. DAS is a fairly straightforward solution for small implementations with few servers. ArcSDE databases with known storage sizes are good candidates for DAS since storage size can be determined fairly accurately thus limiting unused storage capacity. However, as a data center grows, heavy use of DAS among multiple server racks can result in an inefficient use of storage as unused per-server capacity begins to accumulate. Figure 1 provides a picture of a typical DAS configuration.

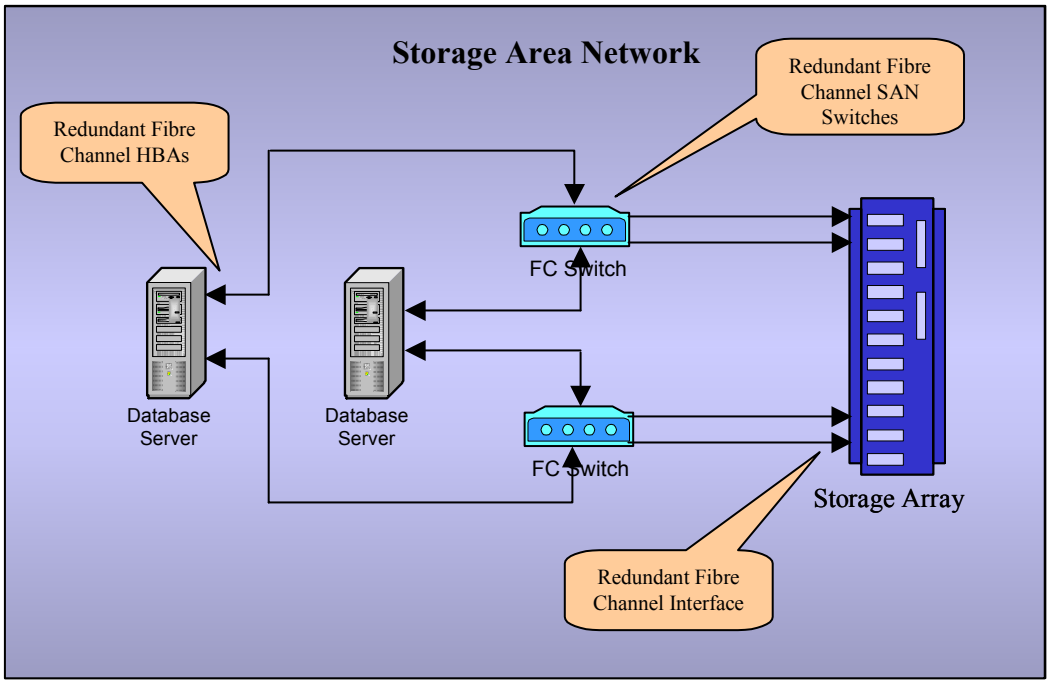
Figure 1
Direct Attached Storage Configuration



Storage Area Networks (SAN)

Storage Area Networks (SAN) are, as the name implies, back-end storage networks that connect multiple hosts through a switched fabric such as Fibre Channel. Fibre Channel is a high-speed fabric that provides up to 200 MB/s of bandwidth in a given Fibre Channel segment or loop. A typical SAN is configured with multiple switches and multiple server Host Bus Adapters (HBA) to create a high availability storage configuration and various RAID solutions are used to protect data at the disk level. Clustering is supported with a SAN in a similar fashion as DAS via shared logical drives. A SAN provides flexibility for "carving" out storage for multiple servers where the servers can be spread out across a data center. Using Fibre Channel and switches, it is possible to configure a SAN with up to 16 million devices with a node distance limitation of 10 KM. The real strength of a SAN is that storage can be assigned and later reassigned as needed to support the changing needs of specific servers. This results in the efficient use of storage and minimizes unused storage capacity for a given server. Initially, building a SAN is more expensive than DAS or NAS and requires expertise with specific hardware and software used to configure the SAN. A SAN should be considered when supporting many servers as part of an overall data center storage design concept. Because a SAN is a back-end storage technology and is distant from the application layer, it can be used to support ArcSDE databases with no integration issues. Figure 2 depicts a typical SAN configuration.

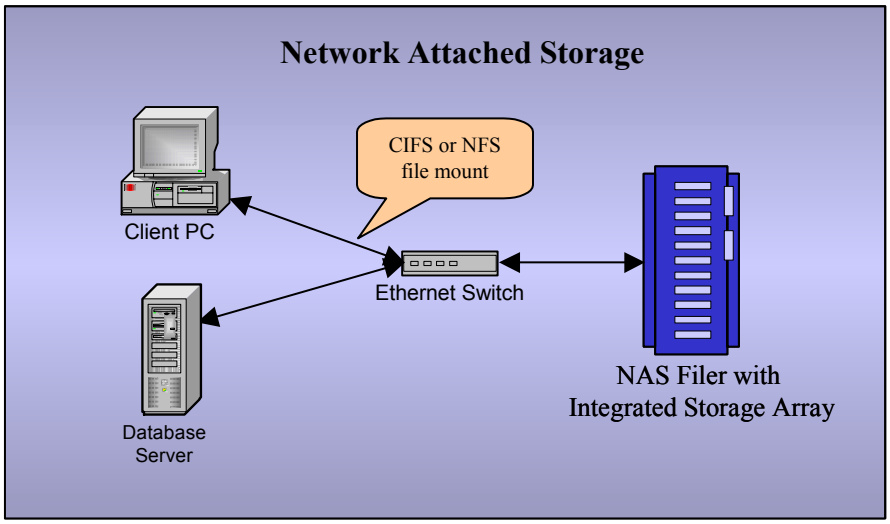
Figure 2
Storage Area Network Configuration



Network Attached Storage (NAS)

Network Attached Storage (NAS) differs considerably from DAS and SAN. A NAS is essentially a file server that has an optimized operating system dedicated to file sharing. One major difference with NAS is that it provides file-level I/O via traditional CIFS and NFS network file shares, while DAS and SAN provide block-level I/O. This is important to consider for performance since there can be performance penalties associated with the use of file-level I/O with a high-performance storage solution. Like DAS and SAN, it does provide disk-level high availability with the use of various RAID solutions. One unique feature of a NAS is that it solves the CIFS/NFS interoperability problem by allowing either Windows or UNIX clients to access its file shares without the need for additional software. NAS devices are easy to implement and offer storage consolidation and file sharing of data over a standard Ethernet network. Since a NAS is a file server, it can be accessed by client systems throughout the enterprise. Though NAS devices are primarily used for flat-file storage, NAS systems are used to support ArcSDE databases as well. However, due to performance considerations with file-level I/O and network protocol overhead, a NAS is generally considered suitable for only small and medium ArcSDE databases. Though cluster solutions usually involve DAS or SAN, a NAS can be used with clustering if specific clustering requirements can be met. For example, some NAS vendors support clustering with the use of software that allows a network drive to appear as a local drive to the database server, which is required for clustering solutions such as Microsoft Cluster Server. Figure 3 depicts a typical NAS configuration.

Figure 3
Network Attached Storage Configuration



Internet SCSI (iSCSI)

SCSI over IP, or Internet SCSI (iSCSI) is an arriving technology that will provide an alternative to traditional SAN fabrics by using standard Ethernet networking hardware to create the back-end storage network while providing high performance with the use of true block-level I/O. Some of the potential benefits of iSCSI as compared to existing SAN fabrics are:

- Provides a much less expensive alternative to traditional SAN fabrics
- Uses standard Ethernet protocols and hardware (1 Gbit now, 10 Gbit future)
- Requires minimal additional expertise to manage
- Is not subject to the 10-KM distance limit of Fibre Channel

Conclusion

Several storage technologies exist to support GIS client/server enterprise requirements. Choosing the correct one depends on several factors including performance, scalability, cost, ease of implementation, and supportability. ESRI software, including ArcSDE, has been successfully deployed in the GIS enterprise using all three aforementioned storage solutions.

Support

Enterprise GIS system design is addressed in the *System Design Strategies* white paper at http://www.esri.com/library/whitepapers/pdfs/sysdesig.pdf. For answers to additional GIS capacity planning and solution questions, contact ESRI Systems Integration at sihelp@esri.com. For technical support, contact ESRI Technical Support at http://support.esri.com.