Indira Gandhi National Open University

MSEL-032

Cyber Attacks, Cloud Security and Data Analytics and Recovery Lab

Cyber Attacks, Cloud Security and Data Analytics and Recovery Lab

Lab Session on Cloud Server Implementation and Security

Lab Session on IT Audit and Penetration Testing

Lab Session on Mobile Phone Auditing and Data Recovery

Lab Session on Hands on open source Data Analytics and Recovery Tools

Expert Committee

Prof. Nilesh K. Modi, SOCS, Dr. BAOU, Ahmedabad

Prof. D.K. Lobiyal, SOCSS, JNU, NewDelhi

Prof. Manoj Kumar
DCSE, DTU, Delhi

Prof. Sushila Madan
LSRC for Women, DU, Delhi

Dr. Anil K Saini
SOMS, GGSIPU New Delhi

Prof. P.V. Suresh, SOCIS, IGNOU

Prof. V.V. Subrahmanyam
SOCIS, IGNOU

Dr. Anup Girdhar
CEO, SST, New Delhi

Dr. Jeetendra Pande, SOCSIT, UOU, Haldwani

Dr. Rachna Agarwal
SOVET, IGNOU

Dr. Geetika Johry
SOVET, IGNOU

Prof. R.S.P Singh
SOVET, IGNOU

Ms Urshla Kant
SOVET, IGNOU

Prof Ashok K Gaba
SOVET, IGNOU

Course Design Committee

Prof. Nilesh K. Modi, SOCS, Dr. BAOU, Ahmedabad

Prof. Sushila Madan
LSRC for Women, DU, Delhi

Prof. Manoj Kumar
DCSE, DTU, Delhi

Prof. Amit Prakash Singh, USICT, GGSIPU, Delhi

Dr. Jeetendra Pande, SOCSIT, UOU, Haldwani

Dr. Anup Girdhar
CEO, SST, New Delhi

Prof. P.V. Suresh, SOCIS, IGNOU

Ms Urshla Kant
SOVET, IGNOU

Prof Ashok K Gaba
SOVET, IGNOU

Course Preparation Team

Course Contributors

(Practical Activity: Unit 1 to 4)

Keshav Kaushik
Assistant Professor,
Cyber Security and Digital Forensics
SCS, UPES, Dehradun

Content Editing

Prof. Sushila Madan
LSR College, DU

Format and Language Editing

Prof. Ashok K.Gaba
SOVET, IGNOU

Proof Reading

Ms Urshla Kant
SOVET, IGNOU

Programme & Course Coordination

Prof. Ashok K.Gaba
Ms Urshla Kant (Programme Co-Coordinator)

Production

School of Vocational Education and Training (SOVET) IGNOU, New Delhi

ISBN- - - -

Printed and Published on behalf of the Indira Gandhi National Open Uiversity, New Delhi by the Registrar, MPDD.
Printed at

LET US BEGIN HERE

The Course on the theme of **Cyber Attack: Use of Technology in Cyberspace** is divided into two Blocks. It comprises eights Units (four units in each block) in all. A schematic representation of the design of the Units is given below:

Unit X

X.0 Objectives
X.1 Introduction
X.2 Section 1 (Main Theme)
X.2.1 Sub-Section 1 of Section 1
X.2.2 Sub-Section 2 of Section 1
.....
.....
.....

Check Your Progress

X.3 Section 2 (Main Theme)
X.3.1 Sub-Section 1 of Section 2
X.3.2 Sub-Section 2 of Section 2
.....
.....
.....

Check Your Progress

X.n Let Us Sum Up

The Key

Check Your Progress:

The section **Learning Outcomes** in each Unit articulates briefly:

- what we expect from you once you complete working on the Unit.

The Units are divided into sections for easy reading and better comprehension. Each section is indicated distinctly by **BOLD CAPITALS** and each sub-section by **relatively smaller but bold** typeface. The significant divisions within sub-sections are in **still smaller but bold** typeface so as to make it easier for you to see their place within sub-sections, and the items which need to be highlighted are numbered [i.e. (i). (ii),etc.]. For purposes of uniformity we have employed the same scheme of 'partitioning' in every Unit throughout the course. Towards the end of each Unit, under the heading 'Let Us Sum Up', we summarise the whole Unit for purposes of recapitulation and ready reference.

Besides, self-check exercises under the captions **Check Your Progress** have been provided at a few places in each of these Units which invariably end with model/sample answers to the questions set in these exercises.

What, perhaps, you would like to do is to go through the Units and jot down important points as you read. This will help you keep track of and assimilate what you have been reading in a particular Unit and answer the 'self-check exercises'. **These exercises are not meant to be submitted to us for correction and evaluation**. The exercises are meant to function as study tools to help you keep on the right track as you read the Units. The points you have jotted down will help you in answering the questions. If required, you may as well take a quick look through the relevant pages to locate the answers.

We would like you to work out the answers in the blank space(s) provided in this booklet. The purpose of giving self-check exercises will be served satisfactorily if you compare your answers with the model ones given at the end of each Unit, after having written your answer in the blank space. **You may be tempted to have a furtive glance at the model answer(s),** as soon as you come across an exercise. But we do hope that you will overcome the temptation, and turn to the model answers (which are not the best answers necessarily) only after you write yours. Each block or combination of blocks will have at least one assignment which should be sent to us for evaluation. In all, you may have to work on two/three assignments for a Course.

The following norms have to be strictly practised while you are working through the assignments.

- The answer should be precise and well-documented.
- Before you put down anything in words, assimilate what you have read, integrate it with what you have gathered from your experience and feed it into your answer.
- Make the best use of the Block and additional reading materials for diligently working through the assignments.
- Write your roll number legibly as indicated in the "Students' Programme Guide".

COURSE INTRODUCTION

This course introduces you to practically work with the most powerful concepts i.e. Cyber Attacks, Cloud Security and Data Analytics and Recovery. After completing this course you will be able to solve problems for various techniques of Cyber Attacks, Cloud Security and Data Analytics and Recovery, you learned in MSEL-032 (Cyber Attacks, Cloud Security and Data Analytics and Recovery) course of this programme.

In order to have better understanding you are advised to firstly get the conceptual clarity of the various concepts presented in the MSEL- 032(Cyber Attacks, Cloud Security and Data Analytics and Recovery) course of this programme, and then implement your understanding to solve the problems given in the various sessions of this course.

This block consists of following Lab sessions:

- **Lab Session on Cloud Server Implementation and Security**
- **Lab Session on IT Audit and Penetration Testing**
- **Lab Session on Mobile Phone Auditing and Data Recovery**
- **Lab Session on Hands on open source Data Analytics and Recovery Tools**

Structure

1.0 Introduction

Cloud computing is in a growing market throughout sectors for digital process transformation. That is why it is preferred among business owners who want to stay up with the latest manufacturing trends. On the other hand, cloud implementation solutions are still too complicated for non-technical people. As a result, cloud-computing providers began providing expert assistance. Cloud Computing guarantees that IT services are implemented consistently across all industries. Cloud computing consulting services make this feasible. They concentrate on how to use cloud computing in a variety of industries. Understanding the definition of cloud computing is the first step in implementing it. Let us have a look at how cloud computing functions. That is preferable before we talk about how to put cloud computing into practice. Cloud computing refers to the use of the internet to supply computing resources. Infrastructure, environment, and software are all part of these computational services. What distinguishes cloud computing from other computing resources? Its ability to be accessed from a distance. Availability of resources via the internet allows corporate data to be accessed from anywhere, at any moment. It will be easier to collaborate if you know how to use cloud computing. This has the potential to reshape the corporate workflow. A company's ability to utilize cloud computing can help with production. This is accomplished by delivering real-time information. As a result, firms will have an easier time monitoring the quality and amount of their output. It is recommended that business owners understand how to use cloud computing services.

1.1 Learning Outcome

This is the first unit of this course, which seeks to introduce and explain the various concepts and definitions generally used in Cloud Computing. This unit covers the fundamentals of Cloud Computing, including the cloud server architecture, why it is essential, what Cloud security specialists do to secure data, and the various case studies. What is cloud computing, and how does it work? We'll offer you an overview of cloud computing in this chapter and some information on the many features of cloud computing and information security in cloud computing.

1.2 Introduction to Cloud Computing

Cloud computing provides on-demand access to computer resources—apps, servers, storage systems, developer tools, communication infrastructure, and more—hosted in a remote data centre controlled by a cloud services provider over the internet (or CSP). The CSP charges a monthly subscription or charges based on the usage of specific resources. Cloud computing is Internet-based computing in which users may access hardware and software components on request. It is a byproduct and repercussion of the internet's easy accessibility to remote computer locations. You may utilize software given via the internet in your browser without installing it, host a program on the internet, build up your online file server and database system, and more using cloud computing. Data storage, data recovery, emails, virtualization software, software programming and implementation, big data analytics, and consumer web apps are among the many use cases for businesses of all sizes and industries using the cloud. Healthcare businesses, for instance, are utilizing the cloud to provide more individualized therapies for their patients. Companies in the financial services industry are turning to the cloud to help them prevent and detect fraud in realtime. Video game developers are also utilizing the cloud to provide online games to thousands of players worldwide.

The cloud provides you quick access to various technologies, allowing you to create more quickly and construct almost anything you can dream of. You may instantly spin up resources as needed, including computation, storage, databases, the Internet of Things, deep learning, data lakes and statistics, and so much more. Information communication technologies may be deployed in minutes, allowing you to get from concept to implementation numerous orders of magnitude quicker than before. This offers you the ability to try new things, explore new ideas, and alter your business. You do not have to over-provision capacity upfront with cloud technology to manage future peak levels of company activity. Instead, you allocate the exact quantity of resources that you require. As your company needs change, you may adjust these capabilities up or down to expand and reduce capacity immediately. You may exchange fixed expenditures for variable expenses on the cloud and only pay for IT as you use it. Furthermore, the variable charges are far less than what you would spend if you did it yourself due to economies of scale. You may quickly grow to new geographic locations and deliver internationally using the cloud.

Service Models

Cloud computing is based on service models. There are three basic categories of cloud computing.

1. Infrastructure as a Service
2. Platform as a Service
3. Software as a Service

Every form of cloud computing offers varying degrees of control, customization, and administration, allowing you to choose the best services collection for your purposes.

1. Platform as a Service (PaaS) - PaaS is a cloud computing model where the cloud service provider delivers hardware and software tools over the internet. It relieves you of the burden of managing underlying infrastructure (most common hardware and software platforms), allowing you to concentrate on the deployment and administration of your applications. This will enable you to be more productive since you will not have to deal with resource procurement, production scheduling, updating software, patching, or any other homogeneous heavy lifting that comes with operating your program. Users have relatively little influence over their software and development environment while using PaaS services. PaaS providers impose a software layer on top of the hardware they provide, compelling customers to interact with that layer. This is not always a bad thing because PaaS vendors make it easier for customers to build their web applications by reducing the technical knowledge required. Google App Engine is an instance of a PaaS service. Google App Engine may be used to host a web application and leverage the database and file storage systems supplied by Google App Engine to supplement the application. Nevertheless, because Google App Engine is a PaaS, the application can only be coded in Java or Python.
2. Infrastructure as a Service (IaaS) - The basic building blocks of cloud computing are contained in IaaS. It usually gives users access to networking capabilities, virtual or dedicated computers, and data storage space. IaaS allows you to have the most flexibility and independence over your IT infrastructure. It is pretty comparable to existing IT resources that many IT organizations and programmers are already familiar with IaaS companies give customers hardware and the bare essentials of software, such as virtual machines and hard drive space, on which they may develop. Users have a lot of control over the services provided by IaaS providers since they may customize settings significantly and use any type of software or coding ecosystem on top of the services. Rackspace, for instance, is an IaaS provider.
3. Software as a Service (SaaS) - SaaS gives you a fully functional product managed and maintained by the service provider. Most of the time, when people talk about SaaS, they are talking about end-user apps like web-based email. You don't have to worry about how the product is updated or how the core architecture is managed when using a SaaS. All you have to do now is consider how you'll utilize the programme. SaaS companies, in essence, give the software to consumers over the internet via a web browser. Google Docs is an example of a SaaS, as it allows users to modify documents using software that is distributed over the internet. The significant benefit of utilizing a SaaS would

be that you don't have to bother about setup, storage space, or loss of data due to PC failures or updates. The cloud vendor takes care of everything and allow you to use the software trough web browser.

Principles of Cloud Computing

- Resource Provisioning: Cloud computing companies take advantage of significant economies of scale by pooling their resources. They assemble a large number of servers and hard drives and implement thesame settings, security, and other features to all of them.
- Elasticity: Adding extra hard drive space or server connectivity on-demand can be accomplished withonly a few mouse clicks. Cloud computing also offers geographic scalability, with the option of replicating data across many data centers across the world.
- Automatic resource deployment: The user merely needs to pick the types and characteristics of the services he requires, and the cloud services provider will set up and hook them up for him effortlessly.
- Virtualization: Customers do not have to be concerned about the physical status of their gear, nor do they have to be concerned about device interoperability, thanks to virtualization.
- Pay what you use: Consumers are only charged for what they use with metered billing.

1.2.1 Study and implementation of infrastructure as Service using Open Stack

OpenStack is a free and open source, cloud computing software platform that is widely used in the deployment of infrastructure-as-a-Service (IaaS) solutions. The core technology with OpenStack comprises a set of interrelated projects that control the overall layers of processing, storage and networking resources through a data center that is managed by the users using a Web-based dashboard, command-line tools, or by using the Restful API. Currently, OpenStack is maintained by the OpenStack Foundation, which is a non-profit corporate organization established in September 2012 to promote OpenStack software as well as its community. Many corporate giants have joined the project, including GoDaddy, Hewlett Packard, IBM, Intel, Mellanox, Mirantis, NEC, NetApp, Nexenta, Oracle, Red Hat, SUSE Linux, VMware, Arista Networks, AT&T, AMD, Avaya, Canonical, Cisco, Dell, EMC, Ericsson, Yahoo!, etc:

1.2.2 Deployment of OpenStack using DevStack

DevStack is used to quickly create an OpenStack development environment. It is also used to demonstrate the starting and running of OpenStack services, and provide examples of using them from the command line. DevStack has evolved to support a large number of configuration options and alternative platforms and support services. It can be considered as the set of scripts which install all the essential OpenStack services in the computer without any additional software or configuration. To implement DevStack, first download all

the essential packages, pull in the OpenStack code from various OpenStack projects, and set everything for the deployment.

To install OpenStack using DevStack, any Linux-based distribution with 2GB RAM can be used to start the implementation of IaaS.

Here are the steps that need to be followed for the installation.

1. Install Git

$ sudo apt-get install git

2. Clone the DevStack repository and change the directory. The code will set up the cloud infrastructure.

$ git clone http://github.com/openstack-dev/devstack $ cd devstack/

/devstack$ ls

accrc exercises HACKING.rst rejoin-stack.sh tests AUTHORS exercise.sh lib run_tests.sh tools clean.sh extras.d LICENSE samples unstack.sh driver_certs files localrc stackrc eucarc functions openrc stack-screenrc

exerciserc functions-common README.md stack.sh

stack.sh, unstack.sh and rejoin-stack.sh are the most important files. stack.sh script is used to setup DevStack.unstack.sh is used to destroy the DevStack setup. If you are on the earlier execution of ./stack.sh, the environment can be brought up by executing the rejoin_stack.sh script.

3. Execute the stack.sh script:

/devstack$ ./stack.sh

Here, the MySQL database password is entered. There' s no need to worry about the installation of MySQL separately on this system. We have to specify a password and this script will installMySQL,and usethispasswordthere.

Finally, we will have the script ending as follows:

```
+       merge_config_group /home/r/devstack/local.conf post-extra
+       local localfile=/home/r/devstack/local.conf
+       shift
+       local matchgroups=post-extra
+       [[ -r /home/r/devstack/local.conf ]]
+       return 0
+       [[ -x /home/r/devstack/local.sh ]]
+       service_check
+       local service
+       local failures
+       SCREEN_NAME=stack
+       SERVICE_DIR=/opt/stack/status
```

+ [[! -d /opt/stack/status/stack]]

++ ls ‘ /opt/stack/status/stack/*.failure’

++ /bin/true

+ failures=

+ ‘ [‘ -n ‘ ’ ‘]’

+ set +o xtrace

- Horizon is now available at http://1.1.1.1/
- Keystone is serving at http://1.1.1.1:5000/v2.0/
- Examples on using the novaclient command line are in exercise.sh
- The default users are: admin and demo
- The password: nova
- This is your host IP: 1.1.1.1

After all these steps, the machine becomes the cloud service providing platform. Here, 1.1.1.1 is the IP of my first network interface.

We can type the host IP provided by the script into a browser, in order to access the dashboard ‘ Horizon’ . We can log in with the username ‘ admin’ or ‘ demo’ and the password ‘ admin’ .

You can view all the process logs inside the screen, by typing the following command:

$ screen –x

Executing the following will kill all the services, but it should be noted that it will not delete any of the code. To bring down all the services manually, type:

$ sudo killall screen

localrc configurations

localrc is the file in which all the local configurations (local machine parameters) are maintained. After the first successful stack.sh run, you will see that a localrc file gets created with the configuration values you specified while running that script.

The following fields are specified in the localrc file:

DATABASE_PASSWORD

RABBIT_PASSWORD

SERVICE_TOKEN

SERVICE_PASSWORD

ADMIN_PASSWORD

If we specify the option OFFLINE=True in the localrc file inside DevStack directory, and if after specifying this, we run stack.sh, it will not check any parameter over the Internet. It will set up DevStack using all the packages and code residing in the local system. In the phase of code development, there is need to commit

the local changes in the /opt/stack/nova repositorybefore restack (re-running stack.sh) with the RECLONE=yes option. Otherwise, the changes will not be committed.

To use more than one interface, there is a need to specify which one to use for the external IP using this configuration:

HOST_IP=xxx.xxx.xxx.xxx

Cinder on DevStackCinder is a block storage service for OpenStack that is designed to allow the use of a reference implementation (LVM) to present storage resources to end users that can be consumed by the OpenStack Compute Project (Nova). Cinder is used to virtualise the pools of block storage devices. It delivers end users with a self-service API to request and use the resources, without requiring any specific complex knowledge of the location and configuration of the storage where it is actually deployed.

All the Cinder operations can be performed via any of the following:

1. CLI (Cinder' s python-cinderclient command line module)
2. GUI (Using OpenStack' s GUI project horizon)
3. Direct calling of Cinder APIs

Creation and deletion of volumes: To create a 1 GB Cinder volume with no name, run the following command:

$ cinder create 1

To see more information about the command, just type cinder help <command>

$ cinder help create

usage: cinder create [--snapshot-id <snapshot-id>]

[--source-volid <source-volid>] [--image-id <image-id>]

[--display-name <display-name>]

[--display-description <display-description>]

[--volume-type <volume-type>]

[--availability-zone <availability-zone>]

[--metadata [<key=value> [<key=value> ...]]]

<size>

Add a new volume.

Positional arguments:

<size> Size of volume in GB

Optional arguments:

--snapshot-id <snapshot-id>

Create volume from snapshot id (Optional,

Default=None)

--source-volid <source-volid>

Create volume from volume id (Optional, Default=None)

--image-id <image-id>

Create volume from image id (Optional, Default=None)

--display-name <display-name>

Volume name (Optional, Default=None)

--display-description <display-description>

Volume description (Optional, Default=None)

--volume-type <volume-type>

Volume type (Optional, Default=None)

--availability-zone <availability-zone>

Availability zone for volume (Optional, Default=None) --metadata [<key=value> [<key=value> ...]]

Metadata key=value pairs (Optional, Default=None)

To create a Cinder volume of size 1GB with a name, using cinder create –display-name myvolume:

$ cinder create -- display-name myvolume 1

+------------------ +--------------------------------------- +

| Property | Value |

+------------------ +---------------------------------- - +

| attachments | [] |

| availability_zone | nova |

| bootable | false |

| created_at | time | | display_description | None |

| display_name | myvolume | | id | id |

| metadata | {} |

| size | 1 |

| snapshot_id | None |

| source_volid | None |

| status | creating |

| volume_type | None |

+---------------------- +-------------------------------- +

To list all the Cinder volumes, using cinder list:

$ cinder list

ID Status Display Name Size Volume type Bootable Attached To id1 Available Myvolume 1 None False id2

Available None 1 None False

To delete the first volume (the one without a name), use the cinder delete <volume_id>command. If we execute cinder list really quickly, the status of the volume going to ‘ deleting’ can be seen, and after some time, the volume will be deleted:

$ cinder delete id2

$ cinder list

ID Status Display Name Size Volume type Bootable Attached To id1 Available Myvolume 1 None False id2 Deleting None 1 None False

Volume snapshots can be created as follows:

$ cinder snapshot-create id2

+---------------------+----------------------------------+

| Property | Value |

+---------------------+----------------------------------+

| created_at | TimeStamp |

| display_description | None |

| display_name | None |

| id | snapshot2 |

| metadata | {} |

| size | 1 |

| status | creating |

| volume_id | id2 |

+---------------------+----------------------------------+.

Check your progress 1

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i) What are the five crucial principles of Cloud Computing?

……

……

……

……

……

……

……

……

1.3Create Custom Apps for Salesforce Classic

Salesforce Sales Cloud is a customer relationship management (CRM) platform designed to support sales, marketing and customer support in both business-to-business (B2B) and business-to-customer (B2C) contexts. Sales Cloud is a fully customizable product that brings all the customer information together in an integrated platform that incorporates marketing, lead generation, sales, customer service and business analytics and provides access to thousands of applications through the AppExchange. The platform is provided as Software as a Service (SaaS) for browser-based access; a mobile app is also available.

Create custom apps to give your Salesforce Classic users' access to everything they need all in one place.

If you're new to custom apps, we recommend using Lightning Platform quick start to create an app. With this tool, you can generate a basic working app in just one step.

If you've already created the objects, tabs, and fields you need for your app, follow these steps. With this option, you create an app label and logo, add items to the app, and assign the app to profiles.

1. From Setup, enter Apps in the Quick Find box, then select Apps.
2. Click New.
3. If the Salesforce console is available, select whether you want to define a custom app or a Salesforce console.
4. Give the app a name and description.

An app name can have a maximum of 40 characters, including spaces.

5. Optionally, brand your app by giving it a custom logo.
6. Select which items to include in the app.
7. Optionally, set the default landing tab for your new app using the Default Landing Tab drop-down menu below the list of selected tabs. This determines the first tab a user sees when logging into this app.
8. Choose which profiles the app will be visible to.
9. Check the Default box to set the app as that profile's default app, meaning that new users with the profile see this app the first time they log in. Profiles with limits are excluded from this list.
10. Click Save

1.3.1 Platform Virtualization

Platform virtualization is a technique that abstracts a single server's existing hardware resources into several virtual data centers, each of which can run various operating systems. The hypervisor is the virtual machine's brain. This is the software that lies between the operating system and the hardware. Its primary function is to allocate system resources. Each of these operating systems acts as though it had complete control over the server's resources. Figure 1 shows how this works. Several of the three operating systems is a virtual

machine image of the different operating systems. Such machine images are snapshots of operating systems then put into the virtual machine environment.

It is crucial to remember that you are not paying for the entire server. Instead, cloud service providers give you a virtual machine image, one of many in the virtual computing ecosystem. Figure 1 shows the diagram for platform virtualization.

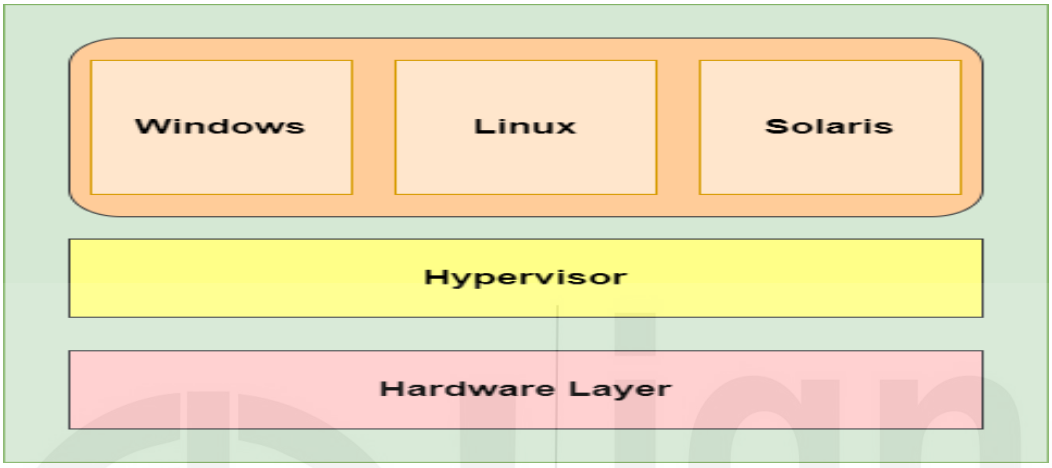


Figure 1: Diagram for Platform Virtualization

1.3.2 Characteristics of Platform Virtualization

Cloud computing providers may deliver cost-effective, quickly deployable, and scalable servers, thanks to platform virtualization's capabilities. Such characteristics are as follows:

- **Server use has increased:** Without platform virtualization, the typical server in a corporate data centre is only used 5 to 10% of the time. Even at maximum load, the highest is approximately 20%. Cloud services use platform virtualization to operate many virtual servers on a single physical server, allowing them to lease out more virtualized resources with the same number of underlying hardware.
- **Hardware constraints have been removed**: Cloud computing has removed the hardware constraint. Now the consumers no longer worry about hardware compatibility concerns. All they have to do now is to concentrate on software packages and contractual arrangements. Furthermore, developers should consider the operating system they require and the amount of efficiency they demand.
- **Removal of software requirements: T**he virtualized environment manages device driver dependencies, so you do not need to worry about them with virtualization.
- **Quick Installation and Teardown:** Deploying a "server" is as simple as loading a virtual machine image with virtualization, and it is also simple to deconstruct a machine picture.

Check your progress 2

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i)What is the Hypervisor layer? What are its functions?

.....

.....

.....

.....

.....

.....

.....

.....

1.4 File Storage in Cloud

The following parts will present the basic ideas of cloud storage, demonstrate their many advantages, and examine some of the existing cloud storage challenges.

1.4.1 Architecture

Cloud storage providers commonly use a three-layer design. In the diagram below, one can see a representation of the overall system architecture and some of the features associated with it in modern cloud storage.

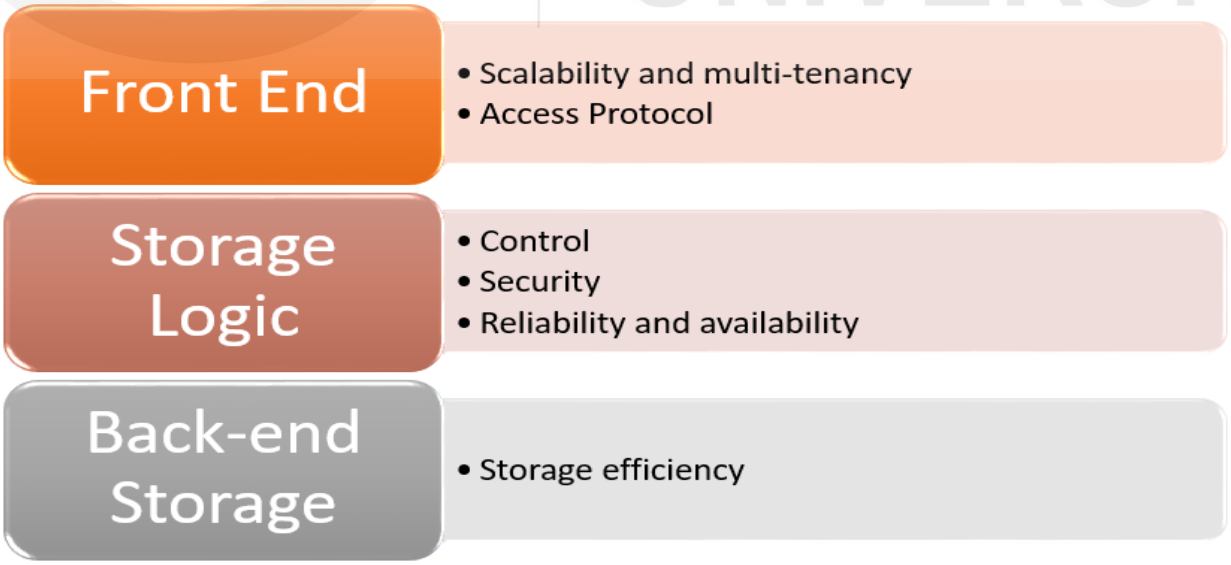


Figure 2: Diagram for cloud architecture

- The front end is responsible for client-server communication, and different APIs are available to access the actual storage. This layer is also concerned with attaining outcomes such as multi-tenancy, which we shall define in the following chapter. Furthermore, it offers a variety of scalability options via various techniques.
- The storage logic layer is in charge of several features and administrative tasks, such as guaranteeing a high level of accessibility and dependability. It's also a type of security fence. In addition, it serves as a cloud storage administrator.
- The back-end is responsible for deploying physical data storage technologies such as GFS (Google File System). It entails various techniques for increasing storage efficiency and decreasing infrastructure expenses.

This section will go over several of the qualities mentioned in the previous chapter's graphic in greater depth. Multi-tenancy relates to the capability of a specific incident of service to serve a large number of clients or renters. It also refers to many tiers of the cloud storage stack, allowing multiple customers to adhere to the same cloud processing capabilities while maintaining security and privacy over their sensitive information. Cloud storage's automated management is an important feature. In general, there are two types of costs: the expense of the physical storage system and the cost of operating it. The management cost is concealed, yet it accounts for a significant portion of the total cost in the long term. The cloud storage system must be capable of adding more storage and dynamically configuring itself to handle it and automatically detecting faults. Since cloud computing is fundamentally about ease, automated management is essential in cloud storage.

One of the main reasons to pick cloud storage over conventional data hosting is the uniformity of functionality throughout the world. Files are typically kept on one server in traditional file hosting, and therefore clients that are far distant from that server would have poor performance. There are two degrees of geographical flexibility with cloud storage. The file is dispersed over numerous servers in the exact location of your original data. Second, some CDNs (content delivery networks) are available on-demand. These are networks with servers located worldwide, allowing for rapid content delivery to customers anywhere around the globe. Cloud storage may reach the same high degree of high-performance electricity worldwide by utilizing CDNs.

One of the key distinctions between cloud services and regular storage is access. Several cloud storage providers now provide numerous access mechanisms, but the Web-Service API remains the most popular. The REST (Representational State Transfer) framework is used to implement these. The structure is used to create protocols that run on top of the HTTP layer, allowing HTTP to be used as a transport medium. APIs that follow this design are stateless and, as a result, reasonably efficient. Larger cloud storage providers such as Amazon (S3) and Microsoft Azure presently use this strategy. Also, there are additional types of access mechanisms, such as file-based APIs like NFS and FTP, which IBM Smart Business Storage Service supports.

One of the fundamentals of cloud storage is high reliability. With today's technical advancements, one would believe that hard disc failures and data loss are no longer a typical occurrence. On the other hand, hardware failures are unavoidable and can be disastrous back-ups are inadequate. To assure dependability, cloud companies often choose one of two ways. The same information is usually kept on many computers at large cloud service providers. Google's cloud back-end storage is often separated into massive clusters and broken down into 64MB bits. Each of these pieces has a unique ID and is duplicated across several servers in their data centres. Additionally, these devices are powered by various power sources.

Certain service providers also use data-reconstruction methods to aid with lost or damaged data. The Information Dispersal Algorithm is one of these algorithms. This technique is capable of constructing a complete collection of data from several pieces of previously scattered data. If the data is partitioned into four pieces, for example, it may still be rebuilt if one of the sites that has one part of the data is destroyed. It is also feasible to use other ratios. For example, 20 pieces will allow for eight failed locations. To lessen the possibilities of all components of the data being lost at the same time, these bits of data are frequently scattered over multiple geographical regions.

Another feature of cloud storage worth highlighting is its reasonable cost-to-storage ratio. Additional data should be stored with the same underlying hardware in order to cut costs. Using data-reduction methods to lower the amount of resources data consumes is a frequent way to do this. There are two methods to this that are notable: De-duplication is the eradication of any identical copies of data detected via the screening of data signatures. Compression is the encoding of data in another more economical format to accomplish data reduction. When storing sensitive data on the cloud, high levels of protection are needed.

Check your progress 2

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i) What are the three layers of cloud architecture?

……

……

……

……

……

……

……

……

1.4.2 Challenges in Cloud Storage

The speed and latency of cloud storage, as well as the integration of cloud applications into current IT infrastructure, are two problems unique to cloud storage.

- A company will need to incorporate cloud storage into their present workflow or other kinds of offline storage sites before using it. File access methods used by traditional file servers and cloud storage resources are not the same. Servers access their storage using block protocol, while cloud storage services typically only give web protocol access via REST-based APIs and SOAP-based APIs, which are APIs built on top of the HTTP protocol to allow more efficient access. To manage the operations, each of the major suppliers has its own set of APIs. This adds a layer of complexity to the situation. Conventional file storage operations in mature businesses are often more sophisticated. Integrating the usage of cloud storage into their present workflows will take a significant amount of time, money, and effort. A younger firm with a less complicated architecture, on the other side, will not encounter the same issue since it will be much easier to incorporate cloud storage into a process that has not yet matured.
- Companies may utilize cloud storage to back up large volumes of data on a regular basis. Data will be sent to a physically distant location as part of these back-up processes. Compared to offline storage technologies, this will invariably be slower. While cloud storage is easier to use, better scalable for businesses, and more trustworthy, it still lags behind offline storage alternatives in terms of performance. In principle, today's cloud storage is geared for processes that require less processing power. Outside of cloud storage, organizations should leave operations with a high-performance need. Real-time bank transactions are one example of this.

1.5Cloud Case Study

Because Amazon EC2 is only a platform virtualization solution, there are difficulties with reliability. To address this issue, Amazon Web Services (AWS) has created a number of additional service options.

1.5.1 Amazon Cloudwatch

For DevOps engineers, developers, site reliability engineers (SREs), IT managers, and product owners, Amazon CloudWatch provides a monitoring and observability solution. To monitor your apps, respond to system-wide performance indicators, and optimize resource use, CloudWatch offers you with data and actionable insights. CloudWatch logs, metrics, and events are used to collect monitoring and operational data. You obtain total visibility of your AWS resources, apps, and services running on AWS and on-premises, as well as a single picture of operational health. To keep your apps operating smoothly, you can use CloudWatch to identify aberrant activity in your environments, trigger alerts, analyze logs and statistics

side by side, take automatic actions, troubleshoot problems, and find insights. Cloudwatch is used in situations where an instance has been overburdened with requests and the application is slowing down as a result of the congested resources. Cloudwatch keeps track of the performance statistic for your zone based on how it is set up. If it fails to meet a given metric requirement (for example, responses lasting longer than 5 seconds), an alarm will sound virtually, and administrators will be notified. At the same time, the developer may define actions that duplicate instances to manage the load if he so desires. Cloudwatch is limited to a certain geographical area.

In the form of performance measures, logs, and occurrences, new advancements, such as those built on microservices frameworks, generate massive amounts of data. Amazon CloudWatch provides you to gather, retrieve, and correlate data from all of your AWS services, apps, and services operating on AWS and on-premises on a single platform, allowing you to break down data silos and swiftly address issues. Set alarms and automates actions based on established criteria or machine learning (ML) algorithms that detect aberrant metrics. To avoid billing overages, you can, for example, initiate Amazon EC2 Auto Scaling automatically or halt an instance. CloudWatch Events may also be used to initiate serverless processes with AWS Lambda, Amazon SNS, and AWS CloudFormation.

1.5.2 Amazon Elastic IP Address

Nonetheless, traffic from all around the world arrives in the actual world. We also want visitors to be intelligently routed to the closest server that hosts our service based on their location. This is exactly what Amazon Elastic IP Address can help us with. Amazon Elastic IP Address analyses incoming traffic, determines the nearest region, and redirects traffic to that region by establishing a common IP address connection to all virtualized instances. As a result, the challenge of horizontal availability has been resolved. A static IPv4 address built for dynamic cloud computing is known as an elastic IP address. Every AWS account is assigned an Elastic IP address, which is yours until you relinquish it. You can disguise the failure of an instance or programme by immediately rebinding the address to some other example in your account using an Elastic IP address. You may also include the Elastic IP address in a DNS record for your website so that it points to your instance. An Elastic IP address is a temporary public IP address that you may assign to any EC2 instance in a given area until you release it.

1.6 Cloud Storage Providers

In this part, we will look at several Amazon Cloud Storage services as instances of cloud storage services. Here are a few of services provided by AWS.

1. Amazon Elastic Block System (EBS) - Amazon EBS, or Amazon Elastic Block System, allows you to attach storage volumes ranging from 1GB to 1TB to EC2 instances. The snapshot functionality of EBS is one of its most essential features. Snapshots, as the name suggests, allow you to save the current state of the EBS, allowing you to simply restore or replicate data. If given permission, such

snapshots may be shared with other users, letting them access the same data like you in your EBS and making collaboration much simpler. For the first year, new customers get 30GB of EBS storage, 2 million I/O, and 1 GB of snapshot storage each month.

2. Simple Storage Service- Simple Storage Service or S3, is a type of storage that Amazon offers to its customers. Buckets are the most common name for this type of storage. S3 lets you store an unlimited number of things, each up to 5TB in size. This is mostly used to back up your data and to collaborate with other services such as EC2. S3 duplicates the data you supply in various facilities in your selected area to guarantee your data is secure and retrievable, ensuring endurance and dependability. If you want to protect your data for security purposes, Amazon also offers an Amazon S3 Encryption Client library.
3. Amazon Import/Export- Amazon Import/Export is a service that allows you to ship a storage device to an address provided by Amazon so that they may immediately transfer the enormous quantities of data on it into or out of S3 or EBS. Direct transmission is more faster than going over the internet, albeit it may be more expensive in terms of shipping charges. You may then rapidly acquire the information you desire, whether you need to retrieve data or share it with other business contacts.
4. The Amazon Storage Gateway- The Amazon Storage Gateway is presently in beta testing. This functionality, on the other hand, allows you to move data from on-premises storage to S3 in order to assure its availability and longevity. To keep things even easier, these data are saved in S3 as EBS snapshots, allowing you to quickly recover them using EBS replication and EC2 instructions to connect the EBS storage. AWS Direct Connect, which establishes a private connection between you and AWS just for the transmission of files, can enhance transfer speed even further.
5. Microsoft Azure, JustCloud, zipCloud, and livedrive are some of the most popular cloud storage solutions. The services I just listed are significantly different from what a typical user may desire, which is to just store some of their data in the cloud. These services allow users to share files using ways that may be utilized in web development or on websites. New users may be unfamiliar with the cloud and would like to try it out before joining the group. In that scenario, we suggest Amazon Web Services, which is both user-friendly and offers a year of restricted free usage as well as a variety of features to assist you in getting started.

1.7Security in Cloud Computing

There is presently no meaningful security standard in place for cloud computing. Cloud service companies use their own unique security requirements and technology. Clients must guarantee that cloud security satisfies their own security protocols and agrees with their regulations through requirements collecting and provider risk assessments in a vendor cloud model. Because of the nature of the subject, most of this material will be presented to big enterprises from the viewpoint of SMBs (Small-medium Businesses). We will look

at the security features supplied by cloud computing companies, the security threats associated with cloud computing, and some cloud security advice.

1.7.1 Security Features

Using cloud computing has a variety of security advantages. The amount of security, nevertheless, is determined by the supplier. As a guide to many of the security mechanisms and benefits of cloud computing, we've gathered some of the most important security characteristics for you from multiple sources in this chapter.

- Economies of Scale: By combining resources on a large scale, security gains are realized in two ways.
- Costs: When security measures are applied on a big scale, they will be less expensive. For the same money, you might get greater security protection, including packet filtering, patch management, and virtual machine fortification, for example.
- Major cloud service providers, including Amazon, Google, Azure, and Rackspace, all have vast resources and experience in the subject of security. They may be able to provide greater security measures than SMBs because of the concentrated work they put into security. For large enterprises, it can also make the process easier.
- Data Security and Management: Using a centralized data model makes data security and management easier.
- Physical data leakage is reduced: Internal data can be lost in a variety of ways, including the loss of business thumb drives, laptops, and back-up discs. We can avoid this difficulty with cloud technology since data is now kept in the cloud, away from physical devices.
- Benefits tracking: It is simpler to govern and monitor with centralized storage. Because there is just one site where targeted assaults may occur, it is both easier and less expensive to establish security safeguards on centralized data than on individual customers. The resources may then be immediately repurposed for screening, traffic management, verification, encryption, and other security procedures, increasing resistance to security threats.
- Incident Investigation: There are two ways that cloud storage may assist speed up the incident investigation process.
- Forensic readiness: IaaS providers can create and deploy a specialized forensic server version in the same cloud. When a security breach occurs, you may have the server up and running almost immediately! You can skip all of the time-consuming hardware provisioning by adopting cloud computing, and you can have your server whenever you need it.
- Evidence transmission time between the hacked server and the forensics server will be lowered since data transit between two servers in same cloud is incredibly quick.

- Logging: For logging, cloud storage separates clumsy setup.
- Automatic logging: Logs are crucial in security examinations and the establishment of new defensive architecture. In many businesses, however, recording is an afterthought. There are frequently few or no logs due to clumsy resource allocation. Cloud storage alters all of this by enabling automated logging for your cloud-based apps.
- Gold Pictures: Cloud storage makes it easier to install and maintain gold images.

1.7.2 Security recommendations and risks

Despite the numerous benefits that cloud computing may potentially provide to a business, cloud computing has unique characteristics that necessitate a customized threat assessment in areas such as data security, resilience, and privacy. A few important topics of concern expressed by various sources will be discussed in this section. This part should serve as a starting point for some security issues that businesses may have with cloud computing, and these aspects should always be taken into account before beginning a cloud computing connection.

- Privileged user access: Sensitive information generated outside of an organization's internal networks has an inherent risk, since internet services circumvent the physical control that an organization would otherwise have. We have no idea who has immediate access to your information. It might be public cloud executives, system administrators, or even hardware department staff. These responsibilities are inescapable in cloud infrastructures, and they pose potential hazards. Negligence or malevolent individuals might cause data disclosure from the inside.
- Data security: Data security refers to the physical security of data through access mechanisms and encryption. Since many cloud providers still utilize the traditional authentication mechanism of username/password, which is a poor secure paradigm, this is a security issue. Furthermore, it lacks any degree of granularity, allowing various levels of accessibility to be granted to different persons.
- The breakdown of methods that separate storage, memory, or other assets between various clients falls under this risk category. Attacks against resource isolation techniques, on the other hand, are less common and, in principle, far more complex for an attacker to execute than attacks on typical operating systems.
- Sanitization refers to the process of properly removing data from a device after it has been inactive for a period of time. In this case, there are two outcomes that may be troublesome. The first scenario involves the removal of hardware due to failure. If the commodity hardware breaks and is discarded, your data may be recovered from the discarded hardware if it is not adequately cleaned before disposal. The second situation occurs when customers cancel their service with a certain cloud provider. Remaining data in the cloud is frequently not destroyed and is retained on purpose.

- Data location: Because of the cloud's scattered structure, you're unlikely to know where its data is stored. You may not even be aware of the country in which it will be stored. You should inquire if providers will commit to keeping and processing data in specified locations and, as a result, jurisdictions, and if your contract allows them to comply with local privacy laws on behalf of their clients. This technique is necessary since various nations may have dramatically varied jurisdictions and processes, which can make things more complicated when security problems occur.
- Data loss and recovery: In the event of data loss, a cloud provider should be upfront about what will occur to the data and service. While the likelihood of a total loss is minimal due to high levels of duplication, things can still go wrong. There's a chance that your data will be lost on numerous servers at same time. As a result, it's critical to keep an eye on agreements with cloud service providers to check if they're transparent about their data recovery methods, such as how long it'll take and how much it'll cost.
- Investigative support: We discussed some of the advantages of IT forensics while utilizing cloud computing in the previous chapter, but detecting inappropriate or illegal activities might be challenging owing to several technical aspects of cloud service providers. If you can't come to an agreement with the suppliers to help you in particular types of research, then research and disclosure procedures are going to be extremely difficult, if not impossible. Before beginning the service, any company should study the fine print on investigative assistance.
- Even if a cloud provider is eager to assist with investigation procedures, they may not be competent of doing so in a timely and effective manner. The cloud's intricacy can make this technique difficult to understand. For instance, one IaaS provider claimed that it took them eight hours to notice and respond to a Denial of Service assault. Before signing a contract with any cloud service, make sure you understand and negotiate the incident response protocols.

1.7.3 Protective recommendation

Cloud computing poses several security threats and problems for IT workers, particularly for public clouds whose architecture and computation power are held by a third party that offers services to the general population. Nonetheless, certain precautions may be taken, and we've included some of the recommendations made by the US National Institute of Standards and Technology (NIST) for government agencies and departments with security requirements as rigorous as major corporations. We've also included some information that may be of interest to the general audience.

- Extend organizational processes to include cloud-specific rules, procedures, and standards. Development, installation, testing, and supervision of delivered or engaged services are examples. To guarantee that organizational practices are followed throughout the system lifespan, audit methods and tools should be created.

- Understand the many types of regulations and laws that place security and privacy requirements on the company, especially those that deal with data location, privacy, and security measures. There's also a need to analyze and appraise cloud provider products in terms of the organizational criteria to be met, and make sure that the terms are suitable.
- Promote contractual transparency by allowing insight into cloud providers' security and privacy policies and procedures, as well as their effectiveness in a constantly changing and transforming risk landscape.
- Examine the cloud provider's security policies, such as access restrictions, encryption techniques, and data handling protocols, and analyze the risks. More significantly, avoid storing data in the cloud.
- To identify the quickest approach to settle an event, comprehend, and execute the contract rules and processes for incident handling required by the business.

Check your progress 4

Note: a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i) What are the protective recommendations against cyber attacks in cloud?

……

……

……

……

……

……

……

……

1.8 Let Us Sum Up

This unit covers the fundamentals of cloud computing, as well as the security aspect of a cloud server. This unit is focused on types of cloud computing providers, generic cloud architecture, and file storage in the cloud. In the later sections, the security recommendations, and risks are discussed along with protective recommendations.

1.9 Session wise- list of lab assignments

Now, we try solving the problems related to cloud server implementation.

Session -1

a) Write down a script that will shutdown all the services manually.

Session -2

a) Study about the Cloud simulation tools. Install CloudSim and analyze the working of CloudSim simulator.

Session -3

a) Study about the Service models with real time examples.

Session -4

a) Study about Management of Cloud Computing Resources.

Session -5

a) Installation and Configuration of Oracle Virtual Box/VMware.

Session -6

a) Installation and Configuration of Single Node Hadoop Cluster.

Session -7

a) Discuss performance evaluation of service over cloud.

Session -8

a) A case study on Google App Engine, Microsoft Azure, Amazon.

Session -9

a) A case study on salesforce.com.

Session -10

a) Create a scenario that uses Amazon S3 as storage on cloud.

1.10 Check your Progress: The Key

1. (i) Below are the five principles of Cloud Computing:

- Resource Provisioning: Cloud computing companies take advantage of significant economies of scale by pooling their resources. They assemble a large number of servers and hard drives and implement the same settings, security, and other features to all of them.
- Elasticity: Adding extra hard drive space or server connectivity on-demand can be accomplished with only a few mouse clicks. Cloud computing also offers geographic scalability, with the option of replicating data across many data centers across the world.

- Automatic resource deployment: The user merely needs to pick the types and characteristics of the services he requires, and the cloud services provider will set up and hook them up for him effortlessly.
- Virtualization: Customers do not have to be concerned about the physical status of their gear, nor do they have to be concerned about device interoperability, thanks to virtualization.
- Pay what you use: Consumers are only charged for what they use with metered billing.

2. (i) The hypervisor has evolved as an indispensable tool for operating virtual machines and promoting innovation in a cloud context as cloud computing becomes more widespread. Hypervisors are a fundamental component of the technology that allows cloud computing because they are a software layer that allows one host computer to handle numerous virtual machines at the same time. Hypervisors enable IT to keep control over a cloud environment's infrastructure, apps, and sensitive data while making cloud-based services available to the user across a virtual environment. By abstracting a computer's software from its hardware, hypervisors enable the development and control of virtual machines (VMs). Virtualization is made feasible by hypervisors, which translate requests between physical and virtual resources. To allow the operating system on a computer to retrieve and use virtualization software, bare-metal hypervisors are often built in the firmware at the same level as the motherboard basic input/output system (BIOS).

3. (i) Below are the three layers of cloud architecture:
   - The front end is responsible for client-server communication. Different APIs will be available to access the real storage. This layer is also concerned with attaining outcomes such as multi-tenancy, which we shall define in the following chapter. Furthermore, it offers a variety of scalability options via a variety of techniques.
   - The storage logic layer is in charge of a number of features and administrative tasks, such as guaranteeing a high level of accessibility and dependability, for example. It's also a type of security fence. In addition, it serves as a cloud storage administrator.
   - The back-end is responsible for the actual deployment of physical data storage technologies such as GFS (Google File System). It entails a variety of techniques for increasing storage efficiency and decreasing infrastructure expenses.

4. (i) Below are the protective recommendations against the cyber attacks in cloud:

- Extend organizational processes to include cloud-specific rules, procedures, and standards. Development, installation, testing, and supervision of delivered or engaged services are examples. To guarantee that organizational practices are followed throughout the system lifespan, audit methods and tools should be created.

- Understand the many types of regulations and laws that place security and privacy requirements on the company, especially those that deal with data location, privacy, and security measures. There's also a need to analyze and appraise cloud provider products in terms of the organizational criteria to be met, and make sure that the terms are suitable.
- Promote contractual transparency by allowing insight into cloud providers' security and privacy policies and procedures, as well as their effectiveness in a constantly changing and transforming risk landscape.
- Examine the cloud provider's security policies, such as access restrictions, encryption techniques, and data handling protocols, and analyze the risks. More significantly, avoid storing data in the cloud.
- To identify the quickest approach to settle an event, comprehend and execute the contract rules and processes for incident handling required by the business.

Structure

2.0 Introduction

In many respects, the introduction of information technology has revolutionized the way we operate, and audit is no exception. However, clearly one of the most successful business tools, the now practically ubiquitous Computer has also brought with its weaknesses in the automated company environment. Credentials and authentication numbers that control access to electronic files, while the paper and pen of mechanical operations have made way for online data entry of computerized applications have substituted the locks and keys of filing systems. Each new vulnerability must be controlled, and determining the efficacy of each control necessitates the use of new auditing methodologies. Most federal agencies, government sector companies, and non-governmental organizations have heavily invested in computerizing their activities during the previous decade. Owing to excessive purchase and operating expenses, computers were first only applicable to large enterprises. Later, with the introduction of the computer and the quick fall in costs, medium-sized businesses were able to use Information Technology for data handling as well. The increasing availability of strong minicomputers and their related bundled software has led to widespread computer adoption by even small businesses. As a result, the chances of data loss and the accompanying organizational expenses have skyrocketed, along with additional risk factors.

2.1 Learning outcome

This is the second unit of this course, which seeks to introduce the concept of IT Auditing and explain the various aspect of penetration testing. This unit covers the introduction to penetration testing, it's types and the steps involved in the penetration testing. Thereafter, the labs related to reconnaissance and footprinting is added in this unit. Tools like NIKTO, theHarvester, and Maltego are discussed under the reconnaissance part. After that, the second phase of hacking i.e. Scanning and Enumeration is discussed and the tools like dnsenum and dig is discussed. This usnit also covers the concept of Man in the middle attack using ARP cache poisoning. Moreover, automated SQL injection using SQLMap is also the part of this unit. At last, the lab on gaining the access of a metasploitable machine using the metasploit framework is added.

2.2 Introduction to IT Auditing

IT audit is a general term that encompasses Forensic Audits, Operational Audits, Information Systems Audit, such as Performance Audit, Specialized Audits evaluating services offered by a third party, such as outsourcing, and Financial Audits to determine the accuracy of an organization's financial statements. However, forming a judgment on the level of dependence that may be put on the IT systems in the inspected firm is a common feature.

The work of both internal and external auditors is affected by the introduction of a computer. They should be familiar with the computer inputs, processing and output in order that they may conduct test checks with understanding. Auditors should also know about some programming languages. The documentation procedures may consist of-

a) Procedure narrative.
b) Flowcharts.
c) Program code with print-outs.
d) Operating instructions for the computer operator.
e) Error routines.
f) Log sheets for recording computer operating and downtime.
g) Program maintenance.

2.2.1 Steps in IT Auditing

The following stages must always be included in the IT audit process:

- Planning
- Defining audit objectives and scope

- Assessing controls
- Gathering and evaluating evidence
- Reporting, and
- Following up.

Check your progress 1

Note: a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

i) What are the steps involved in IT Auditing?

.....

.....

.....

.....

.....

.....

2.3 Introduction to Penetration Testing

Penetration Testing is the process of hacking a system with permission from the owner of that system, to evaluate security, hack value, attacks, exploits, zero-day vulnerabilities, & other components such as threats, vulnerabilities, and daisy chaining. There are basically four types of penetration testing mentioned below:

- Tiger box
    - Collection of OSs and hacking tools.
    - Usually on a laptop.

    - Helps penetration testers and security testers conduct vulnerabilities assessments and attacks.
- White box model
    - Tester is told everything about the network topology and technology.
    - Tester is authorized to interview IT personnel and company employees.
    - Makes tester's job a little easier.
- Black box model
    - Tester is not given details about the network.
    - Burden is on the tester to find the details.
- Gray box model
    - Hybrid of the white and black box models.
    - Company gives tester partial information.

2.3.1 Steps involved in Penetration Testing

Penetration Testing is five step process, in which, firstly we find a vulnerability, then we design the attack (based on the vulnerability identified), and thereafter we appoint a team of ethical hackers to do the penetration testing, after that the ethical hackers will try to find out the kind of data they could steal. Finally, based on the report generated the penetration testers will act on the findings. Figure 1 shows the penetration testing steps.

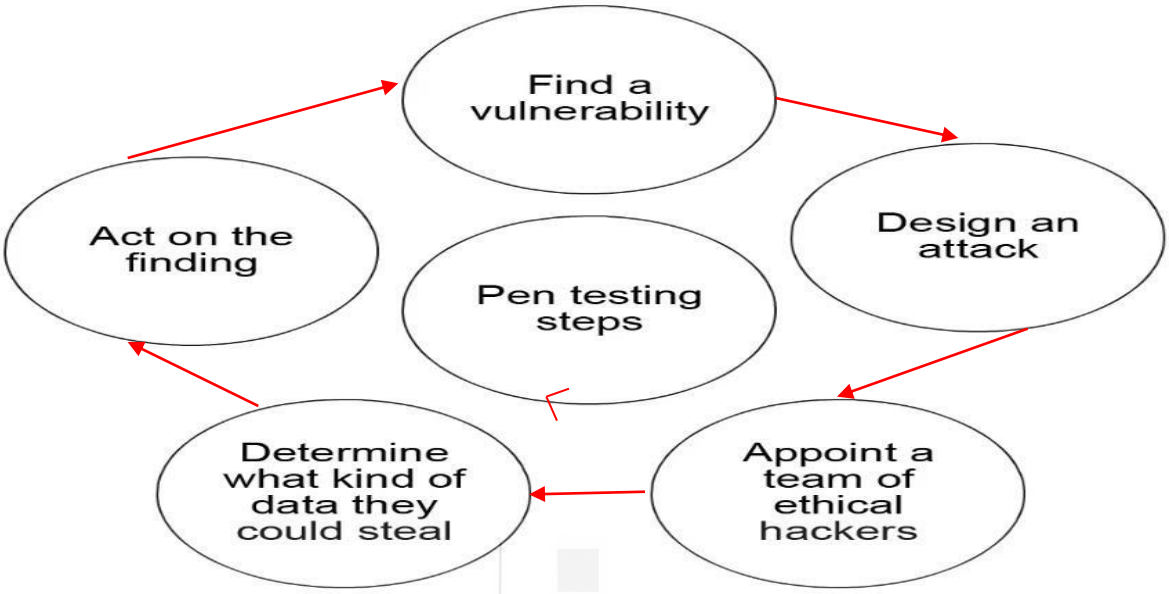


Figure 1: Diagram for Platform Virtualization

2.4 Reconnaissance and Footprinting

Reconnaissance means narrowing down the search to specific targets, tools and techniques. Footprinting is a technique for gathering information on computer systems and the entities they belong to. There are two types of reconnaissance: active and passive reconnaissance. In active reconnaissance, the attacker directly interact with the target by connecting to it whereas in passive reconnaissance, there is non-interactivity with the target, the attacker identifies the indistinguishable public traffic that the target is using.

2.4.1 Footprinting using NIKTO and theHarvester Footpriting using NIKTO

In this lab, we will use the Nikto tool to gather information from a website. Specifically, we are looking for possible vulnerabilities on the website.

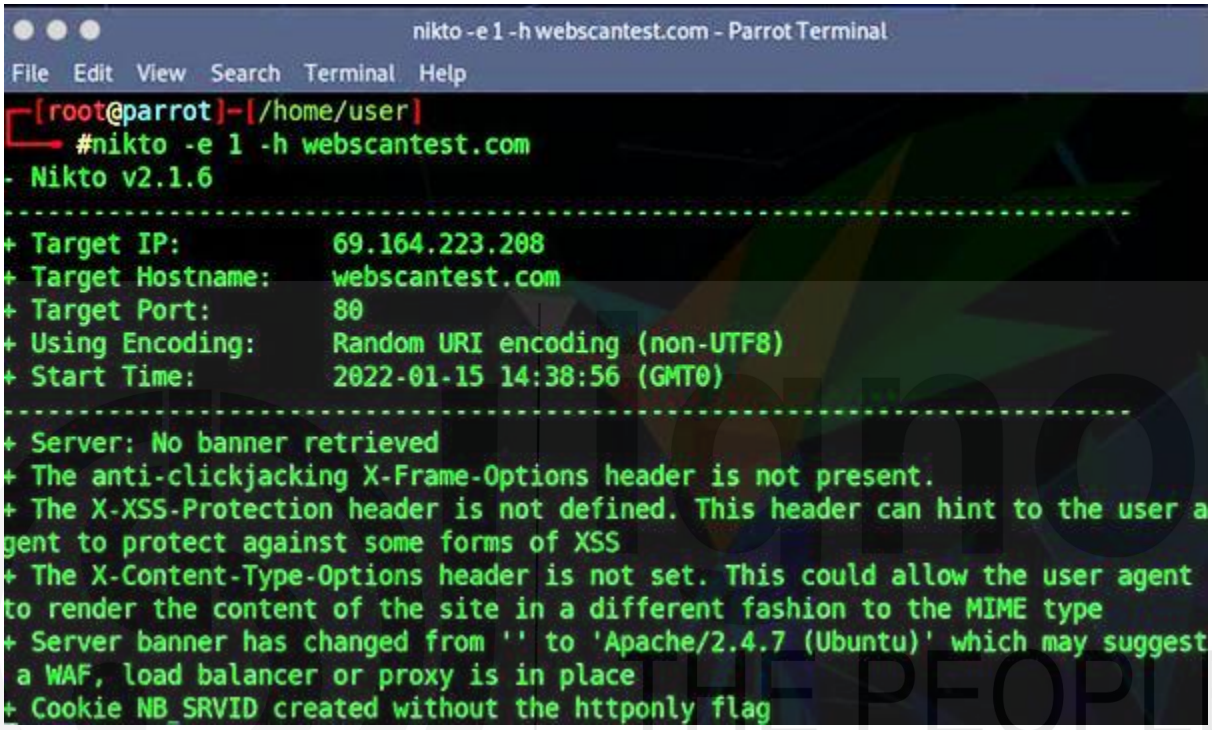
Requirements for the lab: Kali Linux / Parrot OS

Step 1: Launch your Kali virtual machine and log in

Step 2: Launch a Terminal window

Step 3: At the prompt, enter: nikto -e 1 -h webscantest.com

**We are using the evasion switch (-e) and the number 1 (to specify random encoding) to help us be a little bit stealthier when running the scan. We also use -h to define the hostname or IP address. Note: It will take several minutes to run the scan.



Check your progress 2

a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i) Do you see any possible vulnerabilities from the website? If yes, please write some of them below:

.....

.....

.....

Footprinting using theHarvester

In this lab, we are going to to use a tool called theHarvester. This tool is useful for gathering information on subdomains, employee names, emails, open ports, and banners. It gathers the information from public sources, like regular search engines and Shodan. You can read more about it here: https://tools.kali.org/information-gathering/theharvester
Requirements for the lab: Kali Linux / Parrot OS

Step 1: Launch you Kali / Parrot OS machine and login
Step 2: Open a Terminal window
Step 3: At the prompt, type theharvester -h to view the help file for the tool.

Step 4: We are just going to use one of the example statements from the help section.

Please type this command: theharvester -d microsoft.com -l 50 -b google -s -d is the domain or company name to search.

Check your progress 3

a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i). Were you able to see any IP addresses in the output? If yes, write at least two of them here

……

……

……

ii). What other information did you find?

……

……

2.4.2 Footprinting using Maltego

In this lab, we are going to to use a tool called Maltego.

Maltego is an open source intelligence forensic application. Which will help you to get more accurate information and in a smarter way.

In simple words, it is an information-gathering tool.

You can read more about it here:

https://tools.kali.org/information-gathering/maltego-teeth

Requirements for the lab: Kali Linux / Parrot OS

Step 1: Launch you Kali / Parrot OS machine and login

Step 2: Open a Maltego from tools available in Kali and Parrot OS

Step 3: At the prompt, run the CE version of Maltego. . In this experiment, we will be using haveibeenpwned.com database. So, make sure that this integration is available with Maltego.

Step 4: Create a new graph, select the domain as entity from entity panel and the use any domain like linkedin.com (Note: Don't use it on some govt. domains). After that right click on the domain that you have chosen as a target and then run all the transformations related to e-mails.

Step 5: After you get all the email ids, select them and run breach transformation.

Step 6: After you get all the breaches, select anyone to enrich it by running the enrich transformation.

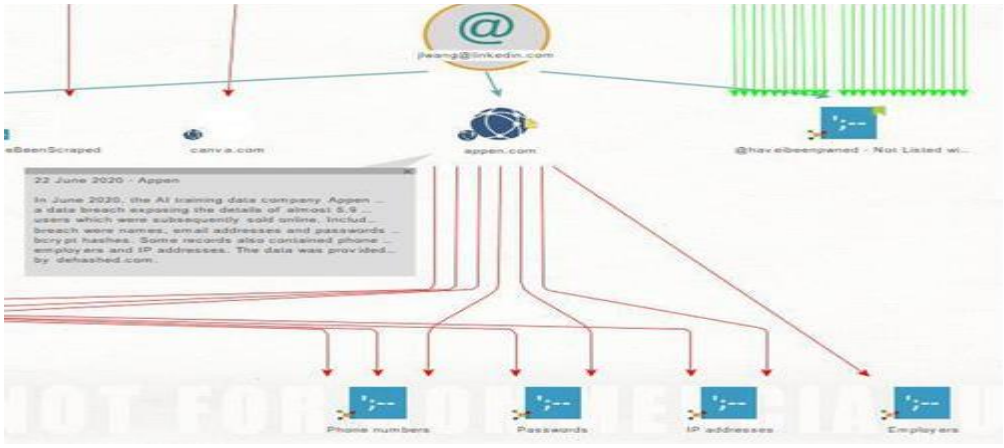


Figure: Footprinting using Maltego

Check your progress

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

1. Were you able to see any breached related to any email id? If yes, write at least three of them here

……

……

……

2. What other information did you find?

……

……

2.5 Scanning and Enumeration

Network scanning is a method of getting network information such as identification of hosts, port information, and services by scanning networks and ports.

2.5.1 Scanning using NMAP

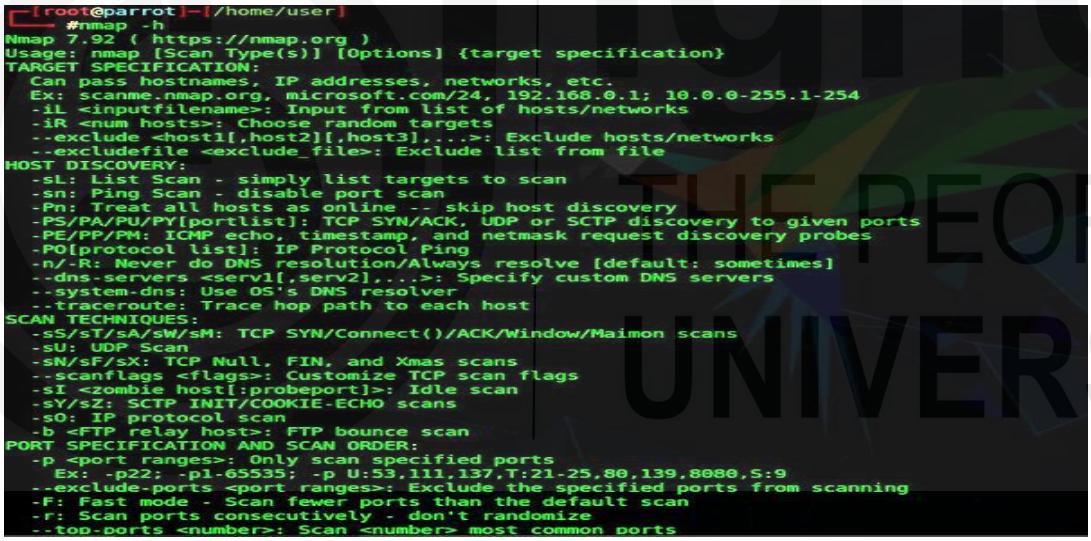
In this lab, we will use the NMAP tool for scanning the network. Nmap has become one of the most popular tools in network scanning by leaving other scanners behind. Many times the hosts in some organizations are secured using {firewalls or intrusion prevention systems which result in the failure of scanning due to the present set of rules which are used to block network traffic. In Nmap, a pentester can easily make use of alternate host discovery techniques to prevent this from happening. It consists of certain features that make the network traffic a little less suspicious. Hence, let us look at various techniques of Host Discovery.

Requirements for the lab: Kali Linux / Parrot OS

Step 1: Launch your Kali virtual machine and log in

Step 2: Launch a Terminal window.

Step 3: At the prompt, enter: nmap –h



Step 4: From the help page explore the various options of nmap on scanme.nmap.org or create a separate metasploitable machine and use that as a target.

```
  --top-ports <number>: Scan <number> most common ports
  --port-ratio <ratio>: Scan ports more common than <ratio>
SERVICE/VERSION DETECTION:
  -sV: Probe open ports to determine service/version info
  --version-intensity <level>: Set from 0 (light) to 9 (try all probes)
  --version-light: Limit to most likely probes (intensity 2)
  --version-all: Try every single probe (intensity 9)
  --version-trace: Show detailed version scan activity (for debugging)
SCRIPT SCAN:
  -sC: equivalent to --script=default
  --script=<Lua scripts>: <Lua scripts> is a comma separated list of
          directories, script-files or script-categories
  --script-args=<n1=v1,[n2=v2,...]>: provide arguments to scripts
  --script-args-file=filename: provide NSE script args in a file
  --script-trace: Show all data sent and received
  --script-updatedb: Update the script database.
  --script-help=<Lua scripts>: Show help about scripts.
          <Lua scripts> is a comma-separated list of script-files or
          script-categories.
OS DETECTION:
  -O: Enable OS detection
  --osscan-limit: Limit OS detection to promising targets
  --osscan-guess: Guess OS more aggressively
TIMING AND PERFORMANCE:
  Options which take <time> are in seconds, or append 'ms' (milliseconds),
  's' (seconds), 'm' (minutes), or 'h' (hours) to the value (e.g. 30m).
  -T<0-5>: Set timing template (higher is faster)
  --min-hostgroup/max-hostgroup <size>: Parallel host scan group sizes
  --min-parallelism/max-parallelism <numprobes>: Probe parallelization
  --min-rtt-timeout/max-rtt-timeout/initial-rtt-timeout <time>: Specifies
      probe round trip time.
  --max-retries <tries>: Caps number of port scan probe retransmissions.
  --host-timeout <time>: Give up on target after this long
  --scan-delay/--max-scan-delay <time>: Adjust delay between probes
```

Step 5: Run the various options listed below:

Ping Sweep (No port scan)

```
┌─[root@parrot]─[/home/user]
└──╼ #nmap -sP scanme.nmap.org
Starting Nmap 7.92 ( https://nmap.org ) at 2022-02-05 13:55 GMT
Nmap scan report for scanme.nmap.org (45.33.32.156)
Host is up (0.0022s latency).
Other addresses for scanme.nmap.org (not scanned): 2600:3c01::f03c:91ff:fe18:bb2f
Nmap done: 1 IP address (1 host up) scanned in 5.49 seconds
```

Types of Scans

UDP Ping Scan

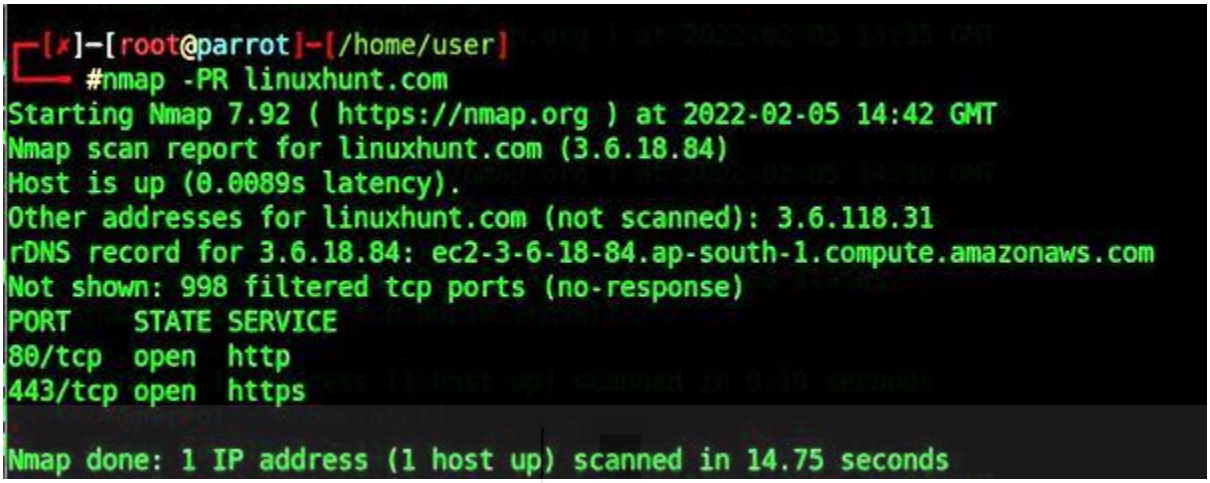
```
┌─[root@parrot]─[/home/user]
└──╼ #nmap -sU 10.0.2.15
Starting Nmap 7.92 ( https://nmap.org ) at 2022-02-05 14:28 GMT
Nmap scan report for 10.0.2.15
Host is up (0.0040s latency).
All 1000 scanned ports on 10.0.2.15 are in ignored states.
Not shown: 1000 closed udp ports (port-unreach)

Nmap done: 1 IP address (1 host up) scanned in 0.62 seconds
```

```
┌─[root@parrot]─[/home/user]
└──╼ #nmap -sU scanme.nmap.org
Starting Nmap 7.92 ( https://nmap.org ) at 2022-02-05 14:12 GMT
Nmap scan report for scanme.nmap.org (45.33.32.156)
Host is up (0.034s latency).
Other addresses for scanme.nmap.org (not scanned): 2600:3c01::f03c:91ff:fe18:bb2f
Not shown: 999 open|filtered udp ports (no-response)
PORT    STATE SERVICE
123/udp open  ntp

Nmap done: 1 IP address (1 host up) scanned in 1587.73 seconds
```

ARP Ping Scan



Check your progress 5

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

i) Do you see any open ports? If yes, please writesome of them below:

……

……

……

……

……

……

2.5.2 DNS Enumeration and NetBIOS Enumeration

In this lab, we will do Enumeration. Enumeration is defined as the process of extracting user names, machine names, network resources, shares and services from a system. Actually, there are various types of enumeration techniques, but the major ones are DNS Enumeration and NetBIOS enumeration. DNS enumeration is the process of locating all the DNS servers and their corresponding records for an organization. DNS enumeration will yield usernames, computer names, and IP addresses of potential target

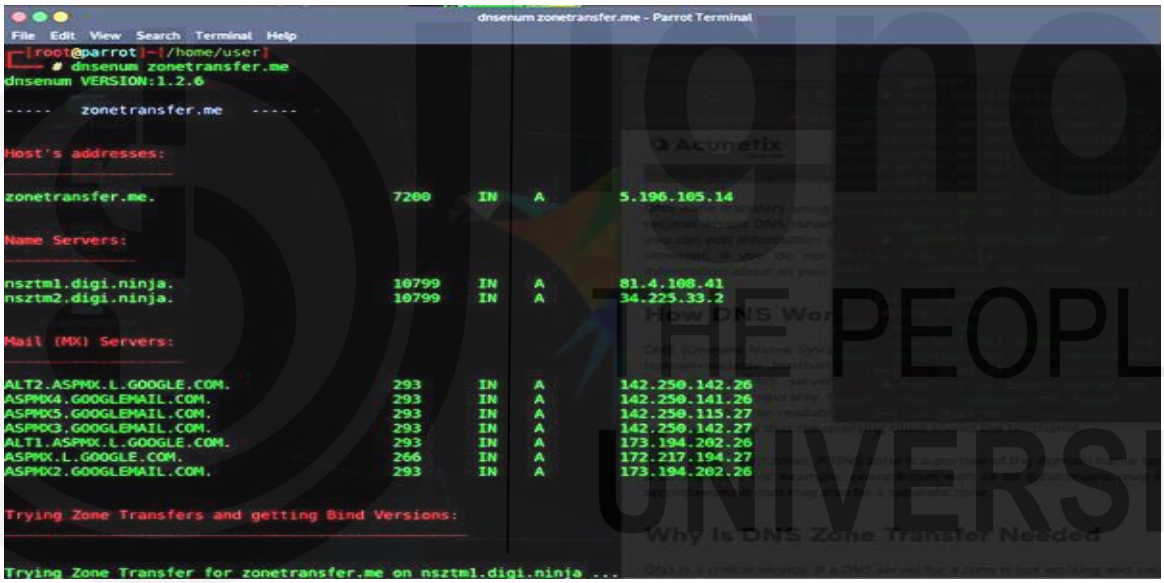
systems. The list of DNS record provides an overview of types of resource records (database records) stored in the zone files of the Domain Name System (DNS). The DNS implements a distributed, hierarchical, and redundant database for information associated with Internet domain names and addresses. DNS Zone Transfer is a process where a DNS server passes a copy of part of its data base (which is called a "zone") to another DNS server. DNS Zone Transfer used to replicate DNS data across a number of DNS servers or to back up DNS files. A user or server will perform a specific zone transfer request from a —name server. If the name server allows zone transfers by an anonymous user to occur, all the DNS names and IP addresses hosted by the name server will be returned in human-readable ASCII text.

Requirements for the lab: Kali Linux / Parrot OS

Step 1: Launch your Kali virtual machine and log in

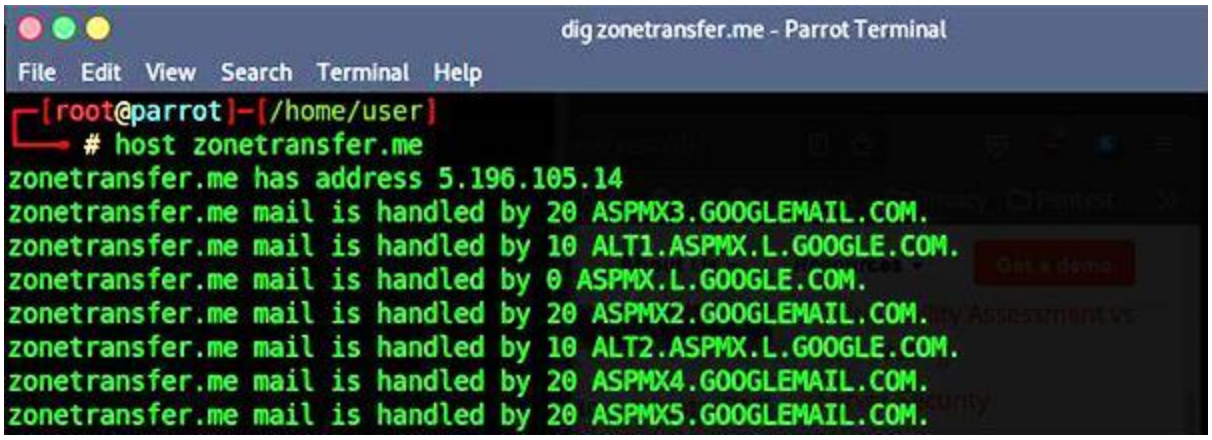
Step 2: Launch a Terminal window

Step 3: At the prompt, enter: dnsenum zonetransfer.me



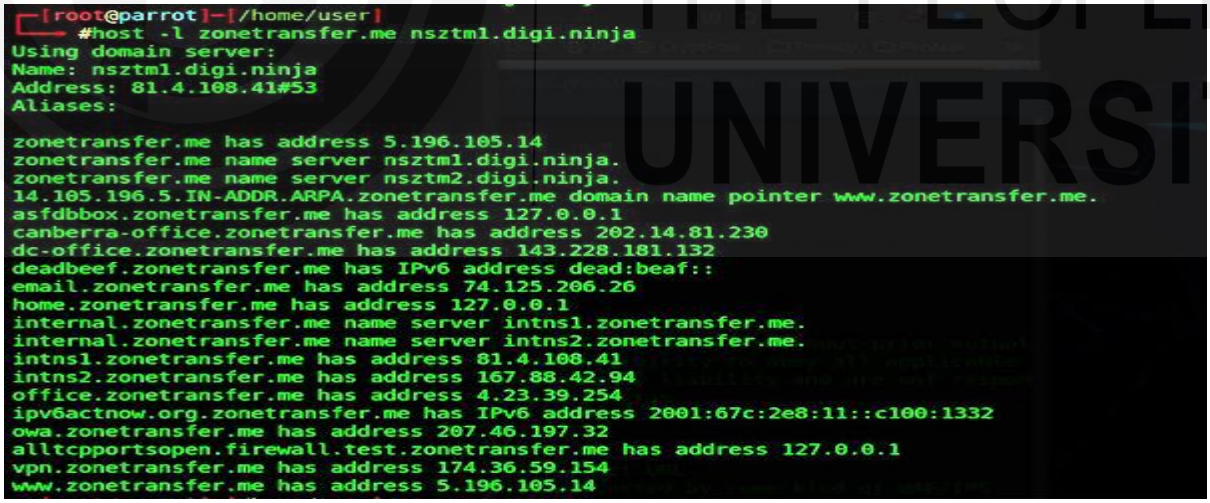
Step 4: Once this is done, write down your observations and answer questions given below, based on your observations.

Step 5: Now, you can use a different command to give some more information, open a separate terminal and type host zonetransfer.me



Step 6: Now, to get information about the name server, type in terminal - host –t ns zonetransfer.me

Step 7: Now, in order to perform zone transfer, write down the name of the name server that you got from the previous command, and type in terminal – host –l zonetransfer.me nsztm1.digi.ninja (The name of the name server is written at last).



Step 8: Now, we can also use another command to do the zone transfer and get some useful information, open a separate terminal and type dig zonetransfer.me NAME OF THE NAME SERVER (e.g. nsztm1.digi.ninja)

```
┌──[root@parrot]─[/home/user]
└──# dig zonetransfer.me

; <<>> DiG 9.16.22-Debian <<>> zonetransfer.me
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 24340
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 1280
;; QUESTION SECTION:
;zonetransfer.me.               IN      A

;; ANSWER SECTION:
zonetransfer.me.        6114    IN      A       5.196.105.14

;; Query time: 112 msec
;; SERVER: 192.168.103.228#53(192.168.103.228)
;; WHEN: Fri Mar 04 11:02:27 GMT 2022
;; MSG SIZE  rcvd: 60
```

Step 9: Now, in order to get some more information about the target, the attacker may use nslookup command. Open a separate terminal and type nslookup zonetransfer.me Note: It will take several minutes to run the scan.

```
┌──[root@parrot]─[/home/user]
└──# nslookup zonetransfer.me

Server:         192.168.103.228
Address:        192.168.103.228#53

Non-authoritative answer:
Name:   zonetransfer.me
Address: 5.196.105.14
```

Check your progress 6

a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i) Do you see any crucial information after doing DNS enumeration? If yes, please write some of them below:

.....

……

……

ii). Do you see any name servers? If yes, please write some of them below:?

……

……

……

2.6 Man-in-the-middle (MITM) attack – ARP Poisoning

In this lab, we will do Man in the middle attack (MITM) using Ettercap. Ettercap is a comprehensive suite for man in the middle attacks. It features sniffing of live connections, content filtering on the fly and many other interesting tricks. It supports active and passive dissection of many protocols and includes many features for network and host analysis.

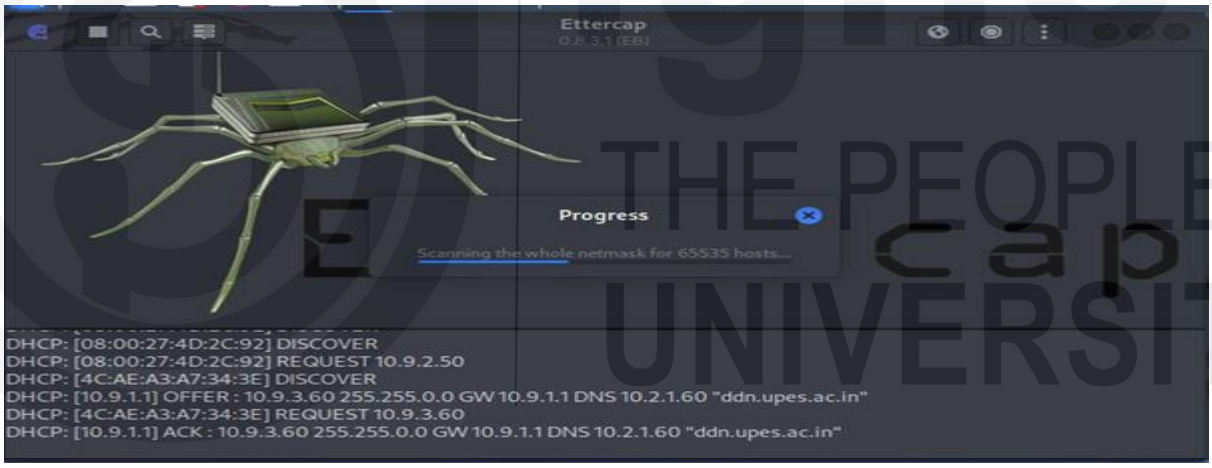
Requirements for the lab: Attacker Machine - Kali Linux /
Parrot OS. Victim / target Machine– Windows Machine
(Win 7/ Win 8)

Step 1: Launch your attacker machine (Parrot virtual machine) and log in

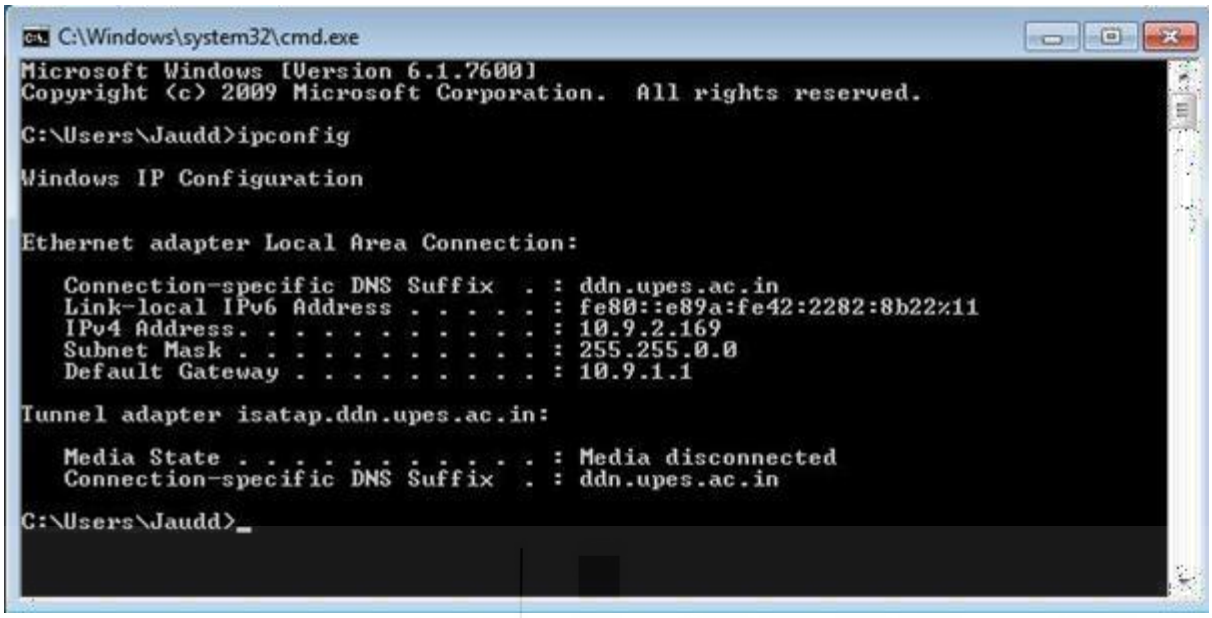
Step 2: Go to Applications -> Pentesting -> Sniffing & Spoofing and then launch Ettercap graphical

Step 3: At the prompt, it will ask for the adapter that you want to choose, you may choose unified or bridged sniffing and then click on start.

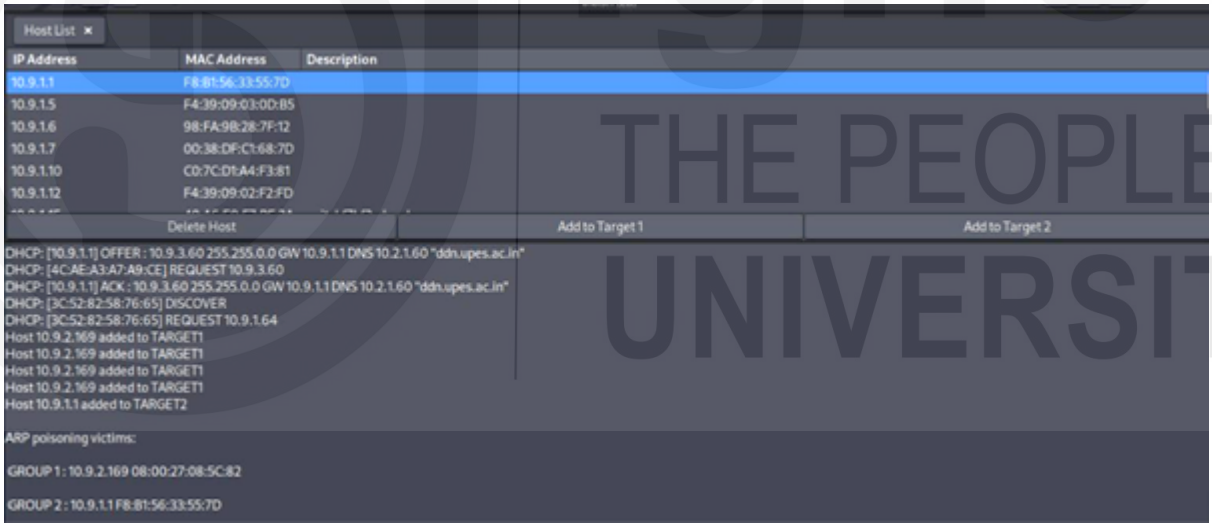
Step 4: Once this is done, scan for the hosts, it will list you the IP address of the default gateway as well as the IP address of the target machine. Make sure that your target machine is up and running before this step.



Step 5: Now, from the lists of the hosts add the targets, add IP address of gateway as target 1 and IP address of victim as the target 2.

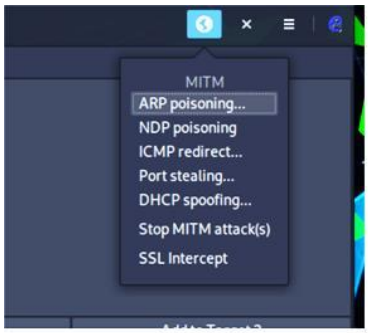


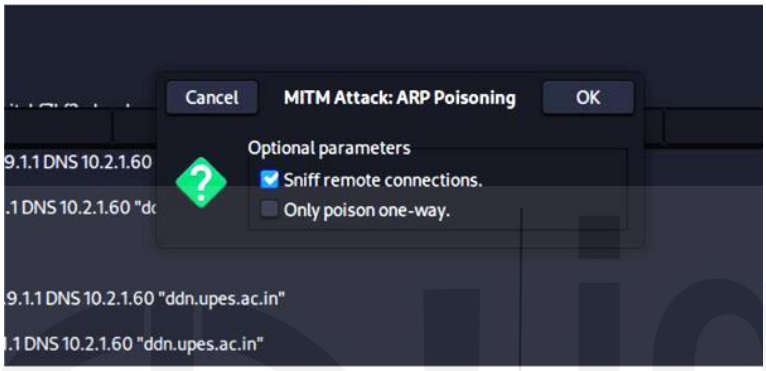
In this we have both default gateway address and IP address of target machine.



Here we added both the Targets

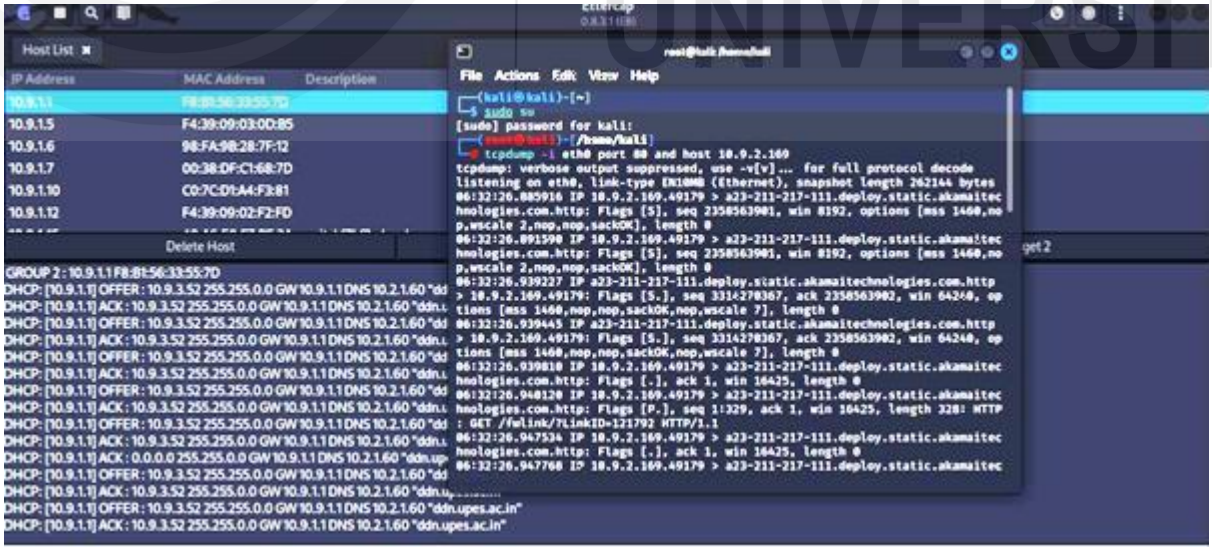
Step 6: Now, from MITM Menu select the ARP Poisoning, and make sure that Sniff remote connections option is checked.





Step 7: Now, Ettercap has poisoned the ARP tables of the gateway and the target. Now, you can perform the MITM and can sniff the traffic going from the target machine.

Step 8: Now, you can launch any network sniffer like tcpdump or wireshark on the attacker machine. I am using tcpdump, I will open a separate terminal and type: Sudo tcpdump –I eth0 port 80 and host IP_Address_of_target.



Step 9: Now, the MITM attack is performed, any network activity on the target machine can be

observed with the tcpdump. Note: It will take several minutes to run the scan.

Check your progress 7

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

i) Do you see any network traffic after performing MITM on tcpdump? If yes, please write some of them below:

.....

.....

ii) Do you see any other active hosts after scanning the hosts on Ettercap-graphical? If yes, please write me of them below:

2.7 Automated SQL Injection using SQLMap

In this lab, we will perform automated SQL Injection using SQLMap. SQL injection is a code injection technique, used to attack data-driven applications, in which malicious SQL statements are inserted into an entry field for execution (e.g. to dump the database contents to the attacker). SQL injection must exploit a security vulnerability in an application's software, for example, when user input is either incorrectly filtered for string literal escape characters embedded in SQL statements or user input is not strongly typed and unexpectedly executed. SQL injection is mostly known as an attack vector for websites but can be used to attack any type of SQL database. sqlmap is an open source penetration testing tool that automates the process of detecting and exploiting SQL injection flaws and taking over of database servers. It comes with a powerful detection engine, many niche features for the ultimate penetration tester and a broad range of switches lasting from database fingerprinting, over data fetching from the database, to accessing the underlying file system and executing commands on the operating system via out-of-band connections.

Requirements for the lab: Attacker Machine - Kali Linux. Target - http://testphp.vulnweb.com/ Step 1: Launch your attacker machine (Parrot virtual

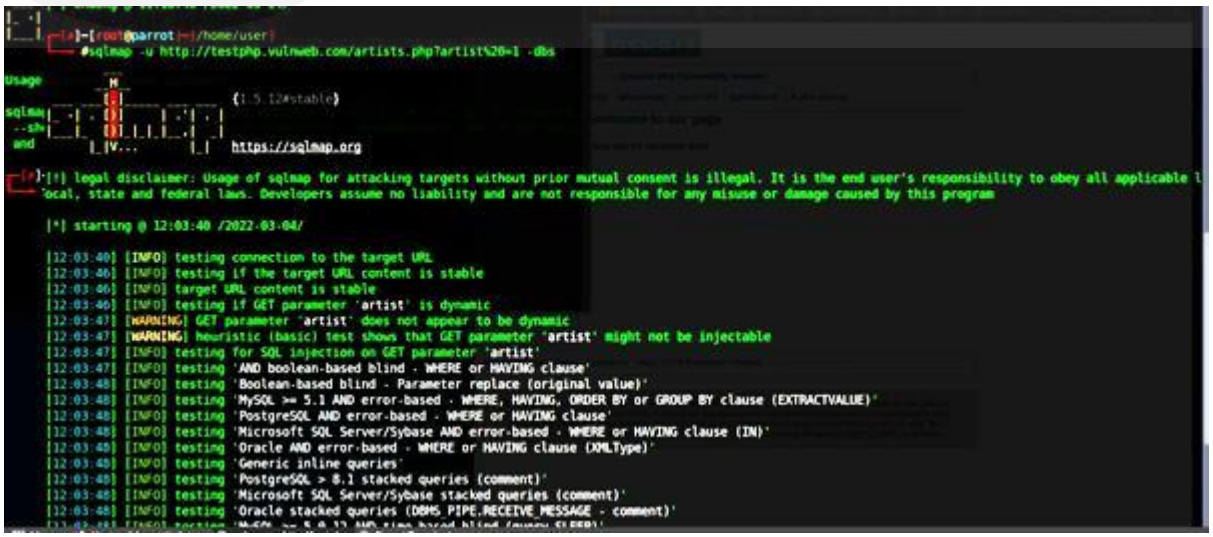
machine) and log in.

Step 2: Open your terminal and type sqlmap to launch the sqlmap.

Step 3: Open a web browser and type - http://testphp.vulnweb.com/ in a tab.

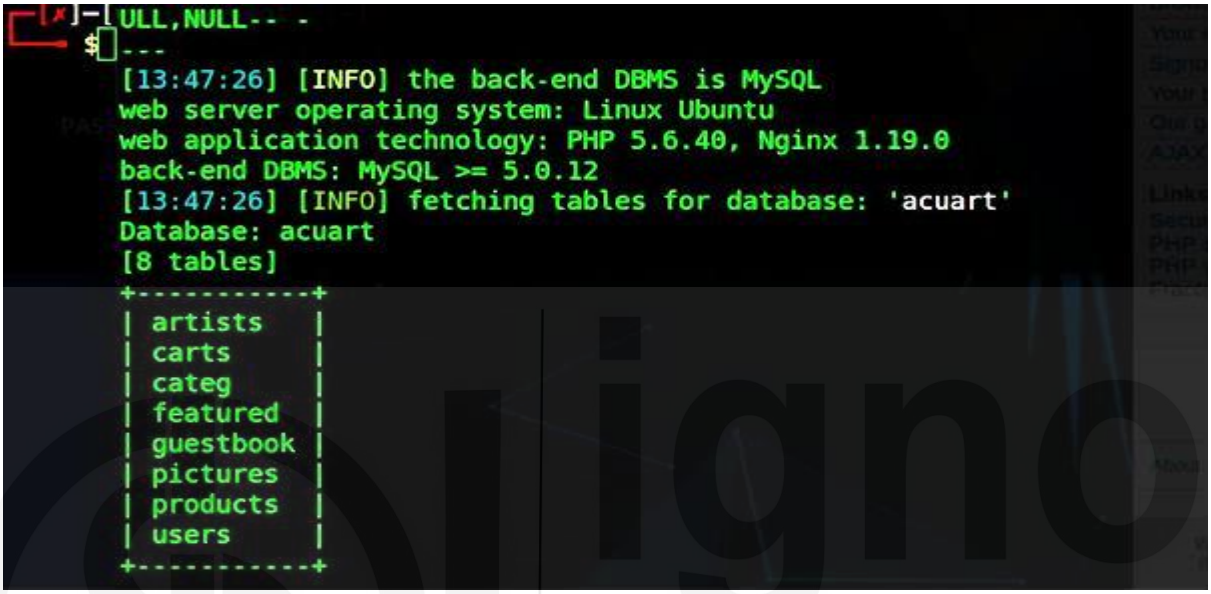
Step 4: After that, find out the link to start the sql injection. For that, you can use google, and type site: http://testphp.vulnweb.com/ php?id= . The first link that has this kind of structure will be your
link for SQL Injection. Example: http://testphp.vulnweb.com/artists.php?artist=1

Step 5: Now, in terminal, type – sqlmap –u http://testphp.vulnweb.com/artists.php?artist=1 –dbs .If it gives you result then the website is vulnerable to SQL Injection.

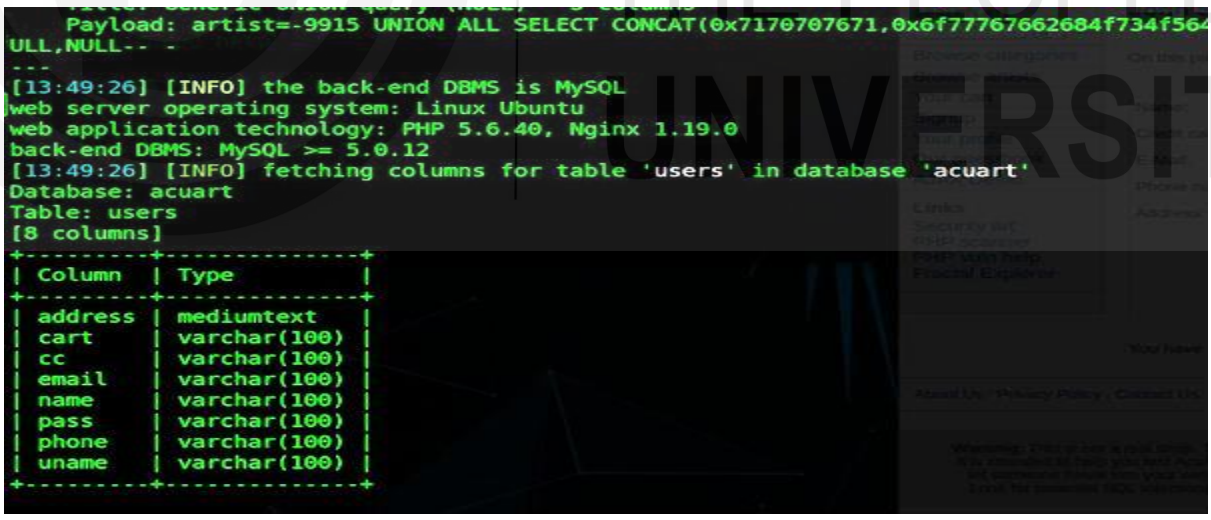


Step 6: Now, you can explore the various options available with sqlmap and find out the tables names, columns name, usernames, passwords and email ids.

These are some of the tables



And these are some of the columns

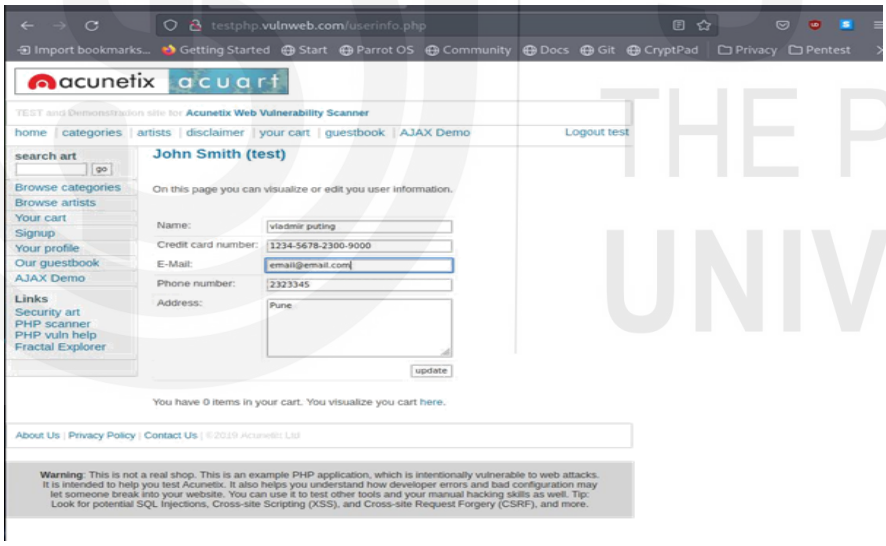


And these are the user credentials

```
[1 entry]
+-------+
| uname |
+-------+
| test  |
+-------+
```

```
Database: acuart
Table: users
[1 entry]
+------+
| pass |
+------+
| test |
+------+
```

Step 7: Use the username and passwords to gain the access of the website.



Check your progress 8

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

i) Do you find any databases and tables using sqlmap? If yes, please write some of them below:

.....

.....

ii) Do you get the access to the database of the target website? What username and password did you get?

If yes, please write some of them below:?

.....

2.8 Gaining Access of Metasploitable machine using Metasploit Framework

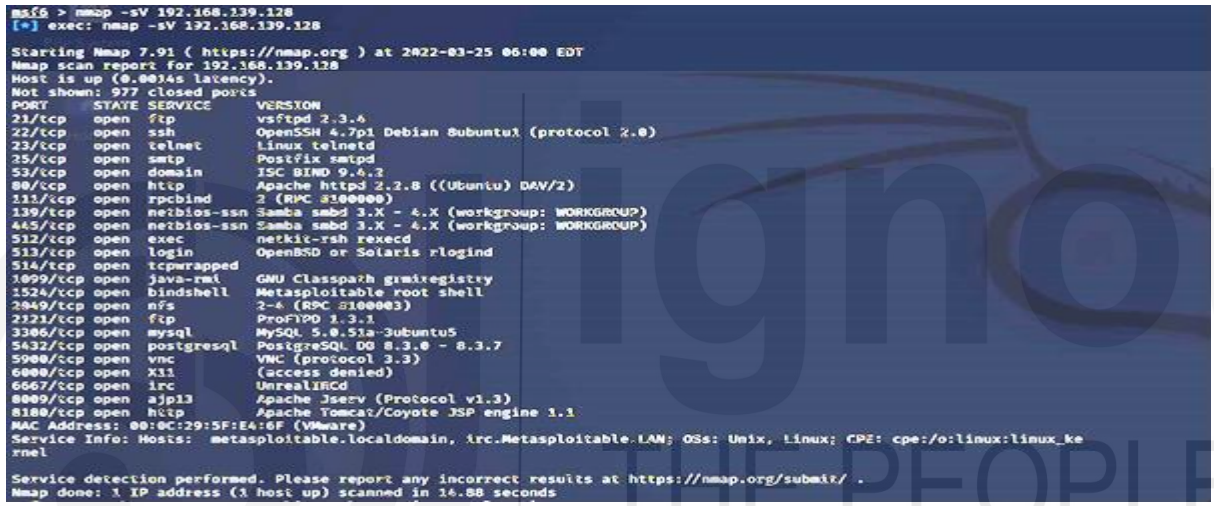
In this lab, we will gain the access of the metasploitable machine using Metasploit Framework. The Metasploit Framework is a Ruby-based, modular penetration testing platform that enables you to write, test, and execute exploit code. The Metasploit Framework contains a suite of tools that you can use to test security vulnerabilities, enumerate networks, execute attacks, and evade detection. At its core, the
Metasploit Framework is a collection of commonly used tools that provide a complete environment for penetration testing and exploit development.
Requirements for the lab: Attacker Machine - Kali Linux /
Parrot OS. Victim / target Machine – Metasploitable VM

Step 1: Launch your attacker machine (Parrot virtual machine) and log in

Step 2: Open your terminal and type msfconsole to launch the metasploit framework. Step 3: Start your target Metasploitable machine and check the ip address using ifconfig.

Step 4: After that, in attacker machine, the msfconsole will launch the metasploit framework and you can explore various functions of metasploit with 'help' command.

Step 5: Now, run the nmap in msfconsole to identify the various services running on the target system. Command: nmap IP_target –sV.



Step 6: Now, you will get the various services running on the target machine. Now, your job is to find out the vulnerable service (e.g. samba) and the exploit corresponding to it. For that, you can type – search name:samba type:exploit platform:unix in your msfconsole terminal. This will give you the list of available

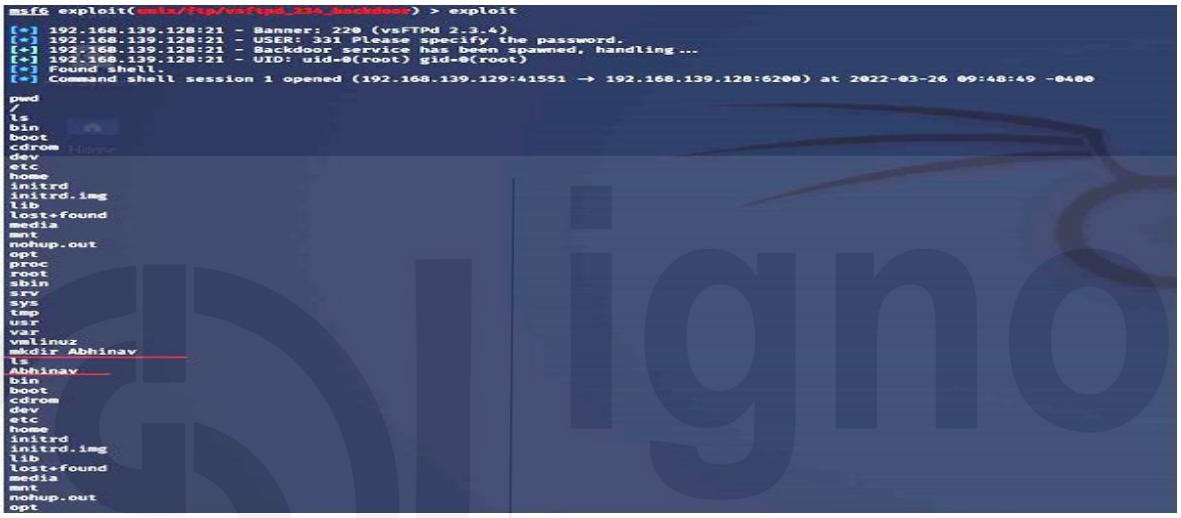
exploit for your target machine. After that you can select the exploit from the list available and can use it by typing – use name_of_exploit.

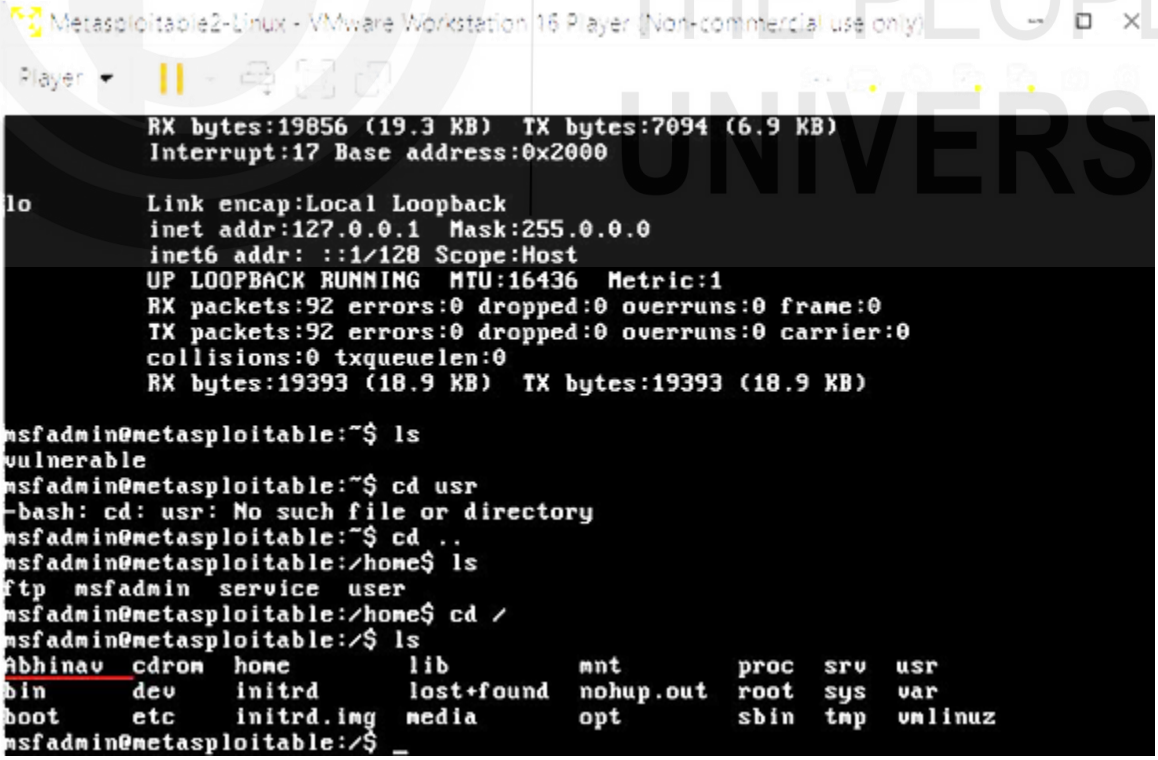
Step 7: After that, you need to set the RHOSTS i.e. the IP address of target. For this, type show options to check various options available with the selected exploit and then set the RHOSTS by

typing – set RHOSTS Target_IP.

```
msf6 exploit(unix/ftp/vsftpd_234_backdoor) > set RHOST 192.168.139.12
8
RHOST => 192.168.139.128
```

Step 8: After that, type exploit and you will have command line access to the target machine. Now, you can run any command on the target machine using the session created by the metasploit. Note: It will take several minutes to run the scan.





Check your progress 9

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

i) Do you various services running on target machine? If yes, please write some of them below:

.....

.....

ii) Do you get the access to the target machine? Are you able to run any command on the shell? If yes, please write some of them below:?

2.9 Let Us Sum Up

This unit covers the fundamentals of IT Auditing, as well as the basics of penetration testing. This unit is focused on hands-on experience related to various phases of ethical hacking like reconnaissance, scanning,
gaining the access (attacking). Various tools like maltego, NMAP, SQL map, theHarvester, metasploit framework are discussed in this unit.

2.10 Session wise- list of lab assignments

Now, the session-wise list of lab experiments is given below:

Session -1

a) Scan the target – scanme.nmap.org using nikto and observe the results.

Session -2

a) Execute theHarvester tool to find out the publically available information about your organization.

Session -3

a) Use NMAP to scan the target – scanme.nmap.org and observe the results.

Session -4

a) Perform footprinting on your organizations website using Maltego.

Session -5

a) Perform DNS Enumeration on target - zonetransfer.me using dnsenum tool in kali linux.

Session -6

a) Perform NetBIOS enumeration using nbtstat tool on your network.

Session -7

a) Use Ettercap tool in kali linux to perform man in the middle attack on your network and observe the results.

Session -8

a) Perform SQL Injection on target - http://testphp.vulnweb.com/ using SQLMap and observe the results.

Session -9

a) Perform SQL Injection on target - http://testphp.vulnweb.com/ using Havij and observe the results.

Session -10

a) Gain the access of Metasploitable machine using Metasploit Framework

Session -11

Describe the man-in-the-middle threat in secure shell and investigate which authentication options it offers.

2.11 Check your Progress: The Key

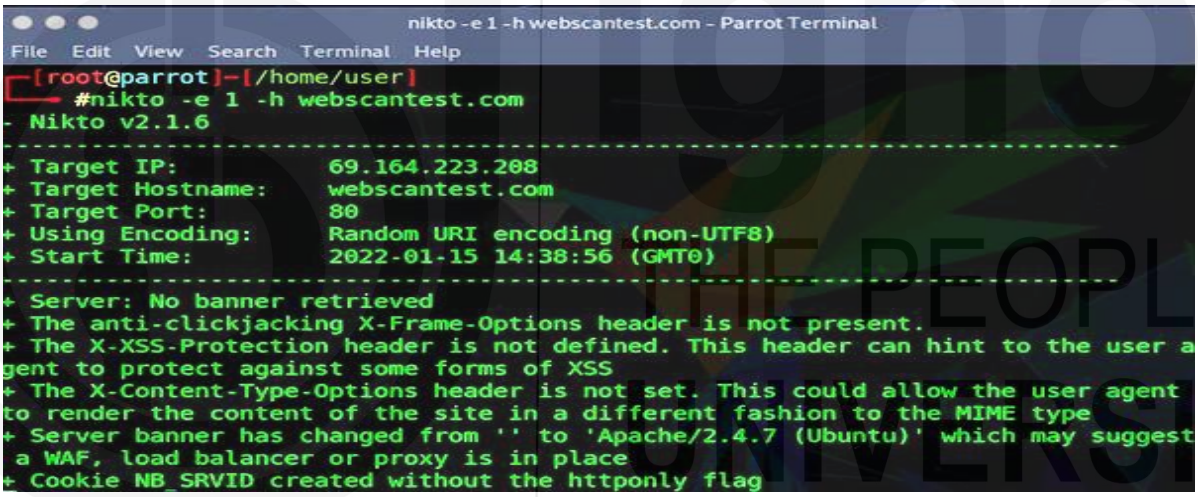
---

1. The following stages must always be included in the IT audit process:

   - Planning

- Defining audit objectives and scope
- Assessing controls
- Gathering and evaluating evidence
- Reporting, and
- Following up.

2.



3. Sample output:

MOBILE PHONE AUDITING AND DATA RECOVERY

Structure

3.0 Introduction

Technology is the backbone of automated audit operations, whether it is assisting with the discovery of anomalies, the analysis of patterns, the examination of trends, or the real-time testing of controls. Controls that are partially effective and can put a business at serious danger. Thus, the advantages of mobile devices like laptops or tablets for auditing "anytime, anywhere" take on more relevance. Printouts of checklists and handwritten evaluations are successful. Additionally, data integrity is more crucial than ever. With today's technology, audit messages can still be stored in offline auditing tools even if network connection is lost. This is especially useful when conducting fieldwork in remote locations with little or no internet. The most recent offline audit solutions make it simple, accurate, and rapid to record audit evidence while also guarding against data loss.

The majority of individuals actually believe that deleted or formatted data cannot be recovered. In actuality, though, the erased files are still kept on the hard drive. Users that are familiar with data recovery are able to retrieve lost data. Furthermore, we will be able to recover lost data if we are familiar with the data storage theory. Data recovery is the process of recovering information from a storage device which is not accessible using the conventional methods because it has been deleted in the past or because the digital medium has been damaged in some way. Various techniques are used to recover the missing files, but only in the event that certain portion of the storage still has the information they contain. Instances where a file has never been transferred to a persistent storage, such as when files were generated but ultimately unable to be saved to the hard disc drive owing to a power outage, are not covered by data recovery.

3.1 Learning Outcome

This is the third unit of this course, which seeks to introduce the concept of Mobile Forensics and explain the various aspect of data recovery. This unit covers the introduction to mobile forensics, android forensics and SQLite Analysis of BlackBerry Messenger on Android. This unit also covers the data recovery by taking the case study of Autopsy which is one of the popular tool in the forensics domain.

3.2 Mobile Phone Auditing

With the use of a tablet with a camera, mobile auditing enables auditors to take photos while out in the field. Previously, these cameras needed to be synced with database records. However, audio recordings may now be instantly converted to text. As a result, auditors can avoid the tiresome chore of entering additional information after coming to the office from the field.

Nowdays, everyone can be seen using laptop, tablet, or smartphone. Because of mobility, information security threats are now present on the workers' mobile devices, which may be used for work-related purpose with or without the organization's permission. While mobile devices promote productivity and efficiency due to their improved capability and mobility, they also pose danger, particularly when global privacy and data security laws and regulations change. Organizations must safeguard mobile device data, and internal audits must confirm, the security measures are effective.

3.2.1 Mobile Phone Forensics

Over the past decade or so, the growth of the mobile market has been the fastest growing segment of the IT industry. We now have portable smart phones and tablets being used by the entire consumer and business market. These add an extra challenge to the forensic investigator, but can also contain a cornucopia of information and evidence of the suspect's activity. These include texts (SMS) messages, emails, browsing activities, installed apps, etc. The overall impact is that the mobile device may be the greatest repository of information on your suspect. The figure below shows the investigative process for digital forensic science.

The mobile forensics process is shown. In order to prevent physical harm to the evidence, the procedure of seizing it include its safe removal and careful transfer to avoid electromagnetic interference, electric shock, and other hazards. This may be done by turning on Airplane Mode and disabling Wi-Fi and Hotspots to prevent any manipulation. The fragmentation of operating systems and item specifications might make it difficult to locate the data. The practise of obtaining the forensic data in a forensically sound manner while preserving its integrity is known as forensic acquisition. The term "Imaging" also applies to this technique. Based on the case history provided by the investigator, the forensic analyst separates the relevant and irrelevant material throughout the investigation phase. The analyst prioritises this data set depending on the ongoing research during the analysis phase by looking for correlations between the pertinent data provided during the examination phase. Reporting is a thorough synopsis of the mobile forensics investigation's findings. This stage also outlines the rationale behind each action taken and the outcome that resulted from it.

3.2.2 Android Forensics

We have sought to just cover one tiny aspect of Android Forensics here because it is a significant and difficult undertaking that deserves its own book, much as Windows, Linux, or Mac forensics. In this article, we'll look at an Android device's.xml manifest file. We want to concentrate on three main aspects while assessing an Android device:

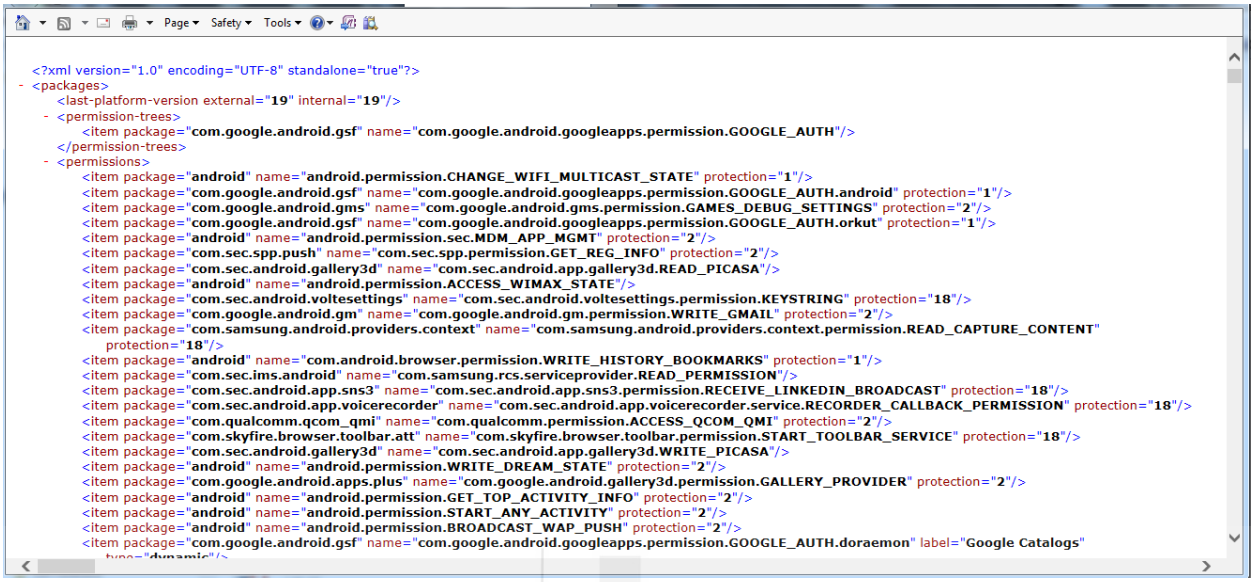
(1) /Root/system/packages

(2) AndroidManifest.xml

(3) any applications themselves. These are .apk files

The manifest file provides an.xml list of every programme that has been installed on the device along with any related permissions. By looking at the permissions of the apps, we may frequently spot malicious packages in this way. In other terms, a programme should be seen as suspicious and is possibly malicious if it has rights that are not necessary for it to function.

We have transferred an.xml file from an Android handset to our PC in this instance. We ought to see a file identical to the one below when we view it in a browser that supports.xml.



Go to the permissions for the ledflashlight programme. It need to be close to the middle of the document. Use the search option if you are having trouble finding it, and it will locate every mention of "ledflashlight". The one we seek has the permissions for the application as shown below.

Note that this application, a flashlight app, has permissions to;

(1) READ_EXTERNAL_STORAGE

(2) access INTERNET

(3) WRITE_EXTERNAL_STORAGE

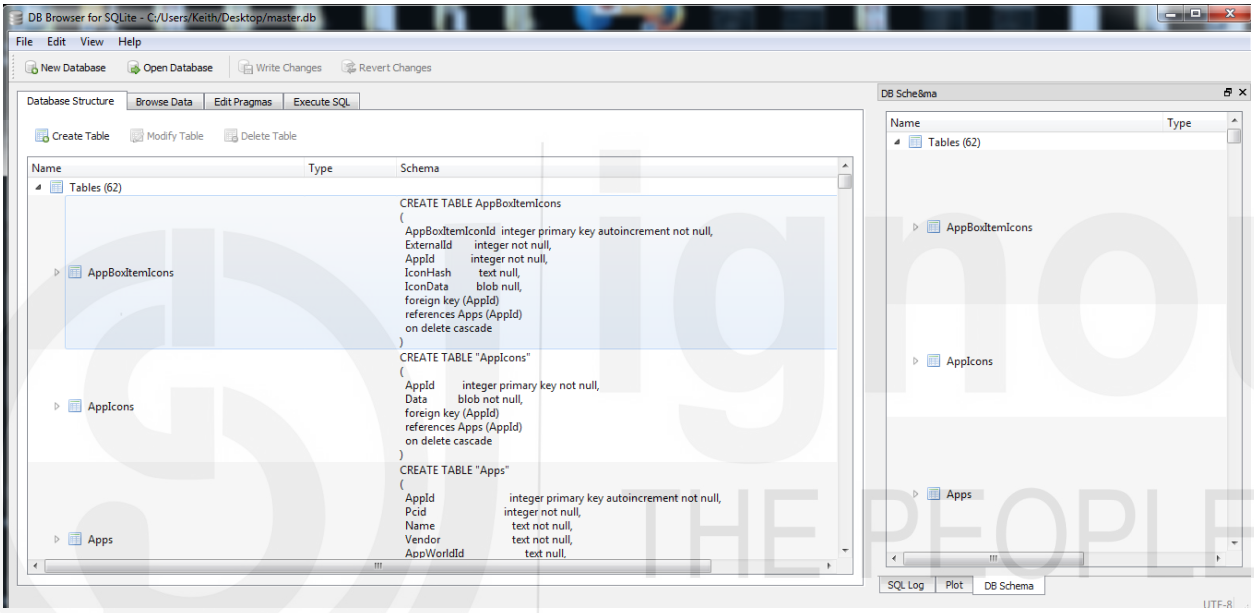
It's doubtful that a trustworthy lighting application would need such rights. We ought to be wary about this application! Malware is most likely what it is.

3.2.3 SQLite Analysis of BlackBerry Messenger on Android

A SQLite database is used by many mobile applications to store data. It is great for mobile devices since SQLite is a comprehensive relational database that is quite lightweight.

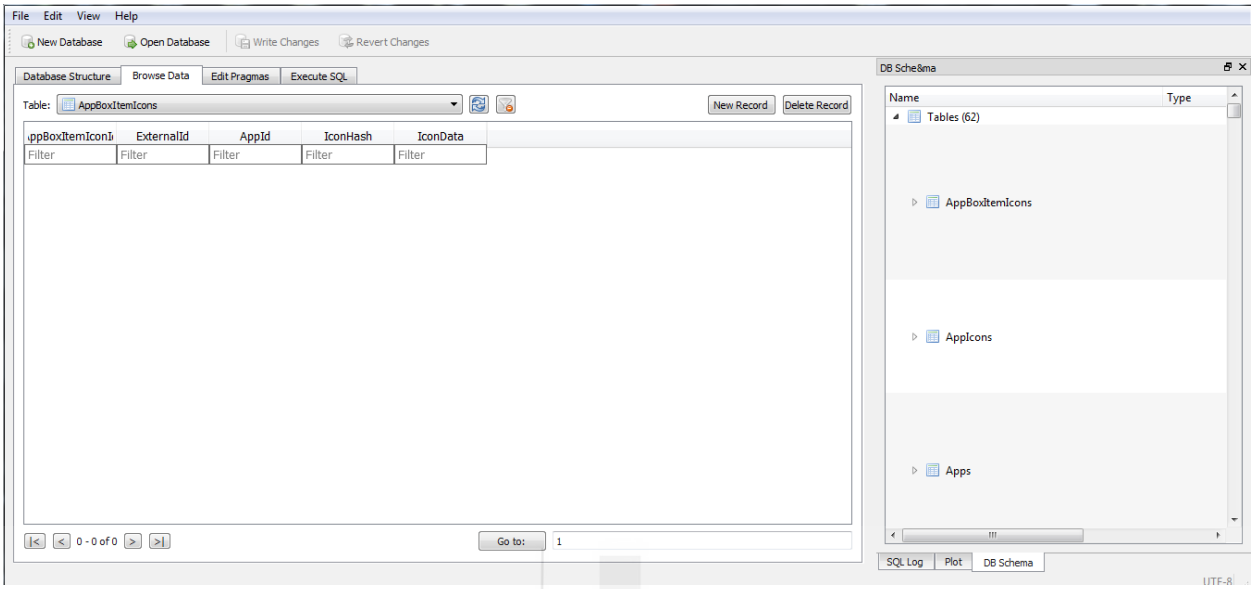
In this lab, we'll look at the SQLite database from an Android device running Blackberry Messenger. The SQLite Browser is required. It comes pre-installed with Kali; if not, you may get it from this page. It should already be set up on your computer if you completed the Browser Forensics training.

Here, we open the master.db from an Android device using the SQLite Browser. Choose File, "Open Database," and then master.db. It ought to resemble the screenshot below.



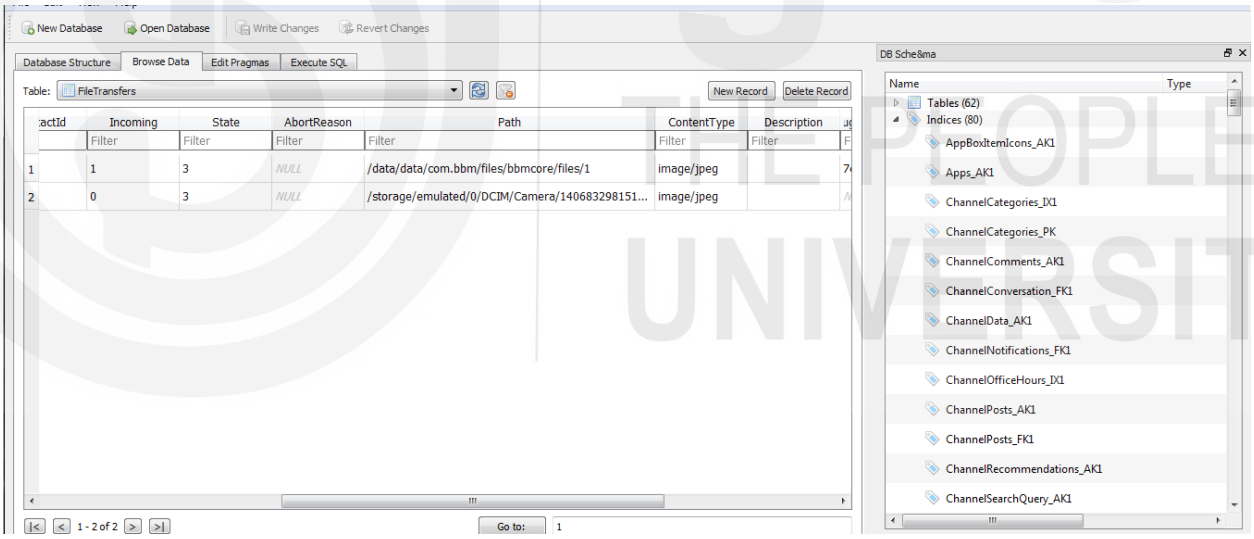
The 62 tables are visible, along with the instructions to construct them, in the main window to the left, under the Schema column.

Next, select the "Browser Data" tab from the main window's top menu. An image of that screen should appear below. The 62 tables are listed in the right pane. Simply choose the table from the "Table" pull-down option at the top of the main window if you want to see the data in a table.

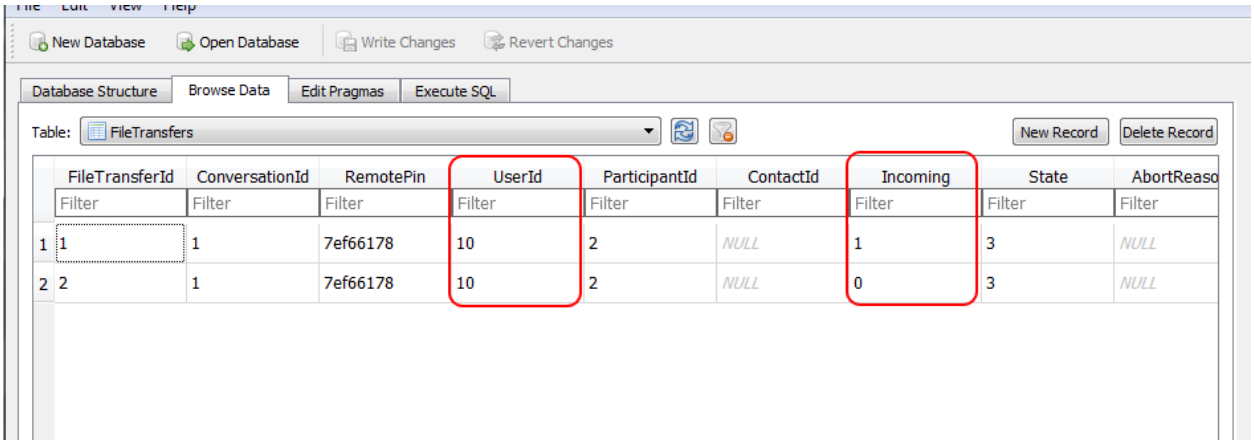


In this instance, access the "File Transfers" table, From the pull-down menu.

Now information from the "File Transfer" table will be added to the main menu. There will be two file transfers. These are.jpeg documents. There location can be seen on the device by enlarging the path column.



These columns will show "UserID" and "Incoming" columns if scrolled left through them. In the image below it can be seen that User ID=10 made both file transfers, and the incoming column shows that the first file transfer was an incoming one and the second one an outgoing one (not incoming). Clearly, this information might be helpful in proving that the suspect used that phone to send or receive a harmful or unlawful file.



Check your progress 1

Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i)What are the findings of mobile forensics?

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

3.3 Introduction to Data Recovery

The practice of recovering inaccessible or corrupted data from digital media that has suffered some sort of harm is known as data recovery. Data may be recovered using data recovery software from a wide range of devices, including hard drives, storage devices, tapes, cell phones, PDAs, floppy discs, CDs, DVDs, data cartridges, Xboxes, and many more. Data may be retrieved from digital media in a variety of ways that can differ substantially; the simplest option frequently just executing some simple software on the storage device in issue. This is usually a bad idea since the information that is being retrieved itself could get overwritten by the recovery data. There are more sophisticated commercial software programmes that can perform this task more expertly.

Prior to imaging the original disk and starting work on a "back-up" of the original programme, no software fixes should be made. If there is an issue with the initial image which is being operated on for restoration, the most reputable businesses will additionally capture a second image. When a customer receives a listing of all the restored files, the data recovery task is often considered complete. After the customer confirms this file listing, the data on the proper medium is subsequently sent to them. A hard drive, floppy disc, CD, or DVD can be used for this. If a file is deemed crucial, another option is to encrypt it before emailing it to the client.

To extract data from a disc drive or other storage device without the aid of the file system that originally produced the file, file carving is a technique used in computer forensics. It is a technique used to recover data and carry out a digital forensic investigation that finds files at unallocated space without any file information. It is also known as "carving," a broad term for the process of separating structured data from unstructured data based on the presence of format-specific properties in the structured data.

File carving methods: Several forms of media need to be examined during digital investigations. Relevant information can be located in computer memory as well as on a variety of networking and storage devices. It is necessary to examine a variety of data formats, including emails, electronic documents, system logs, and multimedia files. The recovery of multimedia files from storage devices or computer memory using the file carving method is the main topic of this text. A recovery method known as "file carving" only takes into account the contents and structures of files, not the file system or other meta-data that is used to arrange data on storage media.

Typical implements for file carving: Because clever harmful users will constantly strive to destroy evidence of their illegal actions, data recovery technologies play a crucial part in the majority of forensic investigations. Tools for data recovery include:

- Scalpel
- FTK

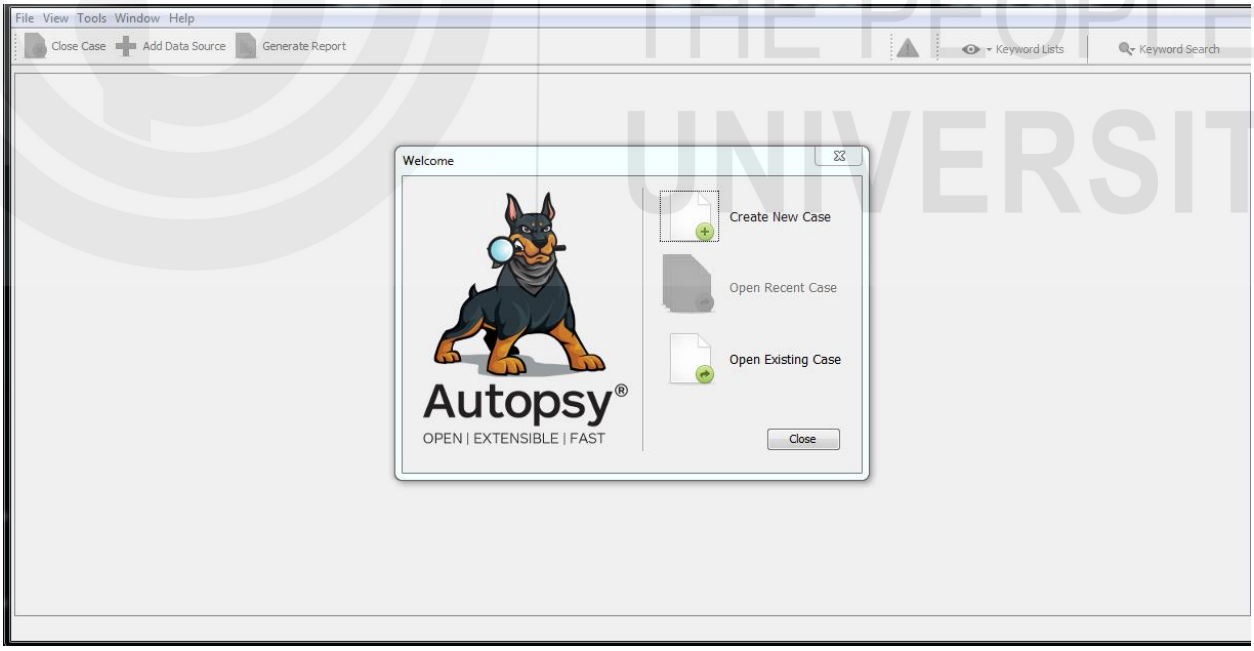
- Encase
- Foremost
- PhotoRec
- Revit
- TestDisk
- Magic Rescue
- F-Engrave

3.3.1 Recovering Deleted Files

Any files that the suspect may have erased can be recovered. The most basic of the abilities required for a forensic investigator is undoubtedly retrieving deleted files. Until replaced, "deleted" files are still there on the storage media. Simply by deleting the files, the cluster becomes ready for overwriting. This implies that the erased evidence files can still be recovered until the point at which the file system overwrites them.

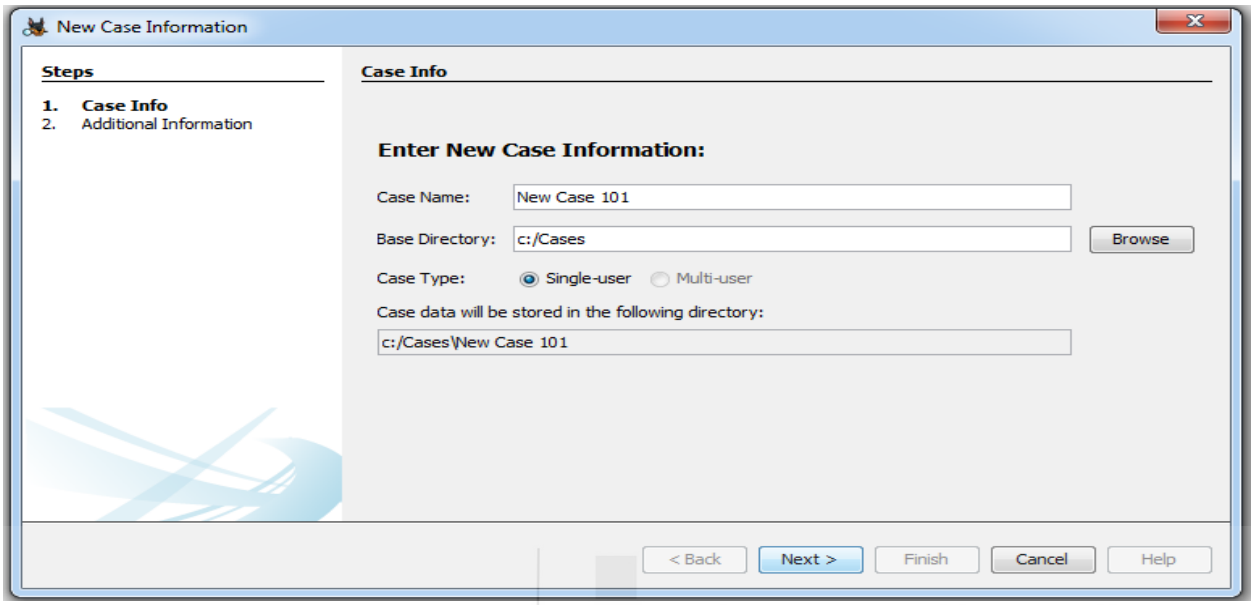
The Sleuth Kit (TSK) is an open-source tool for finding and retrieving deleted data. The Sleuth Kit can be used with the Windows testing environment even though it was originally designed for Linux.We will use the GUI interface known as Autopsy that was created for TSK will be used in this lesson.

Download and install Autopsy using the following link: http://www.autopsy.com/download/



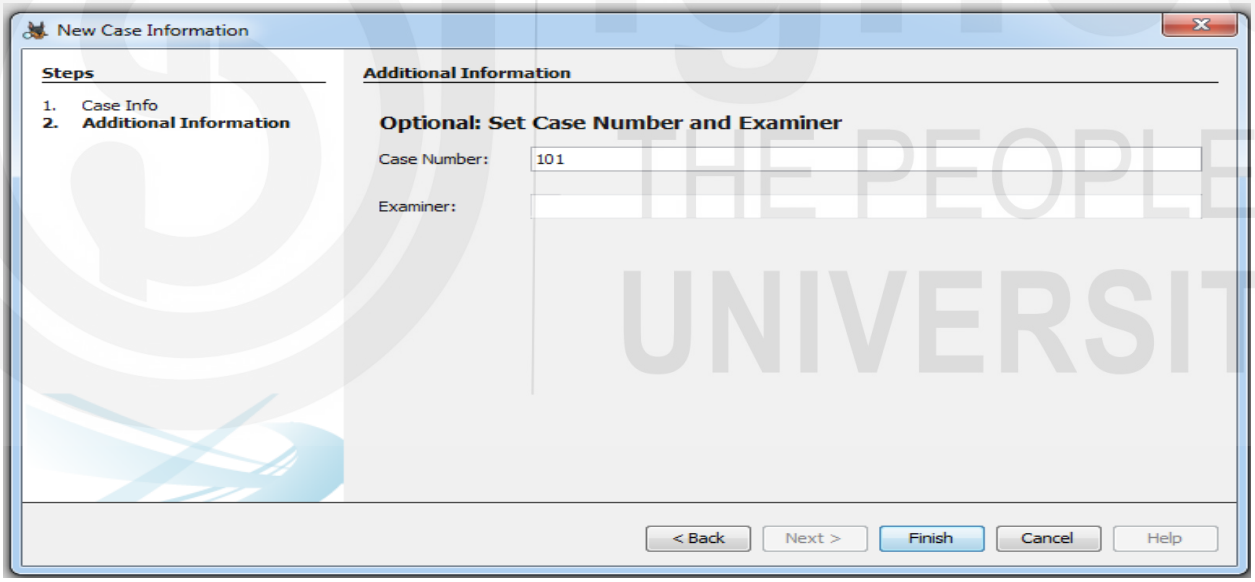
Then select "Create New Case."

A new window will appear asking to name the new case and select the directory where to store your cases. Put the file "New Case 101" in the C:Cases base directory.



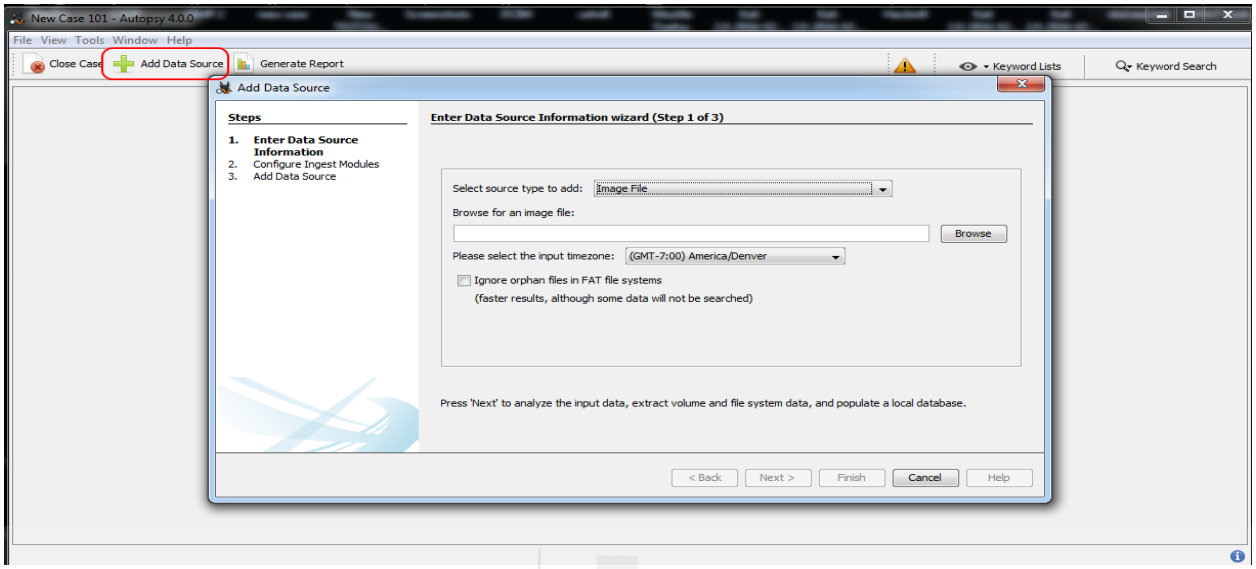
Click Next.

A new window will emerge, requesting the case number and examiner's name. Give it a case number of 101 and the examiner's name or initials.

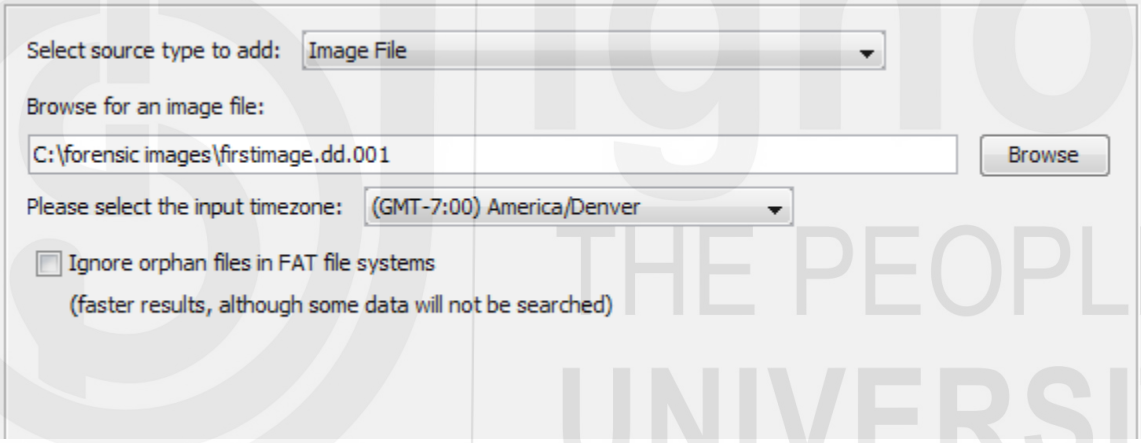


Select "Finish."

Then, select "Add New Data" from the menu in the upper left corner. A popup labelled "Add Data Source" will appear. Select "Image File" and then browse to the image file you prepared in Module 1 because we will be using the one from that module. In the screenshot below the copy is in the location c:forensic pictures. It can vary from system to system.

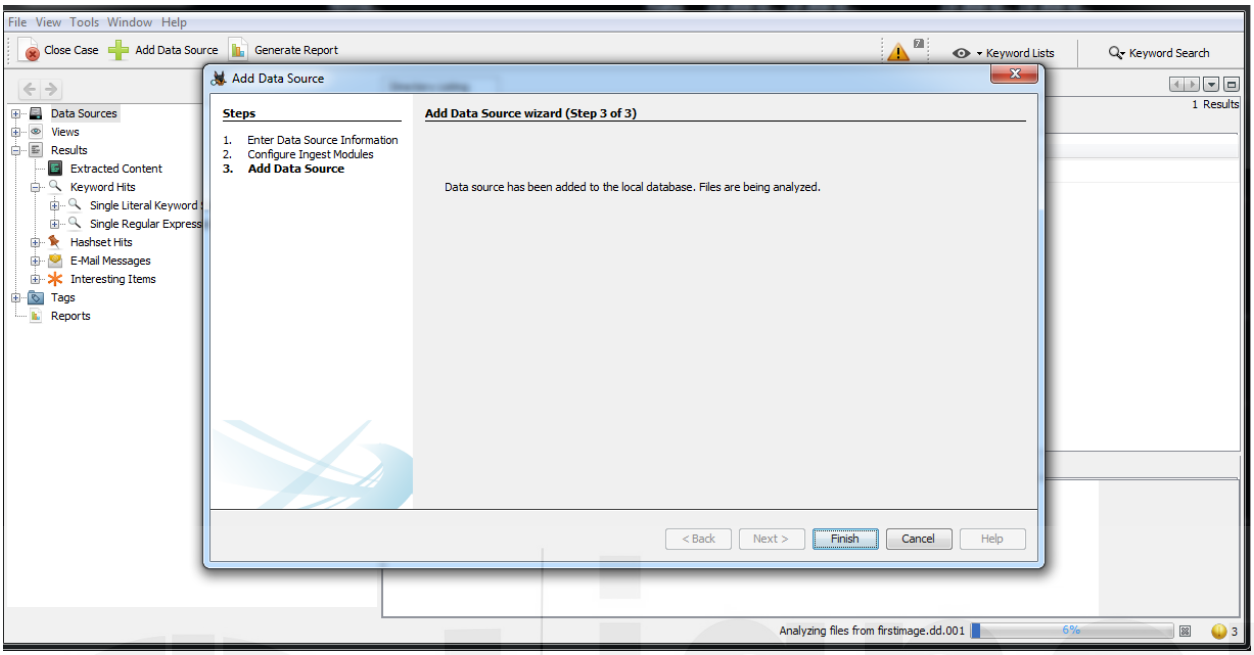


Add first image.dd.001 from the series' first instruction.

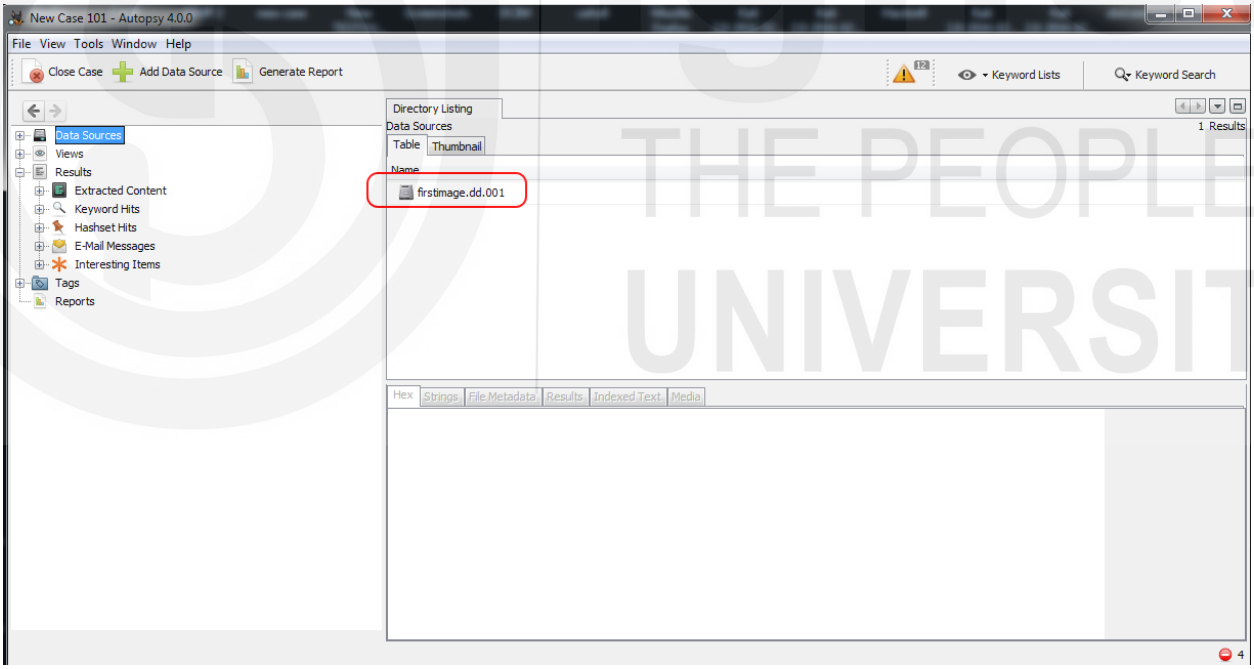


Clicking next after adding the photograph will trigger Autopsy to start analysing it. Eventually a screen similar to the one below will appear.

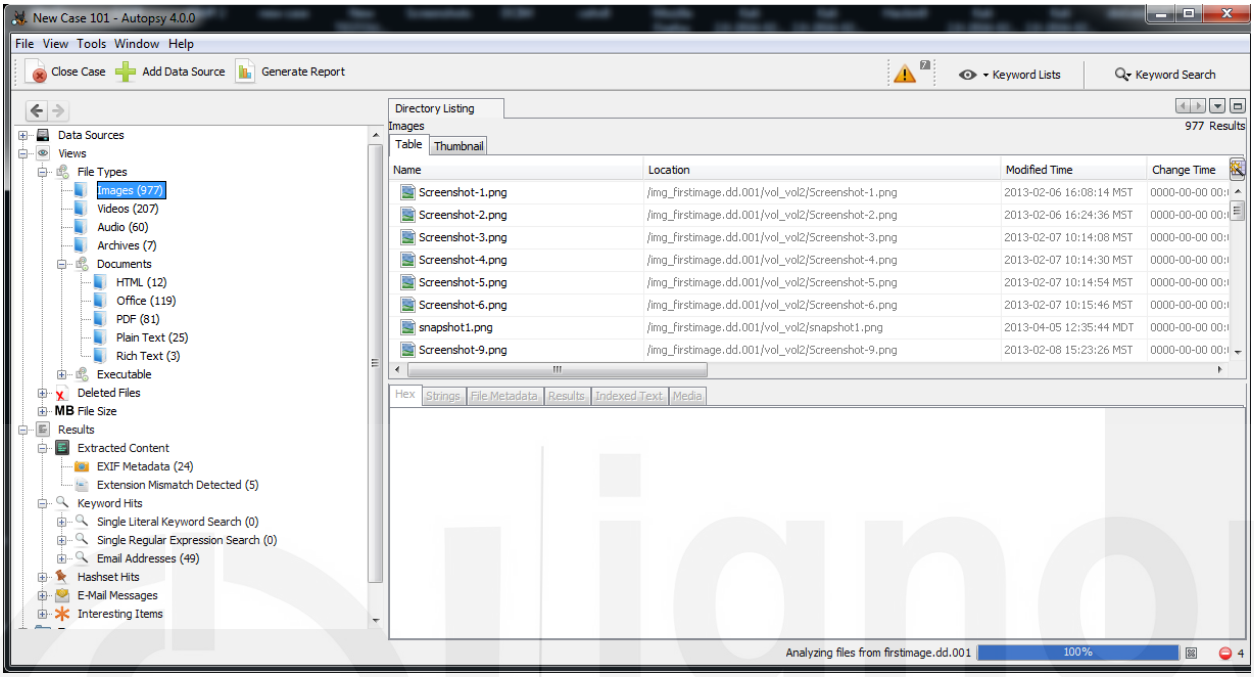
Click "Finish".



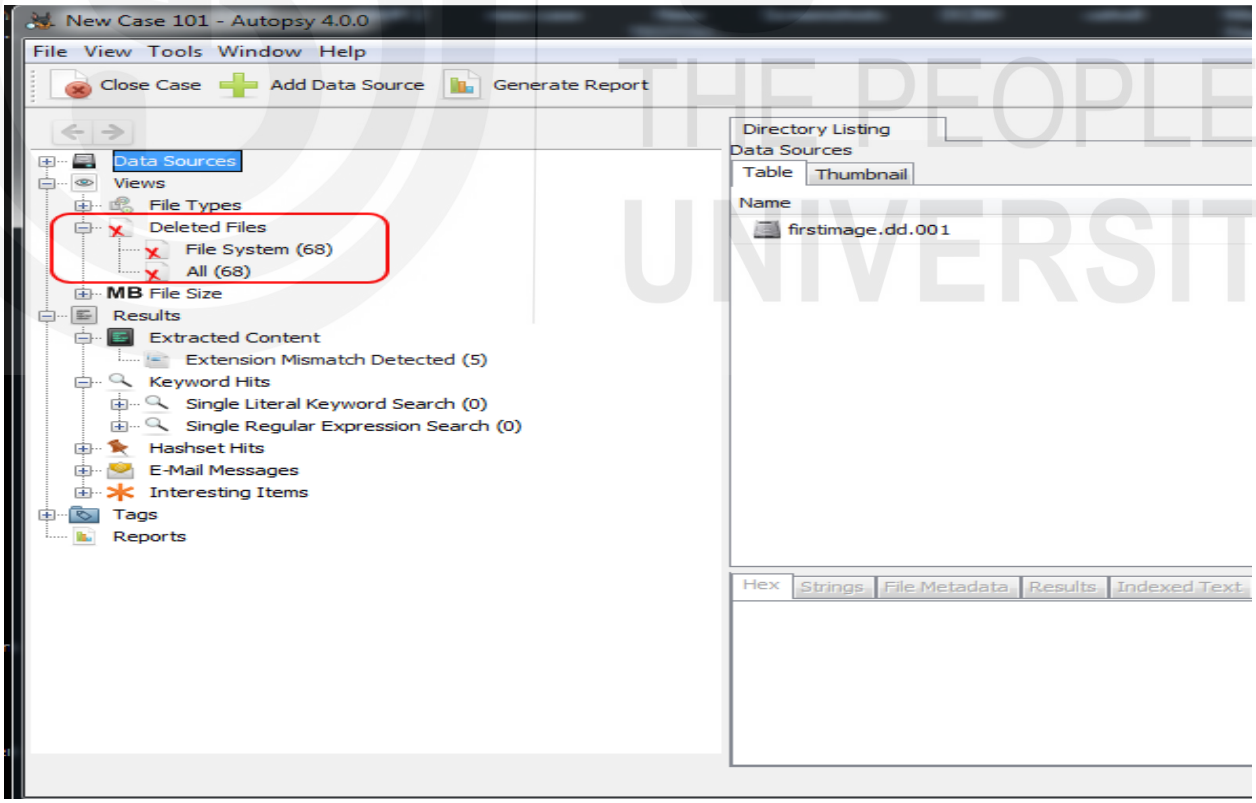
An interface similar to the one below will appear. Keep in mind that your data source should be listed as "firstimage.dd.001".



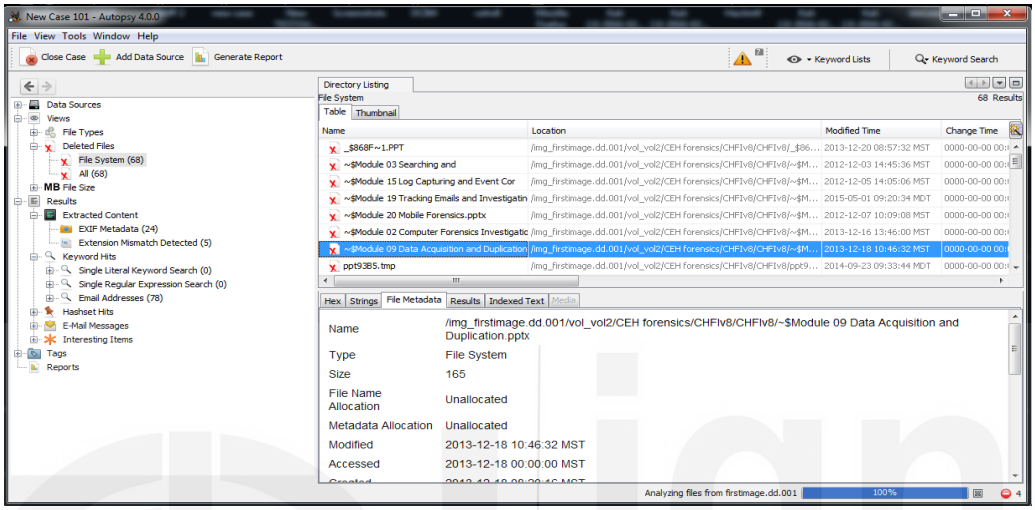
Autopsy will provide all type of files and the number of files in each category if “File Types" is selected in the object explorer.



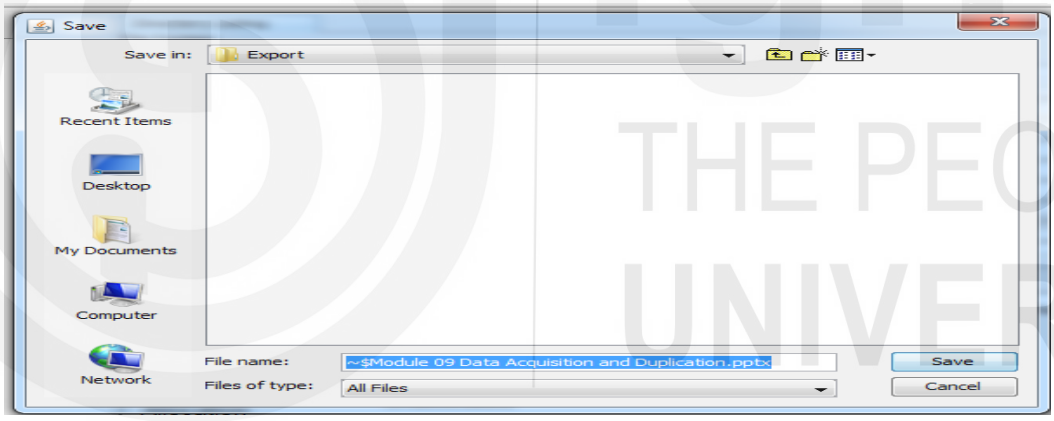
All the deleted files can be seen by using "Deleted Files" option in the object explorer.



After clicking on “deleted files” some options will appear in lower right panel which may be used to perform some analysis. Hex, Strings, File Metadata, Results, and Indexed Text are among the tabs that appear. After choosing the "File Metadata" tab, the file's metadata, which includes the name, type, size, updated date, and generated time, will be displayed (MAC).



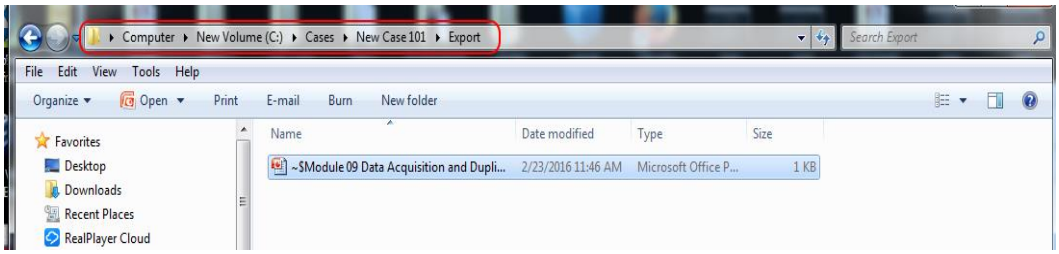
Right click on the file to be recovered and choose export.



Save the deleted file to the Export directory.

Navigate and look for the deleted or exported file;

C:\Cases\New Case 101\Export



Now, double click on the file to launch the proper programme. As forensic investigators, we are aware that the deleted data can be retrieved until the point at which the file system overwrites them. Recovery of these deleted data is simple using programmes like Autopsy and almost any other forensic suite (Encase, ProDiscover, FTK, Oxygen, etc.).

Check your progress 2

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

i) What are the findings of using autopsy for data recovery?

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

3.4 Let Us Sum Up

This unit covers the fundamentals of mobile forensics and data recovery using the autopsy tool. This unit is focused on android forensics specifically,and SQLite analysis of blackberry messenger. In the later sections, the data recovery using autopsy is discussed to make the readers aware about the data recovery process.

3.5 Session wise- list of lab assignments

Now, try solving the problems related to mobile phone forensics and data recovery.

Session -1

b) Study about the fundamentals of mobile phone forensics.

Session -2

b) Use AFLogical OSE to perform mobile forensics.

Session -3

b) Study about the Andriller and perform the forensics on a dummy mobile phone.

Session -4

b) Use Android Data Extractor Lite (ADEL) to draw forensic flowchart from the database of a mobile phone.

Session -5

b) Use WhatsApp Xtractto view the WhatsApp conversation on the computer.

Session -6

b) Perform data recovery using PCInspector on a USB pendrive.

Session -7

b) Perform data recovery using Recuva on a USB pendrive.

Session -8

b) Study about the hardware and software-based write-blockers.

Session -9

b) Use autopsy to recover the deleted items from a USB pendrive

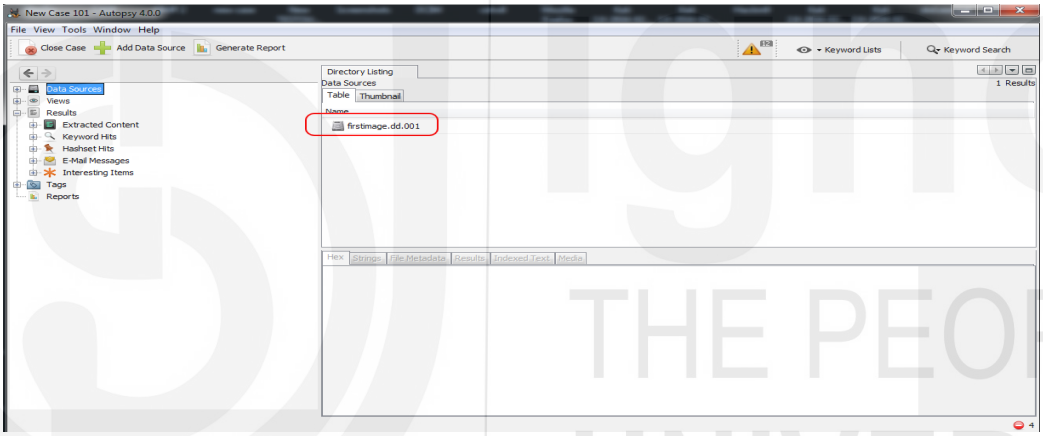
Session -10

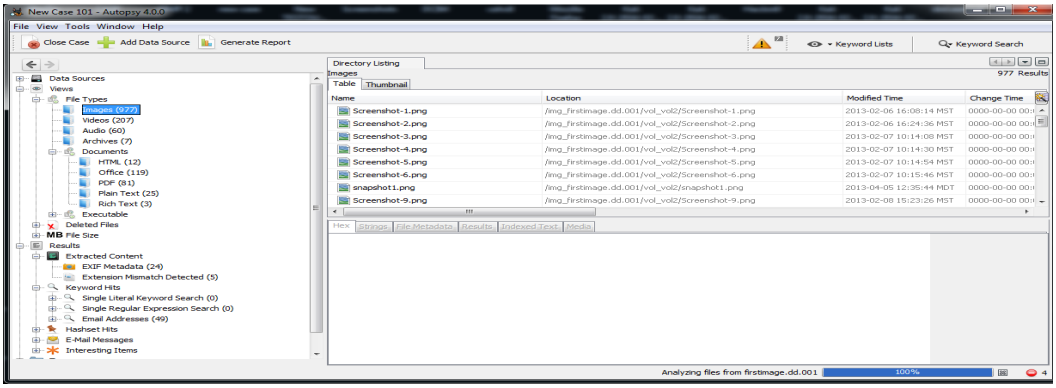
b) Compare the results obtained by autopsy and PCInspector.

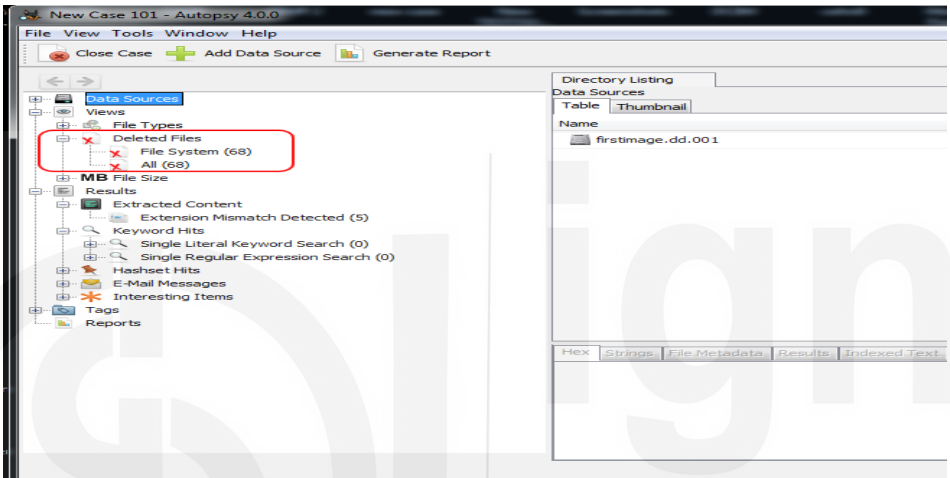
3.6 Check your Progress: The Key

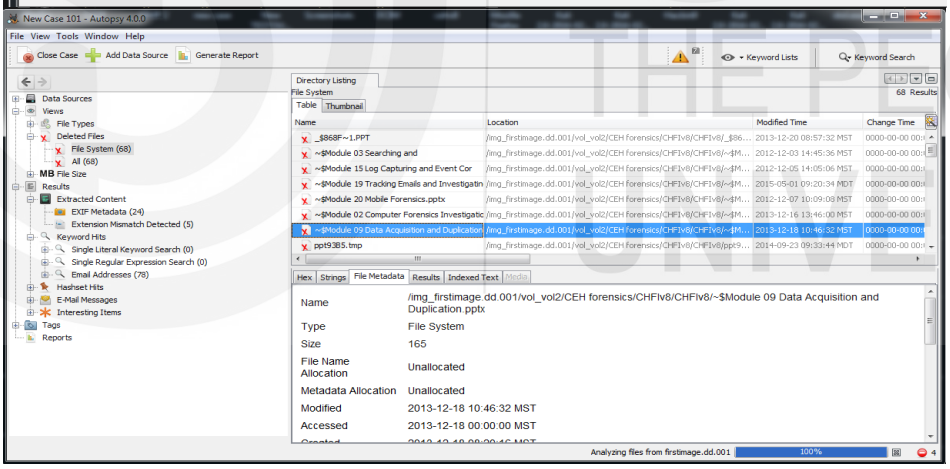
5. Below are a sample of findings of applying mobile forensics on the digital evidence.

6. Below are the screenshots of the findings:









HANDS ON OPEN SOURCE DATA ANALYTICS AND RECOVERY TOOLS

Structure

4.0 Introduction

Data analysis tools are the specialized softwares used by data scientists. These software may create databases by combining massive datasets from several sources. However, a lot of data systems are user-friendly enough for everyone. Data analysis is done by data platforms to provide information about the company process. The outcomes of data analysis assist in forming next company selections. Data visualisation is one of the key components of data analytics solutions. Ordinary consumers cannot go through reams of code and comprehend what is happening. Data is transformed into a variety of graphs, charts, and other graphic representations through data visualisation. It should be simple to set up and use data visualisation. Users may select the data sets and variables for generating graphics by dragging and dropping or by pointing and clicking. To view the changes, simply update the details. A real-time data updating is also possible with data visualisation. This works well when it is presented to the group of people, in an official presentation. They can see the underlying facts reflected in real-time visualisations.

In this unit, two tools are used for performing the data analytics i.e. talend and weka. Talend provides a variety of technologies for data mining and cloud-based data integration. It links information from over 900 sources, such Salesforce, Azure, AWS, and Marketo. It takes relatively little time to set up data warehouses and create data lakes. Complicated relational databases are handled by Talend. IoT analytics and large data processing are both possible with machine learning. Additionally, Talend provides a tonne of free add-on tools. Some of them speed up ETL operations. Another feature is a quick data loader.

Moreover, in the second section of this unit we have explained the PC Inspector, Users may recover lost or damaged data from their personal PCs with the PC Inspector File Recovery system.

Many of its features are shared by other packs like TestDisk and Recover My Files. This software, nevertheless, is categorised as freeware. The cost to download will not be incurred by the user.

4.1 Learning Outcome

This is the fourth unit of this course, which seeks to introduce the concept of open source data analytics and recovery tools. This unit covers the introduction to data analytics followed by the hands-on labs using tools like talend and weka for the data analytics. This unit also covers the data recovery by taking the case study of PC Inspector tools, which is a freeware and is used for recovering the deleted files.

4.2 Introduction to Data Analytics

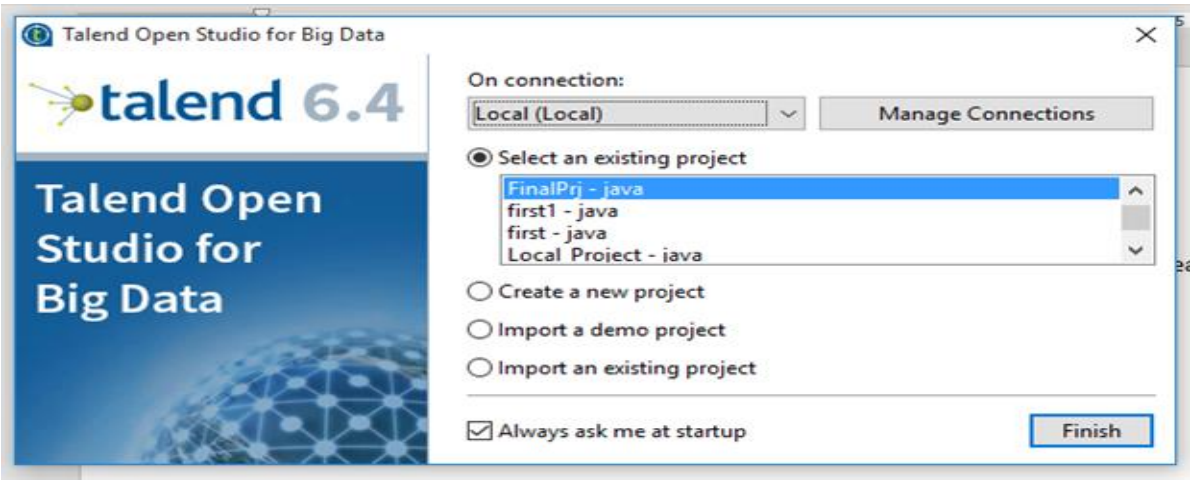
As it is utilised to uncover hidden information, develop reports, conduct market analysis, and enhance business requirements, data analytics plays a crucial part in enhancing a company. The term "data analytics" describes the methods used to analyse data in order to increase productivity and financial benefits. In order to examine different behaviour responses, data is taken from a variety of sources, cleaned up, and classified. The methods and resources employed varies depending on the group or a person. Data analysts transform numerical data into everyday language. By gathering data on certain subjects, interpreting, analysing, and presenting results in thorough reports, data analysts add value to their organisations. Therefore, to become a data analyst one must have the capacity to acquire data from multiple sources, evaluate that data, discover hidden information, and produce reports. The actions taken with the data and the results produced are two important ways that data scientists and data analysts differ from one another. A data analyst will look to solve particular issues that have previously been recognised and are well-known to the company. In order to achieve this, they analyse enormous databases in an effort to spot trends and patterns.

4.2.1 Data Analytics using Talend Studio

Talend Open Studio is a free open source ETL (Extract, Transform and Load) tool sor Data Integration task. It can be downloaded and installed from http://www.telend.com/lp/open-studio-for-data-integration/

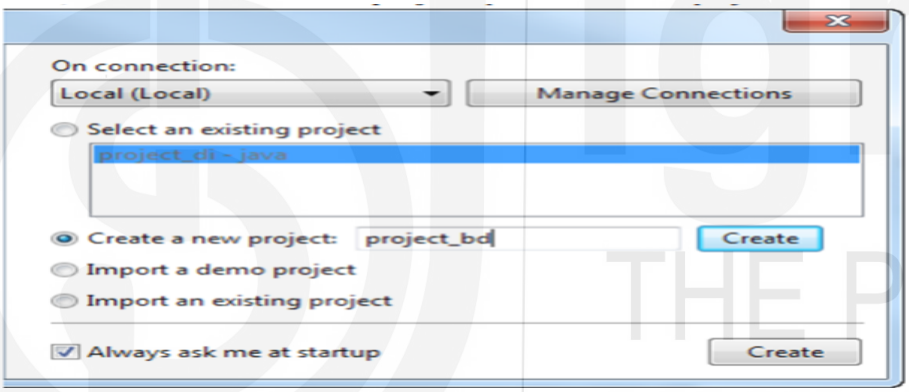
Step 1: Launch Talend Studio.

Step 2: To select an existing project, select the project name from the list and click on Finish. OR



To create a new project after the initial startup of the Studio, do the following:

1. On the login window, select the Create a new project option and enter a project name in the field



2. Click Create to create the project. The newly created project is displayed on the list of existing projects.

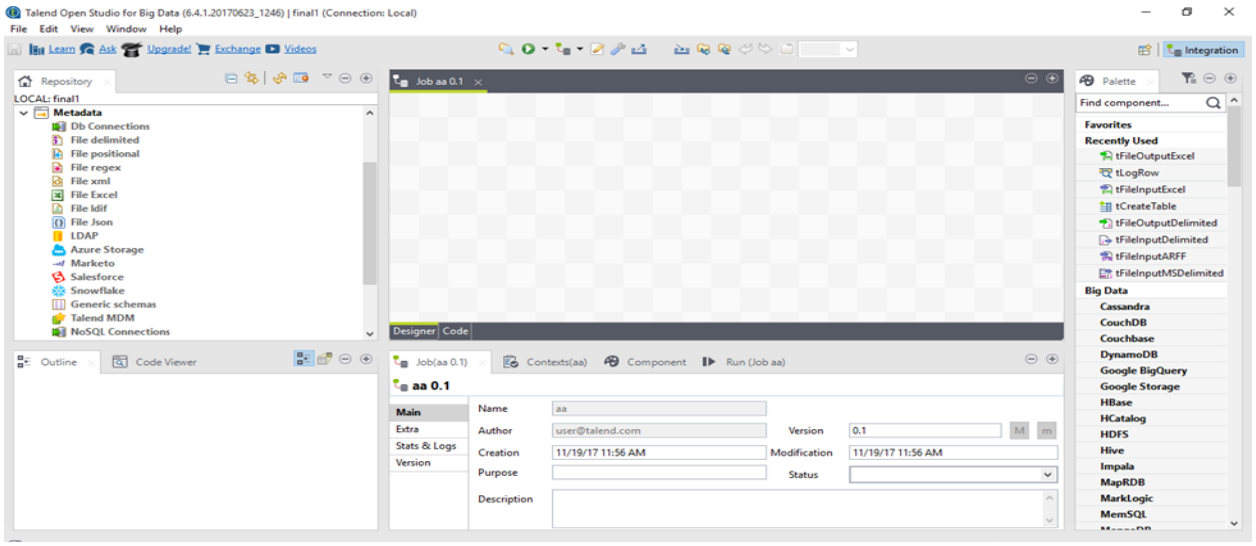
3. Select the project on the list and click Finish to open the project in the Studio.

Step 3: To start a job, do the following:

1. On the main window right click on job design and select the create job option.

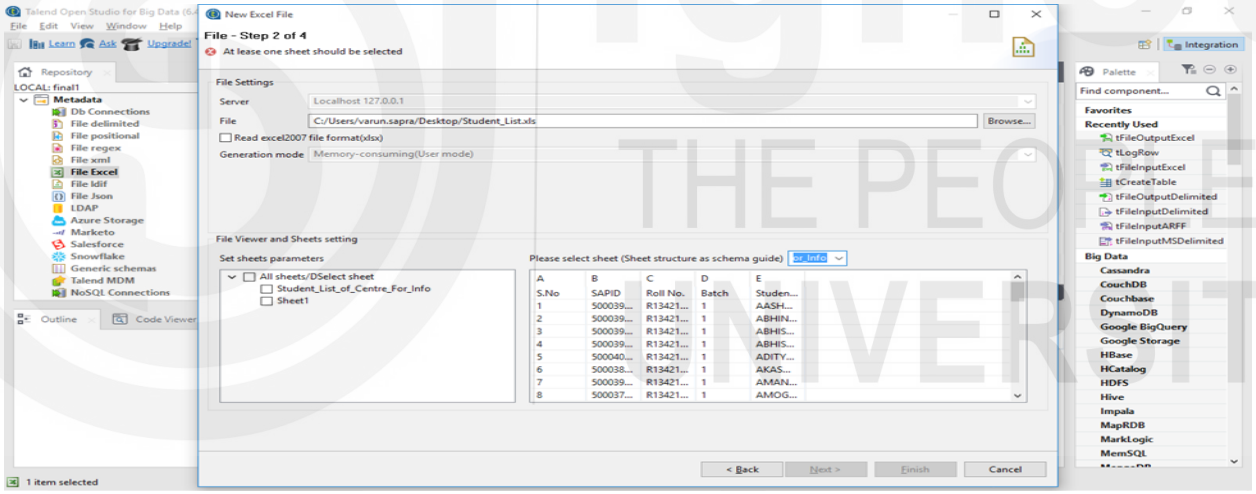
2. In the new dialog box, enter the job descriptions like name, purpose, Version etc. and click on Finish button.

3. In order to read an excel file, choose the option metadata from repository and select excel file.



4. Right click on the excel file option and select Create File. Write name, purpose and click on next.

5. Browse the file to be read. Once done, the list of sheets and preview of data is displayed at the bottom panel. Click on Next.

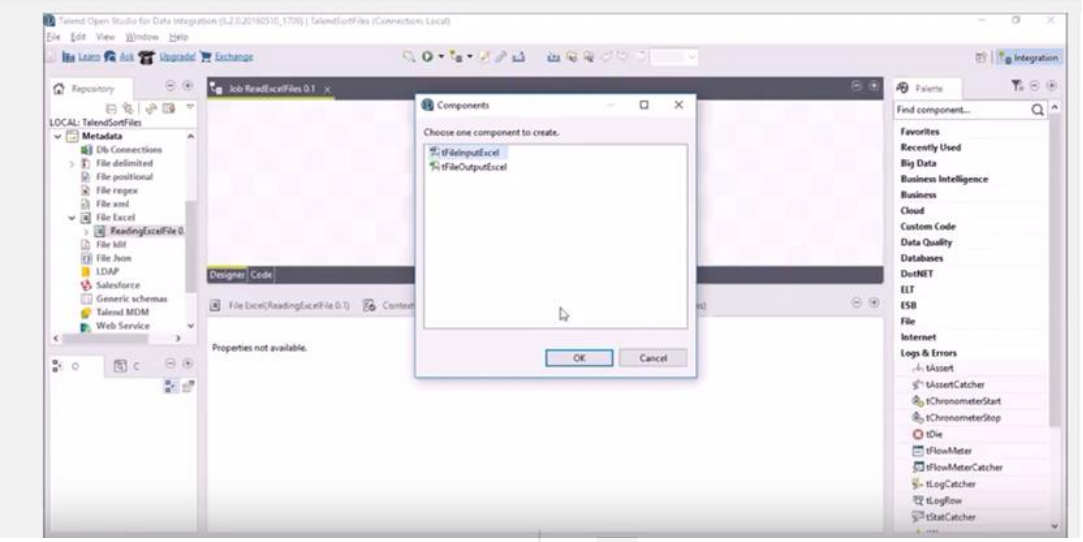


6. Enter First column number and make that header and click next. The metadata will be displayed. Now, Click on Finish and metadata will be displayed on Repository panel.

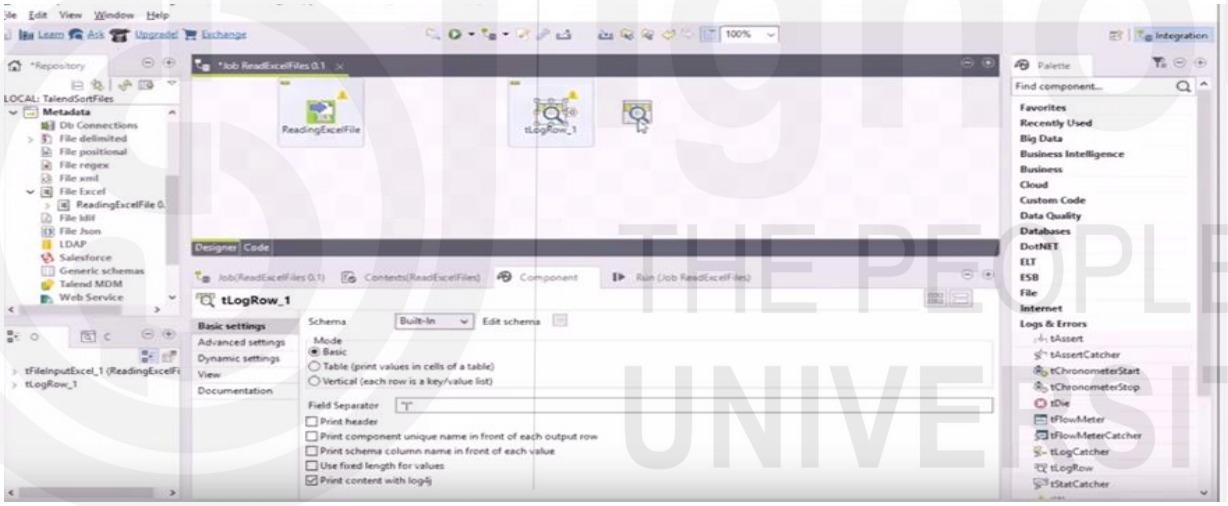
7. Drag and Drop the metadata on the job panel, it will ask that whether it is input or Output file. Select Input File and click OK.

8. To read the file add a component called tLogRow_1 on the working panel.

9. To link the inputfile component with tLogRow_1 right click on the inputfile component. Select Row and Main and drag it to the tLogRow component.



10. To pass the data to other file drag the outputfile component towards the right side component panel.



11. To link the tLogRow_1 component with outputFile right click on the tLogRow component. Select Row and then Main and drag it to the outputFile component.

a) Space is given below for writing your answer.
b) Compare your answer with the one given at the end of the unit.

Check your progress 1

Note a) Space is given below for writing your answer.

c) Compare your answer with the one given at the end of the unit.

i)Discuss the findings after performing data analytics using talend studio?

……
……
……
……
……
……
……
……
……
……
……
……
……
……

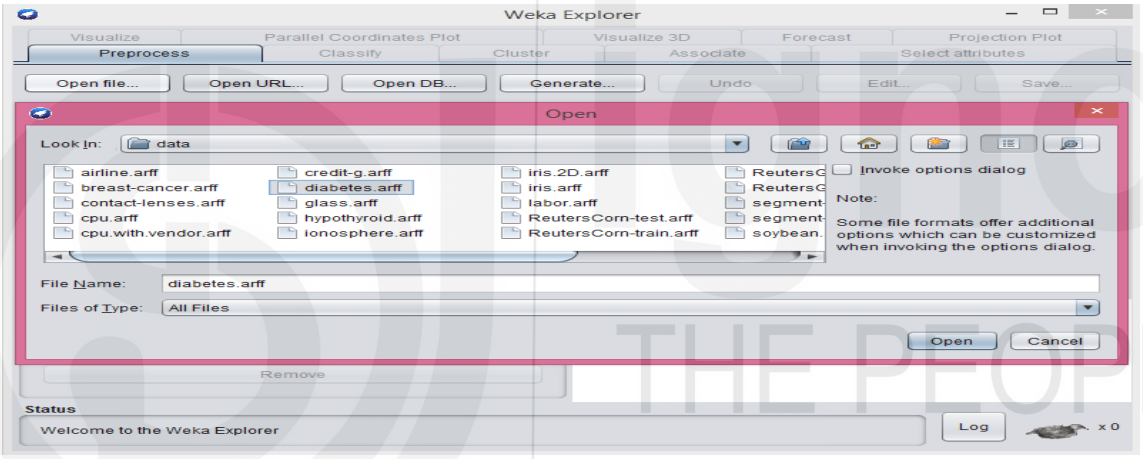
4.2.2 Data Analytics using Weka

Weka is open source software issued under the GNU General Public License. It contains a collection of machine learning algorithms to perform several data analysis and visualization tasks. Input to Weka must be formatted using the Attribute-Relation file format and must have filename extension as .arff. It can be downloaded and installed from http://sourceforge.net/projects/weka/

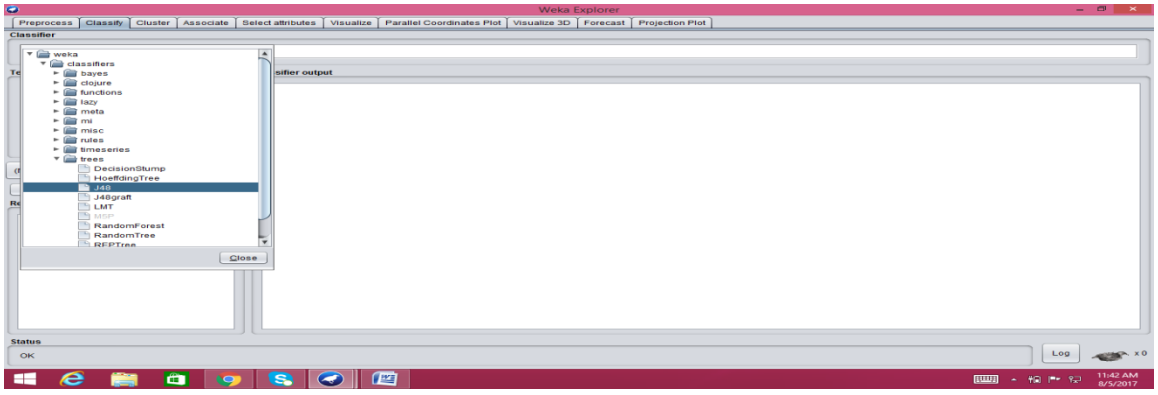
Objective: To Generate Decision Tree (C4.5) using weka

Step 1 : Open the Weka GUI chooser and click on explorer.

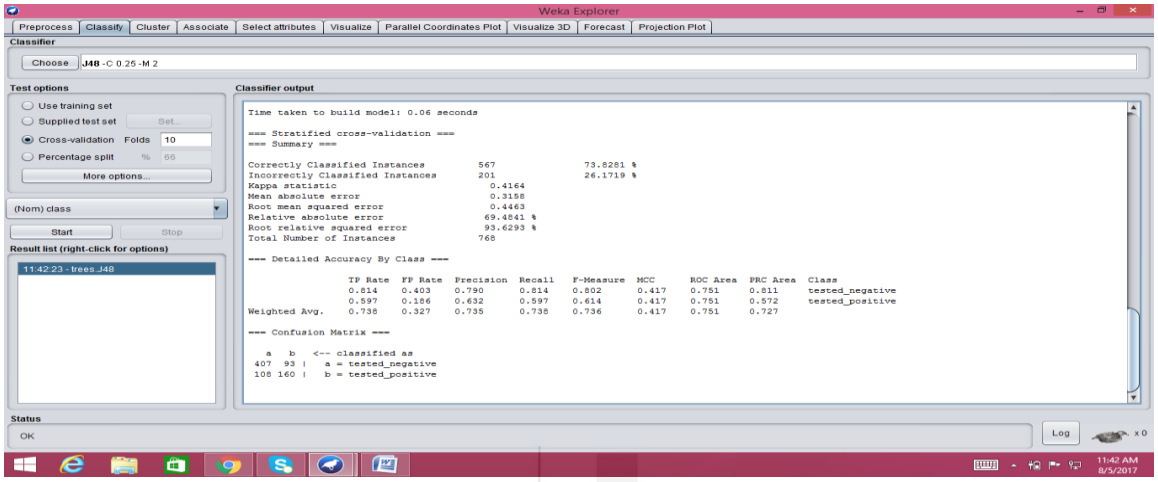
Step 2: While installing Weka, several datasets are loaded in the data folder of Weka. To select pima diabetes dataset from data folder, click on open file and select diabetes.arff. Now, click on the open tab to load the data.



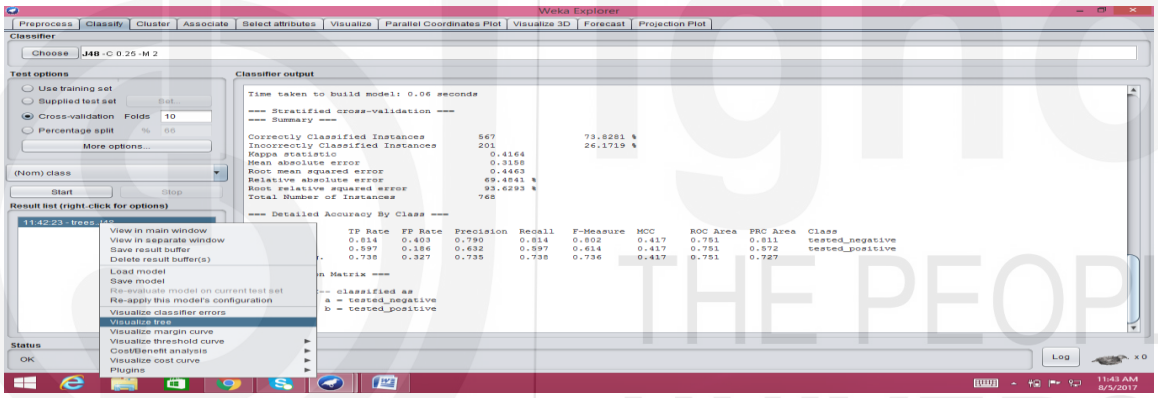
Step 3: Click on the classify tab and choose j48 from the decision tree option. Now initiate the classification using the start button.



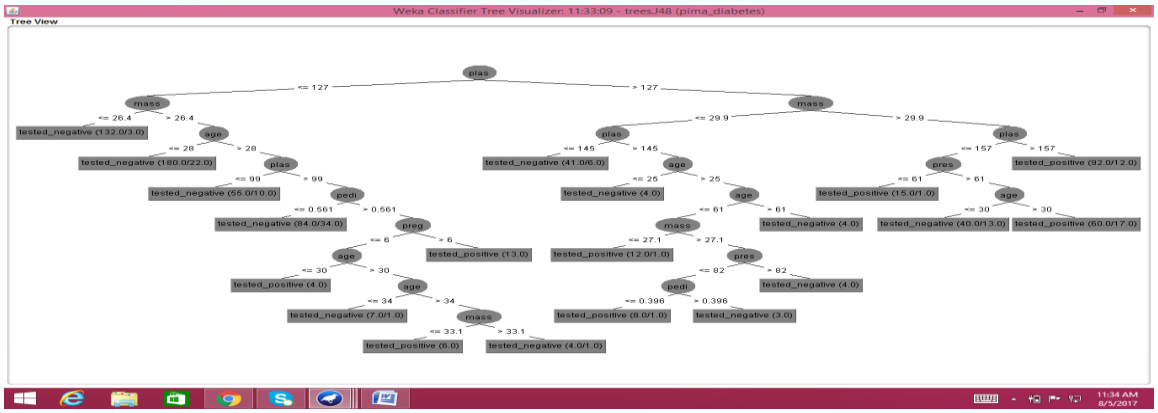
Step 4 : Result displays the confusion matrix and other statistical measures



Step 5 : To visualize the tree right click on the jtree J48 and choose the visualize tree option .

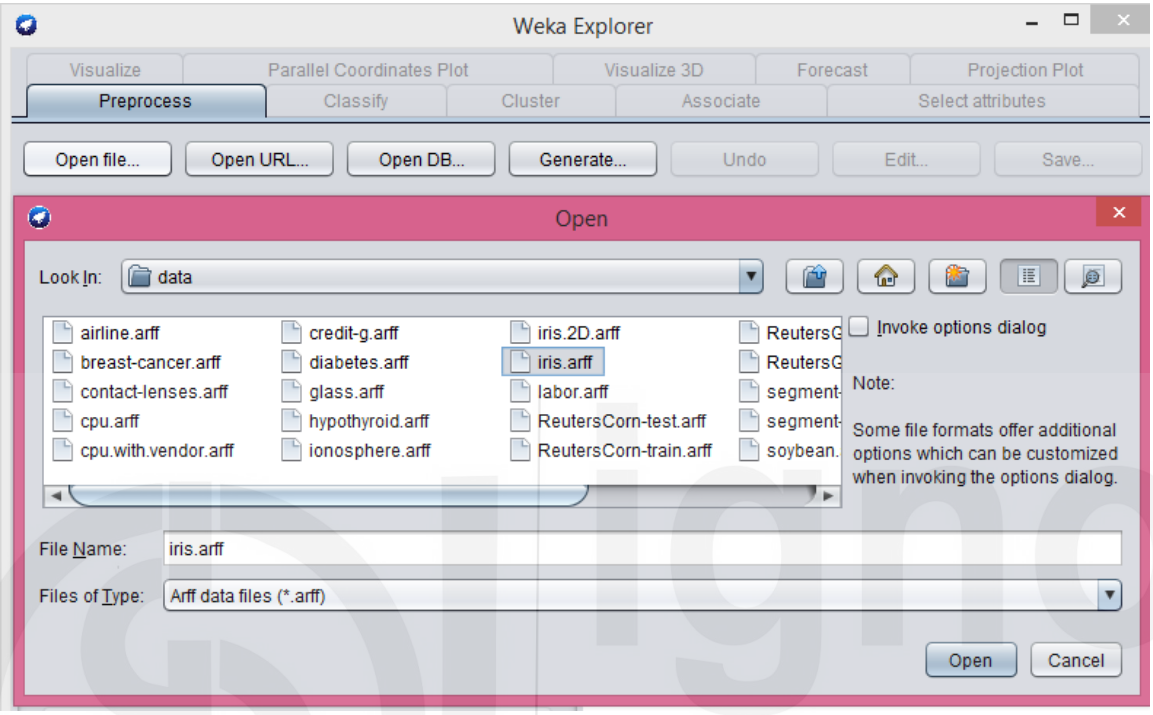


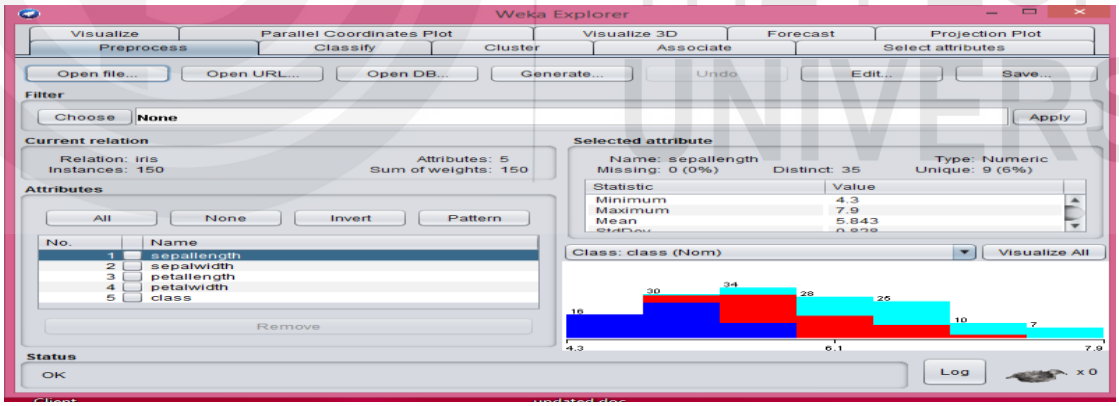
Decision tree will be generated for the diabetes data set



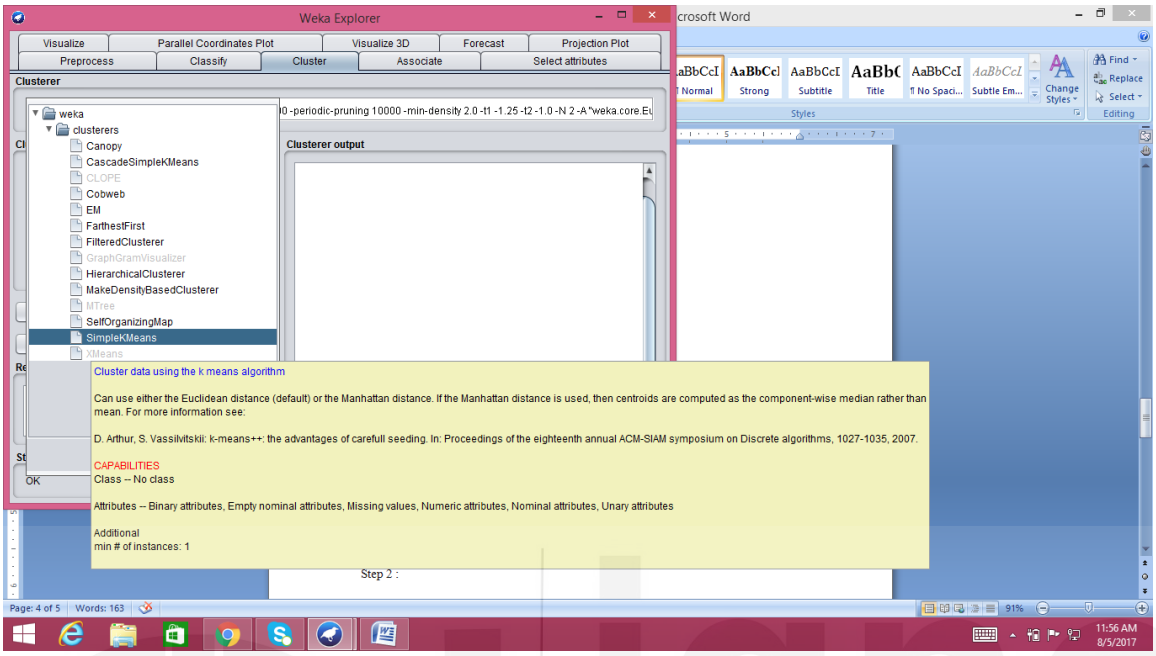
Aim : To Generate two clusters using k-means clustering algorithm

Step 1 : Select the sepal dataset from weka data folder and click on the open tab to load the iris.arff data set.

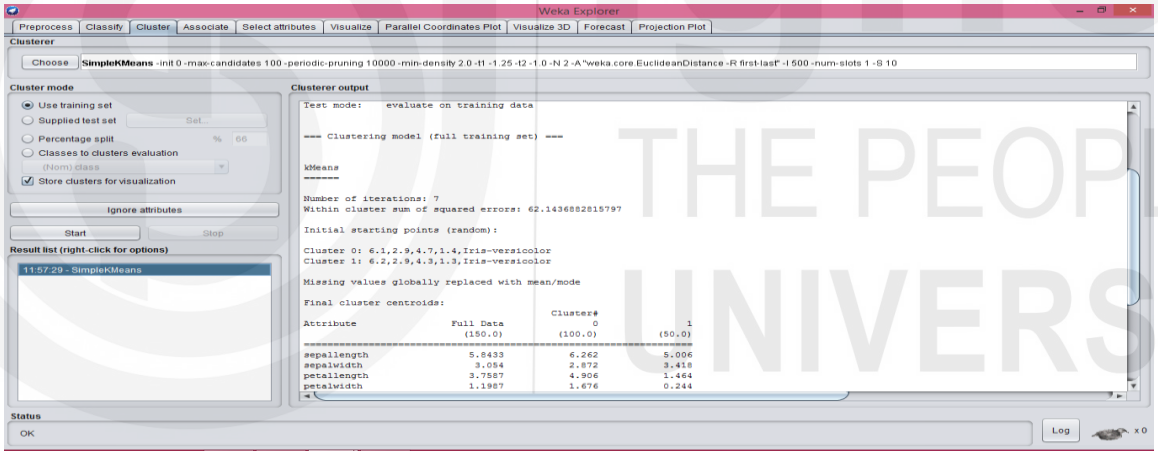




Step 2 : Select the cluster tab and choose SimpleKMeans option to perform K-Means clustering. Now, click on start button.

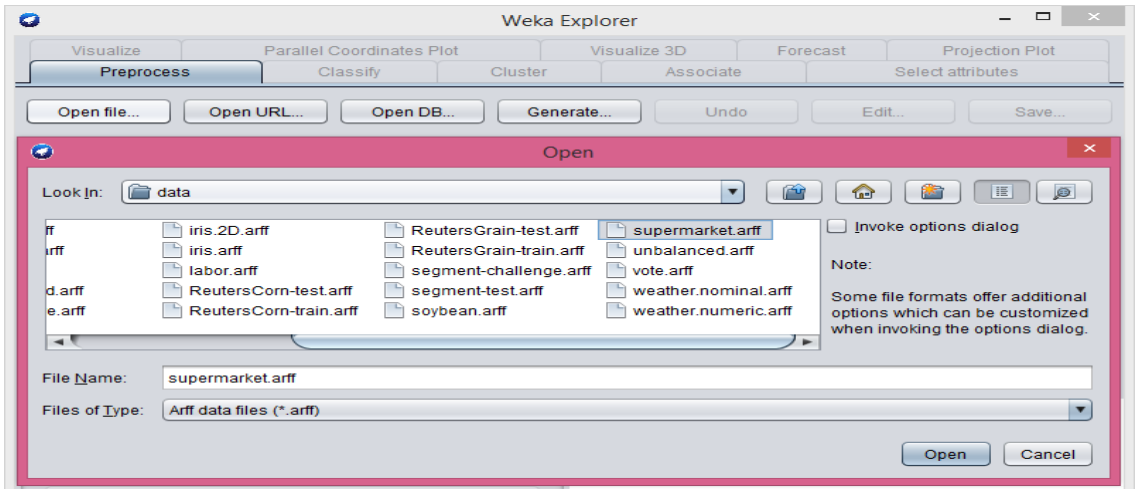


Step 3 Result of K-means Clustering will display the centroid of each of the clusters.

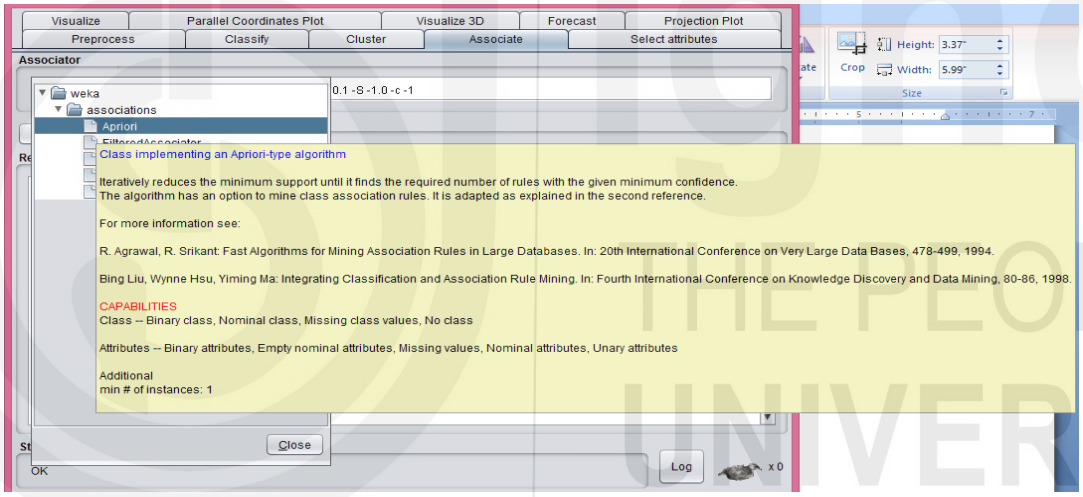


Aim : Apply apriori algorithm using supermarket data set

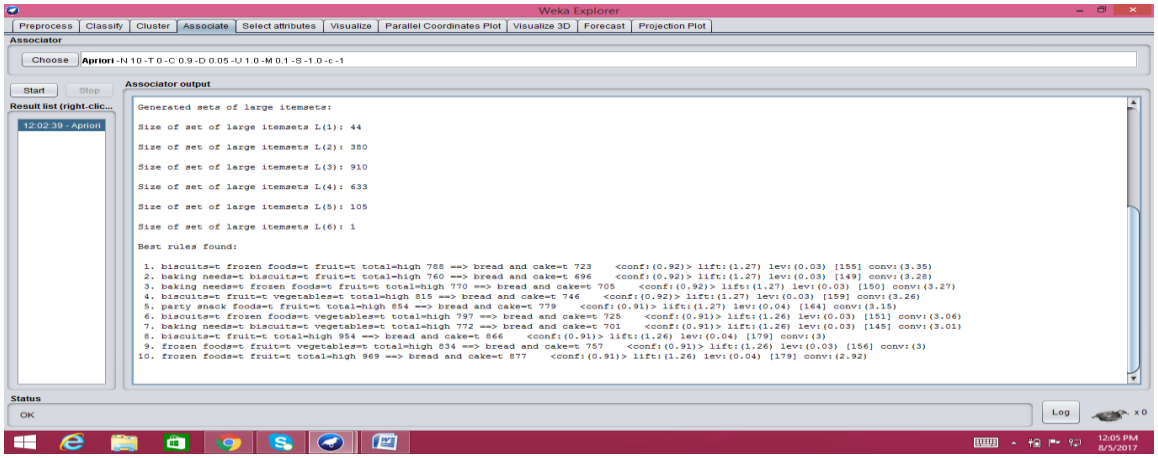
Step 1 : Open Weka explorer and load the supermarket dataset.



Step 2 : choose the Apriori option from Associate tab and then click on start button.

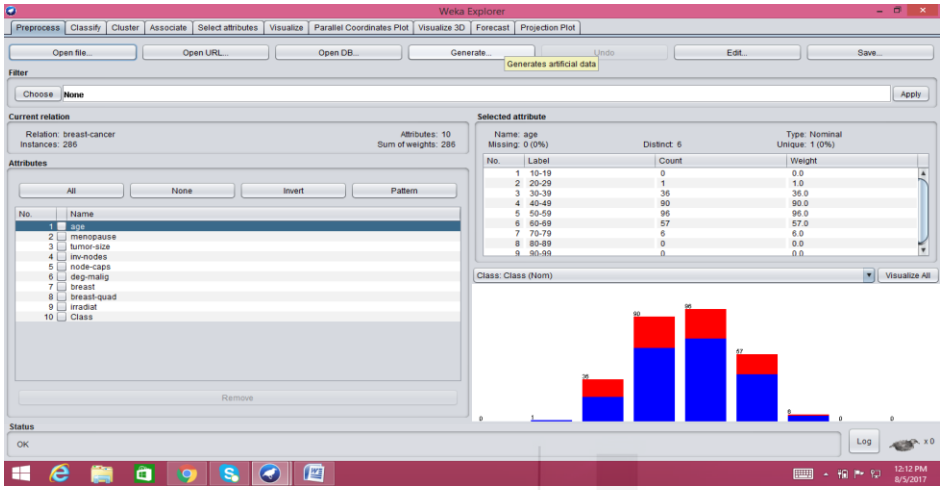


Step 3 Result of Apriori Algorithm

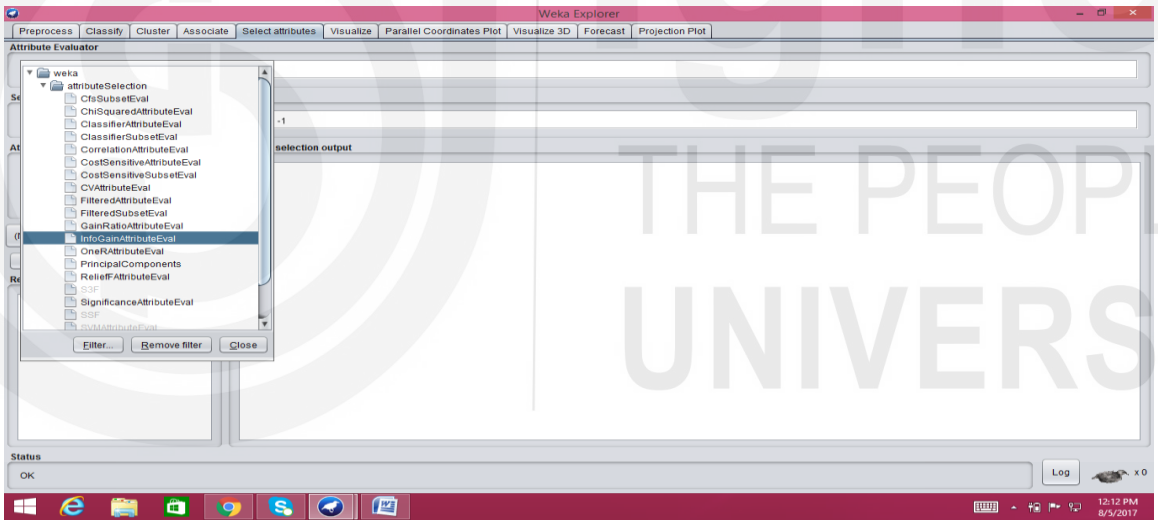


Aim : Implement Informaiton Gain feature selection algorithm

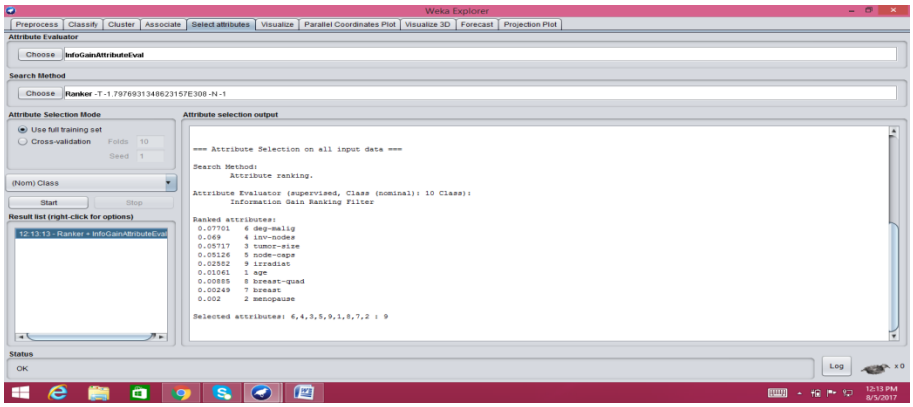
Step 1 : Select the data set



Step 2 : Choose the select attribute tab and select the attribute evaluator and search method by clicking on InfoGainAttributeEval.



Step 3 : Result of the feature selection



Space is given below for writing your answer.

a) Compare your answer with the one given at the end of the unit.

Check your progress 2

Note:a) Space is given below for writing your answer.

b) Compare your answer with the one given at the end of the unit.

i)List and describe the findings on applying K-means clustering and decision tree on the diabetes data set using Weka?

……

……

……

……

……

……

……

……

……

……

……

……

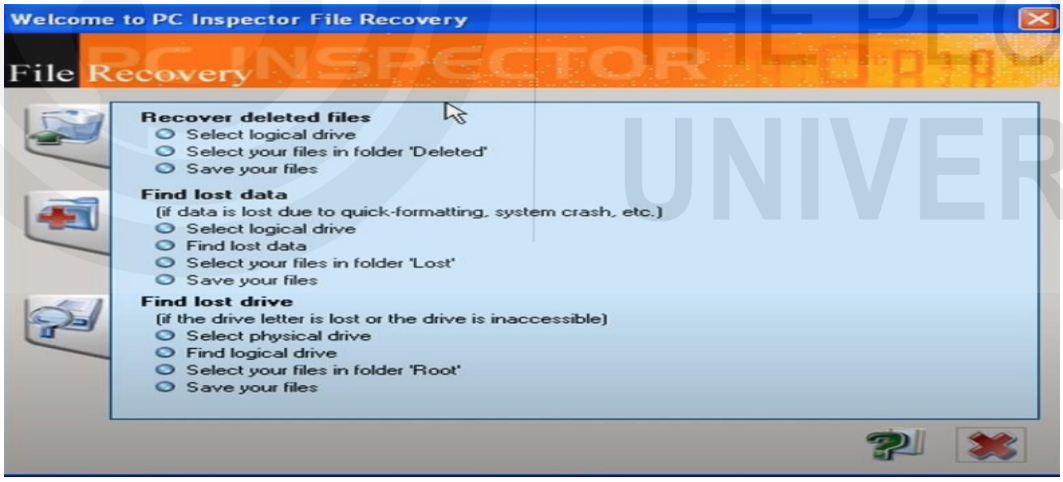
……
……

4.3 Data Recovery using PC Inspector

Crashing of hard drive is an annoying issue. However, the file allocation table, which holds filenames and refers to data on the disc, is typically what is lost when a hard drive crashes rather than the data itself. Consequently, if you have the right tools and the right conditions, it is technically feasible to recover data. PC Inspector File Recovery is a data recovery application which can be downloaded for free. It can be installed in Windows to recover the lost data on any partition on a device other than C. Install the application on some other computer and connect the damaged hard drive as a secondary drive on that machine if the complete hard disc is broken and is unable to boot.

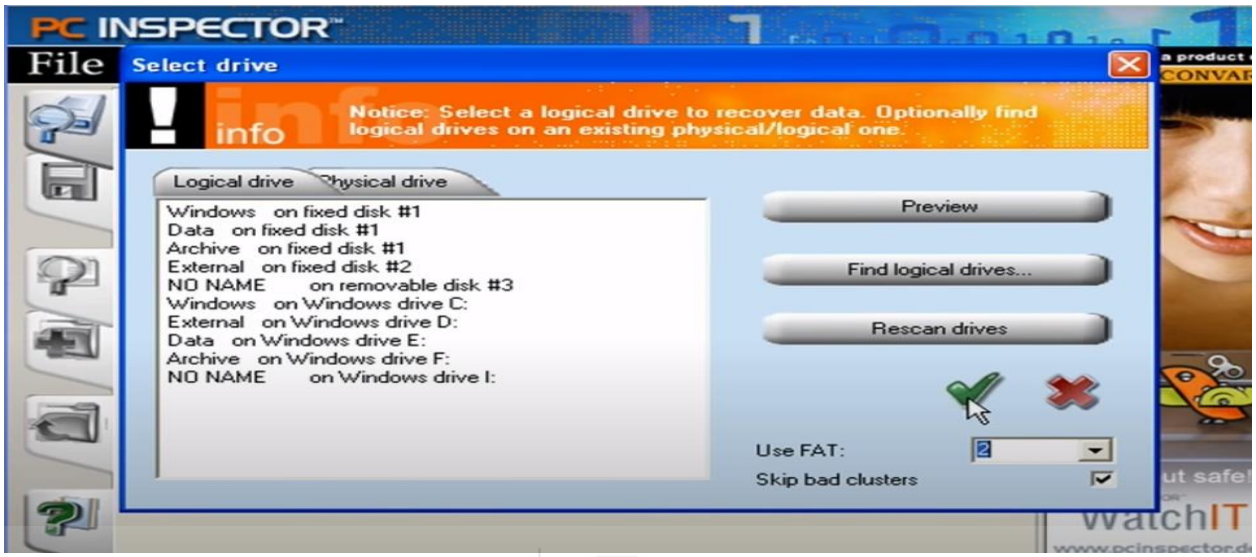
1. Launch the Program

Select English as the language when PC Inspector File Recovery is launched, and then click the green "tick" button. By selecting one of the three options to the left, select the type of data be recovered. In case of Lost Data, go to step 6 or go to step 8 for Lost Drive. Click the symbol in the upper left in order to recover deleted data.



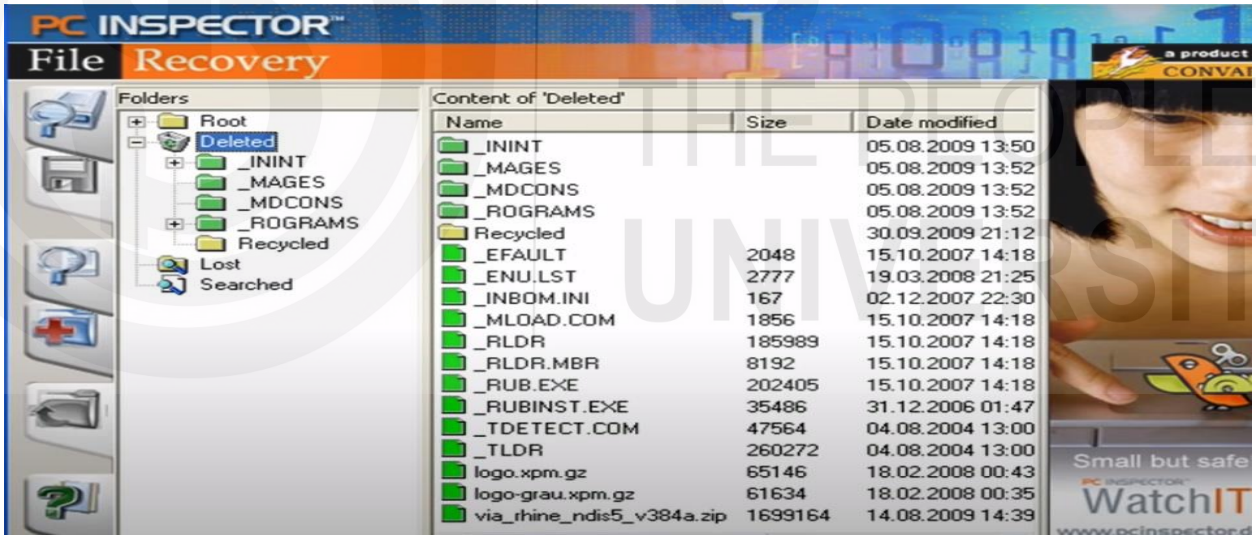
2. Select the Drive

The application will take its time searching the hard disc for any accessible partitions before displaying them. Very likely, the Logical Drives tab will allow you to choose the drive. Pick the entry that contains the drive's letter if it appears twice. Select the Preview button to see the contents of the drive. To move on to the following stage, click the tick button.



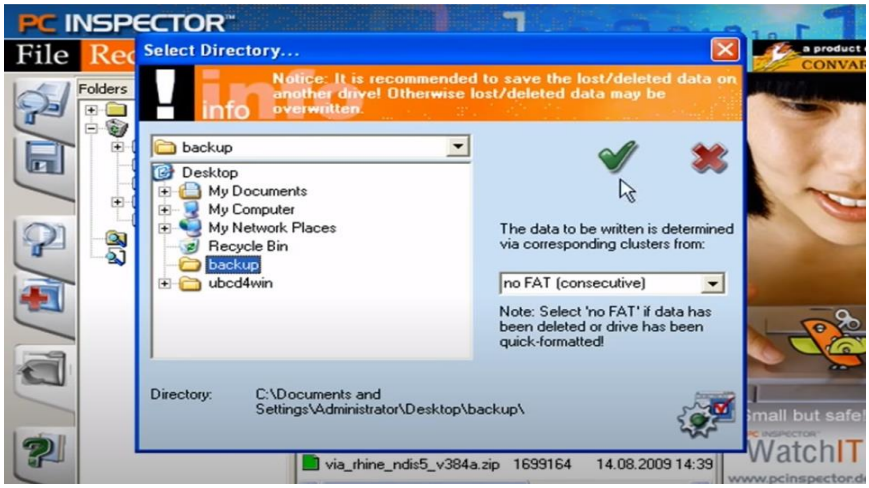
3. Sniffing Out Deleted Files

The chosen drive will now be examined; this process might take a while. The Recycle Bin symbol will be displayed alongside a Deleted folder. To find the files to be restored, search through this folder. It's possible that the actual filename was changed. To search file: choose Object > Find, specify the file type, and then click the green search button.



4. Recovering the File

After the search is complete, a list of files that match will be displayed. The Size and Date Modified columns might be used to assist locate the file being searched for if it doesn't show up (assuming you know these details). To organise these columns properly and make it simpler to find the file, click on the top of each one. Keep clicking on the potential candidates while holding down the [Ctrl] key to choose them. Right-click on them while holding [Ctrl] down, then select Save to. Decide on a place.



5. Check and Rename

Attempt to access the files after they have been restored to determine whether they are the ones being searched for. Once done, you may give them their original names and later copy them back to their original spot. Go to PC Inspector File Recovery and attempt search again if the files are not found yet.

6. Find Your Lost Data

Data can occasionally be lost as a result of a fast format or a system or software crash. In these circumstances, select the Lost Data button (the middle button on the left) from PC Inspector File Recovery's home screen. In this situation, data is recovered in a manner akin to that of deleted files. When the Select Cluster Range dialogue box displays, select the drive from the list of logical drives by clicking on the tick icon. It will take a little time to identify the files that need to be retrieved.

7. Retrieve the Lost Data

Observe hundreds of "lost" files—many of which are fragments—have been recovered by PC Inspector File Recovery. Going through them all will take a lot of time, but it will be worthwhile since important files may be recovered. To find a file and verify its integrity, repeat steps 3 to 5.

8. Find the Lost Drive

One might not be able to see any drives from the impacted hard disc in the list of logical drives if the partition table has been compromised. It need to be searched manually. Select the hard drive, which is typically referred to as fixed disc #1, by clicking on the Physical drive tab in the Select Drive box. To find logical drives, choose Find.

9. Search within Clusters

If the physical location of patition on the disc is known, the sliders may be used to narrow the search to that area rather than scanning the whole drive for missing drives. When the search is

finished, click the tick button again, and then choose the logical disc to recover your files or folders (as explained in the previous steps).

4.4 Let Us Sum Up

This unit covers the fundamentals of mobile forensics and data recovery using the autopsy tool. This unit is focused on android forensics specifically, and SQLite analysis of blackberry messenger. In the later sections, the data recovery using autopsy is discussed to make the readers aware about the data recovery process.

Check your progress 3

What are the findings of using PC Inspector for recovering the deleted data?

……
……
……
……
……
……
……
……
……
……
……
……
……
……

4.5 Session wise- list of lab assignments

Now, we try solving the problems related to data analytics and data recovery.

Session -1

c) Build a Data Warehouse and Explore WEKA.

Session -2

c) Generate a Decision Tree (C4.5) using weka.

Session -3

c) Apply apriori algorithm a supermarket data set.

Session -4

c) Implement Information Gain feature selection algorithm on an appropriate dataset.

Session -5

c) Demonstrate the steps to perform Regression on a data set.

Session -6

c) Demonstrate the steps to perform classification on a data set.

Session -7

c) Demonstrate the steps to perform clustering on a data set.

Session -8

c) Perform data preprocessing and demonstrate the steps to apply association rule mining on an appropriate data set.

Session -9

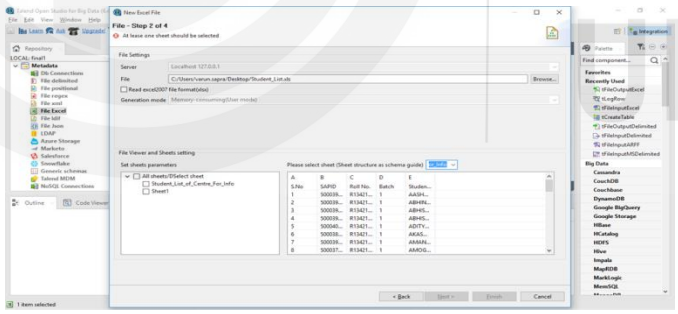
c) Generate two clusters using k-means clustering algorithm.

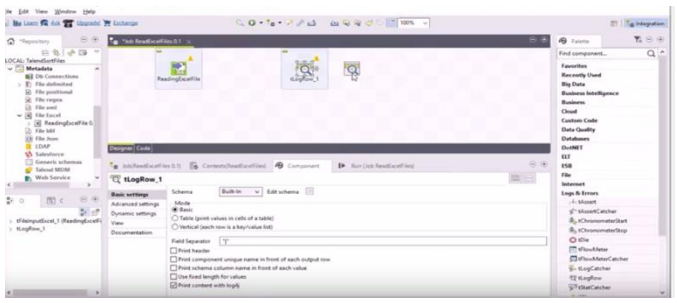
Session -10

b) Explore the functionality of talend.

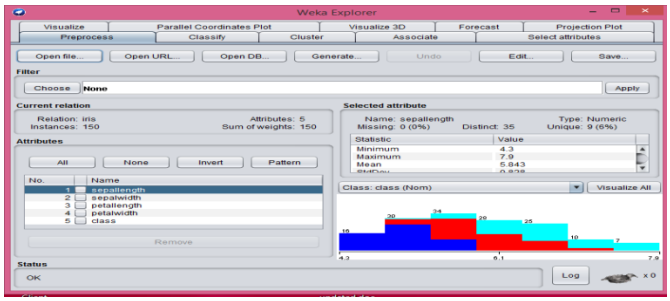
4.6 Check your Progress: The Key

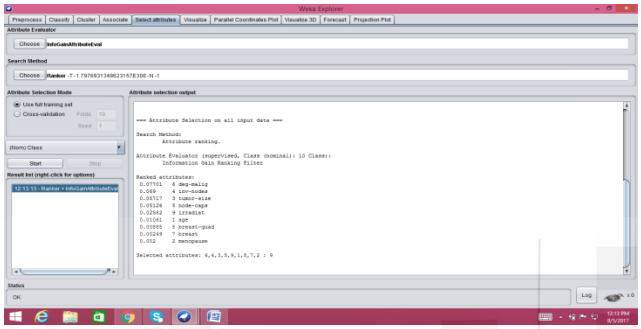
1. Below are a sample of findings of performing data analytics using talend studio.





2. Below are the screenshots of the findings:





3. Below are the findings of using PC Inspector for recovering the deleted data: