ANISH YAKKALA

(408) 872-2279 ayakkala@gmail.com
https://github.com/ayakkala1 https://ayakkala1.github.io/ayakkala.github.io/

PROFESSIONAL SUMMARY

Software Engineer with 4+ years of experience building and operating large-scale data platforms at Apple. Specializes in distributed systems, cloud-native architectures, and petabyte-scale stream processing. Designed and deployed globally distributed services vending 28B+ taskings per month across 1B+ devices, and led Flink-based ETL platforms processing over 3 PB/day. Strong background in data lake architectures, real-time analytics, and cross-functional technical leadership.

WORK EXPERIENCE

Apple Inc. Aug 2021 - Present
Software Engineer - Core Reporting *Cupertino, CA*

- **Device Configuration at Global Scale:** Designed and own a globally distributed configuration service handling 28B+ taskings per month, securely vending telemetry collection policies to over 1B Apple devices. Built using reactive Spring WebFlux with R2DBC and a dual-cache strategy to achieve low-latency, high-availability delivery at global scale.
- **Petabyte-Scale Stream Processing:** Built and run large-scale Apache Flink ETL pipelines processing 3+ PB/day of device telemetry, powering real-time analytics and diagnostic insights for Apples Software Engineering organization.
- **Cloud-Native Data Platform Leadership:** Led Apple Cloud Object Storage (ACOS) adoption for the Software Engineering organization, coordinating across Spark, Flink, Unified Auth, and storage backend teams. Removed critical production blockers and established ACOS as the scalable, compliant backend for petabytes of diagnostic data.
- **Data Lake Modernization:** Architected Iceberg-based data lakes on ACOS/S3, converting compressed JSON archives into queryable Parquet tables. Designed a Splunk sidecar pattern for Flink to enable real-time log streaming and significantly improving pipeline observability.
- **Large-Scale Migrations:** Led large-scale platform migrations from legacy Python 2 MapReduce and Samza jobs to Java 17 Flink and Python 3 PySpark. Drove the data lake transition from HDFS to cloud object storage (S3/Ceph), migrating petabytes of diagnostic logs with zero downtime. Coordinated cross-team efforts to migrate downstream consumers and refactor ETL pipelines for a fully cloud-native architecture.

Boeing Data Science & Analytics Summer 2019
Data Science & Analytics Intern

- Created web applications in R providing live graphics and metrics to VP executives for solving Supply Chain initiatives. Tools adopted for production use by executive leadership.
- Contributed to Machine Learning project using TensorFlow and H2O.ai to optimize Procurement Agent item assignments.

EDUCATION

California Polytechnic State University SLO *Sept 2017 - June 2021*

Bachelor of Science in **Statistics**
Minors in **Data Science** & **Mathematics**

Awards: Helen V. Sandercock Scholarship (Scholastic Excellence), Joyce Curry-Daly and James Daly Scholarship, Herbert E. Collins Scholarship (Promise in Industry). President's List 2017-2020.

TECHNICAL SKILLS

Languages: Java, Python, Scala

Frameworks & Technologies: Apache Flink, Apache Spark, Spring (WebFlux, R2DBC), PySpark, Kafka, Hadoop

Data & Storage: Apache Iceberg, Parquet, S3, Ceph, PostgreSQL, MySQL

Tools & Platforms: Splunk, Prometheus, Docker, Kubernetes