# IBISEN GONÇALVES

*DATA ENGINEER*

## CONTACT

TEL: +55(11)91264-2832
Address: São Paulo-SP, Brazil
Linkedin: linkedin.com/in/ibisenbrito
GitHub: github.com/ibisen
Email: ibisenit@gmail.com

## EDUCATION

B.S.
Electrical and Electronic Engineering
**UNOPAR UNIVERSITY**
Feb 2019 – current
Bahia, Brazil

## HIGHLIGHTED SKILLS

Python
SQL, NoSQL
Microsoft Azure
Databricks
Apache NiFi
Apache Spark
PySpark, Spark SQL
Delta Lake, Lakehouse
Azure Data Factory
Cosmos DB
Microsoft Synapse

## SKILLS

Shell Script/Power Shell
Denodo Data Virtualization
Hadoop
Hive
Hue
AWS
Power BI
Tableau
ETLs, Data Pipelines
Pandas
Matplotlib
Sci-kit Learn
OpenCV
CI/CD, Git/GitLab
Control-M
Scrum, Kanban
Jira/Pipefy

## CAREER OBJECTIVE

Motivated by technology, engineering, data and results, my passion and associate level experience combined will exceed expectations of all companies. Ready to join a team of critical thinkers to build a data stack from scratch for all purposes.
Focused on cloud computing to help companies that use the cloud to improve their performance and especially to migrate companies from their on-premises servers to the power of cloud processing.

## WORK EXPERIENCE

**Act Digital – Senior Data Engineer - current - São Paulo, Brazil**

***Out/2024 - Plard (Start-up developing debt and credit card system fo Santander)***

Senior Data Engineer with experience in high-performance projects on AWS Cloud, focused on data solutions for the banking sector. I work on building and maintaining robust and scalable data pipelines, leveraging cutting-edge technologies such as Python, SQL, Java Spring Boot, Kafka, Apache Flink, Debezium, PySpark, and Spark Streaming.

Develop and maintain batch and streaming data pipelines using PySpark and Spark Streaming to ingest and transform large volumes of banking system data (debit and credit card transactions).

Implement Change Data Capture (CDC) solutions with Debezium to ensure real-time data updates. Design Lambda and Kappa architecture data pipelines to meet diverse business requirements. Orchestrate data workflows with Apache Airflow to ensure task reliability and scheduling. Utilize AWS tools such as EMR (for running Hadoop, Spark, and YARN) and DynamoDB for low-latency data storage and access.

Develop automation scripts and tools in Python and Java to optimize pipeline performance. Implement CI/CD pipelines with GitHub to ensure software quality and development agility.

Use Jira for task organization, Agile methodologies, and SCRUM. Collaborate with cross-functional teams to understand data needs and propose innovative solutions.

**ESS IT – Data Engineer - current - São Paulo, Brazil**

***Feb/2024 - Out/2024- Ultragaz***

Active participation in strategic decisions alongside the architecture team for the implementation of Delta Lake House and Databricks Unity Catalog for centralized and secure data governance.

Created and maintained over 1,200 data pipelines with Azure Data Factory, extracting data from Oracle EBS, Salesforce, AWS MySQL RDS, SQL Server, and Flask APIs.

Configured Azure DevOps, CI/CD with GitHub Actions, and integrated data with Power BI. This work provided insights, reports, and dashboards for various business areas of the largest gas company in Brazil, optimizing strategic decision-making.

Actively involved in the implementation and management of Databricks environments, configuring clusters, defining permissions, and optimizing workflows for processing large volumes of data.

Integrated and automated CI/CD pipelines using Azure DevOps and GitHub Actions for continuous deployment of Databricks scripts and notebooks.

Managed Unity Catalog for centralized data governance, ensuring security, compliance, and refined access control across different environments and development teams.

Collaborated with multidisciplinary teams to integrate data processes with GitHub, facilitating version control and collaboration with multiple teams in distributed workflows. Conducted data analysis and modeling to provide strategic insights, using Spark SQL and PySpark to optimize large-scale data processing.

**NTT DATA LATAM – Data Engineer - current - São Paulo, Brazil**

***Aug/2023 - Fev/2024 - RecordTV***

## SOFT SKILLS

Communication

Collaboration

Analytical Mindset

Problem-solving

Business Mindset

Creativity

Self-learning

## CERTIFICATES

Microsoft Certified:DP-203 Azure Data Engineer Associate

Microsoft AZ-900 Azure Cloud Fundamentals

Microsoft Certified:AI-900 Azure AI Fundamentals

Microsoft Certified:DP-900 Azure Data Fundamentals

Databricks Fundamentals

## LANGUAGES

Brazilian portuguese - Native

English - Upper Intermediate keeping studying

## PROJECTS

Colorize black and white photos with OpenCV.

Bitcoin and Stock Market Price prediction with Tensor Flow.

Business Intelligence and Dashboards (Power BI, Stream Lit & Tableau);

Machine Learning Supervised and unsupervised learning, 'learning' models, scikit-learn (classification and clustering);

Data collection, manipulation and visualization - Web scraping.

I've worked at RecordTV, the second-largest television channel in Brazil. As a digital platform, they encompass PlayPlus, Mundo Record, and R7.com, the third-largest portal in Brazil, boasting over 82 million unique monthly users.

My responsibilities revolve around creating and migrating ETL/ELT processes to the AWS cloud, along with user profile classification and clustering.

I've utilized various tools and languages including Amazon AWS cloud provider, EMR, EC2, AWS DMS, AWS S3, Step Functions, AWS Glue, Amazon Redshift, Amazon DynamoDB, Jupyter HUB, Cognito, Linux, Shell Script, Git, GitLab, Terraform for IaC, Python, Apache Spark, PySpark, SQL, SQL Server, AWS RDS MySQL, AWS RDS PostgreSQL, GA4 – Google Analytics 4, Rest API, Jira, and Confluence.

My work involves data extraction using AWS DMS to fetch data daily from SQL Server, BigQuery, MySQL, and PostgreSQL databases. Additionally, I've set up an SFTP server to receive JSON and CSV data. We've implemented a medallion architecture, utilizing bronze, silver, gold, and diamond layers.

Data is ingested daily or weekly into the bronze layer, transformed using Apache Spark, and persisted into the silver layer using the Parquet format. Moving from the silver to gold layer involves an analytics engineering step, executing various business rules for user profile classification and clustering. Data in the gold layer is persisted using Delta Live Table format and ingested into DynamoDB and Redshift.

Redshift serves as a Data Warehouse for the client's Business Intelligence team to run SQL queries. Furthermore, we've connected PowerBI with Redshift for dashboard and report generation. DynamoDB acts as a reverse ETL, where the advertising management application connects to detect customer profiles and offer customized advertising.

We adhere to a development, testing, and production environment model and employ CI/CD using Git, GitLab, AWS CLI, and Terraform for IaC. We follow Agile Scrum methodologies for project management.

In addition to Jira, Microsoft Teams, Excel, Visual Studio Code, and Notepad++, we use Confluence for documentation.

The project's outcomes include enabling the client to identify their users and understand their usage patterns. Data enrichment via API from Serasa Experian allowed for targeted advertising, resulting in up to a tenfold increase in advertising revenue for every thousand views.

***Mar/2023 - Aug/2023 - Santander Bank***

During my tenure at NTT DATA from March 2023 to August 2023, I had the opportunity to work closely with Banco Santander, a global banking giant with a presence in over 17 countries and serving more than 66.3 million Brazilian customers. As a data engineer specializing in asset and liability management (ALM) and Liquidity Coverage Ratio (LCR), my primary responsibility was to oversee the migration of Santander's ALM On-premise ETL system to the Microsoft Azure cloud.

The decision to migrate to Azure was driven by the exponential growth in data volume, reaching petabyte-scale monthly, necessitating a scalable and cost-effective solution. Our objective was not only achieved but surpassed expectations, resulting in substantial cost savings for the organization annually.

Throughout the migration process, we transitioned away from Informatica Power Center to modern Azure solutions, leveraging tools such as Azure Data Lake Storage Gen2, Azure Data Factory, Azure SQL, and Azure CosmosDB. Azure Databricks, in conjunction with Apache Spark, facilitated efficient data processing and load handling, the project was developed with Python, PySpark and SQL..

Adopting a development lifecycle model, we implemented environments spanning from development to production, incorporating bronze, harmonization, silver, and gold layers for Big Data processing. Our CI/CD pipeline, comprising Git, Gitlab, Jenkins, and other tools, ensured seamless code versioning and deployment.

Agile methodologies, including Scrum and Kanban, guided our project management approach, fostering collaboration and adaptability. Additionally, we embraced various tools like ITSM Service Now, Jira, and Microsoft Teams to streamline communication and workflow.

The transition to Azure not only optimized performance but also enhanced operational efficiency. With faster execution, delivery, and deployment capabilities, we eliminated

pipeline failures due to memory constraints or job concurrency issues, ensuring uninterrupted data processing.

In summary, our successful migration to Microsoft Azure not only met the primary objective of data and pipeline migration but also delivered significant ancillary benefits, including cost reduction, enhanced performance, and improved operational resilience..

For project deployment and code versioning we use CI/CD mat: Git, Gitlab, Maestro, Jenkins, CloudBees, Control-M orchestrator, Azure Data Factory CLI, Databricks Workflow and Job-cluster. Other tools used at work on a daily basis are: ITSM Service Now, Pipefy, Jira, Visual Studio Code, Notepad++, among others.

**CARDIM SISTEMAS – Data Engineer**

**Jan 2022 – Feb/2023 - Bahia, Brazil**

I've established virtual machines with Firebird and high availability to facilitate the integration of data from on-premise systems to the Azure cloud platform.

Collaborating with a senior data engineer, I utilized Python, SQL, Spark, and Databricks to develop cloud-first dashboards.

Utilizing Azure Data Factory, Python, and SQL, I orchestrated the ingestion of data from multiple sources to construct data views tailored for BI tools such as Microsoft Power BI and Tableau, catering to over 70 enterprise clients. Employing Spark in Python, I orchestrated distributed data processing on extensive streaming datasets, resulting in a 67% enhancement in ingestion speed and efficiency.

I spearheaded the development and upkeep of data pipelines responsible for ingesting data into Azure Cloud Blob Storage. I led the development and maintenance of data pipelines responsible for ingesting, cleaning, and transforming data for a Data Warehouse.

Furthermore, I developed and maintained dashboards aimed at providing stakeholders with rapid access to accurate information.

**BAHIA'S COURT HOUSE – IT Technician Specialist**

Jun 2008 – Jan 2022 – Bahia, Brazil

**From Feb/2020 to Jan/2022 - I worked as an internal Data Analyst**

Created and maintained over 500 data pipelines with Azure Data Factory, extracting data from Oracle and Flask APIs.

Configured Azure DevOps, CI/CD with GitLab, and integrated data with Power BI. This work provided insights, reports, and dashboards for various KPIs for one of the biggest Brazil Court House, optimizing strategic decision-making.

Actively involved in the implementation and management of Databricks environments, configuring clusters, defining permissions, and optimizing workflows for processing large volumes of data. Integrated and automated CI/CD pipelines using Azure DevOps and GitLab for continuous deployment of Databricks scripts and notebooks.

Managed Unity Catalog for centralized data governance, ensuring security, compliance, and refined access control across different environments and development teams.

Developed reports and dashboards with Python, SQL and Power BI

Python libraries for data analysis: Pandas, PyCaret, Keras Tensor Flow, NumPy, Matplotlib, Seaborn and others. Oracle database servers was the source of data origin.

**From Jun/2008 to Feb/2020 - I worked as Technician Specialist**

Creation of a system to control arrested defendants, release permits and arrest warrants using Microsoft Access. With this simple program the team was capable of generating hundreds of court orders in less than one single minute and nowadays, even with programs that worth R$ 40 million, they can't do it the same..

Created a control system for dispatched documents using MySQL and Python with Google Cloud. System migration, from physical archives to digital. Dispatch and creation of document templates linked to the database.

I implemented the audiovisual recording system of the hearings and the Jury Court.