

WORK EXPERIENCE

Data Mining Analyst

dataEvo

August 2018 - May 2019

- Analyzing data and creating visualizations from Twitter data for political campaigns.
- Building dashboards with Shiny to present prediction results of dataEvo algorithm (devoBooster)
- Developing and improving unsupervised learning algorithms to understand and improve the performance of the dataEvo algorithm.

Data Engineer

Nubimetrics

May 2019 - December 2019

- Dealing and computing outliers or anomalous values.
- Refactoring Nubimetrics core solution migrating from Spark SQL functions to Spark native functions, creating unit and integration tests.
- Creating automation of ETL and reporting processes to provide information to different areas like marketing and sales.

Data Engineer

Mutt Data

December 2019 - Act.

- Helping with ETL solution of data migration from Teradata to Athena. Creating process in order to calculate deviation between data and report them.
- Experience as teacher in the seminar of Digital House about Apache Airflow and Apache Spark.
- Refactoring ETL process (implemented with Sqoop and Airflow previously) of data migration from Oracle to Impala (using python clients for each database engine and Airflow) in order to improve data and code quality, performance, control and monitoring.
- Building an analytical dashboard that provides an automatic diagnostic applying data analytics and non-supervised techniques with geographical information.
- Building a data product that scores customers using telco data, deploying services to support machine learning models, databases and a dashboard for data query. Create an analytical dashboard to show interpretations about machine learning models.
- Designing a functional PoC to face fraud detection problem in e-wallet platforms.
- Participation in a project to promote best practices in Software and Data Engineering stuff using GCP stack. Improve data pipelines and pyspark jobs in order to have idempotents and deterministic DAGs. Add Data Quality for data pipelines. GCP Stack: Composer, GCS, DataProc, BigQuery, Great Expectations.
- Member of onboarding staff to help new members to integrate to the company. This activity includes code reviews, mentoring following a handbook about Mutt Data practices and technologies which are used in most projects and content creation.
- Facing the tech debt in an Ad-tech project by improving data pipelines according to the best practices. Some of them: Implement the modern python data stack, idempotency, data validations, ETL factories, etc.

Independent Consultant

January 2020 - Act.

- Building a desktop application to compute metrics and to elaborate reports about irrigation systems.
- Experience in the early stages of a data-driven start-up. Productionizing a web scraper to store and transform data using AWS Batch, CloudWatch Events, Lambda, ECS, ECR implementing repositories with CI/CD pipelines.
- Designing solution that integrates social networks data to create analytical dashboards.
- Building an analytical dashboard using Apache Superset (mounted on EC2 and integrated with CI/CD pipelines) to integrate data from different sources such as RDS, Athena and Google Sheet.
- Building a PoC in collaboration with a frontend team to fetch news about health according to certain topics to centralize the content. Deploy in Azure VM to collect raw data into Azure Blob Storage and then centralize the information in a MongoDB database where dashboard gets the info in order to present the news for end users.

EDUCATION

- **Information Systems Engineer** | **Universidad Tecnologica Nacional - Facultad Regional Tucumán** (2013 - 2018)
- **Master in Management + Analytics** | **Universidad Torcuato di Tella** (2018 - 2020)
- **Solutions Architect Course (AWS)** | **Digital House** (2019)

ACHIEVEMENTS AND PARTICIPATIONS

- **7th best average at UTN FRT** (2017) - Average 8.53
- **Apache Airflow Fundamentals** (March 2021) - Astronomer Academy
- **Pipelines de Machine Learning usando datos de Spotify (UTN FRT)** (June 2021) - Topics: how to create an automated data product, which technologies are involved, what is an ETFL, what is a data pipeline. Links: https://youtu.be/QQZn_kEcj1I
- **Apache Airflow DAG Authoring** (August 2021) - Astronomer Academy
- **Desarrollo de productos de datos automatizados utilizando un stack moderno de datos** - Talk given in PyCon CL (November 2021). Recording: <https://youtu.be/2zJ0pA8Z6jk?t=6704>

TECH STACK

- Python (ML libraries, data processing, data viz, dashboarding, web development, scripting, asynchronous processing, data validation, etc).
- Scala (Functional programming, Spark, unit and integration tests).
- R (ML libraries, data processing, Shiny and data viz).
- Bash, SQL and NoSQL.
- Apache Spark, Apache Airflow, FastAPI, Apache Superset, Streamlit, MLFlow, Great Expectations, FastAPI, Docker, Docker-Compose, basics about K8s, Git workflow, basic knowledge about web development, CI/CD.
- AWS, Azure, GCP.

LANGUAGES

- Spanish (native), english (professional competence), italian (professional competence) and french (basic).