

Assignment: Creating ETLs using AWS and Terraform

Objective

The objective of this assignment is to test your ability to create a simple ETL using AWS and Terraform. You will be required to create a data pipeline that extracts data from a source, transforms it, and loads it into a destination using AWS services and Terraform.

Assignment

Your task is to create a data pipeline that extracts data from a source, transforms it, and loads it into a destination using AWS and Terraform.

Phase 1: Data Extraction

1. Identify a publicly available dataset that you would like to use for this assignment. The dataset should be in a format that can be easily extracted, such as CSV or JSON.
2. Use AWS S3 to store the dataset in an S3 bucket. You should write a Terraform script to create the S3 bucket and configure it with the appropriate permissions.

Phase 2: Data Transformation

1. Use AWS Glue to transform the data into a format that is suitable for loading into a destination. You should write a Terraform script to create the Glue job and configure it with the appropriate permissions.
2. Perform any necessary data validation, error handling, and logging in your Glue job.

Submission

Please submit the following materials:

1. Your Terraform scripts for creating the S3 bucket, Glue job, and destination.
2. A README file with instructions for deploying and testing your data pipeline.

Good luck with the assignment! If you have any questions, please do not hesitate to reach out.