**Fase 3 – Arquitetura (pergunta bônus):**

Considere agora que você tenha que montar uma arquitetura em cloud, idealmente na Google Cloud Platform, para realizar de forma recorrente as tarefas acima citadas.

Descreva a arquitetura proposta. Quais soluções/serviços/tecnologias você utilizaria? Essa questão é opcional. Fique à vontade para respondê-la caso você se sinta confortável.

Quando se pensa em arquitetura de dados o contexto da empresa, desde o modelo de negócios, conhecimento dos profissionais atuais da empresa, conhecimento dos profissionais disponíveis no mercado, orçamento e desafios técnicos que a empresa tem de enfrentar.  
  
Especificamente a respeito dos desafios técnicos podemos levantar algumas questões como:  
volume de dados, tipo de processamento batch ou streaming, velocidade necessária de retorno de consulta aos dados (se é obrigatório baixa latência), nível de disponibilidade dos dados, se existem momentos de picos onde a elasticidade seria importante, necessidade de replicação de dados seja por atender com mais velocidade usuários de diferentes regiões, por tolerância a falhas ou ainda aumentar a disponibilidade, se faz mais sentido utilizar aplicações serverless ou com servidor para gerenciar (sendo necessário avaliar a capacidade atual dos profissionais da empresa e/ou do mercado para cuidar de tais servidores bem como orçamento necessário para esses casos).  
  
Acima foram citados apenas algumas pontos que poderiam ser considerados, mas uma análise real ainda iria bem mais além disso. Aqui citarei algumas possibilidades de arquitetura que acho interessante de acordo com alguns contextos.

**ORQUESTRAÇÂO DAS TASKS**

Para o exercício realizado acima, eu utilizaria o **Cloud Composer** para fazer a orquestração dos Jobs em SQL, já que o Airflow é uma ferramenta extremamente popular, de fácil aprendizado e de interface intuitiva, o facilita a utilização por profissionais de tecnologia. Uma outra possibilidade seria utilizar o Airflow hospedado em uma máquina virtual **GCE,** mas nesse caso é um importante salientar que não seria um serviço totalmente gerenciado como no caso do Cloud Composer. Para a escolha da melhor opção seria interessante estipular uma carga de utilização de recursos atual e em um médio prazo para calcular aproximadamente os custos e seus prós e contras (totalmente gerenciado ou não, maior disponibilidade ou não).

**ARMAZENAMENTO**

**CloudSQL** caso se tratasse de dados que dificilmente passariam de 30TB a médio prazo (limite do CLoudSQL) e caso não fosse necessário replicação global desses dados para atender usuários. Por outro lado, caso os dados pudessem passar de 30TB em um futuro de médio prazo ou ainda fosse necessário um banco globalmente distribuído eu utilizaria o **Cloud Spanner** já que o mesmo poderia atender esses requisitos. Uma outra possibilidade ainda seria o **Bigquery**, que permitiria consultas mais rápidas do que CloudSQL**,** um ponto de atenção para essa solução seria a cobrança por query rodada, caso uma query seja muito “pesada” poderia aumentar consideravelmente o custo.

**NECESSIDADE DE CAMADA RAW/WORKED**

Em alguns casos opta-se por salvar uma cópia RAW (dados sem tratamento) dos dados em um arquivo em um sistema de armazenamento e posteriormente dados WORKED (após tratamento), para depois realizar a carga em um banco de dados. Esse sistema faz com que facilmente seja possível o reprocessamento até mesmo em casos que o dado não esteja mais disponível na sua fonte. Nesse caso eu utilizaria como armazenamento desses arquivos o GCS, **Google Cloud Storage** por oferecer armazenamento de arquivo com um excelente custo e podendo ainda definir políticas de ciclo de vida (alteração automática entre as classes de armazenamento) e excelentes controles de acesso e segurança, como políticas IAM e criptografia de dados.

**FERRAMENTA PARA VISUALIZAÇÃO**  
Por fim, como ferramenta de visualização de dados optaria pelo Looker, que é uma excelente ferramenta de Self-service BI e que se integra muito bem com os outros serviços do Google Cloud.