

ERWIN GLENN L. CAPITAN II

BI / Data Analyst | Analytics Engineer | Power BI · SQL · Azure · Databricks · Python
Davao City, Philippines • glcapitan007@gmail.com • +63 955 135 2544 • linkedin.com/in/erwin-glenn-capitan-ii • github.com/glcapitan

PROFESSIONAL SUMMARY

Data and BI analyst who builds end-to-end data platforms on **Azure** and **Microsoft Fabric** — metadata-driven ingestion pipelines with watermark-based incremental loading, Bronze–Silver–Gold medallion architectures in PySpark and T-SQL, and dimensional models with **SCD Type 2** history tracking. I design star schemas from scratch, write production-shaped dbt on Snowflake, validate what I build, and deliver decision-ready Power BI and Tableau dashboards. Currently completing DP-700 (Fabric Data Engineer Associate). Based in Davao and open to remote, hybrid, or onsite Metro Manila roles.

TECHNICAL SKILLS

Azure & Databricks: Data Factory (metadata-driven pipelines, watermark incremental loading, Logic App alerting), Databricks (Auto Loader, Delta Live Tables, Unity Catalog governance), ADLS Gen2, Delta Lake, PySpark, SCD Type 1 & 2

Microsoft Fabric: OneLake, Lakehouse, Warehouse, Data Pipelines, Deployment Pipelines (Dev → Test CI/CD), semantic models, medallion architecture

SQL & Modeling: T-SQL, PostgreSQL, Snowflake, DuckDB — window functions, CTEs, query optimization; dimensional modeling (star and snowflake schemas), Kimball methodology, data quality validation

Analytics Engineering: dbt (staging / intermediate / marts, tests, documentation), ETL and ELT patterns, Python (pandas)

Visualization: Power BI (DAX, Power Query, semantic models, RLS), Tableau (LOD expressions, parameters), Advanced Excel/VBA (Power Query, Power Pivot)

Source Control & Tooling: Git and GitHub (branching, pull requests, code review), dbdiagram.io

PORTFOLIO PROJECTS

Azure Incremental Lakehouse — Azure Data Factory · Databricks · Delta Live Tables · Unity Catalog

2026

- Built a **metadata-driven ADF pipeline** loading five tables from Azure SQL by iterating a JSON parameter array, with watermark-based incremental extraction, conditional branching to discard empty no-op loads, and Logic App failure alerting.
- Streamed Bronze → Silver with **Databricks Auto Loader** into Unity Catalog Delta tables, factoring shared PySpark logic into a reusable transformations class; governed all storage through a managed-identity credential — no account keys or SAS tokens in code.
- Modeled a Gold star schema with **Delta Live Tables** and SCD Type 2 dimensions so fact rows join to the dimension version active at event time; **validated the result** with reconciliation counts and an integrity check confirming exactly one open row per key. Repo: github.com/glcapitan/azureproject

NYC Taxi Analytics Platform — Microsoft Fabric · T-SQL · Power BI · Medallion Architecture

2026

- Built an end-to-end Fabric platform on 6M+ NYC Yellow Taxi records — metadata-driven pipelines with a **processing_log watermark** making each run self-managing, T-SQL stored procedures transforming staging into a historical presentation table, and a semantic model feeding Power BI.
- Implemented the transformation **both as a Dataflow Gen2 and as a stored procedure** and measured both; the stored procedure cut presentation processing from ~2–3 minutes to ~30 seconds and became the production design.

Retail Medallion Pipeline — Microsoft Fabric · PySpark · Delta Lake · CI/CD

2026

- Built a Bronze–Silver–Gold pipeline with PySpark cleansing and a Gold Warehouse star schema, promoted across Development and Test workspaces as a single unit through a **Fabric deployment pipeline**, version-controlled on GitHub.

Supply Chain Analytics Platform — SQL (DuckDB) · Parquet · Power BI · Kimball Star Schema

2026

- Designed a Kimball star schema and Parquet-backed pipeline powering a Power BI dashboard that surfaced **~\$95K in trapped working capital**, with an executive briefing deck for non-technical stakeholders.

Financial Performance Dashboard — Excel FP&A · Power Query · Power Pivot · DAX

2026

- Built a two-fact star schema (actuals + budget) with Power Query unpivoting monthly P&L grids and DAX time-intelligence measures (MTD, trailing 3/6/12, YTD, prior-year) for variance and gross-margin analysis.

WORK EXPERIENCE

Freelance Data & BI Analyst | Self-Employed — Remote

2025 – Present

- Design and deliver data pipelines, semantic models, and dashboards for clients, owning requirements gathering, build, testing, and stakeholder walkthroughs end-to-end.

Teacher I (Data and Research Coordinator) | Department of Education — Davao

Mar 2019 – Apr 2026

- Coordinated school data and research reporting, consolidating enrollment and performance datasets and maintaining recurring documentation for review.

Branch Operations Supervisor | Personal Collection Direct Selling Inc.

Jan 2018 – Mar 2019

- Ran branch operations analytics — warehouse inventory levels, expenses, liquidations, stock monitoring, and staff performance — building daily KPI reporting in **Excel, VBA, and Power BI** to catch and correct operational issues with field teams.

Credit & Risk Analyst | Personal Collection Direct Selling Inc.

Aug 2015 – Jan 2018

- Analyzed the full credit and collections portfolio to assess risk, producing recurring reports that informed branch-level lending decisions.

EDUCATION & CERTIFICATIONS

BA Philosophy — Queen of Apostles College Seminary

- Microsoft Certified: Fabric Data Engineer Associate (DP-700) — In Progress
- Full-Stack Data Analytics Certification — EXCELHelpline (2024)