

Rahul Dekarmakar

Address: Kolkata, India | Mob: 9007793125 | [Github](#) | [Portfolio](#) | Email: rahul.dekarmakar@gmail.com | [Linkedin](#)

Profile Summary

- Data engineering lead with 12+ years of experience in designing/building scalable, optimized data intensive applications using Apache **Spark**, **Hadoop**, **HDFS**, Python (**PySpark**), **Databricks**, **Structured Streaming**, Kafka, Airflow, ETL tools (**Informatica** PC/IICS/BDM, StreamSets), Cloud Services (**Azure** – ADLS/Blob Storage/Databricks/EventHub/DevOps/Synapse etc, **AWS** – S3/Lambda/RDS/SNS/Redshift etc), **RDBMS** (Oracle, Teradata, PostgreSQL, SQL Server, Greenplum), **Delta Lake**, Open File Formats (**Avro**, **Parquet**, **ORC** etc), SaaS Applications (Salesforce, Workday), **CI/CD**, **GitHub** and **UNIX Shell Scripts**.
- Have strong expertise in Data Warehouse (**DWH**) ETL advanced concepts, data profiling, **performance tuning** and **optimization**, handling **big data** sets, performant and **complex SQL** writing. Built big scale systems to **ingest raw** data, perform **exploratory data analysis**, data munging, **outlier** detection, writing business rules for data quality, concept alignment across various sources and ultimately pushing clean data to downstream consumers.
- Experience in **data modelling** and good understanding of both **Data Vault** and **Dimensional** model in BI/DW context.
- Hands-on experience in reporting tool **Tableau / Power BI**.
- **Data Science** and **Machine learning** enthusiast with necessary math, statistics, and algorithm knowledge. Have hands-on knowledge on **image processing** using python modules and **predictive** modelling.
- Full project lifecycle (SDLC) involvement (Waterfall and **Agile**) from POC, customer requirements gathering, analysis, design, documentation, coding, continuous integration, testing, code deployment.
- Solid team leading, team building experience, managed priorities simultaneously and very good experience in connecting with cross-functional teams across geographies.
- Ready to learn new technologies and a resourceful individual with a proven track record in implementing successful deliverables. Curious, self-motivated & a self-starter with a 'can do attitude'.

Work Experience

- **Mindera** Kolkata, India
Lead Data Engineer Jul, 2023-Present
 - Led a team of 7, supervising the design and development of an enterprise-wide data analytics platform which increased **data accessibility** using PySpark, Azure, Databricks, Azure Data Factory, GitHub, ML, Logic Apps, ADLS Gen2 and Power BI.
 - There was 900+ data pipelines that has been developed catering to different **subject areas** (HR, Finance, eCommerce, Demand, Logistics etc) with **integrated data quality** and data governance requirements. Delivered a milestone mvp within a very short span of time ahead of schedule resulting **20% reduction in budget**.
 - Developed **orchestration framework** and **utility functions** to facilitate reducing manual **efforts of reconciliation** during sprint cycles by up to **40%**.
 - Created YML **crawler** functions using python to generate lookups to optimize by refactoring data pipeline workflows which resulted in **20% Databricks cluster cost saving**. Received client appreciation award.
 - Developed and designed application for extracting required key information from contract documents for **Third party Logistics** using opencv, pytesseract, regex etc python modules **image processing** techniques.
 - Launched **machine learning initiatives**, reducing **predicting errors** by **33%** for **demand forecast** in inventory management.
 - Worked with product owners for **sprint backlog refinement**, assigning sprint stories to team according to priority, code review, smooth delivery of tasks in stipulated timeframe with quality and defined metrics. Maintaining **optimum delivery velocity** through iterative sprint improvements and **stakeholder management** across all pillars of business.
 - Provided technical **mentorship** to junior team members, did code and design reviews as well as enforced coding standards and best practices.
- **Maersk** Kolkata, India
Senior Data Engineer Dec, 2022-Present
 - Designed a consumer framework for WDP platform based on **Spark Structured Streaming** framework that is suited for both batch as well as streaming scenarios with **integrated DQ** component.
 - Led a team of 5 Data engineers for end-to-end development for **WDP** data pipeline and dashboard/visualization building using PySpark, Azure, Databricks, ADF, GitHub, ADLS Gen2, Power BI in Agile methodology.
 - Driven **Metric Analysis** work to help **EBIT** improvement, trying to figure out end-to-end data lineage for **Finance** data set and creating a centralized master table for **Geo data/Site** master that can be accessed across organization.
 - Data pipeline/process **optimization** using best practices standards and guidelines to have better ROI.
 - Provided technical leadership and hands-on implementation role in the areas of data techniques including ingestion, modelling, integration, refactoring, and orchestration.
- **Eli Lilly and Company** Kolkata, India
DE Manager - Solution Architect Aug, 2022-Dec, 2022

- Led a cross functional team for end-to-end development for **EDP** (Enterprise Data Program) data pipeline building using **PySpark, AWS, Databricks, BDM, Git** in Agile methodology.
 - Working on **BDM dynamic mapping** template for rapid ingestion framework to data lake.
 - Designed central **data platform architecture** for single source of truth for entire org with built in analytical capabilities.
 - Helped establish best practices standards and guidelines which increased sprint delivery efficiency.
 - Provided technical leadership and hands-on implementation role in the areas of data techniques including ingestion, modelling, integration, refactoring, and orchestration.
-
- **Deloitte Consulting** Hyderabad, India
Lead Data Engineer *Oct,2018-Aug,2022*
 - Performed wide variety of responsibilities as Senior Data engineer for various client engagements for Auto Insurance, Consumer Retail, Pharma and Tech Industry in different types of activities i.e., Ingestion, migration, Refactoring.
 - Designed and developed highly optimized Data Pipelines using **PySpark** on Databricks cluster for **Patient360 analytics** application for a pharma client involving huge data sets (~2.5 billion of records) catering to ~9900 US pharmacy stores.
 - Led and implemented **complex ingestion** (Batch and Streaming), **curation** data pipelines to Azure **Lakehouse** from 25 sources systems (Breakdown, Call, Complaints, Staff, vehicle, Third Party, Insurance etc) involving **1000+ entities** sources (oracle/ SQL server/ PostgreSQL/ csv/ json/ xml) for transactions and **stream of events** within a very short span of time ahead of schedule resulting **20% reduction in project budget**. Received Applause Award.
 - Understanding the requirement and priority from business SME. **Effort estimation** of deliverables, Data **modelling**, creating ETL framework as per business rules, validating data/rules, developed, and **implemented ABC** (Audit, balance, control) mechanism using shell scripts. Created **scheduling** framework using shell scripts, code optimization, prod deployment using **dev-ops** pipeline using **GitHub** repo.
 - For a pharma major, Led data migration from Legacy reporting and analytics capabilities for the Global **Medical Affairs** (GMA) data to Enterprise Data Backbone (EDB) cloud platform in AWS. Delivered the project with **ZERO UAT defects**.
 - Provided technical **mentorship** to junior team members, did code and design reviews as well as enforced coding standards and best practices.
 - Involved in creation of decoupled, distributed, high available **Data Lake** in **Agile** method along with multiple Feature Teams parallelly for tech giant supply chain data. Refactored and resolved recurring issue for concurrency conflicts that garnered millions of saving and 100+ **monthly incident reduction**.
 - Worked for a food chain major helped in creation of **Pricing Analytics solution** for strategic pricing recommendations engine that delivered Franchisee Owner/Operators quarterly pricing updates that enable them to optimize their pricing to drive a more profitable business. Did 30% **performance improvement** of ETL's using **push down optimization** and stored procedure. Created **UDF** using **regular expressions**.
-
- **Tata Consultancy Services** Kolkata, India
Technical Lead *Jun,2016-Oct,2018*
 - Led a team of ten resources and driving the project from ground up for Advanced Money Lending system to create and Basel Lending DataMart BLDM for Banking major as per BASEL II Compliance. Designed/Developed ETL code, updating design documents HLD, LLD, peer review, unit/system testing, code migration and post go live warranty support.
 - Automated manual data reconciliation process between various sources, Increasing the data accuracy and efficiency by 45%.
 - For a pharma client R&D portfolio managed entire project life cycle end to end successfully. Executing the plan for monthly release to production environment to accommodate changes in Datavault model in Teradata EDW. Involved in core group for continuous improvement plan.
 - Created coding standards that helps team members easier to read and review.
-
- **IBM** Vienna, Austria
Onshore Lead *Jan,2015-May,2016*
 - As onshore co-ordinator functioned as a bridge between offshore and client and helped built Kunden Datamart and mobile subscriber stats DM for a Telco Major in Europe.
 - Increased the efficiency of data fetching by approximately 30% using query optimization and indexing.
 - Scheduled to run workflows in most optimized way as per the dependency specified.
 - Implemented SCD2, Fact & dimension data load in Informatica, Teradata in Agile method.
-
- **Cognizant** Kolkata, India
ETL Developer *Jan,2013-Dec,2014*
 - Worked to integrate data from various sources in ODS to Data warehouse for an insurance client.
 - As an ETL developer where I need to develop informatica mappings, sessions, workflow and to do unit testing of data and transformation logic. Also engaged in informatica performance tuning activities. Developed reconciliation mechanism and shell scripts to automate the ETL process. Worked as Offshore lead for a sub project (multi country rollout) under APAC-Asia and understand the requirements from onsite coordinator and deliver the same with minimal UAT defects.
 - Developed alert mechanism using shell scripts to inform customer about the progress of loading and all.
 - Used shell scripts/Cron Scheduler to trigger workflows using PMCMD commands as per dependencies.

- **Mahindra Satyam**

Pune, India

ETL Developer

May,2010-Dec,2012

- Worked as a Developer in providing ETL solutions for EDW of Telco Major. Was Responsible for identifying the business-critical attributes using the mapping sheets and map it to the fields in DWH and using it to extract, load the data and unit testing for RCDR/MCDR, CRM datasets. Developed the mappings using needed Transformations in Informatica tool according to technical specifications.
- Designed entire new source system load (UPSS) for better plan and revenue realization. Performed tuning of several long running queries. Responded with the most feasible solution for change requests. Involved in data validation and control checks. Have done automation using shell scripts for monthly billing cycle updates.

Education/Certification

- **B Tech in Computer Science** and Engg

West Bengal University of Technology; CGPA 8.28

- **Certifications:** Microsoft Certified: Azure Data Engineer Associate, AWS certified: cloud practitioner, Informatica certified: Developer Specialist