

John Zargaryan

Burbank, CA • 818-940-8908 • johnzargaryan@gmail.com • linkedin.com/in/john-zargaryan

EDUCATION

University of Southern California - Los Angeles, CA

- *MS in Applied Data Science, GPA: 4.0* *January 2026 - December 2027*
- Relevant Coursework: Data Management, Machine Learning, Big Data (Spark), Data Mining

California Polytechnic University - Pomona, CA

- *BS in Economics, GPA: 3.82* *September 2023 - May 2025*
- Relevant coursework: Calculus, Econometrics, Forecasting, Machine Learning

WORK EXPERIENCE

Amazon - Seattle, WA

Data Engineer Intern *May 2026 - August 2026*

- Developed an AWS anomaly detection framework to monitor allocated profit-and-loss output data across 42 business slices, enabling automated identification of outliers before downstream reporting.
- Gained hands-on experience with Amazon's internal engineering workflows, AWS tech stack, and production data systems while contributing to a financial data quality and alerting tool.
- Designed the framework to support future ML/agent-based anomaly detection beyond static thresholding based on rolling z-score analysis.

HumanGood - Duarte, CA

Data Engineer Intern *June 2025 - September 2025*

- Improved data pipeline processing time by 35% by optimizing SQL queries and ETL workflows, reducing reporting delays for 200+ internal users.
- Automated ingestion of 15+ data sources using AWS Glue, Lambda, and Amazon S3, reducing manual processing by 20 hours per week.

SKILLS

- Languages: Python, Java, Scala, SQL, R
- Data & AI: PyTorch, PySpark, Hadoop, Airflow, Tensorflow, RAG, LangChain, Vector Databases
- Cloud & Tools: AWS (S3, Glue, EMR, Redshift), Docker, Kubernetes, Git, Linux

CERTIFICATIONS

- AWS Certified Cloud Practitioner • AWS Data Engineer Associate

PROJECTS

SQL DATA WAREHOUSE

Accomplished the design and implementation of a multi-layer data warehouse architecture as measured by the successful load of bronze, silver and gold layers and creating fact/dimension models, and automating SQL load procedures.

- Improved data processing efficiency as measured by faster, error-free table refreshes, by developing stored procedures and data-quality checks that streamlined ingestion from raw sources to analytics-ready datasets.
- Enhanced analytics readiness as measured by clean, query-optimized schemas, by implementing a star schema design and documenting data lineage for BI and reporting use cases.