R&D Software Engineer - Technical Test

Objective: This test aims to assess your proficiency in Python programming, SQL, AWS, ETL processes, and containerization technologies.

Section 1: Python and Data Structures

1. Implement a function in Python that performs a binary search on a sorted list and returns the index of the target element. If the element is not found, return -1.
2. Implement a function that multiplies two matrices and returns the result. Assume the matrices are given as lists of lists.

Section 2: SQL and BigQuery

1. Write an SQL query that retrieves the top 5 customers with the highest total purchase amount from a table orders with the following structure:

Table: orders

- id (Order identifier)
- customer_id (Customer identifier)
- amount (Purchase amount)

Your query:

2. In Google BigQuery, what is the difference between **partitioning** and **clustering** in a table? Provide a brief explanation.

Section 3: AWS and Containerization

1. Describe how you would deploy a Python application using AWS Lambda and API Gateway. Include details about function packaging and API configuration.
2. Write a Dockerfile to containerize a simple Python application that requires numpy and pandas.

Section 4: ETL and Multithreading

1. Describe an ETL pipeline you have worked on, focusing on the data extraction, transformation, and loading processes.
2. Implement a Python script using multithreading that processes a list of numbers by squaring each number in parallel.

Section 5: CI/CD and DevOps

1. Briefly explain how a CI/CD pipeline works and how Jenkins can be used for automation.
2. What are the advantages of using Git for version control in software development?
