Title : working with data as a product

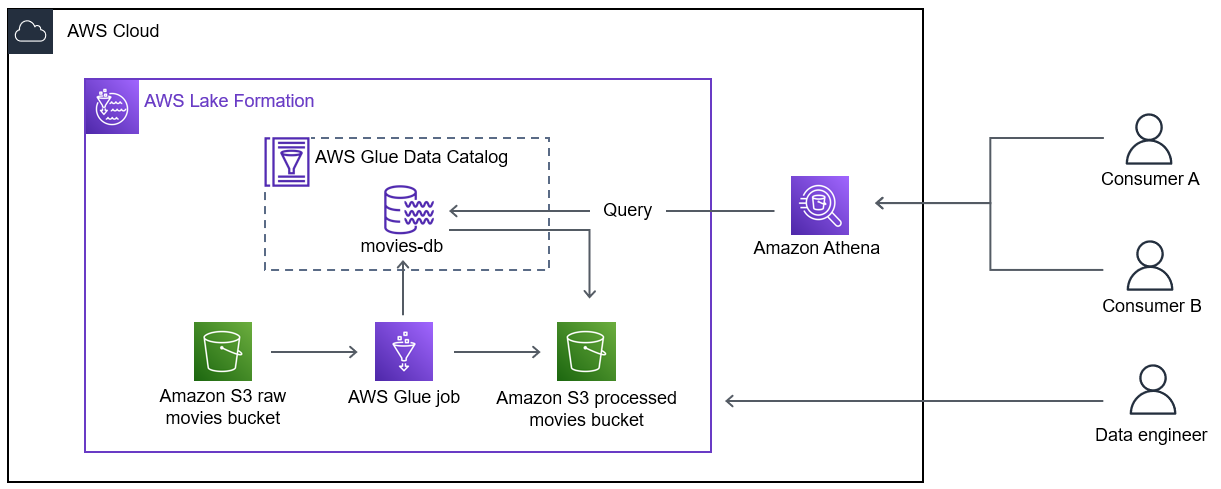
Objectives:

* View an AWS Glue job that maintains a dataset.
* Define LF-tags and apply them to resources.
* Grant LF-TBAC permissions to data consumers.
* Verify consumer-specific data views using Athena.

1>we will be creating LF tags in order to assign specific permissions.

2> we will be applying these LF tags for the schema/

3> we will allot these permissions to specific users.



*Two consumers query the Lake Formation database with Athena. Athena uses the AWS Glue Data Catalog and reads data from the database. A data engineer accesses and works with LF-tags to control the consumers’ access.*

## **Task 1: Test and validate an AWS Glue job**

The AnyCompany application developers want you to view the job and validate the processed data before including the processed data in their application.

1> go to  
AWS glue --> data integration and ETL --> ETL jobs--> <job you created>

2> read the script and analyse the information such as number of nodes and the work they do.

3> under the schedule tab of the location specified in “1”, check the schedule of the job.

In this lab , the job is scheduled to run every 10 minutes.

4> check the runs tab :

The **Runs** tab shows all the recent job runs. You can run jobs automatically using a trigger, or manually when viewing an AWS Glue job. This job is set to run automatically based on a schedule.

TASK 1.2 (validate an AWS glue job):

Query the processed data via athena

TASK 2 (define a set of LF tags):

AnyCompany wants the data tagged with several different keys, including the following keys with their corresponding values:

* Environment: Development, Production
* Customer: Regular, Enterprise
* Confidential: True, False

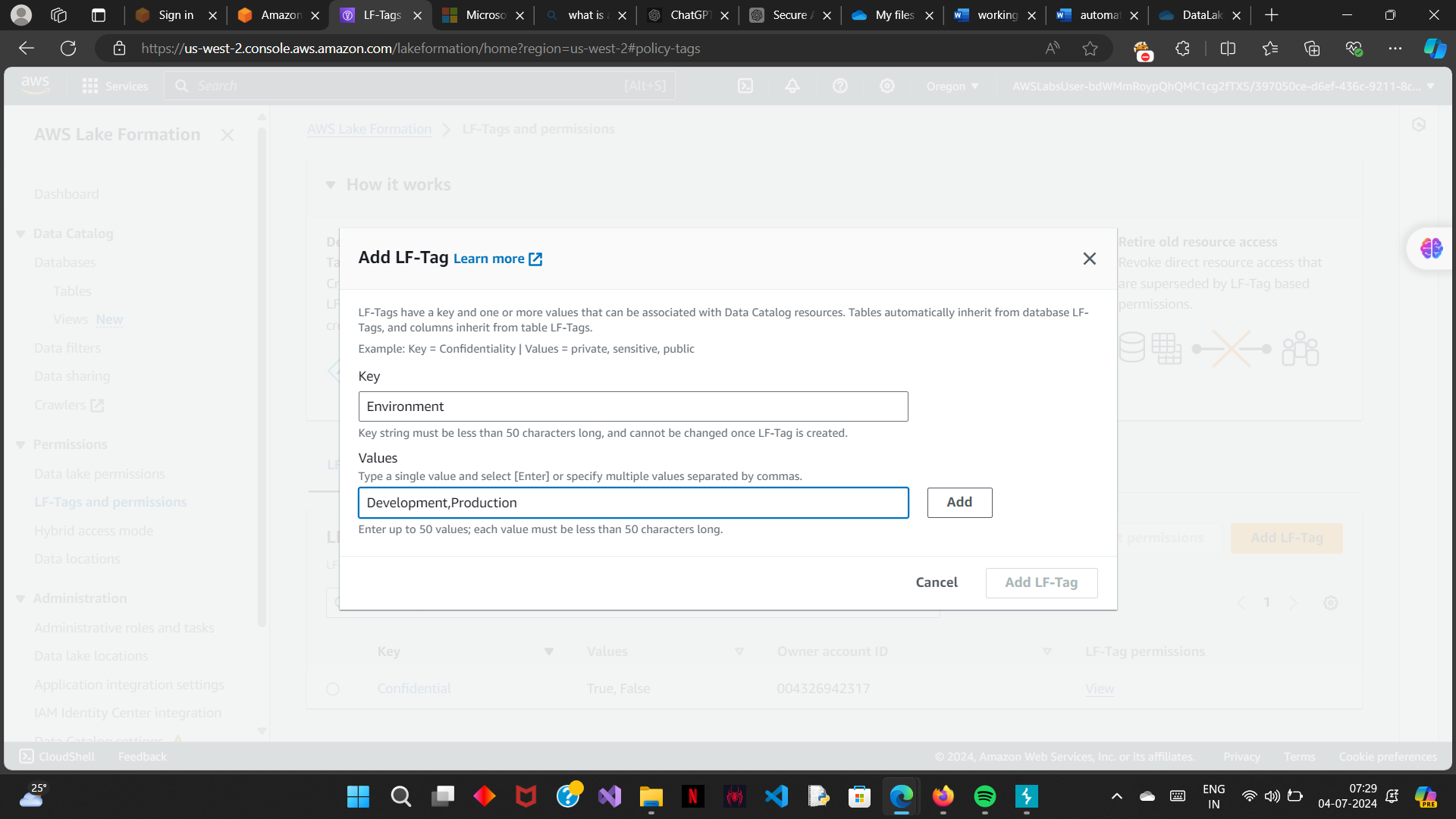
In this task, you define LF-tags in a Lake Formation data lake using your Lake Formation administrator permissions.

TASK 2.1(define an LF-tag)

1>go to:

AWS lake formation --> permissions --> LF tags and permissions --> Add LF tag.

Here, you can add the keys and values mentioned above .



TASK 3 (apply LF tags to resources):

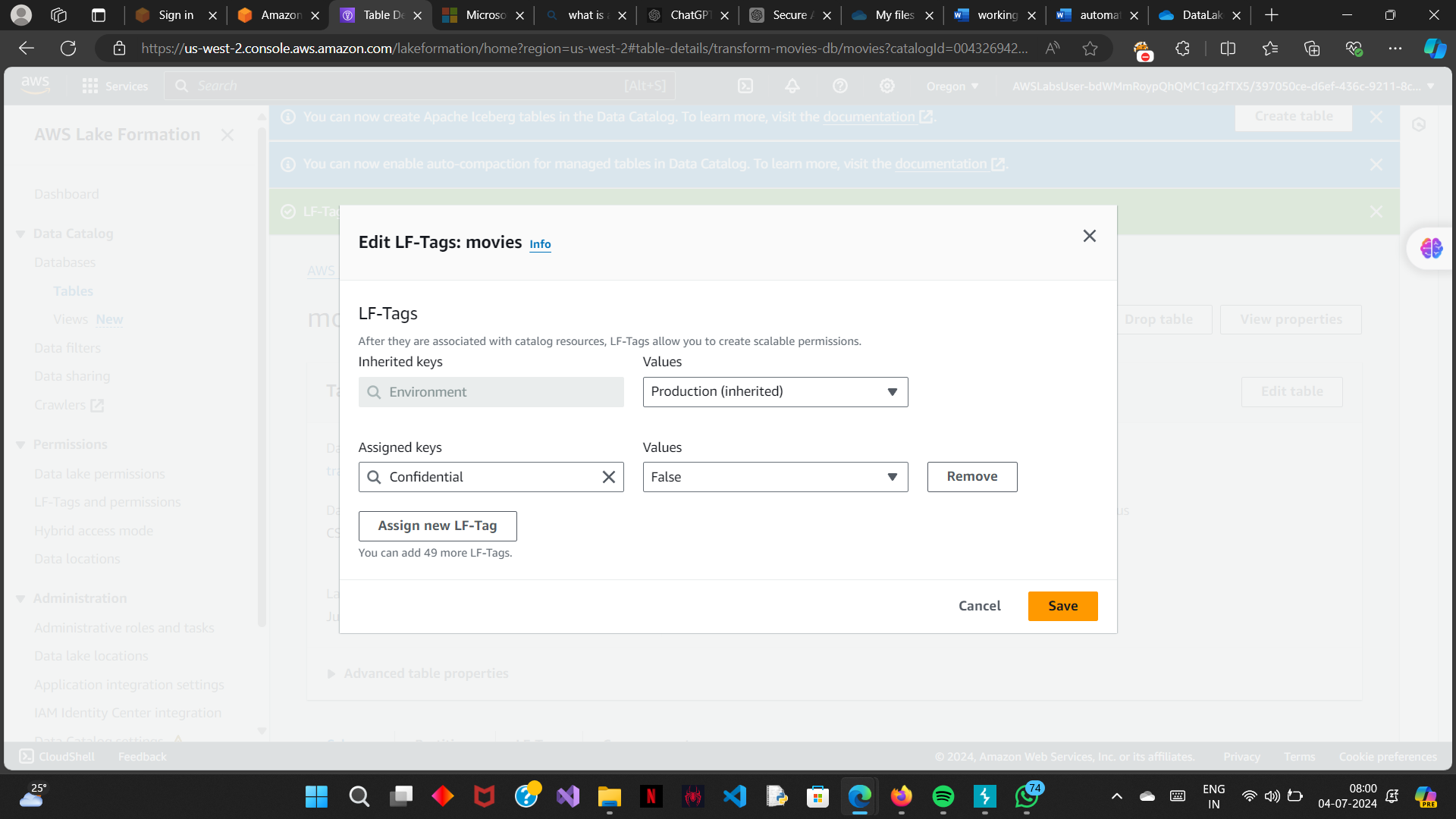
In this task, you use the LF-tags to tag the database, table, and columns in a way that is consistent with AnyCompany’s data sharing plan

TASK 3.1 (apply LF tags to a database):

1>go to:

AWS lake formation --> data catalogue --> databases --> <choose the database> --> actions --> edit LF tags.

Similarly you can go to the tables and can add your lf tags, the lf tags assigned to the database will be inherited by the table.

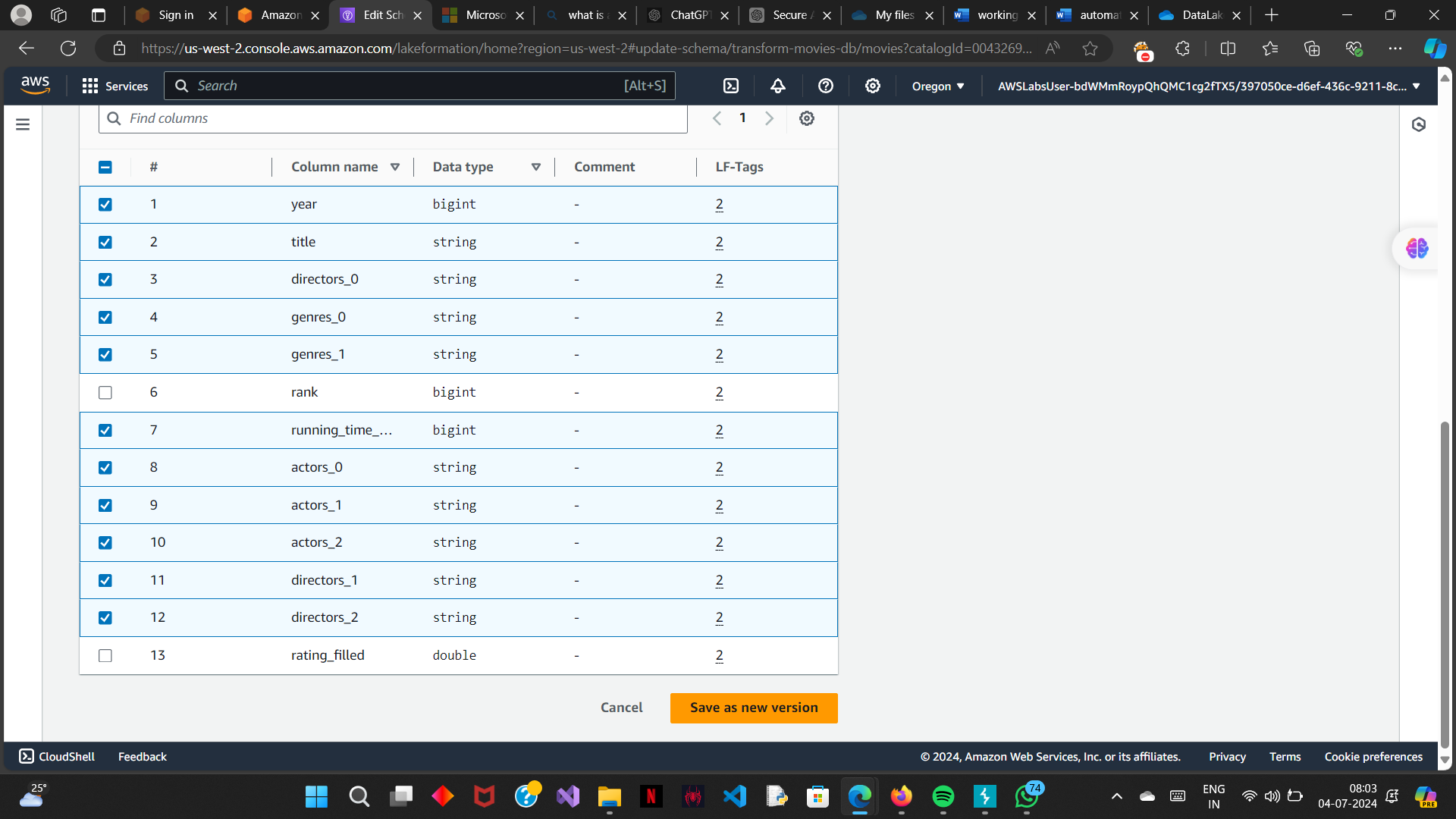


*TASK 3*.2 (Apply LF tags to a column):

AnyCompany wants only their enterprise customers to have access to the **rank** and **rating\_filled** columns in the **movies** table

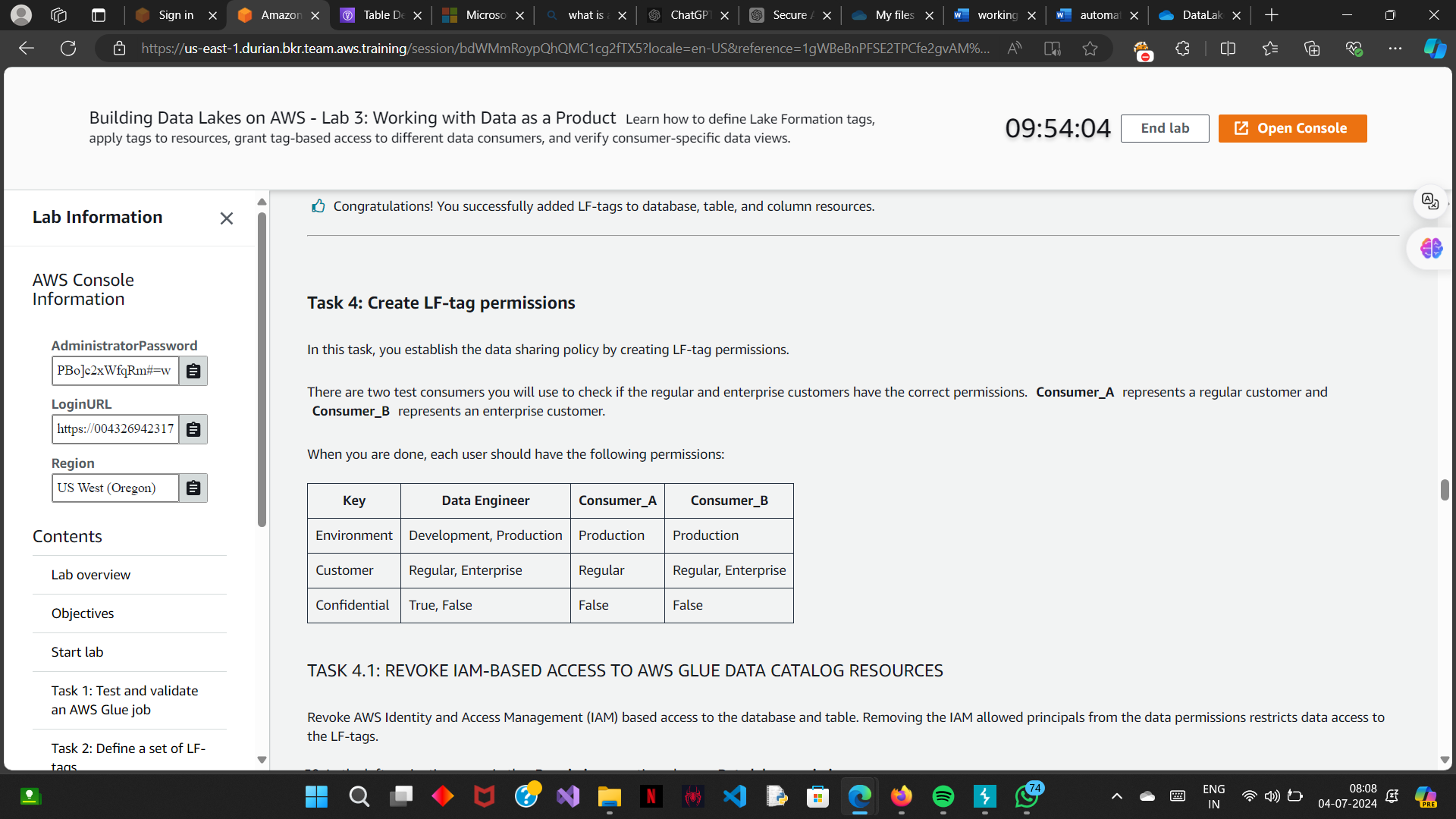
1> go to:

AWL lake formation--> tables --> <choose the table> --> actions --> edit scheme.



TASK 4(create LF tag permissions):

you establish the data sharing policy by creating LF-tag permissions.



TASK 4.1 (Revoke the IAM based access to AWS glue data catalogue resources):

Removing the IAM allowed principals from the data permissions restricts data access to the LF-tags.

Go to:

AWS lake formation --> data lake permissions --> <choose the IAM allowed princliple line> --> click on revoke.

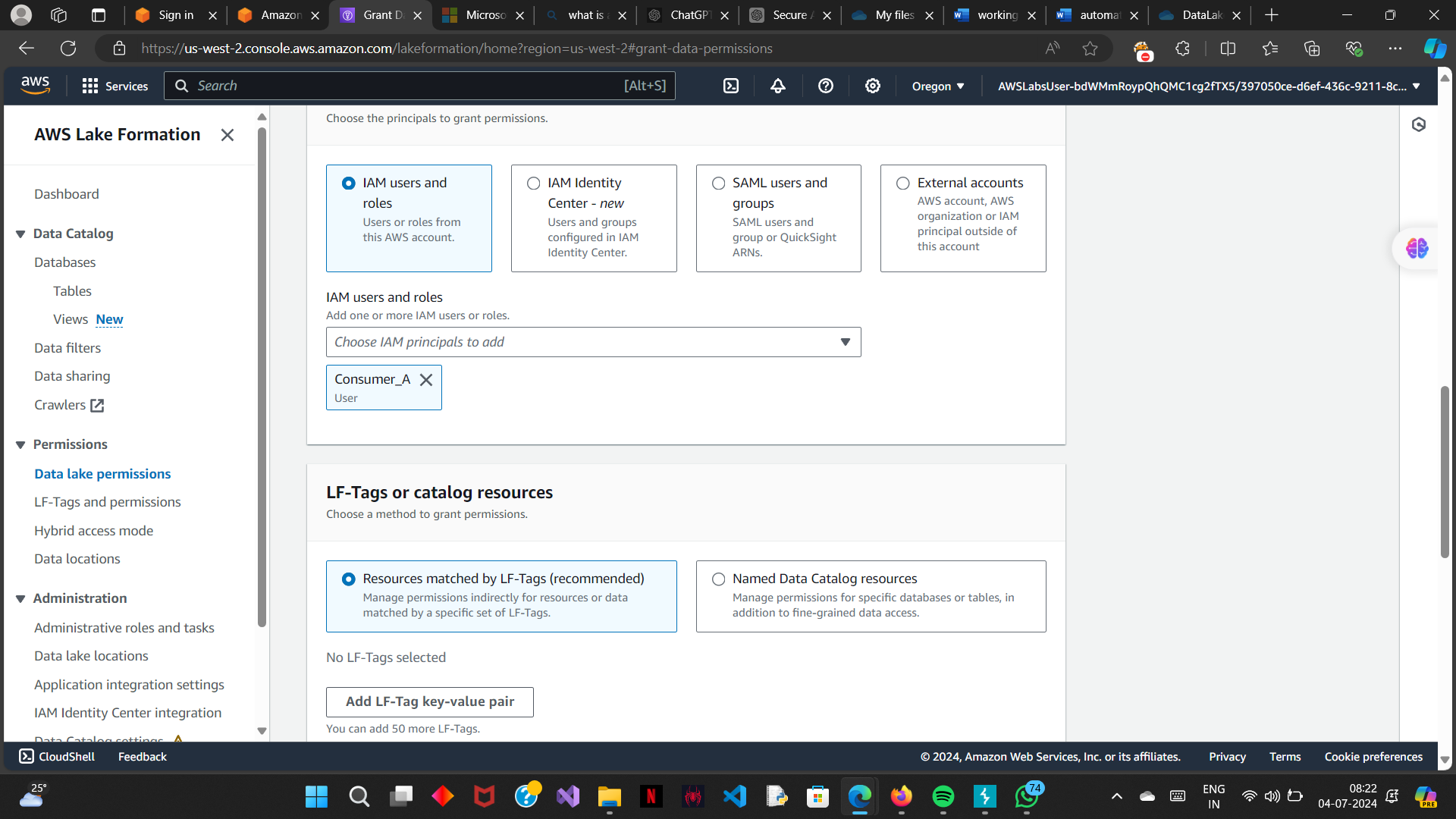
TASK 4.2(create LF tag permissions for a consumer)

1>Go to the above mentioned location and choose GRANT

2> choose consumer A and grant the permissions you want.

When you add two LF-tags in one permission, those tags act as an *AND*, not an *OR*. **Consumer\_A** only has access to tables where **Confidential=False** *AND* **Customer=Regular**. This means **Consumer\_A** is able to see 11 of the 13 columns in the movies table.

For consumer\_B it is : **Confidential=False AND Customer = (Regular OR Enterprise).**



TASK 5 (verify consumer specific data views):

Check if **Consumer\_A** can see the database and table but not the **rank** and **rating\_filled** columns. Then, check if **Consumer\_B** can see the database, table, and all columns.

1>Log in as consumer \_A

2> using athena , query the table.

3> check results. (below are the listed columns for the query)

# year title directors\_0 genres\_0 genres\_1 running\_time\_secs actors\_0 actors\_1 actors\_2 directors\_1 directors\_2

4> sign out

5> sign in as Consumer\_B and repeat the steps.

6> check the results (below are the columns listed for the query):

# year title directors\_0 genres\_0 genres\_1 rank running\_time\_secs actors\_0 actors\_1 actors\_2 directors\_1 directors\_2 rating\_filled