INT 301

Project: Data Extraction Using Open-Source Software

Name	Jatin Rana
Roll No	19
Section	KE010

Submitted To	Manpreet Singh

Content

1. Introduction

    1.1 Objective of the project

    1.2 Description of the project

    1.3 Scope of the project

2. System Description

    2.1 Target system description

    2.2 Assumptions and Dependencies

    2.3 Functional/Non-Functional Dependencies

    2.4 Data set used in support of your project

3. Analysis Report

    3.1 System snapshots and full analysis report

4. Reference/ Bibliograph

Introduction:

Data extraction using open-source software refers to the process of retrieving and extracting data from various sources using freely available open-source tools or libraries. This costeffective and flexible option allows users to access, modify, and customize the software's source code according to their specific needs.

In this project, Testdisk for Linux and FreeRecover for Windows are the chosen open-source software for data extraction.

Objective of the Project:

The main objective of data extraction is to retrieve relevant and accurate data from various sources, convert it into a usable format, and prepare it for further analysis, processing, or storage. The key objectives of data extraction include:

1. Data Retrieval: The project aims to retrieve data from different sources such as databases, websites, files, or APIs. This involves identifying and accessing the relevant data sources, extracting the required data elements, and capturing the data in a standardized format.

2. Data Transformation: Data extracted from various sources may not always be in a format that is directly usable. Data extraction aims to transform the retrieved data into a standardized format that can be easily processed and analyzed. This may involve cleaning, structuring, and validating the data to ensure its accuracy, integrity, and consistency.

Description of the Project:

The data extraction project starts with identifying the relevant data sources, which may include databases, websites, files, APIs, or other data repositories. The project team determines the data elements or attributes that need to be extracted based on the project objectives, requirements, and desired outcomes. Data retrieval techniques are then used to extract the identified data elements from the sources, and the extracted data is stored in a temporary or permanent location for further processing. Once the data has been retrieved, data transformation techniques are applied to convert the data into a standardized format that is compatible with the target system or analysis tools. This may involve cleaning, structuring, validating, and enriching the data to ensure its accuracy, integrity, and consistency. Data integration techniques may also be applied to merge data from multiple sources into a unified dataset, resolving data inconsistencies or discrepancies.

Scope of the Project:

Data Sources: The project scope includes identifying and accessing the relevant data sources, which could include databases, websites, files, APIs, or other data repositories. The scope may also involve evaluating the complexity, diversity, and availability of data sources, and determining the feasibility of data extraction from these sources.

Data Elements: The project scope includes determining the specific data elements or attributes that need to be extracted from the data sources. This may involve defining the data requirements, data mapping, and data modelling, based on the project objectives, requirements, and desired outcomes.

In summary, the scope of this data extraction project includes identifying and accessing relevant data sources and extracting specific data elements from these sources using opensource software. The project also includes data transformation techniques to convert the extracted data into a standardized format for further analysis and processing.

Target System Description

1.

Operating system- Windows 11 64 bit

Storage – 512Gb SSD

Processor – Core i5 9500

Ram- 16Gb

Files Recovered From – Local Disk D

2.

Operating system- Red Hat Linux 64bit

Storage – 30Gb Virtual Disk

Ram- 3Gb Files Recovered From – Local Disk D

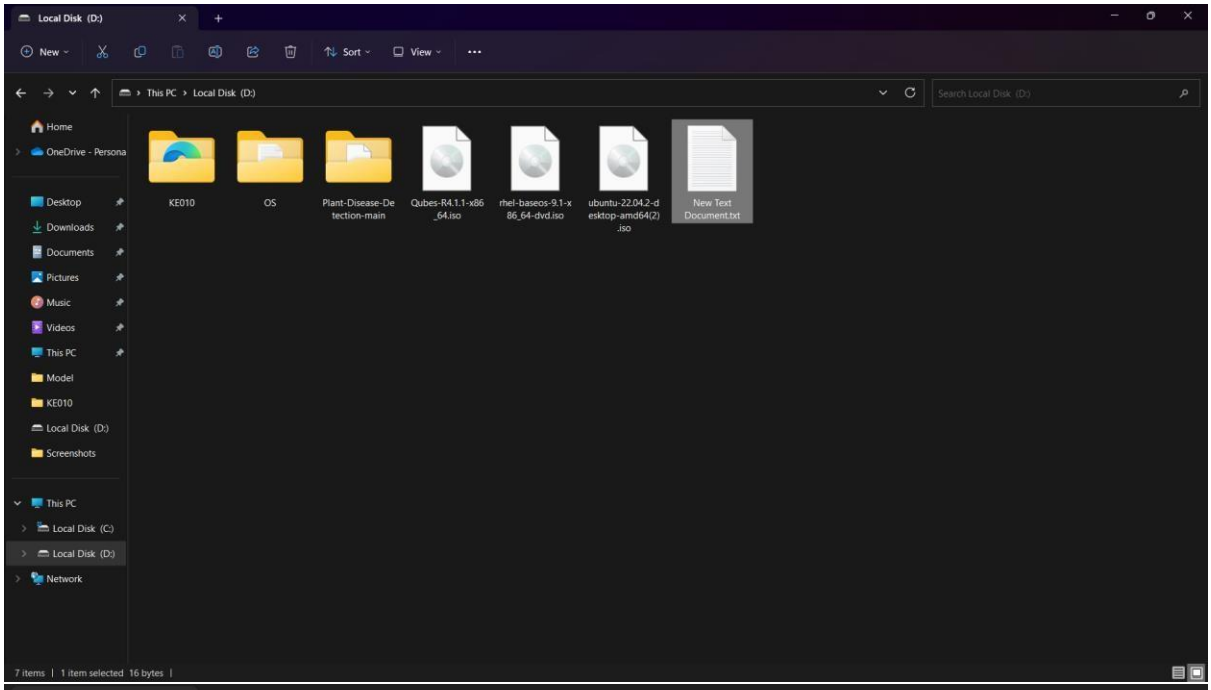
Some brief steps for Data extraction in Linux in Test Disk

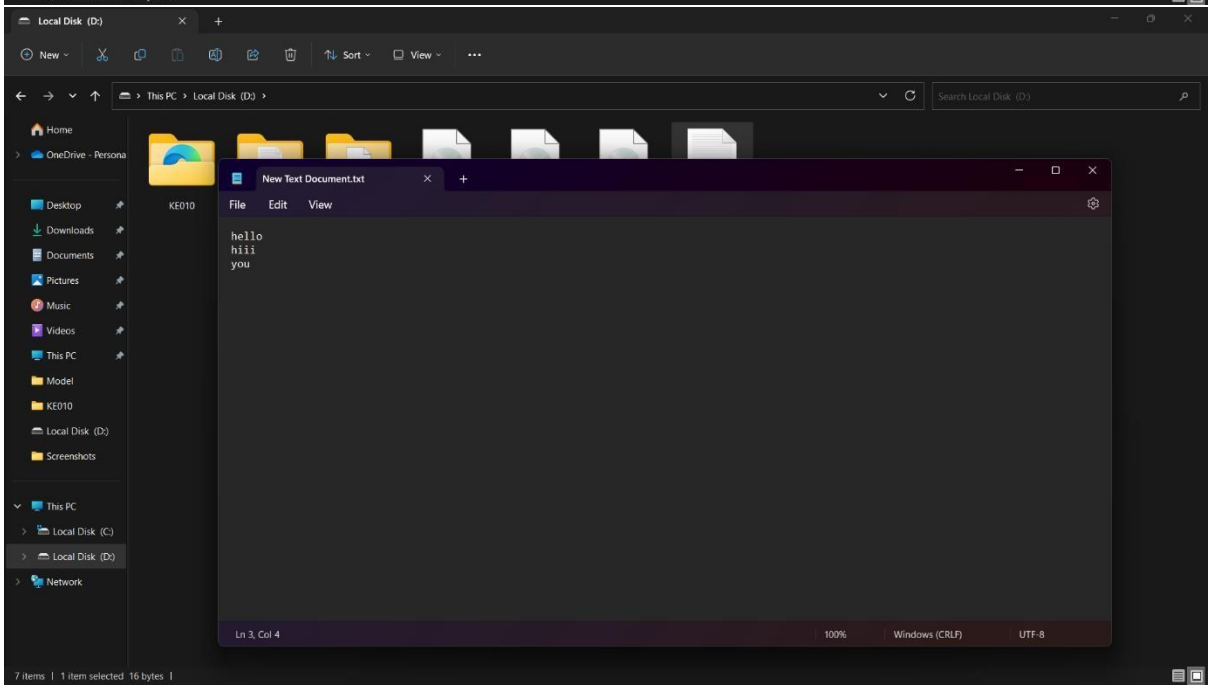
1. Log creation.
2. Disk selection.
3. Partition table type selection. 4. Current partition table status.
5. Quick Search for partitions.
6. Save the partition table or search for more partitions.
7. Partition table recovery.
8. NTFS Boot sector recovery.
9. Recover deleted files.

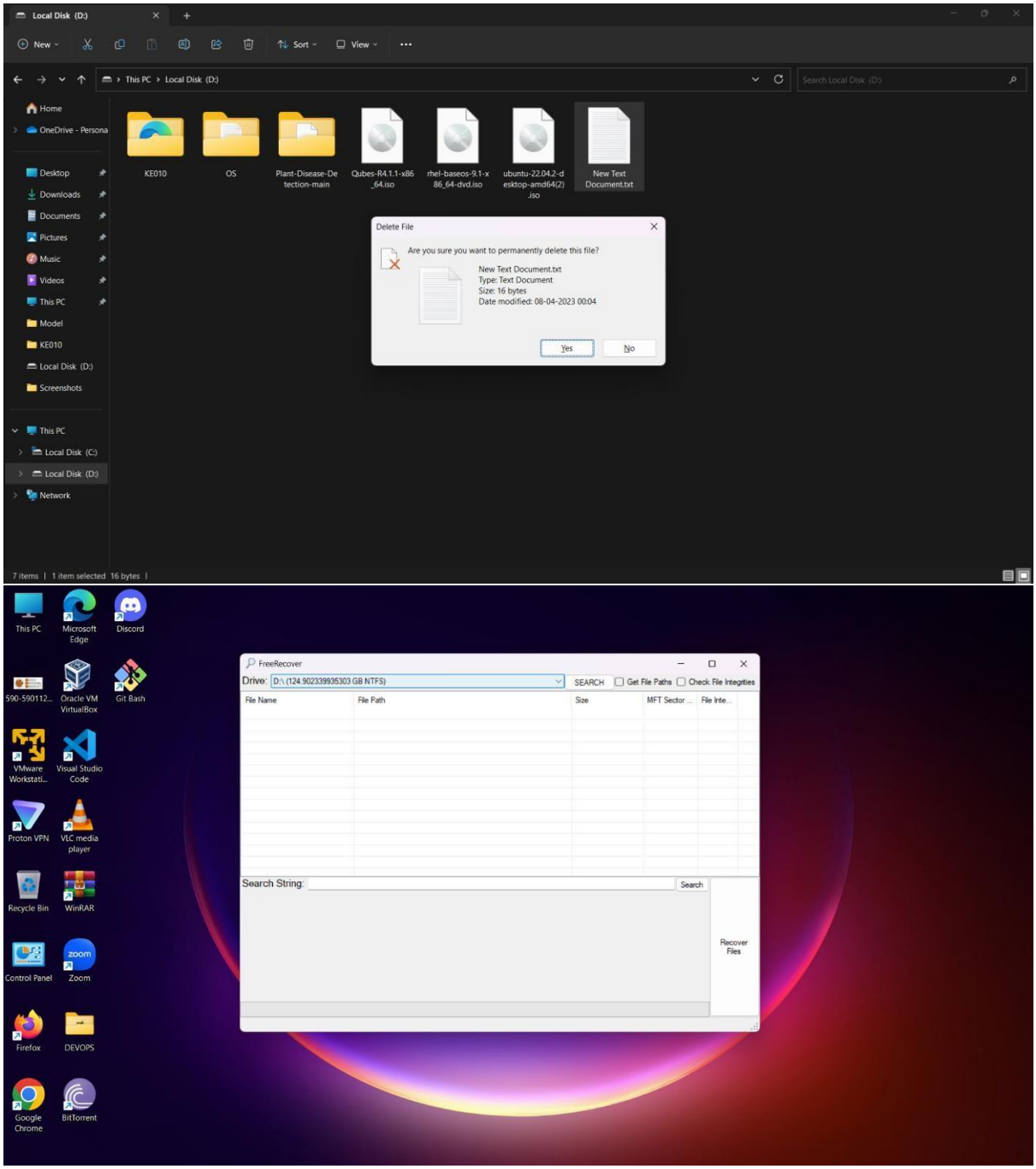
Some brief steps for Data extraction in Windows in FreeRecover

1. Install FreeRecover or run the exe file of the Software.
2. Selecting Disk.
3. Searching for deleted Data.
4. Desired Data selection to recover.
5. Recover deleted Data.

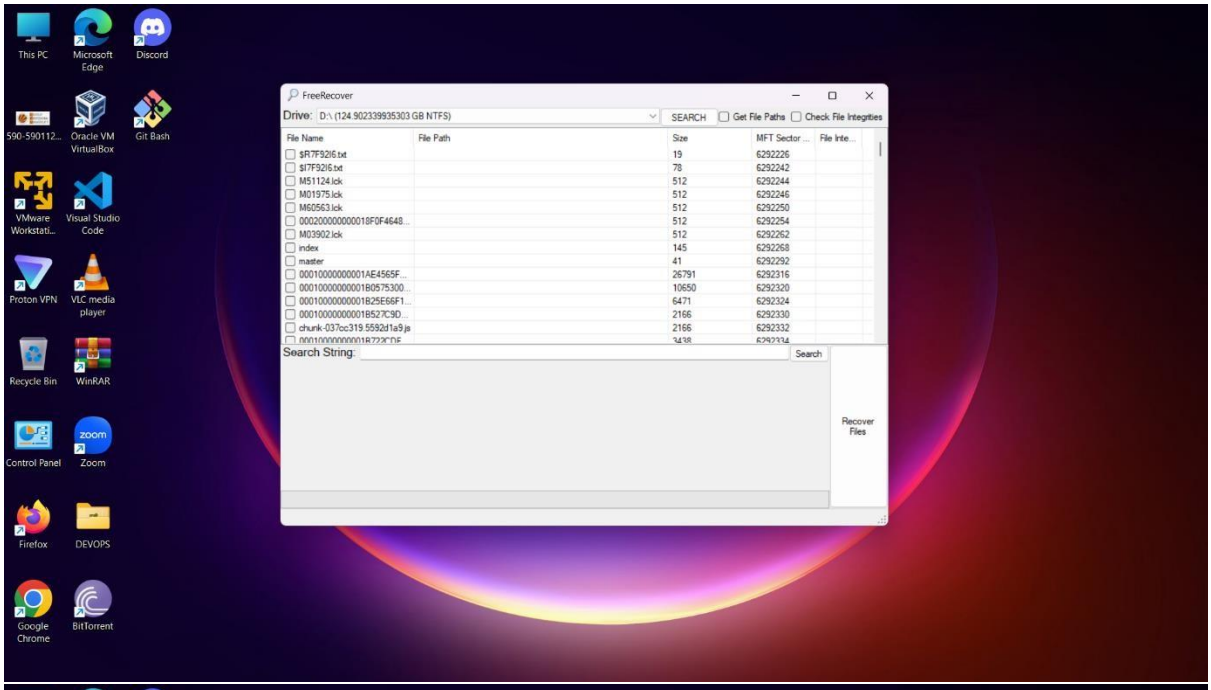
System Snapshots

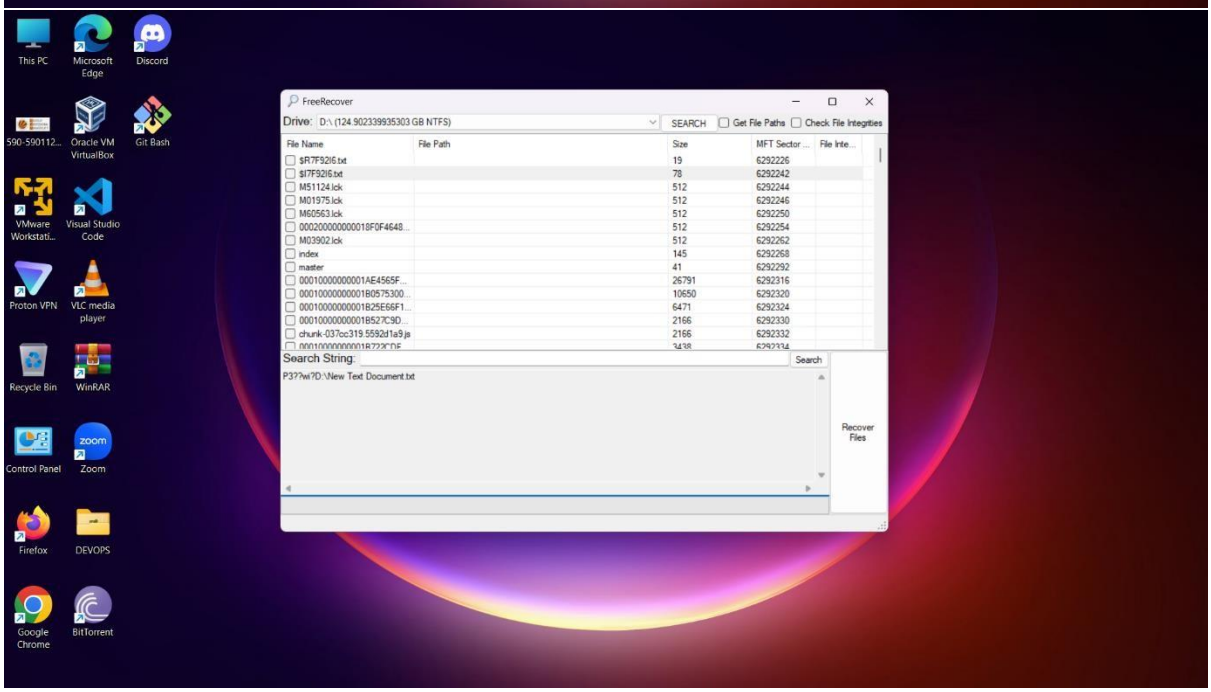


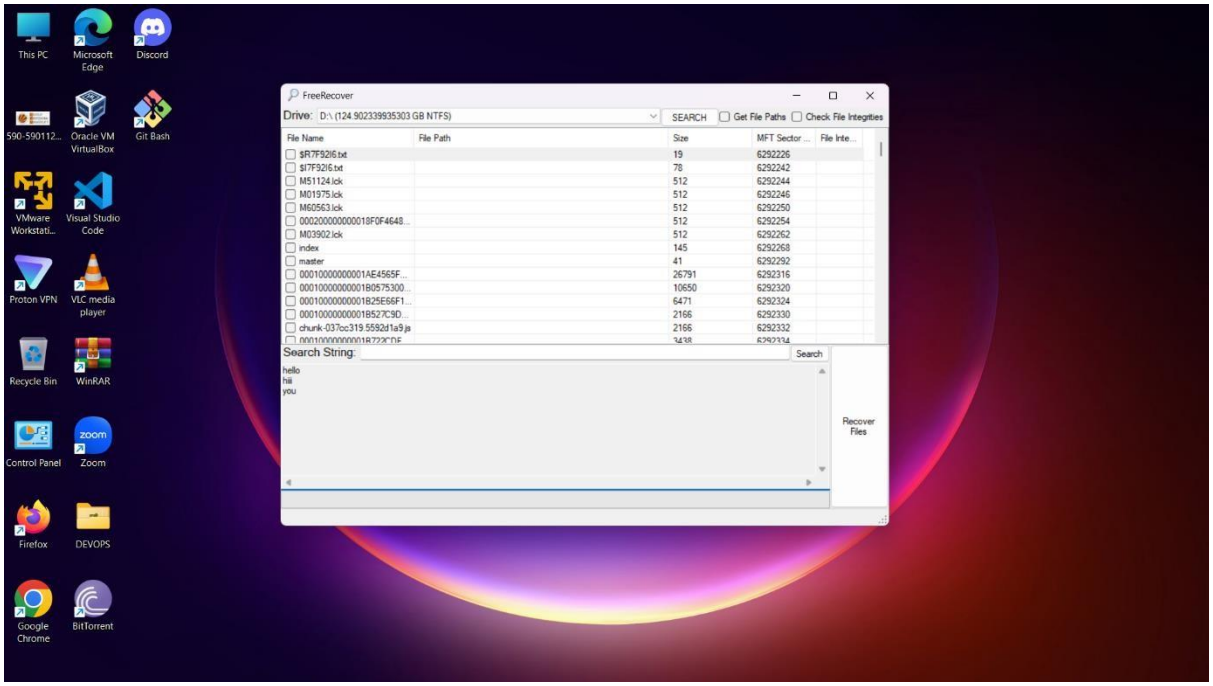



Local Disk (D:)
New
Sort
View
This PC › Local Disk (D:)
Search Local Disk (D:)
Home
OneDrive - Persona
Desktop
Downloads
Documents
Pictures
Music
Videos
This PC
Model
KE010
Local Disk (D:)
Screenshots
This PC
Local Disk (C:)
Local Disk (D:)
Network
KE010
OS
Plant-Disease-Detection-main
Qubes-R4.1.1-x86_64.iso
rhel-baseos-9.1-x86_64-dvd.iso
ubuntu-22.04.2-desktop-amd64(2).iso
New Text Document.txt
Delete File
Are you sure you want to permanently delete this file?
New Text Document.txt
Type: Text Document
Size: 16 bytes
Date modified: 08-04-2023 00:04
Yes
No
7 items | 1 item selected 16 bytes

This PC
Microsoft Edge
Discord
590-590112...
Oracle VM VirtualBox
Git Bash
VMware Workstati...
Visual Studio Code
Proton VPN
VLC media player
Recycle Bin
WinRAR
Control Panel
Zoom
Firefox
DEVOPS
Google Chrome
BitTorrent
FreeRecover
Drive: D:\ (124.902339935303 GB NTFS)
SEARCH
Get File Paths
Check File Integrities
File Name
File Path
Size
MFT Sector ...
File Inte...
Search String:
Search
Recover Files


This PC
Microsoft Edge
Discord
590-590112...
Oracle VM VirtualBox
Git Bash
VMware Workstati...
Visual Studio Code
Proton VPN
VLC media player
Recycle Bin
WinRAR
Control Panel
Zoom
Firefox
DEVOPS
Google Chrome
BitTorrent
FreeRecover
Drive: D:\ (124.902339935303 GB NTFS)
SEARCH
Get File Paths
Check File Integrities
File Name
File Path
Size
MFT Sector ...
File Inte...
$R7F92I6.txt 19 6292226
$I7F92I6.txt 78 6292242
M51124.lck 512 6292244
M01975.lck 512 6292246
M60563.lck 512 6292250
000200000000018F0F4648... 512 6292254
M03902.lck 512 6292262
index 145 6292268
master 41 6292292
00010000000001AE4565F... 26791 6292316
00010000000001B0575300... 10650 6292320
00010000000001B25E66F1... 6471 6292324
00010000000001B527C9D... 2166 6292330
chunk-037cc319.5592d1a9.js 2166 6292332
Search String:
Search
Recover Files


This PC
Microsoft Edge
Discord
590-590112...
Oracle VM VirtualBox
Git Bash
VMware Workstati...
Visual Studio Code
Proton VPN
VLC media player
Recycle Bin
WinRAR
Control Panel
Zoom
Firefox
DEVOPS
Google Chrome
BitTorrent
FreeRecover
Drive: D:\ (124.902339935303 GB NTFS)
SEARCH
Get File Paths
Check File Integrities
File Name
File Path
Size
MFT Sector ...
File Inte...
$R7F92I6.txt 19 6292226
$I7F92I6.txt 78 6292242
M51124.lck 512 6292244
M01975.lck 512 6292246
M60563.lck 512 6292250
000200000000018F0F4648... 512 6292254
M03902.lck 512 6292262
index 145 6292268
master 41 6292292
00010000000001AE4565F... 26791 6292316
00010000000001B0575300... 10650 6292320
00010000000001B25E66F1... 6471 6292324
00010000000001B527C9D... 2166 6292330
chunk-037cc319.5592d1a9.js 2166 6292332
Search String:
Search
P3??wi?D:\New Text Document.txt
Recover Files



GitHub Link:

https://github.com/jatinrana19/INT-301

References

https://forum.cgsecurity.org/phpBB3/

https://www.cgsecurity.org/testdisk.pdf

https://www.cgsecurity.org/wiki/TestDisk_FAQ

https://www.cgsecurity.org/wiki/Undelete_files_from_NTFS_with_TestDis