

PRATIKSHA JAGDHANE

United States | pratikshaj@worktechmail.com | (210) 201-5170 | [linkedin.com/in/pratiksha-jagdhane/](https://www.linkedin.com/in/pratiksha-jagdhane/) | [GitHub](#)

SUMMARY

Data Engineer with 4+ years building cloud-native data platforms on AWS and GCP that reduced data latency by 40% and improved reliability by 95%. Skilled in Spark, Kafka, Airflow, Databricks, dbt, and Snowflake for designing large-scale ingestion and transformation pipelines. Experienced in optimizing SQL, PL/SQL, and Python (Pandas) workflows for advanced analytics and predictive modeling. Focused on delivering high-performance, automated data systems that power real-time insights across healthcare and e-commerce.

SKILLS

Programming: Python (Pandas, NumPy, TensorFlow, PyTorch), SQL, R Language
Data Analysis: Exploratory Data Analysis (EDA), Statistical Analysis, Data Mining, Feature Engineering, Data Modeling
ETL Process: Databricks, dbt, Apache Spark, Apache Kafka, Apache Airflow, Hadoop, Snowflake, Data Cleaning, Data Wrangling
Machine Learning: Predictive Modeling, Classification, Supervised/Unsupervised Learning, Time Series Forecasting, CI/CD for ML
Cloud & DevOps: AWS (S3, EC2, Lambda, Redshift, Glue), GCP (BigQuery, Data Lake), Azure, Spark, Kafka, Docker, Kubernetes
Databases: Oracle, MySQL, PostgreSQL, PL/SQL, PostgreSQL, MongoDB, MS Access
Data Visualization: Power BI (DAX), Tableau, Matplotlib, Seaborn, Looker, Excel (Power Query, Pivot Table, KPIs), Google Data Studio
Other Skills: GitHub, JIRA, Confluence, Agile/Scrum, Visual Studio Code, FastAPI, API Integration

EXPERIENCE

UnitedHealth Group
Data Engineer

United States
January 2025 – Present

- Architected and deployed serverless ETL pipelines using AWS Glue and Lambda to process healthcare claims and enrollment data, achieving a 40% reduction in processing latency for critical regulatory reports compared to legacy SSIS workflows.
- Integrated Databricks Spark workflows to unify transformations and batch orchestration, improving efficiency 30% and enabling scalable data prep for predictive analytics.
- Engineered a high-throughput data ingestion framework utilizing Kafka and Apache Spark to capture and process patient interaction logs in near real-time, supporting an early-warning analytics system and handling peak loads of over 10,000 events per second.
- Led Oracle to AWS Redshift and S3 data lake migration, optimizing schema design and reducing query costs by 25%.
- Implemented automated data quality checks and reconciliation processes using Python’s library like Pandas within Apache Airflow DAGs, minimizing data errors in downstream reports by 95% and ensuring adherence to compliance standards.
- Collaborated with data scientists to define and build production-ready Feature Engineering tables in Redshift, using complex SQL and PL/SQL procedures to ensure feature consistency and directly supporting a Predictive Modeling initiative to forecast hospitalization risk.

Website Vikreta
Data Engineer

India
September 2021 – July 2023

- Engineered end-to-end ETL workflows using Apache Spark, Kafka, and Snowflake to centralize sales and customer interaction data, improving pipeline efficiency by 40% and enabling near real-time analytics.
- Leveraged dbt for modular Snowflake transformations, standardizing data models and reducing ad-hoc SQL maintenance by 35%.
- Conducted exploratory and statistical analysis in Python and SQL on customer behavior datasets, generating insights that drove a 12% increase in lead conversion and a 15% boost in customer retention.
- Designed interactive Power BI dashboards with DAX and Power Query, reducing manual reporting cycles by 35 hours monthly and providing executives with actionable KPIs across finance and marketing.
- Migrated on-prem Oracle databases to AWS Redshift, reducing 28% query response times and lowering 20% infrastructure costs.
- Supported churn prediction proof-of-concept (Random Forest) by preparing features, cleaning datasets, and deploying outputs to Tableau, enabling the sales team to target at-risk customers and reduce churn by 10%.

Tech Mahindra
Data Engineer

India
August 2019 – August 2021

- Designed and maintained highly reliable ETL workflows using Hadoop and Spark for extracting, transforming, and loading customer data from legacy Siebel CRM systems into a centralized GCP BigQuery warehouse, successfully processing daily batches exceeding 5 TB.
- Performed exhaustive performance tuning on MySQL and PostgreSQL databases, optimizing slow-running queries and creating materialized views that collectively reduced the reporting time for executive dashboards by an average of 30%.
- Implemented a rigorous Data Cleaning and Data Wrangling process using Python (NumPy, Pandas) scripts integrated with Docker, ensuring high fidelity and lineage tracking for key transactional data and improving overall data governance compliance by 20%.
- Developed custom API Integration code to ingest third-party vendor data into MongoDB (NoSQL), leveraging Azure cloud functions for serverless execution, which streamlined the acquisition of new data sources and reduced integration time by 50%.
- Validated datasets and performed initial Exploratory Data Analysis (EDA) using Matplotlib and Seaborn visualizations in Python, providing the foundation for Tableau dashboards and uncovering customer usage patterns that resulting in a 10% revenue uplift.

EDUCATION

Master of Science in Data Science

August 2023 – December 2024

University at Buffalo, The State University of New York

United States

Bachelor of Engineering in Computer Engineering

July 2015 – July 2019

Savitribai Phule Pune University

India

CERTIFICATIONS

- [Google Data Analytics](#) April 2025
- [AWS Certified Data Engineer – Associate](#) December 2024
- [Machine Learning, Data Science and Deep Learning with Python](#) June 2022