JUAN DANIEL GARCIA BELMAN

Azure Data Engineer

📍 Celaya | Querétaro, México | 📞 +52 (461) 173-3822 | ✉️ danielgb331@outlook.com
💼 linkedin.com/in/juan-daniel-garcia-belman | 💻 github.com/Daniel-jcVv | 🌐 danieljcvv-portfolio.vercel.app

▸ SELECTED PROJECTS

Insurance Analytics Platform
github.com/Daniel-jcVv/insurance-analytics-medallion
End-to-end analytics solution with Power BI dashboards
Power BI, Python, Microsoft Fabric

Enterprise Data Warehouse
github.com/Daniel-jcVv/modern-datawarehouse-project
Medallion Architecture processing 341K+ records
SQL Server, Python, Docker, ETL

🔗 **VIEW COMPLETE PORTFOLIO → danieljcvv-portfolio.vercel.app**

PROFESSIONAL SUMMARY

Azure Data Engineer with 3+ years of production experience building ETL/ELT pipelines for Fortune 500 clients (Telcel, Citi Banamex). Expert in Data Warehousing with Microsoft Synapse Analytics, Azure Data Factory/Synapse Pipelines orchestration, and Databricks with PySpark for distributed processing (cluster configurations up to 10 worker nodes). Proficient in Azure DevOps CI/CD with ARM Templates for multi-environment deployments (Dev/QA/Prod). Proven track record processing 5M+ daily records with dimensional modeling (Kimball) and layered architectures (Bronze/Silver/Gold).

TECHNICAL SKILLS

Cloud Platforms (Azure): Azure Synapse Analytics, Azure Data Factory/Synapse Pipelines, Azure Data Lake Storage Gen2, Azure SQL Database, Azure Blob Storage

Big Data & Processing: Apache Spark (PySpark), Azure Databricks, Delta Lake, distributed data processing, query optimization

ETL/ELT & Orchestration: Azure Data Factory/Synapse Pipelines, Apache Airflow (DAG development), pipeline monitoring, data quality frameworks

Programming: Python (Pandas, PySpark), SQL (T-SQL - advanced queries, stored procedures, optimization), Shell Scripting

Data Modeling: Kimball methodology, star/snowflake schemas, fact/dimension tables, Bronze/Silver/Gold architecture

Databases: SQL Server, Teradata, Hive, Azure SQL, Parquet, JSON, CSV

DevOps & Automation: Azure DevOps CI/CD, Infrastructure as Code (IaC) in Azure, Git/GitHub, ARM Templates, automated testing

Tools: Power BI, JIRA/Confluence, Agile/Scrum, technical documentation

PROFESSIONAL EXPERIENCE

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

MAY 2024 - APRIL 2025

- Implemented Data Warehousing architecture using Microsoft Synapse Analytics as centralized repository, designing and deploying data pipelines with Azure Data Factory/Synapse Pipelines to orchestrate ingestion, cleaning, and transformation of telecommunications records from on-premise SQL Server systems to Azure Data Lake Storage Gen2 with Parquet format for batch processing workflows
- Developed and optimized PySpark jobs on Azure Databricks (Standard_DS3_v2 clusters with 2-8 worker nodes) for distributed data processing, implementing Delta Lake for ACID transactions, broadcast joins for dimension tables, and DataFrame caching strategies to improve performance on large-volume datasets
- Implemented dimensional data models following Kimball methodology with layered architecture (Bronze raw Parquet/Silver cleansed Delta/Gold aggregated Delta), designing and maintaining fact and dimension tables in Azure SQL Database with columnstore indexes optimized for analytical queries
- Contributed to CI/CD pipeline implementation using Azure DevOps for automated deployment of Azure Data Factory pipelines across environments, working with Infrastructure as Code patterns including ARM Templates and YAML-based pipeline definitions to enable version control and streamline deployment processes
- Performed SQL query optimization using T-SQL on Azure SQL Database, tuning queries with CTEs, window functions, and non-clustered indexes to improve performance on large fact tables with millions of rows
- Monitored production pipelines using Azure Monitor with custom Log Analytics queries and Alert Rules, troubleshooting data loading issues in Azure Synapse and resolving incidents within established SLA timeframes

NTT DATA - Metadata Governance Analyst | Citi Banamex Client

FEBRUARY 2023 - MAY 2024

- Implemented data governance framework using Azure Purview for automated data discovery and classification across Azure Data Lake Gen2, Synapse Analytics, and hybrid environments, developing Python scripts with PyODBC for metadata extraction from Teradata (500+ tables) and Hive (200+ tables) system catalogs to maintain centralized data catalog for banking operations
- Established metadata management processes by querying system tables (DBC.TablesV, INFORMATION_SCHEMA, sys.tables) to extract and standardize technical metadata, maintaining comprehensive data dictionaries in Confluence documenting 800+ critical tables with business definitions, data types, and ownership for regulatory reporting
- Designed and documented data lineage for 150+ ETL workflows using Azure DevOps Wiki and custom Python scripts to trace data flows from source systems through Azure Data Factory pipelines to Teradata analytical tables, ensuring compliance with CNBV banking regulations and supporting audit requirements
- Served as technical liaison between 5 engineering teams and business stakeholders, managing 40+ monthly metadata change requests through JIRA and translating technical data architecture concepts into business-friendly documentation for non-technical audiences

NTT DATA - Data/Business Analyst | Citi Banamex Client

FEBRUARY 2022 - FEBRUARY 2023

- Created detailed source-to-target mapping specifications for banking reconciliation processes, documenting field-level transformation logic (T-SQL CASE statements, aggregations, business rules) and data lineage flows that served as technical specifications for ETL development workflows
- Developed Python automation framework using Pandas and pyodbc to validate 500+ monthly data attributes, querying centralized metadata repository in Azure SQL Database to cross-reference field definitions, enforce naming conventions, and flag schema inconsistencies, reducing validation cycle from 8 hours to 45 minutes and improving data quality
- Collaborated with data engineering team to translate business requirements into technical specifications, ensuring alignment with Azure-based data integration standards and banking regulatory compliance (CNBV reporting requirements)
- Built operational BI solutions by developing 10+ Power BI dashboards connected to Azure SQL Database and Azure Synapse Analytics via DirectQuery, implementing DAX measures for key banking KPIs and enabling real-time, actionable insights for senior leadership.

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

MAY 2018 - MAY 2021

- Implemented real-time data collection system from 12 CNC production lines using Excel VBA macros to evaluate cutting tool performance, analyzing tool life cycles (inserts, drills, end mills) through Power Query transformations and Pivot Tables with statistical metrics (mean, standard deviation), optimizing cutting parameters (speed/feed/depth) that increased production output by 150 pieces/hour and generated 5% annual cost savings ($25K USD)
- Identified 5 critical improvement areas through Six Sigma DMAIC methodology and process flow analysis, implementing strategic changes in tool management procedures and operator training programs that resulted in 30% reduction in monthly quality complaints (from 50 to 35 incidents)

PROFESSIONAL CERTIFICATIONS

- ***Microsoft Azure Data Engineer Associate (DP-203)*** *- Self-study in progress*
- **IBM Data Warehousing Engineer** - IBM (Coursera) | June 2024 coursera.org/account/accomplishments/specialization/87W6HBAU3QDQ
- **Google Data Analytics** - Google (Coursera) | June 2022 coursera.org/account/accomplishments/professional-cert/WLKVK7PBLVVU
- **ETL and Data Pipelines with Shell, Airflow and Kafka** - IBM (Coursera) | January 2023 coursera.org/account/accomplishments/verify/QMZNBTZS4QMU
- **Databases and SQL with Python** - IBM (Coursera) | 2022 coursera.org/account/accomplishments/verify/QA53WCQKDP4N

EDUCATION

Engineering in Manufacturing Technologies - Universidad Politécnica de Guanajuato
JUNE 2015

LANGUAGES

English: Good - Technical documentation, professional communication and data presentation
Spanish: Native Speaker