

Introduction:

The movie data analysis pipeline represents a comprehensive and systematic approach to extracting valuable insights from the vast and diverse landscape of the film industry. In an era where digital technologies have revolutionized the way movies are produced, distributed, and consumed, understanding the intricacies of this dynamic ecosystem has become paramount. The movie data analysis pipeline serves as a strategic framework that harnesses the power of data engineering and analytics to unveil patterns, trends, and correlations within the multifaceted realm of cinematic information.

Objectives:

- To create a data pipeline for imdb movie data and gather insights from the data.
- Use AWS platform services to implement the project
- Implementing Data Quality Checks
- Perform analysis via Amazon Redshift service.

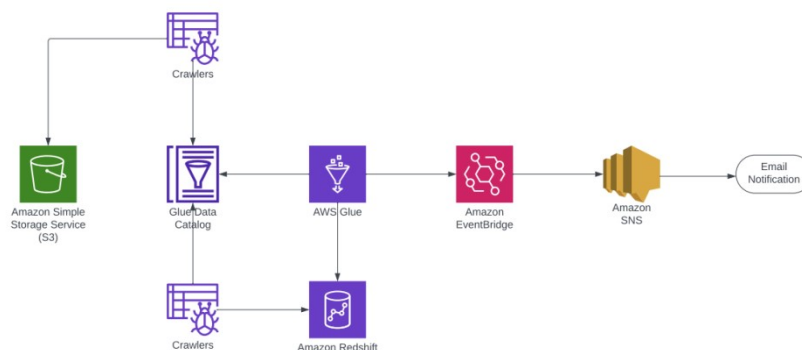
Technology and Services used:

1. **Amazon S3 as Datalake** - Amazon Simple Storage Service (S3) is a scalable and durable object storage service, often utilized as a data lake in AWS. Acting as a central repository, S3 allows organizations to store and retrieve any amount of data, making it a robust solution for storing diverse datasets associated with a data lake architecture. With features like versioning and access controls, S3 provides a secure and scalable foundation for managing large volumes of structured and unstructured data, serving as a crucial component for data storage and retrieval in various analytical and machine learning workflows.
2. **AWS Glue Catalog** - AWS Glue Catalog is a metadata repository that stores structural and operational metadata about data assets. It acts as a centralized location for managing metadata related to databases, tables, and connections. By using Glue Catalog, organizations can discover and understand the structure of their data, facilitating seamless data integration and governance across different AWS services.
3. **Glue Crawlers** - AWS Glue Crawlers are automated tools that analyze data stored in various repositories, infer the schema, and populate the AWS Glue Data Catalog with metadata. These crawlers enable dynamic and efficient discovery of data, automating the process of cataloging diverse datasets. By automating metadata extraction, Glue Crawlers simplify the integration of new data sources into the Glue Data Catalog, ensuring that the metadata remains up-to-date and accurate.
4. **Glue Catalog Data Quality** - AWS Glue Catalog Data Quality is a feature that allows users to define and run data quality checks on their datasets stored in the Glue Data Catalog. It helps ensure the accuracy, completeness, and consistency of the data by defining custom rules and

metrics. By incorporating data quality checks into the data pipeline, organizations can enhance the reliability of their analytics and decision-making processes.

5. **Glue Low Code ETL** - AWS Glue provides a low-code environment for Extract, Transform, and Load (ETL) processes. This service automates the traditionally complex ETL tasks, making it accessible to users with varying levels of technical expertise. With a visual interface, Glue allows users to create ETL jobs, enabling seamless data preparation and transformation without the need for extensive coding, thereby accelerating the data processing workflow.
6. **Amazon Redshift** - Amazon Redshift is a fully managed, petabyte-scale data warehouse service. It is designed to efficiently analyze large datasets using a high-performance and cost-effective architecture. Redshift supports complex queries across structured data, making it a powerful solution for data warehousing and analytics. With features like automatic backups, scalability, and integration with other AWS services, Redshift is a key component for organizations looking to derive insights from their data.
7. **Amazon EventBridge** - Amazon EventBridge is a serverless event bus service that makes it easy to connect various AWS services and applications in a decoupled manner. It allows organizations to build event-driven architectures, enabling seamless communication and coordination between different components. By simplifying event-driven workflows and providing a scalable event bus, EventBridge facilitates the creation of resilient and loosely coupled applications.
8. **Amazon SNS (Simple Notification Service)** - Amazon SNS is a fully managed messaging service that enables the distribution of messages and notifications to a distributed set of recipients or systems. It supports various communication protocols, including SMS, email, and HTTP. SNS simplifies the process of sending messages and notifications, making it a versatile tool for building scalable and decoupled applications that require real-time communication and updates.

Architecture:



Process:

- Amazon S3 serves as the foundational data lake, accommodating our diverse datasets. With a robust structure within S3 buckets, raw and processed data find a secure and scalable abode.
- As the pipeline progresses, AWS Glue Catalog steps into the spotlight, acting as a central repository for metadata management. Glue Catalog captures essential structural and operational metadata about the data assets stored in S3, facilitating seamless data discovery and integration across various AWS services.
- Glue Crawlers take the stage, automating the analysis of data within S3, dynamically inferring schemas, and updating the Glue Data Catalog. This automation ensures that the metadata remains accurate and up-to-date if dataset evolves.
- The journey continues with Glue Catalog Data Quality checks, enhancing the reliability of the data by defining and executing custom rules and metrics.
- Glue's Low Code ETL capabilities come into play, allowing the creation of ETL jobs with a visual interface. These jobs efficiently read data from S3, apply necessary transformations, and then either store the processed data back in S3 or seamlessly transfer it to the analytical powerhouse, Amazon Redshift.
- Redshift, as the central data warehouse, becomes the epicenter for analytical queries, hosting the processed data for deeper insights into industry trends and audience behavior.

Outcome:

1. S3 Bucket

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 (2) [info](#)

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

Name	AWS Region	Access	Creation date
<input type="radio"/> aws-glue-assets-339712801489-ap-south-1	Asia Pacific (Mumbai) ap-south-1	Bucket and objects not public	February 3, 2024, 12:22:49 (UTC+05:30)
<input type="radio"/> movie-data-analysis-proj	Asia Pacific (Mumbai) ap-south-1	Bucket and objects not public	February 3, 2024, 14:32:53 (UTC+05:30)

Amazon S3 > Buckets > movie-data-analysis-proj

movie-data-analysis-proj [info](#)

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

Objects (5) [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](#)

Find objects by prefix

Name	Type	Last modified	Size	Storage class
<input type="checkbox"/> bad_records/	Folder	-	-	-
<input type="checkbox"/> historical_data_rule_outcome/	Folder	-	-	-
<input type="checkbox"/> input/	Folder	-	-	-
<input type="checkbox"/> rule_outcome/	Folder	-	-	-
<input type="checkbox"/> temp/	Folder	-	-	-

2. RedShift Cluster

The screenshot displays the Amazon Redshift console interface. At the top, a navigation bar shows various services like VPC, SecurityGroup, and Endpoints. The main content area is titled 'aws-redshift-cluster' and includes a 'General information' tab. The cluster status is 'Available'. Key details include the cluster identifier 'aws-redshift-cluster', a custom domain name, and a node type of 'dc2.large'. The endpoint URL is 'aws-redshift-cluster.c86rcdap9cnq.ap-south-1.redshift.amazonaws.com:5439/dev'. Below the general information, there are sections for 'Recommendations (0)', 'Alarms (0)', and 'Events (3)'. The bottom of the console shows 'Cluster metrics (11/11)'.

Cluster identifier	Status	Node type	Endpoint
aws-redshift-cluster	Available	dc2.large	aws-redshift-cluster.c86rcdap9cnq.ap-south-1.redshift.amazonaws.com:5439/dev
Custom domain name	Date created	Number of nodes	JDBC URL
-	February 04, 2024, 13:30 (UTC+05:30)	1	jdbc:redshift://aws-redshift-cluster.c86rcdap9cnq.ap-south-1.redshift.amazonaws.com:5439/dev
Cluster configuration	Storage used		ODBC URL
Production	0.26% (0.42 of 160 GB used)		Driver={Amazon Redshift (x64)}; Server=aws-redshift-cluster.c86rcdap9cnq.ap-south-1.redshift.amazonaws.com; Database=dev
	Multi-AZ		
	No		

3. Security and Endpoints

The screenshot shows the Amazon VPC console. The left sidebar contains a navigation menu with options like 'Virtual private cloud', 'Security', 'DNS Firewall', and 'Network Firewall'. The main content area is titled 'Endpoints (1/3)' and displays a table of endpoints. The table has columns for Name, VPC endpoint ID, VPC ID, Service name, Endpoint type, and Status. Three endpoints are listed: 'redshift-s3-endpoint', 'cloudwatch-endpoint', and 'glue-endpoint'. Below the table, the details for the 'redshift-s3-endpoint' are shown, including its VPC ID, status, creation time, and service name.

Name	VPC endpoint ID	VPC ID	Service name	Endpoint type	Status
redshift-s3-endpoint	vpce-061b3d6de2f2ad1c6	vpc-06b3f4dd7de08a760	com.amazonaws.ap-south-1.s3	Gateway	Available
cloudwatch-endpoint	vpce-08e0825eed294cf5	vpc-06b3f4dd7de08a760	com.amazonaws.ap-south-1.monitoring	Interface	Available
glue-endpoint	vpce-0fe0933fb08d91c5	vpc-06b3f4dd7de08a760	com.amazonaws.ap-south-1.glue	Interface	Available

vpce-061b3d6de2f2ad1c6 / redshift-s3-endpoint

Endpoint ID	Status	Creation time	Endpoint type
vpce-061b3d6de2f2ad1c6	Available	Sunday, February 4, 2024 at 13:50:09 GMT+5:30	Gateway
VPC ID	Status message	Service name	Private DNS names enabled
vpc-06b3f4dd7de08a760	-	com.amazonaws.ap-south-1.s3	No

EC2 Dashboard

EC2 Global View

Events

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

New

Images

AMIs

AMI Catalog

Elastic Block Store

Volumes

Snapshots

Lifecycle Manager

Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

Load Balancing

Load Balancers

CloudShell

Feedback

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

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#SecurityGroupsgroupId=sg-02f97dcae1346059

sg-02f97dcae1346059 - default

Details

Security group name: default

Security group ID: sg-02f97dcae1346059

Description: default VPC security group

VPC ID: vpc-06b3f4dd7de08a760

Owner: 339712801489

Inbound rules count: 2

Outbound rules count: 1

Permission entries: 1

Inbound rules (2)

Name	Security group rule...	IP version	Type	Protocol	Port range	Source	Description
-	sgr-04242a00d1106b...	IPv4	Redshift	TCP	5439	0.0.0.0/0	-
-	sgr-025d6ecaee61e0cc7	-	All traffic	All	All	sg-02f97dcae134605...	-

4. Crawlers

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

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 (2)

View and manage all available crawlers.

Last updated (UTC): February 4, 2024 at 10:51:33

Filter crawlers

Name	State	Schedule	Last run	Last run timestamp	Log	Table change...
movie-data-analysis-s3-crawler	Ready		Succeeded	February 4, 2024 at 09:28:...	View log	1 created
redshift-imdb-movie-table-crawler	Ready		Succeeded	February 4, 2024 at 09:45:...	View log	1 created

Console Home | movie-data-ana | Amazon Redshift | Endpoints | VPC | SecurityGroup | Crawlers - AWS | Connectors | Tables - AWS Glue | Editor | Redshift | Runs - Editor - A | Endpoints | VPC | Rules | Amazon | movie-data-ana | + | - | X

ap-south-1.console.aws.amazon.com/gluestudio/home?region=ap-south-1#/connectors

Services Search [Alt+S]

AWS Glue > Connectors

Connectors info

Marketplace connectors
Subscribe to connectors from AWS partners to expand your data sources.
[Go to AWS Marketplace](#)

Custom connectors
Provide your own connector to expand your data sources. [Creating custom connectors](#)
[Create custom connector](#)

Connectors (0) info
You can manage your connectors or use them to create connections.
 < 1 >

Name	Type	Last modified
No connectors		
No connectors to display. You can create a custom connector or get a Marketplace connector . Create custom connector		

Connections (2) info
You can manage your connections or use a connection in a job.
 < 1 >

Name	Type	Last modified
<input checked="" type="radio"/> movie-data-analysis-connection	JDBC	Feb 04, 2024
<input type="radio"/> aws-s3-redshift-connection	JDBC	Feb 04, 2024

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

5. Catalog

Console Home | movie-data-ana | Amazon Redshift | Endpoints | VPC | SecurityGroup | Crawlers - AWS | Connectors - A | Databases | Editor | Redshift | Runs - Editor - A | Endpoints | VPC | Rules | Amazon | movie-data-ana | + | - | X

ap-south-1.console.aws.amazon.com/glue/home?region=ap-south-1#/data-catalog/databases

Services Search [Alt+S]

AWS Glue > Databases

Databases (3)
A database is a set of associated table definitions, organized into a logical group.
 < 1 > ⚙

<input type="checkbox"/>	Name	Description	Location URI	Created on (UTC)
<input type="checkbox"/>	default	-	-	February 1, 2024 at 13:19:30
<input type="checkbox"/>	movie-data-analysis-db	-	-	February 4, 2024 at 09:28:00
<input type="checkbox"/>	s3-glue-redshift	-	-	February 4, 2024 at 08:12:50

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

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

Console Home | movie-data-ana | Amazon Redshift | Endpoints | VPC | SecurityGroup | Crawlers - AWS | Connectors - AI | Databases - x | Editor | Redshift | Runs - Editor - AI | Endpoints | VPC | Rules | Amazon | movie-data-ana | +

ap-south-1.console.aws.amazon.com/glue/home?region=ap-south-1#/v2/data-catalog/databases/view/movie-data-analysis-db?catalogId=3397128...

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.

movie-data-analysis-db

Last updated (UTC)
February 4, 2024 at 10:52:37

Database properties

Name	movie-data-analysis-db	Description	-	Location	-	Created on (UTC)	February 4, 2024 at 09:28:00
------	------------------------	-------------	---	----------	---	------------------	------------------------------

Tables (2)

Last updated (UTC)
February 4, 2024 at 10:52:37

View and manage all available tables.

Filter tables

<input type="checkbox"/>	Name	Database	Location	Classification	Deprecated	View data	Data quality
<input type="checkbox"/>	redshift_dev_movies_imdb_movies_rating	movie-data-analysis-db	dev.movies.imdb_movies_rating	redshift	-	-	View data quality
<input type="checkbox"/>	s3_imdb_movies_rating_csv	movie-data-analysis-db	s3://movie-data-analysis-proj/i	CSV	-	Table data	View data quality

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

6. Glue

Console Home | movie-data-ana | Amazon Redshift | Endpoints | VPC | SecurityGroup | Crawlers - AWS | Connectors - AI | Databases - AWS | Editor | Redshift | Jobs - AWS | Endpoints | VPC | Rules | Amazon | movie-data-ana | +

ap-south-1.console.aws.amazon.com/gluestudio/home?region=ap-south-1#/jobs?selectedInterface=source_and_target

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections

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 Studio

Create job

Author in a visual interface focused on data flow.
Visual ETL

Author using an interactive code notebook.
Notebook

Author code with a script editor.
Script editor

Example jobs

Create example job

Your jobs (1)

Filter jobs

<input type="checkbox"/>	Job name	Type	Last modified	AWS Glue version
<input type="checkbox"/>	movie-data-analysis-etl	Glue ETL	2/4/2024, 3:43:38 PM	4.0

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

Restore

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

Console Home | movie-data-ansi | Amazon Redshift | Endpoints | VPC | SecurityGroup | Crawlers - AWS | Connectors - AWS | Databases - AWS | Editor | Redshift | Visual - Edit | Endpoints | VPC | Rules | Amazon | movie-data-ansi

ap-south-1.console.aws.amazon.com/gluestudio/home?region=ap-south-1#/editor/job/movie-data-analysis-etl/graph

How do you like the new UI? Let us know
The option to toggle to the old UI will be removed on Feb 29th, 2024.

movie-data-analysis-etl
Last modified on 2/4/2024, 3:43:38 PM Try new UI Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Visual

+

Visual ETL workflow diagram showing data flow from source to target, including transformations like Data Quality Check, Rule Set, and Conditional Router.

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

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

7. EventBridge

Console Home | C | movie-data-ansi | Amazon Redshift | Endpoints | VPC | SecurityGroup | E | Crawlers - AWS | Connectors - AWS | Databases - AWS | Editor | Redshift | Visual - Editor - A | Rules | Amazon | movie-data-ansi

ap-south-1.console.aws.amazon.com/events/home?region=ap-south-1#/rules

Amazon EventBridge Rules

A rule watches for specific types of events. When a matching event occurs, the event is routed to the targets associated with the rule. A rule can be associated with one or more targets.

Select event bus

Event bus
Select or enter event bus name
default

Rules (1)
Find rules Any status

Name	Status	Type	ARN	Description
movie-data-analysis-event-bridge	Enabled	Standard	arn:aws:events:ap-south-1:339712801489:rule/movie-data-analysis-event-bridge	

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

Amazon EventBridge console showing the details of the **movie-data-analysis-event-bridge** rule.

Rule details

Property	Value
Rule name	movie-data-analysis-event-bridge
Status	Enabled
Event bus name	default
Type	Standard
Description	
Rule ARN	arn:aws:events:ap-south-1:339712801489:rule/movie-data-analysis-event-bridge
Event bus ARN	arn:aws:events:ap-south-1:339712801489:event-bus/default

Event pattern

```
1 {
2   "source": ["aws-glue-dataquality"],
3   "detail-type": ["Data Quality Evaluation Results Available"],
4   "detail": {
5     "state": ["SUCCEEDED", "FAILED"]
6   }
7 }
```

Buttons: Edit, Disable, Delete, CloudFormation Template

8. SNS Topic

Amazon SNS console showing the details of the **movie-data-analysis-topic**.

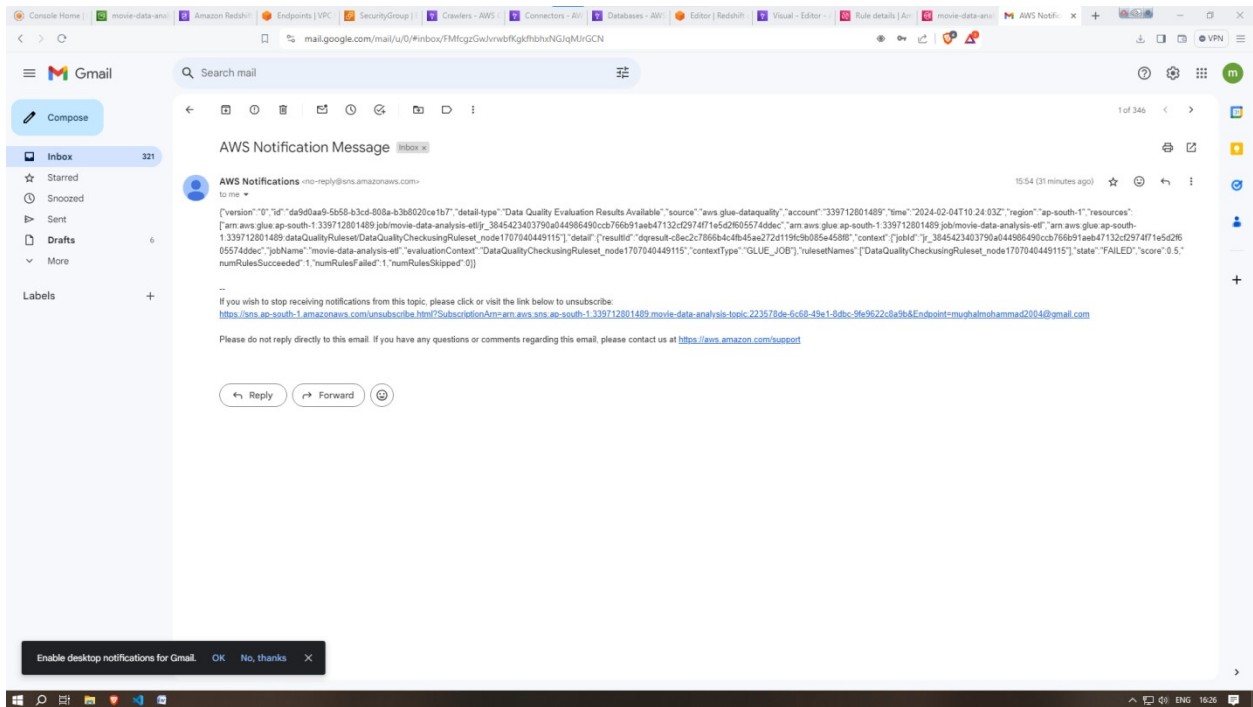
Details

Property	Value
Name	movie-data-analysis-topic
Display name	-
ARN	arn:aws:sns:ap-south-1:339712801489:movie-data-analysis-topic
Topic owner	339712801489
Type	Standard

Subscriptions (1)

ID	Endpoint	Status	Protocol
223578de-6c68-49e1-8dbc-9fe9622c8a9b	mughalmohammad2004@gmail.com	Confirmed	EMAIL

Buttons: Edit, Delete, Publish message



9. Target Redshift table

Redshift query editor v2

Create Load data

Filter resources

aws-redshift-cluster

awsdatacatalog

dev

movies

Tables

imdb_movies_rating

Views

Functions

Stored procedures

public

sales

imdb_movies_rating

Field	Type	NL	CMP
A poster_link	character varying(65535)	NULL	lzo
A series_title	character varying(65535)	NULL	lzo
A released_year	character varying(10)	NULL	lzo
A certificate	character varying(50)	NULL	lzo
A runtime	character varying(50)	NULL	lzo
A genre	character varying(200)	NULL	lzo
A imdb_rating	numeric(10,2)	NULL	az64
A overview	character varying(65535)	NULL	lzo
A meta_score	integer	NULL	az64
A director	character varying(200)	NULL	lzo
A star1	character varying(200)	NULL	lzo

```
1 create schema movies;
2
3 CREATE TABLE movies.imdb_movies_rating (
4   poster_link VARCHAR(MAX),
5   series_title VARCHAR(MAX),
6   released_year VARCHAR(10),
7   certificate VARCHAR(50),
8   runtime VARCHAR(50),
9   genre VARCHAR(200),
10  imdb_rating DECIMAL(10,2),
11  overview VARCHAR(MAX),
12  meta_score INT,
13  director VARCHAR(200),
14  star1 VARCHAR(200),
15  star2 VARCHAR(200),
16  star3 VARCHAR(200),
17  star4 VARCHAR(200),
18  no_of_votes INT,
19  gross VARCHAR(20)
20 );
21
22 select * from movies.imdb_movies_rating limit 10;
23
24 select series_title, genre, overview, meta_score, imdb_rating
25 from movies.imdb_movies_rating
26 where imdb_rating = (select max(imdb_rating) from movies.imdb_movies_rating)
```

Result 1 (1)

Export Chart

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

10. Sample Queries:

- select * from movies.imdb_movies_rating limit 10;

The screenshot shows the Amazon Redshift console. On the left, the schema for the `imdb_movies_rating` table is displayed. On the right, the results of a query are shown.

Field	Type	NL	CMP
poster_link	character varying(65535)	NULL	lzo
series_title	character varying(65535)	NULL	lzo
released_year	character varying(10)	NULL	lzo
certificate	character varying(50)	NULL	lzo
runtime	character varying(50)	NULL	lzo
genre	character varying(200)	NULL	lzo
imdb_rating	numeric(10,2)	NULL	az64
overview	character varying(65535)	NULL	lzo
meta_score	integer	NULL	az64
director	character varying(200)	NULL	lzo
star1	character varying(200)	NULL	lzo

poster_link	series_title	released_year	certificate	runtime	genre	imdb_rating
https://m.media-amazon.c...	The Godfather	1972	A	175 min	Crime, Drama	9.2
https://m.media-amazon.c...	The Godfather: Part II	1974	A	202 min	Crime, Drama	9
https://m.media-amazon.c...	The Lord of the Rings: Th...	2003	U	201 min	Action, Adventure, Drama	8.9
https://m.media-amazon.c...	Schindler's List	1993	A	195 min	Biography, Drama, History	8.9
https://m.media-amazon.c...	Fight Club	1999	A	139 min	Drama	8.8
https://m.media-amazon.c...	Forrest Gump	1994	UA	142 min	Drama, Romance	8.8
https://m.media-amazon.c...	The Lord of the Rings: Th...	2002	UA	179 min	Action, Adventure, Drama	8.7
https://m.media-amazon.c...	Goodfellas	1990	A	146 min	Biography, Crime, Drama	8.7
https://m.media-amazon.c...	One Flew Over the Cuckoo...	1975	A	133 min	Drama	8.7
https://m.media-amazon.c...	Gisaengchung	2019	A	132 min	Comedy, Drama, Thriller	8.6

- select series_title, genre, overview, meta_score, imdb_rating
from movies.imdb_movies_rating
where imdb_rating = (select max(imdb_rating) from movies.imdb_movies_rating);

The screenshot shows the Amazon Redshift console. On the left, the schema for the `imdb_movies_rating` table is displayed. On the right, the results of a query are shown.

Field	Type	NL	CMP
poster_link	character varying(65535)	NULL	lzo
series_title	character varying(65535)	NULL	lzo
released_year	character varying(10)	NULL	lzo
certificate	character varying(50)	NULL	lzo
runtime	character varying(50)	NULL	lzo
genre	character varying(200)	NULL	lzo
imdb_rating	numeric(10,2)	NULL	az64
overview	character varying(65535)	NULL	lzo
meta_score	integer	NULL	az64
director	character varying(200)	NULL	lzo
star1	character varying(200)	NULL	lzo

series_title	genre	overview	meta_score	imdb_rating
The Shawshank Redempt...	Drama	Two imprisoned men bon...	80	9.3

- select genre, count(genre) as count_of_genre
from movies.imdb_movies_rating
group by genre
order by count(genre) desc;

The screenshot shows the Amazon Redshift console. On the left, the schema for the `imdb_movies_rating` table is displayed. On the right, the results of a query are shown.

Field	Type	NL	CMP
poster_link	character varying(65535)	NULL	lzo
series_title	character varying(65535)	NULL	lzo
released_year	character varying(10)	NULL	lzo
certificate	character varying(50)	NULL	lzo
runtime	character varying(50)	NULL	lzo
genre	character varying(200)	NULL	lzo
imdb_rating	numeric(10,2)	NULL	az64
overview	character varying(65535)	NULL	lzo
meta_score	integer	NULL	az64
director	character varying(200)	NULL	lzo
star1	character varying(200)	NULL	lzo

genre	count_of_genre
Drama	80
Drama, Romance	37
Comedy, Drama	35
Comedy, Drama, Romance	31
Action, Crime, Drama	30
Crime, Drama, Thriller	28
Biography, Drama, History	28
Crime, Drama, Mystery	27
Crime, Drama	26
Animation, Adventure, Co...	24
Action, Adventure, Sci-Fi	21
Biography, Crime, Drama	16
Comedy, Crime, Drama	15

- `select genre, avg(imdb_rating) as average_imdb_rating_for_genre`
`from movies.imdb_movies_rating`
`group by genre`
`order by average_imdb_rating_for_genre desc;`

The screenshot shows the AWS CloudShell interface. On the left, a table named `imdb_movies_rating` is displayed with columns: `series_title`, `released_year`, `certificate`, `runtime`, `genre`, `imdb_rating`, `overview`, `meta_score`, `director`, `star1`, `star2`, `star3`, and `star4`. On the right, the query result is shown as a table with two columns: `genre` and `average_imdb_rating_for_genre`. The results are ordered by the average rating in descending order.

genre	average_imdb_rating_for_genre
Animation, Drama, War	8.5
Drama, Musical	8.4
Action, Sci-Fi	8.4
Western	8.35
Drama, Mystery, War	8.35
Crime, Drama, Sci-Fi	8.3
Crime, Drama, Music	8.3
Mystery, Romance, Thriller	8.3
Adventure, Mystery, Thriller	8.3
Drama, Fantasy, Horror	8.3
Comedy, Musical, Romance	8.3
Drama, Family, Sport	8.25
Drama, Family	8.23

- `select director, count(series_title) as count_of_movies_by_director`
`from movies.imdb_movies_rating`
`group by director`
`order by count_of_movies_by_director desc;`

The screenshot shows the AWS CloudShell interface. On the left, the same `imdb_movies_rating` table is visible. On the right, the query result is shown as a table with two columns: `director` and `count_of_movies_by_director`. The results are ordered by the count of movies in descending order.

director	count_of_movies_by_director
Alfred Hitchcock	14
Steven Spielberg	13
Hayao Miyazaki	11
Martin Scorsese	10
Akira Kurosawa	10
Stanley Kubrick	9
Woody Allen	9
Billy Wilder	9
Quentin Tarantino	8
Christopher Nolan	8
Clint Eastwood	8
David Fincher	8
Rob Reiner	7