



JUAN DANIEL GARCIA BELMAN

DATA ENGINEER

Celaya Guanajuato, México | (461) 173 3822 | danielgb331@outlook.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL PROFILE

Data Engineer with 3+ years of experience specialized in designing, developing, and implementing robust and scalable ETL pipelines. Developer enterprise data architectures, including Data Lakes, Data Warehouses, and Data Marts. Solid experience in data process automation, massive data load optimization, and data governance in cloud and on-premise environments. Proven ability to reduce processing times and improve data quality in critical projects.

PROFESSIONAL EXPERIENCE

NTT DATA (CLIENT: TELCEL) | ETL PROCESS DEVELOPER

MAY 2024 – APRIL 2025

- **ETL Pipeline Development:** Designed and implemented scalable ETL solutions using Python (Pandas, Apache Airflow, Apache PySpark) for massive real-time and batch data processing.
- **Automation and Orchestration:** Developed Shell scripts in Unix/Linux environments for critical process automation, implementing orchestration workflows with Apache Airflow.
- **Database Optimization:** Created complex SQL queries in Oracle Database for integrity validation, data transformation, and automated report generation.
- **Transfer Management:** Managed secure file transfers on WinSCP, implementing proactive monitoring protocols to ensure operational continuity.

Key Achievements:

- Reduced critical data processing time by 10%
- Improved accuracy in destination server data loads

NTT DATA (CLIENT: CITI BANAMEX) | DATA ENGINEER & GOVERNANCE SPECIALIST

FEBRUARY 2023 – MAY 2024

- **Data Architecture:** Implemented governance frameworks for Data Lakes, Data Warehouses, and Data Marts, ensuring quality, integrity, and consistency of enterprise data.
- **Metadata Management:** Defined naming standards and managed metadata in Teradata and Hive, optimizing traceability (data lineage) and data catalogs.
- **Technical Documentation:** Developed unified documentation for tables/views and created data-business mapping strategies for alignment with enterprise objectives.
- **Work Methodologies:** Managed user stories and requirements through JIRA/Confluence in Agile/Scrum environments, facilitating collaboration between technical and business teams.

NTT DATA (CLIENT: CITI BANAMEX) | DATA INTEGRATION SPECIALIST

FEBRUARY 2022 – FEBRUARY 2023

- **Data Mapping and Transformation:** Developed complex source-to-target mapping matrices, documenting SQL/Python transformations for high-volume banking reconciliations.
- **ETL Development:** Integrated validation processes with ETL flows in Ab Initio, facilitating precise migrations to enterprise Data Warehouses.
- **Validation Automation:** Implemented Python scripts (OpenPyXL) to validate 500+ monthly attributes, significantly reducing data interpretation errors.
- **Reporting:** Collaborated in developing reports in Cognos Analytics for business stakeholders.

PROJECT PORTFOLIO

[View Projects on GitHub](#)

Practical experience in data architectures, including: Scalable ETL/ELT pipelines, Cloud Data Lakes implementation, Data process automation and Data Migration.

EDUCATION AND CERTIFICATIONS

MANUFACTURING TECHNOLOGIES ENGINEERING

Universidad Politécnica de Guanajuato | June 2015

Professional Certifications:

- **IBM Data Warehousing Engineer** - IBM (Coursera) | June 2024

[Verify Credential](#)

- **Java Back-End Developer** - ONE Oracle Next Education | July 2024

Credential ID: 2a20416b-2197-41db-bfaf-900da6b73c98

- **ETL and Data Pipelines with Shell, Airflow and Kafka** - IBM (Coursera) | Jan 2023

[Verify Credential](#)

- **Google Data Analytics** - Google (Coursera) | June 2022

[Verify Credential](#)

Continuous Learning:

- **Ongoing Professional Development:** Active research and implementation of emerging technologies in Data Engineering, including RAG for intelligent data documentation and n8n for ETL workflow automation

TECHNICAL COMPETENCIES

Languages and Frameworks

- **Python:** Pandas, PySpark, Apache Airflow, OpenPyXL
- **SQL:** Oracle Database, Hive, Teradata, PostgreSQL
- **Java:** Spring Boot for data microservices
- **Shell Scripting:** Unix/Linux automation

ETL and Big Data Technologies

- **ETL Tools:** Informatica Power Center, Ab Initio
- **Big Data:** Apache PySpark, Hive
- **Orchestration:** Apache Airflow, Cron Jobs
- **Workflow Automation:** n8n (In Development - visual pipeline automation)

Cloud Platforms and Infrastructure

- **AWS:** EC2, S3, RDS, Lambda, Glue
- **Azure:** Azure Data Factory, SQL Database,
- **GCP:** BigQuery, Dataflow
- **Containers:** Docker
- **CI/CD:** Jenkins (basic), Github Actions
- **Remote Access:** WinSCP, PuTTY

Databases

- PostgreSQL
- Oracle Database
- Apache Hive

Monitoring

- Prometheus

Emerging Technologies (In Development)

- **MCP (Model Context Protocol):** Discovery new data sources
- **RAG (Retrieval-Augmented Generation):** Implementation for intelligent data documentation and query assistants
- **n8n:** Workflow automation for visual ETL pipeline orchestration

Collaboration Tools

- Git (Version control)
- JIRA/Confluence (Agile project management)
- Microsoft Office (Advanced Excel for analysis)

KEY COMPETENCIES

- **Data Architecture:** Design of Data Lakes, Data Warehouses, and Data Marts
- **Automation:** Development of automated and scalable ETL processes
- **Optimization:** Performance improvement in massive data processing
- **Agile Methodologies:** Experience in Scrum/Agile for data projects
- **Technological Innovation:** Active research in AI applied to data engineering (RAG, intelligent automation. MCP)

LANGUAGES

English - Intermediate Level - Technical reading, documentation, and professional communication