

Imagine que você é um engenheiro de dados em uma startup de tecnologia pet chamada “UOLCatLovers”. A UOLCatLovers está desenvolvendo um aplicativo móvel que fornece fatos interessantes sobre gatos para seus usuários. Os fatos são extraídos da API Cat Facts.

Documentação em: <https://alexwohlbruck.github.io/cat-facts/docs/>

As respostas devem constar em um repositório no GitHub e o link do repositório deve ser compartilhado para a avaliação.

1 - Como a startup foi recém criada, ainda não há uma grande demanda pelos dados, então você precisa desenvolver um script Python simples que extraia os dados de fatos sobre gatos (cat facts) da API e salva em um arquivo CSV local.

2 - Com o tempo, o aplicativo ganhou popularidade e o número de fatos sobre gatos cresceu exponencialmente. Agora, a solução local não é mais viável e é necessário transpor a solução para a nuvem. Você precisa projetar uma arquitetura na plataforma Google Cloud que seja capaz de extrair, armazenar e disponibilizar os dados para os times de analytics. Não é necessário implementar ou codificar, apenas desenhar a arquitetura.

3 - Com o tempo, o time de analytics também sentiu necessidade de realizar suas próprias consultas sobre os dados, como a tecnologia mais conhecida por eles é o BigQuery, você precisa especificar o esquema da tabela de dados de fatos sobre gatos (cat facts), inclua os campos, tipos de campos e quaisquer outras considerações necessárias. A especificação pode ser feita por diagrama ou por código.

Nota: para as questões a seguir, não é necessário criar uma base de dados no BigQuery ou mostrar os resultados das consultas. Basta ter o código SQL escrito.

4 - Apesar dos dados no BigQuery, o time de analytics não está conseguindo realizar as consultas por conta própria e pediu sua ajuda. Crie uma consulta que extraia os fatos que foram atualizados em agosto de 2020 para um estudo de caso demandado por eles.

5 - O time de desenvolvimento soube da sua habilidade com consultas SQL e decidiu se aproveitar da fila de demandas para solicitar uma amostra da base de fatos sobre gatos (cat facts) para popular o ambiente de QA deles. O time solicitou uma consulta SQL que extraia, de forma aleatória, 10% dos registros da base contendo as informações de texto, data de criação e data de atualização. Uma consideração importante feita pelo time de desenvolvimento é que eles precisam da consulta SQL para extrair os dados para um arquivo CSV separado por vírgulas.