

Lucas Cruz

Data Engineer with 3+ years of experience delivering scalable data platforms across cloud (AWS) and on-premises environments. Strong background in Python, SQL, PySpark, Airflow, Trino and Docker, with hands-on experience designing Data Lake, Data Warehouse and Lakehouse architectures. Proven ability to integrate complex data sources (APIs, databases, IoT streams) and deliver reliable analytics for business stakeholders. Currently expanding expertise in Snowflake, Databricks, dbt and Kafka, with a strong interest in modern, cloud-native data engineering roles in the UK.

PROFESSIONAL EXPERIENCES:

Spacecom — Data Engineer;

Brazil | Apr 2023 - Jul 2024 (1y3m)

Led stakeholder discussions to gather requirements and translate business needs into technical data solutions. Designed and implemented an end-to-end batch data pipeline integrating PostgreSQL, SQL Server, APIs and Excel/CSV sources using Python and SQL. Built and maintained a Big Data Lakehouse on AWS S3 following the Medallion Architecture (Bronze, Silver, Gold) with Delta Parquet. Developed scalable data transformations using PySpark and Pandas and exposed curated datasets via Trino. Orchestrated ETL workflows with Apache Airflow, ensuring reliable scheduling and monitoring. Deployed services using Docker on AWS EC2, supporting production-grade workloads. Created analytical dashboards with Qlik Sense and Apache Superset, enabling data-driven decision-making. Supported a platform ingesting data from 90k+ IoT devices, processing hundreds of millions of records per day.

wDiscovery, Brazil — Data Engineer Consultant;

Brazil | Jun 2022 – May 2024 (1y11m)

Delivered multiple end-to-end data projects in collaboration with business stakeholders. Integrated data from payment providers (PagSeguro, Stone, Cielo) and agricultural machinery (John Deere). Ingested and processed data from APIs, SQL Server, PostgreSQL, Excel/CSV files and web scraping pipelines using Python and SQL. Designed and maintained Data Warehouses on SQL Server and Oracle Autonomous Database. Performed data cleansing, transformation and loading using Pandas. Orchestrated ETL pipelines with Apache Airflow and deployed solutions on Oracle Cloud VMs using Docker. Developed Power BI dashboards to support operational and strategic business analytics.

Itaete Group, Brazil — Data Analyst;

Brazil | Jun 2020 – Feb 2023 (2y8m)

Developed data solutions integrating IoT devices, PostgreSQL, SQL Server, APIs and Excel/CSV files. Built data pipelines using Python and SQL, performing data cleansing and transformation with Pandas. Designed and implemented a Data Warehouse in SQL Server to support reporting and analytics. Created dashboards using Power BI and Excel for internal teams and clients. Developed an internal system for inventory management and financial process automation, including invoice generation and automated weekly dashboard distribution. Provided ongoing client support, ensuring data quality, reliability and platform stability.

EDUCATION

Dom Bosco University — Systems Analysis and Development

Brazil | 01/2021 - 08/2023 (2y6m)

Guildford, UK (GU1)

+44 07512 225720 (Whatsapp)

dev.lucascruz@gmail.com

<https://github.com/devzurc>

<https://linkedin.com/in/lucas-cruz>

SKILLS

[0~1 y exp]:

- dbt;
- Kafka;
- Snowflake;
- Databricks;

[1~2y exp]:

- PySpark;
- Trino/Presto;
- Qlik Sense;
- Apache Superset;

[2~3y exp]:

- CI/CD
- Docker;
- Power B.I;
- Apache Airflow;

[3~4y exp]:

- Scrum and Kanban;
- Cloud; (AWS and Oracle);

[4~5y exp]:

- Python;
- SQL and NoSQL;
- Bash Scripts;
- Git;

[5+y exp]:

- Linux/Windows;
- IT Infrastructure;
- Network and Security;
- Advanced MS Office;

LANGUAGES

- Portuguese (Native);
- English (Advanced);
- Spanish (Intermediate);

