MAYUR JAIN

Long Beach, CA | 562 254 7817 | mayurjain333@gmail.com | LinkedIn | Github

PROFESSIONAL SUMMARY

Data Engineer with **5+ 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 & Warehousing: Snowflake, DynamoDB, MySQL, PostgreSQL, MongoDB, Redshift
ETL & Pipelines: Apache Airflow, Batch & Streaming ETL, Data Quality Frameworks
Visualization & Methodologies: Tableau, Power BI, Agile, Scrum, SDLC

INDUSTRY EXPERIENCE

Lead Data Engineer | ZS Associates   *September 2022 – July 2023*
Client: Janssen (Oncology & Immunology), Merck (Oncology)
- Built **ML pipelines** in **PySpark** and **Dataiku** to optimize **HCP targeting** by region, payer, and product.
- Combined **medical & pharmacy claims (IQVIA, Symphony)**, **payer coverage (MMIT)** to track patient journeys, and access barriers.
- Delivered KPI dashboards (**TRx, NRx, adherence, coverage tiers**) in **Tableau**, used by brand teams to measure uptake.
- Automated **10+ Databricks pipelines** with **Azure DevOps**, enabling schema change handling without downtime.
- Ensured **99.9% data accuracy** with **Collibra** and enforced **HIPAA compliance** across 100M+ patient records.
- Improved HCP targeting **using Veeva CRM** patient-level dataset by **25%** through **model tuning** and **A/B testing**.

Senior Data Engineer | Global IT Consulting   *July 2019 – September 2022*
Client: Cardinal Health (Healthcare & Life Sciences)
- Built **Databricks pipelines** to process terabytes of **chargebacks, wholesaler, and pharmacy transactions**.
- Added **anomaly detection** to flag **supply chain and distribution fraud** with **92% accuracy**.
- Migrated legacy warehouse to **Azure Data Lakehouse**, reducing pipeline runtime by **40%**.

Client: Lloyds Banking Group (Banking & Finance)
- Developed **real-time fraud models** on payments and card transactions using **clustering** and **anomaly detection**.
- Built **UAT-ready risk models** with **99.5% data accuracy**, enabling analysts to act on alerts faster.
- Designed **secure ETL pipelines** for **fraud detection**, **risk scoring**, and **regulatory reporting**.

ACADEMIC 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 check **service usage**, helping improve the campus experience.
- Used **Python** and **SQL** to automate **fund tracking and reconciliation**, reducing manual work and boosting accuracy.

EDUCATION

California State University Long Beach	**May 2025**
Master of Science in Computer Science	*California, US*
Guru Tegh Bahadur Institute of Technology	**May 2019**
Bachelor of Technology in Information Technology	*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.

CERTIFICATIONS

Microsoft Certified – Azure Data Engineer Associate   **Microsoft Certified** – Azure DevOps Engineer Expert