

Education

- **Universidade Nova de Lisboa** Lisbon, PT
BA/BS. Geography and Regional Planning 2011 - 2014
 - Degree in Geography and Regional Planning, with a specialization in Geographic Information Systems.
- **ISCTE - Instituto Universitário de Lisboa** Lisbon, PT
PGDip. Computer Science and Business 2015 - 2017
 - Post-graduate degree in Computer Science and Business, only missing the final project to complete the degree.
- **Databricks Certified Associate Developer for Spark 3.0** Lisbon, PT
CERT, Databricks 2020
 - Certification for Apache Spark 3.0, including Spark SQL, DataFrames, Datasets, Structured Streaming, MLlib, and GraphX.
 - Issued by Databricks.
- **Azure Data Engineer Associate** Lisbon, PT
CERT, Microsoft 2021
 - Certification for Azure Data Engineer Associate, including Azure Data Factory, Azure Data Lake, Azure Synapse Analytics, Azure Databricks, Azure Stream Analytics, Azure Cosmos DB, Azure SQL Database, Azure SQL Managed Instance, Azure Database for PostgreSQL, Azure Database for MySQL, and Azure Synapse Analytics.
 - Issued by Microsoft.
- **Terraform Associate** Lisbon, PT
CERT, Terraform 2021
 - Certification for Terraform Associate, including Terraform CLI, Terraform Workflow (Write, Plan, and Create), Provision Infrastructure, Manage Infrastructure, Configure Infrastructure, Secure Terraform Automation, and Understand Terraform Cloud and Enterprise Capabilities.
 - Issued by Terraform.
- **Databricks Certified Data Engineer Associate** Lisbon, PT
CERT, Databricks 2022
 - Certification for Databricks Certified Data Engineer Associate, including Data Engineering Fundamentals, Data Engineering on Azure Databricks, and Data Engineering on AWS Databricks.
 - Issued by Databricks.
- **Certified Kubernetes Application Developer** Lisbon, PT
CERT, CNCF 2023
 - Certification for Certified Kubernetes Application Developer, including Core Concepts, Configuration, Multi-Container Pods, Observability, Pod Design, Services & Networking, State Persistence, and Scheduling & Performance.
 - Issued by the CNCF and the Linux Foundation.

Work Experience

- **Mercedes-Benz.io** Lisbon, PT
Data Engineer *March 2022 - current*
 - Setting up, managing and fixing continuous integration and deployment pipelines for the majority of the project repositories, ensuring that all applications are secure and running.
 - Deploying and managing applications on internal Kubernetes clusters, such as Airflow, GTM, and others.
 - Migrating the current Azure Data Factory orchestration engine to custom Airflow instance, including linked service connections, pipelines, and other objects.
 - Developing and maintaining custom Airflow operators and DAGs for the orchestration of the data pipelines.
 - Building and maintaining data pipelines using Azure Data Factory, Azure Databricks, and Azure DataLake Gen2.
 - Optimizing and maintaining the platform's health, including secrets management, security patches, building monitoring dashboards, and others.
- **United Nations (UNICC)** Lisbon, PT
Data Engineer *December 2021 - March 2022*
 - Support project in the deployment of changes to the infrastructure with Terraform, including Azure VideoIndexer, Azure Key Vault, Azure Data Lake, and others. This includes the creation of new resources, as well as the modification of existing ones. Ultimately, I left this role since I was not performing the job role I was originally hired for. The job role I applied for was as a data engineer, building a data platform from ground-up in a R&D project.
- **Bild Analytics (now Ascent)** Lisbon, PT
Data Engineer *August 2020 - December 2021*
 - Worked as a consultant on several client projects, such as:
 - * **Kantar - Brand Pulse:** engineered an end-to-end ingestion platform that supported different operating modes for each type of input data. The platform was fully controlled by operators via PowerShell scripts, which allowed customers to control different pipelines - each pipeline would extract-transform-load a different source or use case. Internally, scripts would call a main Data Factory pipeline which would compute the data sources on Azure Databricks. The platform had a full featured CI/CD deployment (on different environments such as DEV/TEST/PROD, each with a custom approval process) using ARM templating in conjunction with environment variables and full unit-integration testing using the unittest python module/framework.
 - * **DS Smith - Data Platform:** engineered an end-to-end engine that was fully metadata driven. The engine was controlled using metadata in the form of JSON files, that could take two forms: a form of mainfile (this file had pointers to granular processes) or granular files (which contained the steps of a given process). The metadata-driven was controlled via ADF, which also would generate steps based off a configuration files, along with their connection strings and other objects. The infrastructure was deployed via CI/CD using Terraform. Overall, over 80 different data sources were ingested with this custom metadata driven framework, built with Python.
 - * **New Signature - Research and Development:** development of a early-stages MLOps POC, presented on the Spark Summit 2021; development of a generic Quality-Checks framework, extensible to any Python project; development of common DevOps IAC boilerplate code using Terraform; development of disposable interview labs using Azure Labs and Terraform for external recruitment challenges.

- **Aubay** Lisbon, PT
Data Engineer *November 2019 - August 2020*
 - Tasked with the development of a end-to-end data platform for **UNICRE**, a credit institution in Portugal looking to replace their old engine built with COBOL called CARDPACK. The platform (now live and in-production) was designed to accurately calculate client's credit interests and taxes, with lower execution times and higher resiliency to failures. The project required high collaboration with product analysts and product owners to correctly implement the intricacies of interest time-based calculations and to avoid failing any legal requirements. The project was built with the classic DE stack on Azure, where Azure Data Factory orchestrated each data-source pipeline, which would then compute on Azure Databricks. Results were stored in Azure Data Lake Gen2 and some reports were sent via E-mail to analysts and CARDPACK for validation of results. As a requirement, no schedules could be present on Azure Data Factory, instead we used the company's internal orchestrator called Control-M. At the time, Control-M did not support Azure Data Factory, so, I was tackled with building a custom integration to ADF that allowed to schedule ADF pipeline runs to the internal schedules of the tool. Due to client technical restrictions, the integration had to be built with PowerShell. The first-phase of the project, which I was part of, only had the engine built and unit-testing, so at this stage no CI/CD was implemented. This project constituted my first go-live product.
- **Boost IT** Lisbon, PT
Junior Data Engineer *September 2018 - November 2019*
 - First experience in the Data Engineering field. Worked as a consultant on a couple client projects, such as:
 - * **Carlsberg - Carlsberg Analytics Platform:** Contributed to building and designing an highly scalable data engine using the Azure stack. Part of my contributions there include:
 - Implement data acquisition pipelines for different source systems (SAP ERP, internal API's, Flat files, and others.
 - Developing generic cleaning and transformation classes using PySpark on Azure Databricks.
 - Building a generic cube model generator PoC using C# and PowerShell.
 - Creating and maintaining data models on Azure SQL Data Warehouse.
 - Integrating data science algorithms outputs to the Carlsberg Analytics Platform (CAP) using Azure Data Lake Gen1.
 - * **Bi4All - Various clients:** maintained and developed new features for the following clients:
 - **MetLife:** long-running product for the insurance company MetLife. The product was built using the HADOOP stack of tools, mainly Yarn, Ambari, HDFS, Aginity, Ranger and Spark. In the project, I was given the role of maintaner, and as such I had to daily oversee the extraction and transformation of flatfiles sent over SFTP using Oozie workflows and generate system integration tests to validate data recieved during the various stages of the ETL process.
 - **Antas da Cunha Ecije:** contributed to the data-platform PoC for the legal company called Antas da Cunha Ecije. The PoC used partly the DE stack in Azure such as Azure Data Factory as the orchestration engine, Azure Data Lake Gen2 as the storage layer, Azure SQL Server as the database where all data was stored and finally PowerShell as the temporary IAC tool. The project was pulling flatfiles from a RDC (remote desktop connection) into the storage layer, then process the files on arrival via ADF triggers and data flows and finally store the data on the database. In this project, due to cost restrictions, I had to build a log processing system with ADF and SQL server, using stored procedures and custom API endpoints.

Skills

API's:	●●●●●
AWS cloud:	●●●●●
Airflow:	●●●●●
Apache Spark (PySpark):	●●●●●
Azure DevOps:	●●●●●
Azure cloud:	●●●●●
Bash:	●●●●●
Batch processing:	●●●●●
CI/CD pipelines:	●●●●●
Data warehousing and modelling:	●●●●●
Databricks:	●●●●●
Docker:	●●●●●
ETL pipelines:	●●●●●
GCP cloud:	●●●●●
Git:	●●●●●
Go:	●●●●●
Json:	●●●●●
Kubernetes:	●●●●●
Linux:	●●●●●
PowerShell:	●●●●●
Python:	●●●●●
Rust:	●●●●●
SQL:	●●●●●
Stream processing:	●●●●●
Terraform:	●●●●●
Typescript:	●●●●●
Vim:	●●●●●
Windows:	●●●●●
YAML	●●●●●