

# LUCAS CRUZ

Data Engineer | Gen. AI Engineer

+55 (21) 96552-9308 • dev.lucasacruz@gmail.com • linkedin.com/in/lucas-cruz • devzurc.github.io/portfolio

Curitiba, Brazil • Open to relocation across Europe • Requires visa sponsorship

## PROFESSIONAL SUMMARY

Data & Gen. AI Engineer with 5+ years delivering end-to-end data pipelines, cloud lakehouses, and AI-powered automation across IoT, fintech, and agribusiness. Hands-on with LLMs, n8n agent workflows, and conversational AI on AWS, Azure, and GCP. Fluent English (C1) — actively seeking European opportunities.

## TECHNICAL SKILLS

<b>Gen. AI &amp; LLMs</b>	Claude AI, Google AI Studio, LLM Integration, AI Agents, RAG, Prompt Engineering, Chatbot Development
<b>Data Engineering</b>	Python, SQL, Apache Spark (PySpark), Pandas, dbt, Trino, REST APIs, Web Scraping, Parquet
<b>Pipeline Orchestration</b>	Apache Airflow, n8n
<b>Multi-Cloud Platforms</b>	AWS (S3, EC2, ECS, Lambda, IAM), Microsoft Azure, Google Cloud (GCP), Oracle Cloud (OCI)
<b>Data Platforms &amp; Lakehouses</b>	Snowflake, Databricks, Amazon Redshift, Delta Lake, Oracle Autonomous Database
<b>Relational Databases</b>	PostgreSQL, SQL Server, MySQL
<b>Analytics &amp; BI</b>	Power BI, Qlik Sense, Apache Superset, Looker
<b>Security &amp; Data Governance</b>	LGPD / GDPR Compliance, Data Privacy, Access Control (IAM), RBAC, Data Masking, Encryption at Rest & Transit, Audit Logging, Electronic Security Tags (RFID/NFC)
<b>DevOps &amp; Infrastructure</b>	Docker, Kubernetes, Git, CI/CD Pipelines, Linux, Bash Scripting
<b>Languages</b>	Portuguese (Native) • English (C1 – Advanced) • Spanish (B1 – Intermediate)

## WORK EXPERIENCE

### Gen. AI Engineer – TK Technologies | Curitiba, PR – Brazil (Hybrid) Feb 2026 – May 2026 · Contract

- Built a multi-source web scraping pipeline in Python targeting 10+ automotive parts websites, extracting pricing data and product metadata at scale to deliver competitive market intelligence.
- Designed AI-powered n8n automation workflows (ETL, RPA, chatbot agents) integrating LLMs (Claude AI, Google AI Studio) and GCP APIs (Sheets, Drive) to enrich data and boost team productivity.
- Developed a conversational AI chatbot enabling business users to query and refine pricing datasets via natural language, reducing time-to-insight for non-technical stakeholders.
- Deployed containerised backend services on Azure with Docker, standardising environments and improving deployment reliability by ~25%.

### Data Engineer – Spacecom Monitoramento | Curitiba, PR – Brazil (On-site) Apr 2023 – Jul 2024

- Designed and maintained a Big Data Lakehouse on AWS S3 using Medallion Architecture (Bronze/Silver/Gold) with Delta Lake, Parquet, PySpark, and Trino, supporting analytics and reporting workloads.
- Supported a large-scale IoT platform ingesting data from 95,000+ devices, processing hundreds of millions of records per day with high availability.
- Worked with electronic security tags (RFID/NFC) integrated into an IoT platform serving government court systems, gaining deep hands-on experience with LGPD (Brazil's GDPR equivalent) — implementing data privacy controls, access governance, and secure handling of sensitive judicial data.
- Orchestrated and optimised ETL pipelines with Apache Airflow, reducing failure rates by ~30%; deployed services with Docker on AWS EC2, cutting deployment issues by ~25%.
- Identified and remediated inconsistent pipeline steps, increasing downstream analytics accuracy by ~40%; delivered datasets to Qlik Sense and Apache Superset dashboards.

### Data Engineer Consultant – wDiscover | Curitiba, PR – Brazil (Remote) Jun 2022 – May 2024

- Delivered end-to-end data projects across fintech, agribusiness, and operations — integrating payment providers (PagSeguro, Stone, Cielo), John Deere IoT APIs, REST APIs, and web scraping pipelines.
- Designed Data Warehouses on SQL Server and Oracle Autonomous Database with dimensional modelling; automated ETL with Airflow on Docker/Oracle Cloud, reducing manual effort by ~35%.
- Built Power BI dashboards tracking KPIs, financial metrics, and operational performance for stakeholders across multiple client domains.

### Data Analyst – Itaete Group | Curitiba, PR – Brazil (Hybrid) Jun 2020 – Feb 2023

- Designed a SQL Server Data Warehouse and Python/SQL pipelines with Pandas for cleansing, transformation, and enrichment, integrating IoT devices, PostgreSQL, APIs, and CSV sources.

- Built Power BI dashboards and automated Excel/VBA reporting workflows, reducing data processing time by ~50% and significantly cutting manual operational workload.
- Developed internal systems for inventory management and financial automation including invoice generation and automated weekly report distribution.

## NOTABLE PROJECTS

---

**Automotive Market Price Intelligence Platform** · Python, n8n, LLMs, GCP APIs, Azure, Docker · *TK Technologies – Feb 2026*

- Scraped 10+ automotive parts sites with Python; orchestrated AI agents via n8n with Claude AI, Google AI Studio, and GCP (Sheets, Drive) to automate price analysis and competitive procurement insights. Fully containerised on Azure.

**Modern Cloud Data Pipeline** · Snowflake, dbt, Apache Airflow, Docker · *Personal Project – Jan 2026*

- End-to-end analytics pipeline with Airflow, Snowflake, and dbt; containerised dbt with Docker reducing deployment errors by ~50%; modular tested SQL transformations orchestrated via Airflow DAGs.

**IoT Data Lakehouse on AWS – Medallion Architecture** · AWS S3, Apache Spark, Delta Lake, Airflow, Trino, Apache Superset · *Client Project – Feb 2024*

- Designed Bronze/Silver/Gold lakehouse on AWS S3 ingesting from APIs, databases, and CSV; virtualised Gold layer with Trino; surfaced insights via Apache Superset for analytics consumers.

## EDUCATION

---

**Bachelor of Systems Analysis and Development** – Dom Bosco University, Curitiba, Brazil – Jun 2023

Majors: Information Systems · Business Analytics · Data Analysis · Database Systems · Software Engineering