# Daniel Yanos

## Experience

### Root Insurance Company — Columbus, OH

*Auto insurance startup with IPO in October 2020. Focused on using data and technology to accurately price insurance.*

**Senior Data Engineer** — May 2021 – Nov 2022

- Increased speed to market by developing a platform for insurance actuaries to measure revenue impact when changing pricing algorithms.
  - Utilized AWS Step Functions to create user-friendly workflows that performed multiple data aggregation/transformation steps to produce datasets for actuarial analysis.
  - Overhauled existing AWS Batch and PySpark data pipelines for reliable, timely delivery of data. Boosted application performance by converting feature store data from JSON to partitioned Apache Parquet.
  - Reduced operating costs by over $100k per month via AWS optimizations. Detected/removed unused infrastructure, utilized Datadog to appropriately size instances, shut down resources after hours.
  - Developed tooling to monitor data quality/completeness. Leveraged AWS Glue crawlers/data catalog to enable discovery and querying of S3 data.
- Provided insurance underwriters ability to filter high-risk customers through design and development of automated, fully configurable, rule-based system.
  - Designed reusable state-machine components to automatically mitigate detected risks. Components could be configured per market to ensure compliance. Automated risk mitigation removed potential for human error, ensured ability to process high volume of customers.
  - Developed integration with TransUnion APIs to detect potential fraud early in the customer lifecycle.

### Capital One — Richmond, VA

*Top ten US bank leveraging technology and data analytics to disrupt the financial services industry.*

**Principal Data Engineer** — July 2018 – April 2021

- Simplified data producer experience by developing data ingestion APIs and SDKs for the next generation data processing platform. Eased migration from legacy systems by developing a flexible, configuration-driven AWS Lambda template for data producers.
- Drastically reduced time to market for a team of 25+ developers through implementation of an ECS-based container platform and overhaul of CICD practices.
  - Designed and deployed a self-healing, autoscaling ECS platform. Utilized Terraform to facilitate rapid deployments and provide a consistent user experience across multiple AWS accounts and regions. Implemented key FluentD and Datadog services as containerized applications to ensure high availability.
  - Developed a Jenkins library that enforced versioning and code coverage standards, ran security scans, archived code artifacts, and deployed to the appropriate environment when a GitHub pull request was approved.
  - Eliminated manual intervention and reduced infrastructure costs by refactoring existing applications to automatically scale the number of running containers based on request rate.

**Senior Data Engineer** — July 2016 – June 2018

- Unified enterprise streaming platform and AWS data lake by automating ingestion of 750+ real-time data feeds.
  - Implemented microservice-based solution that automatically detected new data feeds, registered data feeds in the enterprise metadata catalog, and launched tasks to micro-batch events and persist data in S3.
  - Utilized AWS to deploy a scalable, active-active application capable of handling data volumes that exceeded 800 GB per hour. Archived 1 PB of streaming data in under 3 months.
  - Enabled real time alerting via AWS Lambda functions and PagerDuty routing. Utilized ELK stack for log aggregation and analysis. Developed command line interface to diagnose and remediate job failures.
- Led open source launch of ETL tool built on Apache Spark. Released at OSCON 2017 under the name “Hydrograph”.
  - Managed core development team responsible for releasing and maintaining Hydrograph. Created a project website, developed user guides, and fixed reported issues. Grew internal adoption of Hydrograph to 100+ users by leading hands on training and consulting with earlier adopters.

**Data Engineer** August 2015 – June 2016

- Generated Capital One customer profile from 20 disparate data sources using data pipeline powered by Apache Hive to obtain additional insight from existing data.

**Associate Software Engineer** July 2014 – July 2015

- Unlocked 20 years of historical data by creating enterprise frameworks in Java MapReduce and Apache Pig to refine and load data to the on-premise Hadoop lake.

## Education

### University of Michigan **Ann Arbor, MI**

**College of Engineering**

Bachelor of Science in Computer Science Engineering - May, 2014

## Technical Skills

AWS Certified Solutions Architect – Associate (September 2017)

Languages: Python, Java, Scala, Ruby & Ruby on Rails, SQL; some experience with JavaScript, HTML, CSS

Frameworks/Tools: Terraform, Docker, Buildkite, Jenkins, Datadog, Kafka, Apache Spark, Apache Hive, Java MapReduce