

Relatorio de Analise de Carreira

Mariana Vasconcelos Pinto -> Engenheira de Dados Pleno - Fintech de Crédito CreditoJá

72

Bom

Mariana tem experiência sólida em pipelines ETL em produção, modelagem dimensional e BigQuery - exatamente o núcleo técnico da vaga. Seu diferencial é a certificação GCP Professional Data Engineer e a documentação de 80+ tabelas em catálogo de dados. O gap crítico é a falta de experiência comprovada com dbt Core em produção (menciona 'básico') e ausência de projetos com dados financeiros, que é diferencial importante para uma fintech.

PRÓXIMO PASSO

Nos próximos 7 dias, publique no GitHub um mini-projeto dbt + Great Expectations com 3-4 tabelas BigQuery e testes automatizados - isso elimina o maior gap ('dbt básico') e demonstra preparo concreto para a vaga.

PONTOS FORTES

- Experiência comprovada com BigQuery e modelagem dimensional (Kimball) em ambiente de 15M eventos/dia - exatamente a escala que a fintech demanda
- Certificação Google Cloud Professional Data Engineer (2023) alinhada com a stack GCP da vaga
- Documentação e organização de catálogo de dados (80+ tabelas) - conhecimento prático que transfere direto para DataHub
- Automação de ingestão de dados com Python/APIs e colaboração multi-time (produto, marketing) - habilidades de comunicação técnica valorizadas em fintechs

GAPS CRÍTICOS

- dbt Core em produção: currículo indica apenas 'dbt (básico)' - vaga exige mínimo 1 ano produtivo. Precisa detalhar qualquer uso real ou fazer projeto demonstrativo urgentemente
- Nenhuma experiência em dados financeiros ou de crédito mencionada - diferencial importante para fintech. Histórico em marketplace e consultoria não toca setor financeiro
- Testes de qualidade de dados: não menciona Great Expectations ou dbt tests em produção - é requisito obrigatório
- Spark 'básico' enquanto vaga demanda processamento acima de 100GB/dia - precisa aprofundar ou evidenciar projetos Spark em escala

HABILIDADES MAPEADAS

dbt Core em produção

Testes automatizados de dados (Great Expectations / dbt tests)

Arquitetura Lakehouse (BigQuery / Delta Lake)

Spark para grandes volumes (100GB+/dia)

BigQuery

Python avançado para ETL

Apache Airflow

Modelagem de dados / Dimensional

Catálogo de dados (DataHub)

Streaming (Kafka / Pub/Sub)

Experiência com dados financeiros

Inglês intermediário escrito

PLANO DE MELHORIA PRIORITÁRIO

1. Dominar dbt Core em produção com testes automatizados (Great Expectations)

Faça um projeto real ou case study: configure um pipeline dbt em BigQuery com 3-4 tabelas, implemente dbt tests (relationships, not_null, unique) e Great Expectations para validação de qualidade. Publique no GitHub com documentação. Tempo: 2-3 semanas. Recurso: dbt Learn Academy (gratuito) + Great Expectations docs. Isso transforma 'básico' em 'produtivo' e preenche gap crítico.

2. Aprofundar Spark em escala (100GB+) e documentar case de otimização

Crie um projeto no Google Cloud (free tier) processando dataset > 100GB com Spark (Dataproc): leia parquet/CSV,

aplique transformações, salve em BigQuery. Compare performance vs pandas. Documente no GitHub com métricas (tempo, custo). Referência: Spark by Examples ou documentação oficial Apache Spark. Tempo: 2 semanas. Isso evidencia 'Spark profundo' não 'básico'.

3. Explorar dados financeiros via projeto ou certificação especializada

Opção 1: Baixe dataset público de crédito (Kaggle: lending_club) e construa um data lakehouse simulado com modelagem para análise de risco. Opção 2: Course 'Data Engineering for Financial Services' (Coursera ou DataCamp) - ~40h. Mencione aprendizado sobre compliance, KYC, AML na descrição de experiências. Tempo: 3-4 semanas. Diferencial forte para fintech.

4. Conhecer DataHub (ou Collibra) para catalogação de dados

DataHub possui documentação open-source e Docker. Instale localmente, crie um catálogo simulado com 10-15 tabelas (reuse seu projeto Spark/dbt), documente o processo. Tempo: 1 semana. Alternativa: DataCamp tem micro-curso sobre data cataloging. Demonstra expertise além do essencial.

DICA DE APRESENTAÇÃO

Use este texto ao se candidatar no LinkedIn ou por e-mail:

Mariana, sua experiência com modelagem Kimball em 15M eventos/dia e a certificação GCP são exatamente o que essa fintech precisa para evoluir o data lakehouse. O que me atrai na vaga é a oportunidade de aprofundar em dbt em produção com testes automatizados - área onde estou desenvolvendo expertise em projetos recentes. Tenho documentado 80+ tabelas em catálogo de dados e estou familiarizada com arquitetura de dados de escala, pronto para gerar impacto rápido em decisões críticas.