

Marcelo Peres

Big Data Engineer

Email: brmarcelo.peres@gmail.com

LinkedIn: [My Profile](#)

Github: [My Repo!](#)

Sou um Engenheiro de Big Data com 8 anos de experiência, especializado em Python, PySpark e Spark SQL.

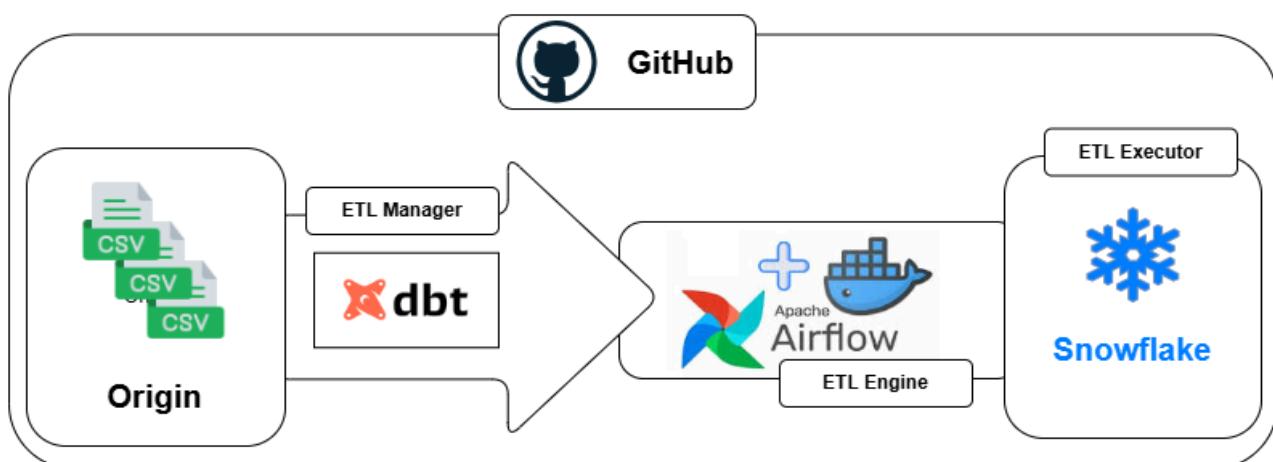
Nos últimos dois anos, tenho trabalhado intensamente no ecossistema Databricks dentro da plataforma AWS, liderando esforços para migrar processos complexos de ETL/ELT de clusters EMR.

Tenho experiência prática com ferramentas de orquestração de dados, particularmente Apache Airflow e Control-M, onde projetei e otimizei fluxos de trabalho para pipelines de dados escaláveis. Tenho me dedicado bastante ao Airflow, o que me permite agendar, monitorar e manter operações de dados confiáveis em sistemas distribuídos de forma eficiente.

Com paixão pela colaboração, destaco-me na promoção do trabalho em equipe, compartilhando conhecimento e boas práticas para ajudar as equipes a crescer e alcançar seus objetivos de negócios. Acredito firmemente que o maior trunfo que um profissional pode ter é a capacidade de compartilhar habilidades e empoderar outras pessoas sem medo da competição.

Meu compromisso com a excelência técnica e o aprendizado contínuo me impulsiona a estar na vanguarda das tecnologias emergentes, sempre agregando valor e mantendo altos padrões de desenvolvimento.

[Github Exemplo de Projeto](#)



Experiências Profissionais

Engenheiro Big Data Senior | Aplos Solutions | 02 ano(s) e 05 meses

Desde Julho de 2023 até o presente momento

Projeto Audlab | 11 meses

De Janeiro de 2025 até o presente momento

Atualmente desenvolvendo um data lake com tabelas Delta Lake e tabelas PostgreSQL para os ativos comerciais da Gerdau. Algumas tabelas no PostgreSQL servem para interação com aplicativos e tomada de decisões por meio de relatórios do Power BI.

Projeto GAB - Manutenção Industrial | 03 meses

De Setembro de 2024 até Dezembre de 2024

Responsável pela migração e reconstrução do processo ETL em PySpark. O escopo original era apenas migrar os scripts e implementá-los em um ambiente CI/CD, além de configurar o agendamento de jobs na plataforma Control-M.

No entanto, devido a um atraso nos fluxos de trabalho do Power BI, o projeto sofreu uma alteração, levando grande parte do processo para o Databricks.

Novas tabelas foram criadas, etapas foram adaptadas e o projeto foi ajustado.

O tempo de faturamento final foi reduzido e o faturamento do Power BI tornou-se mais rápido.

Para três tabelas grandes, os ganhos foram impressionantes, conforme mencionado abaixo:

- A primeira tabela levava 01h20.
- A segunda levava 03h13.
- A terceira levava 01h30.

Agora cada uma das tabelas leva apenas 10 minutos ou menos para carregar os dados para o Power BI.

Projeto Audlab | 01 ano(s) e 02 meses

De Julho de 2023 até Setembro de 2024

Responsável pela ingestão de dados.

A Audlab é um departamento de Auditoria na empresa Gerdau. A Gerdau é uma importante empresa brasileira do setor metalúrgico, que produz diversas ferramentas para diferentes segmentos globais. Neste departamento, eu tinha controle total do GitLab e da implantação de scripts.

O Control-M era responsável pela execução de clusters EMR e scripts em Python e PySpark por meio de diferentes recursos. Produzia scripts de ETL e Ciência de Dados para aprendizado de máquina e transformação de dados. Fui um dos principais responsáveis pela migração de dados das ferramentas da AWS para o Databricks.

Os clusters EMR agora estão replicados corretamente no Databricks. A migração foi um sucesso.

Finalmente, estou trabalhando no Databricks.

Engenheiro de Big Data | A3 Data | 03 meses

Projeto Datalake - Empresa Stellantis | 03 meses

De Dezembro de 2022 até Março de 2023

<< GCP - Platform Google em nuvem >>

Responsável por criar ferramentas para interagir com dados na plataforma.

Ferramentas utilizadas:

- Terraform;
 - Google Storage;
 - BigQuery;
 - Google Functions;
 - Airflow;
 - Python;
 - SQL.
-

Engenheiro de Dados Consultor | Via Consulting | 05 meses

De Abril de 2022 até Setembro de 2022

Projeto - Gol Gaste & Ganhe | 01 mês

De Agosto de 2022 até Agosto de 2022

Smiles S.A.

Função Lambda AWS Python que valida arquivos para serem enviados para uma API.

Recursos:

- Cloud Formation | Python com testes unitários | Github CICD | Jenkins para observabilidade | Sonar para qualidade de código | Finalização do código no repositório.
-

Projeto - Replicação Zendesk | 01 mês

De Julho de 2022 até Julho de 2022

Processo contínuo de envio de dados numa ambiente AWS usando ferramentas como:

- Pyspark Spark Glue jobs | Cloud Formation | AWS Athena | Apache Hudi metadata for data governance.
-

Projeto Unimed - Carteira Recheada | 03 meses

De Abril de 2022 até Junho de 2022

Unimed Seguros

Colaborando com a equipe em processos de ETL usando ferramentas como:

- Informatica Powercenter | PLSQL - Oracle.

A idea do projeto é uma campanha que apoia a pontuação dos corretoras da empresa e posteriormente premiação.

Engenheiro de Dados Consultor | BI4ALL | 10 meses

De Julho de 2021 até Maio de 2022

Projeto - BI na AWS | 10 meses

De Julho de 2021 até Maio de 2022

EmpresaManserv

Responsável pela criação de pipelines de dados para enviar dados para o S3 usando funções AWS Lambda em Python. Utilizando pacotes como awswrangler, xmldict, json e muitos outros. Participando de ótimos projetos, demonstrando as habilidades relevantes, acessando diferentes APIs de vários provedores.

Alguns deles são:

- Volvo
- Nuntec
- Komatsu
- Komtrax
- Caterpillar and many others.

Além disso, atuando em extrações de APIs de retorno XML e JSON, transformando em dados tabulares para serem registrados no formato de tabela Parquet, permitindo um melhor aproveitamento do bucket S3 e ganhando desempenho em um formato de arquivo compactado.

Analista de TI | CLIF | 02 ano(s) e 02 meses

De Abril de 2019 até Junho de 2021

Gestor de Sistemas & desenvolvedor de BI | CLIF | 02 ano(s) e 02 meses

Abril de 2019 até Junho de 2021

CLIF - Centro Logístico Integrado FastCargo S.A

Gestor de Sistemas

- Responsável pela gestão dos sistemas SARA, Logix, My Checklist e Portal de Atendimento ao Cliente;
- Aplicação de pacotes de atualização do sistema alfandegado SARA;
- Atualização de procedimentos armazenados em SQL;
- Atualizações de gatilhos SQL;
- Revisões de módulos, entre outras atividades;
- Criação de um script em Python para inserir e atualizar dados de navios e informações relacionadas no Portal de Atendimento ao Cliente;

Desenvolvedor de BI

Criação de relatórios no Power BI:

- Gestão e controle de cadastros;
- Gestão e controle de unidades;
- Gestão e controle de transportes;
- Gestão e controle de navios.

Os relatórios acima são sincronizados e carregados no Power BI, contendo dados de 2 bancos de dados diferentes e um carregamento de dados de um script Python conectado a uma API disponibilizada pelo Porto Itapoá, contendo as datas de movimentação dos navios. Isso possibilita o cruzamento de todas as informações entre esses diferentes bancos de dados para uma melhor gestão dos ativos dos clientes.
