

Gregory Oliva

Software Engineer

gloliva@proton.me

(916) 468-2263

San Diego, CA

<https://www.linkedin.com/in/gloliva>

EXPERIENCE

Embark Trucks — *Software Engineer, Data Infrastructure*

January 2023 - March 2023

- Improved data offloading system to reprocess data from existing Truck runs to greatly reduce manual triage requests for the team
- Led initiative to catalog ROS bag file type data size and access patterns for historical data stored in S3 with goal to move into tiered cold-storage
- Expanded Airflow utilities module to support greater flexibility of Airflow functionality in code and built out Airflow utils testing framework

Affirm, Inc. — *Software Engineer, Data Infrastructure*

August 2019 - October 2022

- Lead for Storage and Replication data infrastructure team; planned and designed team's projects for the financial year, onboarded 4 engineers and one intern to the team, and provide technical advice on project specs from internal stakeholders
- Contributed to building scalable and self-service database snapshots (100+ jobs), MySQL and DynamoDB CDC replication pipelines (1000+ jobs), and Data Warehousing infrastructure
- Developed framework for replication and warehouse infrastructure and launched 2+ additional countries
- Led team to build our observability framework, scale systems for peak periods (Black Friday/Cyber Monday), migrate off of legacy systems, reduce high priority oncall alerts from 80+ to less than 20 per week, and design thorough documentation
- Contributed to migration from Redshift to Snowflake as our primary datawarehouse and led migration from AWS EMR to AWS Glue for DynamoDB replication ETLs
- Built out deploy infrastructure for DBT and ETL Airflow jobs

PROJECTS

Data Warehouse V2 — *Migration to Snowflake*

- Built internal Snowflake library for loading data to/from S3 and Snowflake
- Developed Terraform module for managing Snowflake network policies and securable objects
- Built automated user onboarding jobs and a self-service permission management framework
- Contributed to setting up networking (AWS PrivateLink, VPC peering, SSO) to connect multiple AWS accounts to our Snowflake accounts

Hardened Observability — *CloudWatch Monitoring for Data Infra*

- Unified metrics, alarms, and dashboards from multiple sources (Kibana, Cabot, Grafana) into CloudWatch
- Developed a framework for emitting runtime and task metrics to CloudWatch from 1000+ replication pipelines
- Implemented Terraform modules to bring up alarms and monitoring dashboards alongside the rest of the stack when bringing up new environments

Managed Replication — *Snapshots and Replication on MWAA*

- Designed and optimized 400+ DAGs for running daily database snapshots, full table re-loads, and CDC incremental replication on AWS Managed Airflow

SKILLS

Python

Terraform

Snowflake

Apache Airflow

Unix, Git, Docker

AWS (DMS, S3, IAM,

CloudWatch, MWAA,

Redshift, DynamoDB, Glue,

Lambda, Aurora, ASG, VPC)

SQL (MySQL, PostgreSQL)

Max/MSP, Gen, Jitter

Education

Bachelor of Science

Electrical Engineering and Computer Science

University of California, Berkeley

August 2015 - May 2019

Berkeley, CA