

Principal AI & Data Engineer

Renan de Lima Andrade

Resumo

Engenheiro Sênior de IA e Dados com mais de 12 anos de experiência em projetos de plataforma de dados em escala corporativa, soluções de IA Generativa e automação inteligente. Especializado na construção de sistemas de IA prontos para produção utilizando LLMs, Agentes de IA, arquiteturas RAG, MLOps e plataformas de dados modernas em multi-cloud (AWS, Azure, GCP, Huawei, OCI).

Ampla experiência no projeto de arquiteturas de dados escaláveis utilizando Databricks, Spark, Airflow, DBT e Python, integrando Machine Learning, Retrieval-Augmented Generation (RAG), bancos de dados vetoriais, IA baseada em agentes e frameworks de orquestração de LLMs. Sólida trajetória na liderança de iniciativas de transformação digital, modernização de sistemas legados e entrega de produtos de IA que reduzem custos operacionais, aprimoram a tomada de decisões e automatizam fluxos de trabalho de negócios complexos. Experiência em governança corporativa, conformidade com LGPD/GDPR, observabilidade, CI/CD e arquiteturas nativas em nuvem.

Experiência

Janeiro de 2025 – Fevereiro de 2026

Senior AI & Data Engineer - IBM

Atuei na modernização da arquitetura de dados de uma empresa de varejo digital, estruturando um Data Warehouse escalável e preparando o ambiente para iniciativas de Machine Learning. Utilizei Spark para processamento distribuído nas camadas Bronze, Silver e Gold (Medallion) com dados provenientes de ERP, CRM, e-commerce e gateways de pagamento. Modelei fatos e dimensões seguindo Kimball, com SCD Tipo II para clientes e padronização de métricas como CAC, LTV e churn. Criação de agentes com N8N.

Evoluí a arquitetura para suportar feature engineering e versionamento de modelos com MLflow, integrando experimentos de ciência de dados ao pipeline de engenharia. Desenvolvi aplicações interativas com Streamlit para visualização de métricas de negócio e monitoramento de qualidade de dados. A orquestração foi feita com Airflow e as transformações com DBT, garantindo governança, rastreabilidade e testes automatizados. Redução de 65% no tempo de geração de relatórios e 40% nas consultas diretas ao banco transacional.

Dezembro de 2022 – Janeiro de 2025

Consultor/Engenheiro de Dados - Nexus Tech IT

Fui responsável pelo desenho e implementação de uma arquitetura Data Lakehouse com AWS, voltada ao processamento de dados financeiros, com foco em escalabilidade,

qualidade, segurança e conformidade regulatória. Estruturei o Amazon S3 como camada central de armazenamento e implementei a arquitetura Medallion, organizando os dados nas camadas Bronze, Silver e Gold. Desenvolvi pipelines distribuídos com Apache Spark no Amazon EMR, realizando ingestão, transformação, enriquecimento e consolidação de grandes volumes de dados. Utilizei o AWS Glue para processos de ETL, catalogação de metadados e integração com o AWS Glue Data Catalog, disponibilizando conjuntos de dados para consultas analíticas com Amazon Athena. A orquestração dos pipelines foi realizada com Apache Airflow, incluindo controle de dependências, monitoramento de SLA, tratamento de falhas e alertas automáticos. Também empreguei serviços como AWS Lambda, Amazon EventBridge e Amazon CloudWatch para automações, observabilidade, registro de logs e monitoramento operacional. Implementei práticas de governança e SI utilizando AWS Lake Formation, IAM, AWS KMS e políticas de acesso baseadas no princípio do menor privilégio, assegurando rastreabilidade, proteção de dados sensíveis e aderência à LGPD. Para garantir confiabilidade, desenvolvi validações de qualidade, integridade e consistência dos dados ao longo dos pipelines. Integrei o MLflow ao fluxo de engenharia para rastreamento de experimentos, versionamento de modelos e gerenciamento do ciclo de vida de soluções preditivas, facilitando a colaboração entre engenheiros e cientistas de dados.

Julho de 2021 – Novembro de 2022

Engenheiro de Dados Sênior - Global Hitss (Claro)

Atuei na melhoria da arquitetura de dados para suporte a campanhas e ações de marketing da Claro. O ambiente envolvia dados on-premise e em nuvem, com processamento batch e streaming. Utilizei Spark para ETL de dados estruturados e não estruturados, integrando com Kinesis e S3. A redução de custos foi obtida com a migração de clusters EC2 para arquiteturas serverless. Desenvolvi um sistema de NLP com Python (NLTK e spaCy) para análise de satisfação de clientes, integrando os resultados a pipelines de feature engineering com MLflow. Para anonimização de dados sensíveis, utilizei AWS Glue e DataBrew. Criei uma aplicação Streamlit para visualização dos resultados das campanhas e intenções de marketing.

Maio de 2019 – Julho de 2021

Engenheiro de Dados - Magna Sistemas (Secretaria da Educação SP)

Desenvolvi algoritmo para predição de evasão escolar utilizando técnicas de Machine Learning. O projeto envolveu raspagem de dados, integração com fontes legadas e criação de Data Marts. Utilizei Spark para processamento dos dados no ambiente Azure (Data Factory, Synapse, SQL Server). Modelei features e versionamento com MLflow, garantindo rastreabilidade dos experimentos. Desenvolvi dashboards interativos no Power BI e uma aplicação Streamlit para consulta rápida de indicadores educacionais, utilizada em reuniões com o Secretário de Educação.

Fevereiro de 2016 – Fevereiro de 2019

Engenheiro de Dados Pleno - Data Self

Maio de 2014 – Abril de 2016

Analista de BI / Big Data - CSU Cardsystem

Maio de 2011 – Junho de 2013

Voice Recognition Developer - Telemidia & Technology International

Artificial Intelligence

- Large Language Models (GPT-5, GPT-4o, Claude 4, Gemini 3, Llama 4, DeepSeek, Qwen)
- AI Agents; Agentic AI; Multi-Agent Systems
- RAG; Hybrid Search; Semantic Search
- Vector Databases; Embeddings
- Prompt Engineering; Context Engineering
- AI Evaluation; LLMOps
- AI Guardrails; AI Automation
- LangChain; LangGraph
- AutoGen; MCP; Bedrock
- OpenAI API; Anthropic API
- Google AI Studio; Azure OpenAI

Formação

Bacharelado Ciências de Dados – (Jul/2024 – Jul/2027)

Certified AWS Cloud Practitioner CLF-C02 – (Jul/2024)

Técnico Ciência de Dados – (Abr/2020 – Out/2020)

Unifal Brasileiro - Analise e Desenvolvimento de Sistemas (2013 - 2015)

Auditor ISO 27001, SI – (Jan/2012 – Jun/2012)

Habilidades

SQL | Python | Spark | Apache Beam | MLflow | Streamlit

MultiCloud (AWS, GCP, Azure, Huawei, OCI)

Ecossistema Hadoop | Airflow | DBT | Docker | Terraform

Governança de Dados | LGPD | GDPR | Modelagem de Dados

Comunicação | Organização | Autodidata | Iniciativa e Liderança

Entre outras...

Ferramentas

Streamlit – Aplicações analíticas e de monitoramento

Apache Beam – Processamento unificado batch/streaming

Spark – Processamento distribuído e engenharia de features

MLflow – Versionamento, experimentação e deploy de modelos

N8N – Agentes e soluções de IA + Automações

Entre outras...

Contato

CNPJ: 55.030.238/0001-51

Rua Purus, 294 – AP 410 – Vila Mazzei – Tucuruvi / SP

CEP: 02308040

Cel./WhatsApp: 11917321380

E-mail: renan.pyd@gmail.com

LinkedIn: [in/limarenanandrade/](https://www.linkedin.com/in/limarenanandrade/)

GitHub: <https://github.com/renanpyd>