

The FAT Dilemma

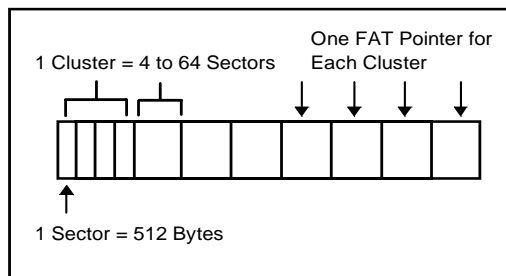
Analysis and Solutions for Disk Storage Efficiency

Now that you've gotten that high capacity hard drive that you've always wanted, you may be wondering what happened to all the space you thought you had. It's not your hard drive. You've got the latest BIOS or BIOS translation utility, so what's wrong? It may be your FAT.

What is a File Allocation Table (FAT)?

DOS, Windows 3.x, and Windows 95 track files stored on the hard disk through directory structures and an index called a file allocation table (FAT). This table contains up to 65,520 entries and keeps track of which disk sectors are bad, in use, or available for assignment. The limitations of FAT and its Windows 95 successor, VFAT, pose a dilemma for the efficient use of storage with the advent of high capacity hard drives. Your hard drive could be losing many megabytes to FAT's built-in file inefficiencies.

Using 16-bit pointers, the FAT allocates file space on your hard disk in clusters. A *cluster* is a group of one or more contiguous sectors and is the smallest unit a file system can allocate. A *sector* is the smallest unit of disk space that the hardware can read or write, and is equal to 512 bytes. These sectors form a volume (an entire hard drive or a partition) that is organized into a file system. See the graphic below.



There are normally 4, 8, 16, 32, or 64 sectors per cluster. Depending on the capacity of the hard drive, the smallest cluster sizes are 2K, the largest 32K. For reasons of compatibility between DOS, Windows 3.x, and Windows 95, the FAT file system has a limit of approximately 65,520 FAT entries. Therefore, as hard drive capacities have doubled in size, so have clusters. The breakpoints in cluster sizes specified by the FDISK command are:

Partition Size*	Cluster Size	Sectors per Cluster
0-127 MB	2K	4
128-255 MB	4K	8
256-511 MB	8K	16
512-1023 MB	16K	32
1024-2047 MB	32K	64

* These numbers are reported in binary megabytes (1,048,576 bytes), not decimal megabytes (1,000,000 bytes). FDISK uses binary megabytes while hard drive capacities are frequently expressed in decimal megabytes.

Some FAT Math

16-bit FAT pointers = 2^{16} = 65,536 pointer values minus those values reserved for special use = 65,520 (the maximum number of locations that can be specified by the FAT).

The 2 GB Barrier

$65,520 \times 64$ sectors per cluster \times 512 bytes per sector = 2,146,959,360 bytes per partition (2.1 GB).

While BIOS upgrades and translating software can break high capacity barriers, they cannot exceed the ultimate 2.1 GB per volume barrier imposed by the DOS/Windows-based FAT system nor the associated breakpoint cluster sizes and sectors per cluster.

Cluster size and sectors per cluster greatly influence overall storage efficiency. Because

operating systems store a file on the hard drive using only whole clusters, hard drive space can be wasted when the file size is considerably smaller than the cluster size. Imagine a 2K file being stored in a 32K cluster—that's 30K of lost space. The larger the cluster size, the greater the likelihood of wasted hard drive space.

Solution: Partition

So what is the solution? Increase the number of partitions? Certainly for hard drive capacities that exceed the 2.1 GB barrier, there is no other choice for DOS/Windows-based systems but to create partitions or virtual drives. The question then becomes—how many? While the FAT is most efficient when drive volumes are formatted for 2K cluster sizes (or 128 MB capacity), this would be impractical for a high capacity drive. The efficiency of large clusters may be the trade-off for fewer but larger virtual disks.

Determine the average length of your work files and the amount of space needed by your applications. You should consider partitioning your 2 GB hard drive into 1 GB partitions, maybe even smaller. Choose a reasonable number of volumes to minimize wasted space and remember these guidelines:

- The smaller the partition size, the smaller the cluster size, therefore, a more efficient use of hard drive space.
- Minimizing the size of clusters maximizes the number of clusters, which can slow file access and increase disk fragmentation.

To Estimate the Amount of Wasted Disk Space on Your Drive

1. Type `CHKDSK C:` to generate a report of hard drive space usage.
2. Take the total number of allocation units and subtract the number of available allocation units.
3. Multiply the result by the number of allocation units used.

The final result is the number of bytes of space allocated.

4. Type `DIR C:\S/A` to determine the cumulative length of all files on disk (except those with special attributes).
5. Take the number of bytes of space allocated result and subtract the cumulative length of files result.

The result is the wasted space due to cluster size.

For service and literature:

714.932.4900 USA
714.932.5000 Outside USA
714.932.4300 DocuFAX
www.wdc.com Internet

Western Digital and Caviar are registered trademarks and WD Caviar, CacheFlow, FIT Lab, and The World's Most Recommended Hard Drive are trademarks of Western Digital Corporation. Other marks may be mentioned herein that belong to other companies. Product specifications subject to change without notice. ©1997 Western Digital Corporation. All rights reserved.

Western Digital Corporation
8105 Irvine Center Drive
Irvine, California 92618

79-880098-000 S0957 4/96