

Pankaj M Sajjanar

Union City, NJ-07087 | Phone: +1 (201) 630-1764 | [Github](#) | pmallika@stevens.edu | [LinkedIn](#)

PROFESSIONAL SUMMARY

Data Engineer with 4+ years of experience in ETL development, cloud solutions (AWS, GCP), and real-time analytics. Proficient in Python, SQL, Spark, and Kafka, with expertise in data modeling, warehousing, and orchestration. Proven ability to optimize pipelines, reduce data latency, improve query performance, and integrate AI/ML models into data workflows.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL (MySQL, PostgreSQL), Java, PySpark
- **Big Data & Streaming:** Apache Spark, Hadoop, Kafka, Flink, Nifi, Real-Time Analytics, Apache Airflow, dbt, Snowflake, Talend
- **Cloud Platforms:** AWS (S3, Redshift, Glue, Lambda, EMR), GCP (BigQuery, Dataflow), Azure (Fabric, Synapse, Data Factory)
- **Data Engineering:** ETL/ELT, Data Modeling, Data Warehousing, Orchestration, Data Governance, NoSQL, Relational Databases
- **Data Visualization & DevOps:** Power BI, Tableau, Looker, Docker, Kubernetes, Terraform, CI/CD (Git, Jenkins), Dagster
- **AI/ML & Predictive Analytics:** Machine Learning Pipelines, Regression Analysis, Anomaly Detection, Time Series Forecasting
- **Security & Tools:** IAM, RBAC, OAuth, GDPR, SOC 2, Scikit-learn, TensorFlow, SageMaker, MongoDB, HIPAA

EXPERIENCE

Data Engineer, TRANSUNION

Aug 2020 - Dec 2022

- Developed ETL pipelines using Python, SQL, Java REST APIs, reducing manual data transformation by 40% and improving processing speed by 35%.
- Migrated high-volume marketing analytics workloads from AWS (Redshift) to GCP (BigQuery), cutting infrastructure costs by 37% (~\$25K annually) while enhancing scalability & performance.
- Integrated ML-based anomaly detection into ETL workflows using Python & SageMaker, reducing data quality incidents by 25%.
- Implemented real-time data processing pipelines with Apache Spark, Kafka and Airflow, reducing latency by 30%.
- Collaborated with data science teams to embed predictive models into marketing data flows, improving campaign performance by 15% and optimizing customer targeting.
- Integrated OAuth authentication for AWS Redshift, ensuring SOC 2 & GDPR compliance to secure sensitive client data.

Associate Software Engineer, TRANSUNION

Jul 2019 - Aug 2020

- Enhanced complex SQL (MySQL, Redshift) queries, improving execution speed by 40% and increasing efficiency of BI dashboards in Tableau & Looker.
- Designed & automated ETL pipelines for Media Mix Modeling (MMM), enabling real-time marketing insights for Fortune 100 clients (Walmart, AT&T, Target), improving marketing ROI by 25%.
- Built CI/CD pipelines using Git, Jenkins, Terraform, reducing deployment time by 35% & increasing release frequency by 2x.

Data Science Intern, SCRIPBOX

Jan 2019 - Mar 2019

- Engineered ETL pipelines in Python & SQL, improving financial data accuracy by 15% and automating investment analytics.
- Generated Tableau dashboards for KPI tracking & portfolio insights, reducing manual reporting efforts by 40%.
- Fine-tuned SQL queries for financial and healthcare datasets, reducing report generation time by 25%.

EDUCATION

Stevens Institute of Technology, Hoboken, New Jersey

May 2025

Master of Science in Data Science

GPA: 3.9

Relevant Courses: Big Data Analytics, Statistical Methods, Optimization for Data Science, Machine Learning, Applied Mathematics

PES University, Bangalore, India

May 2019

Bachelor of Technology in Computer Science and Engineering

GPA: 3.3

PROJECTS

Real-Time Credit Card Transaction Processing Pipeline

Jan 2025

- Developed a real-time fraud detection system using Kafka, Spark, AWS Lambda & Redshift, processing 1M+ transactions/day.
- Improved fraud detection accuracy by 28% while reducing false positives by 15%, enhancing real-time anomaly detection.

ETL Pipeline for Healthcare Data Integration

Dec 2024

- Designed HIPAA-compliant AWS Glue & Lambda pipelines to integrate 200K+ patient records from EHRs & claims databases.
- Built AWS Redshift-based data warehouse, optimizing query performance by 30%.

CERTIFICATIONS

- **AWS Certified Solutions Architect – Associate** 2024
- **IBM Certified Data Engineer** 2025
- **Google Cloud Certified - Professional Data Engineer (In Progress)**