
JUAN DANIEL GARCIA BELMAN

Data Engineer | ETL & Data Pipeline Development

Celaya | Querétaro, México | +52 (461) 173-3822 | danielgb331@outlook.com

LinkedIn: juan-daniel-garcia-belman | GitHub: Daniel-jcVv | Portfolio: danieljcvv-portfolio.vercel.app

PROFESSIONAL SUMMARY

Data Engineer with **3 years** of experience in ETL pipeline development, data governance, and analytics for Fortune 500 clients (Telcel, Citi Banamex) at **NTT DATA**. Skilled in Apache Airflow, PySpark, Python (Pandas, OpenPyXL), SQL Server (T-SQL), and metadata governance frameworks. Experience spans ETL pipeline development processing **millions of daily telecommunications records**, implementing governance standards across Teradata and Hive Big Data environments, and delivering automated source-to-target mapping specifications for banking reconciliation systems. Proficient in translating complex data requirements into actionable business insights using Power BI.

TECHNICAL SKILLS

ETL & Data Engineering: Apache Airflow, PySpark, ETL pipeline development & monitoring, data quality frameworks, workflow orchestration

SQL & Databases: SQL Server (T-SQL), Teradata, Hive, Query Tuning & Optimization, Data Validation

Programming & Automation: Python (Pandas, PySpark, OpenPyXL), Shell Scripting, Git

Business Intelligence: Power BI (DAX, Power Query, data modeling), Excel (Advanced - Pivot Tables, Power Query, Macros)

Data Governance & Tools: Metadata Management, Data Lineage, Data Dictionary Standards, JIRA/Confluence, Agile/Scrum, Technical Documentation

PROFESSIONAL EXPERIENCE

NTT DATA - ETL Process Developer (Data Operations) | Telcel Client

MAY 2024 - APRIL 2025

- Developed and monitored ETL data pipelines using Apache Airflow, processing **millions of daily telecommunications records**, ensuring data quality and consistency for downstream analytics through automated workflow orchestration and pipeline validation
- Performed data validation and SQL query tuning using SQL Server (T-SQL), identifying data anomalies and implementing automated validation rules that improved data integrity by **7%**
- Built Python scripts (Pandas, PySpark) for data transformation, quality validation, and distributed data processing, reducing manual validation time and enabling scalable ETL operations
- Automated operational workflows using Python and Shell scripting for file monitoring, log analysis, pipeline status tracking, and SFTP file transfers to enterprise data lakes

NTT DATA - Metadata Governance Analyst | Citi Banamex Client

FEBRUARY 2023 - MAY 2024

- Implemented data governance framework ensuring data quality, integrity, and consistency across Data Lakes, Data Warehouses (Teradata), and Data Marts (Hive) for banking operations
- Extracted and cataloged metadata using SQL queries against system tables to maintain up-to-date data dictionaries and validate schema consistency across Teradata and Hive environments
- Defined and maintained data dictionary standards for Teradata and Hive Big Data environments, managing naming convention validation by comparing change requests against enterprise standards and documenting naming conventions, data types, and business definitions
- Documented data lineage and field mappings using Excel matrices and Confluence diagrams, ensuring traceability and transparency in data flows for regulatory compliance and audit requirements
- Managed change requests through JIRA/Confluence, acting as liaison between ETL development teams and business stakeholders

NTT DATA - Business Analyst | Citi Banamex Client

FEBRUARY 2022 - FEBRUARY 2023

- Created detailed source-to-target mapping matrices in Excel for banking reconciliation processes, documenting source fields, business transformation logic (SQL queries with business rules), and target field specifications that served as technical specifications for Ab Initio ETL development team
- Developed Python automation scripts using OpenPyXL to validate **500+ monthly data attributes** across mapping matrices, cross-referencing field definitions against data dictionaries, flagging data type mismatches and naming convention violations, and generating validation reports
- Created interactive dashboards and reports using Power BI for executive stakeholders, translating complex data into actionable business insights

East West Industrial Engineering (EWIE) - Tier 1 (ZF TRW) - Industrial Process Analyst

MAY 2018 - MAY 2021

- Implemented real-time data collection system from production lines to evaluate cutting tool performance, analyzing tool life cycles and cutting parameters through Excel advanced analytics (Power Query, Pivot Tables, interactive dashboards), optimizing tool selection and cutting parameters that increased production output by pieces per hour and generated 5% annual cost savings
- Identified **5 critical improvement areas** through process analysis, implementing strategic changes that resulted in 30% reduction in monthly complaints and improved operational efficiency

SELECTED PROJECTS

► View complete portfolio with **4+ data projects** at danieljcvv-portfolio.vercel.app

Insurance Analytics Platform | github.com/Daniel-jcVv/insurance-analytics-medallion

End-to-end analytics solution analyzing insurance claims and customer data with interactive Power BI dashboards for business insights

Power BI, Python, Microsoft Fabric, Lakehouse Architecture

Enterprise Data Warehouse - Medallion Architecture | github.com/Daniel-jcVv/modern-datawarehouse-project

Production-ready Data Warehouse implementing Medallion Architecture (Bronze-Silver-Gold) processing 341K+ records with SQL Server, Python ETL pipelines, and Docker

SQL Server (T-SQL), Python, Docker, ETL Pipelines

Data Migration Pipeline | github.com/Daniel-jcVv/pipeline-data-migration-tracker

Automated data quality monitoring system with Python-based validation rules ensuring data integrity across business processes

Python, SQLite, Data Validation

PROFESSIONAL CERTIFICATIONS

- **IBM Data Warehousing Engineer** - IBM (Coursera) | June 2024
Credential: coursera.org/account/accomplishments/specialization/87W6HBAU3QDQ
 - **Google Data Analytics** - Google (Coursera) | June 2022
Credential: coursera.org/account/accomplishments/professional-cert/WLKVK7PBLVVU
 - **ETL and Data Pipelines with Shell, Airflow and Kafka** - IBM (Coursera) | January 2023
Credential: coursera.org/account/accomplishments/verify/QMZNBTZS4QMU
 - **Databases and SQL with Python** - IBM (Coursera) | 2022
Credential: coursera.org/account/accomplishments/verify/QA53WCQKDP4N
-

EDUCATION

Engineering in Manufacturing Technologies - Universidad Politécnica de Guanajuato
JUNE 2015

LANGUAGES

English: Intermediate (B1) - Technical documentation, professional communication, data presentation
Spanish: Native Speaker