

Biswajit Brahmma

biswajit.brahmma@gmail.com · (+91) 6361590436 · Bengaluru, Karnataka, India
linkedin.com/in/biswa-isdm · github.com/mindwanders1993 · kaggle.com/biswajitbrahmma

Professional Summary

Senior Data Engineer with **9.5 years of total experience**, including **6.5+ years building production ETL/ELT pipelines and scalable data platform infrastructure** across retail, finance, education. Partnered with stakeholders to design **Medallion architectures** and **CDC pipelines** on **Databricks** (Unity Catalog), **Snowflake**, **AWS**; implemented **data governance** and **data quality** frameworks, optimized processing by **70%** and data reliability by **85%**.

Skills

Core Competencies: ETL/ELT Design, Data Lakehouse, Medallion Architecture, Data Observability, Orchestration, Dimensional Modeling, Stakeholder Management, Cross-functional Collaboration

Programming: Python (PySpark, Pandas, Plotly, Dash, Dask, Polars), SQL, Spark SQL

Cloud & Platforms: Databricks (Unity Catalog), Snowflake, AWS (S3, Glue, Lambda, EC2), GCP (BigQuery, GCS), Azure (Data Factory, Cosmos DB, AI Search)

Data Infrastructure: Apache Spark, Airflow, Apache Iceberg, Delta Lake, PostgreSQL, Power BI, Data Studio, dbt

DevOps & AI Tools: Docker, Jenkins CI/CD, Git, Agile/Scrum, Agentic AI (Claude Code, Ollama)

Professional Experience

Genpact

Senior Data Engineer

Bengaluru, Karnataka

August 2023 – Present

Client: Adidas (Retail — Japan)

May 2024 – Present

- Core Data Engineer and Maintainer** for the Japan Data Quantum Lakehouse, partnering with the regional DE Lead to architect, build, and govern end-to-end Medallion data pipelines and enterprise data products across the Japan region.
- Architected an idempotent, config-driven PySpark pipeline for **incremental ingestion** and **CDC (Change Data Capture)** to load wholesale sellout data for **28 partners** into Bronze, Silver, and Gold layers, **reducing execution time by 70%**.
- Developed robust data ingestion frameworks leveraging custom Python utilities (`file_server_utils`, FTP copy operators) with **exponential backoff retry logic**, state-machine audit logging, and automated email alerts.
- Executed complex **dimensional data modeling** (star schemas, 180+ column fact tables) and Gold-layer aggregations for Sales and Demand, integrating Article Master, Store, VAT, and Customer Calendar lookup tables for Power BI dashboards.
- Designed and owned **DAVIS**, an enterprise Python reconciliation framework for KPI validation (Demand, Inventory, Invoice) that reduced data quality issues by **85%** and drove adoption across Emerging Markets (EM).
- Partnered with regional business analysts to define KPI specifications, translating business requirements into Gold-layer schemas and reducing ad-hoc query requests by 40%.
- Led a zero-downtime **Unity Catalog migration** and an **enterprise-wide Bitbucket to GitHub migration** across all data product repositories, establishing strict data governance and CI/CD security compliance.
- GenAI Hackathon (Internal):** Prototyped an enterprise Retrieval-Augmented Generation (RAG) solution utilizing the **Azure OpenAI Service**, **Azure Data Factory**, **Azure Cosmos DB**, and **Azure AI Search** (Vector DB) to automate architectural guidance for legacy Big Data to Databricks migrations.

Client: Franklin Templeton (US Market)

August 2023 – March 2024

- Engineered automated ELT pipelines using an AWS event-driven architecture (EC2, Glue, Lambda, EventBridge, SNS) and Snowflake, processing financial market datasets up to 20 GB daily.
- Achieved a **75% reduction in data processing time** and drove measurable **cost optimization** by improving storage efficiency by **60%** via modern ingestion architectures.
- Reduced pipeline downtime by **45%** and incidents by **30%** by implementing proactive **data observability** and robust governance checks alongside the Data Quality team.

Knoema IT Solutions Pvt Ltd (US & German Clients)

Bengaluru, Karnataka

Senior Data Engineer II

October 2022 – August 2023

- Developed and automated ETL/ELT pipelines for finance, market intelligence, and insurance datasets using Python, Pandas, and SQL, unifying data from **10+ source formats** into analytics-ready schemas.
- Designed and built a scalable AWS S3 data lake architecture; built Apache NiFi pipelines to ingest multi-level compressed files to S3, then transform and load into Snowflake.
- Built custom tools to automate extraction from secure web applications handling 15+ file formats (text, spreadsheet, SQL, document, compressed), reducing ingestion time by **60%**.
- Enforced strict **data contracts** and implemented Slack/Jira API monitoring to alert stakeholders on pipeline failures and schema changes, improving reliability by **30%**.
- Created reusable Python scripts, preset workflow templates, and modular components for common pipeline tasks, reducing new pipeline development time by **50%**.
- Deployed, scheduled, and managed **40+ automated data pipelines** across staging and production environments via Knoema Databricks, Cron jobs, and Bitbucket, maintaining a **95% job success rate**.

Collabera (Client: ZS Associates)

Pune, Maharashtra

Senior Software Engineer

July 2022 – October 2022 (Contract)

- Designed and developed scalable ETL pipelines using Databricks notebooks, Airflow, PySpark, and Spark SQL; extracted data from the data warehouse, applied filters, and loaded processed data into AWS Aurora tables.
- Developed analytics web applications with Python, Dash, Plotly, HTML/CSS, Bootstrap, and PostgreSQL; deployed via CI/CD pipeline to Dash Enterprise with Okta Authentication.

Dhwani Rural Information Systems

Project Manager, Software Application & MIS Development

Gurgaon, NCR

April 2021 – January 2022

- Managed a 7-person engineering team to build a centralized data warehouse and automated ETL pipelines (Python, MySQL, AWS EC2), powering dashboards that delivered data-driven insights to **100,000+ users**.

Lend A Hand India

Officer Technology, Data Analysis & Project Coordination

Bhubaneswar, Odisha

April 2019 – March 2021

- Designed and implemented a GCP-native data platform (BigQuery, GCS, MongoDB) for 4 state education departments, enabling access for over 1,200 teachers and supporting data-driven policy decisions.
- Built ETL pipelines: GCS → MongoDB via BigQuery + cron jobs; ELT pipeline: SurveyCTO → GCS → MongoDB; also built a Google Apps Script teacher reporting tool used by **1,200 teachers** and **2,000 email users**.
- Led team to develop data collection mobile applications (AngularJS, .NET, MySQL) and BI dashboard web apps; authored BRD and Functional Specifications.

SRI Foundation

Co-Founder, Fundraising & Technology

Rewari, Haryana

June 2018 – December 2018

- Co-founded a social impact startup focused on Navodaya School entrance coaching and rural school transformation, successfully raising **INR 9 lakhs** in seed funding.

Piramal Foundation

Gandhi Fellow, School Leadership Development

Jhunjhunu, Rajasthan

July 2015 – May 2017

- Drove educational transformation across 39 schools by analyzing assessment data and advancing teacher capabilities, improving student outcomes from **30% to 60%** over 2 years.

Projects

Cricket Intelligence Platform | [cricket-intelligence-platform](#) *Python, SQL, Airflow, Spark, dbt, Apache Iceberg, DuckDB*

- Engineered an end-to-end cricket data platform applying scalable Medallion architecture, automated data validation, and orchestration to ingest, transform, and serve ball-by-ball performance insights.

NYC Taxi Fare Prediction | [kaggle-nyc-taxi-fare-prediction](#) *Python, Pandas, Scikit-Learn, XGBoost*

- Built an end-to-end Machine Learning pipeline using Linear Regression, Ridge Regression, and XGBoost to predict NYC Taxi fares based on massive geospatial datasets.

Education

Indian School of Development Management (ISDM)

Post Graduate Diploma in Development Leadership (PGP-DL)

Noida, NCR

July 2017 – June 2018

Jadavpur University, Faculty of Engineering & Technology

Bachelor of Engineering in Power Engineering

Kolkata, West Bengal

June 2011 – June 2015