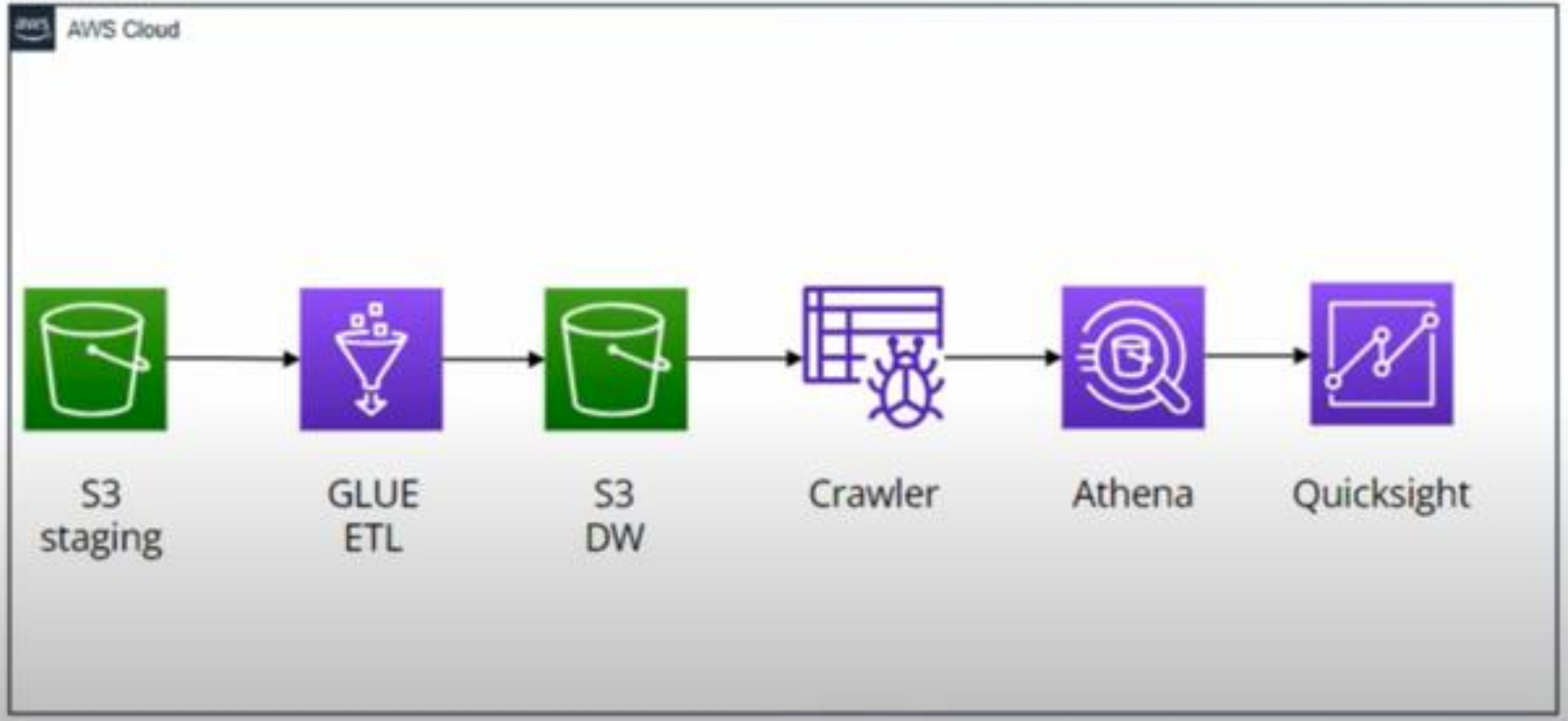
AWS(SERVICES)

PREPARED BY:
S M AALIM HUSSAIN (DS-037/2022)

- Spotify Dataset 2023 (kaggle.com)
- https://www.kaggle.com/datasets/tonygordonjr/spotify-dataset-2023

ARCHITECTURE



IAM USER

- An AWS Identity and Access Management (IAM) *user* is an entity that you create in AWS. The IAM user represents the human user or workload who uses the IAM user to interact with AWS. A user in AWS consists of a name and credentials.

aws Services Search [Alt+S] Stockholm Syed Muhammad Aalim Hussain

Console Home Info

Reset to default layout + Add widgets

Recently visited Info

IAM

Support

View all services

Applications (0) Info

Region: Europe (Stockholm)

Create application

eu-north-1 (Current Region) Find applications

1

Name	Description	Region	Originating a...

No applications

Get started by creating an application.

Create application

Go to myApplications

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Global | Syed Muhammad Aalim Hussain

Identity and Access Management (IAM)

Search IAM

- Dashboard
- Access management
  - User groups
  - Users
  - Roles
  - Policies
  - Identity providers
  - Account settings
- Access reports
  - Access Analyzer
    - External access
    - Unused access

IAM > Dashboard

IAM Dashboard

Security recommendations 1

Add MFA for root user

Add MFA for root user - Enable multi-factor authentication (MFA) for the root user to improve security for this account.

Add MFA

Root user has no active access keys

Using access keys attached to an IAM user instead of the root user improves security.

IAM resources

Resources in this AWS Account

User groups	Users	Roles	Policies	Identity providers
0	0	2	0	0

AWS Account

Account ID

590183725975

Account Alias

Create

Sign-in URL for IAM users in this account

https://590183725975.signin.aws.amazon.com/console

Quick Links

My security credentials

Manage your access keys, multi-factor

https://us-east-1.console.aws.amazon.com/iam/home?region=eu-north-1#/users

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

Step 1
Specify user details

Step 2
Set permissions

Step 3
Review and create

Step 4
Retrieve password

Specify user details

User details

User name

project

The user name can have up to 64 characters. Valid characters: A-Z, a-z, 0-9, and + = , . @ _ - (hyphen)

☑ Provide user access to the AWS Management Console - *optional*
If you're providing console access to a person, it's a best practice to manage their access in IAM Identity Center.

Are you providing console access to a person?

User type

○ Specify a user in Identity Center - Recommended
We recommend that you use Identity Center to provide console access to a person. With Identity Center, you can centrally manage user access to their AWS accounts and cloud applications.

◉ I want to create an IAM user
We recommend that you create IAM users only if you need to enable programmatic access through access keys, service-specific credentials for AWS CodeCommit or Amazon Keyspaces, or a backup credential for emergency account access.

© 2024, Amazon Web Services, Inc. or its affiliates.

Step 1
Specify user details

Step 2
Set permissions

Step 3
Review and create

Step 4
Retrieve password

Set permissions

Add user to an existing group or create a new one. Using groups is a best-practice way to manage user's permissions by job functions. Learn more

Permissions options

- Add user to group
  Add user to an existing group, or create a new group. We recommend using groups to manage user permissions by job function.
- Copy permissions
  Copy all group memberships, attached managed policies, and inline policies from an existing user.
- Attach policies directly
  Attach a managed policy directly to a user. As a best practice, we recommend attaching policies to a group instead. Then, add the user to the appropriate group.

Permissions policies (5/1171)

Choose one or more policies to attach to your new user.

Create policy

Filter by Type

quick | All types | 10 matches | 1

Policy name	Type	Attached entities
AWSQuicksightAthenaAccess	AWS managed	0

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Step 1
Specify user details

Step 2
Set permissions

Step 3
Review and create

Step 4
Retrieve password

Review and create

Review your choices. After you create the user, you can view and download the autogenerated password, if enabled.

User details

User name	Console password type	Require password reset
project	Custom password	Yes

Permissions summary

1

Name	Type	Used as
AmazonAthenaFullAccess	AWS managed	Permissions policy
AmazonS3FullAccess	AWS managed	Permissions policy
AWSGlueConsoleFullAccess	AWS managed	Permissions policy
AWSQuicksightAthenaAccess	AWS managed	Permissions policy

© 2024, Amazon Web Services, Inc. or its affiliates.

User created successfully

You can view and download the user's password and email instructions for signing in to the AWS Management Console.

View user

Step 1
Specify user details

Step 2
Set permissions

Step 3
Review and create

Step 4
Retrieve password

Retrieve password

You can view and download the user's password below or email users instructions for signing in to the AWS Management Console. This is the only time you can view and download this password.

Console sign-in details

Email sign-in instructions

Console sign-in URL
https://590183725975.signin.aws.amazon.com/console

User name
project

Console password
********** Show

Cancel | Download .csv file | Return to users list

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

- Console sign-in details
- Console sign-in URL
- https://590183725975.signin.aws.amazon.com/console
- User name
- project
- Console password
- datawarehouseA1..

You must change your password to continue

AWS account 590183725975

IAM user name project

Old password ••••••••••••••••

New password ••••••••••••••••

Retype new password ••••••••••••••••

Confirm password change

Sign in using root user email

English

Terms of Use Privacy Policy © 1996-2024, Amazon Web Services, Inc. or its affiliates.

CREATE S3 BUCKETS

- An Amazon S3 bucket is a public cloud storage resource available in Amazon Web Services (AWS) Simple Storage Service (S3) platform. It provides object-based storage, where data is stored inside S3 buckets in distinct units called objects instead of files.

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

- End User Computing
- Front-end Web & Mobile
- Game Development
- Internet of Things
- Machine Learning
- Management & Governance
- Media Services
- Migration & Transfer
- Networking & Content Delivery
- Quantum Technologies
- Robotics
- Satellite
- Security, Identity, & Compliance
- Storage

Storage

AWS Backup
AWS Backup centrally manages and automates backups across AWS services

EFS
Managed File Storage for EC2

AWS Elastic Disaster Recovery
Scalable, cost-effective application recovery to AWS

FSx
Fully managed third-party file systems optimized for a variety of workloads

S3
Scalable Storage in the Cloud

S3 Glacier
Archive Storage in the Cloud

Storage Gateway
Hybrid Storage Integration

t to default layout | + Add widgets

Create application

Find applications

1

Region | Originating a...

Applications

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Create bucket Info

Buckets are containers for data stored in S3. Learn more

General configuration

AWS Region

Europe (Stockholm) eu-north-1

Bucket type Info

- General purpose
  Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.
- Directory - *New*
  Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name Info

datawarehouse-project

Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

© 2024, Amazon Web Services, Inc. or its affiliates.

aws | Services | Search [Alt+S] | Global | project @ 5901-8372-5975

Successfully created bucket "datawarehouse-project"
To upload files and folders, or to configure additional bucket settings, choose **View details**.

View details

Amazon S3 > Buckets

Account snapshot

Storage lens provides visibility into storage usage and activity trends. Learn more

View Storage Lens dashboard

General purpose buckets | Directory buckets

General purpose buckets (1) Info

Buckets are containers for data stored in S3. Learn more

Copy ARN | Empty | Delete | Create bucket

Find buckets by name

1

Name	AWS Region	Access	Creation date
datawarehouse-project	Europe (Stockholm) eu-north-1	Bucket and objects not public	January 11, 2024, 05:42:39 (UTC+05:00)

© 2024, Amazon Web Services, Inc. or its affiliates.

Create folder Info

Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more

Your bucket policy might block folder creation

If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.

Folder

Folder name

stagging /

Folder names can't contain "/". See rules for naming

Server-side encryption Info

Server-side encryption protects data at rest.

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws Services Search [Alt+S] Global project @ 5901-8372-5975

Successfully created folder "stagging".

Amazon S3 > Buckets > datawarehouse-project

datawarehouse-project Info

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (1) Info

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Copy S3 URI | Copy URL | Download | Open | Delete | Actions | Create folder | Upload

Find objects by prefix

1

Name	Type	Last modified	Size	Storage class
stagging/	Folder	-	-	-

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Global | project @ 5901-8372-5975

Successfully created folder "datawarehouse".

Amazon S3 > Buckets > datawarehouse-project

datawarehouse-project Info

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (2) Info

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Copy S3 URI | Copy URL | Download | Open | Delete | Actions | Create folder | Upload

Find objects by prefix

1

Name	Type	Last modified	Size	Storage class
datawarehouse/	Folder	-	-	-
stagging/	Folder	-	-	-

CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. Learn more

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (3 Total, 88.6 MB)

Remove | Add files | Add folder

All files and folders in this table will be uploaded.

Find by name

1

Name	Folder	Type	Size
albums.csv	-	text/csv	76.3 MB
track.csv	-	text/csv	10.6 MB
artists.csv	-	text/csv	1.7 MB

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Upload succeeded
View details below.

Upload: status

Close

The information below will no longer be available after you navigate away from this page.

Summary

Destination	Succeeded	Failed
s3://datawarehouse-project/stagging/	3 files, 88.6 MB (100.00%)	0 files, 0 B (0%)

Files and folders | Configuration

Files and folders (3 Total, 88.6 MB)

© 2024, Amazon Web Services, Inc. or its affiliates.

aws | Services | Search [Alt+S] | Global | project @ 5901-8372-5975

stagging/

Copy S3 URI

Objects | Properties

Objects (3) Info

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Copy S3 URI | Copy URL | Download | Open | Delete | Actions | Create folder | Upload

Find objects by prefix

1

Name	Type	Last modified	Size	Storage class
albums.csv	csv	January 11, 2024, 05:47:37 (UTC+05:00)	76.3 MB	Standard
artists.csv	csv	January 11, 2024, 05:47:51 (UTC+05:00)	1.7 MB	Standard
track.csv	csv	January 11, 2024, 05:47:46 (UTC+05:00)	10.6 MB	Standard

© 2024, Amazon Web Services, Inc. or its affiliates.

AWS GLUE

- AWS Glue provides both visual and code-based interfaces to make data integration easier. Users can more easily find and access data using the AWS Glue Data Catalog. Data engineers and ETL (extract, transform, and load) developers can visually create, run, and monitor ETL workflows in a few steps in AWS Glue Studio.

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

AWS Glue

- Getting started
- ETL jobs
  - Visual ETL
  - Notebooks
  - Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- **Data Catalog**
- **Data Integration and ETL**
- **Legacy pages**

- What's New
- Documentation
- AWS Marketplace

Enable compact mode

Enable new navigation

We've added this new task-oriented landing page and a simpler left navigation. Let us know what you think.

Welcome to AWS Glue

Get started by setting up your account and users, cataloging your data, and building ETL jobs to prepare data for analytics.

Prepare your account for AWS Glue

Admins: Grant access to AWS Glue and **set a default IAM role.**

Set up roles and users

Catalog and search for datasets

View your databases & tables and catalog data using Crawlers.

Go to the Data Catalog

Move and transform data

Transform data using a visual, notebook, or code interface.

Author and edit ETL jobs

Resources and tutorials

Getting started with AWS Glue: Documentation | AWS Training

Video on working with AWS Glue Studio: Part 1 | Part 2 | Part 3

Using connectors and connections

Data integration and management

Monitor & debug ETL jobs and track usage

Go to job run monitoring

https://eu-north-1.console.aws.amazon.com/gluestudio/home?region=eu-north-1#/jobs?selectedInterface=source_and_target

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

Untitled job

Job has not been saved | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Data source - S3 bucket
artist

Data preview | Output schema

Start a data preview session

IAM role
To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

Create IAM role.

Additional Settings

Data source properties - S3

Name
artist

S3 source type Info

S3 location
Choose a file or folder in an S3 bucket.

Data Catalog table

S3 URL
s3://datawarehouse-project/ | View

Browse S3

Recursive
Read files in all subdirectories.

Data format

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

Untitled job

Job has not been saved | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

- S3 bucket

Data source - S3 bucket
album

Data preview | Output schema

Start a data preview session

IAM role
To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

Create IAM role.

Additional Settings

Data source properties - S3

Name
album

S3 source type | Info

S3 location
Choose a file or folder in an S3 bucket.

Data Catalog table

S3 URL
s3://datawarehouse-project/ | View

Browse S3

Recursive
Read files in all subdirectories.

Data format

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

Untitled job

Job has not been saved | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

S3 bucket

Data source - S3 bucket
tracks

Data preview | Output schema

Start a data preview session

IAM role
To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

Create IAM role.

Additional Settings

Data source properties - S3

Name
tracks

S3 source type Info
S3 location
Choose a file or folder in an S3 bucket.
Data Catalog table

S3 URL
s3://datawarehouse-project/
View

Browse S3

Recursive
Read files in all subdirectories.

Data format

CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

Untitled job

Job has not been saved | Try new UI | Actions | Save | Run

Visual 1 | Script | Job details | Runs | Data quality | Schedules | Version Control

S3 bucket

Transform - Join
Join Album & Artist

Data preview | Output schema

Start a data preview session

IAM role
To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

Create IAM role.

Additional Settings

Transform

Join Album & Artist

Node parents
Choose which nodes will provide inputs for this one.

Choose one or more parent node

album
S3 - DataSource

artist
S3 - DataSource

Join type
Select the type of join to perform.

Inner join
Select all rows from both datasets that meet the j...

Join conditions
Select a field from each parent node for the join condition.

album		artist
artist_id	=	id

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws Services Search [Alt+S] Stockholm project @ 5901-8372-5975

Untitled job

Job has not been saved Try new UI Actions Save Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Transform - Join

Join with Tracks

Data preview | Output schema

Start a data preview session

IAM role

To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

Create IAM role.

Additional Settings

Transform

Mapping node to rename them and avoid downstream issues.

Custom prefix

Add a prefix to the field names of the parent node on the right

right

Resolve it

Join type

Select the type of join to perform.

Inner join

Select all rows from both datasets that meet the j...

Join conditions

Select a field from each parent node for the join condition.

tracks		Join Album & Artist
track_id	=	track_id

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy Terms Cookie preferences

aws Services Search [Alt+S] Stockholm project @ 5901-8372-5975

Untitled job

Job has not been saved Try new UI Actions Save Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Transform - DropFields

Drop Fields

Data preview | Output schema

Start a data preview session

IAM role

To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

Create IAM role.

Additional Settings

Transform

☐	track_popularity	string
☐	track_name	string
☑	.track_id	string
☐	duration_ms	string
☐	album_name	string
☐	release_date	string
☐	label	string
☐	album_popularity	string
☐	album_id	string
☐	artist_id	string
☐	duration_sec	string
☑	id	string
☐	name	string

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

Untitled job

Job has not been saved | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Transform - Join
Join with Tracks

Transform - DropFields
Drop Fields

Data target - S3 bucket
Destination

Data target properties - S3

Format
Parquet

After you save your job, it will use Glue Studio's optimized Parquet writer.

Compression Type
Snappy

S3 Target Location
Choose an S3 location in the format s3://bucket/prefix/object/ with a trailing slash (/).

s3://datawarehouse-project/ | View

Browse S3

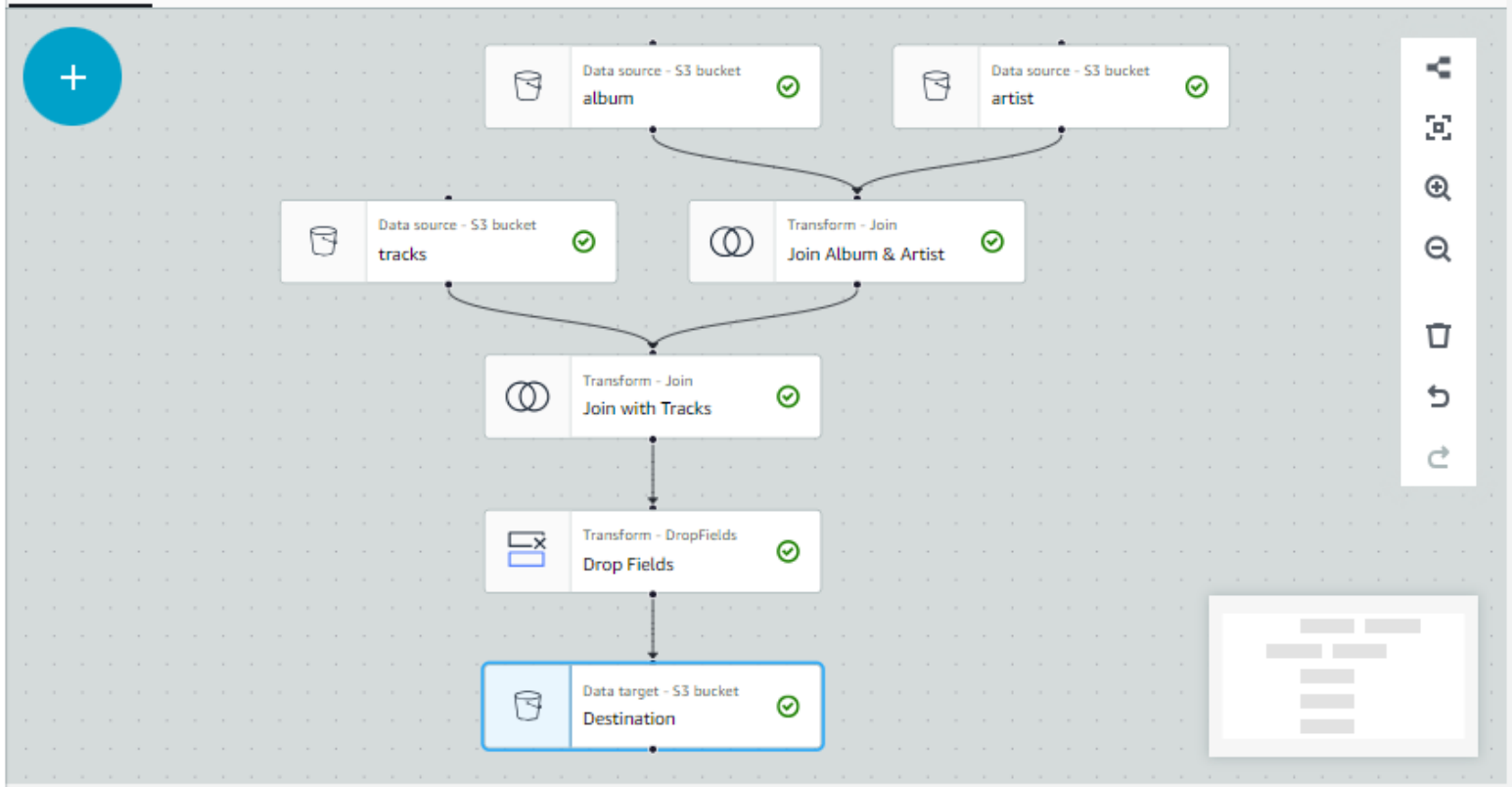
CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | project @ 5901-8372-5975

Spotify Datawarehouse Project

Job has not been saved | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control



Data target properties - S3

Format

Parquet

After you save your job, it will use Glue Studio's optimized Parquet writer.

Compression Type

Snappy

S3 Target Location

Choose an S3 location in the format s3://bucket/prefix/object/ with a trailing slash (/).

s3://datawarehouse-project/ | View

Browse S3

© 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

Spotify Datawarehouse Project

Job has not been saved | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Script (Locked) Info

Download script | Edit script

```
import sys
from awsglue.transforms import *
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job

args = getResolvedOptions(sys.argv, ["JOB_NAME"])
sc = SparkContext()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
job.init(args["JOB_NAME"], args)

# Script generated for node artist
artist_node1704934408435 = glueContext.create_dynamic_frame.from_options(
    format_options={"quoteChar": '"', "withHeader": True, "separator": ","},
    connection_type="s3",
    format="csv",
    connection_options={
```


© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Global | Syed Muhammad Aalim Hussain

Identity and Access Management (IAM)

Search IAM

- Dashboard
- **Access management**
  - User groups
  - Users
  - Roles
  - Policies
  - Identity providers
  - Account settings
- **Access reports**
  - Access Analyzer
    - External access
    - Unused access

IAM > Roles

Roles (2) Info

Delete | Create role

An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.

Search

1

Role name	Trusted entities	Last activity
AWSServiceRoleForSupport	AWS Service: support (Service-Linke	23 hours ag
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service	-

Roles Anywhere Info

Manage

Authenticate your non AWS workloads and securely provide access to AWS services.

Access AWS from your non AWS workloads

X.509 Standard

Temporary credentials

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

Web identity
Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.

SAML 2.0 federation
Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.

Custom trust policy
Create a custom trust policy to enable others to perform actions in this account.

Use case

Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

Service or use case

Glue

Choose a use case for the specified service.

Use case

Glue
Allows Glue to call AWS services on your behalf.

Cancel Next

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Step 1
Select trusted entity

Step 2
Add permissions

Step 3
Name, review, and create

Add permissions Info

Permissions policies (1/907) Info

Choose one or more policies to attach to your new role.

Filter by Type

s3 | All types | 9 matches | 1

Policy name	Type	Description
AmazonDMSRedshiftS3Role	AWS managed	Provides access to manage S3 settings...
AmazonS3FullAccess	AWS managed	Provides full access to all buckets via t...
AmazonS3ObjectLambda...	AWS managed	Provides AWS Lambda functions permi...
AmazonS3OutpostsFullA...	AWS managed	Provides full access to Amazon S3 on ...
AmazonS3OutpostsRead...	AWS managed	Provides read only access to Amazon S...
AmazonS3ReadOnlyAccess	AWS managed	Provides read only access to all bucket...
AWSBackupServiceRoleP...	AWS managed	Policy containing permissions necessar...
AWSBackupServiceRoleP...	AWS managed	Policy containing permissions necessar...

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Global | Syed Muhammad Aalim Hussain

Identity and Access Management (IAM)

Search IAM

- Dashboard
- **Access management**
  - User groups
  - Users
  - **Roles**
  - Policies
  - Identity providers
  - Account settings
- **Access reports**
  - Access Analyzer
    - External access
    - Unused access

Role glue_access_s3 created. View role

IAM > Roles

Roles (3) Info

Delete | Create role

An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.

Search

1

Role name	Trusted entities	Last activity
AWSServiceRoleForSupport	AWS Service: support (Service-Linke	23 hours ag
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service	-
glue_access_s3	AWS Service: glue	-

Roles Anywhere Info

Manage

Authenticate your non AWS workloads and securely provide access to AWS services.

CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

Successfully updated job
Successfully updated job Project. To run the job choose the **Run** Job button.

Project

Last modified on 1/11/2024, 7:41:03 AM | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Basic properties Info

Name

Project

Description - *optional*

Descriptions can be up to 2048 characters long.

IAM Role

Role assumed by the job with permission to access your data stores. Ensure that this role has permission to your Amazon S3 sources, targets, temporary directory, scripts, and any libraries used by the job.

glue_access_s3

Unsaved job found
We found an unsaved job, do you wish to restore it?
Restore

CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

Successfully started job
Successfully started job Project. Navigate to Run details for more details.

Project

Last modified on 1/11/2024, 7:41:03 AM | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Job runs (1/1) Info

Last updated (UTC) January 11, 2024 at 02:43:45 | View details | Stop job run | Table View | Card View

Filter job runs by property

Run status	Retries	Start time (UTC)	End time (UTC)	Duration	Capacity (DP...	Worker type	Glue vers
Running	0	2024/01/11 02:43:42	-	0 s	5 DPUs	G.1X	4.0

Run details | Input arguments (10) | Continuous logs | Run insights | Metrics | Spark UI | Stop job run

Job name	Start time (UTC)	Glue version	Last modified on (UTC)
Project	January 11, 2024 2:43:42 AM	4.0	January 11, 2024 2:43:42 AM

Id	End time (UTC)	Worker type
jr_fb18334f11b03f7789bf3bf3d8336dd3e0769		G.1X

Unsaved job found
We found an unsaved job, do you wish to restore it? Restore

CloudShell | Feedback | https://eu-north-1.console.aws.amazon.com/console/home?region=eu-north-1

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

Project

Last modified on 1/11/2024, 7:41:03 AM | Try new UI | Actions | Save | Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Job runs (1/1) Info

Last updated (UTC) January 11, 2024 at 02:46:16 | View details | Stop job run | Table View | Card View

Filter job runs by property

1

Run status	Retries	Start time (UTC)	End time (UTC)	Duration	Capacity (DP...	Worker type	Glue versio
Succeeded	0	2024/01/11 02:43:42	2024/01/11 02:45:31	1 m 39 s	5 DPUs	G.1X	4.0

Run details | Input arguments (10) | Continuous logs | Run insights | Metrics | Spark UI

Metrics Info

1h | 3h | 12h | 1d | 3d | 1w | Custom | UTC timezone

ETL data movement (bytes)

Data shuffle across executors (bytes)

Unsaved job found
We found an unsaved job, do you wish to restore it?
Restore

CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

Monitoring Info

Date range

7 Day

Job runs summary

Total runs	Running	Canceled	Success	Failed	Success rate	DPU hours
1	0	0	1	0	100%	1

Job runs (1) Info

Actions | View CloudWatch logs | View run details

Filter job runs by property

1

Job name	Run status	Type	Start time (UTC)	End time (UTC)	Run time	Capacity	Worker type	DPU hours
Project	Succeeded	Glue ETL	2024/01/11 02:43:42	2024/01/11 02:45:31	2 minutes	5	G.1X	0.14

Resource usage Info

View service quotas

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws Services Search [Alt+S] Global Syed Muhammad Aalim Hussain

Amazon S3

- Buckets
- Access Grants
- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- IAM Access Analyzer for S3

- Block Public Access settings for this account

- Storage Lens
  - Dashboards
  - Storage Lens groups
  - AWS Organizations settings

Objects (16) Info

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Copy S3 URI | Copy URL | Download | Open | Delete | Actions

Create folder | Upload

Find objects by prefix

1

Name	Type	Last modified	Size	Storage class
1704941104327-part-block-0-r-00000-snappy.parquet	parquet	January 11, 2024, 07:45:13 (UTC+05:00)	3.3 MB	Standard
run-1704941104327-part-block-0-r-00001-snappy.parquet	parquet	January 11, 2024, 07:45:11 (UTC+05:00)	3.3 MB	Standard
run-1704941104327-part-block-0-r-00002-	parquet	January 11, 2024, 07:45:14 (UTC+05:00)	3.3 MB	Standard

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy Terms Cookie preferences

AWS GLUE CRAWLER

- The CRAWLER creates the metadata that allows GLUE and services such as ATHENA to view the S3 information as a database with tables. That is, it allows you to create the Glue Catalog. This way you can see the information that s3 has as a database composed of several tables.

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (0) Info

Last updated (UTC)
January 11, 2024 at 02:52:37

Action Run Create crawler

View and manage all available crawlers.

Filter crawlers

1

Name	State	Schedule	Last run	Last run ti...	Log	Table chang...

No resources

No resources to display.

© 2024, Amazon Web Services, Inc. or its affiliates.

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Set crawler properties

Crawler details Info

Name

datawarehousing_

Name can be up to 255 characters long. Some character set including control characters are prohibited.

Description - *optional*

Enter a description

Descriptions can be up to 2048 characters long.

Tags - *optional*

Use tags to organize and identify your resources.

Cancel Next

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Add data source

Data source

Choose the source of data to be crawled.

S3

Network connection - *optional*

Optionally include a Network connection to use with this S3 target. Note that each crawler is limited to one Network connection so any other S3 targets will also use the same connection (or none, if left blank).

Clear selection | Add new connection

Location of S3 data

- In this account
- In a different account

S3 path

Browse for or enter an existing S3 path.

s3://datawarehousing-project/dataw | View | Browse S3

All folders and files contained in the S3 path are crawled. For example, type s3://MyBucket/MyFolder/ to crawl all objects in MyFolder within MyBucket.

Subsequent crawler runs

This field is a global field that affects all S3 data sources.

Crawl all sub-folders

Cancel | Add an S3 data source

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Choose data sources and classifiers

Data source configuration

Is your data already mapped to Glue tables?

- (●) Not yet
  Select one or more data sources to be crawled.
- ( ) Yes
  Select existing tables from your Glue Data Catalog.

Data sources (1) Info

Edit | Remove | Add a data source

The list of data sources to be scanned by the crawler.

	Type	Data source	Parameters
( )	S3	s3://datawarehousing-pro...	Recrawl all

Custom classifiers - *optional*

A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.

Cancel | Previous | Next

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

AWS Glue > Crawlers > Add crawler

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Configure security settings

IAM role Info

Existing IAM role

glue_access_s3 View

Create new IAM role Update chosen IAM role

Only IAM roles created by the AWS Glue console and have the prefix "AWSGlueServiceRole-" can be updated.

Lake Formation configuration - *optional*

Allow the crawler to use Lake Formation credentials for crawling the data source. Learn more.

☐ Use Lake Formation credentials for crawling S3 data source

Checking this box will allow the crawler to use Lake Formation credentials for crawling the data source. If the data source is registered in another account, you must provide the registered account ID. Otherwise, the crawler will crawl only those data sources associated to the account. Only applicable to S3, Glue Catalog, Iceberg, and Hudi data sources.

Security configuration - *optional*

Enable at-rest encryption with a security configuration.

© 2024, Amazon Web Services, Inc. or its affiliates.

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL

Create a database

Create a database in the AWS Glue Data Catalog.

Database details

Name

spotify

Database name is required, in lowercase characters, and no longer than 255 characters.

Location - *optional*

Set the URI location for use by clients of the Data Catalog.

Description - *optional*

Enter text

Descriptions can be up to 2048 characters long.

Cancel Create database

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL

Databases (1)

Last updated (UTC)
January 11, 2024 at 02:58:11

Edit Delete Add database

A database is a set of associated table definitions, organized into a logical group.

Filter databases

1

Name	Description	Location URI	Created on (UTC)
spotify	-	-	January 11, 2024 at 02:58:10

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

AWS Glue > Crawlers > Add crawler

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Set output and scheduling

Output configuration Info

Target database

spotify

Clear selection | Add database

Table name prefix - *optional*

Type a prefix added to table names

Maximum table threshold - *optional*

This field sets the maximum number of tables the crawler is allowed to generate. In the event that this number is surpassed, the crawl will fail with an error. If not set, the crawler will automatically generate the number of tables depending on the data schema.

Type a number greater than 0

Advanced options

Crawler schedule

You can define a time-based schedule for your crawlers and jobs in AWS Glue. The definition of these schedules uses the

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

AWS Glue > Crawlers > Add crawler

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Review and create

Step 1: Set crawler properties

Edit

Set crawler properties

Name	Description	Tags
datawarehousing_	-	-

Step 2: Choose data sources and classifiers

Edit

Data sources (1) Info

The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
S3	s3://datawarehousing-projec...	Recrawl all

Step 3: Configure security settings

Edit

Configure security settings

IAM role	Security configuration	Lake Formation configuration
glue-access-s3		

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation

One crawler successfully created
The following crawler is now created: "datawarehousing_"

AWS Glue > Crawlers > datawarehousing_

datawarehousing_

Last updated (UTC)
January 11, 2024 at 03:00:05

Run crawler | Edit | Delete

Crawler properties

Name	IAM role	Database	State
datawarehousing_	glue_access_s3	spotify	READY
Description	**Security configuration**	**Lake Formation configuration**	**Table prefix**
-	-	-	-
Maximum table threshold			
-			

Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (0)

Stop run | View CloudWatch logs | View run details

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

- Visual ETL
- Notebooks
- Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- **Data Catalog**
  - Databases
    - Tables
  - Stream schema registries
    - Schemas
  - Connections
  - **Crawlers**
    - Classifiers
  - Catalog settings
- **Data Integration and ETL**
- **Legacy pages**

What's New
Documentation

Crawler successfully starting
The following crawler is now starting: "datawarehousing_"

AWS Glue > Crawlers > datawarehousing_

datawarehousing_

Last updated (UTC)
January 11, 2024 at 03:00:05

Run crawler | Edit | Delete

Crawler properties

Name	IAM role	Database	State
datawarehousing_	glue_access_s3	spotify	READY
Description	**Security configuration**	**Lake Formation configuration**	**Table prefix**
-	-	-	-
Maximum table threshold			
-			

▶ Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (1)

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

- Visual ETL
- Notebooks
- Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- **Data Catalog**
  - Databases
    - Tables
  - Stream schema registries
    - Schemas
  - Connections
  - Crawlers
    - Classifiers
  - Catalog settings
- **Data Integration and ETL**
- **Legacy pages**

What's New
Documentation

Crawler properties

Name	IAM role	Database	State
datawarehousing_	glue_access_s3	spotify	READY
Description	**Security configuration**	**Lake Formation configuration**	**Table prefix**
-	-	-	-
Maximum table threshold			
-			

Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (1)

Stop run | View CloudWatch logs | View run details

The list of crawler runs for this crawler.

Filter data | Filter by a date and time range | 1

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Ta
January 11, 2024 at 03:01:12	-	54 s	Running	-	-

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

Role glue_access_full created. View role

Roles (5) Info

Delete Create role

An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.

Search

1

Role name	Trusted entities	Last activity
AWSServiceRoleForSupport	AWS Service: support (Service-Linke	Yesterday
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service	-
glue_access_full	AWS Service: glue	-
glue_access_s3	AWS Service: glue	1 hour ago
glue_access_s3_proj	AWS Service: glue	-

Roles Anywhere Info

Manage

Authenticate your non AWS workloads and securely provide access to AWS services.

Access AWS from your non AWS workloads

Operate your non AWS workloads using the same authentication and authorization strategy that you use within AWS.

X.509 Standard

Use your own existing PKI infrastructure or use AWS Certificate Manager Private Certificate Authority to authenticate identities.

Temporary credentials

Use temporary credentials with ease and benefit from the enhanced security they provide.

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

AWS Glue

- Getting started
- ETL jobs
  - Visual ETL
  - Notebooks
  - Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- **Data Catalog**
  - Databases
    - Tables
  - Stream schema registries
    - Schemas
  - Connections
  - Crawlers
    - Classifiers
  - Catalog settings
- **Data Integration and ETL**
- **Legacy pages**

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

Crawler successfully starting
The following crawler is now starting: "datawarehousing_"

AWS Glue > Crawlers > datawarehousing_

datawarehousing_

Last updated (UTC)
January 11, 2024 at 05:06:21

Run crawler | Edit | Delete

Crawler properties

Name	IAM role	Database	State
datawarehousing_	glue_access_full	spotify	READY
Description	**Security configuration**	**Lake Formation configuration**	**Table prefix**
-	-	-	-
Maximum table threshold			
-			

▶ Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (6)

The list of crawler runs for this crawler.

Stop run | View CloudWatch logs | View run details

Filter data | Filter by a date and time range | 1

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
January 11, 2024 at 05:06:02	-	01 min 46 s	Running	-	-
January 11, 2024 at 04:58:22	January 11, 2024 at 04:59:07	45 s	Failed	-	-
January 11, 2024 at 03:10:50	January 11, 2024 at 03:11:33	42 s	Failed	-	-
January 11, 2024 at 03:07:34	January 11, 2024 at 03:08:17	43 s	Failed	-	-
January 11, 2024 at 03:05:02	January 11, 2024 at 03:05:45	43 s	Failed	-	-
January 11, 2024 at 03:01:12	January 11, 2024 at 03:03:35	02 min 22 s	Failed	-	-

© 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

AWS Glue

- Getting started
- ETL jobs
  - Visual ETL
  - Notebooks
  - Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- **Data Catalog**
  - Databases
    - Tables
  - Stream schema registries
    - Schemas
  - Connections
  - Crawlers
    - Classifiers
  - Catalog settings
- **Data Integration and ETL**
- **Legacy pages**

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

Crawler successfully starting
The following crawler is now starting: "datawarehousing_"

AWS Glue > Crawlers > datawarehousing_

datawarehousing_

Last updated (UTC)
January 11, 2024 at 05:06:21

Run crawler | Edit | Delete

Crawler properties

Name	IAM role	Database	State
datawarehousing_	glue_access_full	spotify	READY
Description	**Security configuration**	**Lake Formation configuration**	**Table prefix**
-	-	-	-
Maximum table threshold			
-			

▶ Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (6)

The list of crawler runs for this crawler.

Stop run | View CloudWatch logs | View run details

Filter data | Filter by a date and time range | 1

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
January 11, 2024 at 05:06:02	January 11, 2024 at 05:06:58	56 s	Completed	-	1 table change, 0 partition changes
January 11, 2024 at 04:58:22	January 11, 2024 at 04:59:07	45 s	Failed	-	-
January 11, 2024 at 03:10:50	January 11, 2024 at 03:11:33	42 s	Failed	-	-
January 11, 2024 at 03:07:34	January 11, 2024 at 03:08:17	43 s	Failed	-	-
January 11, 2024 at 03:05:02	January 11, 2024 at 03:05:45	43 s	Failed	-	-
January 11, 2024 at 03:01:12	January 11, 2024 at 03:03:35	02 min 22 s	Failed	-	-

© 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

You can now create Apache Iceberg tables in the AWS Glue Data Catalog. To learn more, visit the documentation.

Create table

AWS Glue > Tables

Tables

A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

Tables (1)

Last updated (UTC)
January 11, 2024 at 05:55:19

Delete | Add tables using crawler | Add table

View and manage all available tables.

Filter tables

1

Name	Database	Location	Classification	Deprecated	View data	Data quality
datawarehouse	spotify	s3://datawarehousing-project/(	Parquet	-	Table data	View data quality

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

AWS Glue > Tables > datawarehouse

datawarehouse

Last updated (UTC)
January 11, 2024 at 05:06:58

Version 0 (Current version) Actions

Table overview | Data quality New

Table details | Advanced properties

Name	Description	Database	Classification
datawarehouse	-	spotify	Parquet
Location	**Connection**	**Deprecated**	**Last updated**
s3://datawarehousing-project/datawarehouse/	-	-	January 11, 2024 at 05:06:58

Input format	Output format	Serde serialization lib
org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat	org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat	org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe

Schema | Partitions | Indexes

Schema (15)

View and manage the table schema.

Edit schema as JSON Edit schema

Filter schemas

1

#	Column name	Data type	Partition key	Comment
1	followers	string	-	-
2	track_id	string	-	-
3	artist_popularity	string	-	-
4	artist_id	string	-	-
5	album_id	string	-	-
6	duration_ms	string	-	-

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

s3://datawarehousing-project/datawarehouse/ | - | - | January 11, 2024 at 05:06:58

Input format	Output format	Serde serialization lib
org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat	org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat	org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe

Schema | Partitions | Indexes

Schema (15)

View and manage the table schema.

Edit schema as JSON | Edit schema

Filter schemas

1

#	Column name	Data type	Partition key	Comment
1	followers	string	-	-
2	track_id	string	-	-
3	artist_popularity	string	-	-
4	artist_id	string	-	-
5	album_id	string	-	-
6	duration_ms	string	-	-
7	album_name	string	-	-
8	name	string	-	-
9	duration_sec	string	-	-
10	track_name	string	-	-
11	track_popularity	string	-	-
12	label	string	-	-
13	release_date	string	-	-
14	album_popularity	string	-	-
15	genre	string	-	-

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

AWS Glue > Tables > datawarehouse > Edit schema as JSON

Edit schema JSON: datawarehouse

Name	Version	Description
datawarehouse	0	-

```
[
  {
    "Name": "followers",
    "Type": "string"
  },
  {
    "Name": "track_id",
    "Type": "string"
  },
  {
    "Name": "artist_popularity",
    "Type": "string"
  },
  {
    "Name": "artist_id",
    "Type": "string"
  },
  {
    "Name": "album_id",
    "Type": "string"
  },
  {
    "Name": "duration_ms",
    "Type": "string"
```

JSON Ln 1, Col 1 Errors: 0 Warnings: 0

Choose file Copy

Cancel Save as new table version

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Edit schema: datawarehouse

Schema (15)

View and manage the table schema.

Delete Edit Add

Filter schemas

1

#	Column name	Data type	Partition key	Comment
1	followers	string	-	-
2	track_id	string	-	-
3	artist_popularity	string	-	-
4	artist_id	string	-	-
5	album_id	string	-	-
6	duration_ms	string	-	-
7	album_name	string	-	-
8	name	string	-	-
9	duration_sec	string	-	-
10	track_name	string	-	-
11	track_popularity	string	-	-
12	label	string	-	-
13	release_date	string	-	-
14	album_popularity	string	-	-
15	genre	string	-	-

Cancel Save as new table version

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

Crawler successfully starting
The following crawler is now starting: "datawarehousing_"

AWS Glue > Crawlers

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (1) Info

Last updated (UTC)
January 11, 2024 at 05:59:37

Action Run Create crawler

View and manage all available crawlers.

Filter crawlers

1

Name	State	Schedule	Last run	Last run timestamp	Log	Table changes from last ...
datawarehousing_	Ready		Succeeded	January 11, 2024 at 05:06...	View log	1 created

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

ATHENA

- Amazon Athena is an interactive query service that makes it simple to analyze data directly in Amazon S3 using standard SQL. Athena is server less, so there is no infrastructure to setup or manage, and you can choose to pay based on the queries you run or compute needed by your queries.

Analytics

Amazon Athena

Start querying data instantly.

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 and other federated data sources using standard SQL.

Begin querying your data

Discover the query editor and start querying right away.

Explore the query editor

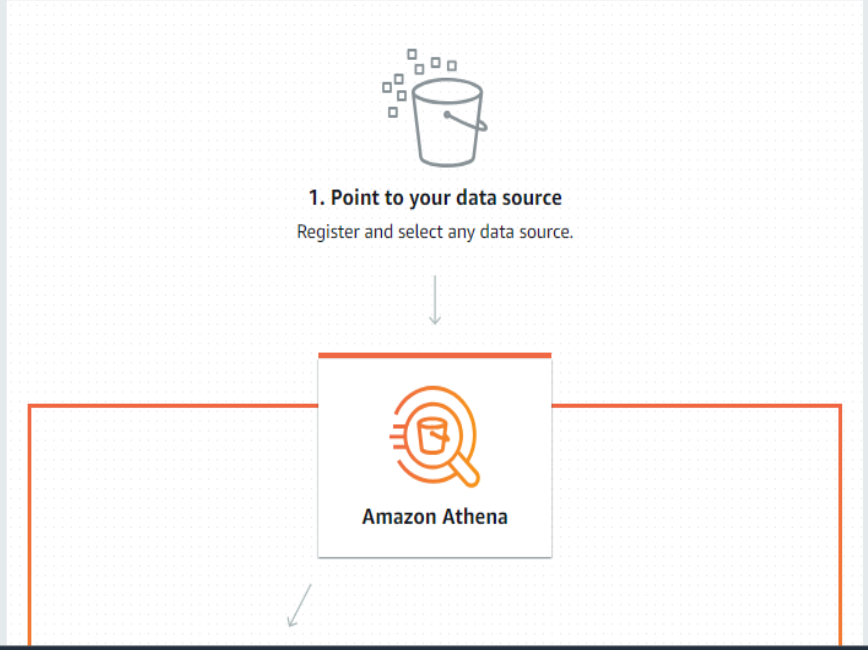
Pricing

Region: EU (Stockholm)

SQL queries per TB scanned	$5.00 per TB
SQL queries on Provisioned Capacity	$0.31 per DPU hour

Cost calculator

How it works



Getting started

Add the data source and identify the table and column names. Then you can begin querying your data right away.

Learn more

More resources

Documentation

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

Workgroup query engine upgrade complete
One or more workgroups have been upgraded to Athena engine version 3. To see the workgroups that have been upgraded, use the Workgroup list page. For information about new features, see the Athena Engine Version Reference.

Amazon Athena > Query editor

Editor | Recent queries | Saved queries | Settings

Workgroup: primary

Before you run your first query, you need to set up a query result location in Amazon S3. Edit settings

Athena now supports typeahead code suggestions to speed up SQL query development
Typeahead suggestions are turned on by default. You can change this setting in query editor preferences. Edit preferences

Data

Data source: AwsDataCatalog

Database: spotify

Tables and views | Create

Filter tables and views

Tables (1)

- datawarehouse
  - followers string
  - track_id string
  - artist_popularity string
  - artist_id string

Query 1

SQL Ln 1, Col 1

Run | Explain | Cancel | Clear | Create

Reuse query results
up to 60 minutes ago

Query results | Query stats

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Create bucket Info

Buckets are containers for data stored in S3. Learn more

General configuration

AWS Region

Europe (Stockholm) eu-north-1

Bucket type Info

- General purpose
  Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.
- Directory - *New*
  Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name Info

athena-output-spotify

Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Choose bucket

Format: s3://bucket/prefix

Object Ownership Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

- ACLs disabled (recommended)
  All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.
- ACLs enabled
  Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws Services Search [Alt+S] Global Syed Muhammad Aalim Hussain

Successfully created bucket "athena-output-spotify"
To upload files and folders, or to configure additional bucket settings, choose **View details**.

View details

Amazon S3 > Buckets

Account snapshot

Storage lens provides visibility into storage usage and activity trends. Learn more

View Storage Lens dashboard

General purpose buckets | Directory buckets

General purpose buckets (5) Info

Buckets are containers for data stored in S3. Learn more

Copy ARN | Empty | Delete | Create bucket

Find buckets by name

1

Name	AWS Region	Access	Creation date
athena-output-spotify	Europe (Stockholm) eu-north-1	Bucket and objects not public	January 11, 2024, 11:04:22 (UTC+05:00)
aws-glue-assets-590183725975-eu-north-1	Europe (Stockholm) eu-north-1	Bucket and objects not public	January 11, 2024, 07:41:05 (UTC+05:00)
datawarehouse-project	Europe (Stockholm) eu-north-1	Bucket and objects not public	January 11, 2024, 05:42:39 (UTC+05:00)
datawarehousing-project	Europe (Stockholm) eu-north-1	Bucket and objects not public	January 11, 2024, 07:32:25 (UTC+05:00)
spotify-project-datawarehouse	Asia Pacific (Mumbai) ap-south-1	Bucket and objects not public	January 11, 2024, 07:15:20 (UTC+05:00)

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Introducing cost-based optimizer
Cost-based optimizer uses table and column statistics from AWS Glue to improve your Athena query execution plans and performance. To get started, choose Generate statistics in the options for a table.

Amazon Athena > Query editor > Manage settings

Manage settings

Query result location and encryption

Location of query result - *optional*
Enter an S3 prefix in the current region where the query result will be saved as an object.

s3://athena-output-spotify | View | Browse S3

You can create and manage lifecycle rules for this bucket
Use Amazon S3 lifecycle rules to store your query results and metadata cost effectively or to delete them after a period of time.
Learn more

Lifecycle configuration

Expected bucket owner - *optional*
Specify the AWS account ID that you expect to be the owner of your query results output location bucket.

Enter AWS account ID

☐ Assign bucket owner full control over query results
Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

☐ Encrypt query results

Cancel Save

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Settings successfully updated.

0 0 1 1 0

Amazon Athena > Query editor

Editor | Recent queries | Saved queries | Settings

Workgroup primary

Query result and encryption settings

Manage

Query result location and encryption

Query result location	Encrypt query results	Expected bucket owner	Assign bucket owner full control over query results
s3://athena-output-spotify/	-	-	Turned off

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

Workgroup query engine upgrade complete
One or more workgroups have been upgraded to Athena engine version 3. To see the workgroups that have been upgraded, use the Workgroup list page. For information about new features, see the Athena Engine Version Reference.

0 0 1 1 0

Amazon Athena > Query editor

Editor | Recent queries | Saved queries | Settings

Workgroup: primary

Athena now supports typeahead code suggestions to speed up SQL query development
Typeahead suggestions are turned on by default. You can change this setting in query editor preferences.

Edit preferences

Data

Data source: AwsDataCatalog

Database: spotify

Tables and views | Create

Filter tables and views

Tables (1) 1

- datawarehouse
  - followers string
  - track_id string
  - artist_popularity string
  - artist_id string
  - album_id string
  - duration_ms string

Query 1

```
select * from datawarehouse limit 10
```

SQL Ln 1, Col 1

Run again | Explain | Cancel | Clear | Create

Reuse query results
up to 60 minutes ago

Query results | Query stats

Completed | Time in queue: 128 ms | Run time: 512 ms | Data scanned: 3.29 MB

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws Services Search [Alt+S] Stockholm Syed Muhammad Aalim Hussain

Tables (1) 1

- datawarehouse
  - followers string
  - track_id string
  - artist_popularity string
  - artist_id string
  - album_id string
  - duration_ms string
  - album_name string
  - name string
  - duration_sec string
  - track_name string
  - track_popularity string
  - label string
  - release_date string
  - album_popularity string

Views (0) 1

SQL Ln 1, Col 1

Run again | Explain | Cancel | Clear | Create

Reuse query results
up to 60 minutes ago

Query results | Query stats

Completed Time in queue: 128 ms Run time: 512 ms Data scanned: 3.29 MB

Results (10)

Copy | Download results

Search rows

1

#	followers	track_id	artist_popularity	artist_id	album_id	duration_ms	album_name
1	117	1zyBaGJEEixy5hSsaGGFsD	0	0ceNYzSeAMXGnuLXneG49h	2dLkCMPpWtieqtPzMUNx0t	192560	Prince Of Blue
2	79	6Rzxnu8yWwk8jjsKHjxVmG	6	6jC1TbQkVD16NiBazw5kKO	0l3a1XlsWOKwiPPaSTRzsX	21000	Toy Box Bop
3	301	5xBse5G3yBDdCiCKbw6Ci1	4	7EaaTl4FJ7bIAAW3jdzNIv	1qSS858EPX3oXH72oPcwVT	235920	5Th Quarter
4	9597639	2MAZwbhIikhYFrDkQ6yWiq	83	4iHNK0tOyZPYnBU7nGAgpQ	0v1DRRYBXYg1uVN1CIsyy0	304440	The Rarities
5	18656	0UAqpibXkggwQUFopRkK8Z	29	2YcJC1hnLjOIvNN3naeAlG	4npSDPvKBgqZOcLIxWO4Ji	212157	Enigma
6	5447095	7zdoKz94Bse1ZM41eQPLFz	71	3b8QkneNDz4JHKKKlLgYZg	0uqAyKT7oMZDFMyteJOJWI	198106	Here'S To The Good Times...This
7	59573	78XehnZc7G8UtKdLkDhrqF	53	3mGCLimuzOdod4Lnegny4d	2OdvUn7zi6YraS19YW3ZmF	223728	Lullaby Versions Of Taylor Swift
8	608638	6iRppEgR1kYag2ZAXViIbP	54	656RXuyw7CE0dtjdPgjJV6	6bfaFsYD6mFqdM6M5mUaRk	344706	Haydn: Stabat Mater, Symphonie
9	95859165	3yk94DUDanXyVhRLVjmI3w	100	06HL4z0CvFAxyc27GXpf02	6WzAiEDGTU7KmEyGwLpBXB	214706	Willow (The Witch Collection)
10	1686	1WzLZSvd3ZHifRlWoLWD1j	25	5aChaGbuNya16HNAqdgwfR	241guzPerNxjWcu0cAB6jV	296306	Sbi Karaoke Superstars - Deep P

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Services | Search [Alt+S] | Stockholm | Syed Muhammad Aalim Hussain

Workgroup query engine upgrade complete
One or more workgroups have been upgraded to Athena engine version 3. To see the workgroups that have been upgraded, use the Workgroup list page. For information about new features, see the Athena Engine Version Reference.

Amazon Athena > Query editor

Editor | Recent queries | Saved queries | Settings

Workgroup: primary

Athena now supports typeahead code suggestions to speed up SQL query development
Typeahead suggestions are turned on by default. You can change this setting in query editor preferences.

Edit preferences

Data

Data source: AwsDataCatalog

Database: spotify

Tables and views | Create

Filter tables and views

Tables (1)

- datawarehouse
  - followers string
  - track_id string
  - artist_popularity string
  - artist_id string
  - album_id string
  - duration_ms string

Query 1

```
select name,track_id from datawarehouse limit 10
```

SQL Ln 1, Col 21

Run again | Explain | Cancel | Clear | Create

Reuse query results
up to 60 minutes ago

Query results | Query stats

Completed — Time in queue: 72 ms — Run time: 473 ms — Data scanned: 1.61 MB

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

aws Services Search [Alt+S] Stockholm Syed Muhammad Aalim Hussain

Filter tables and views

Tables (1) 1

- datawarehouse
  - followers string
  - track_id string
  - artist_popularity string
  - artist_id string
  - album_id string
  - duration_ms string
  - album_name string
  - name string
  - duration_sec string
  - track_name string
  - track_popularity string
  - label string
  - release_date string
  - album_popularity string

Views (0) 1

SQL Ln 1, Col 21

Run again | Explain | Cancel | Clear | Create

Reuse query results
up to 60 minutes ago

Query results | Query stats

Completed Time in queue: 72 ms Run time: 473 ms Data scanned: 1.61 MB

Results (10)

Copy | Download results

Search rows

1

#	name	track_id
1	Future	5syD6f38Jhckk5g9DgoCta
2	red go-cart	7tsRQ15JtGulaCO5Emk0z1
3	Raphael Saadiq	0Lp9zP5kalhwnGyzyMJie4
4	Sabrina Carpenter	6dgUya35uo964z7GZXM07g
5	Luciano	56I2y9GyTYmCZo6vrH1PPS
6	Johann Sebastian Bach	67AaJQRJaraBeBeUa7sAXy
7	Various Artists	4rkWdZtTvWWOPbUAEl7Ahd
8	Various Artists	3Du6vp5sLopdsCGaCkR4bQ
9	Behzad Ranjbaran	6Vq4yZ2oEf2z3icVZXxvM2
10	Gustav Mahler	66jo1ye2aPa1zHdTBy8Uy2

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Amazon S3 > Buckets > athena-output-spotify > Unsaved/ > 2024/ > 01/ > 11/

11/

Copy S3 URI

Objects | Properties

Objects (4) Info

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Copy S3 URI | Copy URL | Download | Open | Delete | Actions | Create folder | Upload

Find objects by prefix

1

Name	Type	Last modified	Size	Storage class
186168fb-49c9-4bcb-8609-839b1db522dc.csv	csv	January 11, 2024, 11:10:30 (UTC+05:00)	433.0 B	Standard
186168fb-49c9-4bcb-8609-839b1db522dc.csv.metadata	metadata	January 11, 2024, 11:10:30 (UTC+05:00)	119.0 B	Standard
f38270a3-3735-417d-9781-179f6c0d24c2.csv	csv	January 11, 2024, 11:06:33 (UTC+05:00)	2.6 KB	Standard
f38270a3-3735-417d-9781-179f6c0d24c2.csv.metadata	metadata	January 11, 2024, 11:06:33 (UTC+05:00)	828.0 B	Standard

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

QUICKSIGHT

- QuickSight is a fast, easy-to-use, cloud-powered business analytics service that makes it easier for all employees within an organization to build visualizations, perform ad-hoc analysis, and quickly get business insights from their data, anytime, on any device.

QuickSight

Your AWS Account is not signed up for QuickSight. Would you like to sign up now?

AWS Account **590183725975**

Sign up for QuickSight

To access QuickSight with a different account, log in again.

aws

QuickSight

Create your QuickSight account

Enterprise edition offers the Paginated Reports add-on

When you sign up for the Enterprise edition, you have the option to add Paginated Reports to your subscription. Reporting enables your business users to generate paginated documents to be sent to recipients in a number of formats.

Learn more

Edition	Enterprise	Enterprise + Q Learn more
Team trial for 30 days (4 authors)*	FREE	FREE
Author per month (yearly)**	$18	$28
Author per month (monthly)**	$24	$34
Readers (pay-per-Session)	$0.30 / session (max $5)****	$0.30 / session (max $10)****
Additional SPICE per month	$0.38 per GB	$0.38 per GB
QuickSight Q regional fee	N/A	$250 / mo / region
Natural language query with QuickSight Q	N/A	INCLUDED
Single Sign On with SAML or OpenID Connect	✓	✓
Connect to spreadsheets, databases & business apps	✓	✓
Access data in Private VPCs	✓	✓
Row-level security for dashboards	✓	✓
Secure data encryption at rest	✓	✓
Connect to your Active Directory	✓	✓
Use Active Directory groups***	✓	✓
Send email reports	✓	✓
Embed QuickSight	✓	✓
Capacity-based pricing	✓	✓
Supported regions	Learn more	Learn more

* Trial authors are auto-converted to month-to-month subscription upon trial expiry
** Each additional author includes 10GB of SPICE capacity
*** Active Directory groups are available in accounts connected to Active Directory
**** Sessions of 30-minute duration. Total charges for each reader are capped at $5 per month. Conditions apply.
Sign up for Standard Edition here.

Continue

Create your QuickSight account

Enterprise

Back

Get Paginated Report add-on

Create, schedule, and share operational reports and data exports from a single fully-managed business intelligence (BI) cloud solution.

Generate multi-page reports

- PDF reports will render all the rows of a table or pivot table.
- Add headers, footers, and page breaks to the PDF report.
- Add system parameters such as page numbers and timestamps.
- Add text and images inside of the header and footer sections.

Automate delivery with schedules

- Send regular email reports with attachments and download links.
- Schedule data exports at scale in CSV format

Get started fast

- Build all insights from a single source of truth dataset.
- QuickSight offers one authoring experience for building and accessing dashboards and reports.
- No additional training or experience needed.

Pricing*

500 unique report units / month , $500 / monthly**

No, Maybe Later | Yes

Monthly charges begin immediately

*First month charges and usage are prorated. Annual plan is available after sign up. Visit the "Your Subscriptions" page to switch.

**A unique report unit is defined to be up to 100 pages long (PDF) or 100MB in size (CSV/Excel). For example, a 200 page report constitutes 2 unique

QuickSight

Create your QuickSight account

Enterprise

Back

Authentication method

- (selected) Use IAM federated identities & QuickSight-managed users
  Authenticate with single sign-on (SAML or OpenID Connect), AWS IAM credentials, or QuickSight credentials
- Use IAM federated identities only
  Authenticate with single sign-on (SAML or OpenID Connect) or AWS IAM credentials
- Use Active Directory
  Authenticate with Active Directory credentials
- Use AWS IAM Identity Center
  Authenticate using AWS IAM Identity center

QuickSight region

Select a region

US East (N. Virginia)

Account info

QuickSight account name
You will need this for you and others to sign in

datawarehousing-project

Notification email address
For QuickSight to send important notifications

syedmuhammadaalimhussain@gmail.com

QuickSight access to AWS services

Make your existing AWS data and users available in QuickSight. Learn more

IAM Role

IAM Role

- (●) Use QuickSight-managed role (default)
- ( ) Use an existing role

Allow access and autodiscovery for these resources

- [ ] Amazon Redshift
- [ ] Amazon RDS
- [x] IAM
- [ ] Amazon S3
  Select S3 buckets
- [x] Amazon Athena
  Make sure you've chosen the right Amazon S3 buckets for QuickSight access
- [ ] Amazon S3 Storage Analytics
- [ ] AWS IoT Analytics
- [ ] Amazon OpenSearch Service
- [ ] AWS DynamoDB
  Select tables
- [ ] Amazon SageMaker
- [ ] Amazon Timestream
- [ ] AWS SecretsManager
  Select secrets

SPICE auto-purchase is based on usage

Your QuickSight account will auto purchase SPICE capacity when it exceeds the free bundled amount. Go to your account settings if you'd like to turn this off.
Learn more

Finish

aws

QuickSight

Congratulations! You are signed up for Amazon QuickSight!

Access QuickSight with the following information
Account name: datawarehousing-project

Go to Amazon QuickSight

QuickSight

Find analyses & more

- Favorites
- Recent
- My folders
- Shared folders
- Dashboards
- Analyses
- Datasets
- Topics New
- Community New

Datasets

New dataset

Name		Owner	Last Modified
Web and Social Media Analytics	SPICE	Me	a minute ago
Business Review	SPICE	Me	a minute ago
People Overview	SPICE	Me	a minute ago
Sales Pipeline	SPICE	Me	a minute ago

QuickSight access to AWS serv

Make your existing AWS data and users available in QuickSight. Learn mor

IAM Role

- (●) Use QuickSight-managed role (default)
- ( ) Use an existing role

Allow access and autodiscovery for these resources

- [ ] Amazon Redshift
- [ ] Amazon RDS
- [x] IAM
- [ ] Amazon S3
  Select S3 buckets
- [x] Amazon Athena
  Make sure you've chosen the right Amazon S3 buckets for QuickSight access
- [ ] Amazon S3 Storage Analytics
- [ ] AWS IoT Analytics
- [ ] Amazon OpenSearch Service
- [ ] AWS DynamoDB
  Select tables
- [ ] Amazon SageMaker
- [ ] Amazon Timestream
- [ ] AWS SecretsManager
  Select secrets

Save | Cancel

Select Amazon S3 buckets

S3 Buckets Linked To QuickSight Account | S3 Buckets You Can Access Across AWS

Select the buckets that you want QuickSight to be able to access.

Selected buckets have read only permissions by default. However, you must give write permissions for Athena Workgroup feature.

[■] Select all

	S3 Bucket	Write permission for Athena Workgroup
[x]	athena-output-spotify	[x]
[ ]	aws-athena-query-results-us-east-1-590183725975	
[ ]	aws-glue-assets-590183725975-eu-north-1	
[ ]	datawarehouse-project	
[ ]	datawarehousing-project	
[ ]	spotify-project-datawarehouse	

Cancel | Finish

QuickSight

Datasets

SPICE capacity for this region: 0 bytes of 0 bytes

Create a Dataset

FROM NEW DATA SOURCES

- Upload a file (.csv, .tsv, .clf, .elf, .xlsx, .json)
- Salesforce — Connect to Salesforce
- S3 Analytics
- S3
- Athena
- RDS
- Redshift — Auto-discovered
- Redshift — Manual connect
- MySQL
- PostgreSQL
- ORACLE
- SQL Server
- Aurora
- MariaDB
- Presto
- Spark
- Teradata — Provided by Teradata
- Snowflake
- Google BigQuery
- Amazon OpenSearch Ser… — Successor to Amazon Elasticsearch Ser…
- Exasol
- Databricks
- Trino
- Starburst
- GitHub
- Jira
- ServiceNow
- Adobe Analytics

QuickSight

Datasets

SPICE capacity for this region: 0 bytes of 0 bytes

Create a Dataset

FROM NEW DATA SOURCES

- Upload a file (.csv, .tsv, .clf, .elf, .xlsx, .json)
- Salesforce — Connect to Salesforce
- Athena
- RDS
- Redshift — Auto-discovered
- Redshift — Manual connect
- MySQL
- PostgreSQL
- ORACLE
- SQL Server
- Aurora
- MariaDB
- Presto
- Spark
- Teradata — Provided by Teradata
- Snowflake
- Google BigQuery
- Amazon OpenSearch Ser… — Successor to Amazon Elasticsearch Ser…
- Exasol
- Databricks
- Trino
- Starburst
- GitHub
- Jira
- ServiceNow
- Adobe Analytics

New Athena data source

Data source name

spotify

Athena workgroup

[primary]

Validate connection SSL is enabled Create data source

QuickSight

Datasets

SPICE capacity for this region: 0 bytes of 0 bytes

Choose your table

spotify

Catalog: contain sets of databases.

AwsDataCatalog

Database: contain sets of tables.

spotify

Tables: contain the data you can visualize.

- datawarehouse

Edit/Preview data | Use custom SQL | Select

Create a Dataset

FROM NEW DATA SOURCES

- Upload a file (.csv, .tsv, .clf, .elf, .xlsx, .json)
- Salesforce — Connect to Salesforce
- Athena
- RDS
- Redshift — Auto-discovered
- PostgreSQL
- ORACLE
- SQL Server
- Presto
- Spark
- Teradata — Provided by Teradata
- Snowflake
- Google BigQuery
- Amazon OpenSearch Ser... — Successor to Amazon Elasticsearch Ser...
- Exasol
- Databricks
- Trino
- Starburst
- GitHub
- Jira
- ServiceNow
- Adobe Analytics

FROM EXISTING DATA SOURCES

QuickSight

Datasets

SPICE capacity for this region: 0 bytes of 0 bytes

Finish dataset creation

Table: datawarehouse
Data source: spotify
Schema: spotify

- (●) Import to SPICE for quicker analytics — ✓ 0 bytes available SPICE
- ( ) Directly query your data

[✓] Email owners when a refresh fails

Edit/Preview data | Augment with SageMaker | Visualize

Create a Dataset

FROM NEW DATA SOURCES

- Upload a file (.csv, .tsv, .clf, .elf, .xlsx, .json)
- Salesforce — Connect to Salesforce
- Athena
- RDS
- Redshift — Auto-discovered
- Manual connect
- PostgreSQL
- ORACLE
- SQL Server
- Aurora
- MariaDB
- Presto
- Spark
- Teradata — Provided by Teradata
- Snowflake
- Google BigQuery
- Amazon OpenSearch Ser… — Successor to Amazon Elasticsearch Ser…
- Exasol
- Databricks
- Trino
- Starburst
- GitHub
- Jira
- ServiceNow
- Adobe Analytics

FROM EXISTING DATA SOURCES

QuickSight | datawarehouse analysis

File Edit Data Insert Sheets Objects Search

ACTUAL SIZE PUBLISH

ADD:

NEW LOOK

Data

Dataset

SPICE datawarehouse

Search fields

+ CALCULATED FIELD

- album_id
- album_name
- album_popularity
- artist_id
- artist_popularity
- duration_ms
- duration_sec
- followers
- genre
- label
- name
- release_date
- track_id
- track_name
- track_popularity

Visuals

+ ADD

CHANGE VISUAL TYPE

ADD DATA

Add a dimension or measure

Sheet 1

AutoGraph

Add 1 or more fields to build a visual.

THANK YOU!