Problem Statement

There are 3 tasks in this technical challenge:

- Task 1: Provide a set of SQL scripts to generate a synthetic EHR dataset (in any SQL-based environment) based on the common data model version 5.4. The steps and processes are in the repositories listed below. Please also add steps to:
    - Create 2 users: user1 and readonly. The first user should have write and read access to all the schemas and tables, whereas readonly should only have read access
    - Create a query that joins at least 2 tables and uses at least 1 WHERE clause, analyze it and explain how to improve its performance

- Task 2: Design a basic python package to query the synthetic EHR dataset. The package needs to have the following requirements:
    - Should not have the credentials to connect to the DB hardcoded. The package should be able to read the credentials from the user system.
    - A function that receives as input an SQL query and outputs the result as a pandas data frame. i.e get_query("SELECT * FROM person")
    - A function that queries the person table with a list of patients ids as a parameter and outputs the result as a pandas data frame. i.e get_person([1000, 2000, 3000])
    - The above functions need to have proper documentation and the appropriate unit tests

- Task 3: Please upload your python package to git and provide a link to the repo. The repository needs to have GitHub actions to build and test the package

References:

- https://ohdsi.github.io/ETL-Synthea/
- https://github.com/synthetichealth/synthea
- https://ohdsi.github.io/TheBookOfOhdsi/