Data Deluge

Advances in PC-based mass storage devices respond to problems of distribution and storage of large databases

By Michael L. Sena

Remember the early days of personal computing (less than 10 years ago), when developers struggled to restrict their software to use under 64K of random access memory and users labored with a 300K floppy drive as their only file storage device?

Fortunately, PC hardware manufacturers have stretched the limits of their computers since then, increasing clock speeds to 33MHz and adding to file storage options up to 32M of RAM, 100M of internal disk storage, and 2M floppies.

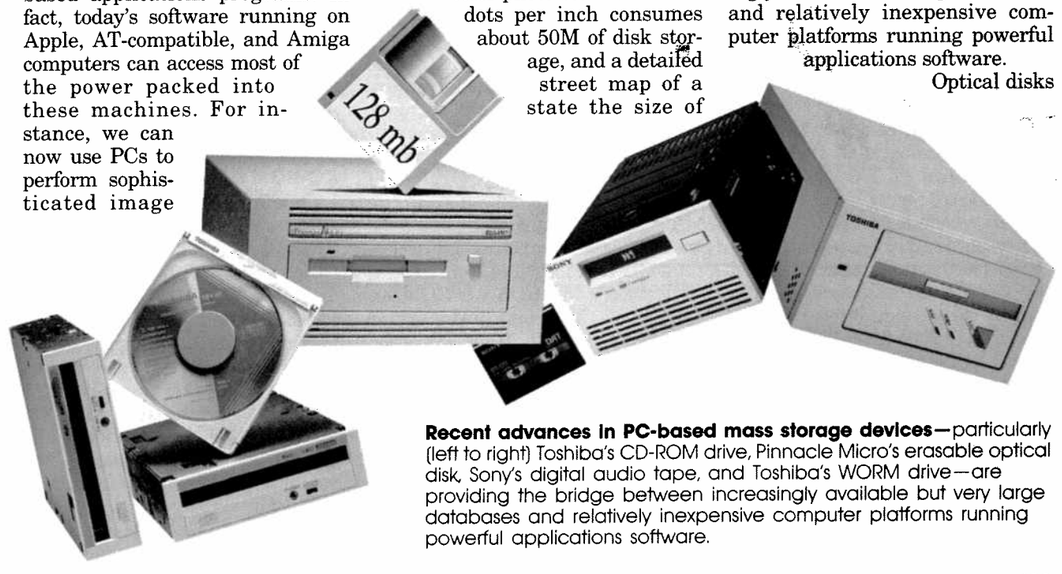
Meanwhile, as PCs have become more powerful, so too have PC-based applications programs. In fact, today's software running on Apple, AT-compatible, and Amiga computers can access most of the power packed into these machines. For instance, we can now use PCs to perform sophisticated image analysis on satellite data or retouch a scanned color photo and separate the photo for four-color printing.

But despite such technological advances in hardware and software, problems do exist that somewhat limit the use of PCs for these and other data-intensive applications: How do we distribute and store on-line the huge amounts of data generated and/or needed to run the software, and how do we back up large files for archival storage? One detailed office building floor plan generated on a CAD system, for instance, might need from 2M to 4M of storage space; a large spreadsheet can easily run up to 500K. But an 8-by-10-inch color photo scanned at 300 dots per inch consumes about 50M of disk storage, and a detailed street map of a state the size of California approaches 1 gigabyte.

Database producers need a way to distribute their data securely, efficiently, and cost-effectively, whether that data is an entire 20-volume encyclopedia or a detailed street map of the US. For such demanding applications, traditional hard drives, floppy disks, and standard magnetic tapes won't do.

Offering a response to the problems of distribution and storage are recent advances in PC-based mass storage devices, particularly CD-ROM (Compact Disk-Read Only Memory), WORM (Write Once, Read Many), and erasable optical disks and digital audio tape. These technologies are providing the bridge between increasingly available but large databases and relatively inexpensive computer platforms running powerful applications software.

Optical disks



Recent advances in PC-based mass storage devices—particularly (left to right) Toshiba's CD-ROM drive, Pinnacle Micro's erasable optical disk, Sony's digital audio tape, and Toshiba's WORM drive—are providing the bridge between increasingly available but very large databases and relatively inexpensive computer platforms running powerful applications software.

come in a wide assortment of sizes and storage capacities, ranging from the standard 4¾-inch disk that can hold up to 660M of data to a 14-inch disk with a capacity of almost 7 gigabytes. An optical disk consists of a glass or plastic substrate coated with a thin metallic film. Data is stored as a sequence of 1s and 0s, just like on magnetic media (such as tape, floppies, and magnetic disks). However, instead of using magnetic fields, data on optical disks is recorded on grooves or pits formed on the metallic coating with precision optics and focusing systems using a beam of light from a low-powered laser. When the disk is read, the laser beam is reflected back from the surface to a detector with light-sensing diodes. Voltage changes caused by the reflected light meeting or not meeting a 1 or a 0 are converted to computer-readable on/off bits.

Promising Technologies

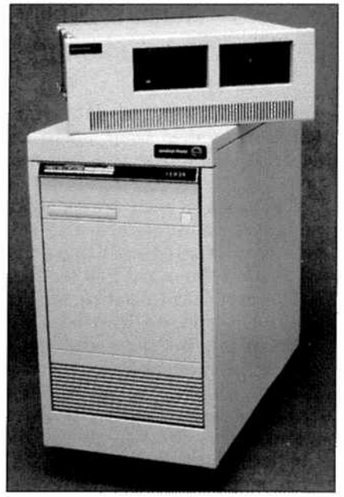
During the first half of the 1980s, companies such as Kodak and 3M promoted the larger 12- to 14-inch optical disks for use in computer-aided retrieval systems, primarily for drawing and document image storage and management functions. Although the larger-format optical disks offer advantages over magnetic media in capacity, stability, and durability, they are too expensive for widespread, non-commercial use.

As such, CD-ROM, WORM, and erasable optical disks appear to be the most promising optical disk technologies for broad-based data storage applications.

Research on compact disk technology dates back to 1976, when N.V. Philips of The Netherlands and Japan's Sony Corporation first began research on the CD format. Then in 1979, Philips introduced the first CD prototype. Collaboration between Philips and Sony resulted three years later in a specification called the "Red Book," which became the standard for the digital audio disk, the now-common CD that has displaced LP records as the preferred medium for distributing and listening to music. The following year, the two companies issued the "Yellow Book" of guidelines for CD-ROM data storage, and in 1985, Digital Equipment Corp. (Maynard, MA) manufactured the first Philips-designed CD-ROM drive.

Contributing editor Michael L. Sena is president of Matrix Consultants (Boston).

CD-ROM drives are similar to CD audio players. The main differences are the absence of computer interfaces and driver software on the CD player, and the lack of analog audio circuitry on the CD-ROM reader (although adapters can turn the CD-ROM drive into an audio player). Another difference is price. CD-ROM drives cost two to three times that of CD players: $600 to $1000 vs. $200 to $300 for portable CD players.



A unique storage solution which transparently integrates optical and magnetic disks is Epoch Systems' Epoch-1 Series of InfiniteStorage servers.

The major standard for accessing CD-ROM devices from DOS is Microsoft's MS-DOS CD-ROM extensions. The original version worked only with High Sierra-format disks, but the recently released version 2.0 can read ISO 9660 format as well. A CD-ROM drive is treated like any other logical disk drive, except that it is read-only.

Small, durable, and stable, CD-ROM solves the problem of distributing large volumes of data to many users. Once a CD-ROM master is made at a cost of around $2000, copies can be mass-produced for about $2 per disk. Because the disks are read-only, there is no danger that data can be unintentionally (or intentionally) erased. Since they are non-magnetic, they cannot be destroyed by machinery or X-ray devices. Unlike fixed disks, they can be removed and mocked up. The combination of high capacity and compact size makes it easy and inexpensive to ship them by mail.

CD-I is the third specification issued by Philips and Sony. Called the "Green Book," it is intended as a consumer product, a self-contained CD-ROM system with a built-in Motorola 68000 series chip and a real-time operating system. CD-I is designed for use in interactive video applications, combining graphics, audio, text, and possibly animation. The computer-in-a-disk is a perfect size for slipping into the dashboard of a car, and one application for CD-I under development at Philips is an in-vehicle navigation system called Carin. The audio capability of CD-I is a necessary component of Carin, since it will use computer-synthesized driver instructions.

Another benefit of adding a CD-ROM reader to your PC configuration is the ability to use the same CD-ROM drive to read your data and access everything from clip art for desktop publishing to information in an encyclopedia. It may also be the preferred future medium for software distribution; Microsoft is reported to be considering using CD-ROM to deliver software and documentation.

CD-ROM Samples

One company that offers CD-ROM subsystem kits is Toshiba (Irvine, CA). The XM-3201A1, which comes with complete plug-and-play CD-ROM attachments for IBM PC/XT/AT/PS/2 and Apple Macintosh computers, provides users with the ability to access 600M of data, or 200,000 pages of text, in seconds. In addition, the XM-3201A1 provides high-fidelity audio output for applications that combine text and audio for instructional purposes.

Headstart Technologies (Great Neck, NY), a division of North

American Philips Corp., also provides CD-ROM products but incorporates them into a complete computer system. For instance, the company offers an MS DOS-compatible PC with built-in, 680M CD-ROM drives. The computer also has a 40M hard drive and is packaged with several CD-ROM databases, including the New Grolier Electronic Encyclopedia, Microsoft Bookshelf, and a national directory of 100,000 business names, addresses, and telephone and fax numbers.

Another data storage option ideal for archival storage is the standard WORM. Unlike with CD-ROM, users can write data to a WORM disk. Like CD-ROM, once data is written, it cannot be erased. Standard WORM drives in a 5¼-inch size from companies such as Storage Dimensions (San Jose, CA), Ricoh (Santa Clara, CA), and Toshiba can store from 600M to over 1 gigabyte of data. Besides higher storage capacities due to larger size, WORM drives have lower access times than CD-ROMs, mostly in the 100 millisecond range, compared to between 450 milliseconds to one second for CD-ROM.

Single WORM drives such as Toshiba's WM-S070 can operate in stand-alone mode, or they can be daisy-chained together for very large archival requirements. Multidisk players, called jukeboxes, are appearing for the smaller-format WORM disks that provide almost limitless capacity. A single jukebox configuration starts in the vicinity of $20,000.

It's unlikely that WORM media and disk drives will become widespread for distribution of databases, unless drives can be developed to read CD-ROM and WORM disks. However, for file backup and storage, they offer the stability, durability, and capacity of an optical disk, with the added advantage of removability for safe storage.

Erasable optical disks, formally known as magneto-optical disks, differ from CD-ROM and WORM in that data can be both written and removed. An ultra-thin magnetic layer covers the platter. On a blank disk, data storage spots are magnetized to face in the same direction. To write data, an infrared laser heats selected data storage spots, and a magnetic coil switches the direction of the heated spots, signifying 1s and 0s. When reading the data, laser light is reflected back from the spots and polarized clockwise or counterclockwise, depending on the direction of the spot, signifying a 1 or a 0. To erase data, the laser heats all data spots on a selected sector of the disk and reorients the spots to the same direction.

Erasable optical disks are precisely what the satellite and orthophoto image analysis market has needed ever since it has been possible to run this application on PCs. Image analysis is both computer memory- and data storage-intensive. MS-DOS memory and disk access barriers were breached several years ago, but the cost of high-capacity fixed devices—600M or more is a minimum for working with a single Landsat scene—has limited the spread of image analysis for mapping.

CD-ROM and WORM offer the stability, durability, and capacity of optical disks, with the added advantage of removability.

Erasable optical disks are about to change this. They combine very large storage capacities, from 256M to 1.8 gigabytes, with removability. They function like an internal hard drive and have access rates in a respectable 30 millisecond range, slower by one-third than standard Winchester fixed drives but 15 to 35 times faster than CD-ROMs. They are being positioned by IBM (Rye Brook, NY) and Next Inc. (Palo Alto, CA) as the future replacements for Winchester disks. Their only problem is that the disks are erasable and therefore lack the security of CD-ROM as a distribution medium.

Several companies provide various product offerings in this category. For instance, Pinnacle Micro (Irvine, CA) recently announced a 3½-inch, 128M erasable optical drive for systems from IBM, Apple, Sun, and Digital Equipment Corp. Called the REO-130, the drive is a SCSI device with a 28 millisecond seek time. Each 3½-inch disk holds 128M of data and reportedly has a shelf life of 10 years. The REO-130 drives are available for around $3000, and media costs $129 per optical disk.

Offering a unique storage solution for today's data-intensive environments is Epoch Systems (Westborough, MA), with its Epoch-1 series of InfiniteStorage servers. The Epoch-1 is designed as a central data storage resource for networks requiring fast, on-line access to high-capacity disk storage over an Ethernet network. It is NFS- and TCP/IP-compatible and uses hierarchical storage techniques to transparently integrate optical and magnetic disks. Systems can be configured with from 1 gigabyte to 1000 gigabytes of online storage; a 20-gigabyte system starts at $82,900.

In addition, Pioneer Communications of America (Upper Saddle River, NJ) recently announced delivery of its Rewritable/WORM Multifunction Disk Drive for both rewritable magneto-optical and WORM media. According to the company, the disk drive allows optical disk system users to select between temporary or permanent storage, and the option to reuse storage space for maximum efficiency. The drive can also write data to and read data from both types of disks, as well as switch modes through commands generated automatically by the host.

Optical disks are not the only mass storage devices that are helping to open new possibilities for end users of large databases. Digital audio tape (DAT) and 8mm tape cartridges provide prodigious data storage capacity in a portable and removable package. DATs store up to 1.3 gigabytes on a cassette the size of a credit card and only 1 cm thick; 8mm cartridges

hold up to 2.2 gigabytes.

There are two competing standards for DAT technology. Hitachi (Allendale, NJ) and Fujitsu (San Jose, CA) developed Data/DAT, which can do random reads and writes to a tape. Hewlett-Packard (Palo Alto, CA) and Sony (Park Ridge, NJ) are promoting the DDS (Digital Data Storage) format. DDS allows for random writes to tape, but not random reads. The ability to update a tape in place is unique among tape formats, since tape is normally sequential. The method used to read and write to DAT is called Helical Scan. Instead of laying data in a linear pattern along the length of the tape, diagonal bands are written across 4mm media.

The major disadvantages of these tape options are the cost of drives and, like other magnetic media, the possibility of accidental erasure. DAT drives cost between $2500 and $6000, while 8mm drives run over $7000.

Which data storage device should you choose for your application? CD-ROM is best for distribution and use of prepackaged databases; WORM offers the promise of permanent backup and secure storage; and erasable optical disk or DAT is best for applications that need large storage capacity for working files. The ideal mass storage device would be one that could work with all three optical disk media and DAT; even better would be a single medium that combines the advantages of each. Neither is likely, given the differences in the technologies.

Growing Incentives

With the cost of CD-ROM drives dropping, with standards finally emerging for CD-ROM data, and with more applications software and databases being delivered on this medium, there are growing incentives for at least adding a CD-ROM drive to your personal computer arsenal. Even a casual PC user can probably find a subject in the growing list of CD-ROM titles that would make an investment in a CD-ROM reader worthwhile.

Consider your application, data levels, and operating environment carefully before selecting one of the other mass storage options, though. It isn't practical to use a WORM drive as you would an internal hard drive. Since a WORM drive cannot be erased, only archival files should be stored. If you're having a problem with backing up a network of PCs or finding a place to store backup tapes and floppies, a WORM drive could be an excellent alternative.

Finally, erasable optical disks are an option—although an expensive one—to a fixed Winchester disk drive. They are practical and cost-effective if your application demands that you require more than 200M of data on a disk at any given time. If you produce many data sets larger than 25M and most of these files can be discarded when a final result is achieved, erasable optical disks will significantly reduce the amount of time working with smaller disks and backing up to tape. *CGW*