AsiansGroup

Assessment for Application Support Developer Role

You are given reports of intermittent performance issues from a customer’s website. Your task is to analyze simulated logs, debug a script, query a database, and propose a solution—all while demonstrating your ability to monitor, script, and problem-solve effectively.

Provided Materials

1. **nginx_access.log**: A sample log file with 7200 entries.
2. **python_monitor.py**: A partially broken Python script meant to analyze logs.
3. **traffic.db**: A SQLite database file with a table request_logs containing fields: timestamp, ip_address, status_code, response_time_ms, bytes_sent.
4. **Instructions**: This document with tasks and deliverables.

These can be accessed from our GitHub Repository.

Tasks

1. Nginx Log Analysis

Analyze the provided nginx_access.log file (Go through some initial logs to understand the pattern). Write a Python script to:

- Identify the **top 5 IP addresses** by request count.
- Calculate the percentage of requests with status codes in the 400-599 range.
- Find the average response size in bytes for *GET* requests.

Output: The script should generate a readable output (console or text file).

2. Debugging the Python Script

The provided **python_monitor.py** script is meant to monitor a log file and alert if error rates exceed 10% in a 5-minute window. However, it contains intentional bugs. Your task is to:

- **Fix the script** and ensure it works correctly.
- **Write a short explanation** describing the bugs you identified and how you fixed them.

3. Attack Pattern Detection

Write a **Bash or Python script** to detect potential attack patterns in the logs. Your script should analyze the logs and flag suspicious IPs or behaviors.

Output: The script should generate a summary of potential attack patterns, including flagged IPs, abnormal request frequencies, or other detected anomalies.

4. SQL Querying (Database Analysis)

Using the provided **traffic.db** SQLite database, write SQL queries to:

- **Find the hour of the day** with the **highest average response time**.
- **Identify any IPs** that sent **more than 100 requests** with a **429 status code** (rate-limited).
- **Calculate the total bytes sent** for requests where **response time > 500ms**.

Output: Submit both the SQL queries and their results.

5. (Bonus) CDN Performance Optimization

Based on your findings in Tasks 1-3, suggest **one mitigation step** to improve CDN performance for this customer. Keep it concise (2-3 sentences).

Deliverables

Candidates must submit:

1. All scripts (.py, .sh, .sql files).
2. A results.txt file containing:
    - Output from Task 1 (Top IPs, error percentage, avg response size).
    - Output from Task 3 (Potential attacking IPs).
    - Output from Task 4 (SQL query results).
    - Any additional observations.
3. A short write-up (1-2 pages) covering:
    - Bug fixes for Task 2.
    - Explanation of attack detection script (Task 3).
    - Optional CDN optimization recommendation (Task 5).

To submit this assessment, create a fork of the GitHub Repository and issue a Pull Request with your solutions, under a dedicated folder of your full name (hiphenized). Additionally, send an email reply to your assessment notifying about the completion, along with the link to the Pull Request.