Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

Assignment No. 1

Title :- Import the legacy data from diffrent sources such as (Excel, SQL server, Oracle etc.) and load in the target system.

Objectives :- To introduce the concepts and components of Business Intelligence (BI)

Outcome :- Apply basic priaciples of elective subjects to problem solving and modeling.
Use tools and techniques in the area of software development to build mini projects.

Pre - requisites :-
1. Basics of dataset extensions.
2. Concept of data import.

Theory :-

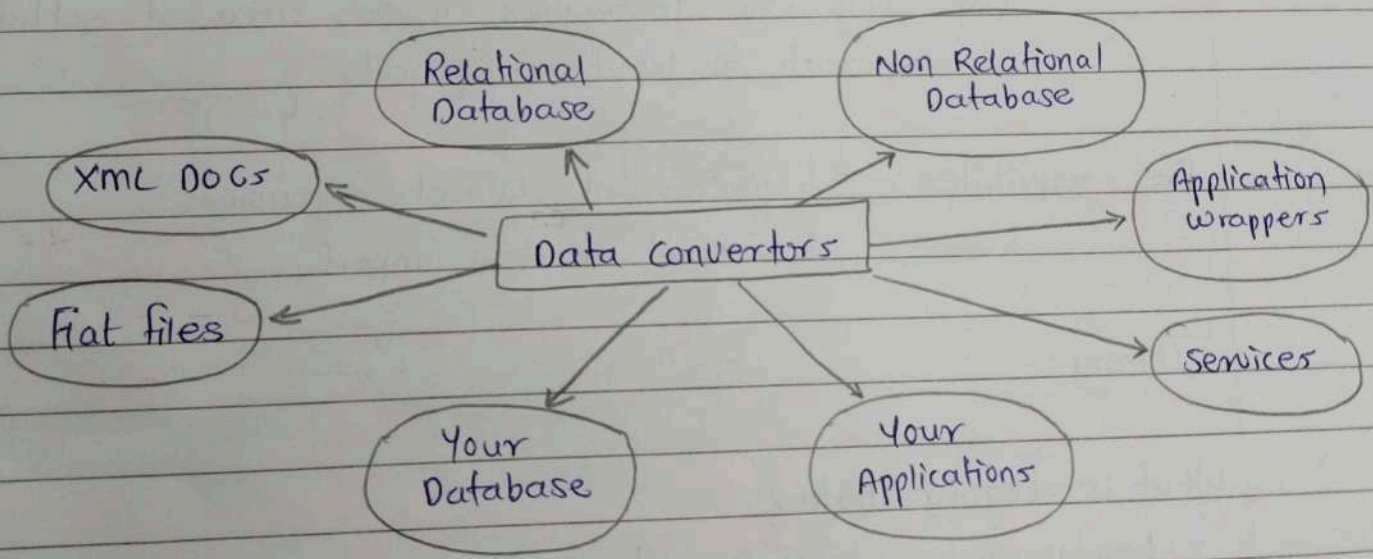
What is Legacy Data?

- Legacy data, according to Business Dictonary, is "Information maintained in an old or out of-date formator computer system that is consequently challenging to access or handle.

Sources of legacy data..

- There are many sources from which you may obtain legacy data..

This includes existing databases, often relational, although non-RDBs such as hierrarchical, network, object, XML, object/relational databases, and NoSQL databases.
Files, such as XML documents or "flat files" such as configuration files and comma-delimited text files, are also common sources of legacy data.
Software, including legacy applications that have been wrapped and legacy services such as web services such as web services or CICS transactions, can also provide access to existing information. The point to be made is that there is often far more to gaining access to legacy data than simply writing an SQL query against an existing relational database.



Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

Assignment No.2.

Title :- Perform the Extraction Transformation and Loading (ETL) process to construct the database in the sqlsever..

Objective :- To introduce the concepts and components of Business Intelligence.

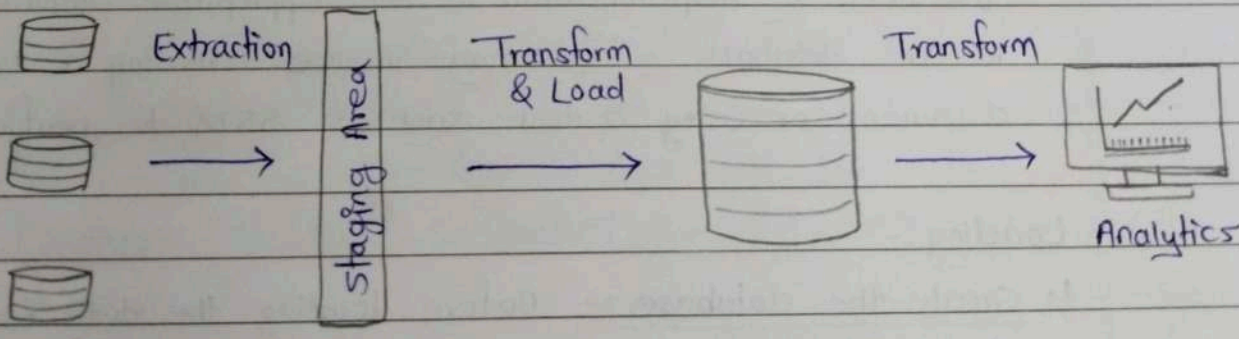
Outcome :-
1. Apply basic principles of elective subjects to problem solving and madelling.
2. Use tools and techniques in the area of software develoment to build mini projects.

Pre-requisites :-
1. Basics of ETL tools.
2. Concept of SQL sever.

Theory :-

ETL (Extract, Transform, and Load)

ETL is a process in Data warehousing and it stands for Extract, Transform & Load..



- It is a process, in which an ETL tool extracts the data from various data source systems, transforms in the staging area and then finally, loads in into the data warehouse system.

Extraction.

1. Identify the data sources: The first step in the ETL process is to identify the data sources.
2. Extract the Data: Once the data sources are identified we need to extract the data from them. This may involve writing queries to extract the relevant data or using tools such as SSIS to extract data from files or databases.
3. Validate the data:- After extracting the data, its important to validate it to ensure that it's accurate and complete. This may involve performing data profiling or data quality checks.

Transformation.

1. Clean & Transform the data :- The next step in the ETL process is to clean and transform the data. This may involve removing duplicates, fixing invalid data or converting data types. We can use tools such as SSIS or SQL transcripts to perform these transformations.
2. Map the data :- Once the data is cleaned and transformed, we need to map the data to the appropriate tables and columns in the database. This may involve creating a data mapping document or using a tool such as SSIS to perform the mapping.

Loading :-

1. Create the database :- Before loading the data, we need to create the database and the appropriate tables.

2. Load the data :- Once the database and tables are created we can load the data into the database. This may involve using tools such as SSIS or writing SQL scripts to insert the data into the appropriate tables.
3. Validate the Data :- After loading the data, it's important to validate it to ensure that it was loaded correctly. This may involve performing data profiling or data quality checks to ensure that the data is accurate & complete.

Steps to setup database in SQL server.

1. Open SQL server management studio to restore backup files.
2. Right click on Databases Restore Database.
3. Click on towards end of device box
4. Click on Add select path of backup files
5. Select both files at a time
6. Click ok and in select backup devices window Add both files of Adventure works
7. Open SQL database server Data Tools.
   - select file New project Business Intelligent Integration Services Project & give appropriate project name.
8. Right click on connection managers in solution explorer and click on New connection manager. Add the SSIS connection manager window
9. Select OLEDB connection manager and click on Ad.
10. Configure OLEDB connection manager window appear click on New.
11. Select server name from drop down and databasename and click on test connection.

If the test connection succeded, click on Ok.

12. Click on Ok.
    - Connection is added to connection manager.
13. Drag and Drop Data flow task in control flow tab.
14. Drag OLE DB source from other sources and drag into Data flow tab.
15. Double click on OLE DB
    select [sales].[Store] table from drop down → Ok.
16. Drag OLE DB destination in the data flow tab & connect both.
17. Double click on OLE DB destination.
18. Click on start
19. Go to SQL server management studio
20. Execute the following query to get output..

```
>>> Select [Adventure works 2012] GO
    SELECT [Business Entity ID],
    [Name],
    [SalesPerson ID],
    [Demographics],
    [rowguid],
    [Modified Date],
    FROM [dbo].[OLE DB Destination] GO
```

Conclusion :- In this way we can perform ETL process to construct a database in SQL server..

Pune District Education Association's
College Of Engineering
Manjari (Bk.), Hadapsar, Pune-412307.
Accredited by NAAC

Assignment No. 3

Title :- Create the cube with suitable dimension and fact tables based on ROLAP, MOLAP and HOLAP model.

Objective :- To introduce the concepts and Business Intelligence

Outcome :-
1. Apply basic principles of elective subjects to problem solving and modelling.
2. Use tools and techniques in the area of software development to build mini projects.

Theory :-

What is a fact Table ?
- A Fact table is a table that stores quantitative data or facts about a business process or activity. It is a central table in a data warehouse that provides a snapshot of a business at a specific point in time.
- For example - A fact table in a retail business might contain sales data for each transaction, with dimensions such as date, product, store and customer. Analysts can use the fact table to analyze trends and patterns in sales, such as which products are selling the most, which stores are performing well, and which customers are buying the most.

What is ROLAP, MOLAP & HOLAP model.

- ROLAP, MOLAP and HOLAP are 3 layers of models used in BI for organizing and analyzing data :

1. ROLAP (Relational Online Analytical Processing).

- In this model, data is stored in a relational database, and the analysis is performed by joining multiple tables. ROLAP allows for complex queries and is good for handling large amounts of data, but it may be slower due to the need for frequent joins.

2. MOLAP (multidimensional online Analytical Processing).

- In this model, data is stored in a multidimensional database which is optimized for fast query performance. MOLAP is good for analyzing data in multiple dimensions, such as time, geography and product but may be limited in its ability to handle large amounts of data..

3. HOLAP (Hybrid Online Analytical Processing).

- This model combines elements of both ROLAP and MOLAP. It stores data in both a relational & multidimensional database, allowing for a efficient analysis of both large amounts of data and complex queries. HOLAP is good compromize between the other two models, offering both speed & flexibility..

BI

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

a-1

Assignment No. 4

Title :- Import the data warehouse data in microsoft excel and create the Pivot table and Pivot chart.

Objective :- To introduce the concept & components of BI

Outcome :-

1. Apply basic principl-es of elective subjects to problem solving and modelling.
2. Use tools and techniques in the area of software development to build mini project.

Theory :-

What is data warehouse?

- A data warehouse is a centralized repository of integrated and transformed data from multiple sources within an organization. It is designed to support business Intelligence (BI) activities, such as data analysis, reporting and decision-making.

What is Pivot Table & Pivot Chart?

- A pivota table is a powerful tool in spreadsheet software that allows you to summarize and analyze large datasets by grouping and summarizing data in diffrent ways. Pivot tables allows you to quickly create tables that show a summary of data based on specific criteria or dimensions.

For example, you can use a pivote table to summarize sales data by region or by product category. A pivot chart is a graphical representation of the data in a pivot table.
- Pivot charts are especially useful when dealing with large amounts of data, as they can help identify patterns and trends that might not be immediately obvious from raw data..

3. Steps for creating a pivot table in Google sheets.
-
1. Open a Google sheets document with the data you want to use for the pivot table.
2. Select the range of data you want to use for the pivot table.
3. Click on the "Data" tab in the top menu. then click on "Pivot table".
4. In the "Create Pivot Table" dialog box, select the range of data you want to use for the pivot table & choose where you want to place the pivot table
5. Click on "create".
6. In the pivot table editor, drag & drop, the columns you want to use for the pivot table into the "Rows", "Columns" & "Values" sections.
7. To ~~do~~ add a filter to the pivot table, drag a column into filter section..
8. To customize the values in pivote table, click on the drop-down menu in the "Values" section & choose the type of calculation..
9. customize any additional options in the pivot table editor
10. click on "Update" to apply the changes and create the pivot table..

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

Assignment - 01

~~Title :-~~

4. Steps for Creating a Pivot chart in Google Sheets.

1. Open a Google Sheets document with the data you want to use for the pivot chart.
2. Select the range of data you want to top menu, click on "Pivot table".
3. Click on the "Data" tab in the top menu.
4. In the "Create Pivot table" dialog box, slect the range of data you want to use for the pivot table & choose where you want to place the pivot table & choose where you want to.
5. Click on "Create".
6. In the pivot table editor, drag & drop the columns you want to use for.
7. Click on the "chart" tab in the pivot table editor.
8. Choose the type of chart you want to use for the pivot chart from the drop-down menu.
9. Customize the chart options to your liking.
10. Click on "update" to apply the changes and create the pivot chart.

Conclusion :- In this way we pivot table and pivot chart using Google Spreadsheet / Excel..

5-1

Assignment No. 5

Title :- Perform the data classification using classification algorithm. or perform the data clustering using a clustering algorithm.

Objective :- Perform to introduce the concepts & components of Business Intelligence.

Outcome :-

1. Apply basic principles of elective subjects to problem solving & modeling.
2. Use tools and techniques in the area of software development to build mini projects.

Theory :-

What is Clustering and Classification ?

- Clustering and classification are two important techniques used in bioinformatics to analyze biological data. Clustering is the process of grouping similar objects or data points together based on their similarity or distance from each other.
- In bioinformatics clustering is often used to group Genes or proteins based on their expression patterns or sequences. clustering can help identify patterns and relationships between different genes or protiens, which can provide insights into their biological function & interactions.

Classification, on the other hand, is the process of assigning a label or category to a new observations based on its features or characteristics. In bio informatics, classifications often used to predict function or activity of a new gene or protien based on its sequence or structure.

- Classification can help identify new drug targets or biometers for disease diagnosis and treatment.

Both clustering and classification are important tools for analyzing a large and complex biological datasets and can provide valuable insights into the underlying biological processes.

Clustering in Tableau.

1. Connect to the data: Connect to the data set that you want to cluster in tableau.
2. Drag & Drop the data fields: Drag & drop the data fields into the view, and select the data points that you want to cluster.
3. Choose a clustering algorithm: Select a clustering algorithm from the analytics pane in tablue.
4. Define the number of clusters: Define the number of clusters that you want to create. You can do this manually or let Tablue atomatically determine the optimal number of clusters.
5. Analyze the clusters: Visualize the clusters and analyse them using Tableaus built-in visualizations and tools.

Classification in Tableau:

1. Connect to data: connect to the data set that you want to classify in tableau.

5.2

2. Drag and drop the data fields: Drag and drop the data fields into the view and select the target variable that you want to predict.
3. Choose a classification algorithm: select a classification algorithm from the analytics pane in Tableau. Tableau provides several built in classification algorithms, such as Decision trees & random forest
4. Define the ~~dem~~ model parameters: Define the model parameters, such as the maximum tree depth or the number of trees to use in the forest.
5. Train the model - Train the model on a subset of the data using Tableaus built in cross validation functionality.
6. Evaluate the model - Evaluate the model accuracy of using Tableau's built-in metrics, such as confusion matrix, precision, recall, and F1 score.
7. Predict the target variable - Use the trained model to predict the target variable for new data.
8. Visualize the results - Create visualizations to communicate the result of the classification analysis using Tablue's built-in visualization tools.

Conclusion :- In this way we implement classification & clustering using Tableau.

B1

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

Mini Project

Problem Statement:- A BI report must be prepared outlining the following steps

a) Problem definition, identifying which data mining task is needed.

b) Identify and use a standard data mining dataset available for the problem.

Problem Description :- In this project, we will develop and evaluate the performance and the predictive power of a model trained and tested on data collected from houses in Boston's suburbs. Once we get a good fit, we will use this model to predict the monetary value of a house located at the Boston's area.

- A model like this would be very valuable for a real estate agent who could make use of the information provided in a daily basis.
- Dataset used is the Boston Housing Dataset.

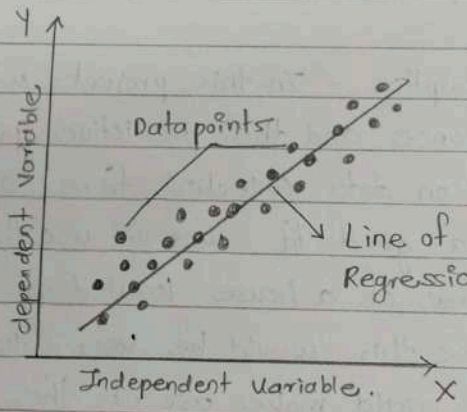
Theory :-

Linear Regression in ML

- Linear regression is one of the easiest and most popular ML algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continous/ real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm allows a linear relationship between a dependent (y) and one or more independent (y) variables hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables.



Mathematically, we can represent a linear regression as,

$$y = a_0 + a_1 x + \varepsilon$$

here,

Y = Dependent variable (Target variable)

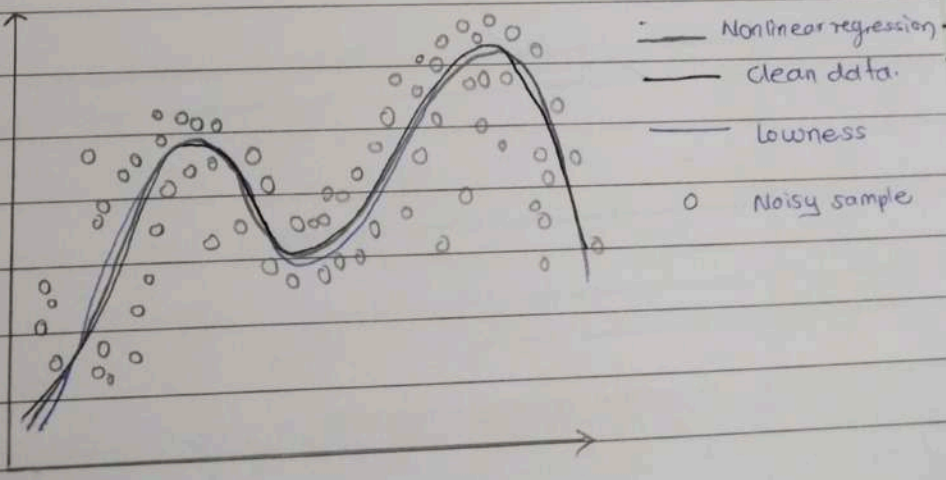
X = Independent variable (predictor variable)

a_0 = intercept of the line (Gives an additional degree of freedom)

a_1 = Linear regression coefficient (scaler factor to each input value).

Support Vector Regression.

- It is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyperplane that has the maximum number of points.
- Unlike other Regression models that try to minimize the error between the real & predicted value. The SVR tries to fit the best fit line within a threshold value. The threshold value is the distance between the hyperplane & boundry line. The fit time complexity of SVR is more than quadratic with the number of samples which makes it hard to scale to datasets with more than a couple of 10000 samples.
- For large datasets, linear SVR or SGD Regressor is used. Linear SVR provides a faster implementation than SVR but only considers the linear kernel. The model produced by support vector Regression depends only on a subset of the training data, because the cost function ignores samples whose prediction is close to their target.



Conclusion:- We have made a machine learning regression project from end to end and we learned and obtained several insights about regression models and how they are developed..

SDN

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

1-1

Assignment No. 1

Aim :- Prepare setup for miniet network evaluation environment with the help of Virtual Box and minimet. Demonstrate the basic commands on mininet & emulate diffrent custom network ~~topot~~ topology. view flow tables.

Software Requirements :- Ubuntu 16.04

Hardware Requirement :- Pentinum 7v system with least configuration

Theory :-

The mininet vm is meant to speed up mininet installation. plus make it easy to run on non-linux platforms. Then vm works on coindow, MAC linux though vmware, VirtualBox QEmv & KVM.

After downloading the vm, you'll can a few steps to customize it for your setup. This wonit take long vm setup.

1. Download the mininet vm.

- Download the mininet vm from https://github.com/mininet/wiki/mininet-vm-images

The vm comes out to 1 GB compressed & ≈ 2GB uncompressed. It is an OVF (open visualization format) virtual machine image which can be imported by most virtual machine monitors.

Pune District Education Association's College Of Engineering Manjari (Bk.)

Download & install a visualization program as: Vmware workstation, for windows or linux or Vmware fusion for mac. Virtual Box (free/GPL). or any platform, or (free/GPL) for Linux. Virtual Box is free to download & distribute which is definite advantage?
Boot vm.
Add the vm & start it up, in the virtualization program of your choice.

Virtual Box :-
1. Visually you can just double click on the out file, import it.
2. If you get errors importing the .ouf file, you can simply create a new vm of the appropriate type (eg. Linux, Ubuntu - 64 bit) & use the .vmdk file as the virtual hard disk for the new vm.
3. select "settings" & add an additional host only network adaptor that you can use log in to the vm image. Start the vm.
4. For more information on setting up workint in virtualBox you may wish to check out these virtualBox specific instructions.

Vmware: Import the OVF. file, then start the vm VMWare may ask you to install vmware tools on the vm - if it asks, everything graphically the tutorial is done via X - forwarding through SSH (in fact the vm doesn't have a desktop manager installed), so the vmware tools are unnecessary unless you wish to install an X11/G have / etc. environment in your vm.

1-2

@emu/Kvm:
for @emu, something like the following should work:
qemu-system-x86_64 -m 2048 mininet-vm-disk1.vmdk
-net nic-model=virtio -net
user, net=192.168.101.0/24, host fwd=tcp::8022-22

for kvm
Sudo qemu-system x86-64. machine accel=kvm. m2048
mininet-vm-disk. vmdk-net nic. model=virtio-net
user, net=192.168.101.0/24. host/fwd=tcp:: 8022-22

The above commands will setup ssh forwarding, from the vm to host port 8022.
For 32 bit vm image, use qemu-system-i386.
Parallels: Use parallels transporter to convert the .vmdk file to an .hdd image that parallels can use & then create a new vm using that .hdd image as its virtual drive. start the vm. log in to vm.
log in to the vm, using the following name & password.

mininet-vm login: mininet
Password: mininet

(some older vm images may use openflow / openflow instead).
The root account is not enabled for login, you can use sudo to run a command with super-uses privileges.
SSH # into vm.

first, find the vm's IP address, which for vm-ware is probably in the range 192.168.x.y. In the vm console.

if config. etho.

Note: VirtualBox users who have setup a host only network on eth1 should use.

sudo dhclient eth1 # make sure that eth1 hash an IP address
ifconfig. eth1

you many want to add the address to your host PC's /etc. hosts files to be able to SSH in by name if its Unix like. for example, add a line like this for OS X.

192.168.x.y. mininet-vm.

where 192.168.x.y. is replaced by the vm's IP address. SSH into the vm, we assume the vm is running locally & that the additional precautions of ssh-x are unnecessary. ssh-y also has no authentication timeout by default.

ssh -y mininet@ mininet-vm

Conclusion:- Thus we learnt mininet network emulation environment with the help of virtual box & mininet.

SDN

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

Assignment No. 2.

Aim :- After studying, open source Pox & flood light controller. Install controller & run custom topology using remote controller like Pox & flow/light controller, recognize inserted flows by controllers.

Software Requirements :- Ubuntu 16.04

Hardware Requirements :- Pentium 2v system with latest configuration.

Theory :-

Using the Pox SDN controller

Pox components.

- Pox components are additional python programs that can be invoked when pox is started from the command line.

These components implements the network functionality in the software defined network.

Pox comes with some stock components already available.

The Pox stock components are documented in the ~~box~~ pox wiki & the code - for each component can be found in the ~/pox/pox directory on the mininet 2.2 vm image.

for example. the forwarding, l2 - learning, components is the ~/Pox/ Pox/ forwarding directory as seen below..

```
$ cd Pox/ Pox
$ ls
boot. py datapaths info log protolk. py
boot. py6 forwarding. init - py messenger. py. py topology.
core. py help. py_ init_ pyc min py. pyc web
core. pyc host_ tracker /ib. Openflow samples
$ cd forwarding..
$ le
hub. pyl2_ flowvisor. py l2_px. pyl3_ learning. py
_init_ . py l2_learning. py l2_Px_ self _learning . pyl3_ learning py
_init_ - pyc l2_ multi py l2_ pairs. py topo - proactive. py
$
```

Programming for Pox.

The general purpose of all SDN controllers, including Pox, is to allow users to write their ow applications that use the controller as an intemediary - or abstraction layer - between network application & the network equipment.

To learn how to write applications for Pox, developers may study the stock, Pox components as examples that show how to write their own components or they any review the Pox API documentation to learn how to write networking applications that use.

the POX python API.

Installing Pox :
Pox comes already installed on the mininet 2.2 VM image In this tutorial, we will use the VM image. see my previous post about. setting up the mininet 2.2 VM.

If you wish to install mininet & pox on your own linux system - either hardware or a virtual machine - you may use the mininet install script, which also installs the pox controller when it installs mininet.

Running Pox
- start Pox by running the pox.py program & specifying the Pox components to use. for example, to run pox so it makes the switches it controls emulate the behaviour of ethernet learning switches run the command.

mininet@ mininet-vm: ~ $ sudo ~/pox/pox.py forwarding.12 learning.

The Pox console.
The Pox console is the terminal session from which we can run the pox controller, after pox starts it will display log information & may optionally display an interactive python command line. if the pox is started with the py component.

Quite Pox.
- To Quit pox, use the control-c key combination in the pox console.

Highlights :-
- easy to setup with minimal dependencies.
- supports a board range of virtual & physical openflow switches.
- can handle mixed openflow & non-openflow networks.
- It can manage multiple "islands" of openflow hardware switches.
- Designed to be high performance..

Installation:
- Floodlight master class has been updated on 04/03 to Java 8.

Prerequisites:-
- Java development kit.
  - JDK 8 for floodlight master & above.
  - JDK 7 for floodlight V1.2 & below
  - Ant to build
  - python development package.
- To download dependencies for floodlight master & above.
  sudo apt-get install / build-essential / python-dev.
- To download dependencies for floodlight V.12. & below.
>>> Sudo apt-get install build essential openjdks. 7-jdk ant maven python-dev eclipse.

Download & build:
- Floodlight is simple to download from github & install. The "git clone" step below uses the master version of floodlight.

Pune District Education Association's
College Of Engineering
Manjari (Bk.), Hadapsar, Pune-412307.
Accredited by NAAC

Type the below command in SDN hub terminal.
$ git clone git://github.com/floodlight/floodlight.git
If you are having JDK 7 use a specific version specify the version branch in the "git clone".

$ git clone -b v1.2 git://github.com/floodlight/floodlight.git.
After installation complete, follow the steps to build floodlight controller.

we create single topology with 3 host & switches & are one controller.
https://127.0.8.1.8080/ui/index.html
In topology tab we can see our topology
switch can be veiw in switches tap.
Hostr can view ~~the~~ host seperately by xtream in mininet
mininet> xtam hin2.

ping each other host by ping command with host IP in node
h1 $ ping 10.0.0.2.
In node h2 $ ping 0.0.0.1.

Conclusion:- Thus we learnt symmetric & assymmetric implementation of particle swarm optimization for travelling salesman problem..

SDN

Pune District Education Association's
College Of Engineering
Manjari (Bk.), Hadapsar, Pune-412307.
Accredited by NAAC

3-1

Assignment No. 3

Title :- Create a SDN environment on mininet and configure a switch to provide a firewall functionality using pox controller.

Objectives :- 1. learn installation of mininet
2. Configured a switch to provide firewall functionality using POX controller.

Theory :-

SDN environment.

- An SDN environment is a network architecture that seperates the network control plane from the data plane allowing for centrallized management and programmability In an SDN environment, the network control plane is managed by a software controller, which communicates with the network devices using Openflow protocol.

To create an SDN environment, several components are required, including :

1) SDN controller : The SDN controller is the central management component that communicates with the network devices using the Openflow protocol.
2) Network devices : The network devices, such as switches & routers, are responsible for forwarding data packets based on the instructions from SDN controller.

3. Openflow protocol : The openflow protocol is used to communicate between the SDN controller and the network devices.

4. Network applications : Network applications can be developed to run on top of SDN environment. providing additional network services & functions.

5. Network topology : The network topology is the physical or logical arrangement of the network devices & links.

Firewall Functionality.

- It is a security feature that is used to protect computer networks from unauthorized access and malicious traffic. In the context of SDN, a firewall can be implemented using an SDN controller, which controls the network switches and enforces security policies.

- The firewall functionality implemented by an SDN controller typically involves setting up flow rules on the network switches to block or allow traffic based on certain criteria, such as the source and destination IP address, the transport protocol used, and the ports being used. For example, as SDN controller can setup flow rules to block traffic from certain IP addresss, or to block traffic on certain ports that are known to be used by malware.

POX controller.

- POX is an open-source SDN controller that is written in python and is designed to support the openflow protocol. POX is

designed to be lightweight and flexible, and it provides a number of useful features for building and managing SDN networks.

Some of the key features of POX include:

1. Modular design: POX is designed to be modular and extensible, which makes it easy to add new features.

2. Python based: POX is written in python, which makes it easy to read, understand, and modify the code. This also means that it is easy to integrate with other Python based tools and libraries.

3. Openflow support: POX supports the Openflow protocol which is used to communicate between the controller and the switches in an SDNetwork.

4. Debugging tools: POX includes a number of built-in debugging tools that make it easy to test and troubleshoot SDN applications.

5. Flexible API: POX provides a flexible API that can be used to build custom SDN applications and services.

Conclusion :- We learn how to install mininet and how to created an SDN environment on mininet and configured a switch to provide firewall functionality using POX controller.

4c1

Assignment No. 4.

Title :- Using mininet as an emulator and POX controller, build your own internet router. Write simple outer with a static routing table. The router will receive raw ethernet frames and process the packet forwarding them to correct outgoing interface. You must check the Ethernet frames are received and the forwarding logic is created so packets go to the correct interface.

Objective :- 1. To understand the concept of SDN
2. To use mininet & POX controller to find build an internet router.

Theory :-

Router

- A router is a device that connects diffrent computer network together. It helps to direct traffic on the internet by sending data packets to the correct destination.
- when a data packet is sent on the internet, it contains information about the sender and receipient as well as the actual data being sent. The router receives the packet and reads the information about the receipient. It then checks its own routing table to see where the reciepent is located.

- The routing table is a list of destinations and the corresponding paths that the router should used to send the data packets to the correct destination. Once the router has determined the correct path, it sends the packets to the next router or directly to the destinations if it is located on the same network.

- Routers are used to connect diffrent types of networks together such as local area networks (LANs) or wide area networks. They are alsoused to connect devices within a network, such as computers, printers another devices.

- Routers can perform other functions besides routing data packets. for example, they can act as firewalls to protect the network from unauthorized access, or they can perform network address translation to hide the internal network addresses from the outside world..

Ethernet frames :-

- These are the basic units of data in Ethernet networks They contain important information about the data being transmitted and how it should be forwarded to its destination. Here are some key points to explain Ethernet frames.

- Ethernet frames are composed of several fields that contain information about the data being transmitted. Some of these fields include:

SDN

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

1. Destination MAC address : The MAC address of the device that the frame is being sent to.
2. Source MAC address : The MAC address of the devices that sent the frame.

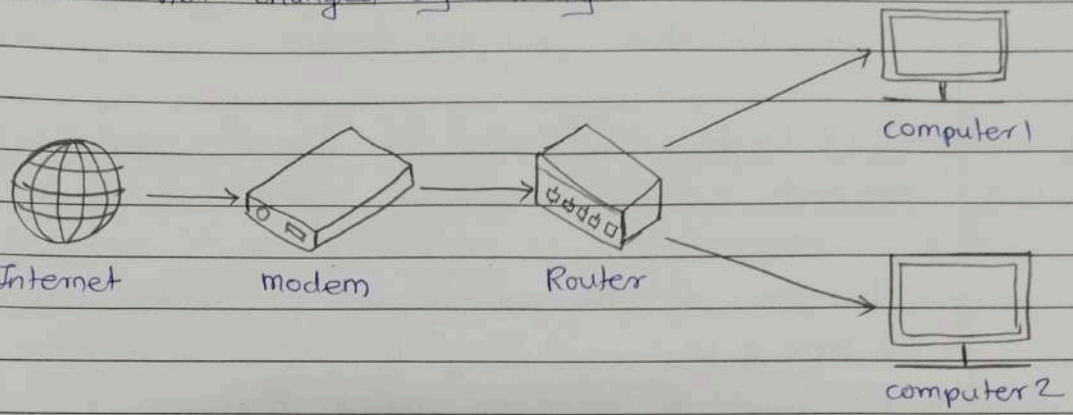
Type / length : Indicates the type of contained in the frame.
Payload : The data being transmitted.

- The size of Ethernet frames is limited to a maximum of 1518 bytes, including the header & trailer.
- Ethernet frames are used to transmit data between devices on an Ethernet network, such as a local area network (LAN).
- When an Ethernet frame is transmitted, it is first sent to the switch or router closest to the source device. The switch or router then examines the destination MAC address in the frame to determine which interface the frame should be forwarded to.

- If the destination MAC address is not known, the switch or router may broadcast the frame to all devices on the network to try to find the devices with the correct MAC address.

Routing tables :-

- Routing tables are used by routers to determine the next hop for packets based on their destination IP address.
- A static routing table, is one that is manually configured and does not changed dynamically.



Conclusion :- Thus, we learned how to build our own internet router using mininet as an emulator or POX controller.

SDN

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

Assignment No. 5

Title:- Emulate and manage a Data center via a Cloud Network Controller : create a multi-rooted tree-like (clos) topology in mininet to emulate a data center. Implement specific SDN application on top of the network controller in order to orchestrate multiple network tenants within a data center environment. in the context of the network visualization and management.

Objectives :- 1. To understand the concept of Software Defined Networking (SDN).

2. To emulate and manage a data center environment using a cloud network controller.

3. To learn how to create a multi-rooted tree-like (clos) topology in minenet & implement specific ~~on~~ SDN applications to orchestrate multiple network tenants within a data center environment.

Requirements: 1. Computer with any Linux OS installed.

2. Python Runtime environment.

3. Mininet Emulator

4. Pox controller.

Theory :-

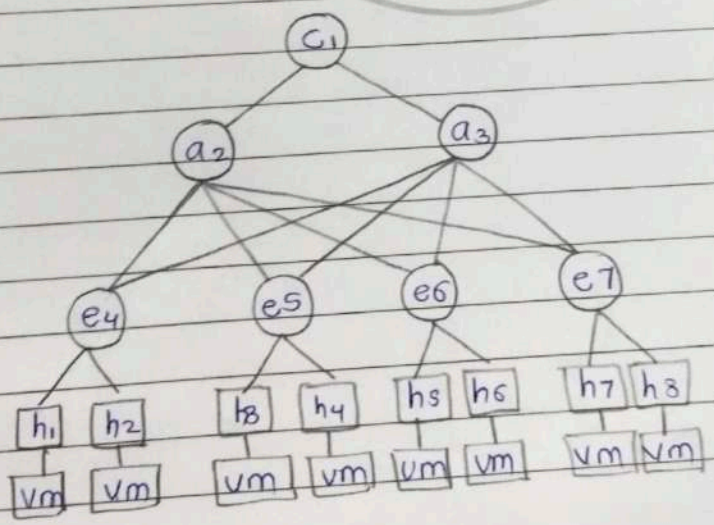
Data center:

- A data center is a large facility used to house computer systems and related equipment, such as storage devices & networking equipment.
- It is designed to provide a secure and controlled environment for these systems to operate in, as they are critical to the functioning of many businesses & organizations.
- Data centers are used to store, process and manage large amounts of data and applications, such as websites, mobile apps & databases.
- They are typically equipped with backup power supplies, cooling systems, and fire suppression system to ensure the equipment is protected and can operate continously.
- Data centers can range in size from small server rooms to large, multi-story facilities that can house thousands of servers.
- The location and design of a data center can impact its efficiency and environmental impact, so many data centers are designed to be energy-efficient and environmentaly friendly.

Data Center (clos) Topology :

- The clos topology is a network topology that is commonly used in data centers to provide high levels of scalability and redundancy.
- It is based on a multi-rooted tree-like structure, with multiple layers of switches or routers, each of which is interconnected in a specific way to provide maximum performance & reliability.
- At the core of clos topology are three layers of switches or routers known as spine layers, the leaf year & the host layer.
- The spine layer consists of a set of high-capacity switches or routers that are interconnected in a full-mesh topology, providing

high-speed interconnectivity between the leaf layer switches.
- The host layer consists of the individual devices themselves which are connected to the leaf switches. In this way, the host layer can scale up or down as needed, while the spine and leaf layers remain relatively static.
- one of the key advantages of the clos topology is th ?ts ability to provide high level of redundancy and fault tolerance. By using a full-mesh spine layer, traffic can be easily rerouted around any failed switches or links, ensuring that the network remains operational even in event of a failure.
- Another advantage of the clos topology is its ability to scale horizontally, allowing additionally leaf switches to be added as needed to support increasing number of devices.
- However, one potential disadvantage of the clos topology is its complexity. Because of the multiple layers and interconnectivity between switches, it can be callenging to manage and troubleshoot in large-scale deployments.



Conclusion :- Thus, we learned how to emulate and manage a data center via a cloud network controller using mininet & SDN application. We created a multi-rooted tree-like (clos) topology in mininet to emulate a data center with multiple hosts and switches. We implemented specific SDN applications on top of the network controller in order to orchestrate multiple network tenants within a data center environment.
In the context of network virtualization & management. we also experimented with diffent traffic pattern and routing algorithms to optimize the network performance..

Pune District Education Association's
College Of Engineering
Manjari (Bk.), Hadapsar, Pune-412307.
Accredited by NAAC

1.1

Assignment No. 1

Title :- Design and implement Parallel Breadth First search and Depth first search based on existing algorithms using OpenMP. Use a Tree or undirected graph for BFS and DFS.

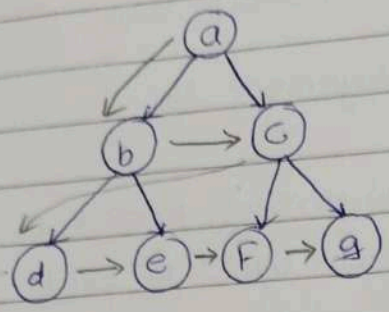
Objective :- Students should be able to perform parallel BFS & DFS based on existing algorithms using OpenMP.

Software Requirements :-
1. Operating system : Linux / windows / macOS
2. Compiler : GCC with OpenMP support or MinGW

Theory :-

What is BFS ?
- It is a Graph traversal algorithm used to all nodes of a graph or tree, starting from root node and visiting all the neighbouring nodes at the current depth level before moving to the next depth level.
- Algorithm uses queue data structure.
- BFS is commonly used in many applications, such as Finding the shortest path between two nodes, solving puzzles, and searching through tree or a graph..
- The basic idea of the BFS algorithm is to visit all the nodes at a given level before moving on the next level..

Diagram :-



Traversal Pattern :- a b c d e f g.

What is DFS ?

- It is a graph traversal algorithm that explores as far as possible among each branch before backtracking.. This algorithm can be used to find the shortest path between two vertices or to traverse a graph in systematic way.
- The algorithm starts at the root node and explores as far as possible along each branch before backtracking..

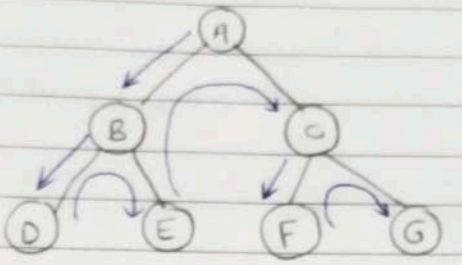
The backtracking is done to explore the next branch that has not been explored yet.

- DFS can be implemented using either a recursive or iterative approach. The iterative approach uses a STACK to keep track of nodes to be explored and is preferred for larger graphs..
- DFS can also be used to detect cycles in graph. If a cycle exist in a graph, the DFS algorithm will eventually reach a node that has already been visited, indicating that a cycle exists..
- A standard DFS algorithm puts each vertex in one of two categories :

1) Visited
2) Not visited..

1-2

example:



Traversal Pattern :- A B D E C F G.

OpenMP :-

- OpenMP (Open multi-Processing) is an application programming interface (API) that supports shared-memmory parallel computing in C, C++.
- It is used to write parallel programs that can run on multicore processors, multiprocessors systems, and parallel computing clusters.
- Openmp provides a set of derivatives and functions that can be inserted into the source code of a program to parallelize its execution. These directives are simple and easy to use and they can be applied to loops, sections, functions and other program constructs.
- The compiler then generates parallel code that can run on multiple processors concurrently.
- Openmp is widely used in scientific computing, engineering and other fields that require high-performance computing.

Parallel BFS using OpenMP

1. Divide the node exploration tasks among multiple threads.
2. Implement dynamic workload management using OpenMP constructs.
3. Use #pragma omp parallel to enable parallel execution.
4. Synchronize shared data using OpenMP loks and barriers.

Parallel DFS using OpenMP.

1. Implement parallelization by assigning independent branches to multiple threads.
2. Use openmp tasks to create new tasks for each recursive call.
3. Manage shared resources using mutex locks..

Conclusion :- This assignment provides hands-on experience in implementing and analyzing parallel graph traversal algorithms.

Assignment No. 2.

Title :- Write a program to implement Parallel Bubble sort and Merge sort using OpenMP. Use existing algorithms and measure the performance of sequential and parallel algorithms.

Objective :- Student should be able to write a program to implement Parallel Bubble sort and merge sort and can measure the performance of sequential & parallel algorithms.

Software Requirements :-
1. Operating system : Linux / window / macOS.
2. Compiler : GCC with openMP support or mingw.

Theory :-

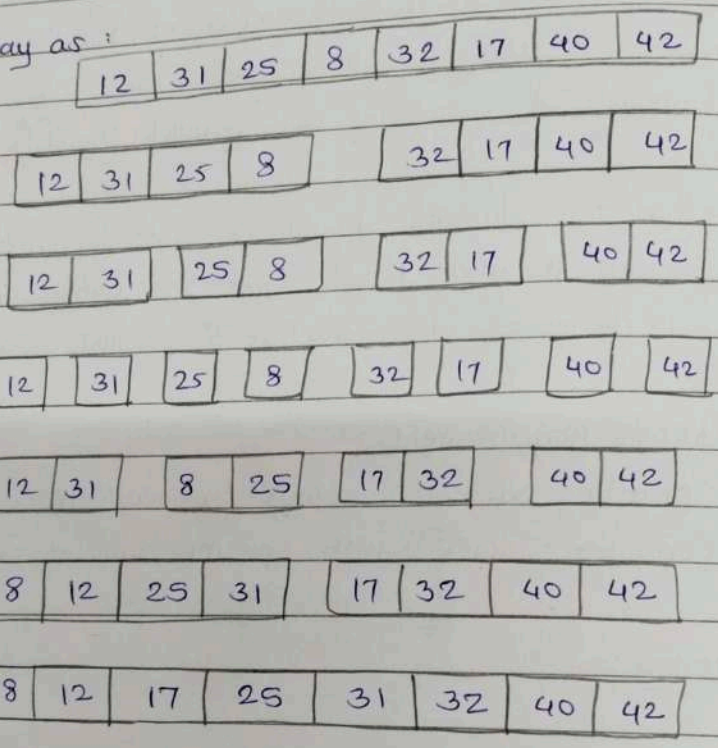
Merge Sort :-
- It is a sorting algorithm that uses a divide-and-conquer approach to sort an array or a list of elements. The algorithm works by recursively dividing the input array into two halves, sorting each half and then merging sorted halves to produce a sorted output.

The merge sort algorithm :

1. Divide input array into two halves.
2. Recursively sort the left half of the array.

3. Recursively sort the right half of the array..
4. Merge the two sorted halved into a single sorted output array..

eg.. let array as :



How to measure the performance of sequential & parallel algorithm?

1. Execution Time :- amount of time it takes for the algorithm to complete its sorting operation..
2. Speedup :- Ratio of the execution time of sequential merge sort algorithm to the execution time of parallel merge sort algorithm.. A speedup of greater than 1 indicates that parallel algorithm is faster than sequential..

3. Efficiency :- Ratio of speedup to the no. of processors or cores used in the parallel algorithm..
4. Scalability :- The ability of algorithm to maintain its performance as the input size and number of processors or cores increase.

What is Bubble Sort?

- A simple sorting algorithm that works by repeatedly swapping adjacent elements if they are in wrong order..
- It is called "bubble" sort because the algorithm moves the larger elements towards the end of the array in a manner that resembles the rising of bubbles in a liquid..

Algorithm :-

1. Start at the beginning of the array.
2. Compare the first two elements. If the first element is greater than 2nd swap them.
3. move to the next pair of elements and repeat step 2.
4. Continoue the process until the end of the array is reached..
5. If any swap were made in step 2-4, repeat the process from step 1.

The time complexity is $O(n^2)$

- For every small database, Bubble sort can be an efficient sorting algorithm, as its overhead is relatively low..

rs summed in parallel.
+8] ⇒ [3, 7, 11, 15]

How to measure the performance of sequential & parallel algorithm?

1. Implement both the sequential & parallel Bubble sort algorithms.
2. Choose a range of test cases, such as arrays of diffent sizes and diffrent degrees of sortedness to test the performance of both algorithms.
3. Use a reliable timer to measure the execution time of each algorithm on each test case
4. Record the execution time & analyze the results..

Conclusion :- In this way we can implement Parallel Bubble sort and merge sort using OpenMP also come to know how to measure performance of serial and parallel algorithm..

HPC

Pune District Education Association's
College Of Engineering
Manjari (Bk.), Hadapsar, Pune-412307.
Accredited by NAAC

Assignment No. 3

Title :- Implement Min, Max, Sum and Average operations using Parallel Reduction.

Objective :- To understand the concept of parallel reduction and how it can be used to perform basic mathematical operations on given data sets.

Software Requirements :-
1. Operating System : Linux / Windows / MacOS
2. Compiler : GCC with openMP support or MinGW

Theory :-

Parallel Reduction.
- It is a technique used in a parallel computing to efficiently compute a single result (such as a sum, maximum or product) from a large datasets using multiple processors or threads.
- It involves reducing the dataset step-by-step using a tree-like hierarchical structure, where partial results are combined in parallel until the final result is obtained.

eg. let array of numbers = [1, 2, 3, 4, 5, 6, 7, 8]
let compute the sum.
step1 : each pair of ~~sum~~ numbers summed in parallel.
ie. [1+2, 3+4, 5+6, 7+8] ⇒ [3, 7, 11, 15]

step 2: The partial results are summed again:
[3+7, 11+15] ⇒ [10, 26]
step 3: Final sum
ie. 10+26 = 36.

1. Min_Reduction Function.
- Function takes a vector of integers as input and finds the minimum value in the vector using parallel reduction.
- The minimum value found by each thread is reduced to the overall minimum value of the entire array.

2. Max_Reduction Function.
- The function takes in a vector of integers as input and finds the maximum value in the vector using parallel reduction.
- The maximum value found by each thread is reduced to the overall maximum value of the entire array.

3. Sum_Reduction Function.
- The function takes in a vector of integers as input and finds the sum of all the values in the vector using parallel reduction.
- The sum found by each thread is reduced to the overall sum of the entire array..

4. Average_Reduction Function.
- The function takes in a vector of integers as input and finds the average of all the values in the vector using parallel reduction.
- The final sum is divided by the size of the array to find the average.

5. main-function.

- The function initializes a vector of integers with some values.
- The final minimum, maximum, sum & average values are printed to the console.

Conclusion :- We have implemented the min, max, sum and Average operations using parallel reduction in C++ with OpenMP. Parallel reduction is the powerful technique that allows us to perform these operations on large arrays more efficiently by dividing the work among multiple threads running in parallel.

Assignment No. 04

Title:- Write a CUDA program for.
1. Addition of two large vectors
2. Matrix multiplication using CUDA C.

Objective:- Students should be able to perform CUDA program for addition of two large vectors & matrix multiplication using CUDA C.

Software Requirements:-
Operating System:- Linux / windows / macOS
Compiler :- GCC with openmp support or minGW.

Theory :-

What is CUDA?

- CUDA (compute Unified Device Architecture) is a parallel computing platform and programming model developed by NVIDIA. It allows developers to use the power of NVIDIA graphics processing units (GPUs) to accelerate computation tasks in various applications, including scientific computing, machine learning and computer vision.
- CUDA provides a set of programming APIs, libraries and tools that enable developers to write and execute parallel code on NVIDIA GPUs.

- It supports popular programming languages like C, C++ & Python and provides a simple programming model that abstracts ~~every~~ away much of the low-level details of GPU architecture.

- Using CUDA, developers can exploit the massive parallelism and high computational power of GPUs to accelerate computationally intensive tasks, such as matrix operations, image processing, and deep learning.
CUDA has become an important tool for scientific search and is widely used in fields like physics, chemistry, biology & engineering.

Steps for addition of two large vectors using CUDA.

1. Define the size of the vectors
2. Allocate memory on the host
3. Initialize the vectors
4. Allocate memory on the device
5. Copy the input vectors from host to device
6. Launch the kernel
7. Copy the result vector from device to host
8. Free memory on the device
9. Free memory on the host.

Steps for matrix multiplication using CUDA.

1. Matrix Initialization
2. memory Allocation
3. Data transfer
4. Kernel launch
5. Device Synchronization
6. Data Retrieval.
7. memory Deallocation.

Conclusion :- This assignment provides hands-on-experience on implementing CUDA program for addition of two large vectors and matrix multiplication using CUDA C.

Pune District Education Association's
College Of Engineering
Manjari (Bk.), Hadapsar, Pune-412307.
Accredited by NAAC

Assignment No. 2A.

Title :- Multiclass classification using Deep Neural Network
Example: Use the OCR letter recognition dataset.

Objective :- Students should be able to solve multiclass classification using deep neural network solve.

Theory :-

What is Multiclass classification?

- multiclass classification, also known as multiclass classification or multiclass classification problem, it is a type of classification problem where the goal is to assign input data to one of three or more classes or categories. In other words instead of binary classification, where the goal is to assign input data to one of two classes, multiclass classification involves assigning input data to one of several possible classes or categories.

- some examples of multiclass classification problems include image classification, where the goal is to classify images into one of several categories, and next classification, where the goal is to classify text documents into one of several categories (eg. news topics, sentiment analysis).

Examples of multiclass classification.

- Here are a few examples of multiclass classification problems.

Image classification:- The goal is to classify images into one of several categories. For example an image classification model might be trained to classify images of animals to categories such as cats, dogs and birds.

Text classification:- The goal is to classify text documents into one of several categories. For example, a text classification model might be trained to classify news articles into categories such as politics, sports & entertainment.

Disease Diagnosis: The goal is to diagonse patient with one of several disease based on their symptoms and medical history. For example, a disease diagnosis model might be trained to classify patients into categories such as diabetes, cancer & heart disease.

Speech Recognition: The goal is to transcribe spoken words into text. A speech recognition model might be trained to recognize spoken words in several languages or dialects.

Credit risk analysis: The goal is to classify loan applicants into categories such as low risk, medium risk and high risk. A credit risk analysis model might be trained to classify loan applicants based on their credit score, income & other factors..

In all of these examples, the goal is to assign input data to one of several possible classes or categories. multiclass classification is a common task in ML & can be approached using a variety of algorithms, including decision trees, support vector machines & deep neural networks.

Conclusion :- In this way we can do multi classification using ONN.

Assignment No. 2B

Title:- Binary classification using Deep Neural Networks Example: Classify movie reviews into positive "reviews" and "negative" reviews, just based on the text content of the reviews. Use IMDB dataset.

Objectives:- Students should be able to Classify moview reveiws into positive reveiws and "negative reviews on IMDB dataset.

Theory:-

What is classification?

- It is a type of supervised learning in ML that involves categorizing data into predefined classes or categories based on a set of features or characteristics. It is used to predict the class of new, unseen data based on the patterns learned from the labeled training data.
- In classification, a model is trained on a labled dataset where each data point has a known class label. The model learns to associate the input features with the corresponding class labels and can then be used to classify new, unseen data.

- For example, we can use classification to identify whether an email is spam or not based on its content and metadata to predict whether a patient has a disease based on their

medical records and symptoms, or to classify images into different categories based on their visual features.

- Classification algorithms can vary in complexity, ranging from simple models such as decision trees and k-nearest neighbors to more complex models such as support vector machines & neural networks.
The choice of algorithm depends on the nature of the data, the size of the dataset, and the desired level of accuracy & interpretability.

- Classification is a common task in deep neural networks, where the goal is to predict the class of an input based on its features. Here's an example of how classification can be performed in a deep neural network using the popular MNIST dataset of handwritten digits.

- The MNIST dataset contains 60000 training images & 10000 testing images of handwritten digits from 0 to 9. each image is a grayscale 28 x 28 pixel image, and the task is to classify each image into one of the 10 classes corresponding to the 10 digits.

- We can use a convolutional neural network (CNN) to classify the MNIST dataset. A CNN is a type of deep neural network that is commonly used for image classification tasks.

How deep Neural Network work on classification.

- Deep neural networks are commonly used for classification tasks because they can automatically learn to extract relevant features from raw input data and map them to the correct output class.

The basic architecture of a deep neural network for classification consists of three main parts. an input layer one or more hidden layers, and an output layer. The input layer receives the raw input data, which is usually preprocessed to a fixed size and format.

IMDB Dataset - IMDB dataset is a large collection of movie reviews collected from the IMDB website. which is a popular source of user generated movie ratings & reviews. The dataset consists of 50000 movie reviews, split into 25000 reviews for training and 25000 reviews for testing.

Each review is represented as a sequence of words, where each word is represented by an integer index based on its frequency in the dataset. The labels for each review are binary. with 0 indicating a negative review and 1 indicating a positive review.

Conclusion :- In this way we can Classify the movie Reveiws by using DNN.

DL

Pune District Education Association's

College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

3-1

Assignment No. 3.

Title :- Use MNIST Fashion dataset and create a classifier to classify fashion clothing into categories.

Objective : Students should be able to use MNIST fashion dataset and create a classifier to classify fashion clothing into categories.

Prerequisite :- 1) Basic of programming language.
2) Concept of classification
3) Concept of Deep Neural Network.

Theory :-

What is classification ?

- Classification is a type of supervised learning in ML that involves categories data into predefined classes or categories based on a set of features or characteristics.
- In classification a model is trained on a labelled dataset where each data point has a known class table. The model learns a associate the input features with the corresponding class tables & can then be used to classify new unseen data.
- For example, we can use classification to identify whether an email is spam or not not based on its content & meta data to predict whether a patient has a diesease based on their medical records & symptoms

or to classify images into diffent categories based on their virtual feature.

The MNIST dataset contain 60000 training images 10000 testing images of hand written digits from a tag Each image is a gray scale 28×28 pixel image, & the task is to classify each image into one of the 10 classes corresponding to 10 digits.

What is CNN ?

- Convolutional neural network are commonly used for image classification tasks & they are designed to automatically learn & extract features from input images. let's consider an example of using a CNN to classify image of handwritten digits.
- Each filter produces a feature map that highlights area of the image that match the filter. The filter are learned during training so the network can automatically learn which features are not most relevant for the classification task.
- The fully connected layers in the CNN take the flatened output from the last pooling layers & perform a classification task by outputting a probalbility distribution over the 10 possible digits.

CNN, have a wide range of applications in various fields, some of which are:

1) Image classification : CNN are commonly used for image classification tasks such as identifying objects in images & recognizing faces.

Object detection : CNN can be used for object detection in images & videos which involves partitioning an image into segments & assigning each segment a semantic label.

Natural language Processing : CNN can be used for natural language tasks and such as sentiment analysis & text classification.

Medical Imaging : CNN are used in medical imaging for tasks such as diagnosing diseases from x-rays & identifying tumors from MRI scans.

Autonomous Vehicles :- CNN are used in autonomous vehicles for tasks such as object detections & lane detection.

How deep neural network work on classification using CNN :
- Deep NN using CNNs work on classification tasks by learning to automatically extract feature from input images & using those features to make predictions
Here's how it works :
Input layer : Input layer of the network tasks in the image data as input.

Convolutional layers :- The convolutional layers apply filters to the input layers images to extract relevant features.

Activation function: Activation function is applied to the input of each convolutional layer to introduced non-linearity into the network.

Pooling layers: The pooling layers down sample the feature maps to reduce the spatial dimensions of the data.

Dropout layer: Dropout is used to prevent overfitting by randomly dropping out a percentage of the neurons in the network during training.

Fully connected layer: The fully connected layers takes the flatned output from the last pooling layer & perform classification task over the possible classes.

optimization: An optimization algorithm, such as stochastic gradient descent is used to minimize the loss function by adjusting the values of the network parameter.

Training: The network is trained on a large dataset of labeled images, adjusting the values of the parameters to minimize the loss function.

Conclusion:- In this way we can classify fashion clothing into categories using CNN.

Mini Project

Title:- Implement Human Face Recognition.

Objective:- To build a deep neural network model that can recognize the ~~fuma~~ human faces.

Theory :-

- Human face recognition using deep neural networks involves building a neural network model that can take an image of a human face as input and accurately recognize the person in the image.
- Human face recognition using deep neural networks (CNNs) involves training a neural network to identify and distinguish between diffrent faces. The process typically involves the following steps :

1) Data collection :- A large dataset of face images is collected, including images of diffrent individuals and under diffrent lighting and pose conditions.

2) Data Preprocessing :- The face images are preprocessed to remove noise, align the faces, and normalize the illumination.

3) Feature Extraction :- The preprocessed face images are then fed into a deep neural network to extract high-level features that captures the important characteristics of a face.
- The neural network typically consists of several layers of convolutional and pooling operations, followed by fully connected layers that produce a feature vector.

4) Training :- The extracted features are then used to train the neural network to distinguish between diffrent faces. This is typically done using a supervised learning approach, where the network is trained on a labeled dataset of face images and their corresponding identities.

5) Testing :- After the neural network has been trained, it can be tested on a seperate dataset to evaluate its performance. This typically involves measuring the accuracy of the network in correctly identifying the individuals in the test dataset.

6) Deployment :- Once the neural network has been trained and tested, it can be deployed in a real-world application for face recognition. This typically involves capturing a face image, preprocessing it, and then feeding it into the neural network to obtain a feature vector.

Overall, human face recognition using DNNs is a complex process that requires a large amount of data sophisticated neural architecture, and careful preprocessing and training. However, with the increasing availability of large datasets

and powerful computing resources, DNN - based face recognition system have become increasingly accurate and effective in real world applications.

Conclusion :- In this way Human Face Recognition Implemented..