Daniel Yanos

danyanos.github.io

Experience

Capital One Financial — Richmond, VA

Principal Data Engineer July 2018 – Present

- Simplified 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.
- Ensured consistent data delivery and increased customer confidence by stabilizing the batch data platform.
    - Captured detailed application metrics and enabled real-time alerting, consolidated deployment pipelines, fixed security vulnerabilities, identified and removed unused code, and implemented strategies to save AWS costs.
- 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 catalogue, 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.
    - Leveraged suite of tools to automate platform monitoring and troubleshooting. 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 and designed for the cloud. Released at OSCON 2017 under the name “Hydrograph”.
    - Managed core development team responsible for releasing and maintaining Hydrograph. Led effort to create a project website, develop user guides, and fix reported issues. Streamlined security analysis, build, and deployment process through implementation of CICD pipelines.
    - Grew internal adoption of Hydrograph to 100+ users across the enterprise. Fostered growth of developer community by leading hands on training, consulting with early adopters.

Senior 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. Gained hands-on experience developing data-driven solutions in a highly regulated environment.
- Led weeklong software development courses for classes of 30+ college hires. Created training materials and hands-on exercises to demonstrate techniques for writing clean, maintainable code.

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 Hadoop lake.
- Remediated space issues on the Hadoop lake by compacting 1.2 million small files to 25 thousand files. First production application at Capital One to use Oozie.

General Electric **Van Buren, MI**

Software Engineer Intern Summer 2013

- Automated Teradata billing process; replaced manual spreadsheets with code that generates dynamic reports. Improved data quality, user efficiency by working with a global user-base to deactivate unused database schemas.

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: Proficient with Java, Scala, Python; experience with SQL, JavaScript, HTML, CSS

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