

1st Meeting: 11/29 (Wed) @9pm (20 min.)

Agenda_1 Decide dataset.

Amazon Product Dataset 2023

https://www.kaggle.com/datasets/asaniczka/amazon-products-dataset-2023-1-4m-products?select=amazon_products.csv

Code example) <https://www.kaggle.com/code/alancano/amazon-products-2023>

Agenda_2

Goal: Analyze customer ratings to find top-rated products

→ Analyzing customer ratings is typically straightforward. It involves sorting or filtering data based on ratings and drawing conclusions.

Agenda_3

Q#1. A data lake or data warehouse - defend your choice. Why are you choosing to store your data in one over the other?

A: Data Lake

Reasoning:

- Flexibility: Data lakes can store raw, unstructured, and structured data, providing flexibility in handling various data formats.
- Scalability: Data lakes, especially cloud-based solutions, can scale horizontally to accommodate growing datasets.
- Cost-Efficiency: Cloud-based data lakes often offer a pay-as-you-go model, making it cost-effective for storing large amounts of data.

Q#2. Connect data lake to AWS

Connect to a Distributed Cloud Service - Choice: Amazon Web Services (AWS)

Reasoning:

- Widely Used: AWS is a popular and widely adopted cloud service with extensive documentation and community support.
- Comprehensive Services: AWS provides a variety of services suitable for data storage, processing, and analysis, including S3 for data lake storage, and EMR for running Spark applications.

In class on 11/30 from prof. Wu

- Grading point=Design ETL system on AWS
- The best case below link;

<https://github.com/kamranadil436/bigdata-finalproject/blob/main/Big%20Data%20Final%20Project.pdf>

2nd Meeting: 12/4 (Mon) @9pm (20 min.)

Agenda Q#3. Run your Spark application over those distributed services.

A: We are using AWS Glue to design ETL data pipeline. AWS Glue is a fully managed extract, transform, and load (ETL) service that uses Apache Spark under the hood. When you create and run an AWS Glue ETL job, it automatically runs a Spark application in a managed Spark environment. *// find a image of Spark UI from below captured images*

ETLjob-bd

The screenshot displays the AWS Glue console interface for a specific ETL job. At the top, there is a navigation bar with tabs: 'Visual', 'Script', 'Job details' (which is the active tab), 'Runs', 'Data quality - updated', 'Schedules', and 'Version Control'. Below the navigation bar, the 'Basic properties' section is visible, containing several configuration fields:

- Name:** ETLjob-bd
- Description - optional:** A text area for a description, with a note below stating 'Descriptions can be up to 2048 characters long.'
- IAM Role:** A dropdown menu showing 'my-bd-glue' and a refresh button.
- Type:** A dropdown menu showing 'Spark', with a note: 'The type of ETL job. This is set automatically based on the types of data sources you have selected.'
- Glue version:** A dropdown menu showing 'Glue 4.0 - Supports spark 3.3, Scala 2, Python 3'.
- Language:** A dropdown menu showing 'Python 3'.
- Worker type:** A dropdown menu showing 'G 1X (4vCPU and 16GB RAM)', with a note: 'Set the type of predefined worker that is allowed when a job runs.'

Agenda Role: *//Feel free to modify*

Ramis- report

Nafia-ppt

Gahyeon-aws

Ravi-aws

In the next meeting,

Share the script, visualized results and discuss how to write the final report.

Final meeting: 12/7 (Thur) Before presentation

Agenda Finalize presentation, powerpoint

//Shared powerpoint documentation link here;

//Shared final report link here;

Finalize documentation during weekend (12/9-10) and,

Submit Final Report, Ppt, dataset, Github link etc. on Canvas, before **Dec 10, 2023 11:59 PM**

Documentation - Report Structure:

- Introduction:

Briefly introduce the goal of the project..

Explain the significance of analyzing customer ratings for business insights

- Datasets Used:

Describe the source of the data (Amazon product dataset).

Mention key features included in the dataset (e.g., product ratings, qty of reviews, etc.).

- Technical Approach:

Discuss the choice of a data lake (AWS S3) for storing the dataset.

Detail the steps taken to set up and configure AWS services (S3, Glue, Athena).

Explain how Spark is utilized for processing and analyzing the customer ratings.

- Results:

Present findings on top-rated products based on customer ratings.

- Challenges and Solutions:

Discuss any challenges encountered during the project and how they were addressed.

//(1)

Challenge: After Glue ETL job, we need to load table Metadata since AWS automatically saves the table as a set of schema and metadata separately.

Solution: The command below discovers the partitions in data and updates Athena metadata accordingly.

MSCK REPAIR TABLE your_table_name;

//(2)

Challenge: Wanted to visualize more through Quicksight (dashboard visualization tool). But, it is \$24/month. So, we used only Athena.

Solution: Money...?

- Conclusion:

Summarize the key insights gained from the analysis.

Reflect on the significance of the findings for businesses.

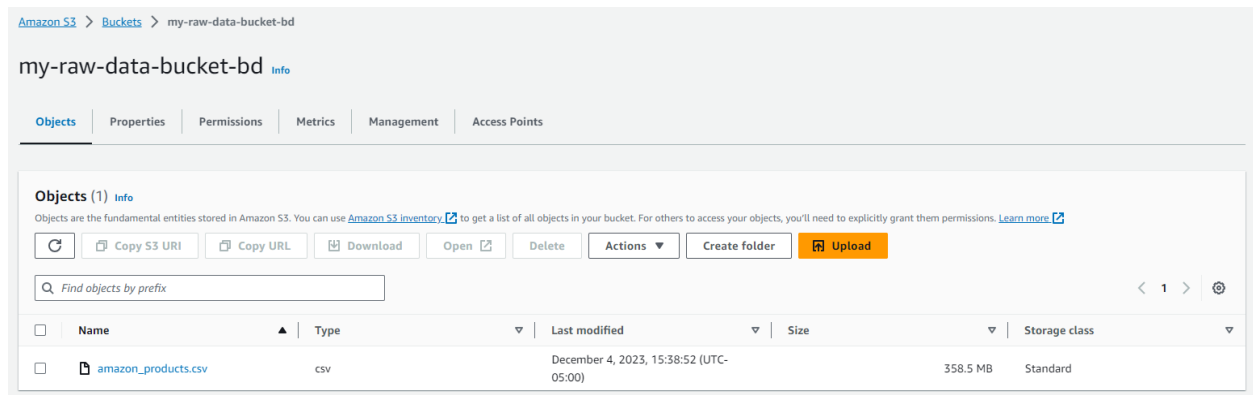
- Appendix:

Include any additional technical details, code snippets, or configurations used.

ScreenShots of each steps on AWS

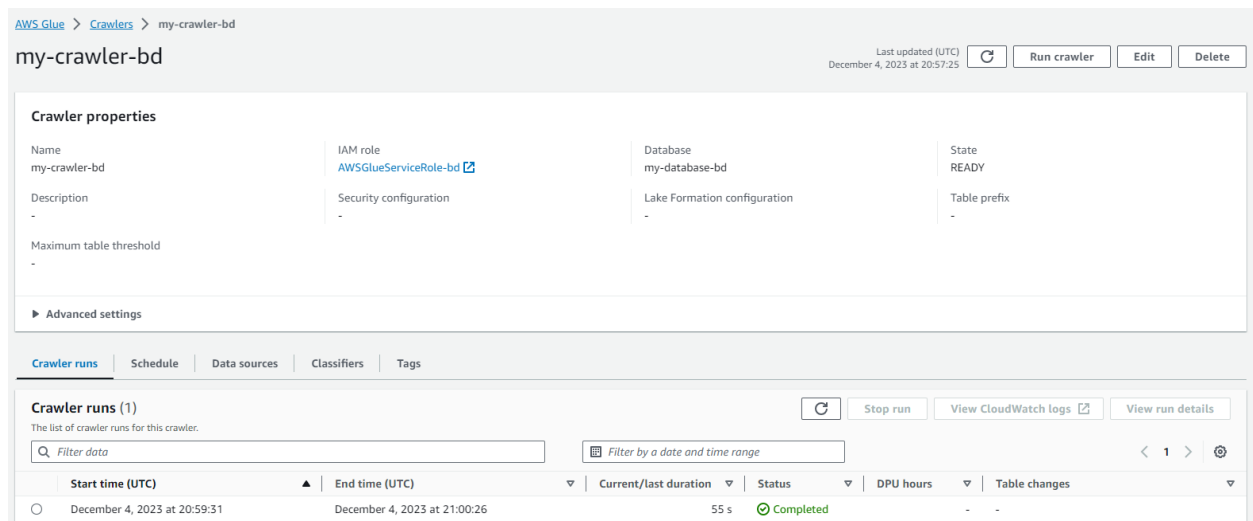
1. Set up S3 bucket and upload .csv file:

Create an S3 bucket using the AWS Management Console. S3 is used as a storage solution for your data. Uploading the .csv file to S3 makes the data accessible for processing by various AWS services.



2. Add Glue Crawler and run:

Create a Glue Crawler using AWS Glue Console. Specify the S3 bucket and path where your data is stored. Run the Glue Crawler to discover the schema and metadata of your data. The Glue Crawler is used to automatically infer the schema and metadata of your data. This is important for the subsequent steps where you'll be processing and transforming the data.



Running an AWS Glue Crawler automatically creates or updates tables in the Glue Data Catalog, streamlining the process of managing metadata and making your data easily accessible for analysis and transformation.

[AWS Glue](#) > [Tables](#) > amazon_products_csv

amazon_products_csv

Last updated (UTC)
December 6, 2023 at 23:22:19

Version 0 (Current version) ▼

Actions ▼

Table overview

Data quality New

Table details

Advanced properties

Name

amazon_products_csv

Location

s3://my-raw-data-bucket-bd/amazon_products.csv

Description

-

Connection

-

Database

my-database-bd

Deprecated

-

Classification

CSV

Last updated

December 6, 2023 at 23:22:19

Input format

org.apache.hadoop.mapred.TextInputFormat

Output format

org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat

Serde serialization lib

org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

Schema

Partitions

Indexes

Column statistics - new

Schema (11)

Edit schema as JSON

Edit schema

View and manage the table schema.

Q Filter schemas

< 1 > ⚙

#	Column name	Data type	Partition key	Comment
1	asin	string	-	-
2	title	string	-	-
3	imgurl	string	-	-
4	producturl	string	-	-
5	stars	double	-	-
6	reviews	bigint	-	-
7	price	double	-	-
8	listprice	double	-	-
9	category_id	bigint	-	-
10	isbestseller	boolean	-	-
11	boughtinlastmonth	bigint	-	-

3. IAM Role and S3 bucket permission:
- (1) Create an IAM role with the necessary permissions for your Glue ETL job to access S3 and other AWS services. The IAM role ensures that your Glue ETL job has the necessary permissions to interact with other AWS services.

[IAM](#) > [Policies](#) > my-bd-project

my-bd-project [Info](#)

Delete

Policy details

Type
Customer managed

Creation time
December 05, 2023, 14:50 (UTC-05:00)

Edited time
December 05, 2023, 14:50 (UTC-05:00)

ARN
 arn:aws:iam::563460431563:policy/my-bd-project

Permissions

Entities attached

Tags

Policy versions (1)

Access Advisor

Permissions defined in this policy [Info](#)

Edit

Summary

JSON

Permissions defined in this policy document specify which actions are allowed or denied. To define permissions for an IAM identity (user, user group, or role), attach a policy to it

Q Search

Allow (4 of 402 services)

Show remaining 398 services

Service	Access level	Resource	Request condition
Athena	Full access	All resources	None
Glue	Full access	All resources	None
QuickSight	Full access	All resources	None
S3	Full access	All resources	None

[IAM](#) > [Roles](#) > my-bd-glue

my-bd-glue [Info](#)

Delete

Allows Glue to call AWS services on your behalf.

Summary

Edit

Creation date
December 05, 2023, 14:55 (UTC-05:00)

ARN
 arn:aws:iam::563460431563:role/my-bd-glue

Last activity
-

Maximum session duration
1 hour

Permissions

Trust relationships

Tags

Access Advisor

Revoke sessions

Permissions policies (1) [Info](#)

Simulate

Remove

Add permissions ▾

You can attach up to 10 managed policies.

Q Search

Filter by Type
All types ▾

< 1 >

<input type="checkbox"/>	Policy name	Type	Attached entities
<input type="checkbox"/>	my-bd-project	Customer managed	1

(2) Ensure that the S3 bucket has the correct permissions, allowing the IAM role to read and write data. Proper S3 bucket permissions are crucial for reading and writing data during the ETL process.


[Amazon S3](#) > [Buckets](#) > my-raw-data-bucket-bd

my-raw-data-bucket-bd [Info](#)

[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)

Bucket policy

The bucket policy, written in JSON, provides access to the objects stored in the bucket. Bucket policies don't apply to objects owned by other accounts. [Learn more](#)

**Public access is blocked because Block Public Access settings are turned on for this bucket**

To determine which settings are turned on, check your Block Public Access settings for this bucket. [Learn more about using Amazon S3 Block Public Access](#)

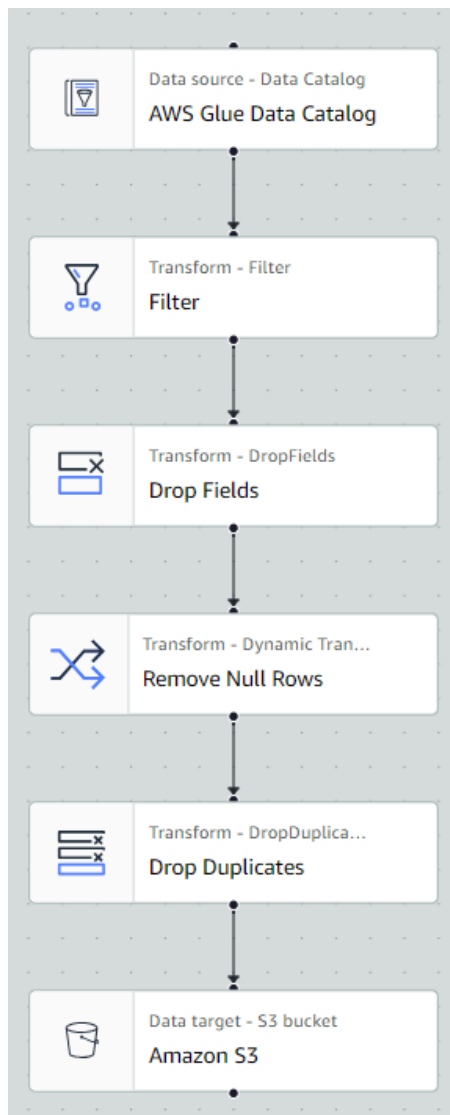
```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "glue.amazonaws.com"
      },
      "Action": [
        "s3:GetObject",
        "s3:PutObject"
      ],
      "Resource": "arn:aws:s3:::my-raw-data-bucket-bd/*"
    }
  ]
}
```

4. Design Visual Glue ETL job:

- Use the AWS Glue Console to create a new ETL job.
- Define the source (Data Catalog table from glue crawler), target (S3), and transformations needed using a visual interface or Glue Python/Spark scripts.
- Configure the job to output the data in .csv format.

The Glue ETL job is designed to transform and move data from source to target. The visual interface simplifies the ETL process, and you can specify the output format as .csv for your use case.

(1) Visual Mode



[←ETL Data Pipeline]

Transform

Builds a new output by selecting records from the input data that satisfy a specified predicate function

☒ **Global AND**
All filter conditions will be applied as a global "AND."

☐ **Global OR**
All filter conditions will be applied as a global "OR."

Filter condition | [Info](#)

Specify your filter condition by choosing the key, operator, and entering a value.

Key | **Operation**

stars | >=

Value

4

Remove

Key | **Operation**

reviews | >=

Value

10000

Remove

Key | **Operation**

boughtinlastmonth | >=

Value

10000

[↑Filter]

Transform

Name

Drop Fields

Node parents

Choose which nodes will provide inputs for this one.

Choose one or more parent node

Filter X

Filter - Transform

DropFields

	Field	Data type
<input type="checkbox"/>	price	double
<input type="checkbox"/>	reviews	long
<input type="checkbox"/>	stars	double
<input checked="" type="checkbox"/>	category_id	long
<input checked="" type="checkbox"/>	imgurl	string
<input checked="" type="checkbox"/>	listprice	double
<input type="checkbox"/>	isbestseller	boolean
<input checked="" type="checkbox"/>	producturl	string
<input checked="" type="checkbox"/>	asin	string
<input type="checkbox"/>	title	string
<input type="checkbox"/>	boughtinlastmonth	long

[↑Drop Fields]

Drop Duplicates: Eliminate duplicate rows in your dataset to ensure each product is unique.

Drop null Fields: Remove represent null value including empty string, null string, -1 integer

Format

Parquet

Compression Type

Snappy

S3 Target Location

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

Q s3://my-result-data-buck X View Browse S3

Data Catalog update options | [Info](#)

Choose how you want to update the Data Catalog table's schema and partitions. These options will only apply if the Data Catalog table is an S3 backed source.

☐ Do not update the Data Catalog

☒ Create a table in the Data Catalog and on subsequent runs, update the schema and add new partitions

☐ Create a table in the Data Catalog and on subsequent runs, keep existing schema and add new partitions

Database

Choose the database from the AWS Glue Data Catalog.

my-database-bd

Use runtime parameters

Table name

Enter a table name for the AWS Glue Data Catalog.

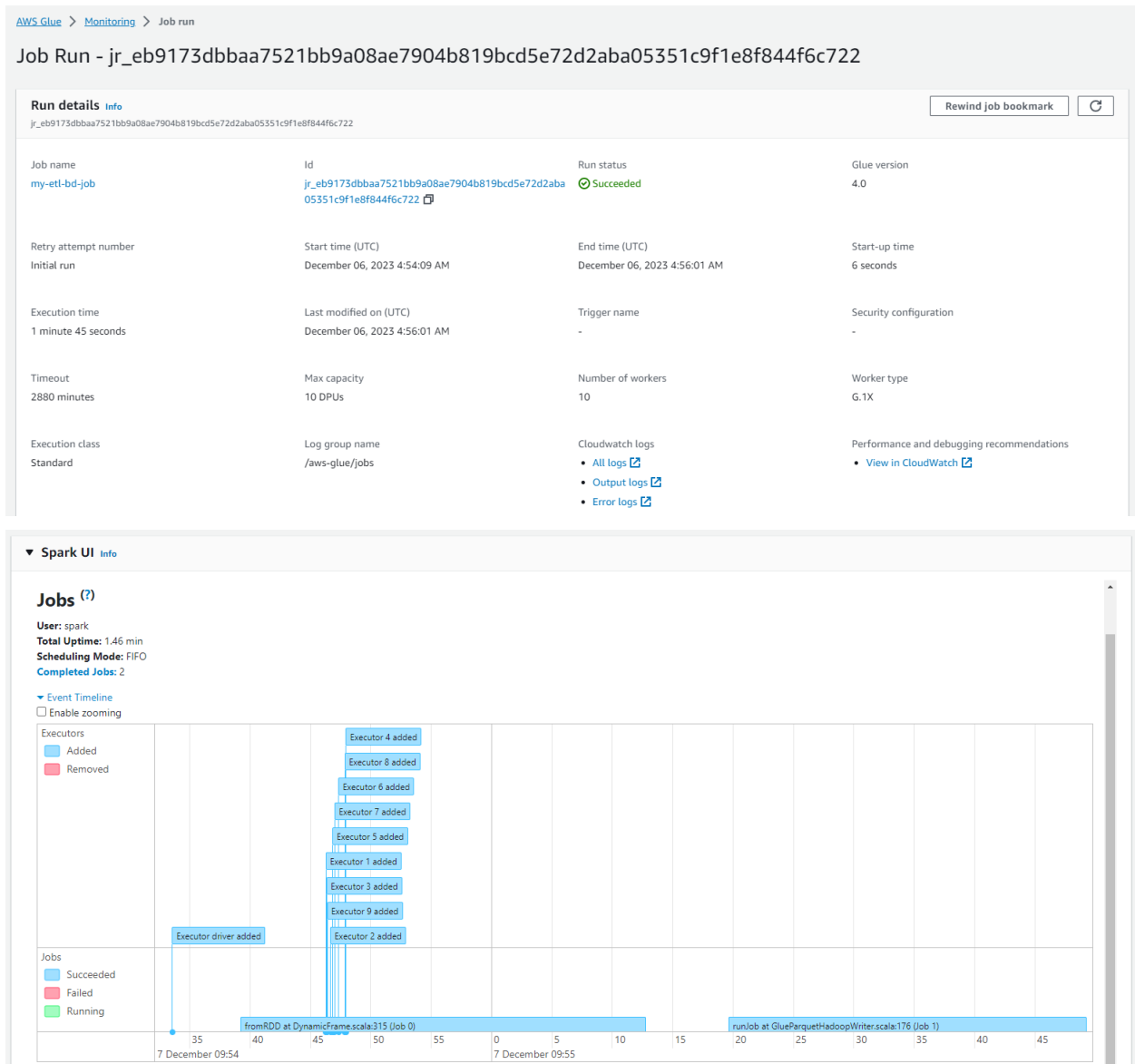
glued_data_bd

[←Data Target - S3]

(2) Python script here; <https://github.com/Duggieeeeey/AWSglue-ETL/blob/main/my-etl-bd-job.py>

5. ETL Result

(1) Spark UI



(2) Before and After in excel view

—Before (size 358MB)—

	A	B	C	D	E	F	G	H	I	J	K	L
1	asin	title	imgUrl	productURL	stars	reviews	price	listPrice	category_ic	isBestSeller	boughtIn	LastMonth
2	B014TMV5	Sion Softsich	https://m.r	https://ww	4.5	0	139.99	0	104	FALSE	2000	
3	B07GDLQ	Luggage Se	https://m.r	https://ww	4.5	0	169.99	209.99	104	FALSE	1000	
4	B07XSCCZ	Platinum El	https://m.r	https://ww	4.6	0	365.49	429.99	104	FALSE	300	
5	B08MVFKG	Freeform H	https://m.r	https://ww	4.6	0	291.59	354.37	104	FALSE	400	
6	B01DJLKZ	B Winfield 2	https://m.r	https://ww	4.5	0	174.99	309.99	104	FALSE	400	
7	B07XSCD2	F Maxlite 5 S	https://m.r	https://ww	4.5	0	144.49	0	104	FALSE	500	
8	B07MXF4G	Hard Shell	(https://m.r	https://ww	4.5	0	169.99	0	104	FALSE	400	
9	B07H515V	(Maxporter	https://m.r	https://ww	4.5	0	299.99	0	104	FALSE	100	
10	B08BXCNI	Omni 2 Har	https://m.r	https://ww	4.5	0	112.63	137.04	104	FALSE	500	
11	B089K44X	T Luggage Se	https://m.r	https://ww	4.4	0	209.99	0	104	FALSE	200	
12	B07QZLHT	I Crew Versa	https://m.r	https://ww	4.7	0	271.99	319.99	104	FALSE	100	
13	B01HIATG	E Chatelet H	https://m.r	https://ww	4.6	0	259.6	413.33	104	FALSE	100	
14	B07QXD3J	E Crew Versa	https://m.r	https://ww	4.5	0	323.99	0	104	FALSE	100	
15	B081JDHN	J Centric 2 H	https://m.r	https://ww	4.5	0	399.92	699.97	104	FALSE	100	
16	B07DMBG	Xpedition 3	https://m.r	https://ww	4.3	0	42	0	104	FALSE	300	
17	B082X1BD	I Stratum 2.	(https://m.r	https://ww	4.5	0	89.95	139.99	104	FALSE	300	
18	B00GI357	2 Women's S	https://m.r	https://ww	4.6	0	164.99	174.99	104	FALSE	200	
19	B084P184	M Ascella X S	https://m.r	https://ww	4.5	0	198.41	209.99	104	FALSE	100	
20	B094LVR	CZ Boren Poly	https://m.r	https://ww	3.9	0	99.99	0	104	FALSE	300	
21	B072Z7K9	E Clear PVC S	https://m.r	https://ww	4.2	0	17.99	19.99	104	FALSE	1000	
22	B09WXL	YG Aspire DLX	https://m.r	https://ww	4.6	0	130.29	159.99	104	FALSE	200	
23	B07FDK2F	V Winfield 3	https://m.r	https://ww	4.3	0	366.53	779.97	104	FALSE	50	
24	B07FNSKM	Oxford Exp	https://m.r	https://ww	4.5	0	119.47	0	104	FALSE	100	
25	B09R35YK	2 Chatelet H	https://m.r	https://ww	4.5	0	276	369.99	104	FALSE	50	
26	B084LP1J	L Anzio Softs	https://m.r	https://ww	4.1	0	98.62	149.99	104	FALSE	300	
27	B0040A6B	1 Amsterdam	https://m.r	https://ww	4.2	0	89.9	121.25	104	FALSE	200	
28	B074ZKXZ	74 Kix Expar	https://m.r	https://ww	4.6	0	78.83	99.99	104	FALSE	300	
29	B08KYF6C	N 4010 Softs	https://m.r	https://ww	4.5	0	123.25	144.99	104	FALSE	200	
30	B0812RQ9	' Platinum El	https://m.r	https://ww	4.4	0	399.49	469.99	104	FALSE	50	

—After(size 59KB)—

	A	B	C	D	E	F	G
1	price	reviews	stars	isbestseller	title	boughtin	lastmonth
2	18.99	27263	4.7	TRUE	Shynerk Ba	10000	
3	24.31	28518	4.5	FALSE	Pure Protei	10000	
4	9.71	21854	4.7	TRUE	Edge Shave	30000	
5	18.99	42849	4.6	TRUE	Etekcitcity Inf	10000	
6	8.99	25420	4.4	TRUE	OLANLY Lu:	10000	
7	4.97	98769	4.5	TRUE	NYX PROFE	10000	
8	9.99	12636	4.6	TRUE	e.l.f. Power	40000	
9	8.71	15828	4.7	FALSE	Neutrogena	20000	
10	4.99	20341	4.4	FALSE	essence L	10000	
11	7.2	55040	4.4	FALSE	Lâ€™™OrÃ©	10000	
12	9.5	37968	4.4	FALSE	e.l.f. Porele	10000	
13	16.49	183726	4.6	TRUE	Schick Hydr	10000	
14	15.95	25280	4.6	TRUE	Lilly's Love	10000	
15	44.95	122450	4.3	TRUE	Pharmedoc	10000	
16	15.45	33115	4.6	FALSE	MAJESTIC F	10000	
17	4.44	40005	4.7	TRUE	Bright Start	10000	
18	14.1	12769	4.8	FALSE	Ddrops Org	10000	
19	19.99	16615	4.5	FALSE	Massage Gi	10000	
20	4.99	28988	4.4	FALSE	essence L	10000	
21	6.01	49923	4.6	TRUE	Palmer's Cc	10000	
22	39.99	14705	4.5	TRUE	RAEMAO Iv	10000	
23	6.47	21095	4.4	FALSE	Dr. Schollâ	10000	
24	14.99	22157	4.5	FALSE	Adams Plus	10000	
25	489.99	24023	4.8	TRUE	Xbox Series	10000	
26	99.95	89852	4.7	TRUE	NOCO Boos	20000	
27	24.99	39441	4.6	FALSE	ANCEL AD3	10000	
28	9.99	12548	4	FALSE	Gel Nail Po	10000	
29	41.95	36945	4.6	FALSE	nbpure Ma	10000	
30	8.93	59015	4.7	TRUE	Garnier Ski	20000	

6. Queries on Athena

Q: What are the most top-rated products? ⇒ *The goal of our project*

: To retrieve the top 100 most highly rated products based on quantity of actual reviews and sales volume, write query selects all columns from the preprocessed data and orders the results based on the rating points first in descending order, then by qty of reviews and sales both in descending order as well.

// SQL script;

```
✓ Question#1 :  
1 SELECT title, stars, reviews, boughtinlastmonth  
2 FROM glued_data_bd  
3 ORDER BY stars DESC, reviews DESC, boughtinlastmonth DESC  
4 LIMIT 100;
```

// Result for top 100 products in .csv file finds here;

<https://github.com/Duggieeeeey/AWSglue-ETL/blob/main/Result/TopRatedProduct100inAmazonMarket2023.csv>

A: The most highly rated product based on reviews and sales in 2023 was 'The Legend of Zelda: Tears of the Kingdom - Nintendo Switch (US Version)'. ⇒ *The result of our project*

—Athena view—

The screenshot displays the AWS Athena console interface. On the left, the 'Data' sidebar shows the data source 'AwsDataCatalog', database 'my-database-bd', and a list of tables including 'amazon_products_csv', 'glued_data', 'glued_data_bd', and 'my_raw_data_bucket_bd'. The 'glued_data_bd' table is expanded, showing columns like 'price', 'reviews', 'stars', 'isbestseller', 'title', and 'boughtinlastmonth'. The main panel shows a SQL query (Query 10) that selects all columns from 'glued_data_bd' and orders them by 'stars' (DESC), 'reviews' (DESC), and 'boughtinlastmonth' (DESC), with a limit of 100. Below the query, the 'Query results' tab is active, showing a 'Completed' status with a 'Results (100)' table. The first row of the results table shows the title 'The Legend of Zelda: Tears of the Kingdom - Nintendo Switch (US Version)' with a star rating of 4.9. The console also displays query statistics: 'Time in queue: 101 ms', 'Run time: 1.098 sec', and 'Data scanned: 86.57 KB'.

title	stars
The Legend of Zelda: Tears of the Kingdom - Nintendo Switch (US Version)	4.9

—Excel View—

	A	B	C	D
1	title	stars	reviews	boughtinlastmonth
2	The Legend of Zelda: Tears of the Kingdom - Nintendo Switch (US Version)	4.9	15084	10000
3	The Legend of Zelda: Tears of the Kingdom - Nintendo Switch (US Version)	4.9	15084	10000
4	Let's Count Soft Book - World of Eric Carle the Very Hungry Caterpillar Baby on the Go Clip Teething Crinkle Soft Sensory Book	4.9	10476	10000
5	Let's Count Soft Book - World of Eric Carle the Very Hungry Caterpillar Baby on the Go Clip Teething Crinkle Soft Sensory Book	4.9	10476	10000
6	PlayStation DualSense Wireless Controller â€” Midnight Black	4.8	90649	10000
7	PlayStation DualSense Wireless Controller â€” Midnight Black	4.8	90649	10000
8	MEGA BLOKS Fisher-Price Toddler Block Toys, Big Building Bag with 80 Pieces and Storage Bag, Blue, Gift Ideas for Kids Age 1+	4.8	77797	10000
9	MEGA BLOKS Fisher-Price Toddler Block Toys, Big Building Bag with 80 Pieces and Storage Bag, Blue, Gift Ideas for Kids Age 1+	4.8	77797	10000
10	Graco TurboBooster 2.0 Backless Booster Car Seat, Denton	4.8	65909	20000
11	Graco TurboBooster 2.0 Backless Booster Car Seat, Denton	4.8	65909	20000
12	Sassy Stacks of Circles Stacking Ring STEM Learning Toy, Age 6+ Months, Multi, 9 Piece Set	4.8	56107	20000
13	Sassy Stacks of Circles Stacking Ring STEM Learning Toy, Age 6+ Months, Multi, 9 Piece Set	4.8	56107	20000
14	Neutrogena Makeup Remover Wipes Singles, Daily Facial Cleanser Towelettes, Gently Removes Oil & Makeup, Alcohol-Free Ma	4.8	54029	60000
15	Neutrogena Makeup Remover Wipes Singles, Daily Facial Cleanser Towelettes, Gently Removes Oil & Makeup, Alcohol-Free Ma	4.8	54029	60000
16	LEGO Classic Medium Creative Brick Box 10696 Building Toy Set - Featuring Storage, Includes Train, Car, and a Tiger Figure, and	4.8	45455	10000
17	LEGO Classic Medium Creative Brick Box 10696 Building Toy Set - Featuring Storage, Includes Train, Car, and a Tiger Figure, and	4.8	45455	10000
18	Infantino Lil' Nibbles Textured Silicone Teether -Sensory Exploration and Teething Relief with Easy to Hold Handles, Orange Car	4.8	42153	10000
19	Infantino Lil' Nibbles Textured Silicone Teether -Sensory Exploration and Teething Relief with Easy to Hold Handles, Orange Car	4.8	42153	10000
20	NOW Supplements, Vitamin D-3 5,000 IU, High Potency, Structural Support*, 240 Softgels	4.8	40845	20000
21	NOW Supplements, Vitamin D-3 5,000 IU, High Potency, Structural Support*, 240 Softgels	4.8	40845	20000
22	MunchkinÂ® Float & Play Bubblesâ„C Baby and Toddler Bath Toy, 4 Count	4.8	27099	10000
23	MunchkinÂ® Float & Play Bubblesâ„C Baby and Toddler Bath Toy, 4 Count	4.8	27099	10000
24	Xbox Series X	4.8	24023	10000
25	Xbox Series X	4.8	24023	10000
26	Nature Made Folic Acid 400 mcg (665 mcg DFE), Dietary Supplement for Nervous System Function, 250 Tablets, 250 Day Suppl	4.8	18554	10000
27	Nature Made Folic Acid 400 mcg (665 mcg DFE), Dietary Supplement for Nervous System Function, 250 Tablets, 250 Day Suppl	4.8	18554	10000
28	Fine Ground Celtic Sea Salt â€” (1) 16 Ounce Resealable Bag of Nutritious, Classic Sea Salt, Great for Cooking, Baking, Pickling,	4.8	18110	10000
29	Fine Ground Celtic Sea Salt â€” (1) 16 Ounce Resealable Bag of Nutritious, Classic Sea Salt, Great for Cooking, Baking, Pickling,	4.8	18110	10000
30	Ddrops Organic Baby 400 IU 90 Drops - Daily Vitamin D Liquid for Infants. Supports Teeth & Bone Health. No Preservatives, No	4.8	12769	10000

// Further data analysis idea:

Classify the categories of top rated product and visualize with pie chart to get some Amazon e-commerce market insight in 2023

Q: [FEEL FREE TO ADD MORE](#)

//SQL script link here;