ADITYA GURUNG

Email: aditya.gurung03@outlook.com | Phone: (701) 990-9633 | Location: Bangalore
LinkedIn: linkedin.com/in/aditya-gurung | GitHub: github.com/Adityagurung

PROFESSIONAL SUMMARY

Data Engineer with a decade of experience in building and deploying scalable ELT pipelines in production. Seeking a role that blends data pipeline development and building agentic RAG pipelines.

TECHNICAL SKILLS

Programming: Python, SQL, Java

Storage: Data Lakehouse, PostgreSQL, BigQuery, MySQL, DB2, SQLite, Snowflake, Azure SQL Pool

Orchestration: Automic, Mage, Airflow

Data Ingestion: Databricks, Azure Data Factory, Talend, Azure Logic Apps, Denodo

Visualization: Power BI, matplotlib, seaborn, Grafana

Transformation: DBT, PySpark, ODI, GCP Dataproc, BigQuery

Cloud & Infrastructure: Azure, GCP, Infrastructure as Code, Docker, Terraform, Kubernetes

Data Governance: Great Expectations, Immuta

LLMs: OpenAI, Hugging Face, Mistral 7B, Phi3, Gemma, Ollama

AI Tools: Langchain, LangGraph, LlamaIndex, Agentic RAG, n8n

Vector DBs: Qdrant, Pgvector

PROFESSIONAL EXPERIENCE

Big Data Engineer
Tredence Analytics - Bangalore | 12/2023 - till now

Databricks Unity Catalog Implementation - BFSI

- Migrated 1,650+ data objects including pipelines, notebooks, and workflows to Databricks Unity Catalog within 2 months
- Achieved 60% reduction in deployment time through comprehensive automation strategies
- Created and deployed centralized feature store using Databricks Feature Store with Unity Catalog governance, creating 100+ reusable features for credit risk and fraud detection models, establishing feature lineage tracking and automated data quality checks that reduced model development cycle from 6 weeks to 3.5 weeks
- Developed feature engineering framework using PySpark and Delta Live Tables to compute time-series aggregations, behavioral patterns, and risk indicators from transaction data, delivering consistent features across offline model training and real-time fraud detection systems with millisecond response times

Order-to-Cash Data Migration – CPG

- Led and executed migration of OTC domain data (customer, product, sales, delivery, billing, transportation) from multiple SAP HANA to an Azure EDP, delivering reconciled, governed, and performant datasets for enterprise analytics.
- Extracted and loaded SAP Datasphere and SAP HANA data into ADLS Gen2 using scalable, secure pipelines with CDC, Delta Lake, and medallion layering, enabling efficient data management and near real-time OTC insights.
- Engineered high-throughput extractions from SAP HANA using partition pruning, filter pushdown, and parallel reads; introduced CDC/watermarking to cut full loads and reduced SAP system load by 60 percent.

Senior Data Engineer
IBM - Bangalore | 10/2021 - 12/2023

Enterprise Data Platform Development - Consumer Electronics

- Led a team of 4 Data Engineers, providing mentorship and guidance on project execution while partnering with a Solution Architect to develop and implement architectural solutions. Managed data ingestion from diverse sources including AWS S3, MS CDM, SharePoint, storage accounts, and on-premises servers, and proposed a GIT branching strategy and repository structure for a client new to Azure.
- Developed Python scripts for data transformation and validation processes across multiple consumer electronics product lines
- Implemented Spark-based data processing jobs for product analytics and performance metrics

Enterprise Data Platform Built and Migration - BFSI

- Developed end-to-end Enterprise Data Platform using Azure Data Lake Storage Gen2 and Delta Lake architecture

- Built automated ingestion pipelines for heterogeneous sources including JSON APIs, CSV files, Parquet data, XML documents, PostgreSQL, and Oracle DB
- Created metadata-driven framework with dynamic control tables for parameterized loading
- Implemented CDC patterns and scalable ADF pipelines with Databricks clusters for distributed PySpark transformations
- Configured Snowflake as the cloud data warehouse solution for centralized financial reporting and analytics workloads

CX Build and Data Migration - HLS

- Integrated Salesforce with Einstein Analytics for enhanced reporting capabilities
- Executed patient data migration ensuring HIPAA compliance and data accuracy
- Built PySpark jobs for patient data cleansing and transformation during the migration process
- Set up Airflow workflows to automate daily data synchronization between healthcare systems

Data Engineer
Accenture - Bangalore | 04/2020 - 10/2021

Credit Union Data Migration - BFSI

- Analyzed churn risk for Credit Union using advanced analytics, achieving 30% reduction in churn risk
- Developed key funnel metrics and deployed SSIS packages for data transformation
- Built Python scripts for data quality assessment and member profile validation, processing 150+ credit union member records
- Used Spark for processing transaction data and calculating aggregated metrics for business reporting

ETL Developer
Lanesquare Technology - Bangalore | 06/2018 - 03/2020

- Managed data ingestion from 5 databases and configured Power BI Gateway with Azure
- Monitored data pipelines ensuring quality, governance compliance, and reliable accessibility
- Conducted unit testing and code reviews for data process accuracy

Technical Consultant
Concentrix - Gurgaon | 12/2015 - 03/2018

- Developed stored procedures, functions, packages, triggers, and views in distributed environments
- Performed database optimization and executed SQL scripts in BSS model

EDUCATION

Master of Science in Computer Science
Dr. C V Raman University, Kota | 2012-2014

Bachelor of Science in Computer Science
Graphic Era University, Dehradun | 2009-2012

CERTIFICATIONS

- dbt Labs: dbt Fundamentals (08/2025)
- Databricks: Accredited Gen AI Fundamentals (11/2023)
- Databricks: Accredited Lakehouse Fundamentals (12/2022)
- Microsoft Certified: Azure Data Engineer Associate (05/2022)

PERSONAL PROJECTS

The Football Pundits (2025) - GitHub Link

- Built full-fledged end-to-end data pipeline from scratch that feeds into interactive PBI dashboard
- Delivers football performance metrics helping decision-makers

Brahman.ai - Your Smart Travel Assistant (2025) - GitHub Link

- Built advanced Retrieval-Augmented Generation (RAG) platform for seamless travel planning and destination exploration
- Integrates vector databases, diverse search techniques, and large language models
- Delivers precise, context-aware, and insightful travel recommendations