

1. Problem

Copy the data in the file 'web-server-access-log.txt.gz' to the table 'access_log' in the PostgreSQL database 'template1'.

The file is available at the location : "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DB0250EN-SkillsNetwork/labs/Bash%20Scripting/ETL%20using%20shell%20scripting/web-server-access-log.txt.gz".

The following are the columns and their data types in the file:

- a. timestamp - TIMESTAMP
- b. latitude - float
- c. longitude - float
- d. visitorid - char(37)

and two more columns: accessed_from_mobile (boolean) and browser_code (int)

The columns which we need to copy to the table are the first four columns : timestamp, latitude, longitude and visitorid.

NOTE: The file comes with a header. So, use the 'HEADER' option in the 'COPY' command.

The problem may be solved by completing the following tasks:

Task 1: Start the Postgres server.

Task 2: Create the table.

Create a table named access_log to store the timestamp, latitude, longitude and visitorid.

Task 3. Create a shell script named cp-access-log.sh and add commands to complete the remaining tasks to extract and copy the data to the database.

Task 4. Download the access log file.

Task 5. Unzip the gzip file.

Task 6. Extract required fields from the file.

Task 7. Redirect the extracted output into a file.

Task 8. Transform the data into CSV format.

Task 9. Load the data into the table access_log in PostgreSQL

Task 10. Execute the final script.

Task 11. Verify by querying the database.