SURYA DUTTA

📞 (708) 712-9625 · ✉ surya@suryadutta.com · in linkedin.com/in/suryadutta · ⌨ github.com/suryadutta

TECHNICAL SKILLS

Data Platform: Airflow, Kafka/Kafka Streams/Kafka Connect, Spark/Spark Streaming, Hive, Iceberg, Trino, Redshift

Programming: SQL, Python/PySpark/PyData stack, shell scripting, Kotlin, Go, C#, JavaScript/React

Infrastructure: AWS, Kubernetes/Helm, Terraform, Docker, Concourse CI, Git/GitOps, Prometheus/Grafana

EXPERIENCE

Senior Data Engineer II
SpotHero, Data Platform Team

Chicago, IL
Mar 2022 – Present

- Enable analysts, engineers, and data scientists to create, deploy, and monitor hundreds of data pipelines integrating internal and external systems. Leverage Apache Airflow for scheduling batch workflows, streaming data (Kafka), a data lake (Trino/Hive/Iceberg), a data warehouse (Redshift), and various custom applications
- Designed and implemented end-to-end streaming ML service to probabilistically exclude search results when parking facilities are expected to experience capacity or other supply-side issues
- Maintain an in-house service that provides a low-code interface to generate full and incremental SQL materializations in SpotHero's data lake and data warehouse, complete with dependency management, monitoring, and alerting
- Evaluate and document different architectures for applications & services, provision resources and deploy containerized applications in AWS using Terraform and Helm, and fix variety of issues during on-call rotations
- Collaborate with stakeholders and the Product team to determine best practices for creation and deployment of data models and data flows. Work with stakeholders to define service level agreements, and enforce through data contracts, monitoring, alerting, and dependency management
- Promoted from Senior Data Engineer I to Senior Data Engineer II in Mar 2023

Senior Data Engineer
McMaster-Carr Supply Company, Digital Analytics

Elmhurst, IL
Aug 2018 – Mar 2022

- Architected and productionized data warehouse combining 5 years of clickstream data & transactions with internal customer and product data. Used to report aggregate North Star metrics to company and product leaders
- Leveraged data build tool (DBT) to prototype & document tables/views that made complex datasets accessible
- Led remote team of 7 junior engineers in building & orchestrating robust pipeline jobs
- Promoted from Data Engineer to Senior Data Engineer in Feb 2021

TALKS & PROJECTS

SQL Materializations in Trino & Redshift using Airflow | *Chicago Apache Airflow Meetup, Apr 2023* | Slides
- Delved into the architecture behind the engine that powers incremental and full SQL materializations at SpotHero

Fairness and Bias in Machine Learning | *McMaster-Carr Developer Talk, Aug 2021* | Slides
- Described how algorithmic bias in ML is introduces & amplified; explored how to measure & mitigate these biases

Cognition Tracker | *Capstone Project, Masters in Data Science, Apr 2021* | Website
- Developed mobile application that measures cognition and progression of dementia through spontaneous speech
- Achieved best-in-class accuracy with audio ensemble ML model (LSTM + speech-to-text BERT embeddings)
- Scaled backend architecture in AWS using Docker and Tensorflow Serving to return instantaneous score to app

EDUCATION

University of California, Berkeley
Master of Information and Data Science (MIDS) | GPA 3.88/4.00

Berkeley, CA
Jan 2019 – Apr 2021

Yale University
Bachelor of Science, Physics (Intensive) | GPA 3.74/4.00

New Haven, CT
Aug 2014 – May 2018