

IBISEN DE BRITO GONÇALVES

Engenheiro de Dados

Tel: (11)91264-2832

Email: ibisenit@gmail.com

Linkedin: <https://linkedin.com/in/ibisenbrito>

Reside em: São Paulo
Brasileiro, 40 anos.

Habilidades e destaques:

- Microsoft Azure
- Databricks
- Apache Spark
- PySpark
- Python
- SQL
- Spark DataFrames
- Delta Tables

Formação Acadêmica:

8º Semestre de Engenharia Elétrica com
foco em Eletrônica - data de formação
Dez/2025

Certificações:

- Microsoft Azure Data Engineer Associate DP-203;
- Microsoft Azure Fundamentals AZ-900;
- Microsoft Azure AI Fundamentals AI-900;

Idiomas:

Inglês – Intermediário

Cursos Específicos:

- Curso Intensivo de Python e Ciência de Dados
- Big Data & Analytics
- Business Intelligence (BI)
- Microsoft Power BI

Cursos Stack Academy - Felipe Santana

- AutoML com Pycaret
- Visualização de Dados com Power BI
- Engenharia de Dados do Zero
- Data Pipelines com Azure

Cursos Cloud Expert - Josué Vidal

- Engenheiro Cloud

Soft Skills

- Comunicação
- Colaboração
- Mentalidade Analítica
- Solução de problemas
- Mentalidade de Negócios
- Criatividade
- Auto-aprendizagem
- Flexibilidade
- Resiliência
- Disciplina

Hard Skills

- Azure Data Factory
- Azure SQL
- Azure Data Lake
- Azure Synapse
- Microsoft Power BI
- Framework Stream Lit

Resumo Profissional:

Engenheiro de dados há mais de 4 anos com fortes conhecimentos nas linguagens e ferramentas: Python, SQL, Microsoft Azure, Azure Data Factory, Azure Data Lake Gen2, Databricks, Linux, Shell Script, Apache Spark, PySpark

Experiências Profissionais:

Empresa: Act Digital – atual. Cargo: Engenheiro de Dados

Out/2024 – Atual

Senior Data Engineer com experiência em projetos de alta performance na nuvem AWS, focado em soluções de dados para o setor bancário. Atuo na construção e manutenção de pipelines de dados robustos e escaláveis, utilizando tecnologias de ponta como Python, SQL, Java Spring Boot, Kafka, Apache Flink, Debezium, PySpark e Spark Streaming.

Responsabilidades: Desenvolvimento e manutenção de pipelines de dados batch e streaming, utilizando PySpark e Spark Streaming, para ingestão e transformação de grandes volumes de dados de sistemas bancários (cartão de débito e crédito).

Implementação de soluções de captura de dados de mudança (CDC) com Debezium para atualização de dados em tempo real.

Criação de pipelines de dados com arquitetura lambda e kappa para atender diferentes requisitos de negócios.

Orquestração de workflows de dados com Airflow para garantir a confiabilidade e o agendamento das tarefas.

Utilização de ferramentas AWS como EMR (para execução de Hadoop, Spark e YARN) e DynamoDB para armazenamento e acesso a dados de baixa latência.

Desenvolvimento de scripts e ferramentas em Python e Java para automatizar tarefas e otimizar o desempenho dos pipelines.

Implementação de esteiras de integração e entrega contínua (CI/CD) com GitHub para garantir a qualidade e a agilidade do desenvolvimento.

Utilização de Jira para organização do trabalho, aplicação de metodologias ágeis e SCRUM.

Interação com outras áreas da empresa para entender as necessidades de dados e propor soluções inovadoras.

Empresa: NTT DATA – atual. Cargo: Engenheiro de Dados

Set/2023 – Out/2024 – RecordTV

Sobre o cliente: A RecordTV é o 2º maior canal de televisão do Brasil, como plataforma digital, eles possuem o PlayPlus, Mundo Record e o R7.com que é o 3º maior portal do Brasil, com mais de 82 milhões de usuários únicos mensais.

Funções: Criação e Migração de processos ETL/ELT para a nuvem AWS, classificação e clusterização de perfis de usuários.

Ferramentas e linguagens utilizadas: Provedor de nuvem Amazon AWS, EMR, EC2, AWS DMS, AWS S3, Step Functions, AWS Glue, Amazon Redshift, Amazon DynamoDB, Jupyter HUB, Cognito, Linux, Shell Script, Git, GitLab, IaC com Terraform, Python, Apache Spark, PySpark, SQL, SQL Server, AWS RDS MySQL, AWS RDS Postgree, GA4 – Google Analytics 4, Rest API, Jira e Confluence

Desenvolvimento do trabalho: extração de dados utilizando AWS DMS para extrair dados diariamente dos bancos de dados SQL Server, BigQuery, MySQL e Postgree, Criação de servidor SFTP para receber dados JSON e CSV. Arquitetura medalhão, utilizando as camadas bronze, silver, gold e diamante.

Os dados são ingestados diariamente ou semanalmente na camada bronze, transformados utilizando o Apache Spark e persistidos na camada silver utilizando o formato Parquet. Da camada silver para a camada gold é uma etapa de analytics engineer, onde são executadas várias regras de negócios para classificação e clusterização dos perfis de usuários, para a camada gold os dados são persistidos utilizando format Delta Live Table, bem como, são ingestados no DynamoDB e Redshift.

O Redshift como Data Warehouse para que o time de Business Intelligence do cliente possa fazer consultas SQL, bem como, conexão do PowerBI com o Redshift para geração de dashboards e reports. O DynamoDB serve como ETL reverse, onde a aplicação de gerenciamento de publicidade se conecta para detector o perfil do cliente e oferecer propaganda customizada pelo perfil do lead. Trabalhamos no modelo de ambientes de desenvolvimento, homologação e produção.

Para implantação de projetos e versionamento de código, utilizamos CI/CD: Git, Gittlab, AWS CLI e Terraform para IaC.

Resultados atingidos: com o desenvolvimento do projeto foi possível o cliente

- Tableau
- Microsoft SQL Server Integration Service – SSIS
- Microsoft SQL Server Data Tools
- Data Pipelines
- Processos ETL/ELT
- Esteira CI/CD
- Git
- GitLab
- Docker
- Linux
- Shell Script
- Apache Airflow
- Pandas
- NumPy
- Matplotlib
- Seaborn
- PyCaret
- Keras
- TensorFlow
- scikit-learn
- OpenCV

identificar seus usuários, bem como, o perfil e comportamento dos mesmos na utilização dos serviços digitais, os dados foram enriquecidos via API através do bureau de dados Serasa Experian, de posse da classificação e perfil do usuário, quando ele acessa os serviços digitais da RecordTV o Sistema de Ads Manager se conecta ao Amazon DynamoDB para recuperar o perfil do usuário e entregar conteúdo, bem como, propaganda direcionada ao perfil do usuário, desta forma, foi possível a RecordTV multiplicar em até 10 (dez) vezes o valor cobrado para cada milhar de visualização de publicidade.

Mar/2023 – Ago/2023 – Banco Santander

Sobre o cliente: o Banco Santander é o 17º maior banco do mundo, e o 6º banco brasileiro em número de contas correntes, atendendo mais de 66,3 milhões de clientes.

Funções: Criação e Migração de processos ETL/ELT para a nuvem Microsoft Azure

Ferramentas e linguagens utilizadas: Provedor de nuvem Microsoft Azure, Apache Spark, Databricks, PySpark, SQL, Azure Data Lake Storage, Azure Data Factory, Informatica Power Center Design, Linux, Shell Script, Git, GitLab, Blaze, Pipefy, Jenkins, CloudBees, Maestro

Desenvolvimento do trabalho: estamos em processo de descontinuação do uso da Informática Power Center como ferramenta ETL. Assim, o Informática Power Center Design serve como suporte para a compreensão do processo ETL combinado com dicionários técnicos, Confluence, diagramas UML e outras documentações.

Para realizar esta tarefa, utilizamos o Microsoft Azure e a sua vasta gama de soluções, como Azure Data Lake Storage Gen2, Azure Data Factory, Azure SQL, entre outras soluções Azure. Também usamos Azure Databricks para trabalhar com Apache Spark para manipulação e ingestão dos dados.

O projeto é feito em Python, PySpark e SQL e durante o processo usamos Spark Dataframes para a etapa de transformação e Delta Tables para salvar os dados.

Para implantação de projetos e versionamento de código, utilizamos CI/CD: Git, Gitlab, Maestro, Jenkins, CloudBees, Azure Data Factory CLI, Databricks Workflow e Job-cluster.

Empresa: Cardim Sistemas

Cargo: Engenheiro de Dados

Início: Jan/2022 – Fev/2023

Funções: Configurações de máquinas virtuais com Firebird com alta disponibilidade para integrar dados locais com a nuvem Microsoft Azure.

Ferramentas utilizadas: Linux, Docker, Shell Script, Python, SQL, Apache Airflow, Azure Data Factory, SSIS, Azure SQL, Azure Data Lake, Databricks, Spark, PySpark, Python, SQL para processos de ETL/ELT.

Pandas, NumPy, Matplotlib, Seaborn, PyCaret, Keras, TensorFlow, scikit-learn, SQL, Spark, PySpark, Tableau e Microsoft PowerBI para elaborar data viz, relatórios e dashboards para os mais de 70 clientes corporativos atendidos pela Cardim Sistemas.

Desenvolvimento do Trabalho: desenvolver e manter pipelines de dados para ingestão de dados em Azure Cloud Blob Storage. Desenvolvo e mantenho pipelines de dados para ingestão, limpeza e transformação de dados de camada raw para Datawarehouse.

Elaborei e mantive relatórios e dashboards para gerar insights e informações para os diversos administradores e gestores que trabalham com os produtos da Cardim Sistemas terem informações precisas e tomada de decisões estratégicas, certas e rápidas. Trello para aplicação da Metodologia Ágil Kanban

Empresa: Tribunal de Justiça da Bahia

Cargo: Técnico Judiciário

Início: Jun 2008 – Jan 2022

De Ago/2020 até Jan/2022

Engenheiro de Dados

Funções: Criação de Pipelines de dados para Microsoft Azure, conexão do Linked Service com o Oracle para extração das bases de dados utilizando o Azure Datafactory como ferramenta de extração e orquestração, aplicação da arquitetura medalhão, Databricks para manipulação de dados entre as camadas bronze e gold;

Desenvolvimento de relatórios e Dashboards com Apache Spark, SQL, Power BI, bibliotecas Python para análise de dados: Pandas, PyCaret, Keras Tensor Flow, NumPy, Matplotlib, Seaborn entre outras.