

PRAKHAR MITTAL

9823186723
prakharmittal1510@gmail.com

SUMMARY

Accomplished Data Engineer with 7+ years of experience in designing and optimizing ETL pipelines and Big Data solutions. Expertise in Spark, Hadoop (Hive), SQL, and data modeling, complemented by strong skills in Linux shell scripting and AWS. Focused on building scalable data architectures to enhance business insights and drive efficiency in enterprise data environments.

SKILLS

- Spark
- Scala
- SQL
- Hadoop (Big Data/Hive)
- Informatica
- Workload Schedulers
- Data Modeling
- Data Warehousing
- Data Migration
- Data Analysis
- Teradata and Oracle Utilities
- ETL Pipeline Processing
- Linux Shell Scripting
- AWS (Amazon Web Services)
- Performance Tuning
- Business Intelligence
- SQL and Databases

EXPERIENCE

IDFC FIRST BANK

Mumbai, India

Data Solution Analyst

02/2026 to Current

- Leading end-to-end development of scalable ETL pipelines using **Amazon Web Services EMR, PySpark, and Jupyter Notebooks**, processing high-volume banking datasets; achieved ~30% improvement in data processing efficiency and reduced latency across critical workflows.
- Spearheaded migration of enterprise data workloads from AWS (S3/EMR) to on-prem **HDFS**, ensuring seamless data transition, enhanced governance, and ~25% reduction in cloud infrastructure costs.
- Engineered automated data orchestration using **Apache Airflow**, designing DAGs for complex dependencies; reduced manual intervention by ~70% and improved pipeline reliability and scheduling efficiency.
- Optimized PySpark transformations and execution strategies (partitioning, caching, join tuning), achieving ~35% reduction in job runtime and improved cluster resource utilization.
- Developed interactive data analysis and validation frameworks using AWS Jupyter Notebooks, enabling faster debugging and reducing data validation effort by ~40%.
- Implemented data quality checks, reconciliation logic, and monitoring mechanisms across pipelines, improving data accuracy and consistency by ~85%.
- Collaborated with business stakeholders and analytics teams to design scalable data solutions, accelerating data delivery timelines by ~20% for critical reporting use cases.
- Managed and supported production data pipelines with proactive monitoring and issue resolution, maintaining ~99.9% SLA adherence for enterprise data platforms.

LTIMINDTREE

Pune, Maharashtra

Specialist - Data Engineering

05/2019 to 02/2026

- Managed and supported 100+ enterprise-scale ETL applications across the African continent using a diverse technology stack **Hadoop (HDFS, Hive, Tez), Teradata, Informatica, Unix/Linux, TWS Tivoli, CA WADE, and Airflow** ensuring 99.9% SLA adherence for critical production workloads.
- Led **Informatica migration from v10.1 to v10.4 Virtual Server Component**, optimizing mappings, sessions, and pushdown queries; achieved a 40% improvement in data processing speed and enhanced system scalability for high-volume banking data.
- Designed and optimized ETL pipelines with data quality rules, partition pruning, and join optimizations, improving data accuracy and consistency by 80 - 85% across downstream reporting and compliance systems.
- Developed and deployed scalable ETL scripts (**Python, Spark, Sqoop, Shell**) to integrate structured and semi-structured data into Hadoop, increasing data availability and enabling near real-time analytics for customer insights and fraud detection.

- Refactored legacy data models by introducing partitioning, bucketing, clustered tables (CT), and replacing Spark execution with Tez engine; achieved a 35% reduction in query runtime and 25% better cluster resource utilization.
- Drove migration of legacy **on-prem Hadoop data marts to Dell ECS (S3 compatible object storage, hybrid cloud model)** by leveraging HDFS - S3A DistCP, data tiering, lifecycle policies, encryption with JCEKS credential providers, and replication strategies. This migration: Reduced on-prem storage overhead by 45%. Enabled elastic scaling and policy-driven archiving for regulatory data. Improved data durability (11 nines) and disaster recovery capabilities with geo-replication across regions.
- Onboarded 10+ business units onto the Big Data platform, providing ingestion frameworks, access governance (Ranger/Atlas), and scheduling automation, leading to a 50% increase in enterprise-wide adoption of Hadoop/ECS for advanced analytics.

LTIMINDTREE (ON-SITE) / ABSA (BARCLAYS AFRICA GROUP)

Johannesburg, South Africa

Senior Data Engineer

06/2023 to 12/2023

- Initiated and led the Data Rail Program, transforming legacy Teradata warehouses into a scalable Hadoop-based architecture, enabling SQL-to-Spark SQL translation and reducing query processing time by 45%.
- Managed a team of 12 Big Data Engineers across on-site and offshore locations, ensuring seamless coordination between development and support, and maintaining > 98% delivery compliance.
- Converted Teradata marts and extracts for 12 African countries into Hadoop marts using Hive, Spark, and Teradata utilities, resulting in 50% faster data ingestion and 30% cost savings in infrastructure and licensing.
- Built complex business and financial analytics pipelines by integrating multi-source data, which doubled data-driven revenue potential (2x) for ABSA through improved decision-making capabilities.

EDUCATION

EXECUTIVE MBA: BUSINESS ADMINISTRATION AND MANAGEMENT

05/2026

Symbiosis International University (SCMHRD), Pune, MH

BACHELOR OF SCIENCE: INFORMATION TECHNOLOGY

05/2019

SRM Institute of Science And Technology, Chennai, TN

PUBLICATIONS

Event Study of the Effect of U.S. 2025-2026 Tariffs on the Indian Stock Market, MBA Dissertation Research Paper, 02/01/2026, This dissertation studies the measurable effect of U.S. tariff-policy announcements on Indian equity prices, utilizing a hand-collected dataset and event-study methodology.