
Juan Eduardo Domingos

Lead Data Engineer

São Paulo, Brazil

+55 (19) 99770-5273

juan.eduardo@outlook.com.br

SUMMARY

I'm an experienced and passionate Data Engineer, with 10 years of experience in the software industry, where from these 10 years 6 were working on Data Environment Development. I've already worked with many technologies, architecting and building Data Platforms and Pipelines from scratch, working along with my pairs to find the best solutions to answer the business needs, conciliating delivery, value and efficiency.

My current technologies stack is composed by:

- Python (5 years of experience)
- Apache Spark (6 years of experience)
- Databricks (4 years of experience)
- Apache Airflow (4 years of experience)
- Terraform (5 years of experience)
- dbt (3 years of experience)
- AWS (AWS, MSK, EKS, EMR, SSM, Secrets Manager, S3, etc) (6 years of experience)
- Relational Databases (Oracle, MySQL, PostgreSQL) (10 years of experience)
- Apache Hadoop Ecosystem (YARN, HDFS, Hbase) (5 years of experience)
- Apache Kafka (5 years of experience)
- Data Quality with Great Expectations (3 years of experience)
- Data Observability with MonteCarlo (2 years of experience)

I also have experience on:

- Java 8+
- Scala Lang
- Microservices Architecture
- Oracle PL/SQL
- Google Cloud Platform (Big Query, Workflows, Dataproc, Storage, Cloud Functions)
- Observability tools like Datadog, Sentry, Graphana

In short, I'm a Data Engineer with a software development and architecture background, and my main goal in the Data World is to bring to this context all of the resilience defended by the software development best practices and create a reliable and efficient Data Environment. Last but not least, I'm a lifelong learner. I always like to be exposed to new technologies and scenarios that challenge me, push me to get the bar higher, and incentivize me to learn new things and optimize my current skills.

LANGUAGES

- Portuguese (Fluent, Native)
- English (Advanced, Write & Speak)

WORK EXPERIENCE

Windifferent (It's a Bairesdev Company), Florida, United States - Lead Data Engineer

From March 2022 - Nowadays

WinDifferent is a company focused on growth and lead prospect services using the B2B model; the company works on acquiring leads for the BDev Ventures partners looking to ramp up the startups funded by the BDev Ventures Company.

- Lead Data Engineer in a 10-member team, giving them technical directions and helping to translate the business needs to the data platform technical implementations, including architecture drawings and component designing. I also help my peers develop themselves with career tips and practical peering.
- Design a Data Platform to attend to the company's goals and support the company's data-driven products. We've used Delta Lake and Databricks to build a Lake House and store all the data that we have (some Terabytes)
- In the end-to-end pipelines (from Data Ingestion, Transformation, and Servering), I work acquiring Data from different sources, mainly from relation databases using CDC strategies with Debezium, Kafka, and Spark for Data processing, and also transforming the data to fit the business needs mainly because our main product is data-based, this modeling is made using Spark with Databricks and dbt for some analytical cases.
- Ensure the Data Quality from the end-to-end perspective, using tools like Great Expectations for on-time checks (Completeness, Typing, etc.) and Monte Carlo for live data monitoring, e.g., identifying data spikes.

-
- Work on cost reduction strategies by remodeling and re-designing the processes to be efficient and save as many resources as possible in all of our infrastructure spheres (AWS, Databricks, Airflow Providers, etc.)
 - Manage Cloud Resources on AWS, looking to keep a healthy amount of AWS quotas for services like EC2, Lambda, etc., and also working on the permissions using IAM policies and in the S3 life-cycle maintenance.

Some Highlights that can be mentioned:

- Foundation of Company-Wide's Data and Analytics team planning and providing a Data Platform and making it available across the company, working to help the teams increase their efficiency and to create a single source of truth for its Datasets according to the business requirements and global KPIs
- Reduction of 85% of the cost and 70% of the execution time of one of the main company's pipelines using strategies of data modeling, infrastructure dimensioning, and Spark Tuning Strategies.
- Built a consistent cost report using a fine tagging strategy in the cloud resources, which helped the company to better understand its ROI and give better directions to the business and helped our vendor's contracts negotiations.

Technologies Stack:

- Python
- SQL
- Apache Spark
- MongoDB
- Databricks
- dbt
- Rest API
- AWS
- Terraform
- Kafka
- Apache Airflow
- Metabase
- Great Expectations for Data Quality

PicPay, São Paulo, Brazil - *Senior Data Engineer*

From 2021 August - To 2022 January

PicPay is a Brazilian Fintech with ~50 M users, growing around 30% per year; the main company goal is to provide frictionless payment and wallet services. In

2021, PicPay bought a company called Guia Bolso, looking to create the biggest Brazilian Open Banking case.

- Open Banking Foundation: I've worked as the first data engineering member of the open banking team, identifying the ecosystem entities and the technologies involved and designing the data pipelines to bring this data to the analytical context. I've also worked on the governance definitions along with the Corporate Governance team, considering important requisites like GDPR and data privacy.
- After identifying the entities and defining the standards, my mission was to create the Pipelines and ingest data from scratch; we've decided to use technologies like Debezium with CDC to acquire data from the transactional layer, airflow to orchestrate our pipelines and Databricks as our data processing engine where were Spark Jobs running and for some cases, EMR workloads were present too.
- When the company bought Guia Bolso, the challenge was to integrate the companies' data and enable Guia Bolso's data usage in the PicPay ecosystem. I've worked in entity mapping, user base correlation, and cloud infrastructure integration to make it possible to use the Guia Bolso data to enrich PicPay's users' data. At the end of this process, we used a cloud cross-account integration strategy with well-defined IAM roles to reduce costs and keep the juridical requirements of the data custodian.
- After the integration, we worked on the business visibility of the Open Banking products results, defining the main KPIs and the Data Analysts, always helping them keep in touch with the business areas. In the technical field, we used Tableau for C-Level reports and Looker, which is more user-friendly for business analytics to do their ad-hoc queries. The metrics here were taken from different sources, like application logs, user store comments, call center metrics, etc.

Technologies Stack:

- Python
- SQL
- Apache Spark
- EMR
- Terraform
- Apache Airflow
- Databricks
- AWS
- Tableau
- Looker Studio

Itaú Unibanco Bank, São Paulo, Brazil - Data Engineer

From December 2020 - To August 2021

Itaú is the biggest Latin America Private Bank, with ~60M clients. The bank operations represent 25% of the Brazilian financial transactions.

- Being part of a team of 6 Engineers responsible for building the bank PFM (Personal Financial Management) Platform, my Squad was responsible for cleaning and preparing the Data Received from the Mainframe integrator in a Streaming model, using some key technologies like Apache Spark with Scala for Data processing, Kafka as our Data bus and S3 in a medallion structure to support our data.
- Apply GDPR requirements to the Data using anonymization strategies.
- Create a data server layer to integrate with the customer's are which applied a Featuring model over the data and used it to feed the Machine Learning models used in the PFM.

Technologies Stack:

- Java
- Scala
- SQL
- Apache Spark
- AWS
- Rest API
- Shell Script
- Terraform
- Hadoop
- Kafka

Dextra (CI&T), Campinas, São Paulo, Brazil - Data Engineer | Full Stack Software Engineer (Big Data)

From December 2019 - To December 2020

Dextra was a successful agile outsourcing company in Campina's metropolitan region, which was bought by the global company CI&T in 2021. The company has big customers in its portfolio, like Embraer, Dafiti, GoFundMe, Serasa Experian, Claro, etc.

- Being part of the Serasa Experian enrichment platform project, I've worked in an eight engineers squad responsible for end-to-end development, from Database submission to Data processing and customer feedback. We've deployed an async message-oriented enrichment engine using Angular as FrontEnd technology, Spark over Hadoop Yarn as a processing engine, and AWS as a cloud provider to host the services and store the processing results.
- The Growth Platform was my second challenge at Dextra, where I led a project to enhance the data ingestion and transformation strategies for data coming from Google Analytics and App Stores, like Apple App Store and Google PlayStore; that was a really notable project where assuming a tag translation strategy in the Data Pipelines we've

managed to reduce the ML models implementation time from 2 weeks to 2 hours all the infrastructure was hosted on Google Cloud and the Processing engine was based on Python with Apache Spark, in a parametrized application configured via a YAML file.

Technologies Stack:

- Java
- Spring Boot
- Spring Data
- Rest API
- SQL
- Scala
- Apache Spark
- Databricks
- AWS
- Terraform

Veiling Holambra, Santo Antônio de Posse, São Paulo, Brazil - Full Stack Software Engineer

From November 2014 - To December 2019

Veiling Holambra is the largest flower and plant Cooperative in Latin America, with partners around the world in countries like Belgium, Netherlands, and Germany.

- In the back office, the building of a rental management system was my first project at Veiling, where I had a mission to develop a system to control the cooperative's customers' rents. This system had integrations with the company's ERP (Oracle EBS) to generate all bills and get aligned with the current fiscal laws. The system was developed in .NET with a PL/SQL backend to do the ERP integrations and a desktop application interface, and it was also responsible for generating the documents dynamically when a new contract was created.
- In the sales area, I worked in the Vendor's integration services, which were responsible for supporting the supply chain of the cooperative to coordinate all the product transactions, which would be later used in the company's auctions (The common sales method for Veilings around the world) and also for sales intermediated by sellers; these integrations were basically done using SOAP APIs with RabbitMQ queues, with applications written in Java in an event-driven architecture.
- I've also done a lot of projects for the logistics area, where I developed systems to track the products in the company's logistics chain and also to define a strategic order in the auctions to ensure the

best sales strategy; for these projects, I've worked with some IoT technologies, e.g., RFID and also with Windows Mobile applications along with PL/SQL to develop the ERP integrations.

Technologies Stack:

- Java
- Groovy (Groovy on Rails - GRAILS)
- SQL
- C#
- JBoss
- SOAP
- Rest API
- RabbitMQ
- Oracle DB
- Oracle EBS
- PL/SQL

EDUCATION

PUCRS - Pontifícia Universidade Católica Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil - *Post Graduation on Software Architecture field*

From August 2023 - Nowadays

PUCPR - Pontifícia Universidade Católica Paraná, Curitiba, Paraná, Brazil - *Post Graduation on Data Science and Distributed Systems architecture field*

From March 2021 - To June 2022

UNIFAJ - Centro Universitário de Jaguariúna, Jaguariúna, São Paulo, Brazil - *Computer Science Bachelor's Degree*

From January 2017 - To December 2020