

Os Pilares da Solução no Google Cloud (O que Pedir e Discutir)

Esta é a lista de produtos que você deve apresentar como os componentes da sua arquitetura. Cada um resolve uma parte específica do nosso sistema.

a) Para Hospedar a Aplicação (O "Motor"): Cloud Run

- **O que é:** Um serviço "serverless" para executar nossos contêineres Docker.
- **Por que usar:** É a solução mais moderna e custo-efetiva. Você simplesmente envia a imagem Docker que construímos, e o Google cuida de todo o resto. Ele escala automaticamente (se muitos usuários acessarem, ele cria mais instâncias; se ninguém usar, ele pode escalar até zero, economizando muito dinheiro).
- **O que dizer:** "Planejamos empacotar nossa aplicação Flask em um contêiner Docker e executá-la no **Cloud Run** para aproveitar a escalabilidade automática e o gerenciamento simplificado."

b) Para o Banco de Dados Relacional (O "Arquivo Central"): Cloud SQL for PostgreSQL

- **O que é:** Um serviço de banco de dados PostgreSQL totalmente gerenciado pelo Google.
- **Por que usar:** É a nossa migração do PostgreSQL local para a nuvem. O Google cuida de backups, segurança, atualizações e disponibilidade. É a escolha profissional que garante que seus dados estejam seguros e sempre acessíveis.
- **O que dizer:** "Para nosso banco de dados principal, utilizaremos o **Cloud SQL for PostgreSQL**, aproveitando a robustez do Postgres com a confiabilidade e a manutenção gerenciada pelo Google."

c) Para os Documentos dos Casos (O "Arquivo Físico"): Google Cloud Storage (GCS)

- **O que é:** Um serviço de armazenamento de objetos infinitamente escalável.
- **Por que usar:** Não podemos salvar os arquivos de upload (PDFs, imagens) dentro do contêiner. O GCS é o lugar certo. É seguro, barato e feito para armazenar terabytes de dados. Cada arquivo terá um caminho que incluirá o tenant_id para garantir o isolamento (ex: bucket-advocacia-ia/escritorio_123/caso_456/documento.pdf).
- **O que dizer:** "Todos os documentos enviados pelos usuários serão armazenados de forma segura no **Google Cloud Storage**, com uma estrutura de pastas que isola os dados de cada tenant."

d) Para a "Memória" da IA (A Busca Vetorial): Vertex AI Vector Search

- **O que é:** Um banco de dados vetorial de altíssima performance, totalmente gerenciado.
- **Por que usar:** O ChromaDB, que usamos localmente, é ótimo para desenvolvimento, mas não para um SaaS escalável. O Vertex AI Vector Search é a solução de produção do Google para buscas por similaridade (RAG). É o que permitirá que a IA encontre rapidamente os documentos relevantes em meio a milhares de casos. Os vetores de cada tenant serão indexados com metadados (tenant_id, case_id) para garantir o isolamento.
- **O que dizer:** "Para a funcionalidade de RAG (busca em documentos), substituiremos nosso ChromaDB local pelo **Vertex AI Vector Search** para garantir buscas semânticas rápidas e escaláveis, mantendo o isolamento dos dados vetoriais de cada tenant."

e) Para a Autenticação dos Usuários (A "Portaria"): Cloud Identity Platform ou Firebase Authentication

- **O que é:** Um serviço gerenciado para identidade de usuários.

- **Por que usar:** Embora tenhamos construído nosso próprio sistema, usar um serviço gerenciado do Google para isso é muito mais seguro e robusto. Ele cuida de tudo: registro, login, redefinição de senha, autenticação de dois fatores, logins sociais (com Google, etc.), e se integra perfeitamente ao resto do ecossistema.
- **O que dizer:** "Para a autenticação, planejamos utilizar o **Cloud Identity Platform** para gerenciar nossos usuários de forma segura e escalável."