**Learning AWS through Projects – Day 1**

In the tech field, it’s often said that the best way to learn any tool or language is by building projects around it, and I couldn't agree more! Learning through hands-on projects is one of the most effective ways to truly understand and master a skill.

In this series, I’ll guide you through various AWS services and tools by building projects. Every day, I’ll complete a project, document the process, and share it here. The projects will range from basic to advanced, integrating multiple services to achieve desired outcomes.

Day 1 Project: Data Analysis using Amazon Athena  
Project: Query data stored in Amazon S3 using Amazon Athena.

Steps:

1. Upload a CSV file to S3.
2. Use AWS Glue to catalog the data.
3. Create a SQL query in Athena to retrieve insights from the data.
4. (Optional) Visualize the query results using QuickSight.

Key Services: S3, Glue, Athena, QuickSight.

Here’s a quick overview of the four AWS services used in your project:

**1. Amazon S3 (Simple Storage Service):**

* **Purpose:** Object storage service to store and retrieve any amount of data.
* **Use in Project:** Stores the CSV file that will be queried.
* **Key Features:**
  + Highly scalable and durable.
  + Cost-effective storage for structured and unstructured data.
  + Supports integration with other AWS services like Athena, Glue, etc.

**2. AWS Glue:**

* **Purpose:** Managed ETL (Extract, Transform, Load) service to prepare and load data.
* **Use in Project:** Catalogs the data in S3 and creates metadata for the Athena queries.
* **Key Features:**
  + Automatically discovers and catalogs metadata for your data.
  + Generates ETL code to transform the data.
  + Serverless and scales automatically.

**3. Amazon Athena:**

* **Purpose:** Interactive query service to analyze data directly in S3 using standard SQL.
* **Use in Project:** Executes SQL queries to extract insights from the data stored in S3.
* **Key Features:**
  + No infrastructure to manage (serverless).
  + Pay only for the queries you run (per TB of data scanned).
  + Supports standard SQL.

**4. Amazon QuickSight (Optional):**

* **Purpose:** Scalable business intelligence (BI) service for data visualization.
* **Use in Project:** (Optional) Visualizes the results from the Athena queries in dashboards.
* **Key Features:**
  + Rich set of visualization options.
  + Scales to hundreds of thousands of users without infrastructure management.
  + Pay-per-session pricing model.

Data Source: https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

The file athlete\_events.csv contains 85258 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events). The columns are:

Summer Olympics

1. **ID** - Unique number for each athlete (67474)
2. **Name** - Athlete's name
3. **Sex** - M or F
4. **Age** – Integer (12-71)
5. **Height** - In centimeters
6. **Weight** - In kilograms
7. **Team** - Team name
8. **NOC** - National Olympic Committee 3-letter code (210)
9. **Games** - Year and season (5, 2000-4-8-12-14)
10. **Year** - Integer
11. **Season** - Summer or Winter
12. **City** - Host city
13. **Sport** - Sport
14. **Event** - Event
15. **Medal** - Gold, Silver, Bronze, or NA