DATA RECOVERY BOOK V1.0

FOREWORD

The core of information age is the information technology, while the core of the information technology consists in the information process and storage. Along with the rapid development of the information and the popularization of the personal computer, people find information more and more useful and need it ever more than the past. Owning to the massive data, there is a huge challenge to the data storage technology. So how to save so many documents and how to visit document as fast as possible become the key point. We know we need storage devices to save data, while there are so many kinds of storage devices and modes to save data. What's more, when saving the data and information, it is more important to ensure the storage security as well as the accuracy, usability, reliability of data. Often, what is invaluable is that invisible data.

CONTENTS

Elementary

Ⅰ.Elementary knowledge of data recovery

1.Connotation of data

Connotation of data is comprehensive, it includes not only multi-media files such as data documents, images, voices that stored in file system or data base, but also hardware information, network addresses and network services, which are used to deposit and manage those information.

2.The essence of data recovery

Data recovery means retrieving lost, deleted, unusable or inaccessible data that lost for various reasons.

Data recovery not only restores lost files but also recovers corrupted data.
On the basis of different lost reason, we can adopt different data recovery methods. There are software and hardware reasons that cause data loss, while we can recover data by software and hardware ways.

Being different from prevention and backup, data recovery is the remedial measure. The best way to insure the security of your data is prevention and backup regularly. To operate and use your data according to the normative steps, you can reduce the danger of data loss to the lowest.

3.The scope of data recovery

There are so many forms and phenomenon on data problem, we can divide the objects or scope of data recovery according to different symptoms.

System problem
The main symptom is that you cannot enter the system or the system is abnormal or computer closes down. There are complex reasons for this, thus we need adopt different processing methods. Reasons for this symptom may be the key file of system is lost or corrupted, there is some bad track on hard disk, the hard disk is damaged, MBR or DBR is lost, or the CMOS setting is incorrect and so on.

Bad track of hard disk
There are logic and physical bad track. Logic bad track is mainly caused by incorrect operation, and it can be restored by software. While physical bad track is caused by physical damage, which is real damage, we can restore it by changing the partition or sector. When there is physical bad track, you'd better backup your data for fear that the data can not be used any more because of the bad track.

Partition problem

If partition can not be identified and accessed, or partition is identified as unformatted, partition recovery tools such as Partition Table Doctor can be used to recover data.

Files loss

If files are lost because of deletion, format or Ghost clone error, files restoring tools such as Data Recovery Wizard can be used to recover data.

Password loss

If files, system password, database or account is lost, some special decryption tools that correspond to certain data form such as Word, Winzip can be used.

Files repair

For some reasons, some files can not be accessed or used, or the contents are full of troubled characters, the contents are changed so as they can not be read. In this condition, some special files restoring tools can be tried to restore the files.

4.The principle of data recovery

Data recovery is a process of finding and recovering data, in which there may be some risk, for no all situations can be anticipated or prearranged. It means maybe there will be some unexpected things happen. So you need reduce the danger in data recovery to the lowest:

Backup all the data in your hard disk
Prevent the equipment from being damaged again
Don't write anything to the device on which you want to recover data
Try to get detailed information on how the data lost and the losing process
Backup the data recovered in time.

Ⅱ.Data loss

Actually, there are various reasons that cause data loss; software, hardware, factitious, natural, intended, unintended, all may cause data loss or damage on storage devices.

Generally, There are two main reasons for data problem: software and hardware whose corresponding reasons are software reason and hardware reason.

1.Software reason

Virus, format, mis-partition, mis-clone, mis-operation, network deletion, power-cut during operation all may be the software reasons. The symptoms are usually mis-operation, read error, can not find or open file, report no partition, not formatted, password lost and troubled characters.

A: Computer Viruses: some malicious virus programs will destroy data, overwrite, or erase the data contents.

B: Mis-format: fast or completely format partition, thus changing the file system form (NTFS, FAT32) of partition.

C: Mis-Clone: when backing up the hard disk, mis-clone or overlay the original data on hard disk.

For these, we can use software tools to recover it. So called soft recovery means data can be recovered by software, not referring to hardware fixing operation for its fault is not because of hardware failure.

The following are prompts that system can not start up normally:

Invalid Partition Table: Invalid partition table information.

Missing Operating System: “55AA” mark in DOS boot sector lost or DBR corrupted.
Disk Boot Failure: System file read failure.

Bad or missing command interpreter: Can not find command.com file or ‘COMMAND.COM’ file corrupted.

Invalid system disk: DOS boot record corrupted.

Type the name of the command, Interpreter: DOS partition mark in partition table error or ‘COMMAND.COM’ file lost, corrupted.

Error Loading Operating System: Main boot startup program read boot sector unsuccessfully.
Not found any active partition in HDD: Active partition mark in partition table changed as inactive partition mark.

2.Hardware reason
Sometimes data loss is because of hardware, such as bad sector in hard disk, power cut, head damage, circuit panel problem, etc.

When your hardware has some problems, you probably will find: the speed of hardware become slow, you cannot operate successfully; you cannot read data, etc, which are most often physical bad track failures.

Correspondingly, data recovery in hardware fix is considered as hard recovery, such as memory medium damage, track damage, hard disk scrape, head damage, electric machinery damage, chip burnout and so on..

The most distinct feature or difference between soft recovery and hard recovery is whether the memory medium itself can be normally accessed by fixing or replacing parts.

Ⅲ.Data Protecting Technologies

Data security and fault freedom of storage are paid more and more attention. People are attaching more and more importance to developing new technologies to protect data.

1.SMART Technology

SMART, also called Self-Monitoring Analysis and Report Technology, mainly protects HD from losing data when there is some problems on the HD. SMART drive can reduce the risk of data loss, it alarms to predict and remind thus enhancing the data security.

2.SPS

Shake Protecting System, can prevent the head from shaking thus enhancing the anti-knock characteristics of HD, avoiding damages caused by shake.

3.DFT

DFT, a kind of IBM data protecting technology, can check hard disk via using DFT program to access the DFT micro codes in hard disk. By DFT, users can conveniently check the HD operation.

4.Floppy disk array technology

Originally 'Redundant Arrays of Inexpensive Disks'. A project at the computer science department of the University of California at Berkeley, under the direction of Professor Katz, in conjunction with Professor John Ousterhout and Professor David Patterson.

The project is reaching its culmination with the implementation of a prototype disk array file server with a capacity of 40 GBytes and a sustained bandwidth of 80 MBytes/second. The server is being interfaced to a 1 Gb/s local area network. A new initiative, which is part of the Sequoia 2000 Project, seeks to construct a geographically distributed storage system spanning disk arrays and automated libraries of optical disks and tapes. The project will extend the interleaved storage techniques so successfully applied to disks to tertiary storage devices. A key element of the research will be to develop techniques for managing latency in the I/O and network paths.

The original ('Inexpensive') term referred to the 3.5 and 5.25 inch disks used for the first RAID system but no longer applies.

The following standard RAID specifications exist:

RAID 0 Non-redundant striped array

RAID 1 Mirrored arrays

RAID 2 Parallel array with ECC

RAID 3 Parallel array with parity

RAID 4 Striped array with parity
RAID 5 Striped array with rotating parity
The basic idea of RAID (Redundant Array of Independent Disks) is to combine multiple inexpensive disk drives into an array of disk drives to obtain performance, capacity and reliability that exceeds that of a single large drive. The array of drives appears to the host computer as a single logical drive. The Mean Time Between Failure (MTBF) of the array is equal to the MTBF of an individual drive, divided by the number of drives in the array. Because of this, the MTBF of a non-redundant array (RAID 0) is too low for mission-critical systems. However, disk arrays can be made fault-tolerant by redundantly storing information in various ways.

5.SAN

SAN, called Storage Area Network or Network behind servers, is specialized, high speed network attaching servers and storage devices. A SAN allows "any to any" connection across the network, using interconnect elements such as routers, gateways, hubs and swithes. It eliminates the traditional dedicated connection between a server and storage, and concept that the server effectively "owns and manages" the storage devices. It also eliminates any restriction to amount of data that a server can access, currently limited by the number of storage devices, which can be attached to the individual server. Instead, a SAN introduces the flexibility of networking to enable one server or many heterogeneous servers to share a common storage "utility", which may comprise many storage devices, including disk, tape, and optical storage. And, the storage utility may be located far from the servers which use it.

6.NAS

NAS is Network Attached Storage. It can store the quick-increased information
.Backup means to prepare a spare copy of a file, file system, or other resource for use in the event of failure or loss of the original. This essential precaution is neglected by most new computer users until the first time they experience a disk crash or accidentally delete the only copy of the file they have been working on for the last six months. Ideally the backup copies should be kept at a different site or in a fire safe since, though your hardware may be insured against fire, the data on it is almost certainly neither insured nor easily replaced.

7.Backup

Backup in time may reduce the danger and disaster to the lowest, thus data security can be most ensured. In different situations, there are different ways. Both backing up important data of system with hardware and backing up key information with cloning mirror data to different storage device can work well.

Ⅳ.Elementary knowledge of hard disk

1.History of hard disk development

The hard disk drive has short and fascinating history. In 24 years it evolved from a monstrosity with fifty two-foot diameter disks holding five MBytes (5,000,000 bytes) of data to today's drives measuring 3 /12 inches wide and an inch high (and smaller) holding 400 GBytes (400,000,000,000 bytes/characters). Here, then, is the short history of this marvelous device.

Before the disk drive there were drums... In 1950 Engineering Research Associates of Minneapolis built the first commercial magnetic drum storage unit for the U.S. Navy, the ERA 110. It could store one million bits of data and retrieve a word in 5 thousandths of a second..

In 1956 IBM invented the first computer disk storage system, the 305 RAMAC (Random Access Method of Accounting and Control). This system could store five MBytes. It had fifty, 24-inch diameter disks!

By 1961 IBM had invented the first disk drive with air bearing heads and in 1963 they introduced the removable disk pack drive.

In 1970 the eight inch floppy disk drive was introduced by IBM. My first floppy drives were made by Shugart who was one of the "dirty dozen" who left IBM to start their own companies. In 1981 two Shugart 8 inch floppy drives with enclosure and power supply cost me about $350.00. They were for my second computer. My first computer had no drives at all.

In 1973 IBM shipped the model 3340 Winchester sealed hard disk drive, the predecessor of all current hard disk drives. The 3340 had two spindles each with a capacity of 30 MBytes, and the term "30/30 Winchester" was thus coined.

In 1980, Seagate Technology introduced the first hard disk drive for microcomputers, the ST506. It was a full height (twice as high as most current 5 1/4" drives) 5 1/4" drive, with a stepper motor, and held 5 Mbytes. My first hard disk drive was an ST506. I cannot remember exactly how much it cost, but it plus its enclosure, etc. was well over a thousand dollars. It took me three years to fill the drive. Also, in 1980 Phillips introduced the first optical laser drive. In the early 80's, the first 5 1/4" hard disks with voice coil actuators (more on this later) started shipping in volume, but stepper motor drives continued in production into the early 1990's. In 1981, Sony shipped the first 3 1/2" floppy drives.

In 1983 Rodime made the first 3.5 inch rigid disk drive. The first CD-ROM drives were shipped in 1984, and "Grolier's Electronic Encyclopedia," followed in 1985. The 3 1/2" IDE drive started its existence as a drive on a plug-in expansion board, or "hard card." The hard card included the drive on the controller which, in turn, evolved into Integrated Device Electronics (IDE) hard disk drive, where the controller became incorporated into the printed circuit on the bottom of the hard disk drive. Quantum made the first hard card in 1985.

In 1986 the first 3 /12" hard disks with voice coil actuators were introduced by Conner in volume,

but half (1.6") and full height 5 1/4" drives persisted for several years. In 1988 Conner introduced the first one inch high 3 1/2" hard disk drives. In the same year PrairieTek shipped the first 2 1/2" hard disks.

In 1997 Seagate introduced the first 7,200 RPM, Ultra ATA hard disk drive for desktop computers and in February of this year they introduced the first 15,000 RPM hard disk drive, the Cheetah X15. Milestones for IDE DMA, ATA/33, and ATA/66 drives follow:

1994 DMA, Mode 2 at 16.6 MB/s
1997 Ultra ATA/33 at 33.3 MB/s
1999 Ultra ATA/66 at 66.6 MB/s

6/20/00 IBM triples the capacity of the world's smallest hard disk drive. This drive holds one gigabyte on a disk which is the size of an American quarter. The world's first gigabyte-capacity disk drive, the IBM 3380, introduced in 1980, was the size of a refrigerator, weighed 550 pounds (about 250 kg), and had a price tag of $40,000.

2.Main technical specification and parameter of hard disk

Capacity
We can see the capacity in two aspects: the total capacity and the capacity of one disk. The whole capacity is made up of each disk capacity.
If we increase the disk capacity, we would not only improve the disk capacity, improve the speed of transmission, but also cut the cost down.

Rotate speed.
Rotate speed is the speed disk rotate. It is measured by RPM (Round Per Minute).The rotate speed of IDE hard disk are 5400RPM, 7200RPM etc.

Average Seek Time
The average seek time gives a good measure of the speed of the drive in a multi-user environment where successive read/write request are largely uncorrelated.
Ten ms is common for a hard disk and 200 ms for an eight-speed CD-ROM.

Average Latency
The hard disk platters are spinning around at high speed, and the spin speed is not synchronized to the process that moves the read/write heads to the correct cylinder on a random access on the hard disk. Therefore, at the time that the heads arrive at the correct cylinder, the actual sector that is needed may be anywhere. After the actuator assembly has completed its seek to the correct track, the drive must wait for the correct sector to come around to where the read/write heads are located. This time is called *latency*. Latency is directly related to the spindle speed of the drive and such is influenced solely by the drive's spindle characteristics. This operation page discussing spindle speeds also contains information relevant to latency.

Conceptually, latency is rather simple to understand; it is also easy to calculate. The faster the disk is spinning, the quicker the correct sector will rotate under the heads, and the lower latency will be. Sometimes the sector will be at just the right spot when the seek is completed, and the latency for that access will be close to zero. Sometimes the needed sector will have just passed the head and in this "worst case", a full rotation will be needed before the sector can be read. On average, latency will be half the time it takes for a full rotation of the disk.

Average Access Time

Access time is the metric that represents the composite of all the other specifications reflecting random performance positioning in the hard disk. As such, it is the best figure for assessing overall positioning performance, and you'd expect it to be the specification most used by hard disk manufacturers and enthusiasts alike. Depending on your level of cynicism then, you will either be very surprised or not surprised much at all, to learn that it is rarely even discussed. Ironically, in the world of CD-ROMs and other optical storage it is the figure that is universally used for comparing positioning speed. I am really not sure why this discrepancy exists.

Perhaps the problem is that access time is really a derived figure, comprised of the other positioning performance specifications. The most common definition is:

Access Time = Command Overhead Time + Seek Time + Settle Time + Latency

The speed with which data can be transmitted from one device to another. Data rates are often measured in megabits (million bits) or megabytes (million bytes) per second. These are usually abbreviated as Mbps and MBps, respectively.

Buffer Size（Cache）

A small fast memory holding recently accessed data, designed to speed up subsequent access to the same data. Most often applied to processor-memory access but also used for a local copy of data accessible over a network etc.

When data is read from, or written to, main memory a copy is also saved in the cache, along with the associated main memory address. The cache monitors addresses of subsequent reads to see if the required data is already in the cache. If it is (a cache hit) then it is returned immediately and the main memory read is aborted (or not started). If the data is not cached (a cache miss) then it is fetched from main memory and also saved in the cache.

The cache is built from faster memory chips than main memory so a cache hit takes much less time to complete than a normal memory access. The cache may be located on the same integrated circuit as the CPU, in order to further reduce the access time. In this case it is often known as primary cache since there may be a larger, slower secondary cache outside the CPU chip.

The most important characteristic of a cache is its hit rate - the fraction of all memory accesses which are satisfied from the cache. This in turn depends on the cache design but mostly on its size relative to the main memory. The size is limited by the cost of fast memory chips.