

# KUNAL DAS

[iamkunaldas333@gmail.com](mailto:iamkunaldas333@gmail.com) • [+91-9748799116](tel:+91-9748799116) • [Kunal Das | LinkedIn](#) • Kolkata, India

---

## PROFILE SUMMARY

**Data Engineer with 5 years of experience in Data Engineering**, specializing in **SQL, Python, and AWS** for designing, building, and optimizing **ETL pipelines, data integration, and analytics workflows**. Proficient in **PySpark, AWS Glue, Amazon Redshift, S3, and Tableau** with expertise in **data modeling, database optimization, real-time data processing, automation, and data visualization**. Skilled in developing **scalable data architectures**, implementing **data quality frameworks**, and enabling **business intelligence (BI)** through efficient **reporting solutions and dashboarding**.

---

## WORK EXPERIENCE

### [KPMG Assurance and Consulting Services LLP, India](#)

Kolkata, India

**Job Title: *Consultant***

**June 2025 - Present**

- Working as a **Lead Data Engineer** for a major UK client, guiding a team of 7 in building end-to-end data pipelines on an on-prem Hadoop platform.
- Alongside project delivery, actively involved in architecting and developing a reusable Data Quality Platform on GCP, enabling standardized validation, monitoring, and automated data quality checks across domains

### [DELOITTE USI](#)

Kolkata, India

**Job Title: *Consultant***

**May 2025 – June 2025**

- Revamped the legacy on-prem IBM MQ integration by developing a modern data pipeline using AWS SQS for Reltio data ingestion into the ODS, enhancing reliability and reducing lost event debugging efforts by 50%.
- Designed and implemented an automated notification system using AWS EventBridge, Lambda, and SNS to alert users about the expiration of ACM Private Certificates.

**Job Title: *Analyst***

**May 2023 – Apr 2025**

- Engineered a cloud-native AWS architecture (EC2, S3, RDS Aurora, Redshift, Lambda) to integrate Salesforce data, laying the foundation for a scalable and secure data ecosystem used by 3 business units.
- Designed and Developed Event-Based CDC process to integrate near real time Salesforce Data into Operational Database (RDS Aurora).
- Constructed data pipelines using PySpark and AWS Glue to automate the ingestion, transformation, and delivery of 150GB+ structured and semi-structured datasets.
- Implemented JWT-based authentication mechanisms across 3+ internal APIs to standardize access control and enhance identity security in alignment with OAuth2 protocols.
- Developed a reconciliation system between ODS and Data Mart to match source & target metrics which improved data accuracy by 40 % and reduction of manual efforts by 20 %.
- Improved Materialized View performance by **95%** and implemented post Go-Live enhancements by **collaborating with Salesforce and MuleSoft teams** to boost system efficiency and user experience.
- Directed the team, ensuring 100% on-time delivery of project milestones by coordinating cross-functional efforts, managing daily workflows, and serving as the SME for ongoing tasks.

**Title: *Associate Analyst***

**Nov 2020 - Apr 2023**

- Streamlined and maintained database CI/CD deployments across all Non-Prod and production environments, enabling seamless migration of 300+ tables and schema updates through automated pipelines.
- Integrated 30+ CSV and JSON files from 6+ different source systems, using AWS Lambda to enable scalable querying through Athena.

- Developed and deployed interactive Tableau dashboards for an insurance client, eliminating manual effort by 60% by replacing legacy Excel-based reporting with automated visual analytics.
- Designed and implemented automated Daily Load Status dashboards, reducing report generation time by 80% by transitioning from static reports to dynamic Tableau dashboards.

---

## PROJECT

- Designed and built a cloud-based data pipeline to ingest, process, and transform sales data using PySpark and AWS S3, demonstrating end-to-end data engineering capabilities on cloud infrastructure.
- Utilized AWS S3 for storing raw and transformed Parquet files, ensuring high durability and scalability across multiple data partitions.
- Implemented PySpark jobs to read, transform, and write large datasets, optimizing data processing performance by applying schema evolution handling and lazy evaluation techniques.
- Resolved inconsistencies in data schema by integrating schema validation logic, preventing pipeline failures and ensuring continuity when new columns were introduced.

---

## EDUCATION

**Bachelor of Computer Application**  
Institute of Engineering and Management

**May 2017 - Jun 2020**

**Class XI & XII**  
The Aryans School

**Apr 2015 - Mar 2017**

---

## SKILLS

**Programming Languages:** Python, SQL

**Big Data & Cloud:** Apache Spark, Hadoop, AWS (S3, EC2, EMR, Lambda, **RDS Aurora**, Redshift), BigQuery

**Databases:** PostgreSQL, MySQL, SQLite, **AWS RDS Aurora**, Redshift

**ETL & Data Engineering:** Extract, Transform, Load (ETL), Apache Kafka, AWS Glue, OLAP & OLTP

**Visualization & Analytics:** Tableau, Matplotlib, Seaborn

---

## ACHIEVEMENTS & CERTIFICATIONS

- **Secured** Runner-Up position in Deloitte Hackathon (2024)
- **Received the Outstanding Awards 3 times**, for leading high-impact cloud data engineering initiatives.
- **Honored with Applause Awards 5 times**, for consistently accelerating team delivery timelines.
- **Director's Award for Academic Excellence** for consistently ranking in the top 5% of performers.
- **Academy Accreditation – AWS Databricks Platform Architect** (2025)
- **Academy Accreditation – Databricks Fundamentals** (2025)
- **AWS Certified Cloud Practitioner** (2023)
- **Tableau Desktop Specialist** (2021)
- **Google Data Analytics Professional Certificate** (2021)