Networks, System Architecture, and Storage

Nearly all digital image collections will be created and distributed to some extent over networks. A **network** is a series of points or nodes connected by communication paths. In other words, a network is a series of linked computers (and data storage devices) that are able to exchange information or can "talk" to one another, using various languages or protocols such as **TCP/IP** (Transmission Control Protocol/Internet Protocol), **HTTP** (Hypertext Transfer Protocol, used by the World Wide Web), or **FTP** (File Transfer Protocol). The most common relationship between computers, or, more precisely, between computer programs, is the **client/server model**, in which one program—the **client**—makes a service request from another program—the **server**—that fulfills the request. Another model is the **peer-to-peer** (P2P) relationship, in which each party has the same capabilities, and either can initiate a communication session. P2P offers a way for users to share files without the expense of maintaining a centralized server. Music file-sharing has made P2P both popular and controversial at the turn of the twenty-first century, with some copyright owners asserting that the technology facilitates the circumvention of copyright restrictions.

Networks can be characterized in various ways, for instance by the size of the area they cover: local area networks (**LAN**); metropolitan area networks (**MAN**); wide area networks (**WAN**); and the biggest of all, the Internet (from International Network), a worldwide system. They can also be characterized by who is allowed access to them: **intranets** are private networks contained within an enterprise or institution; **extranets** are used to securely share part of an enterprise's information or operations (its intranet) with external users. Devices such as **firewalls** (programs that examine units of data and determine whether to allow them access to the network), user **authentication**, and virtual private networks (**VPN**), which "tunnel" through the public network, are used to keep intranets secure and private.

Another important characteristic of a network is its bandwidth—its capacity to carry data, which is measured in bits per second (bps). Older modem-based systems carry data at only 24 or 56 kilobits per second (Kbps), while newer **broadband** systems can carry exponentially more data over the same time period. One of the problems faced by anyone proposing to deliver digital images (which are more demanding of bandwidth than text, though much less greedy than digital video) to a wide audience is that the pool of users attempting to access these images is sure to have a varying range of bandwidth or connection speeds to the Internet.

Many different network configurations are possible, and each method has its advantages and drawbacks. Image servers might be situated at multiple sites on a network in order to avoid network transmission bottlenecks. A digital image collection might be divided among several servers so that a query goes to a particular server, depending on the desired image. However, splitting a database containing the data and metadata for a collection may require complex routing of queries. Alternatively, redundant copies of the collection could be stored in multiple sites on the network; a query would then go to the nearest or least busy server. However, duplicating a collection is likely to complicate managing changes and updates. Distributed-database technology continues to improve, and technological barriers to such systems are diminishing. Likely demand over the life cycle of a digital image collection will be a factor in deciding upon network configuration, as will the location of users (all in one building or dispersed across a campus, a nation, or throughout the world).

Storage is becoming an increasingly significant component of networks as the amount of digital data generated and stored each day increases almost exponentially. It is often differentiated into three types: **online**, where assets are directly connected to a network or computer; **offline**, where they are stored separately (perhaps as shelved tapes or optical disks such as **CD**- or **DVD-ROMS**) and are not readily accessible; and nearline, where assets are stored offline but are available in a relatively short time frame if requested for online use. Nearline storage systems often use automated "jukebox" systems, where assets stored on media such as optical disks can be retrieved on demand. Other mass-storage options include magnetic tape, which is generally used to create backup copies of data held on hard disk drives, or Redundant Arrays of Independent Disks (**RAID**), which are systems of multiple hard disks, many holding the same information.

Online storage, now known also as **storage networking**, has become a serious issue as the volume of data that is required to be readily accessible increases. The essential challenge of storage networking is to make data readily accessible without impairing network performance. Two approaches to this challenge gaining currency of late are storage area networks (**SAN**) and the less sophisticated network-attached storage (**NAS**). The two are not mutually exclusive: NAS could be either incorporated into or be a step toward a SAN system, where high-speed subnetworks of storage devices are used to hold data, thus unburdening servers and releasing network capacity for other purposes. Higher-end storage systems offer sophisticated file management that includes continuous error checking, failover mirroring across physically separate storage devices, and durable pointers to objects so that they can be stored once but referenced from many locations.

Whatever storage system is employed, because of the ephemeral nature of digital objects, and because no one yet knows the best preservation strategy for them, it is extremely important to keep redundant copies of digital assets on different media—for instance: CD-ROM, magnetic tape, and hard disk—under archival storage conditions and in different locations (see *Long-Term Management and Preservation*).