

Superstore Sales Analysis using AWS Serverless Analytics Stack

To analyze retail sales performance using AWS services for scalable, serverless, and real-time business insights. The goal is to derive actionable insights such as top-performing regions, sales trends, discount impact, and profit margins.

AWS Services Used:

Service	Purpose
IAM	Secure access control
S3	Data lake for storing raw sales data
AWS Glue	Schema detection and ETL cataloging
Athena	Serverless querying using SQL
QuickSight	Data visualization and dashboarding

Using modern AWS analytics stack for serverless data analysis and each service plays a key role in analyzing something like Superstore sales data.

1. Amazon IAM:

Firstly created new user for this analysis, as we should not perform any kind of analysis and work on root user. So as practice created new user for analysis through Amazon Identity Access Management (IAM) and assign admin role policy for that user.

Have to create custom roles and policies for

1. S3 access
2. Glue Data catalog access
3. Athena Query Execution
4. QuickSight Data Source Access

2. Amazon S3 (Storage Layer)

In this Service, I created object named Orders, inside that created folders like structure in the customised manner (partitioned with Year and month folder)
Uploaded the files accordingly into that folders.

3. AWS Glue

In this service, created Glue crawler to scan the s3 bucket.

It inferred schema automatically:

order_id, order_date, category, sales_price, profit, etc.

It created a table in Glue Data Catalog named as orders.

It enabled partitioning based on year and month (from order_date column)

4. Amazon Athena (Query Engine)

It supports serverless Querying using SQL and after that saved results to S3 for reporting.

5. Amazon QuickSight (Dashboards)

It is used for making interactive dashboards.

It is connected to Athena Dataset and built an interactive dashboard with it.

Such as Total Sales, Profit, Quantity sold, Region-wise performance, Discount vs Profit Impact etc.

Data Analysis using Athena:

1. What is the monthly trend of total sales and profit?

```
SELECT
  extract(month FROM date_parse(order_date, '%c/%e/%Y')) AS month,
  round(SUM(sales),0) AS total_sales,
  round(SUM(profit),2) AS total_profit
FROM orders
GROUP BY 1
ORDER BY 1;
```

2. Which product categories and sub-categories are most profitable?

```
SELECT
  category,sub_category,
  round(SUM(sales),0) AS total_sales,
  round(SUM(profit),2) AS total_profit
FROM orders
GROUP BY 1,2
ORDER BY total_profit desc;
```

3. Which cities or regions are underperforming in terms of profit?

```
SELECT
  region,city,
  round(SUM(profit),2) AS total_profit
FROM orders
GROUP BY 1,2 having sum(profit)<0
ORDER BY total_profit;
```

4. What is the effect of discounts on profit?

```
SELECT
  ROUND(discount, 2)*100 AS discount_level,
  COUNT(*) AS order_count,
  round(SUM(sales),0) AS total_sales,
  round(SUM(profit),2) AS total_profit
FROM orders
GROUP BY 1
ORDER BY 1;
```

5. Who are the top 10 most valuable customers?

```
select customer_id,customer_name, sum(sales) as total_sales,sum(profit) as
total_profit from orders
group by 1,2
order by total_profit desc limit 10
```

Project Outcomes:

It is serverless so no server or cluster to manage and it is fully scalable.
Using Athena, it is pay per query policy and no ETL pipelines required ,So it's cost effective.
QuickSight enables rapid dashboard creation.
Secure data access with IAM and S3 policies.