

Best Technology Talent For Companies Around The World



Fabio

Senior Data Engineer

Summary

A passionate Data Engineer driven by curiosity and creativity, approaches data engineering as a bridge between distributed systems, DevOps, and machine learning rather than a single discipline. Thrives on solving complex technical challenges, exploring new tools, and continuously building scalable data solutions. His work spans ETL design, real-time ingestion, cloud orchestration, and infrastructure automation using technologies like Terraform, Kubernetes, and Nx. Also lead a microservice-based data engineering project where experiments with lakehouse architectures and real-time pipelines. Always learning and innovating, aims to merge system-level thinking with hands-on experimentation to create efficient, future-ready data ecosystems.

TECHNICAL SKILLS:

- Golang, Python, Rust, Node.js, Lua, PySpark, SQL
- Microservices, Event-Driven, Distributed Systems, DSLs in Go, AI-enabled Pipelines
- RabbitMQ, Kafka, Pub/Sub
- Airflow, BigQuery
- Postgres, MongoDB, ElasticSearch, Redis, Neo4J
- Terraform, Kubernetes, Docker, AWS, GCP, GitLab CI/CD
- Web scraping unstructured data (PDF, CSV, XML, Power BI, HTML)
- Delta, Parquet, and Iceberg format
- PyTorch

Professional Career

Senior Data Engineer

2022 – Present | São Paulo, Brazil

Project Info:

Brazil's leading big data and analytics company, specializing in providing data-driven solutions to businesses across various industries. The company processes up to 385 million new records daily across its multiple solutions.

Responsibilities:

- Enhanced data reach by 10% through automated web scraping of critical public data, such as arrest warrants and court cases, using Python microservices for risk, compliance, and legal.
- Improved Golang microservices to effectively solve captchas (e.g., Tencent, AWS WAF), increasing automated data extraction coverage by 5% from protected sources.
- Innovated a captcha bypass tool reCAPTCHA and Cloudflare Turnstile, completely eliminating captcha-related expenses for the company.
- Developed and optimized data treatment and enrichment pipelines with Golang and PySpark to ensure accuracy, consistency, and usability of vast datasets for legal and compliance purposes.
- Optimized DevOps processes for bot operations by refining CI/CD pipelines, enhancing deployment efficiency and system reliability, and eliminating storage, which reduced GitLab runner costs by 20% per month.
- Indexed curated data into Elasticsearch, enhancing data accessibility, retrieval capabilities, and search performance to support sophisticated risk and compliance analysis.
- Developed a stealth web scraping framework: Integrated Selenium and Scrapy, maintaining bot stealth while leveraging Scrapy's metrics and functionalities.
- Led the development of internal tools for Data Engineers, integrating GenAI services to scale data collection with minimal human intervention, increasing efficiency and enabling new data ingestion strategies.
- Developed CLI tools in Go to streamline common data engineering tasks, improving team efficiency and reducing onboarding time for new pipelines.
- Designed and maintained Golang-based microservices interacting with Kafka, RabbitMQ, and Kubernetes, processing high-throughput workloads in a secure and distributed architecture.
- Extended internal Golang DSL by introducing advanced array operations, boosting developer productivity and optimizing entity generation pipelines used in data modelling for compliance and risk systems.
- Led the design and implementation of an AI-powered data governance platform, structured on graph-based architectures and integrated with Next.js, React, and Motia, enabling intelligent data lineage visualization, semantic search, and policy enforcement through LLM-driven workflows.

Software and technologies:

Python, Golang, PySpark, Airflow, Kafka, RabbitMQ, AWS, ElasticSearch, Terraform

Data Engineer (Freelance, part-time)

2021 – 2023 | São Paulo, Brazil

Project Info:

Forward-thinking company focused on transforming education through data-driven solutions and technology, helping schools and educational systems optimize student learning outcomes and operational efficiencies.

Responsibilities:

- Enhanced operational efficiency by 40% by automating data flows, integrating CRM and ERP systems, and replacing manual processes to ensure data accuracy
- Reduced reporting time by 70% by designing scalable data pipelines with PySpark and Step Functions for processing commercial and financial data
- Collaborated with cross-functional teams to refine data workflows, increasing data processing speed by 15% and improving integration with downstream applications, resulting in smoother, more efficient operations across the organization
- Conducted a comprehensive study of machine learning processes, laying the foundation for structuring a dedicated, sustainable data science area to drive predictive analytics and enhance decision-making

Data Analyst

2021 – 2022 | São Paulo, Brazil

Project Info:

Technology company focused on providing sustainability solutions through data-driven platforms. Their products help businesses achieve environmental compliance and drive sustainable practices by leveraging advanced data analytics and blockchain technology

Responsibilities:

- Developed and maintained data warehouses using SQL, improving data retrieval time by 60%, enabling more efficient long-term analysis and decision-making.
- Saved 15 hours per week by automating repetitive tasks and workflows with Python, reducing manual effort by 80%.
- Created interactive dashboards using Power BI and Streamlit, reducing report generation time by 30% and empowering stakeholders to make timely, data-driven decisions.
- Led squads through project management and training, increasing team efficiency by 25% and improving collaboration in delivering data-driven solutions.

Software and technologies:

Python, SQL, Power BI, Streamlit

Senior Data Engineer

2019 – 2021 | São Paulo, Brazil

Project Info:

Forward-thinking company focused on transforming education through data-driven solutions and technology, helping schools and educational systems optimize student learning outcomes and operational efficiencies.

Responsibilities:

- Designed and optimized data pipelines using PySpark, Airflow, and PubSub, increasing data processing capacity by 40% to handle larger volumes efficiently.
- Drove lead generation growth by 1,000 new users through the integration of Sencon's data into TC's data lake, which greatly improved cross-company analysis and decision-making.
- Automated data flows by integrating external APIs (e.g., GitHub, Hotjar, ERP), improving access to critical business data for analysis and reporting.
- Reduced query time by 80% by optimizing dashboards with data awareness techniques, significantly enabling faster access to key insights.
- Developed interactive dashboards using Looker, providing stakeholders with real-time insights for financial, commercial, and customer experience data, reducing reporting time by 40%.
- Recurring Revenue, 15% improvement in customer retention, 5% boost in Daily Active Users, and 15% in Monthly Active users.
- Headed the development of a Feature Store to streamline machine learning operations, reducing model development time by 30% and improving data accessibility for predictive analytics.
- Crafted robust API to integrate machine learning models with financial systems, enhancing scalability and enabling real-time financial data analysis impacting 1000 users.
- Collaborated with cross-functional teams such as commercial and customer experience to implement machine learning models such as churn prediction and NLP-based models, boosting customer retention and engagement.

Software and technologies:

Python, PySpark, ML, API, Airflow, Neo4j, GCP, PubSub, BigQuery

Education

2014 – 2020

University of Sao-Paulo, Brazil

Bachelor's Degree in Chemical Engineering

Certifications:

- Data Scientist with Python Track – DataCamp - 2020
- Applied Text Mining in Python - Coursera Course Certificates- 2020
- Applied Machine Learning in Python - Coursera Certified Courses
- Applied Plotting, Charting & Data Representation in Python - Coursera Certified - 2020
- Data Science Math Skills - Duke University | Coursera - 2020
- Introduction to Big Data - Coursera Certified Courses - 2020
- Introduction to Data Science in Python - Coursera Certified Courses - 2020
- Python Data Structures - Coursera Certified Courses - 2020

Foreign Languages Skills

English: Upper-Intermediate

Portuguese: Native