

MAYUR JAIN

Long Beach, CA | 562 254 7817 | mayurjain333@gmail.com | [Github](#) | [LinkedIn](#)

Data Engineer with 6+ years of experience designing **scalable ETL pipelines**, optimizing **data models**, deploying **real-time data solutions**. Skilled in **PySpark, SQL, Cloud**, and **DevOps automation**. Delivered improved **data quality**, automated **workflows**, and enhanced **analytics efficiency**. Experienced in **predictive analytics, data validation**, and **dashboard development** to support **data-driven decisions**.

TECHNICAL SKILLS

Languages & Tools: Python, SQL, PySpark, Java, YAML, REST APIs, Linux

Cloud & DevOps: AWS (EC2, S3, Lambda), Azure, Docker, Kubernetes, Jenkins, Git, CI/CD

Data Engineering: Kafka, Spark, Azure Data Factory, Databricks, Delta Live Tables, DBT, Data Lakehouse

Databases: Snowflake, DynamoDB, MySQL, PostgreSQL, MongoDB, Redshift

ETL & Pipelines: Apache Airflow, Batch & Streaming ETL, Data Quality Frameworks

Visualization & Methodologies: Tableau, Power BI, Agile, Scrum, SDLC

PROFESSIONAL EXPERIENCE

Data Engineer (Student Assistant) | Beach Shops

April 2024 – May 2025

- **Managed data for 40K+ student, staff, and faculty accounts**, tracking **funds** and **campus services** like food, sports, and events.
- Built **dashboards and reports** to monitor **service usage**, helping improve the campus experience.
- Used **Python** and **SQL** to automate **fund tracking and reconciliation**, reducing manual work and boosting accuracy.

Lead Data Engineer | ZS Associates

September 2022 – July 2023

- Developed **real-time ETL workflows** using **Kafka** and **PySpark** to process EHR and claims data, enabling timely clinical insights.
- Ingested secured sensitive data in **Snowflake** with **HIPAA-compliant masking**, using compliant analytics for health management.
- Created **Tableau dashboards** tracking ER visits, and preventive care, contributing to **15% reduction** in avoidable hospital admissions.
- Used **PySpark SQL** for data validation across multiple sources, ensuring **99.5% accuracy** in critical healthcare reporting pipelines.
- Collaborated with analysts and clinicians to deliver actionable insights, improving **HEDIS scores** and care coordination efforts.

Senior Data Engineer | Tata Consultancy Services

July 2019 – September 2022

- **Led migration** of a **legacy fraud detection system** to an **on-premises server**, improving control over sensitive financial data.
- Built and deployed **UAT-ready financial models** with **feature selection** and **testing**, maintaining **99.5% data integrity** in production.
- Designed **fraud detection pipelines** using **PySpark** and **Databricks**, reducing **processing time by 30%**.
- Rebuilt **ETL workflows** on **Azure Databricks** and **Snowflake**, boosting **data performance** and enabling faster reporting.
- **Mentored 5 junior engineers** and introduced **CI/CD automation**, improving team delivery speed and code quality.

Software Development Engineer | ClickIndia Infomedia

August 2017 – July 2019

- Developed automated **ETL pipelines** using **Python** and **SQL** for retail transaction and sales data (5M+ rows weekly).
- Integrated and cleaned data from **APIs, web clickstreams, and store systems**, enabling **12% higher campaign ROI**.
- Wrote optimized **SQL queries** to reduce manual reporting time by **60%** with **aggregations and multi-source joins**.

EDUCATION

California State University Long Beach

Master of Science in Computer Science

May 2025

California, US

Guru Tegh Bahadur Institute of Technology

Bachelor of Technology in Information Technology

May 2019

Delhi, IN

SELECTED PROJECTS

Credit Card Fraud Detection | https://github.com/mayurjainf007/Credit_Card_Fraud_Detection

- Developed Spark and Airflow based event alerting system reducing false positives and improving detection latency.

Real-Time Sentiment Analysis | <https://github.com/mayurjainf007/Apache-Airflow-and-Machine-Learning>

- Built Twitter API-based NLP pipeline for brand monitoring with real-time data ingestion and processing using Airflow.

Heart Disease Detection | https://github.com/mayurjainf007/Heart_Disease_Detection

- Streamed patient vitals via Kafka and scored risk in near real-time with SparkML.

CERTIFICATIONS

Microsoft Certified – Azure Data Engineer Associate

Microsoft Certified – Azure DevOps Engineer Expert