

CANDIDATA: Isabel Priscila Chaves da Silva Rebello

RESOLUÇÃO DESAFIO BEMOL: PARTES 3 E 4

Parte 3

Pergunta 1

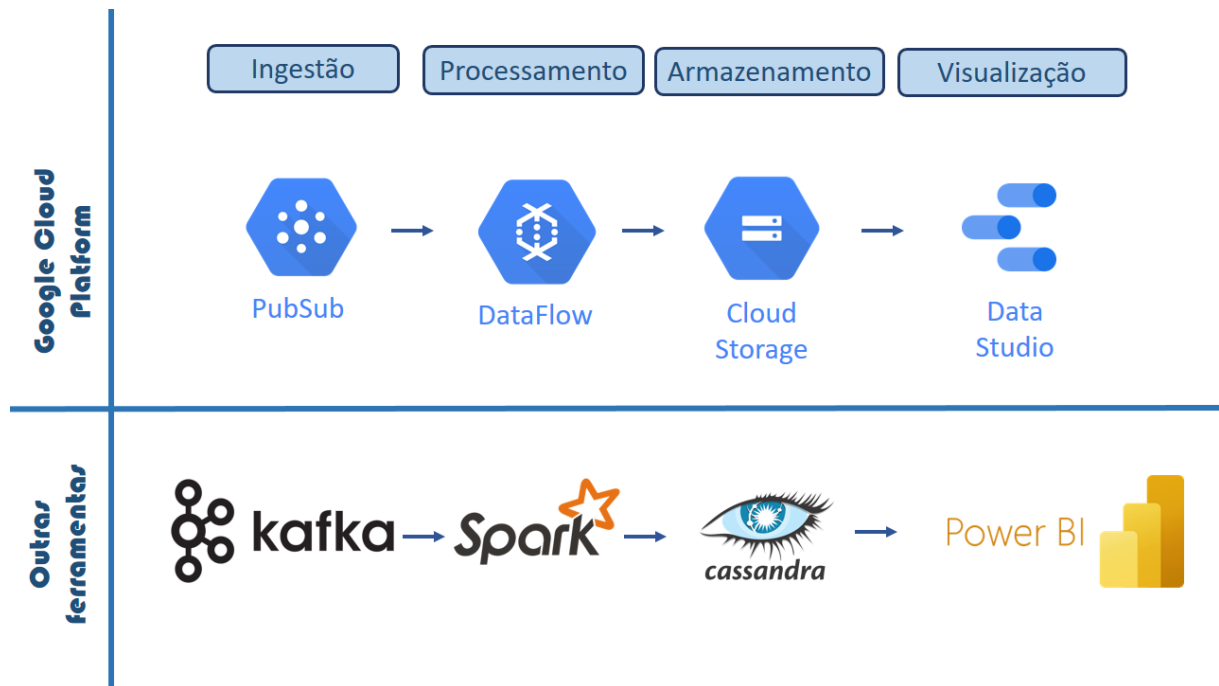
Resposta: Considerando que a empresa mantém uma tabela no banco de dados contendo uma ampla gama de informações, algumas das quais podem ser sensíveis, é possível criar uma visualização (view) dessa tabela, filtrando apenas os dados necessários para as operações diárias e garantindo que não inclua informações sensíveis. Essa view pode então ser disponibilizada para os demais funcionários da empresa. Esse procedimento assegura um gerenciamento correto de acessos dentro da empresa e preserva a segurança dos dados.

Pergunta 2

Resposta: A solução mencionada na pergunta anterior pode ser útil neste caso, porém uma alternativa seria adotar a computação em nuvem, que oferece uma ampla gama de ferramentas para lidar com grandes volumes de dados. Uma dessas ferramentas permite que o responsável pela gestão dos dados especifique o acesso de cada usuário. Isso garante que apenas as pessoas necessárias tenham acesso às informações, garantindo a segurança dos dados sensíveis e reduzindo o risco de perda ou roubo. É importante destacar que a plataforma de computação em nuvem escolhida deve ser escalável, segura, confiável e eficiente.

Parte 4

Sugiro duas arquiteturas: uma que utilize computação em nuvem, aproveitando suas vantagens na adaptação a variações de demanda, facilidade de implementação, segurança, confiabilidade e integração entre ferramentas. A segunda opção seria adotar ferramentas de código aberto, que são gratuitas, complementadas pelo Power BI, uma ferramenta de visualização de baixo custo.



Explicando a escolha de cada item:

- Google Cloud Pub/Sub ingestão de dados em tempo real devido à sua escalabilidade, durabilidade e baixa latência. Pub/Sub permite configurar tópicos para diferentes tipos de dados, o que facilita a organização;
- Google Cloud Dataflow processa dados em tempo real ou em lote. Dataflow permite criar pipelines de dados escaláveis e paralelas, facilitando o processamento eficiente dos dados em diferentes cenários;
- Google Cloud Storage oferece armazenamento de objetos altamente escalável e durável, permitindo organizar os dados em buckets de acordo com a necessidade do projeto;
- Google Data Studio possui integração com os serviços da GCP e facilidade de uso. Além de permitir conexão com os dados armazenados no GCS ou em outros serviços da GCP, oferecendo uma maneira conveniente de visualizar e analisar os dados na plataforma;
- O Apache Kafka é um ótimo sistema de ingestão de dados em tempo real, sendo utilizado por diversas grandes empresas. O Kafka é capaz de lidar com grandes volumes de dados, oferece alta disponibilidade e tolerância a falhas;
- Apache Spark possui capacidade de lidar com grandes conjuntos de dados e processamento em escala, além de oferecer suporte a vários tipos de processamento;
- Apache Cassandra tem alta disponibilidade, é extremamente tolerante a falhas e lida muito bem com cargas de trabalho de alta velocidade;

- Power Bi como ferramenta de visualização, pela sua facilidade de uso, análise em tempo real para dados e dashboards personalizáveis.