Lab overviewYou are a data engineer at AnyCompany. Your company is developing a new application that uses movie data. As part of the application development process, you were asked to set up permissions for two different types of customers: standard subscription and enterprise subscription. You set up these permissions using AWS Lake Formation tags (LF-tags) and then confirm if you grant the right access using Lake Formation tag-based access control (LF-TBAC) permissions.In this lab, you view an AWS Glue job that maintains a processed dataset. You also configure access to support data discovery using LF-tags and then set up custom access for two different consumers. You use Amazon Athena to access the curated data product.Objectives By the end of this lab, you will be able to do the following: • 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.

Environment overview The following diagram shows the basic architecture of the lab environment:In the preceding diagram, an Amazon Simple Storage Service (Amazon S3) bucket containing raw movie data is registered to Lake Formation to create a data lake. The data is loaded in AWS Glue and is processed using an AWS Glue job. The AWS Glue job adds the data to the database. The database is connected to the Amazon S3 bucket. 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.

**Lab Information**

AWS Console Information

* **AdministratorPassword**

Copy AdministratorPassword secret to clipboard

* **LoginURL**

Copy LoginURL secret to clipboard

* **Region**

Copy Region secret to clipboard

Contents

1. [Lab overview](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#lab-overview)
2. [Objectives](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#objectives)
3. [Start lab](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#start-lab)
4. [Task 1: Test and validate an AWS Glue job](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#task-1-test-and-validate-an-aws-glue-job)
5. [Task 2: Define a set of LF-tags](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#task-2-define-a-set-of-lf-tags)
6. [Task 3: Apply LF-tags to resources](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#task-3-apply-lf-tags-to-resources)
7. [Task 4: Create LF-tag permissions](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#task-4-create-lf-tag-permissions)
8. [Task 5: Verify consumer-specific data views](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#task-5-verify-consumer-specific-data-views)
9. [Conclusion](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#conclusion)
10. [End lab](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#end-lab)
11. [Additional resources](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#additional-resources)
12. [Appendix](https://us-east-1.durian.bkr.team.aws.training/session/urSUNrW5oXUSi2Pyx274zV?locale=en-US&reference=6ETN6pEWyuDo6sSTwwt94C%3A%3A5a08216e-df15-416b-9cbd-cda11c8d663c#appendix)

Task 1: Test and validate an AWS Glue jobYou have an AWS Glue job that updates and appends new movie data in a data lake when any new movies are added to the underlying raw data. The AnyCompany application developers want you to view the job and validate the processed data before including the processed data in their application.In this task, you view an AWS Glue job that maintains processed data in a data lake. You view the job definition in AWS Glue and then query the data using Athena.Task 1.1: Test an AWS Glue jobAn AWS Glue database, table, job, and trigger are already defined for the movies data. Test the AWS Glue job and review how the job transforms the data.3. At the top of the AWS Management Console, in the search bar, search for and choose AWS Glue.4. In the left navigation pane, in the Data Integration and ETL section, choose Jobs. 5. In the Your jobs section, choose transform-movies. The transform-movies AWS Glue job appears. Consider: Take a moment to read through the AWS Glue job script. How many nodes are contained in the job?Each node completes a task. In this lab, there are eight nodes in the AWS Glue job. The nodes complete the following tasks:

• Node 1: Inputs data from the data/movies\_csv/movies.csv file in the S3 bucket. • Node 2: Inputs data from the AWS Glue Data Catalog.• Node 3: Fills missing data in the rating column and creates a rating\_filled column using FillMissingValues from the awsglueml.transforms package.• Nodes 4 and 5: Applies mapping to the tables.• Node 6: Uses a custom PySpark transformation to append new rows to the movies data if there are any new rows added to the raw movies data.• Node 7: Selects the new rows dataframe from the collection returned by the custom transformation.• Node 8: Outputs data to the AWS Glue Data Catalog.

6. Choose the Schedules tab

This AWS Glue job is scheduled using an AWS Glue trigger called transform-movies-trigger. The job is scheduled to run every 10 minutes and is defined by a cron expression.i Learn more: Refer to AWS Glue triggers in the Additional resources section for more information about scheduling AWS Glue jobs.7. Choose 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.Refresh: If a job does not appear or you want to see the current Execution time, choose therefresh icon. 8. Wait until at least one AWS Glue job has a Run status of Succeeded. Note: Each transform-movies AWS Glue job takes 4-5 minutes to run. Note: If any of the AWS Glue jobs errors with a max concurrency error, view the full list of AWSGlue jobs until you find the job that succeeded. This AWS Glue job only runs one job at a time.You viewed the AWS Glue job and tested it by letting the trigger launch the AWS Glue job automatically.

Task 1.2: Validate an AWS Glue job Now that you viewed your AWS Glue job, query the processed movie data using Athena.9. At the top of the AWS Management Console, in the search bar, search for and choose Athena.Note: If the Athena menu appears, choose Launch query editor. 10. In the Workgroup primary settings window, choose Acknowledge.In the Data section, the AWSDataCatalog data source and transform-movies-db database are automatically selected. In the Tables section, you see one table listed. 11. In the query editor, enter the following: SELECT \* FROM "transform-movies-db"."movies" limit 10; 12. Choose Run.Ten records appear. There are 13 columns shown. As the engineer, you should see all the columns and records in the dataset. Note: If records do not appear, the AWS Glue job has not finished yet. Wait until the AWS Gluejob has finished and run the query again.

Next, count the number of rows in the dataset. 13. Choose the plus sign + to create a new query. 14. In the query editor, enter the following: SELECT COUNT(\*) AS number\_of\_movies FROM "movies"; 15. Choose Run. In the Results section, you see 4609 returned as the count for number\_of\_movies. Congratulations! You successfully tested and validated an AWS Glue jo

Task 2: Define a set of LF-tagsAnyCompany wants the data tagged with several different keys, including the following keys with their corresponding values: • Environment: Development, Production • Customer: Regular, Enterprise • Confidential: True, FalseYou will use these keys and values later in the lab to control access to the movie database, table, and columns.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 16. At the top of the AWS Management Console, in the search bar, search for and choose AWS Lake Formation.17. In the Welcome to Lake Formation popup window, make sure that Add myself is selected and then choose Get started.18. In the left navigation pane, in the Permissions section, choose LF-Tags. 19. Choose Add LF-tag.

20. For Key, enter Confidential. 21. For Values, enter True,False. 22. Choose Add. 23. Choose Add LF-tag. You added an LF-tag with a key of Confidential and values of True and False.Challenge A: Create more LF-tagsAdd two more LF-tags for Environment (Development, Production) and Customer (Regular, Enterprise). Hint: Repeat the steps in the previous task to add more LF-tags. Answer: Navigate here for a solution.

Congratulations! You successfully defined a set of LF-tags

Task 3: Apply LF-tags to resourcesIn 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 The movies database is ready for production. Add an LF-tag that provides access to the database for production consumers.24. In the left navigation pane, in the Data catalog section, choose Databases. 25. Choose transform-movies-db. 26. Choose Actions ▼. 27. Choose Edit LF-tags. 28. Choose Assign new LF-Tag. 29. For Assigned keys, choose Environment. 30. For Values, choose Production. 31. Choose Save. A tag appears in the LF-Tags section for the transform-movies-db (database) resource.

Challenge B: Apply LF-tags to a table

The movies table is ready for production and is not confidential. Add LF-tags that provide access to the database for production consumers who have access to data that is not confidential.Hint: The Environment tag of Production is inherited from the database, so you only need to add a Confidential tag of False to the table. Answer: Navigate here for a solution. Task 3.2: Apply LF-tags to a columnAnyCompany wants only their enterprise customers to have access to the rank and rating\_filled columns in the movies table. Edit the Customer LF-tag for the rank and rating\_filled columns to only include enterprise customers.32. In the left navigation pane, in the Data catalog section, choose Tables.

33. Choose movies. 34. Choose Actions ▼. 35. Choose Edit schema.36. Select the year, title, directors\_0, genres\_0, genres\_1, running\_time\_secs, actors\_0, actors\_1, actors\_2, directors\_1, and directors\_2 columns.37. Choose Edit tags. 38. Choose Assign new LF-Tag. 39. For Assigned keys, choose Customer. 40. For Values, choose Regular. 41. Choose Save. 42. Deselect all the columns. 43. Select the rank and rating\_filled columns. 44. Choose Edit tags. 45. Choose Assign new LF-Tag. 46. For Assigned keys, choose Customer. 47. For Values, choose Enterprise.

48. Choose Save. 49. Choose Save as new version. Only enterprise customers can view the rank and rating\_filled columns. Congratulations! You successfully added LF-tags to database, table, and column resources

Task 4: Create LF-tag permissions In this task, you establish the data sharing policy by creating LF-tag permissions.There are two test consumers you will use to check if the regular and enterprise customers have the correct permissions. Consumer\_A represents a regular customer and Consumer\_B represents an enterprise customer.When you are done, each user should have the following permissions: KeyData Engineer Regular, EnterpriseEnvironment Development, Production Production CustomerConfidential True, FalseRegular FalseConsumer\_A Consumer\_B ProductionRegular, Enterprise FalseTask 4.1: Revoke IAM-based access to AWS Glue Data Catalog resourcesRevoke AWS Identity and Access Management (IAM) based access to the database and table. Removing the IAM allowed principals from the data permissions restricts data access to the LF-tags.50. In the left navigation pane, in the Permissions section, choose Data lake permissions.

51. Choose one of the IAMAllowedPrincipals lines.

52. Choose Revoke. 53. Choose Revoke. 54. Choose the other IAMAllowedPrincipals line. 55. Choose Revoke. 56. Choose Revoke. You revoked the IAM permissions.Task 4.2: Create LF-tag permissions for a consumer Grant Consumer\_A access to production databases. 57. Choose Grant. 58. For IAM users and roles, select Consumer\_A.

59. Choose Add LF-Tag. 60. For Key, choose Environment. 61. For Values, choose Production. 62. In the Database permissions section, for Database permissions, choose Describe. 63. Choose Grant. Consumer\_A has access to production databases. Grant Consumer\_A access to data that is not confidential and is for regular customers. 64. Choose Grant. 65. For IAM users and roles, select Consumer\_A. 66. Choose Add LF-Tag. 67. For Key, choose Confidential. 68. For Values, choose False. 69. Choose Add LF-Tag. 70. For Key, choose Customer. 71. For Values, choose Regular. 72. In the Table permissions section, for Table permissions, choose Select and Describe.

73. Choose Grant. Consumer\_A has access to tables that are not confidential and are for regular customers.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.i Learn more: Refer to Lake Formation tag-based access control permissions model in the Additional resources section for more information about describe and associate permissions.Challenge C: Create LF-tag permissions for another consumerGrant describe permissions for Consumer\_B, an enterprise customer in the production environment who does not have access to confidential data.Hint: Set up Consumer\_B the same way as Consumer\_A, changing only the values chosen for the Customer key.

Hint: Choose the Regular and Enterprise values for the Customer key. Answer: Navigate here for a solution. Congratulations! You successfully created LF-tag permissions

Task 5: Verify consumer-specific data viewsThe movies table, as part of the Lake Formation data lake, is a product available to all users and roles that have been granted permissions. You created different access permissions for two sets of consumers, represented in this lab by users Consumer\_A and Consumer\_B. The purpose of this task is to verify that both consumers have access to the data product as targeted for their consumer type.In this task, you verify the consumer-specific data views for two test consumers using Athena. To test the LF-TBAC, 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.Task 5.1: Verify the data view for the first consumerVerify that Consumer\_A can see the database and table but not the rank and rating\_filled columns. 74. On the top-right of the screen, choose your username. 75. Choose Sign out. 76. Copy the LoginURL value listed to the left of these instructions. 77. Paste the LoginURL into your browser tab. The AWS login page appears. 78. For IAM user name, enter Consumer\_A.

79. For Password, copy and paste the AdministratorPassword listed to the left of these instructions.80. Choose Sign in. You are now signed in as the Consumer\_A user. Note: Make sure the AWS Region listed in the console matches the Region value listed to theleft of these instructions.81. At the top of the AWS Management Console, in the search bar, search for and choose Athena.Note: If the Athena menu appears, choose Launch query editor. 82. In the Workgroup primary settings window, choose Acknowledge. 83. Choose the plus sign + to create a new query. 84. In the query editor, enter the following:

SELECT \* FROM "transform-movies-db"."movies" limit 10; 85. Choose Run.Ten records appear. There are 11 columns shown. As Consumer\_A, a regular customer, you cannot see the rank and rating\_filled columns. You have successfully tested the tag-based permissions applied to Consumer\_A.Task 5.2: Verify the data view for second consumer Verify that Consumer\_B can see the database, table, and all columns. 86. On the top-right of the screen choose your username. 87. Choose Sign out. 88. Copy the LoginURL value listed to the left of these instructions. 89. Paste the LoginURL into your browser tab. The AWS login page appears. 90. For IAM user name, enter Consumer\_B.91. For Password, copy and paste the AdministratorPassword listed to the left of these instructions.92. Choose Sign in. You are now signed in as the Consumer\_B user.

Note: Make sure that the AWS Region listed in the console matches the Region value listed tothe left of these instructions.93. At the top of the AWS Management Console, in the search bar, search for and choose Athena.Note: If the Athena menu appears, choose Launch query editor. 94. In the Workgroup primary settings window, choose Acknowledge. 95. Choose the plus sign + to create a new query. 96. In the query editor, enter the following: SELECT \* FROM "transform-movies-db"."movies" limit 10; 97. Choose Run.Ten records appear. There are 13 columns shown. As Consumer\_B, an enterprise customer, you can see all of the columns, including rank and rating\_filled. You have successfully tested the tag-based permissions applied to Consumer\_B.

Congratulations! You successfully verified the consumer-specific data views for both consumers

Conclusion Congratulations! You now have successfully: • Viewed an AWS Glue job that maintains a dataset. • Defined LF-tags and applied them to resources. • Granted LF-TBAC permissions to data consumers. • Verified consumer-specific data views using Athena.End lab Follow these steps to close the console and end your lab. 98. Return to the AWS Management Console. 99. At the upper-right corner of the page, choose AWSLabsUser, and then choose Sign out. 100. Choose End lab and then confirm that you want to end your lab.

Additional resources • AWS Glue triggers • Assigning LF-Tags to Data Catalog resources • Overview of data filtering • Lake Formation tag-based access control permissions model

AppendixChallenge A solution Add an LF-tag for Environment (Development, Production). 101. Choose Add LF-tag. 102. For Key, enter Environment. 103. For Values, enter Development,Production. 104. Choose Add. 105. Choose Add LF-tag. Then, add an LF-tag for Customer (Regular, Enterprise).

106. Choose Add LF-tag. 107. For Key, enter Customer. 108. For Values, enter Regular,Enterprise. 109. Choose Add. 110. Choose Add LF-tag. You successfully created two more LF-tags. To continue this lab, go to Task 3.Challenge B solutionAdd LF-tags that provide access to production consumers who have access to data that is not confidential. 111. In the left navigation pane, in the Data catalog section, choose Tables. 112. Choose movies. 113. Choose Actions ▼. 114. Choose Edit LF-tags.The Environment tag of Production is inherited from the database, so you only need to add a Confidential tag of False. 115. Choose Assign new LF-Tag. 116. For Assigned keys, choose Confidential. 117. For Values, choose False. 118. Choose Save. Two tags appear in the LF-Tags section for the movies (table) resource. Consider: Take a moment to think about how you can use inheritance with LF-tags in your own data lakes to accurately tag databases, tables, and columns.i Learn more: Refer to Assigning LF-Tags to Data Catalog resources in the Additional resources section for more information about assigning LF-tags to resources.You successfully added LF-tags that provide access to production consumers who have access to data that is not confidential. To continue this lab, go to Task 3.2.

Challenge C solution Grant Consumer\_B access to production databases. 119. Choose Grant. 120. For IAM users and roles, select Consumer\_B. 121. Choose Add LF-Tag. 122. For Key, choose Environment. 123. For Values, choose Production. 124. In the Database permissions section, for Database permissions, choose Describe. 125. Choose Grant. Consumer\_B has access to production databases. Grant Consumer\_B access to data that is not confidential and is for enterprise customers. 126. Choose Grant. 127. For IAM users and roles, select Consumer\_B. 128. Choose Add LF-Tag. 129. For Key, choose Confidential. 130. For Values, choose False. 131. Choose Add LF-Tag. 132. For Key, choose Customer. 133. For Values, choose Enterprise and Regular. Note: A principal that is granted permissions on a LF-tag with multiple values can access AWSGlue Data Catalog resources with either of those values134. In the Table permissions section, for Table permissions, choose Select and Describe.135. Choose Grant.Consumer\_B has access to tables where Confidential=False AND (Customer=Enterprise OR Customer=Regular). This means Consumer\_B can see all the columns in the movies table.© 2023, Amazon Web Services

You successfully granted permissions for Consumer\_B. To continue this lab, go to Task 5.For more information about AWS Training and Certification, see https://aws.amazon.com/training/.Your feedback is welcome and appreciated. If you would like to share any feedback, suggestions, or corrections, please provide the details in our AWS Training and Certification Contact Form.