

Julia acabou de concluir a realização de uma competição de programação e ela precisa da sua ajuda para montar o quadro de líderes!

Construa um pipeline de dados que leia os arquivos de exemplo que foram fornecidos, trate os dados lidos e faça a carga em uma base de dados relacional.

Observe os formatos dos arquivos e o conteúdo de cada um.

A base de dados relacional deverá ser modelada e criada por você. Modele da forma que julgar mais eficiente.

Com os dados carregados na base (ETL em python), escreva uma consulta em SQL para imprimir o hacker_id e o nome dos hackers que alcançaram pontuações máximas em mais de um desafio.

Ordene a saída em ordem decrescente pelo número total de desafios nos quais o hacker obteve pontuação máxima. Se mais de um hacker obteve pontuações máximas no mesmo número de desafios, então ordene-os em ordem crescente pelo hacker_id.

Requisitos: O candidato deverá realizar o teste utilizando as tecnologias: Python, SQL e Docker;

Bônus 1: O candidato deverá fazer o deploy em um serviço de livre escolha da Google Cloud;

Bônus 2: Crie uma arquitetura de DW / Lake / Data Mesh e sua respectiva modelagem;

Bônus 3: O Candidato deverá disponibilizar endpoint de consumo dos dados analíticos;

#####

Os arquivos fornecidos contém a seguinte estrutura além dos dados da competição:

Hackers: O hacker_id é o identificador do hacker, e o nome é o nome do hacker.

hacker_id : Integer
name : String

Difficulty: O difficult_level é o nível de dificuldade do desafio e score é a pontuação do desafio para o nível de dificuldade.

difficulty_level : Integer
score : Integer

Challenge: O challenge_id é o identificador do desafio, o hacker_id é o identificador do hacker que criou o desafio, e difficulty_level é o nível de dificuldade do desafio.

challenge_id : Integer
hacker_id : Integer
difficulty_level : Integer

Submission: O submission_id é o identificador da submissão, hacker_id é o identificador do hacker que fez a submissão, challenge_id é o identificador do desafio ao qual a submissão pertence, e score é a pontuação da submissão.

submission_id: Integer
challenge_id : Integer
hacker_id : Integer
score : Integer

Artefatos que devem ser entregues:

- 1 - saída do resultado esperado
- 2 - código do pipeline de dados
- 3 - código de criação da base de dados
- 4 - código da consulta para obtenção do resultado
- 5 - dockerfile
- 6 - documentação técnica do teste