AWS ETL Project Breakdown

- Integrated with Amazon Athena for Real-Time Data Querying

Objective:

Analyze streaming data stored in S3 using serverless SQL with Amazon Athena.

Steps Taken:

- Set up Athena, selected S3 output bucket
- Created 'etl_sales' database
- Defined tables referencing raw CSV data in S3
- Queried using SQL to analyze and summarize transaction data

Real-World Insight:

This is how businesses do ad-hoc analytics using a data lake in S3.

- Tried Connecting Tableau for Visualization

Objective:

Visualize real-time sales data from Athena using Tableau Desktop.

Steps Taken:

- Installed Athena JDBC Driver and added to Tableau directory
- Attempted to connect Tableau to Athena with access credentials
- Faced configuration errors; resolved IAM and S3 access policies

Next Step:

Tableau can now visualize Athena datasets for live dashboards.

- Navigated IAM, Access Keys, and Permissions

Objective:

Manage AWS permissions and secure data access.

AWS ETL Project Breakdown

Steps Taken:

- Created IAM user (ETL_hand), configured access key/secret
- Set up boto3 and AWS CLI credentials
- Debugged IAM errors (e.g., PutUserPolicy, Quicksight access)
- Wrote JSON IAM policies to grant limited secure access

Real-World Insight:

IAM mastery is essential for roles in DevOps, Cloud Data Engineering.

- Summary Table

Module	Tools Used	Skill Developed
ETL Pipeline	Python, Faker, Boto3, S3	Data generation, scripting, upload automation
Athena SQL Analytics	AWS Athena, SQL	Serverless querying, table creation
Visualization	Tableau, Athena JDBC	BI integration, live dashboard prep
Access Management	AWS IAM, Policies	Secure access setup, key management