Create AWS Glue Database & Table from S3

# Introduction

This tutorial guides you through creating a database and tables in AWS Glue using JSON data stored in an S3 bucket. Learn to configure a crawler for automatic schema detection, manage IAM roles, and adjust table schemas, enabling efficient data integration and analysis within the AWS ecosystem.

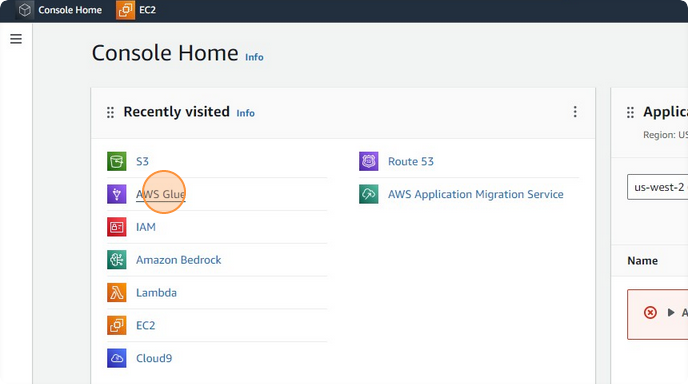
# Let’s get Started

**Access AWS Management Console:**

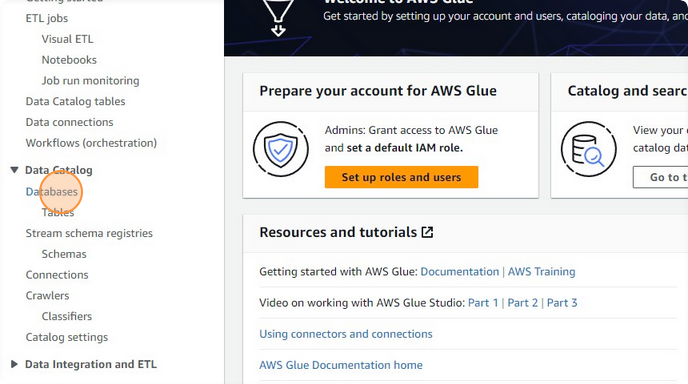
* Navigate to the [AWS Management Console](https://us-west-2.console.aws.amazon.com/console/home?region=us-west-2) for the US West (Oregon) region using the provided URL

**Open AWS Glue Service**

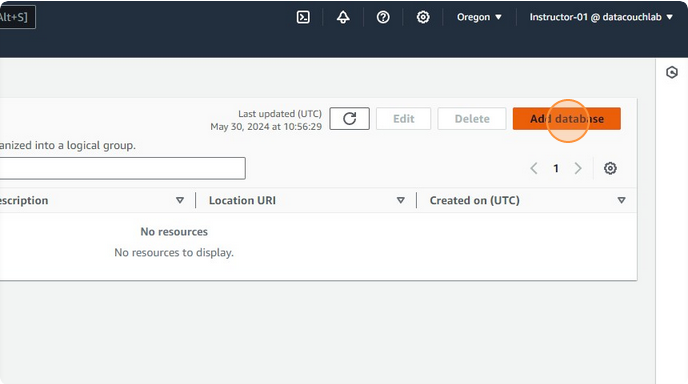
* Click on "**AWS Glue**" from the service menu to access the AWS Glue console where you can manage databases, crawlers, and jobs



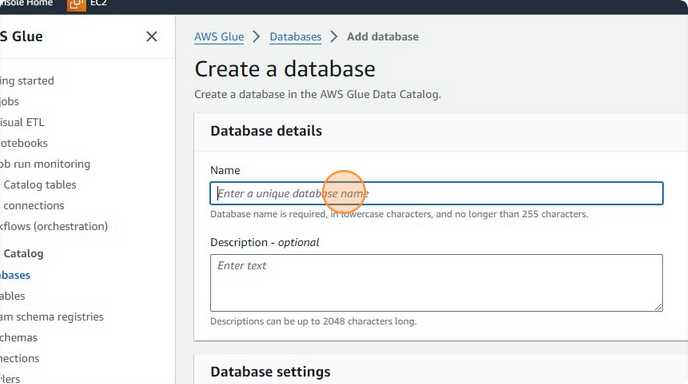
* Click on "**Databases**" to view existing databases or to create a new one



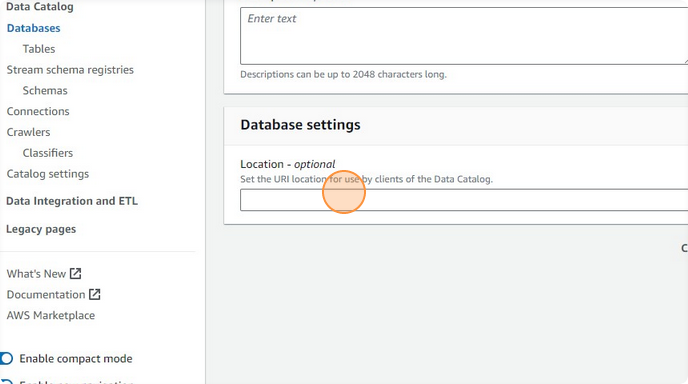
* Click "**Add database**" to initiate the creation of a new database which will store metadata related to your data assets.



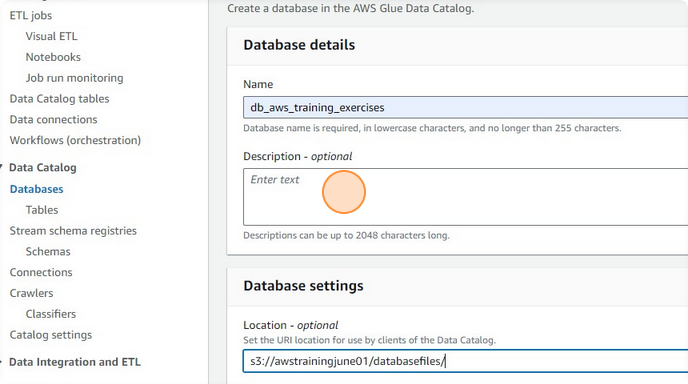
* Click on the "**Name**" field twice and enter a unique name for your database, such as "**db\_aws\_training\_exercises**"



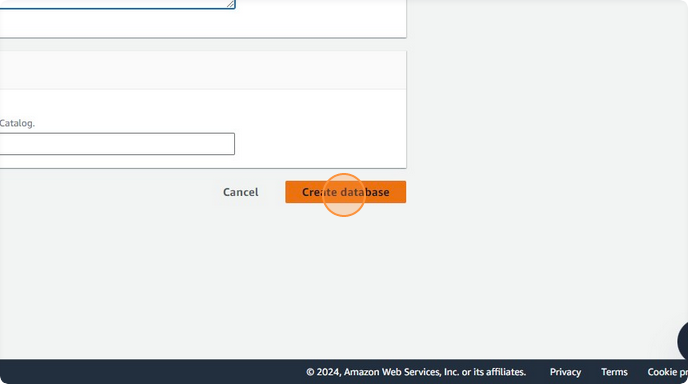
* Click the "**Location-optional**" field and if necessary, paste (ctrl + v) a predefined S3 path where metadata for this database should be stored.



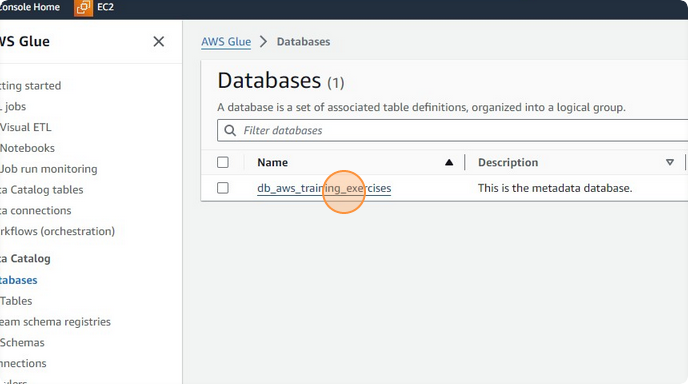
* Click on the "**Description-optional**" field and type a brief description, such as "**This is the metadata database**"



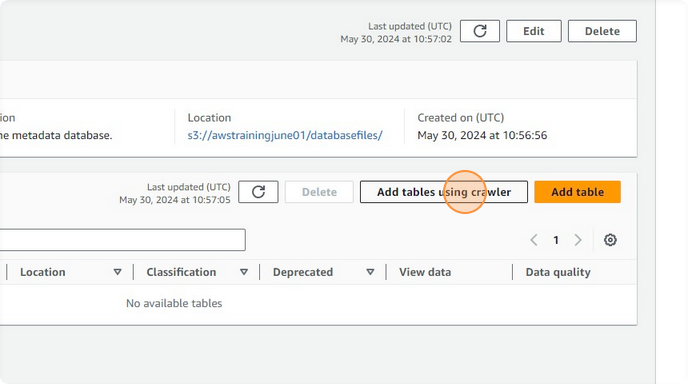
* Click the "**Create database**" button to finalize the creation of the Glue database.



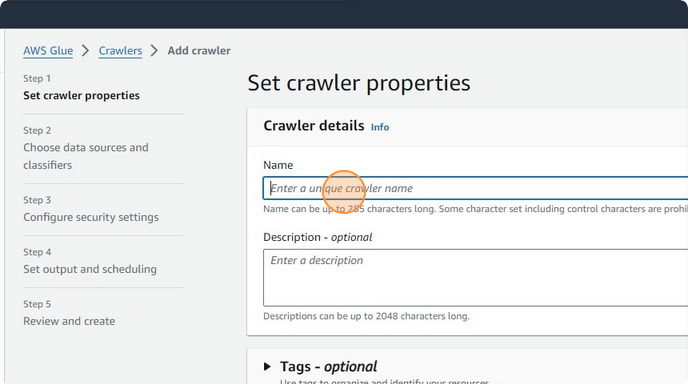
* Click on the newly created database "**db\_aws\_training\_exercises**" to start adding tables.



* Click "**Add tables using crawler**" to set up a crawler that will explore your data and automatically propose table structures.

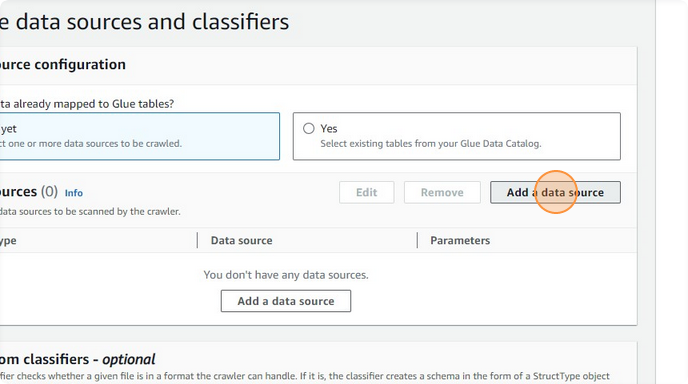


* Click the "**Name**" field and enter a name for the crawler, e.g., "**crwl\_Customers\_Orders**".

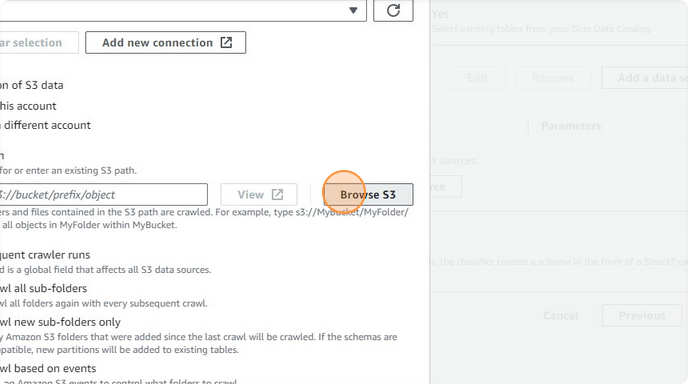


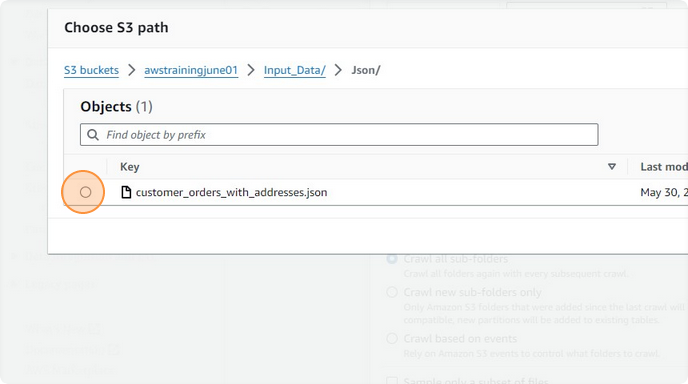
**Add Data Source:**

* Click "**Add a data source**" and then **"Browse S3**" to select the bucket and folder where your JSON files are stored.

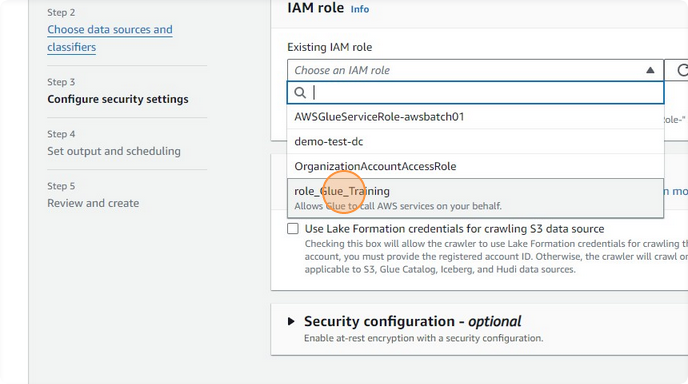


* Navigate through “**Browse S3**” > "**awstrainingjune01**" > "**Input\_Data/**" > "**Json/**" and select this path.

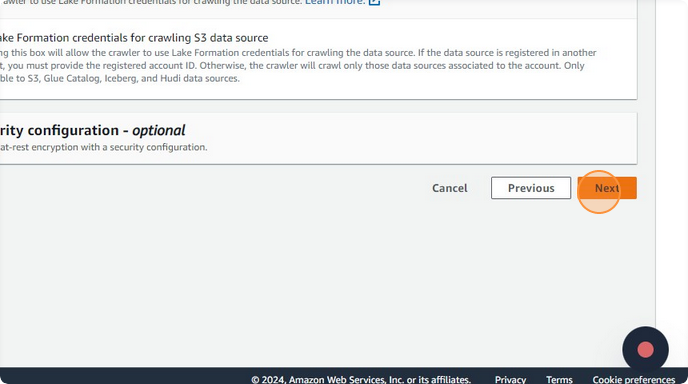




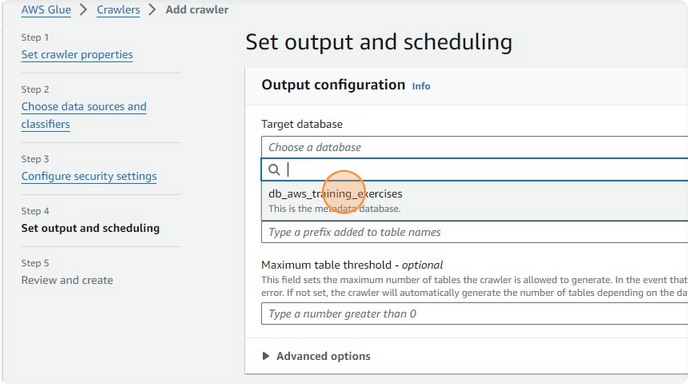
* Click "**Choose an IAM role**" and select "**role\_Glue\_Training**" which has the necessary permissions to access the S3 data and create resources



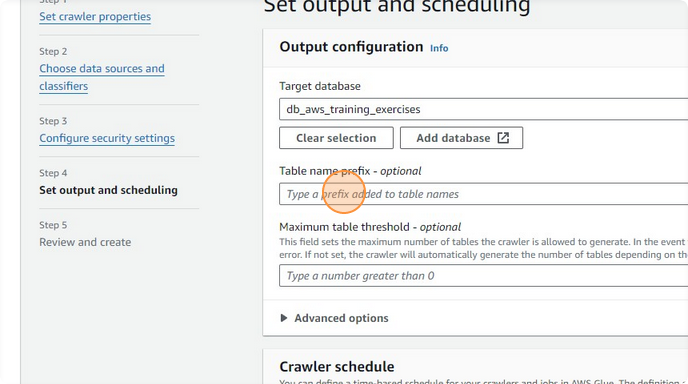
* Click "**Next**"



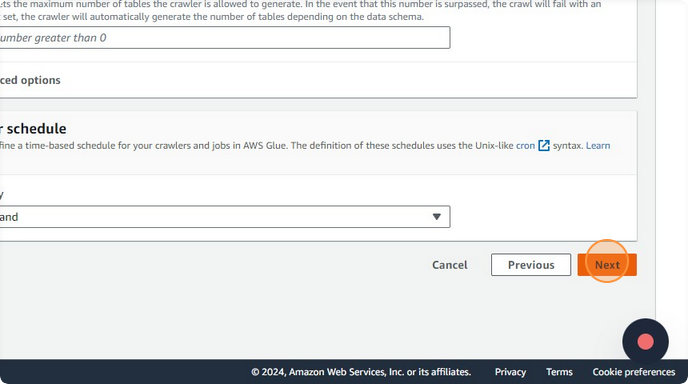
* Click "**Choose a database**" and select "**db\_aws\_training\_exercises**" to specify where the new tables should be created



* Click the "**Table name prefix - optional**" field and type "**tbl\_**" to prefix all created table names for easier identification.

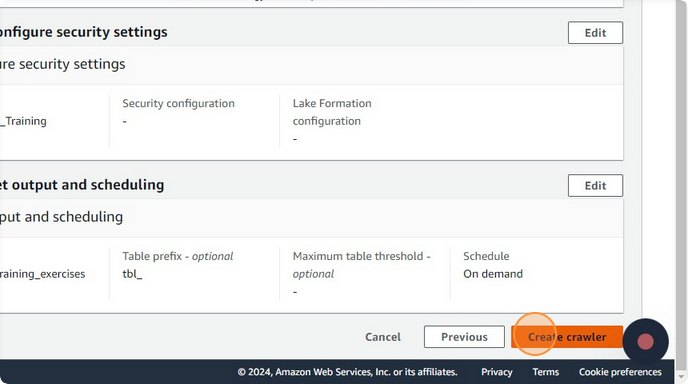


* Click "**Next**"

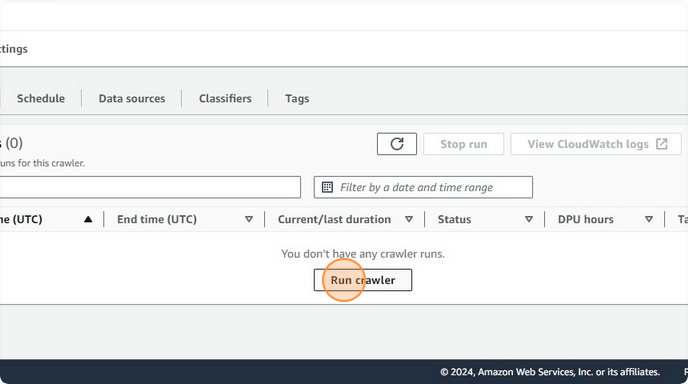


**Create and Run the Crawler:**

* Click "**Create crawler**"

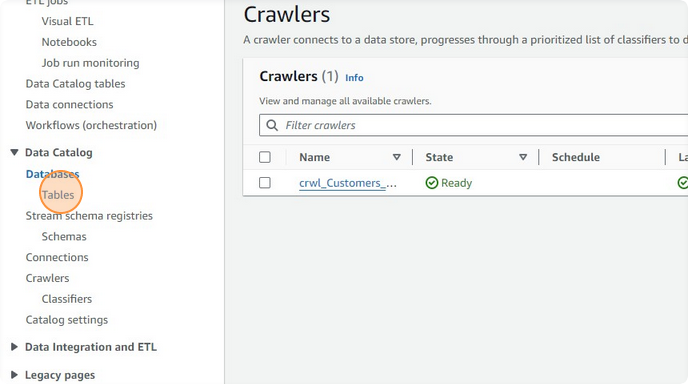


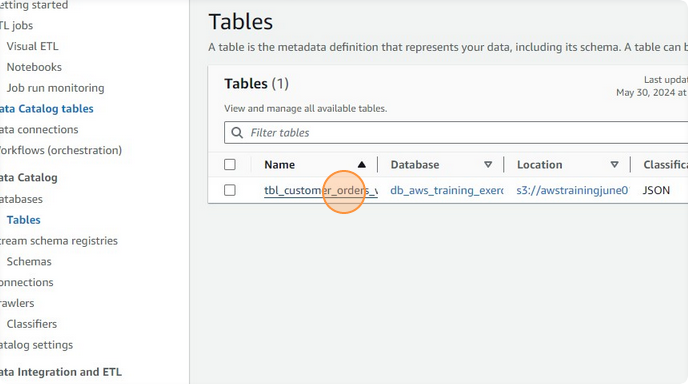
* Navigate to crawler created and then "**Run crawler**" to start the process of crawling the S3 JSON data and creating table schemas based on its structure



**Review Created Tables:**

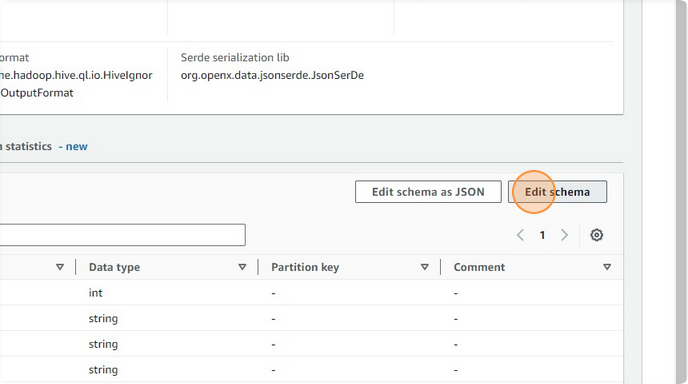
* After the crawler completes, navigate to "**Tables**" and view the created tables, such as "**tbl\_customer\_orders\_with\_addresses\_json**"





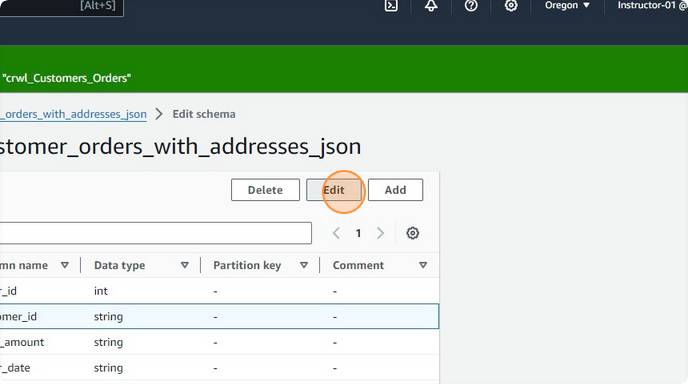
**Edit Table Schema:**

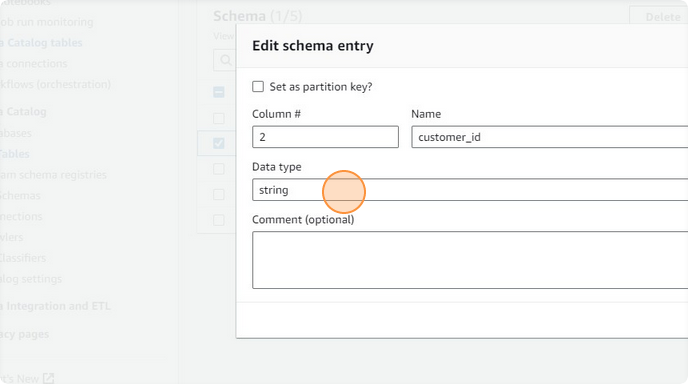
* Select a table and click "**Edit schema**" to make necessary adjustments to the data types and fields to accurately reflect the data structure and requirements.



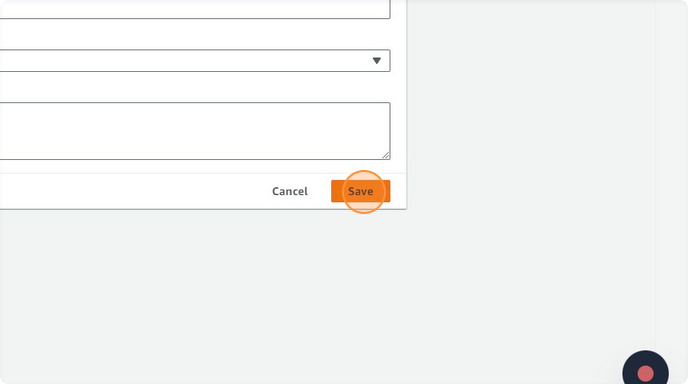
**Final Adjustments and Save:**

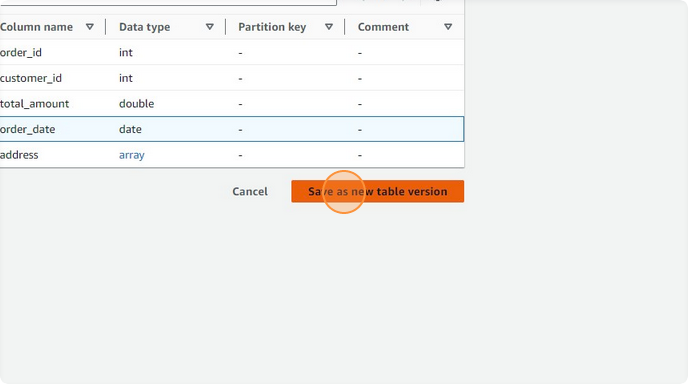
* Make final adjustments like changing data types from "string" to "int", "double", or "date" as necessary for each field, ensuring that the schema accurately represents the data types contained in the JSON files by clicking on **Edit** button



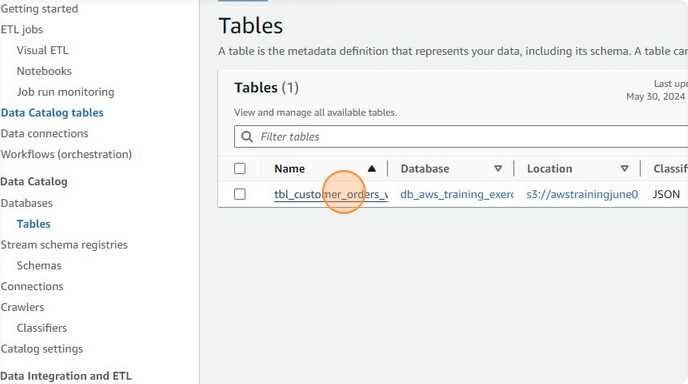


* Click "Save" and then "Save as new table version" to preserve these changes.





* Go to "tbl\_customer\_orders\_with\_addresses\_json" table again



**Voila!!** We have successfully completed this exercise