

MODI SHRESHTHA(190130111081)

Data Anonymization pipeline

Eternal Soft Solutions ✦ GEC Gandhinagar ✦ 2023
(EC)

About

Eternal is a AWS partner web development based consulting company based in Ahmedabad and UK that provides cloud based product design and development solutions for businesses and startups



401 Satyam Mall Nr Mansi Circle, Satellite, Ahmedabad,
Gujarat 380051

Training



Training

05

Week 5

aws



07

Week 7

Linux™



06

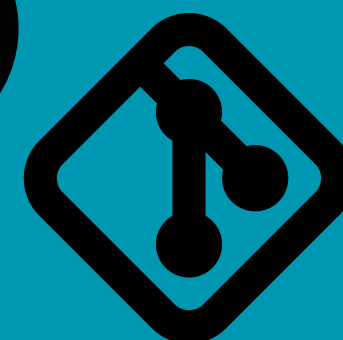
Week 6

Linux™



08

Week 8



Week 1

01

Installing python on Linux, virtual environments in python and the code skeleton of python, introduction to vscode intellisense ide and shortcuts for vscode

03

basic and nested loops, switch statements, basic calculator and a mad libs game

02

Data types, variables, typecasting

04

String, list, dictionary and array methods such as slicing, indexing and appending

05

Functions and modules, global and local scope, basics of understanding the errors and debugging them in python

Shreshtha Modi 190130111081

Week 2

01

Classes and inheritance and understanding of scope

03

Global and local variables and iterators

02

Modules and various types of modules, introduction to the pandas and numpy modules

04

Pandas (importing the csv data, understanding the csv data, getting the columns and other methods), working with csv data

05

FAPI and apis using python, pip and python file handling

Shreshtha Modi 190130111081

Week 3

01

Introduction to web scraping using beautiful soup using selenium

03

Python refresher + basics of computer networking components used on aws

02

What is cloud computing, advantages of cloud as compared to on prem and major cloud providers, models of cloud computing

04

Introduction to aws and the aws free tier, aws core values and principles

05

AWS networking components, basic services of aws storage and networking

Shreshtha Modi 190130111081

Week 4

01

AWS computing services

03

AWS S3, EC2 and Lambda and their various tiers

02

AWS billing and pricing (budgets, cost explorer, cloudwatch and cloud logs) and understanding the budget analytics

04

AWS shared responsibility model, importance of data privacy and protecting your data on the cloud

05

Components of data privacy and data anonymization techniques and aws solutions for data privacy

Shreshtha Modi 190130111081

Week 5

01

Project brief and aws cdn services and reliability

03

Misc aws services such as sns, inspector and macie for threat detection, Introduction to docker

02

aws database services, aws services for big data and machine learning

04

Docker components, architecture and docker basics

05

Frequently used docker commands and building a local docker instance from an image

Shreshtha Modi 190130111081

Week 6

01

Introduction to Operating Systems and kernels

03

Installing ubuntu 22.04 on a local machine and virtual machine and explanation about various components of the linux os

02

Introduction to linux and it's various distributions

04

working with files, directories and linux file structures

05

shell training to learn more about installing and updating the packages, debugging dependency errors and using regex in shell to automate things, package manager

Shreshtha Modi 190130111081

Week 7

01

Managing system units and logs, and managing users

03

open ssh and numerical permissions

02

bash history, output redirection, variables

04

scp and rsync

05

Introduction to version control, history of version control and basics of git

Shreshtha Modi 190130111081

Week 8

01

Installing github and local and online repository, git file structure and git pull, push and add

03

ssh git and other basics of github and project brief introduction

02

git rebash, git commit and git clone

04

Installing arx and running the arx instance locally

05

arx documentation understanding and sample anonymization

Shreshtha Modi 190130111081

Project Abstract

Internet has an approximate of 4.6 billion users worldwide and they are growing at an astronomical rate. With the rise in users, privacy concerns also pop up. Data privacy is a complex area which happens to be a vital part of the internet ecosystem. This project aims at protecting the privacy of the user data by anonymizing the data using various statistical tools and techniques and proposing an end to end solution of the anonymized data on the AWS. The proposed solution's schema can be modified according to the need of the user and the type of the data with ease

Shreshtha Modi 190130111081

Thank You So Much!

Eternal soft solutions ✦ GEC Gandhinagar(EC) ✦ 2023