# Title: Real-time dashboards from a source database to a cloud data warehouse

By T.Gomahalakshmi - 2200080020



# Abstract

This project explores the architecture and implementation of near real-time dashboards by streaming data from an on-premises or transactional source database to a cloud-based data warehouse on Amazon Web Services (AWS). Leveraging AWS services such as AWS DMS (Database Migration Service), Amazon Kinesis, AWS Glue, and Amazon Redshift, the system ensures low-latency data ingestion, transformation, and availability for business intelligence tools like Amazon QuickSight or Tableau.

The solution addresses key challenges including data freshness, scalability, latency, and fault tolerance. This paper presents a robust, scalable pipeline that minimizes data lag while maximizing query performance and dashboard responsiveness. The architecture supports continuous data sync and enables stakeholders to monitor critical KPIs in near real-time, ultimately enhancing operational efficiency and strategicagility.

This setup empowers business users to access up-to-date KPIs, monitor system performance, and make data-driven decisions with confidence. Key considerations such as data consistency, fault tolerance, scalability, and cost-efficiency are addressed through cloud-native solutions and best practices. The resulting system delivers a robust, maintainable, and real-time analytics platform that enhances operational visibility and drives strategic.

# Modules

### 1. ****Data Ingestion / Streaming****

Modules used to capture and move data from the source database to the cloud in near real-time.

* **AWS DMS (Database Migration Service)**  
  For real-time replication (CDC - Change Data Capture) from source databases (e.g., MySQL, PostgreSQL, Oracle, etc.) to targets like Amazon S3, Redshift, or Kinesis.
* **Amazon Kinesis Data Streams / Firehose**  
  For streaming data ingestion. Firehose can directly load data into S3, Redshift, or Elasticsearch.
* **AWS Glue (Streaming ETL)**  
  Can also be used to process streaming data and convert formats (e.g., JSON to Parquet) on the fly.

### 2. ****Data Processing / Transformation****

To clean, transform, or enrich your data before storing.

* **AWS Glue (Batch or Streaming Jobs)**  
  Managed ETL service for transforming data.
* **Amazon Kinesis Data Analytics**  
  Run SQL queries on streaming data from Kinesis.
* **AWS Lambda**  
  Lightweight, serverless compute that can transform data in real time.
* **Apache Spark on Amazon EMR**  
  For more complex or large-scale transformations, if needed.

### 3. ****Data Storage****

Where your processed data lands, ready for analysis or dashboards.

* **Amazon Redshift**  
  Fully managed data warehouse—ideal for analytics and dashboards.
* **Amazon S3**  
  Often used as a staging or landing zone for raw or processed data.
* **Amazon RDS / Aurora**  
  If you need to keep data in relational format for operational dashboards.

### 4.****Visualization****

Used to build real-time or near-real-time dashboards.

* **Amazon QuickSight**  
  AWS-native BI tool. Connects to Redshift, RDS, Athena, S3, etc.
* **3rd Party Tools (integrated)**  
  Like Tableau, Power BI, or Looker—can connect to Redshift or other storage options.

### 5. ****Security & Governance****

To ensure compliance, encryption, and access control.

* **AWS IAM** (Identity and Access Management)  
  For managing access control.
* **AWS KMS** (Key Management Service)  
  For encrypting data.
* **AWS Lake Formation**  
  For fine-grained data access and governance on S3.

# Real-time data pipeline Architecture Overview

This project focuses on building a scalable, near real-time data pipeline that extracts data from a source database and delivers it to a cloud data warehouse on **AWS** for analytics and dash boarding. The aim is to ensure that business users can access up-to-date insights via dashboards with minimal latency. The architecture involves capturing changes from the source database using **AWS Database Migration Service (DMS) or AWS Glue with Change Data Capture (CDC).**

[AWS Database Migration Service](https://aws.amazon.com/dms/) (AWS DMS):-

  It is a cloud service that makes it easy to migrate relational databases, data warehouses, NoSQL databases, and other types of data stores. You can use AWS DMS to migrate your data into the AWS Cloud or between combinations of cloud and on-premises setups. AWS DMS also helps you replicate ongoing changes to keep sources and targets in sync.

Change data capture (CDC) refers to the process of identifying and capturing changes made to data in a database and then delivering those changes in real time to a downstream system. Capturing every change from transactions in a source database and moving them to the target in real time keeps the systems synchronized, and helps with real-time analytics use cases and zero-downtime database migrations.

[Amazon Kinesis Data Streams](https://aws.amazon.com/kinesis/data-streams/) :-

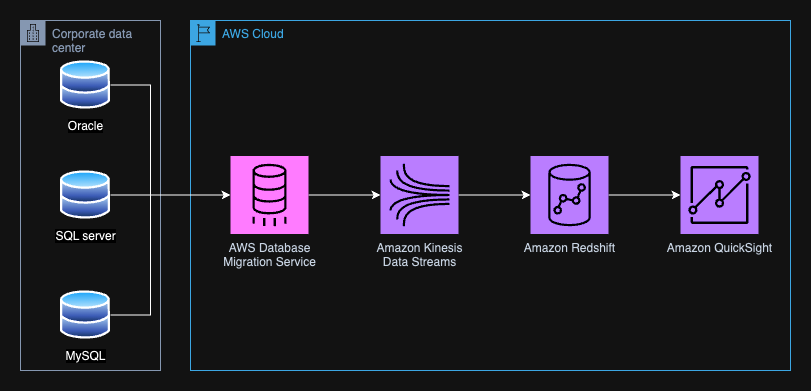
It is a fully managed streaming data service. You can continuously add various types of data such as clickstreams, application logs, and social media to a Kinesis stream from hundreds of thousands of sources. Within seconds, the data will be available for your Kinesis applications to read and process from the stream.

[Amazon Redshift](https://aws.amazon.com/redshift/) :-

It is a fully managed cloud data warehouse that’s used by tens of thousands of customers for price-performance, scale, and advanced data analytics. Amazon Redshift enables data warehousing by seamlessly integrating with other data stores and services in the modern data organization through features such as [zero-ETL](https://aws.amazon.com/what-is/zero-etl/), data sharing, streaming ingestion, data lake integration, and [Redshift ML](https://aws.amazon.com/redshift/features/redshift-ml/).

[Amazon QuickSight](https://aws.amazon.com/quicksight/) :-

It is a fully managed, cloud-centered BI service that makes it straightforward to connect to your data, create interactive dashboards, and share these with tens of thousands of users, either within the QuickSight interface or embedded in software as a service (SaaS) applications or web portals.



*Solution architecture of the solution described in this post. The major components are AWS Database Migration Service, Amazon Kinesis Data Streams, Amazon Redshift, and Amazon QuickSight.*

## Services Used in the Project:

### ****1. AWS Database Migration Service (DMS)****

**Purpose:**  
AWS DMS is used to migrate data from on-premises databases like Oracle, SQL Server, and MySQL to AWS in near real-time.

**Key Features:**

* Supports Change Data Capture (CDC) for real-time replication
* Minimal downtime during migration
* Supports heterogeneous migrations (e.g., Oracle to Redshift)

### ****2. Amazon Kinesis Data Streams****

**Purpose:**  
Acts as a real-time data streaming platform that ingests data changes from DMS and pushes them further into the pipeline.

**Key Features:**

* Handles high-throughput streaming data
* Ensures low-latency delivery
* Scalable and durable stream storage
* Can be consumed by multiple downstream applications in parallel

### ****3. Amazon Redshift****

**Purpose:**  
Serves as the cloud **data warehouse** where transformed data is stored and made available for analytics.

**Key Features:**

* Optimized for complex queries and large-scale analytics
* Supports SQL-based querying
* Integrates easily with BI tools
* High performance with columnar storage and massively parallel processing (MPP).

### ****4. Amazon QuickSight****

**Purpose:**  
 Used to create **interactive dashboards and visualizations** from the data stored in Amazon Redshift.

**Key Features:**

* Cloud-native BI service
* Real-time analytics and reporting
* Supports ML-based insights (like anomaly detection)
* Scalable for enterprise users

# Screenshots

# C:\Users\DELL\OneDrive\Desktop\CLOUD\SDC\Project\Screenshot 2025-04-05 211939.png

# *Figure 1. Amazon RDS Create database screen with options for database creation method (Easy create selected) and configuration- of MySQL selected.*

# C:\Users\DELL\OneDrive\Desktop\CLOUD\SDC\Project\Screenshot 2025-04-05 215250.png

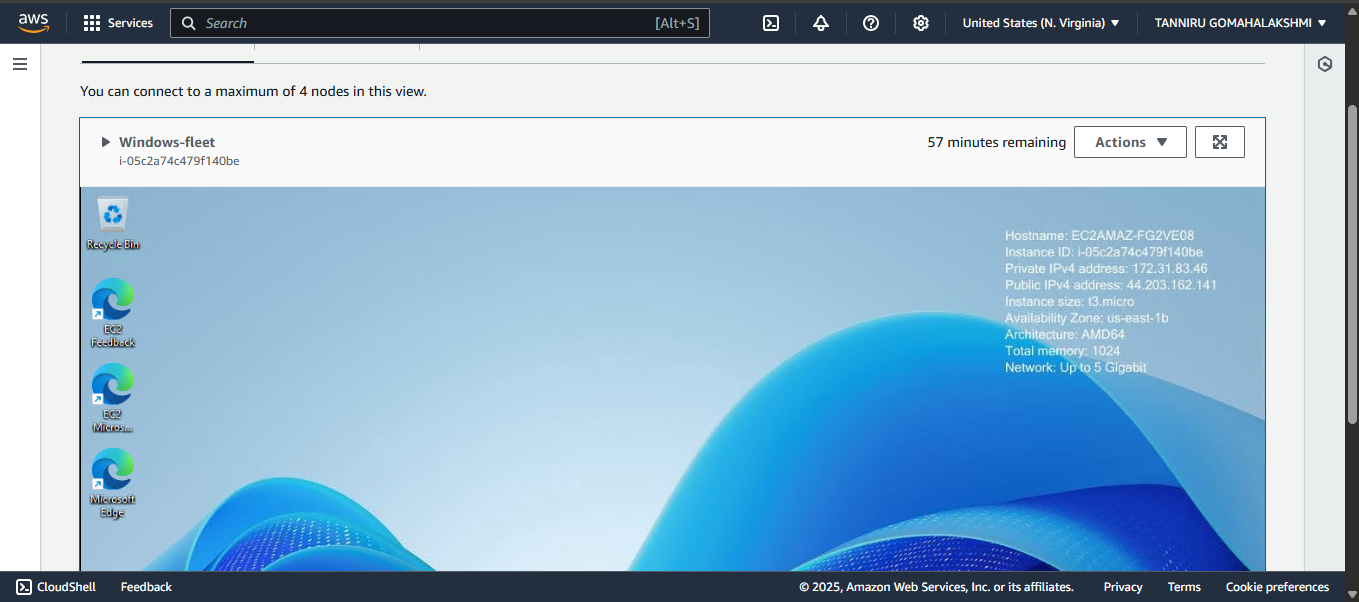
# Figure 2. Amazon RDS Created database.

# C:\Users\DELL\OneDrive\Desktop\CLOUD\SDC\Project\Screenshot 2025-04-06 084434.png

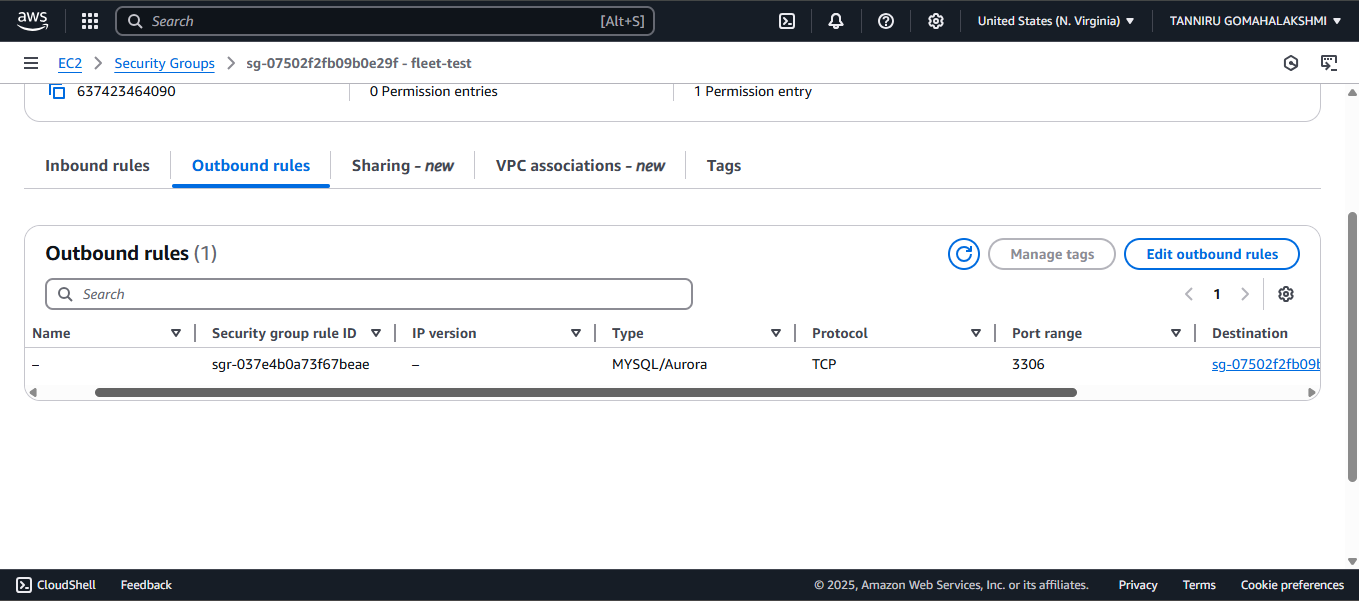
# *Figure 3. Amazon EC2 screen for launching an instance, with Windows AMI selected.*

# 

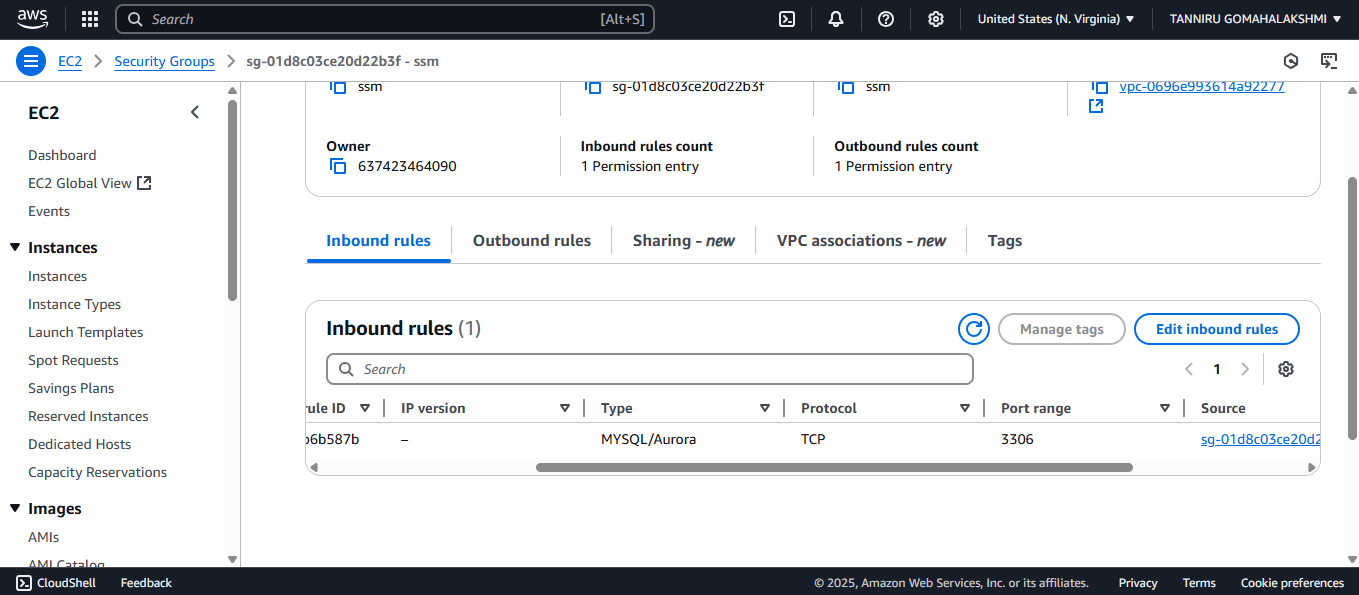
# Figure 3. Amazon EC2 created instance.



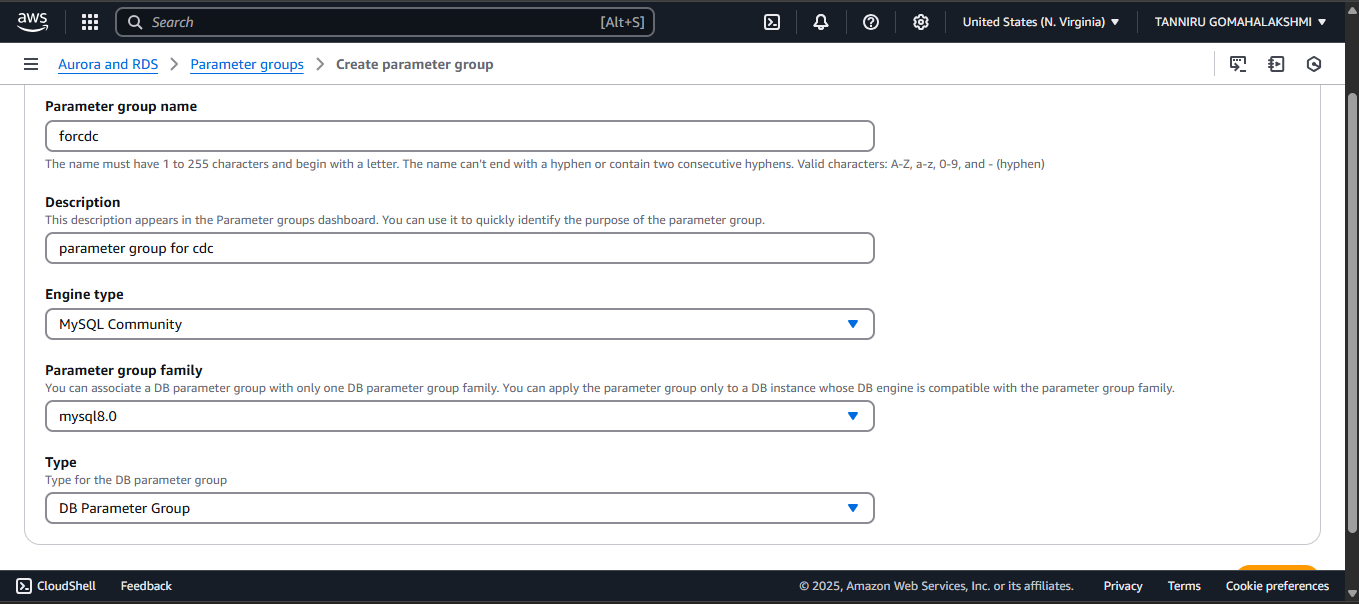
**Figure 4. Connected instance using fleet manager.**



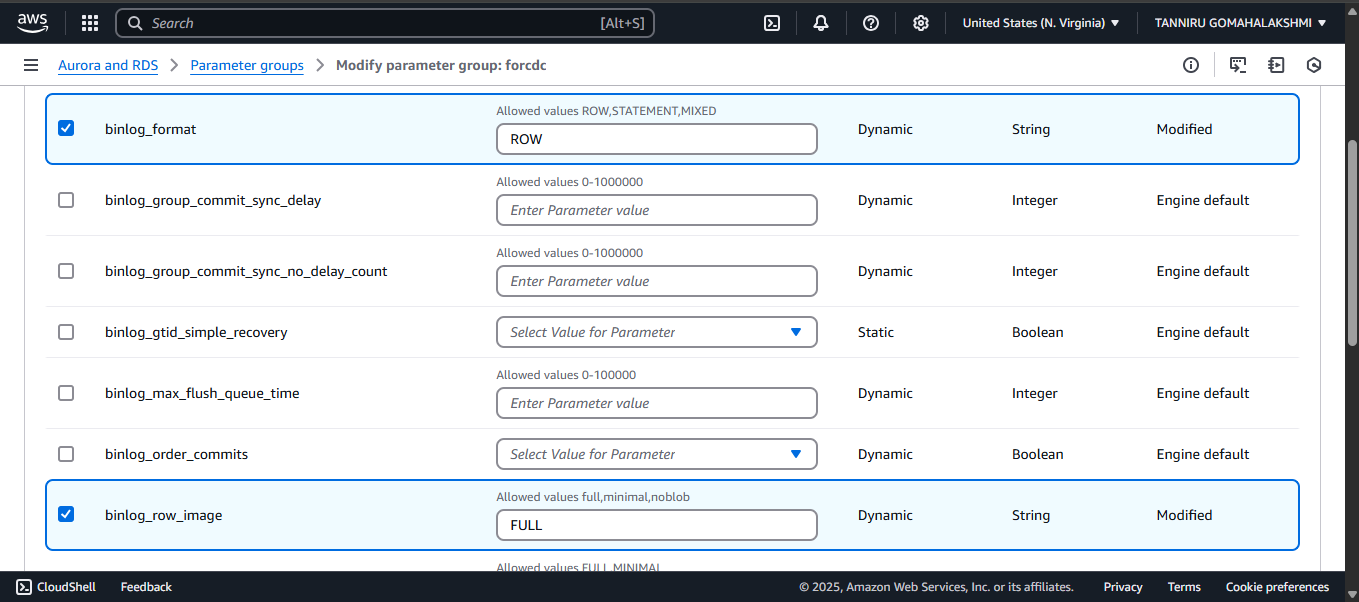
**Figure 5. The Amazon EC2 security group should allow outbound traffic type MYSQL/Aurora TCP with port = 3306  to the destination = security group of the RDS (no inbound rules)**

****

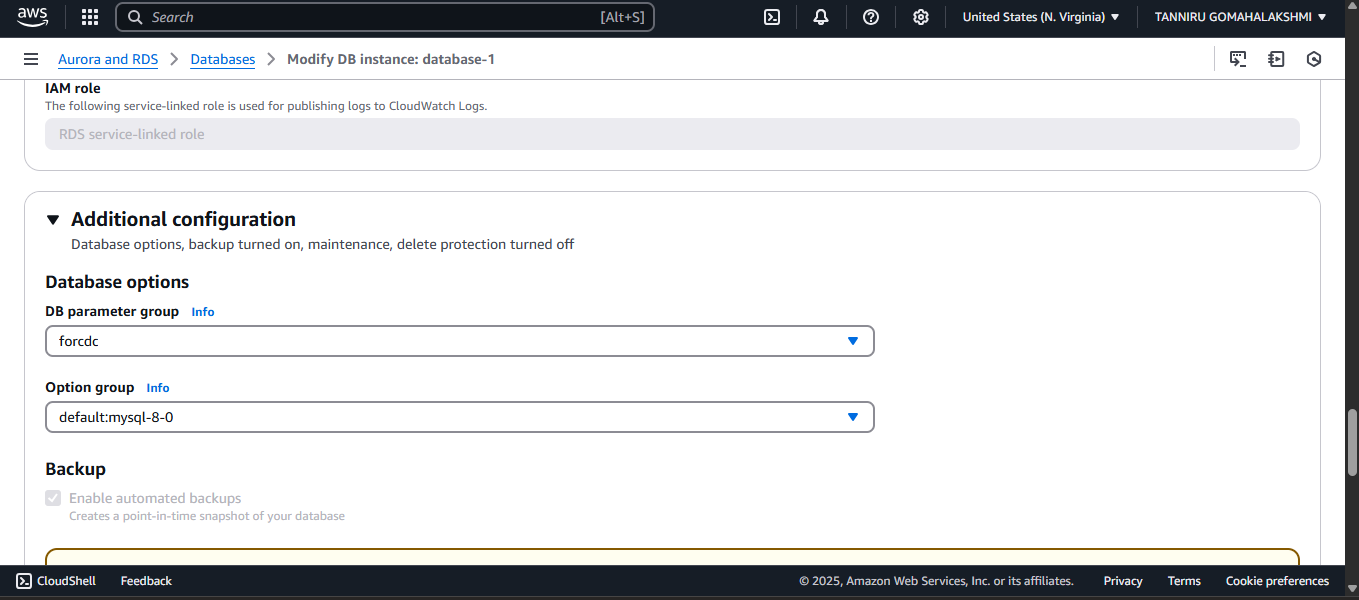
**Figure 6**. **The security group of the Amazon RDS instance should allow inbound of the type** **MYSQL/Aurora TCP with port = 3306** **from** **the security group of the EC2 instance.**



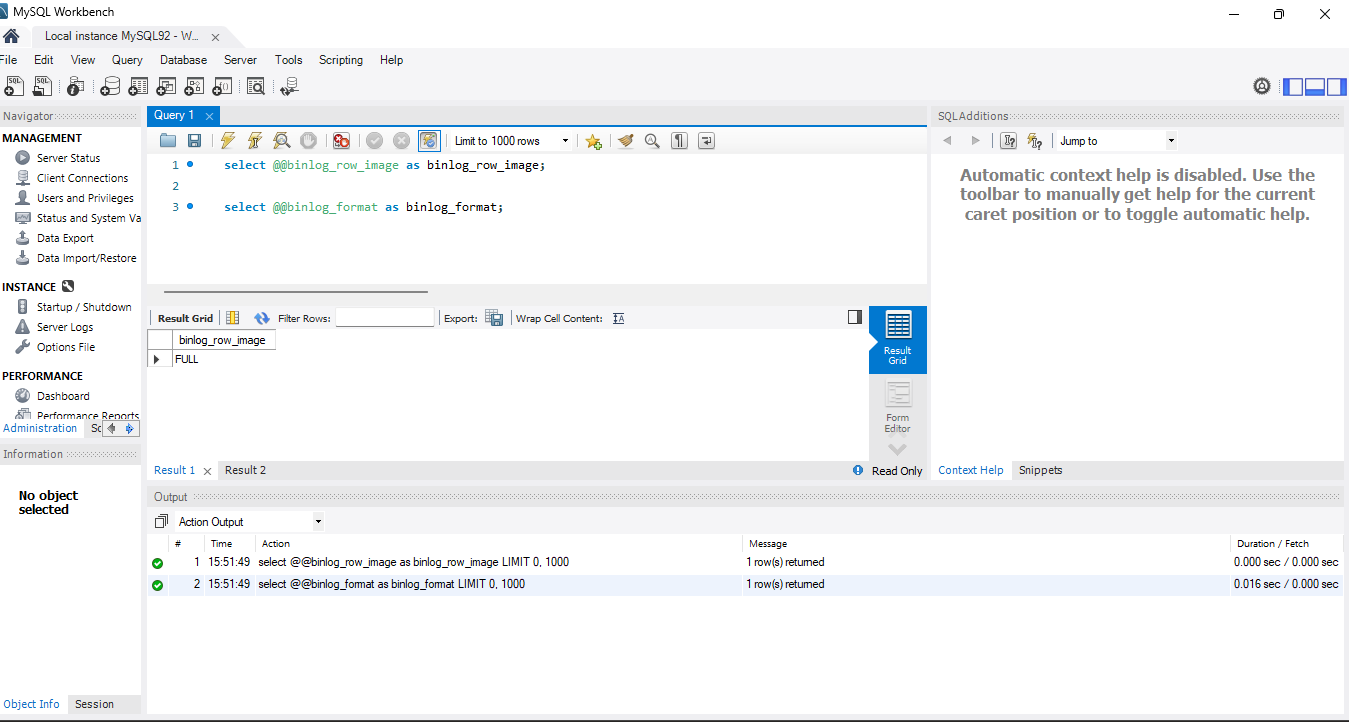
**Figure 7. *Parameter group details screen includes group name, description, engine type, parameter group family, and type fields and dropdown menus.***



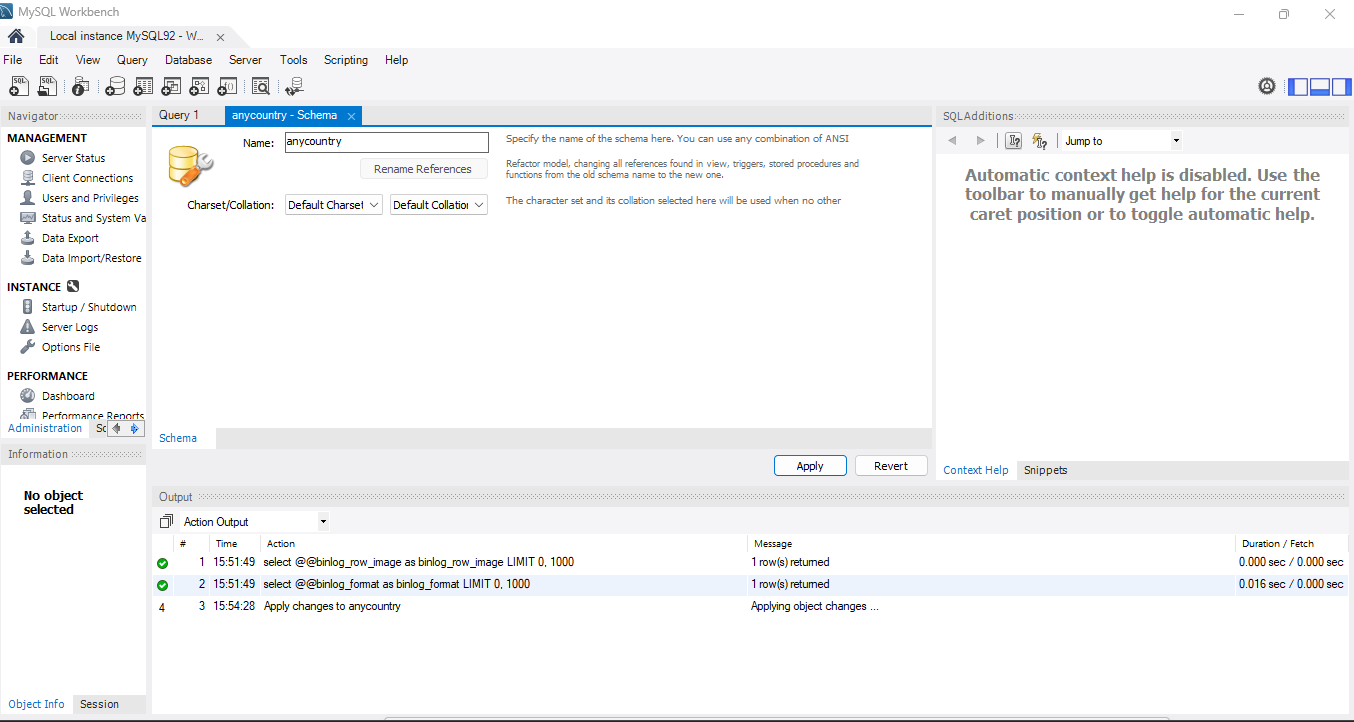
***Figure 8. Editing parameters for binlog\_format and binlog\_row\_image.***



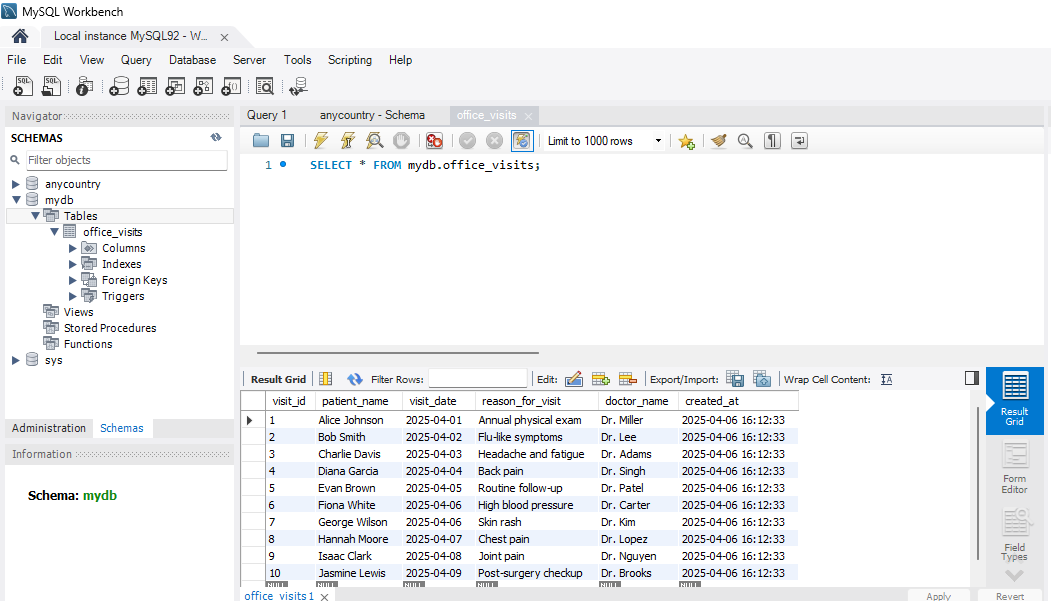
***Figure 9. Changing the DB parameter group.***

****

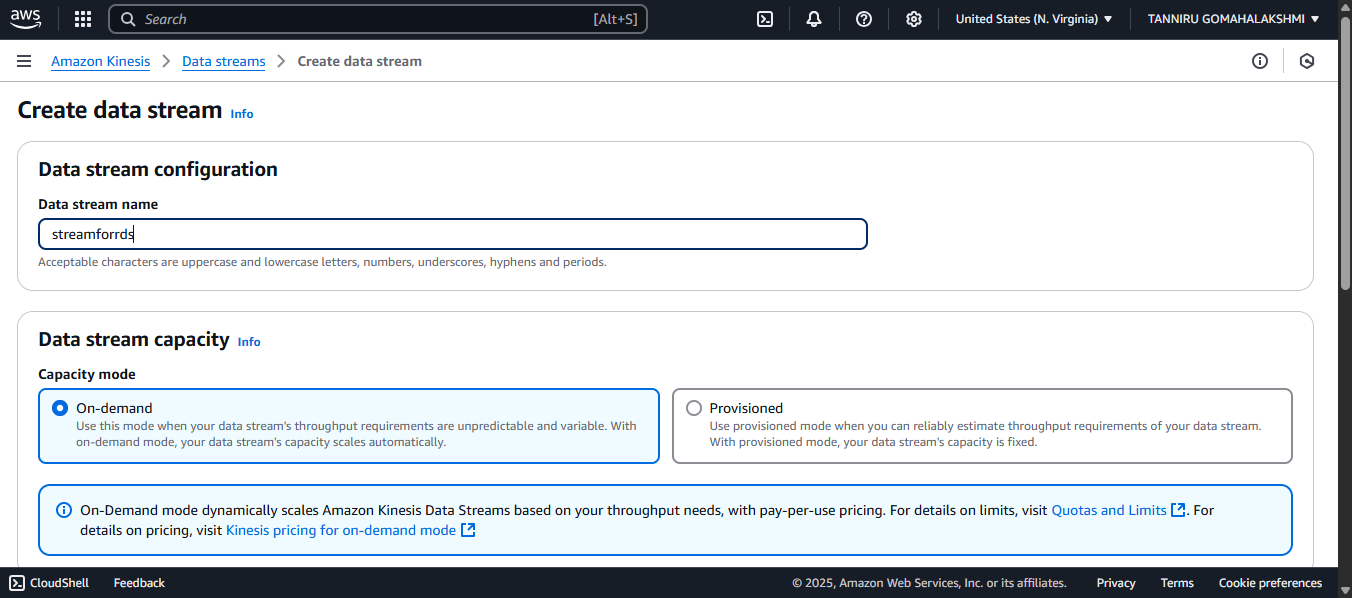
***Figure 10. MySQL Workbench with the queries run.***

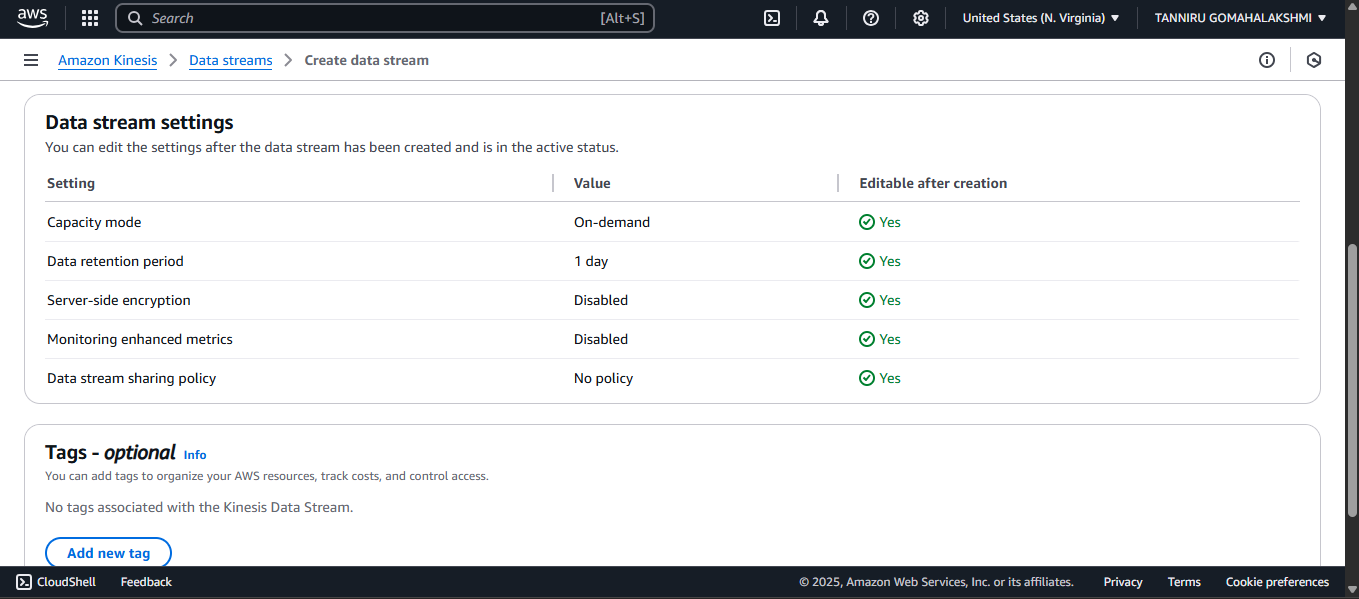
******

***Figure 11. Naming new schema in MySQL Workbench.***

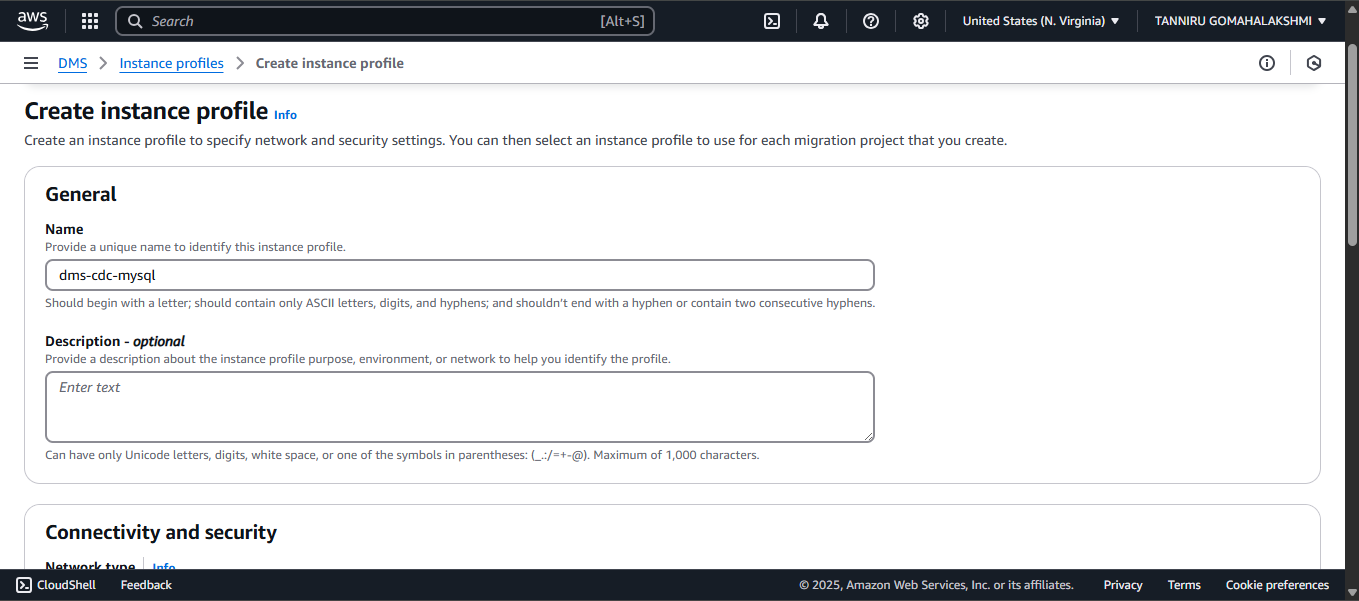
****

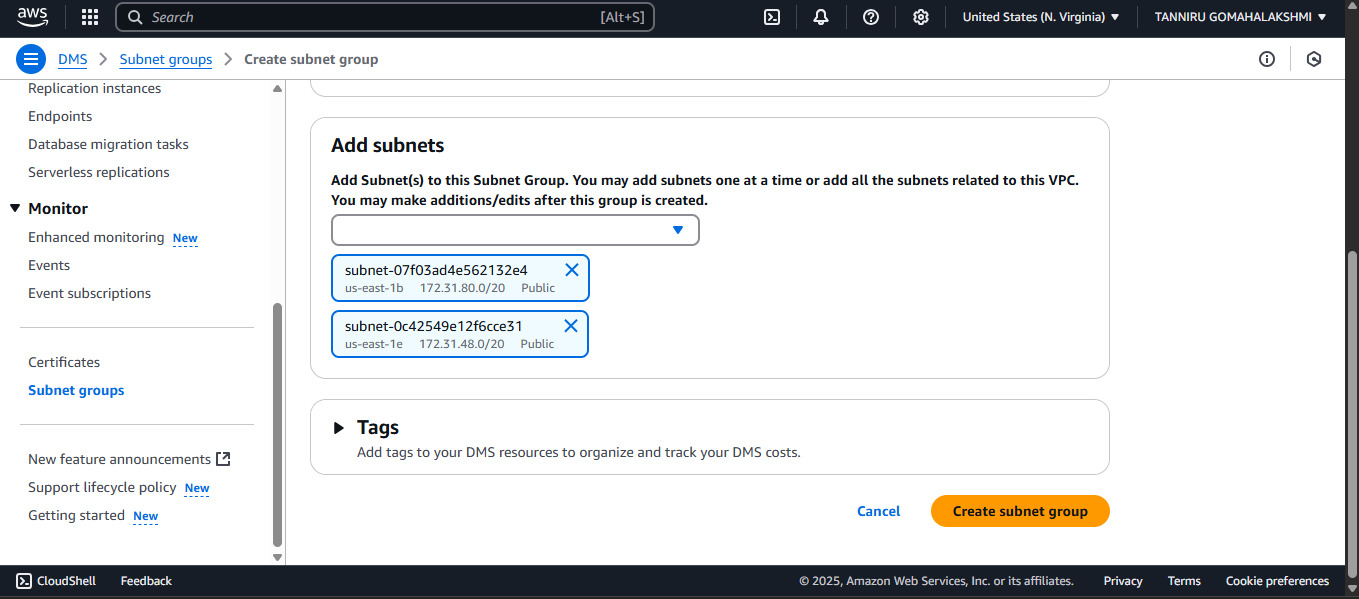
***Figure 12. Adding data to the table in MySQL Workbench.***

****

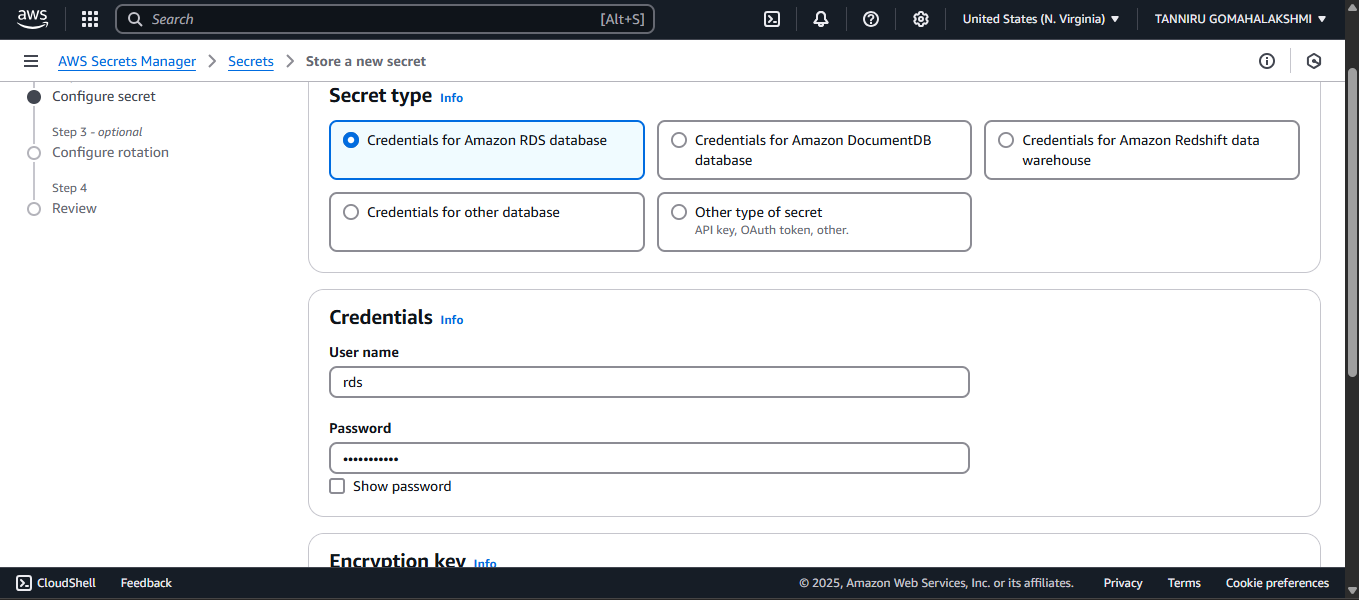
****

***Figure 13. Creating a data stream in Kinesis Data Streams.***

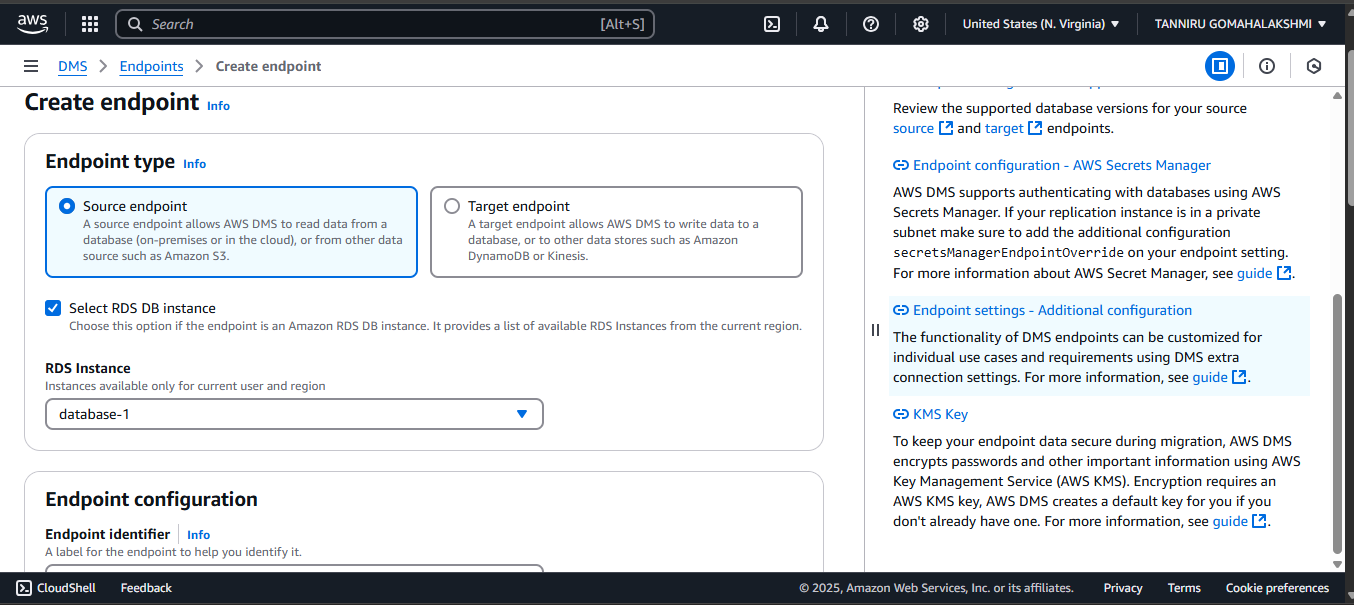


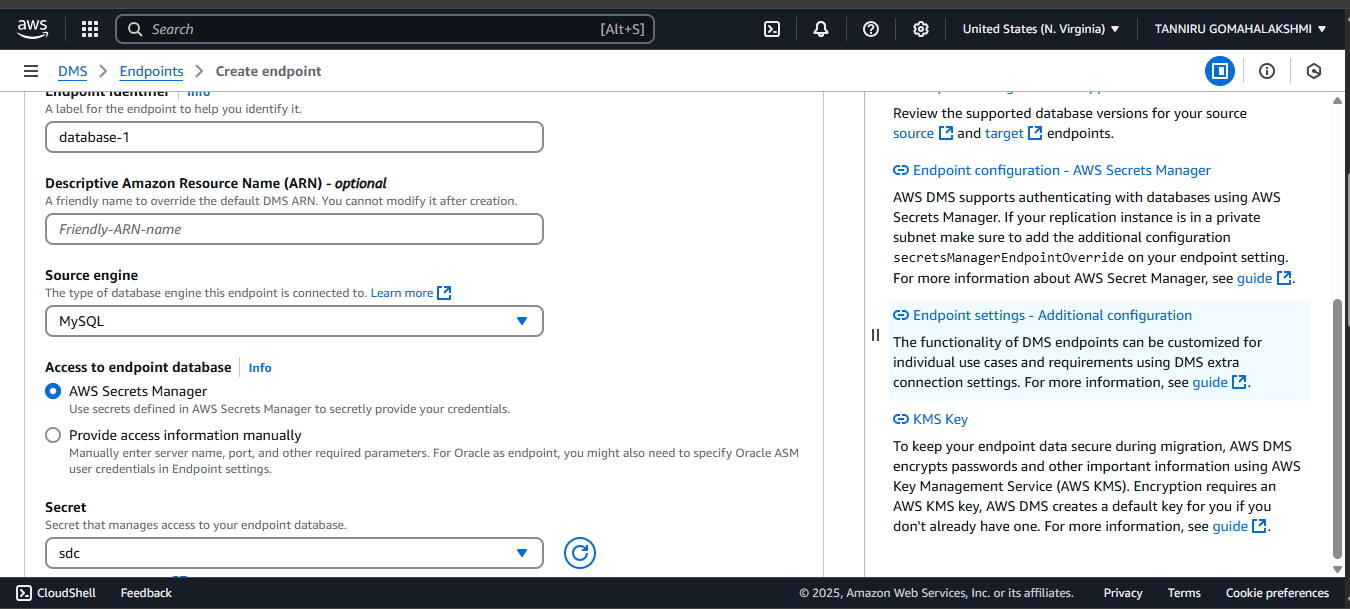


***Figure 14. Creating an instance for CDC.***

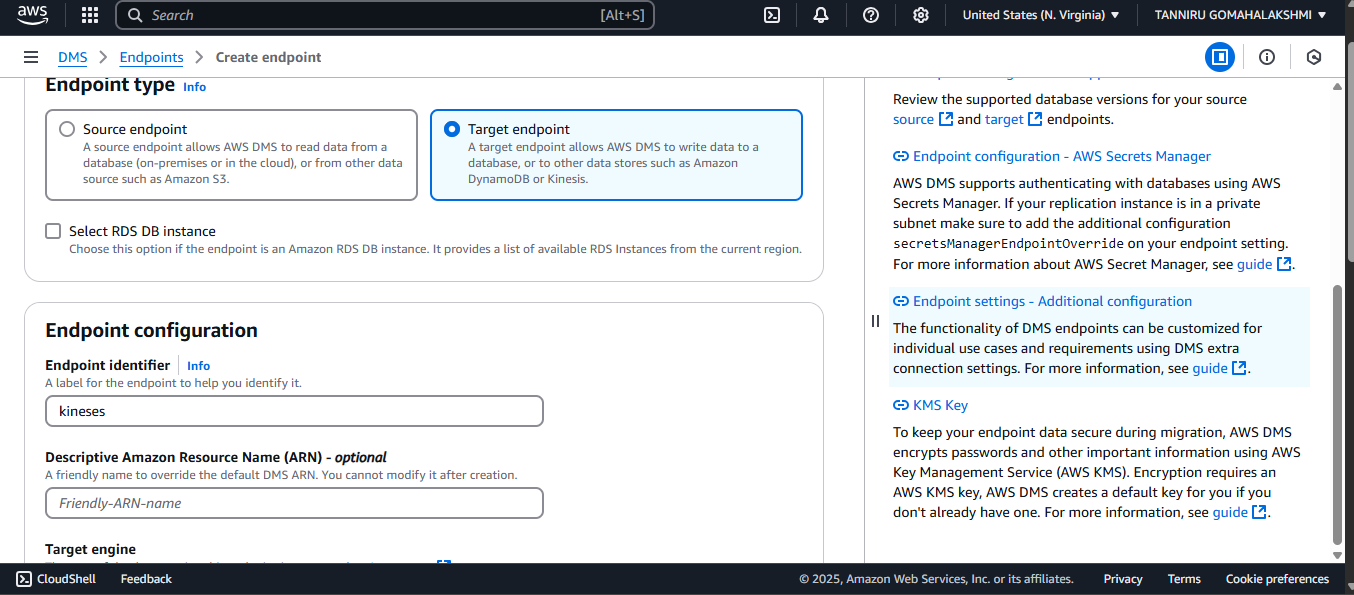


***Figure 15. Creating a secret for Amazon RDS.***

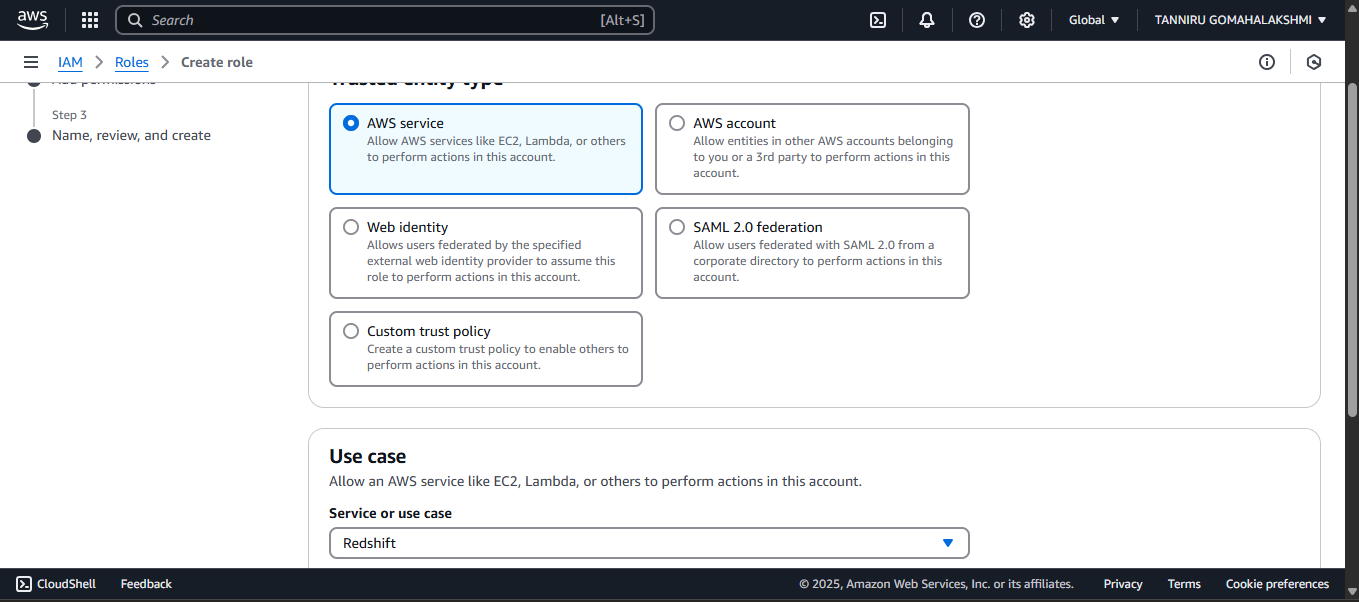




***Figure 16. Creating DMS endpoint to the RDS MySQL database.***



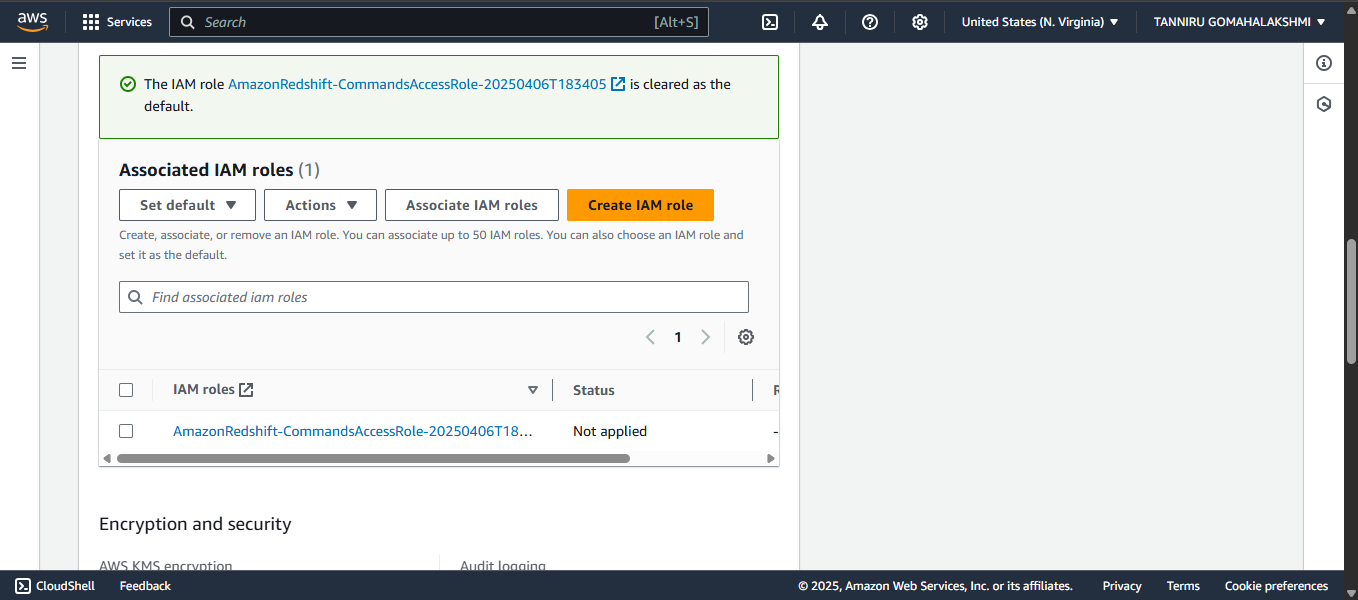
***Figure 18. Creating target endpoint.***



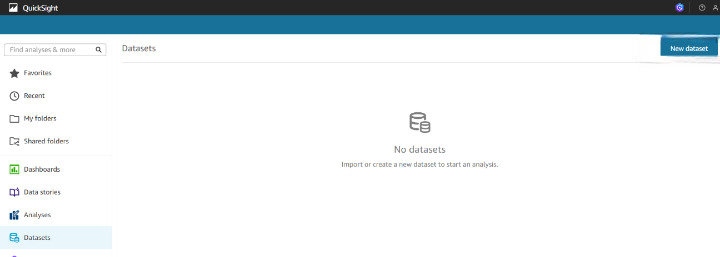
***Figure 20 . Creating an IAM Role for Amazon******Redshift Serverless.***



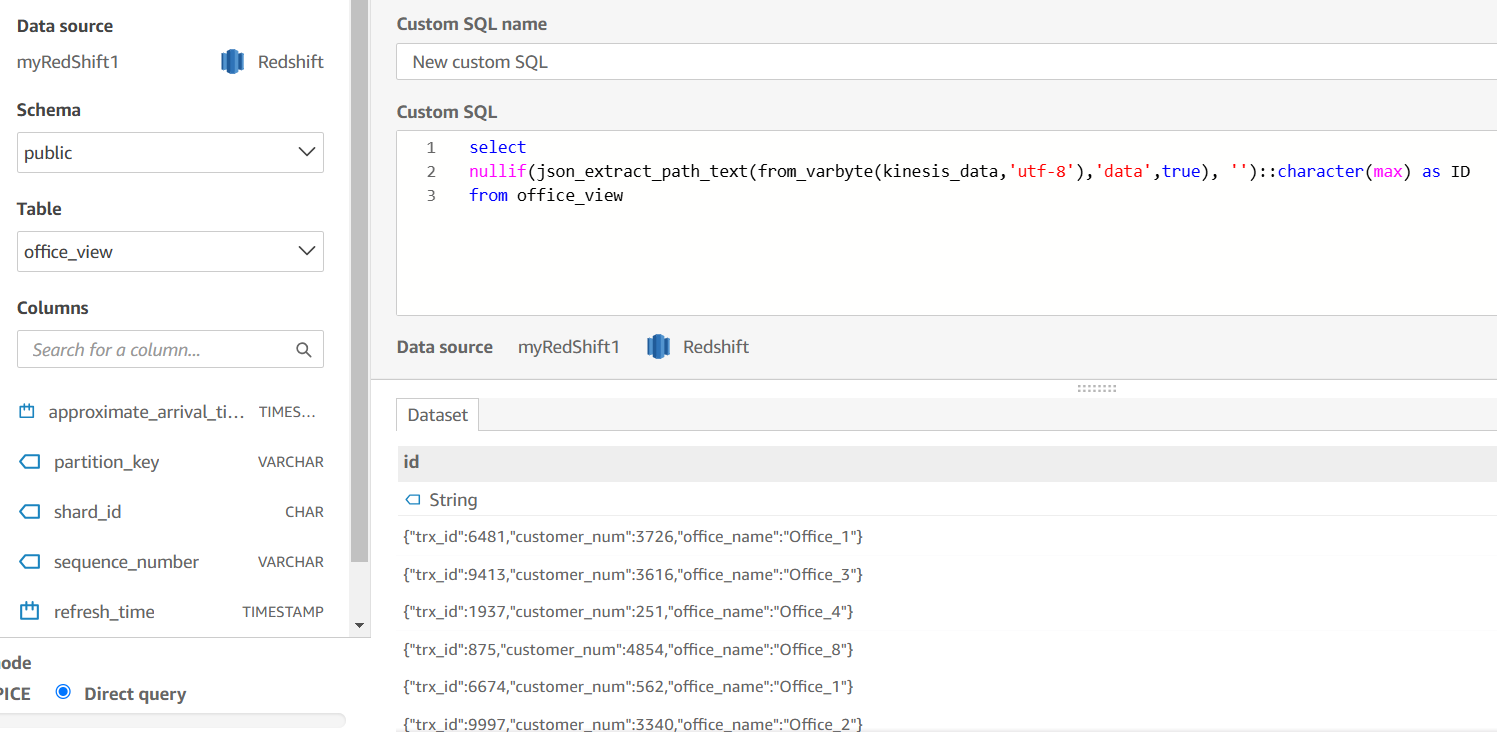
***Code for RedShift IAM Role***



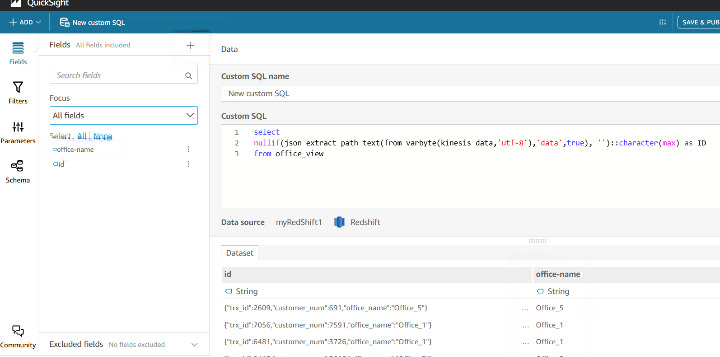
**Figure 22. Creating and Associating *default Amazon Redshift IAM role.***

****

***Figure 23. Selecting “New dataset” button on top right of Amazon QuickSight Datasets page.***



***Figure 24. Entering query to define new dataset.***

****

***Figure 25. Adding field for office names in Amazon QuickSight.***

### Testing for Output

### https://d2908q01vomqb2.cloudfront.net/9e6a55b6b4563e652a23be9d623ca5055c356940/2025/02/26/aws_dms_figure34.png

### *Figure 26. Vertical bar chart for office wait times.*

# Conclusion

# 

With Amazon Kinesis Data Streams as an AWS DMS target, together with a managed, serverless cloud data warehouse and a cloud business analytics service, you now have a powerful way to stream change data from a database directly into your data warehouse without impacting your source database while providing near real-time dashboards of your business operations. You can use this method to stream change data from any sources supported by AWS DMS to perform real-time data processing.

This architecture enables seamless, near real-time data movement from multiple on-premises databases such as Oracle, SQL Server, and MySQL into the AWS Cloud. By leveraging **AWS Database Migration Service**, **Amazon Kinesis Data Streams**, **Amazon Redshift**, and **Amazon QuickSight**, organizations can build a robust and scalable data pipeline. This setup ensures continuous data ingestion, real-time processing, and up-to-date visualization through interactive dashboards. It empowers businesses to make timely, data-driven decisions while reducing operational complexity and accelerating analytics performance in a cost-effective, cloud-native environment.