JAYA SRAVAN KALA

Sunnyvale, California | 7033895214 | jaysravan245@gmail.com | linkedin.com/in/sravank117

Summary

As an experienced computer vision data engineer, engineered scalable ETL pipelines using Python and PySpark to enhance processing efficiency at Sam's Club. Architected Delta Lake ingestion layers, reducing data drift by 20% and improving query speed by 30%. Aiming to leverage advanced data engineering skills to optimize data workflows and improve accuracy in analytics and reporting.

WORK EXPERIENCE

Sam's Club **Feb 2025 - Present**

Computer Vision Data Engineer *Remote*

- Engineered scalable ETL pipelines on the Element platform using Python, PySpark, and DVC to process multi-camera imagery and metadata for analytics and retraining workflows.
- Architected Delta Lake ingestion and transformation layers to unify raw, curated, and model-ready datasets, improving query speed by 30% and reducing data drift by 20%.
- Enforced schema validation, quality thresholds, and version control to ensure 100% reproducibility across all experiments.
- Integrated MLflow for metadata tracking and lineage, automating dataset, model, and metric registration for complete experiment traceability.
- Built monitoring and alerting systems for pipeline health and drift detection, cutting production issues by 30% via early anomaly detection and auto-recovery.
- Collaborated with data analysts, labeling, and MLOps teams to reconcile inconsistencies, strengthening pipeline reliability and governance by implementing robust data version control and experiment tracking processes
- Resolved critical data-quality issues in the computer vision pipelines by developing custom overlap-detection and polygon-matching logic, ensuring accurate segmentation outputs across multiple camera systems.

Ucode Technologies **Sep 2024 - Jan 2025**

Software Trainee *Sterling, Virginia*

- Designed an end-to-end Azure data pipeline for analytics and reporting using ADF, ADLS Gen2, HDInsight (Hive), and Azure SQL, which improved reporting accuracy and efficiency
- Configured ADF Data Flows with Pivot, Join, Lookup, Sort, Group, and Select transformations to validate and enrich datasets, enhancing data quality and validation processes
- Orchestrated complex ETL workflows with dependency management, tumbling triggers, and CI/CD pipelines in Azure DevOps, reducing deployment time by 40%.
- Automated multi-environment deployments (Test & Prod) using ARM templates and established real-time monitoring with Azure Monitor and Log Analytics (KQL).
- Partnered with business analysts to validate data outputs and maintain consistency across reporting layers, leading to improved accuracy in business reporting
- Conducted performance tuning on HDInsight/Hive queries and optimized ADF sink/lookup operations, lowering pipeline runtimes by 25% during test cycles.

Cognizant **Feb 2021 - Jul 2022**

Program Analyst Trainee *Hyderabad, India*

- Created and optimized Snowflake and MySQL ETL workflows using SQL transformations to deliver dashboards and monthly reports.
- Streamlined data cleansing and preprocessing with Excel (formulas, pivot tables, macros), improving reporting turnaround time by 20%.
- Assisted visualization teams by preparing production-ready datasets for Tableau dashboards and executive reporting, ensuring timely and accurate data visualization
- Established job dependencies and alerting for monthly and ad-hoc batch jobs, ensuring SLA compliance and quick failure recovery.
- Improved SQL query performance and refined data extraction logic for recurring reporting workflows, decreasing monthly refresh time by 15%.
- Investigated data discrepancies across Snowflake and MySQL sources, performing root-cause analysis and coordinating fixes that improved data consistency for executive dashboards.
- Resolved broken Tableau refresh jobs by tracing failures back to missing Snowflake partitions and incorrect Excel preprocessing steps, ensuring all client dashboards updated without delays.

TECHNICAL SKILLS

- **Programming**: Python (Pandas, NumPy), SQL, Scala
- **Data Engineering & Orchestration**: PySpark, Airflow, Kafka, Delta Lake, Azure Data Factory (ADF), Azure Synapse, Databricks, HDInsight, Event Hub
- **MLOps & Automation**: MLflow, DVC, Docker, Kubernetes, Git, YAML Pipelines, CI/CD (Azure DevOps), Experiment Tracking, Model Registry
- **Cloud Platforms**: Azure (ADLS Gen2, Synapse, ADF, Monitor), AWS (S3, Glue, Lambda, Redshift)
- **Databases & Warehousing**: Snowflake, PostgreSQL, MySQL, Azure SQL DB, MongoDB, DynamoDB
- **Computer Vision & Deep Learning**: YOLOv11, Object Detection, Segmentation, PyTorch, Transformers, CLIP/VILT, IoU/Confidence Tuning, Custom Post-processing (NMS, mask IoU)
- **Tools**: Jupyter, Linux, Postman, GitHub, VS Code

PROJECTS

Formula 1 Data Pipeline

- Built batch and incremental ingestion pipelines with ADF, Delta Lake, and Databricks to enable time-series analytics and historical trend exploration.
- Structured Delta Tables with schema enforcement, ACID compliance, and time travel for reproducible data management.
- Implemented incremental merges and refined Delta Lake partitioning strategies, improving query latency by 25%.
- Designed modular PySpark workflows to ingest, transform, and model telemetry data for race analytics and forecasting.
- Applied SparkSQL and dimensionality reduction (PCA/t-SNE) to analyze correlations in race and pit stop data for visualization.

Fake News Detection (Multimodal CLIP + ViLT Pipeline)

- Built a multimodal detection system using CLIP and ViLT to evaluate alignment between article text and thumbnail images for misinformation screening.
- Engineered a cross-modal similarity module using cosine distance and projection heads to quantify visual-textual coherence with high recall on mismatched pairs.
- Added drift-monitoring hooks and lightweight retraining triggers to adapt the model when content distributions shifted over time.
- Delivered an interpretability dashboard with attention heatmaps (Grad-CAM/VILT attention) to help users inspect which regions influenced mismatch decisions.

CERTIFICATIONS

- **Microsoft Certified: Azure Data Engineer Associate (DP-203): Aug 2024**
- **AWS Certified Cloud Practitioner: Jan 2024**

EDUCATION

George Mason University, Fairfax, VA **Aug 2022 - May 2024**
M.S., Computer Science
- **GPA:** 3.83/4.0

Institute of Aeronautical Engineering, Hyderabad, India **Aug 2017 - May 2021**
B.Tech, Information Technology
- **GPA:** 3.78/4.0