MIGUEL OZANA DE CAMPOS SILVA

Data Engineer

Tel: +55 21 98631-8960
E-mail: miguelozana@gmail.com
GitHub: https://github.com/m1guelozana

DATA ENGINEER

SUMMARY

Data Engineer with a strong focus on **Data Quality, Data Reliability, and Data Governance** in large-scale Big Data environments. Experienced in designing and implementing validation, profiling, and comparison frameworks using **PySpark** and **Azure Synapse**, ensuring data consistency, traceability, and trust across Data Lake architectures.

CORE SKILLS

- Data Engineering & Big Data Processing
- PySpark (batch processing)
- Azure Synapse Analytics
- Azure Data Lake Storage (ADLS)
- Delta Lake, Parquet, CSV
- Data Quality & Data Validation
- Data Profiling & Schema Management
- Data Governance & Metadata Management
- Azure Purview
- ETL / ELT Pipelines
- Analytics Engineering Foundations

TECHNICAL EXPERIENCE

Data Engineer

Big Data & Data Lake Engineering

- Developed and maintained PySpark notebooks in **Azure Synapse** for large-scale data processing.
- Worked with multi-layer **Data Lake architectures (bronze, silver, gold)**.
- Processed and validated datasets stored in **Delta Lake, Parquet, and CSV** formats.

Data Quality & Reliability

- Designed **generic and reusable data validation frameworks**.
- Implemented automated **data profiling** for schema, nullability, and data type analysis.
- Performed structured **comparisons between heterogeneous data sources**.
- Handled schema drift, missing columns, and data type mismatches.
- Produced **conformity and non-conformity metrics** to support audit and governance processes.

Reporting & Transparency

- Automated the generation of **technical validation reports (Excel)** for stakeholders.
- Delivered clear diagnostics and quantitative indicators for data quality issues.

Data Governance Integration

- Integrated technical validation processes with **Azure Purview**.
- Worked with metadata, asset mappings, and catalog references.
- Ensured alignment between **physical data assets** and **cataloged metadata**.

TOOLS & TECHNOLOGIES

- **Languages:** Python
- **Big Data:** Apache Spark (PySpark)
- **Cloud Platform:** Microsoft Azure
- **Analytics & Storage:** Azure Synapse, Azure Data Lake
- **Data Formats:** Delta Lake, Parquet, CSV
- **Governance:** Azure Purview
- **Reporting:** Excel (automated generation)

PROFESSIONAL PROFILE

- Strong analytical mindset focused on **data correctness over appearance**.
- Comfortable working with complex, high-volume, and imperfect datasets.
- Emphasis on **deterministic validation, observability, and trust** in data pipelines.
- Bridges the gap between **data engineering and data governance**.

TARGET ROLES

- Data Engineer
- Junior Data Engineer (Data Quality / Reliability)
- Data Reliability Engineer (DRE)
- Data Quality Engineer