ADITYA GOYAL

DATA ARCHITECT

goyaladitya1994@gmail.com (91) 979 986 6777 linkedin.com/in/goyaladitya777/ Gurgaon

Senior Data Engineer and Data Architect with 9+ years of experience designing and owning scalable data platforms, ETL pipelines, and analytics products for global consulting clients including McKinsey and BCG. Expertise in Azure Databricks, PySpark, SQL, cloud-native data architectures, CI/CD, and data quality frameworks. Proven ability to integrate machine learning models into production pipelines, improve performance by 30–40%, and lead cross-functional engineering teams. Strong background in Revenue Growth Management (RGM), Marketing Mix Modeling (MMM), and optimization-driven analytics.

RELEVANT SKILLS

Programming Languages: Python, SQL, R
Big Data & Processing: Apache Spark, PySpark, Azure Databricks, ETL Pipelines
Data Platforms: Unity Catalog, Azure Data Lake Storage (ADLS), PostgreSQL, SQL Analytics
Cloud & DevOps: Microsoft Azure, AWS (EC2, Lambda), CI/CD Pipelines, GitLab
Data Quality & Testing: Unit Testing, Integration Testing, Regression Testing, SonarQube, Code Coverage
Analytics & ML: Forecasting, Optimization, Random Forest, Marketing Mix Modeling (MMM), Revenue Growth Management (RGM)
BI & Tools: Power BI, Jira, Confluence

WORK EXPERIENCE

McKinsey & Company, Gurgaon (Contract via Pratham Software)
Data Architect
2022 - Present

RGM ETL Platform (Azure Databricks)
Multi-client Revenue Growth Management (RGM) data platform supporting sales and P&L analytics across consulting engagements.

- Architected a configurable ETL platform on Azure Databricks to ingest sales and P&L data from ADLS and publish curated datasets to Unity Catalog and PostgreSQL.
- Enabled concurrent execution of multiple client pipelines within a single environment, allowing parallel processing on shared infrastructure.
- Implemented JSON/YAML-driven configuration to support client-specific data specifications, transformation logic, and formula changes without code modifications.
- Integrated a Random Forest–based imputation module to estimate elasticities and coefficients for new products during ETL reruns.
- Led a version-2.0 overhaul of the ETL framework, improving end-to-end pipeline performance by 30–40% while increasing stability and configurability.
- Established unit tests, Databricks-based integration tests for catalog I/O, and end-to-end regression tests benchmarked against baseline datasets to achieve a coverage above 80%.
- Automated orchestration using Databricks Workflows and automated job creation for scheduled and on-demand execution reducing cost by 30%.
- Built a data-sanitization module to anonymize sensitive client data and meet confidentiality requirements.

Data Quality, CI/CD & Coverage Tooling

- Designed an integration testing framework for Databricks pipelines and implemented CI/CD workflows to validate, test, and deploy ETL jobs reliably.
- Integrated SonarQube to enforce code-quality standards and surface technical debt early.
- Delivered a proof-of-concept test coverage framework based on the Python coverage library, customized to measure line-level coverage for functions and classes defined in Databricks notebooks.

People & Platform Leadership (Pratham Software)

- Managed 3–4 engineers in a general managerial capacity, supporting staffing, performance feedback, and technical guidance across multiple client engagements.

Boston Consulting Group (via Pratham Software)
Senior Data Analyst
2017 - 2022

Marketing Catalyst – MMM-Based Marketing Investment Platform
End-to-end marketing investment planning and optimization platform based on Marketing Mix Modeling (MMM).

- Led development of Marketing Catalyst, supporting strategic, tactical, and hybrid (mix) planning use cases for enterprise clients.
- Built ratio-based optimization models in R for annual, market-level planning and gradient-based optimization for granular weekly planning across markets, channels, campaigns, and activities.
- Redesigned optimization algorithms in R to improve performance by 8× for large-scale tactical planning scenarios.
- Migrated core tactical optimization logic from R to Python (Pandas, NumPy), implementing a custom gradient-based algorithm that delivered an additional 4× performance improvement.
- Owned end-to-end product delivery for the Mix module, acting as Product Manager to gather requirements, define roadmap, and lead a 4–5 member team from conception to release.
- Served as technical lead for the optimization service across modules, mentoring 1–2 developers and guiding implementation decisions.

Cloud & Platform Optimization Initiatives

- Migrated batch workloads from AWS EC2 to an event-driven Lambda architecture, reducing infrastructure costs by ~30% while maintaining reliability.
- Streamlined internal reporting and data operations through automation using Python, SQL, and VBA, saving 50+ hours of manual effort.
- Instituted standardized Git workflows and documentation practices to improve collaboration and long-term maintainability.

People & Platform Leadership

- Balanced engineering leadership with business-analysis responsibilities, managing timelines, requirements, and stakeholder expectations.

EDUCATION

Indian Institute of Technology, Roopnagar
B. Tech.
Mechanical Engineering
2012 - 2016