Exploring Windows File Systems

Tom Burt

with material from
Art Tanaka

Advanced Windows SIG

February 21, 2002

Topics

- Master Boot Record Overview
- FAT File System
  - File System Features and Structures
  - FAT Integrity Problems
  - FAT Capacity vs Performance
- New Technology File System
  - File System Features
  - Metadata Files
  - MFT and Allocation Bitmap
- Good Links for More NTFS Information
- FAT 32 vs NTFS Trade-offs
- Tom's Good Hard Disk Practices

Hard Disk Master Boot Record

Absolute Disk Sector 0

Master Boot Record and Partition Table

Up to 4 Partitions

Partition Table

P0 Type	P0 Start	P0 Size
P1 Type	P1 Start	P1 Size
P2 Type	P2 Start	P2 Size
P3 Type	P3 Start	P3 Size

- Each disk drive has a master boot record and partition table
- BIOS reads MBR from designated boot drive (usually Ctrlr 0, D0)
- MBR finds Active, Primary Partition and reads abs sector 0 of that partition as the OS boot record; then executes that boot code.
- Partition’s boot record actually starts up Windows or other OS.
- Every formatted partition has a minimal boot record.

Three Kinds of FAT

FAT 12 (Mainly for Floppy Disks)

- 12 bits or 1-1/2 bytes per entry
- 12 bits, 4085 clusters can be tracked (10 units reserved)
- Approx 2 MB, each cluster equal to 1 sector

FAT 16 (DOS 6, Win 3.1, Win 95)

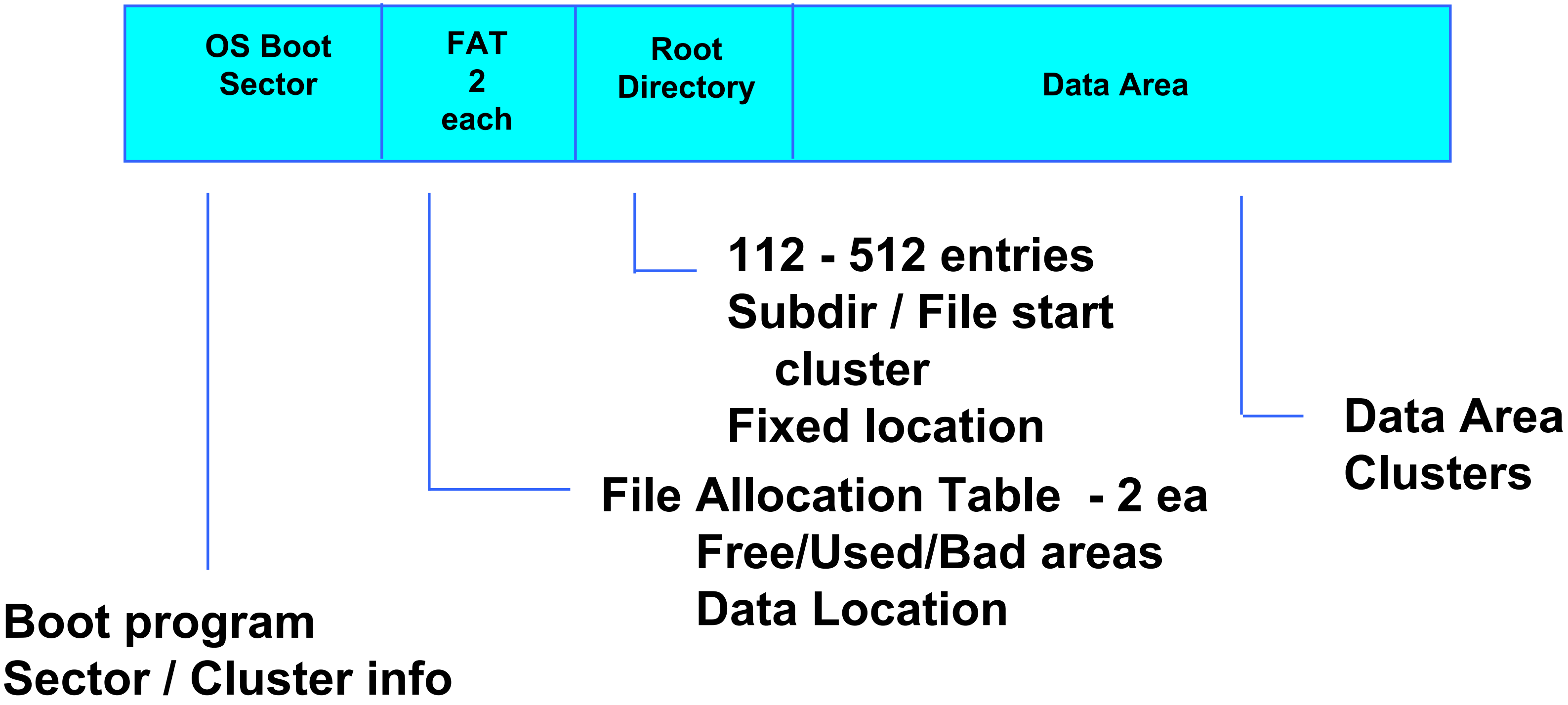
- 16 bits or 2 bytes per entry
- 16 bits, 65,525 clusters can be tracked (10 units reserved)
- Approx 2 GB when each cluster equal to 64 sectors (32K)

FAT 32 (Win 95 OSR2, Win 98, Win ME, Win 2K, Win XP)

- 32 bits or 4 bytes per entry
- 28 bits available, 4 bits reserved
- 28 bits = 2^28 = 268,435,456 clusters can be tracked
- Approx 8 TB (Terabytes) - 28 bits, 64 sector clusters (32K)
- FAT 32 size can now be approx 1 GB! (4 bytes * 2 ^ 28)

FAT12 / 16 Partition Structures

DOS FDISK and FORMAT - partitions, clusters & storage info



OS Boot Sector

FAT 2 each

Root Directory

Data Area

112 - 512 entries
Subdir / File start cluster
Fixed location

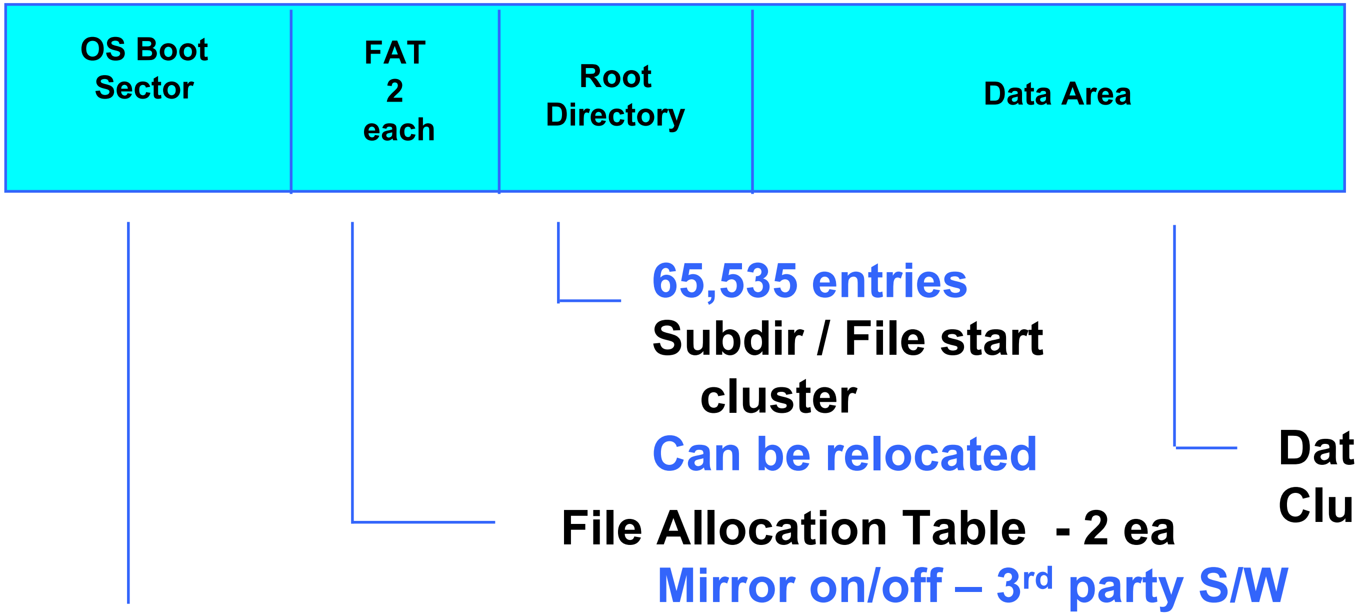
Data Area
Clusters

File Allocation Table - 2 ea
Free/Used/Bad areas
Data Location

Boot program
Sector / Cluster info

FAT 32 Partition Structures

DOS FDISK and FORMAT - partitions, clusters & storage info



OS Boot Sector

FAT 2 each

Root Directory

Data Area

65,535 entries
Subdir / File start cluster
Can be relocated

Data Area Clusters

File Allocation Table - 2 ea
Mirror on/off – 3rd party S/W
Free/Used/Bad areas
Data Location

Boot program
Sector / Cluster info
Root Directory cluster location

FAT 16 and 32 Cluster Sizes

Drive Size	FAT 16 Cluster Size	FAT 32 Cluster Size
256 MB – 511 MB	8 KB	Not Supported
512 MB – 1023 MB	16 KB	4 KB
1024 MB – 2 GB	32 KB	4 KB
2 GB – 8 GB	Not Supported	4 KB
8 GB – 16 GB	Not Supported	8 KB
16 GB – 32 GB	Not Supported	16 KB
> 32 GB	Not Supported	32 KB

Note – Win XP Limits Fat 32 Partition size to 8BG

FAT Structure Contents

There’s one FAT entry for each cluster on drive

Each FAT 12, 16 or 32 bit entry describes one of four states

- Cluster is available for use
- Cluster contains a bad sector, can’t be used
- If the file has more than one cluster, provides next cluster number in the chain
- Cluster is the last one of the file

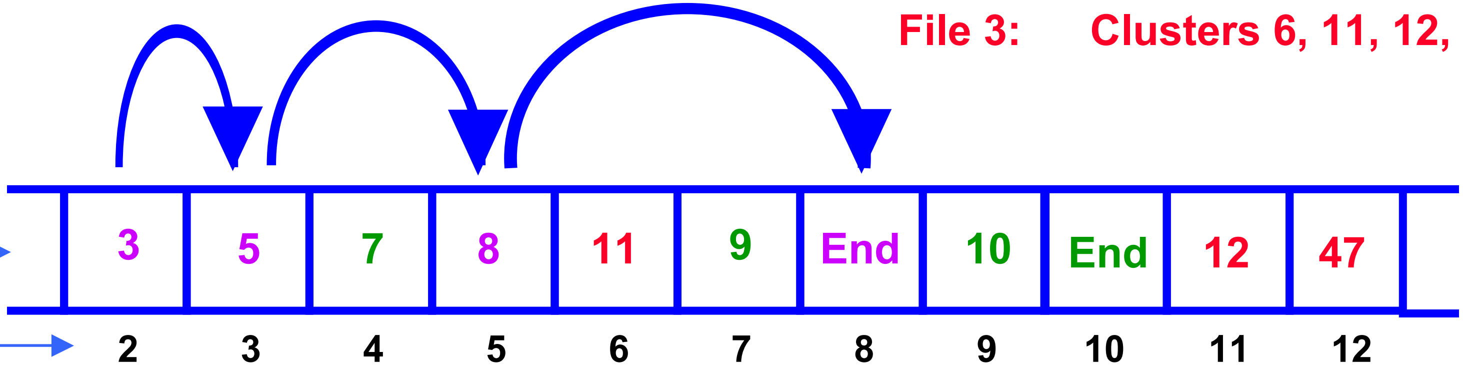
FAT also sets aside reserved space for system

FAT Cluster Chain

FAT Entries

Cluster information
12, 16, or 32 bit values

File 1: Clusters 2, 3, 5, 8
File 2: Clusters 4, 7, 9, 10
File 3: Clusters 6, 11, 12, …



2	3	4	5	6	7	8	9	10	11	12
3	5	7	8	11	9	End	10	End	12	47

FAT Entry Number / Cluster Number

Cluster numbers start at zero
1st two entries reserved
Corresponds to data cluster number

FAT File System Directories

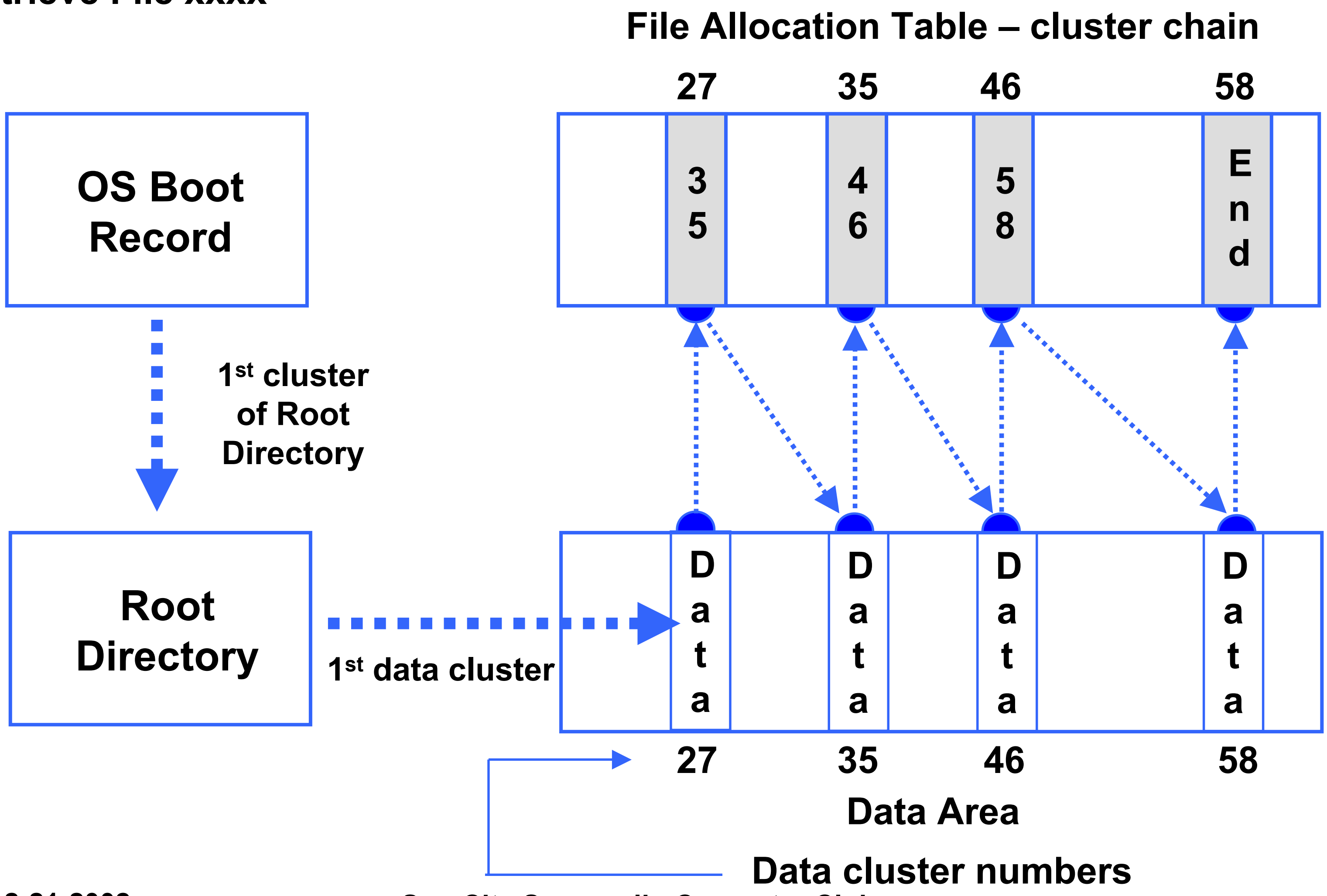
Root directory entries

360 KB / 720 KB Floppy	112
1.2 MB / 1.44 MB Floppy	224
2.88 MB Floppy	448
Hard Disk / DOS-Win95	512
Win95 OSR2 and up	65,535

Each FAT directory is 32 bytes long

Putting it all Together

Retrieve File xxxx



Short and Long File Names

Short File Name (SFN) 8.3 format

- Only upper case A-Z, numerals 0- 9, and $ % ` ! { } ^ # &
- Space possible in DOS name, but DOS programs unable to process
- Entries in ASCII

Long File Name (LFN) Up to 255 Characters

- Mix of lower and upper case
- Spaces in name
- As many periods in the name as desired, last period separates name and extension
- Adds + , ; = [ ] to above SFN symbol list
- Entries in Unicode (2 Bytes / Character)

Long File Names – FAT File System

- Use a sequence of multiple directory entries
- If LFN, 26 bytes (13 Unicode chars) per LFN entry
- Also use an entry for the 8.3 short file name
- Example: "An example of a long filename.doc" (33)
  - Uses: 1 directory entry for the 8.3 file name
  - Uses: 3 directory entries for the long file name
  - Total: 4 directory entries
- In FAT 16 using LFNs in root directory can quickly exhaust available root entries.

FAT Integrity Problems

Orphan Cluster

- Cluster is marked as being used in FAT
- Cluster not part of a chain associated with any directory entry
- Recoverable using ScanDisk or ChkDsk

Cross-linked file

- FAT entry indicates cluster is part of a chain associated with more than one directory entry
- Perverse – deleting one file of cross-linked pair frees the cluster to be reallocated to yet another file. Need to run Scandisk / Chkdsk
- Less common under Win 98, Win ME, Win 2K, Win XP

File size error

- Size entry in Directory is not consistent with size determined from number of clusters in FAT

FAT - Capacity vs Performance

Hard Disk Partition Size Issues

- HD Capacities > 100GB Overwhelm FAT architecture
- FAT gets so large, OS reliability and performance suffer
- Cluster size varies with HD size (see earlier table)
- Large cluster size wastes space with small files

Some FAT 32 Considerations (4K, 8K clusters)

- Better utilization of HD space, gaps efficiently filled
- But - Files more fragmented, reducing performance

Solutions

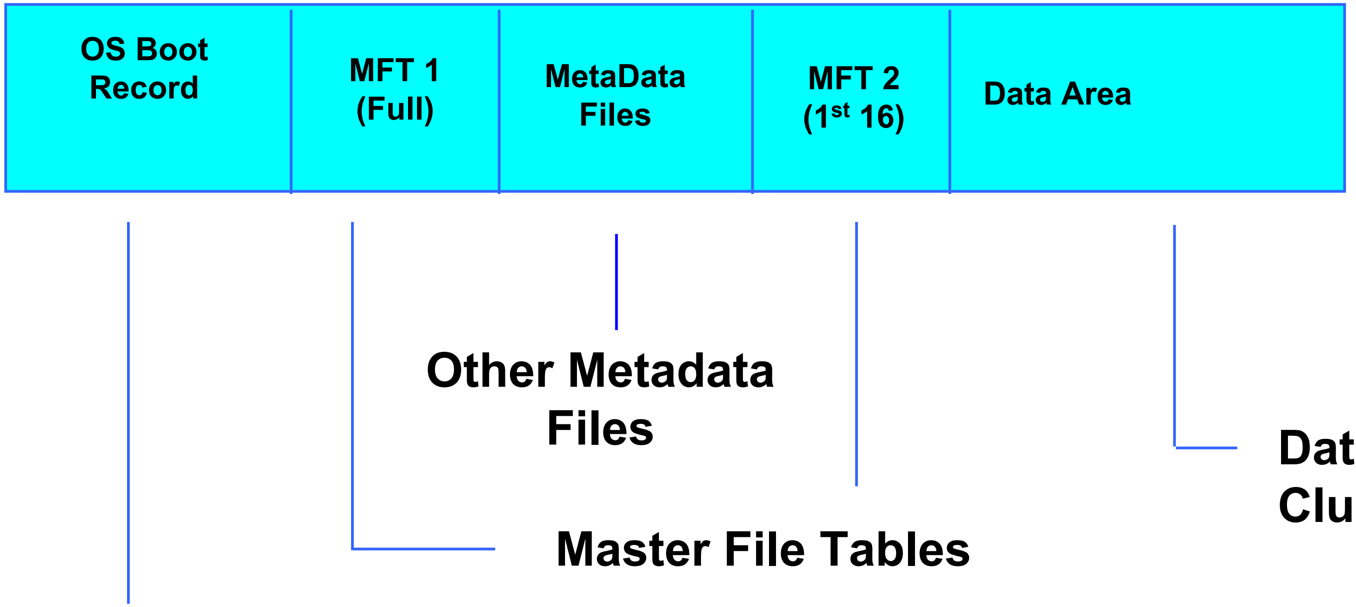
- Partition HD to reduce cluster size, reduce space waste
- Reduce cluster size with Fat 32, reduce space waste
- Provide Defrag tool, improve efficiency

New Technology File System

- Originated with Windows NT
- Enhanced in Windows 2000 and Windows XP (NTFS V3.x)
- Supports User and Group based access controls
- Win 2K and XP versions support file and folder compression
- Win 2K and XP versions support file and folder encryption
- High reliability – all writes to the NTFS partition are first written to the Volume Log File; Chkdsk can recover lost data
- Supports very large hard disk partitions
  - 2^64 (16 billion billion) 4K clusters = 64K Petabytes
- All file names in Unicode (16 bits per character)
- Supports multiple data streams per file (like Mac OS)

NTFS 3.x Partition Structures

Win NT/2K/XP FDISK and FORMAT - partitions, clusters & storage info



OS Boot Record

MFT 1 (Full)

MetaData Files

MFT 2 (1st 16)

Data Area

Other Metadata Files

Master File Tables

Data Area Clusters

Boot program
Sector / Cluster info
MFT 1 / 2 cluster locations

NTFS 3.x Default Cluster Sizes

Partition Size	Default Cluster Size
< 512 MB	512 bytes
512 MB – 1023 MB	1024 bytes
1024 MB – 2047 MB	2048 bytes
>= 2048 MB	4096 bytes

- **4096 bytes (4K) is *Optimum* cluster size (larger allowed)**
- **512 bytes is *Minimum* cluster size**
- **When *Converting* FAT 32 partition to NTFS, Win XP almost always chooses 512 as cluster size, regardless of partition size**

NTFS 3.x Metadata Files (Hidden)

Name	MFT Record	Description
$MFT	0	Master File Table - NTFS's command central
$MFTMIRR	1	Copy of the first 16 records of the MFT
$LOGFILE	2	Transactional logging file
$VOLUME	3	Contains volume serial number, creation time,and dirty flag
$ATTRDEF	4	Attribute definitions
.	5	Root directory of the disk
$BITMAP	6	Contains drive's cluster bit-map (in-use vs. free)
$BOOT	7	Boot record of the drive
$BADCLUS	8	Lists bad clusters on the drive
$QUOTA	9	Contains user quota information (unused before Win 2K / Win XP NTFS)
$UPCASE	10	Maps lowercase characters to their uppercase version

NTFS Master File Table

- MFT uses 1K records for each file or subdirectory (folder)
- MFT is mapped as a file (MFT$) so it can grow and shrink
- MFTMirror$ file is a backup of 1st 16 records of the MFT
- MFT Record has Header, series of Attributes & Data fields
- MFT always has Filename, Security flags and “standard info”
- Small files (< 700 bytes) stored directly in an MFT record
- Larger Files tracked in Extent List
  - Each Extent List element has Starting Cluster No, Number of Clusters
  - Fragmented partition forces lots of extents
- One MFT record can chain to another if file info exceeds 1K
- Directory Records organized as a Binary Tree (fast)

Good Links for NTFS Info

http://www.winternals.com

http://www.storageadmin.com/Articles/Index.cfm?TopicID=135

http://www.win2000mag.com/Articles/Index.cfm?DepartmentID=1

http://www.wd-mag.com/link/subject149.htm?topic=links

http://linux-ntfs.sourceforge.net/regis/

http://www.google.com/ Search: “Windows” “NTFS” “Internals”

Trade-offs NTFS vs FAT 32

FAT 32 – Compatible w Win 98 / ME

- Can boot from Win 9x Emergency floppy to make repairs
- For NTFS on Win 2K / Win XP Must Boot Recovery Console

NTFS has richer security and reliability features

- Individual & Group file ownership
- Encryption
- Journaling allows recovery of written data after [rare] crash

NTFS makes better use of disk space – esp. on large drives

- Smaller clusters
- Bitmap to manage allocation of clusters (32x smaller than FAT32)

Avoid Converting FAT 32 to NTFS – Usually get 512 byte clusters

Tom’s Good Hard Disk Practices

- Keep partition sizes reasonable (8GB to 16GB)
  - Smaller number of clusters
  - smaller system tables
- On main drive:
  - Create a primary partition for OS and Apps
  - Create one or more secondary partitions for data
  - Less likely to lose all data if OS gets corrupted forcing reinstall
- If can afford it, have a second drive:
  - Create one partition to hold an image of the OS partition (Ghost)
  - Create another partition to back up volatile data & for Temp & paging
- Try to format NTFS with 4096 (4K) byte clusters
  - Best tradeoff of space efficiency and performance
- Avoid really long file names
- Defrag Often
- Avoid “dirty” shut downs

Questions and Answers