

# Debasish Chakraborty

debasish2392000@gmail.com — +91-8102062588 — linkedin.com/in/chakrabortydebasish/

**Senior Data Engineer | Cloud Data Engineer** — Data Engineer with hands-on experience designing, building, and operating large-scale batch and real-time data platforms. Specialized in cloud-native ETL architectures, high-throughput data ingestion, and analytics-ready data warehouses using Spark, Kafka, Airflow, AWS, and GCP. Proven ability to deliver reliable, secure and scalable data systems processing 50 M+ records/day for telecoms and marketing analytics use cases.

## Skills

**Programming** Python, SQL, Shell Scripting, Java.

**Big Data** Apache Spark, Kafka, Hadoop, NiFi.

**ETL** Apache Airflow, DBT, Talend.

**Cloud Platforms** AWS (S3, Glue, Lambda, Step Functions, CloudWatch), GCP (BigQuery, Dataflow, Cloud Functions).

**DevOps** Docker, Git, GitHub Actions, CI/CD, Linux.

**Databases** BigQuery, Snowflake, PostgreSQL, MySQL, Oracle, Teradata.

**Soft skills** Scalable architecture design, Problem Solving, Critical thinking.

## Experience

### Comviva

**Data Engineer (Senior Engineer – Product Development)**

Dec 2024 – Present

- Designed and operated **cloud-native ETL pipelines** processing **50 M+ records daily** across multiple telecom clients using Talend, Spark, and SQL, with built-in validation, de-duplication, and audit logging.
- Architected scalable ingestion platforms on **AWS** leveraging S3, Lambda, Glue, and Step Functions, enabling schema discovery, fault tolerance, and automated recovery.
- Built **near real-time data pipelines** using **Kafka + Spark Streaming**, reducing event processing latency and improving system resiliency.
- Orchestrated batch and streaming workflows using **Apache Airflow**, implementing retry logic, SLA monitoring, and auto-healing DAGs.
- Migrated on-premise data pipelines to **GCP BigQuery**, implementing **DBT-based transformation layers** and CI/CD-driven deployments.
- Implemented IAM-based access control, encryption-at-rest, and secure data handling aligned with telecoms data privacy and compliance requirements.
- Developed SQL-based analytical datasets and KPIs supporting campaign analytics, funnel tracking, and cohort-based reporting.
- Deployed Dockerized services with **GitHub Actions CI/CD**, ensuring repeatable, version-controlled releases across environments.

**Engineer (Product Development)**

Dec 2023 – Dec 2024

- Improved data pipeline efficiency by **40%** by optimizing Talend job design, parameterization, and partitioning strategy.
- Developed Kafka-Spark ingestion pipelines for behavioral analytics and real-time trigger generation.
- Designed automated **GCP-based data lakes** integrated with downstream BI and analytics platforms.
- Applied Python-based ML models (Scikit-learn) for churn prediction and user segmentation to enhance campaign targeting accuracy.
- Designed **API-integrated solutions** that ingested event streams and powered dashboard metrics.

**Engineer Trainee**

Jul 2023 – Dec 2023

- Implemented credential masking and encryption in ETL workflows to secure sensitive telecoms customer data.
- Optimized SQL queries and schema across PostgreSQL, MySQL, and Oracle, improving reporting performance.
- Supported monitoring, logging, and audit trail frameworks to ensure data lineage and traceability.

## Projects

---

### MRTM

*Mobilytix Real-Time Marketing*

JUL 2023 – Present

Comviva Product

- Core data engineer for a real-time marketing analytics platform serving 7+ telecom clients.
- Designed Kafka + Spark pipelines processing high-volume behavioral events.
- Built cloud-native ETL workflows on AWS/GCP with Airflow orchestration.
- Enabled A/B testing, audience segmentation, and campaign analytics use cases.
- Built real-time logging and monitoring dashboards reducing incident resolution time by 40%.

### VOID

*An University Related Chatbot to answer any Query*

Jan 2022 – Jan 2023

[ASSIC-IEEE]

- Developed an NLP chatbot backend using Flask and RESTful APIs to serve real-time educational queries.
- Implemented intent classification and entity recognition using NLTK and Scikit-learn with accuracy of 96%.
- Built contextual dialog flow for better query resolution and conversation memory.
- Trained model on a custom dataset to capture university-specific questions and structured topics.
- Integrated fallback handling, session control, and routing logic to manage edge cases.
- Deployed the application to a cloud server with proper logging and fail-safe response handling.
- Published research as an IEEE paper outlining system architecture, dataset preparation, and NLP evaluation.

---

## Education

**C.V. RAMAN GLOBAL UNIVERSITY, Bhubaneswar, Odisha, India**

2019 – 2023

*Bachelor of Technology in Computer Science and Engineering*

GPA : 8.87/10

### Publications

IEEE Author - "VOID - An Assistant (Chatbot) to Education Queries System." ([link](#))

### Certifications

- Cloud - AWS Certified – Academy Cloud Foundation.
- CCNAv7 - Introduction to Networks
- CCNAv7 - Switching, Routing and Wireless Essentials
- CCNAv7 - Enterprise Networking, Security and Automation