Final Project

Soham Chakraborty EmployeeID: 2276590

**Project Topic :-**

Real time energy consumption analysis using Kinesis and Redshift

**Requirements :-**

**Dataset** - generated using python script inside which we have installed and imported boto3 which is an AWS SDK for python

**Technology used** -

* Python (Boto3)
* Kinesis Data Stream
* Redshift
* IAM
* Lambda
* SNS
* DynamoDB

**Introduction and Description:-**

**Real-time energy consumption analysis is crucial for several reasons:**

* **Efficiency Optimization:** By analyzing energy consumption in real-time, organizations can identify inefficiencies and areas where energy is being wasted. This allows them to take immediate corrective actions to optimize energy usage, leading to cost savings and reduced environmental impact.
* **Demand Response:** Real-time analysis enables organizations to participate in demand response programs where they can adjust their energy consumption in response to signals from the grid operator. This can help balance supply and demand on the grid, prevent blackouts, and potentially earn revenue through incentives.
* **Predictive Maintenance:** Real-time analysis can reveal patterns in energy consumption that may indicate equipment malfunction or degradation. By detecting issues early, organizations can schedule maintenance proactively, preventing costly breakdowns and downtime.
* **Peak Load Management:** Understanding real-time energy consumption helps organizations manage peak loads more effectively. By shifting or reducing energy usage during peak periods, they can avoid high demand charges or penalties imposed by utility companies.
* **Resource Planning:** Real-time data on energy consumption helps organizations make informed decisions about resource allocation and future investments in energy-efficient technologies. It provides insights into which areas require upgrades or modifications to improve energy efficiency.
* **Environmental Impact:** Monitoring energy consumption in real-time allows organizations to assess their environmental footprint accurately. By reducing energy usage, they can lower greenhouse gas emissions and contribute to environmental sustainability efforts.

**Python & Boto3 -**

Boto3 is the Amazon Web Services (AWS) SDK for Python, allowing developers to interact with AWS services using Python code. It provides a simple and intuitive interface to access a wide range of AWS services programmatically, such as Amazon S3 (Simple Storage Service), EC2 (Elastic Compute Cloud), DynamoDB (NoSQL database service), Lambda (serverless computing service), and many others.

With Boto3, developers can create, manage, and automate AWS resources and services directly from Python scripts or applications. It simplifies tasks such as provisioning infrastructure, managing data storage, triggering serverless functions, and orchestrating cloud-based workflows. Boto3 abstracts away the complexities of interacting with AWS APIs, making it easier for developers to integrate AWS services into their Python applications.

**Kinesis Data Stream -**

Amazon Kinesis Data Streams is a fully managed service provided by AWS for ingesting, processing, and analyzing real-time streaming data at scale. It enables users to collect and process large volumes of data records in real-time, supporting applications such as log and event data collection, IoT telemetry data ingestion, clickstream analysis, and real-time analytics. With automatic scaling and high durability, Kinesis Data Streams seamlessly integrates with other AWS services like AWS Lambda, Amazon Kinesis Data Analytics, and Amazon Kinesis Data Firehose, allowing users to build robust, real-time data processing pipelines and derive actionable insights from streaming data.

**Amazon Redshift -**

Amazon Redshift is a fully managed, petabyte-scale data warehousing service provided by Amazon Web Services (AWS). It allows users to efficiently analyze large datasets using SQL queries and business intelligence tools. Redshift is based on a columnar storage architecture and is optimized for high performance, enabling quick querying and processing of data. It offers features such as automatic backups, encryption, and integration with various data sources and analytics tools. With its scalability, cost-effectiveness, and ease of use, Amazon Redshift is widely used by organizations for data warehousing, analytics, and reporting purposes.

**Amazon Redshift Streaming Ingestion** is a feature that allows users to continuously load streaming data into Amazon Redshift tables in real-time. It leverages services like Amazon Kinesis Data Streams or AWS Lambda to capture and process streaming data, and then inserts this data directly into Redshift tables without the need for intermediate storage or transformation. This enables users to analyze and query streaming data alongside their existing structured data in Redshift, providing near real-time insights and analysis. Redshift Streaming Ingestion simplifies the process of ingesting and analyzing streaming data, making it easier for organizations to incorporate real-time data into their analytics workflows.

**Lambda -**

Lambda is a serverless computing service that allows you to run code without provisioning or managing servers. It enables you to execute code in response to events triggered by other AWS services or custom events. Lambda functions are event-driven, meaning they can be triggered by events such as changes in data in a DynamoDB table, incoming HTTP requests, or messages published to an SNS topic.

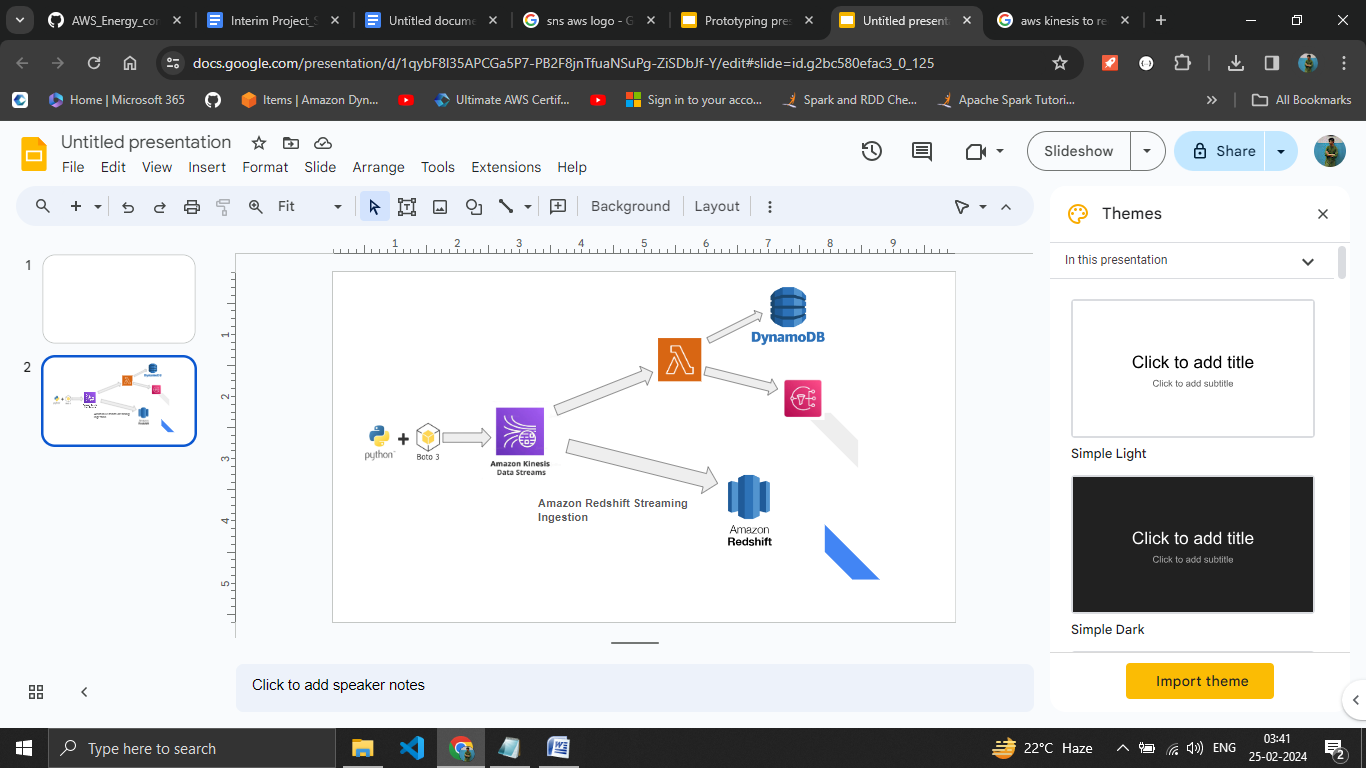
**DynamoDB -**

DynamoDB is a fully managed NoSQL database service provided by AWS. It offers seamless scalability, high availability, and low-latency performance for applications requiring fast and predictable performance at any scale. DynamoDB is commonly used for storing and retrieving structured data such as user profiles, session data, and metadata. It integrates well with AWS Lambda through services like DynamoDB Streams, which allows Lambda functions to be triggered in response to changes in DynamoDB tables.

**SNS -**

SNS is a fully managed pub/sub messaging service that enables you to send messages or notifications to a large number of subscribers or endpoints. It supports a variety of messaging protocols, including HTTP, HTTPS, email, SMS, and more.

**Steps of the project -**

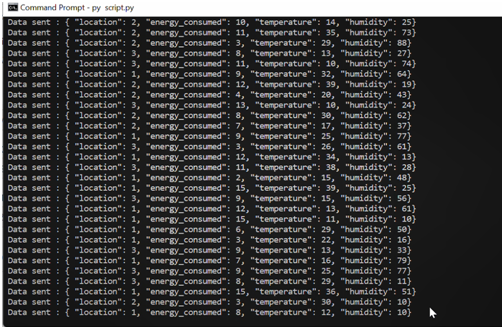


1. Creating script using boto3 and generating the data .

We are generating random data using a python script like location , temperature , energy consumed , humidity and using boto3.client to send the data to kinesis data stream in our aws console with the name energy\_data\_stream

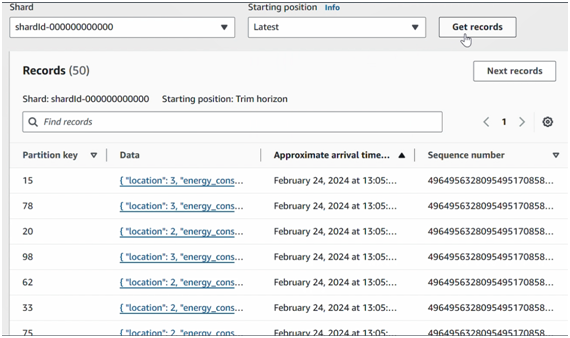


The data being generated looks like this



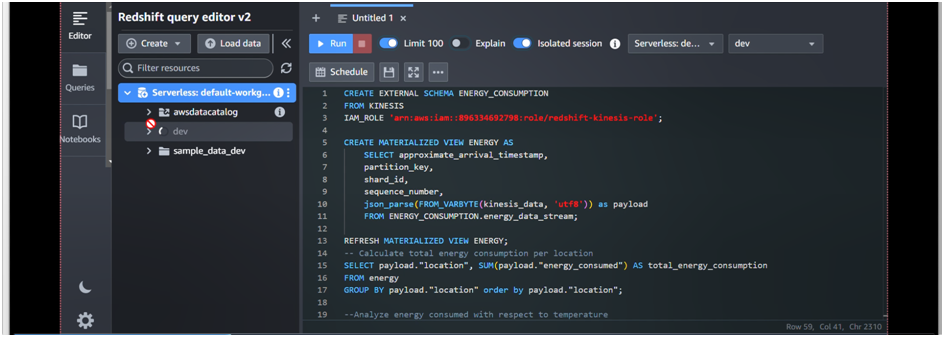
1. Data Ingestion using Kinesis Data Stream .

We are ingesting the data from our script into our energy\_data\_stream in kinesis in real time .



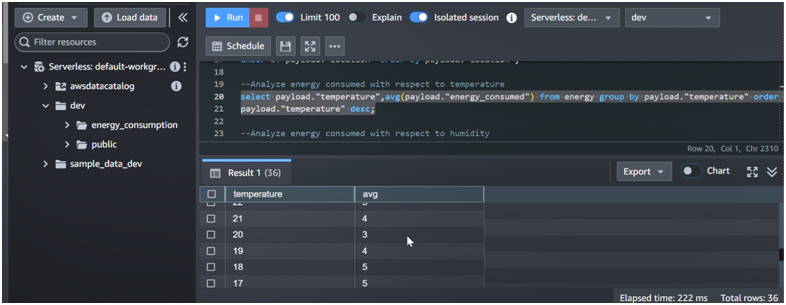
1. Sending the real time data from Kinesis Data Stream to Amazon Redshift using Amazon Redshift Streaming Ingestion.

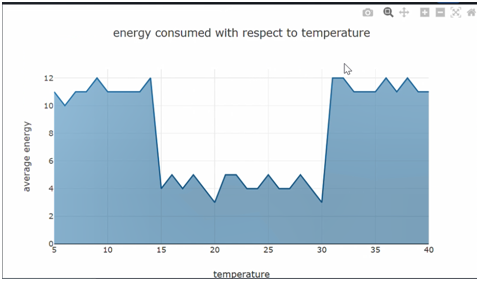
**Amazon Redshift Streaming Ingestion** is a feature that allows users to continuously load streaming data into Amazon Redshift tables in real-time. It leverages Amazon Kinesis Data Streams to capture and process streaming data, and then inserts this data directly into Redshift tables without the need for intermediate storage or transformation. This enables users to analyze and query streaming data alongside their existing structured data in Redshift, providing near real-time insights and analysis. Redshift Streaming Ingestion simplifies the process of ingesting and analyzing streaming data, making it easier for organizations to incorporate real-time data into their analytics workflows.



1. Performing data cleaning , data analyzing and visualization using Amazon redshift to gain data driven business insights from the data.

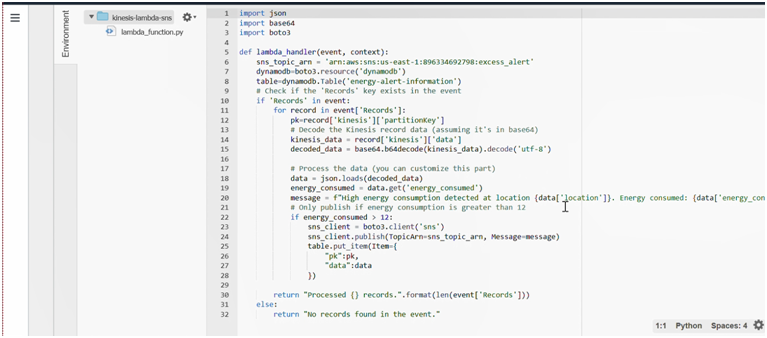
We are performing different data cleaning operations , analyzing the data by running different queries and visualizing the data to gain data driven insights from the data . For example screenshot is shown below





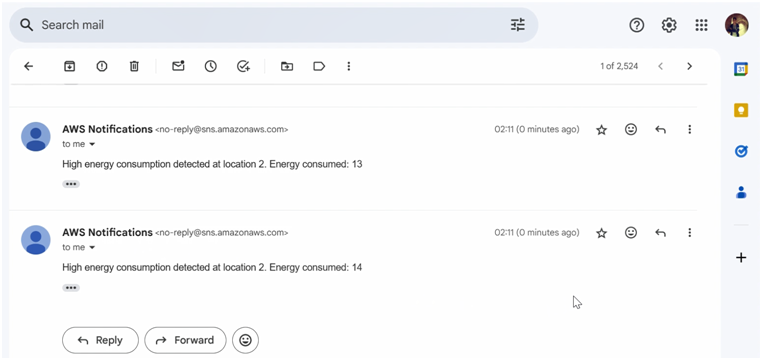
Like this we have performed 9 more queries to analyze the data to derive business insights from the energy data

1. Creating a lambda function with the Kinesis data stream as the trigger to make sns and dynamodb deal with alerts and reporting if data ingested is above a certain maximum value.

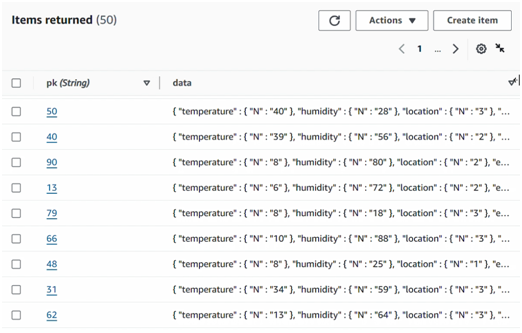


1. Creating SNS topic and subscription to send custom alerts to subscribed email id .

When we will receive data where energy consumed is more than 12 units then it will invoke the lambda function to send the alert message to the subscribed email id.



1. Creating a dynamo db table to collect and store the data that is above the certain maximum value of energy consumed i.e, 12 kWh as alert information for future analysis and making reports for saving energy and cost .



**Conclusion -**

In conclusion, my project successfully demonstrated the capabilities of Amazon Redshift and Amazon Kinesis in analyzing energy consumption data. Through the integration of these two powerful AWS services, we were able to efficiently ingest, process, and analyze large volumes of real-time streaming data related to energy consumption. By leveraging Amazon Kinesis Data Streams, we captured data from various sources and seamlessly fed it into Amazon Redshift for further analysis. With Amazon Redshift's scalable data warehousing capabilities and SQL-based querying, we were able to perform complex analytics on the data, uncovering valuable insights into energy consumption patterns, trends, and anomalies. This project not only showcased the effectiveness of using cloud-based solutions for energy data analytics but also highlighted the potential for leveraging AWS services to drive data-driven decision-making and optimize energy usage in various sectors. Moving forward, further enhancements could include implementing machine learning models for predictive analytics and integrating with additional AWS services to enhance the overall analytics pipeline