**Integrating AWS, Snowflake, and Tableau for Data-Driven Insights**

*By Bhavishya Koluguri*

**Overview**

This project demonstrates the integration of Amazon Web Services (AWS), Snowflake, and Tableau for data storage, transformation, and visualization. The objective was to analyze renewable energy usage and cost savings across various regions and energy sources. The process involved setting up a data pipeline from AWS S3 to Snowflake and visualizing the data in Tableau to provide insights.

Integrating AWS, Snowflake, and Tableau makes it easier to store, process, and analyze data. AWS provides a secure place to store data, Snowflake helps organize and prepare it, and Tableau turns it into clear, interactive dashboards.

When implemented in industries, this integration facilitates better resource management, drives cost optimization, and enables data-driven strategies across departments. Whether monitoring energy usage, forecasting demand, or analyzing customer behavior, this pipeline offers a robust foundation for scalable analytics solutions.

**Process WorkFlow**

**Step 1: Creating and Loading Data to S3 Bucket**

Created an S3 bucket named tableauproject in AWS to store renewable energy data. Uploaded the file Renewable\_Energy\_Usage\_Sampled.csv to the bucket. The data was in CSV format to ensure compatibility with subsequent processing steps. The data was successfully uploaded to S3 and prepared for integration with Snowflake.

A screenshot of a computer

Description automatically generated

**Step 2: Creating IAM Roles for Snowflake and Tableau**

Two IAM roles were created in AWS: Snowflake-Test-Role and Tableau-Role. These roles were configured to securely enable data access for Snowflake and Tableau. The AmazonS3FullAccess policy was attached to both roles, granting them the necessary permissions. Trust relationships were configured to allow Snowflake and Tableau to assume these roles securely. This setup ensured seamless and secure access to the S3 bucket.

A screenshot of a computer

Description automatically generated

**Step 3: Creating the Integration Object in Snowflake**

A storage integration object was created in Snowflake to link it with the S3 bucket. Using Snowflake SQL, the integration was established by executing the following command shown in the snip:

The integration object was verified using the command desc integration tableau\_Integration. This step successfully established a connection between Snowflake and the S3 bucket.

A screenshot of a computer

Description automatically generated

**Step 4: Loading Data into Snowflake**

The data was loaded into Snowflake from the S3 bucket. A database and schema were created in Snowflake:

CREATE DATABASE tableau;

CREATE SCHEMA tableau\_Data;

A table was defined in Snowflake to store the dataset:

CREATE TABLE tableau\_dataset (

Household\_ID STRING,

Region STRING,

Country STRING,

Energy\_Source STRING,

Monthly\_Usage\_kWh FLOAT,

Year INT,

Household\_Size INT,

Income\_Level STRING,

Urban\_Rural STRING,

Adoption\_Year INT,

Subsidy\_Received STRING,

Cost\_Savings\_USD FLOAT

);

A stage was created in Snowflake to link the S3 bucket:

CREATE STAGE tableau\_Data.tableau\_stage

URL = 's3://tableauproject'

STORAGE\_INTEGRATION = tableau\_Integration;

The data was then copied into the Snowflake table using the following command:

COPY INTO tableau\_dataset

FROM @tableau\_stage

FILE\_FORMAT = (TYPE=CSV FIELD\_DELIMITER=',' SKIP\_HEADER=1)

ON\_ERROR = 'CONTINUE';

This step successfully loaded the data into Snowflake, making it ready for analysis.

A computer screen shot of a black screen

Description automatically generated

**Step 5: Data Understanding and Transformation in Snowflake**

Data understanding and transformation were performed using Snowflake SQL. Queries were executed to analyze data distribution and completeness. For example, the count of records per region was determined using the following query:

A screen shot of a computer program

Description automatically generated

These transformations prepared the data for visualization in Tableau.

**Step 6: Connecting Snowflake to Tableau**

The Snowflake database was connected to Tableau to visualize the transformed data. A live connection was established using the database tableau, schema tableau\_data, and table energy\_consumption. The connection was verified by previewing the data in Tableau.

A screenshot of a computer

Description automatically generated

**Step 7: Creating Visualizations in Tableau**

Various visualizations were created in Tableau to analyze monthly energy consumption and cost savings. Charts were built to display:

* Monthly Usage (kWh) by Region.
* Monthly Usage (kWh) by Country.
* Monthly Usage (kWh) by Energy Source.
* Cost Savings (USD) by Region.
* Cost Savings (USD) by Country.
* Cost Savings (USD) by Energy Source.

Tableau’s drag-and-drop interface was used to design dynamic and interactive charts, providing insights into energy consumption trends.

**Step 8: Creating the Dashboard**

A dashboard was created in Tableau to summarize the insights in a single view. Multiple charts were combined into an interactive dashboard with filters for Region, Country, and Energy Source. This enabled users to perform specific analyses based on their requirements. The dashboard was published for stakeholders, offering a comprehensive visualization of energy consumption and cost-saving trends.

A screenshot of a graph

Description automatically generated

**Insights from the Dashboard**

1. Monthly Usage (kWh) by Country:
   * Australia, New Zealand, and Canada have the highest usage, indicating high demand or population density.
   * Chile, Germany, and Nigeria show lower energy usage.
2. Cost Savings (USD) by Country:
   * New Zealand, Australia, and Canada lead in savings, reflecting efficiency or effective energy-saving measures.
   * Chile and Germany have lower savings, suggesting higher energy costs or less efficiency.
3. Monthly Usage (kWh) by Energy Source:
   * Wind and Hydro dominate, followed by Solar and Biomass, showing their availability or preference.
   * Geothermal energy has the lowest usage, likely due to adoption or geographical limits.
4. Cost Savings (USD) by Energy Source:
   * Solar and Biomass offer the highest savings, while Geothermal shows minimal cost benefits.
5. Monthly Usage (kWh) by Region:
   * Europe, Africa, and South America consume the most energy, reflecting greater access or demand.
   * Australia and North America show comparatively lower usage.
6. Cost Savings (USD) by Region:
   * South America and Africa excel in cost savings, hinting at better energy policies or subsidies.
   * North America and Australia lag in savings, suggesting room for efficiency improvements.

This dashboard provides actionable insights into energy consumption and savings patterns across different dimensions. Stakeholders can use these insights to focus on regions or energy sources with the potential for improvement, implement targeted policies, and encourage the adoption of cost-effective renewable energy solutions.

**Key Deliverables**

Key deliverables of the project included the S3 bucket for data storage, IAM roles for secure access, Snowflake integration with S3 for data loading, transformed data in Snowflake, and an interactive Tableau dashboard for decision-making.

**Conclusion**

This project successfully demonstrated the integration of AWS, Snowflake, and Tableau for renewable energy data analysis. The resulting dashboard enables stakeholders to understand energy usage and cost-saving trends interactively, providing actionable insights.