TANISH ARORA

Riverside, CA | (951) 850-2020 | arora.tanish97@gmail.com | linkedin.com/in/tanish-arora

PROFESSIONAL SUMMARY

Data Engineer and AI/ML specialist with 5+ years of experience building enterprise-scale data pipelines, implementing machine learning solutions, and driving cost optimization initiatives for testing environments. Proven track record of delivering $1.6M+ in cost savings through AI-based data scrubbing solutions. Expert in cloud-based data architectures (AWS, GCP), test data management, and real-time fraud detection systems. Transitioning focus toward AI/ML engineering with hands-on experience in probabilistic modeling, predictive analytics, and intelligent test data generation.

CORE COMPETENCIES

AI/ML & Data Science	**Cloud & Big Data**	**Development & Tools**
• Python/PySpark	• AWS (S3, Glue, Lambda)	• ETL/Data Pipelines
• Probabilistic Modeling	• GCP (BigQuery, Dataflow)	• C#/.NET/Scala
• Anomaly Detection	• Hadoop/Hive/Spark	• Power BI/Tableau
• Fraud Prediction	• Real-time Streaming	• CI/CD Automation

PROFESSIONAL EXPERIENCE

Associate Data Engineer-AI

Jorie AI | Oak Brook, IL (Feb 2026 – Present)

- Designing and developing interactive **Revenue Cycle Management** (RCM) dashboards to track KPIs including AR aging, denial rates, reimbursement trends, and collection performance.
- Establishing secure SQL Server connections via RDP environments to extract, transform, and model healthcare billing datasets for analytics workflows.
- Implementing automated data validation checks to ensure billing and claims data accuracy prior to dashboard publication.
- Collaborating with data science and operations teams to translate RCM business requirements into scalable reporting and visualization solutions.

Software Developer / Data Engineer - Arena Testing Team

American Express | Phoenix, Arizona (April 2021 – Feb 2026)

- **Saved $1.6M annually by developing AI-based data scrubbing solution for banking and credit card personas, eliminating dependency on external vendor contracts for test data generation**
- Enhanced fraud detection accuracy by 35% in test environments through implementation of probabilistic matching algorithms and anomaly detection models for real-time validation of 130K+ daily test transactions
- Developed and deployed machine learning-powered risk analysis dashboards in Power BI, enabling real-time fraud pattern identification and reducing false positives by 28%
- Built predictive analytics models for customer behavior analysis, improving risk assessment accuracy and strategic planning capabilities across fraud detection teams
- Architected and implemented enterprise-scale test data lake on AWS, consolidating data from Oracle, Teradata, mainframes, and RDBMS sources into HDFS, supporting 100+ QA engineers and test automation frameworks
- Optimized ETL pipeline performance by 60% through implementation of parallel processing and incremental loading strategies using PySpark and AWS Glue, reducing processing time from 8 hours to 3.2 hours
- Automated 15+ manual data workflows using SSIS and Python, eliminating 20+ hours of weekly manual effort and improving data accuracy from 94% to 99.7%
- Designed and deployed real-time streaming data pipelines using AWS Lambda and Kinesis for test environment fraud detection, enabling sub-second validation for 130K+ daily test credit card transactions

- Led multi-cloud migration initiative (AWS to GCP) for test data infrastructure, reducing infrastructure costs by 20% while improving query performance by 45% through BigQuery optimization
- Managed offshore development team of 6 engineers in India to ensure compliance with RBI data localization requirements for test data, successfully migrating 5TB+ of test datasets while maintaining zero downtime
- Implemented comprehensive CI/CD framework for data pipelines, reducing deployment time from 4 hours to 15 minutes and cutting production incidents by 67%

Software Engineer Intern / Data Analyst

CE-CERT, UC Riverside | Riverside, CA (June 2020 – January 2021)

- Developed automated data collection system for Eco-Vissim driving simulation project, processing sensor data from Volvo trucks to optimize fuel efficiency algorithms
- Built ETL pipelines using Tableau and Python for Eco-Drive technology research, analyzing 500GB+ of vehicle telemetry data

EDUCATION

Master of Science in Computer Engineering

University of California, Riverside | Graduated: December 2020

Bachelor of Science in Computer Engineering

University of California, Riverside | Graduated: August 2019

CERTIFICATIONS & TRAINING

- HTML, CSS and JavaScript for Web Developers (Coursera)
- AWS Certified Solutions Architect – Associate (In Progress)
- Claude with the Anthropic API
- Claude with Amazon Bedrock

AWARDS & RECOGNITION

- UCR Achievement Scholarship Recipient (2017-2019)