# UNIT–V Data Analysis

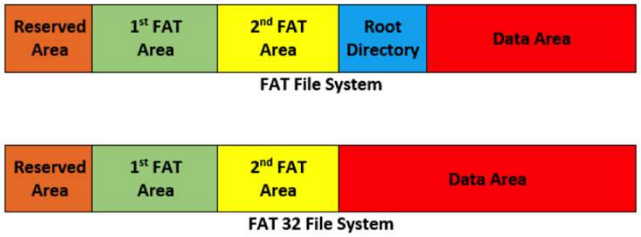
# File System

- A file system in a computer is the manner in which files are named and logically placed for storage and retrieval. It can be considered as a database or index that contains the physical location of every single piece of data on the respective storage device, such as hard disk, CD, DVD or a flash drive. This data is organized in folders, which are called directories. These directories further contain folders and files.
- For storing and retrieving files, file systems make use of metadata, which includes the date the file was created, data modified, file size, and so on. They can also restrict users from accessing a particular file by using encryption or a password.
- Files are stored on a storage media in “sectors”. Unused sectors can be utilized for storing data, typically done in sector groups known as blocks. The file system identifies the file size and position and the sectors that are available for storage.

# FAT/FAT32 File System

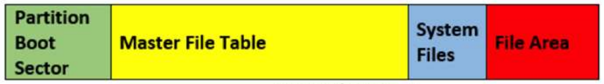
- FAT or File Allocation Table is a file system used by operating systems for locating files on a disk. Due to fragmentation, files may be scattered around and divided into sections.

- FAT32 is an advanced version of the FAT File system and can be used on drives ranging from 512 MB to 2 TB.



# NTFS File System

- •It provides folder and file security. This is done by passing on NTFS permission to files and folders. Security works at local as well as network level. Every file and folder in the list has an Access Control List that includes the users, security identifier, and the access privileges that are granted to the users.
- •Files and partition sizes are larger in NTFS than those of FAT. An NTFS partition can be of a size as large as 16 Exabytes, but practically it is limited to 2TB. File size can range from 4GB to 64 GB.
- •It provides up to 50% file compression
- •It is a reliable and recoverable file system which makes use of transaction logs for updating files and folders automatically.
- •It provides bad-cluster mapping. This means that it can detect bad clusters or erroneous space in the disk, retrieve the data in those clusters, and then store it in another space. To avoid further data storage in those areas, bad clusters are marked for errors.



NTFS File System

# EXT File Systems

- Extended file system (EXT), Second Extended file system (EXT2) and Third Extended file system (EXT3) are designed and implemented on Linux. The EXT is an old file system that was used in pioneer Linux systems. EXT2 is probably one of the most widely used Linux file systems. EXT 3 also includes same features as EXT 2, but also includes journaling.
  Here we will talk about the most commonly used EXT2. With the optimizations in kernel code, it provides robustness along with good performance whilst providing standard and advanced Unix file features.
- **Features**
- Supports standard file types in Unix i.e. regular files, device special files, directories, symbolic links
- Can manage file systems created on huge partitions. Originally, file system size was restricted to 2 GB, but with recent work in VFS layer, this limit has now increased to 4 TB.
- Reserves about 5 percent of blocks for administrator usage, thus allowing the admins to recover from situations of overfilled processes.
- Allows for secure deletion of files. Once data is deleted, the space is overwritten with random data to prevent malicious users from gaining access to the previous data.

# File system forensics process

- **Acquisition**
- The system should be secured to ensure that all data and equipment stays safe. In other words, all media required for forensic analysis should be acquired and kept safe from any unauthorized access. Find out all files on the computer system including encrypted, password-protected, hidden and deleted (but not overwritten) files. These files must be acquired from all storage media that include hard drive and portable media. Once acquired, forensic investigators have to make a copy of them so that the original files are kept intact without the risk of alteration.
- This can be done in four ways:
- Disk-to-Image: This is the most common method as it provides more flexibility and allows to create multiple copies.
- Disk-to-Disk: Used where disk-to-image is not possible.
- Logical: it captures only the files that are of interest to the case. Used when time is limited.
- Sparse: It gathers fragments of deleted or unallocated data.

# File system forensics process

- **Validation and discrimination**
- Before you analyze an image, you need to validate it to ensure the integrity of the data.
- Hashing algorithms help forensic investigators determine whether a forensic image is exact copy of original volume or disk. This validates the integrity of an evidence and conforms to its admissibility into the court.

# File system forensics process

- **Extraction**
- Next comes data extraction, which involves the retrieving of unstructured or deleted data and needs to be processed for forensic investigation. Many computer users think that a file, once deleted, will disappear forever from the hard disk. However, this is not true. Deleting files only removes it from the disc contents table. In FAT systems it is called the File Allocation Table, while in NTFS it is called the Master File Table. Data is stored in clusters on the hard disc and consists of a certain number of bits. Parts of files are mostly scattered throughout the disc, and deleting the files makes it difficult to reconstruct them, but not impossible. With increased disk capacity, it now takes longer for all fragments of a file to be overwritten.
- In many cases, the criminals may have hidden the data that can turn out to be useful for forensic investigation. Criminals with basic technical knowledge have many options available for hiding data such as disk editor, encryption, steganography, and so on. Recovering and reconstructing this data can be time consuming, but generally it produces fruitful evidence.
- Extracting data from unallocated space is file carving. It is a helpful technique in digital forensics that finds deleted or hidden files from the media. A hidden file can lie in any areas such as slack space, unallocated clusters or lost clusters of the digital media or disk. For using file carving, a file should have a header which can be located by performing a search which continues till the file footer is located. Data that lies between these two points is extracted and then analyzed for file validation.

# File system forensics process

- **Reconstruction**
- Extracted data can be reconstructed using a variety of available software tools that are based on various reconstruction algorithms such as bottom-up tree reconstruction and inference of partition geometry. Reconstructed data is thoroughly analyzed for further evidence and put forth in the form of a report.