

Jazem Siddiqui

(510) 697-8805 | jazemf.s@gmail.com | San Francisco, CA | [linkedin.com/in/jazems](https://www.linkedin.com/in/jazems)

EXPERIENCE

Capital One

San Francisco, CA

Software Engineer

Nov 2025 – Current

- Overhauled legacy vector search infrastructure with an optimized **OpenSearch** KNN pipeline for dataset recommendations, improving hit rate by ~50% and reducing runtime by 63%
- Designed a property graph model for users, assets, and interactions; built 12 **PySpark** ETL processors on EMR ingesting **50M+ records** into NeptuneDB
- Architected an ML training dataset schema and pipeline on **Databricks** and **Snowflake** with retry logic, multi-asset extensibility, and end-to-end observability
- Accelerated developer efficiency **by 90%** by building **custom agentic AI skills** to autonomously handle infrastructure provisioning, repository synchronization, and vulnerability patching
- Built **React** and Lit frontend components for the Data Discovery platform homepage, translating mockups into reactive UI for dataset recommendations, FTUX flows, data management tools, and notifications, with state management, a client-side data layer, API integration, and event emitting/handling

Associate Software Engineer

Feb 2024 – Nov 2025

- Built an algorithms-driven dataset **recommendations pipeline 0→1** for an internal data discovery platform, optimizing user-dataset ranking via extensive offline/online trials and feature weight tuning
- Developed an event-driven **Go-based Kafka** consumer and **DynamoDB** upsert handler achieving **15K+ TPS**, enabling deprecation of a third-party payment vendor and slashing infrastructure costs
- Led org-wide migration of secrets management from HashiCorp Vault to **AWS Secrets Manager**, completing cutovers for high-traffic ECS and Lambda services

Software Engineer Intern

Jun 2023 – Aug 2023

- Reduced cross-team coordination by ~1 week per sprint by designing and iteratively decoupling RBAC permission logic from upstream services within a new workflow management platform
- Led onboarding of the platform's first clients by transforming custom workflows into configurable setups

Anaplan

San Francisco, CA

Software Engineer Intern

Jun 2022 – Aug 2022

- Integrated an in-memory cache to throttle user traffic across API instances and optimize data retrieval times by **80%** for downstream APIs with **Redis** and **Java Spring Boot**
- Designed exhaustive tests for a rate limit filter by utilizing asynchronous programming, multi-threading, and locks

TECHNICAL SKILLS

Languages: Python, TypeScript/JavaScript, Golang, Java, SQL

Technologies: AWS (ECS, Lambda, S3, DynamoDB, EMR, RDS, Neptune, SQS, CloudFormation), Spark, Kafka, Snowflake, Databricks, OpenSearch, Redis, Docker, NumPy, Pandas, React

EDUCATION

University of California, Berkeley

Year: 2023

B.A. Computer Science

Major GPA: 3.9

- Relevant Coursework: Operating Systems, Database Systems, Computer Security, Algorithms, Machine Structures