

# ADITYA KULKARNI

Sunnyvale, CA · adi2188@gmail.com · 5189515772 · www.adityakulkarni.info

## EDUCATION

---

**Carnegie Mellon University** Pittsburgh, PA  
MS Information Networking (Computer Science + Elec & Comp Engineering) Aug 2010 - May 2012

**University of Pune - Vishwakarma Institute of Technology** Pune, India  
BEEngg Electronics and Telecommunications Engineering Aug 2005 - May 2009

## WORK EXPERIENCE

---

**Apple, Inc.** Cupertino, CA  
*Senior Software Engineer - Data @ AI/ML org* Nov 2020 | Present

- Building config driven framework to generate unified Spark and Flink applications to process batch and realtime data from 1B+ Apple devices for various use cases like Machine hearing, Natural Language Processing, Speech Recognition, User Experience Measurement, etc. This enables creating 100s of individual batch and streaming data processing applications with a small team of engineers.
- Design, build, and operate highly performant and scalable batch and stream data processing infrastructure and solutions to support day to day ML operations including training, serving, evaluation and experimental systems
- Providing Data Warehousing solutions using Spark/Flink/Presto at Petabyte scale with ownership of all Machine Hearing, Device specific Siri functionalities, Machine Translation Derived Datasets, powering all Machine Hearing, Device specific Siri functionalities, Machine Translation model building and analytics.
- Building tooling for Deployment, Orchestration, Stability, Monitoring and Observability for these Spark and Flink Data Pipelines.
- Skills: Java, Python, Scala, Spark, Flink, Presto, Airflow

**Rubrik, Inc.** Palo Alto, CA  
*Member of Technical Staff* May 2019 | Nov 2020

- Working on massive data processing pipelines in Spark/MapReduce for telemetry callhome data from Rubrik deployments. Launched a new, Generic Ingest Data Pipeline to ingest any json data to Snowflake. This serves the company wide use case to enable any sort of analytics, data visualization for any data generated by any teams and providing value added services like Capacity recommendations.
- Working on building telemetry infrastructure handling billions of time-series data points and log messages. This data powers proactive alerting - we detect a large fraction of issues before our customers notice them.
- Providing Data Warehousing solutions using Snowflake at Petabyte scale with ownership of all Product Metrics Datasets, powering all Product Management, Support, Sales usecases.
- Created tooling for Deployment, Orchestration, Stability, Monitoring and Observability (logging and job metrics) for these Spark Data Pipelines.
- Led the AWS Cost Optimization efforts to optimize the utilization of EMR and EC2 resources and cut team AWS spend by 40 percent (\$X MM).
- Led the cloud services availability monitoring and Observability project. Driving the SLA definition and enforcement process for all cloud services.
- Skills: Python, Scala, Spark, AWS (EMR, EC2/ECS), Snowflake, Elasticsearch, Airflow, Apache Druid

**LinkedIn** Sunnyvale, CA  
*Senior Software Engineer* Dec 2017 | May 2019

- Worked on massive data flow architectures for LinkedIn Marketing Solutions (LinkedIn Ads)
- Mentored intern project to develop tooling to automate tasks around creation of dataflow archetypes required for data pipelines in the LinkedIn Stack
- Creating scalable Data Warehousing solutions based on Spark and Presto.

- Working on Autonomous Data Quality/Consistency Frameworks
- Part of Proposal Review Committee for Microsoft MLADS(Machine Learning and Data Science) Conference - Fall 18
- Skills: Java, Hadoop, Hive, Pig, Spark, Azkaban, Gobblin, Apache Calcite, Presto

### **Drawbridge (acquired by LinkedIn)**

*Engineering Manager (prev. Senior Software Engineer)*

San Mateo, CA

Aug 2016 | Dec 2017

- Led and Managed the backend reporting and analytics team of Drawbridge's Ad Platform products. Also let the ads targeting team for brief time as a joint responsibility.
- Both hands on development of the data reporting platform to develop near real time capabilities and also managing a team to achieve state of the art reporting capabilities.
- Worked on reporting pipelines/Data pipelines that process more than 100 Billion events per day. Product facing enhancements for the data pipeline.
- Drove and engineered a Presto based Data Warehouse, that enabled Business Analysts, BI engineers and Data Scientists to query the datasets on HDFS directly as against a serving layer columnar relational DB thereby increasing the amount people can query and use in their models.
- Key Project Contributions : 1. Conversion Tracking/Attribution <https://drawbridge.com/blog/p/the-drawbridge-approach-to-attribution> 2. Cross Device Insights with Real time Attribution Metrics <https://drawbridge.com/c/insights&attribution>
- Skills: Java, Python, Hadoop, Hive, Pig, Spark, Hbase, Couchbase, Oozie, Columnar DBs, Presto etc.

### **WalmartLabs**

*Software Engineer III (prev. Software Engineer II)*

Sunnyvale, CA

Dec 2013 | Aug 2016

- @Labs - WMX (Walmart Exchange) - Data Mining and Analytics for Online Ads - Ad Impression Ingestion and Processing : As an early stage member on the team implemented and fully owned (dev, test, and product) the Ingestion pipeline processing upto 1B impressions per day from various sources.
- Designed and implemented petabyte scale analytic pipeline
- Worked in a small team to build a lambda-architecture realtime analytics and OLAP platform from the ground up
- Pioneered use of Apache Spark on the team for improved pipeline performance and maintainability
- Built a querying system using Apache Solr to build analytics item sets using Clustering of Short Strings in Large Datasets
- Skills: Java, Python, Hadoop, Hive, Spark, Storm, Airflow/Luigi

### **Yahoo! Inc.**

*Software Engineer*

Sunnyvale, CA

July 2012 | Dec 2013

- Development related to Content Enrichment for all ingested content in the Content Agility pipeline, and served up on all of Yahoo Media properties (Yahoo News, Yahoo Sports, Yahoo Finance, etc.)
- End to end Design and Development related to selection of most relevant Canonical URLs and contextual clickthrough URLs for content. These power all Yahoo hosted content urls on all content streams including the Yahoo homepage.
- Sole developer for Canonical URL generation for all content in the Enrichment Workflow to achieve Search Engine Optimization for each content (story, photo, video)
- Independently Developed Content Enrichment Libraries for Categorization, Content Quality marking, Geo Classification, Named Entity Recognition, etc. on content (stories, photos, videos) through Feed Ingestion and Editorial Ingestion Workflows using Yahoo's Contextual Analysis Platform.
- Development within Category Taxonomy Management Workflow for Yahoo Content Taxonomy.

- Setting up of Continuous Integration Environment for Content Track's Libraries.

**Carnegie Mellon University**  
*Graduate Teaching Assistant*

Pittsburgh, PA  
 Aug 2011 | Dec 2011

- Graduate Teaching Assistant for Fundamentals of Embedded Systems

**Avaya Inc.**  
*Associate Software Engineer*

Pune, India  
 Jul 2009 | Jul 2010

- Worked on the Expert Systems and HealthCheck products. Expert System was a rule based engine (rule engine) in Java with the Rete II rule engine. Used OPSJ language for building the rules.

## SKILLS

---

Languages:	Java, Scala, Python, SQL, C
Frameworks:	Spark, Flink, Hadoop, Hive, Pig
Others:	Clean Code Practices, Code Reviews, Software Engineering, CI/CD, Observability
Fundamentals:	Distributed Systems, Machine Learning, Data Mining

## AWARDS

---

<b>Narotam Sekhsaria Fellowship</b>	Narotam Sekhsaria Foundation
Awarded to 15 individuals from across India every year	Jun 2010
<b>Dhirubhai Ambani Scholarship</b>	Reliance Foundation
Ranked 2nd amongst 150,000 students	May 2005
<b>National Talent Search Scholarship</b>	Government of India
Awarded to top 750 students (from 1M applicants) in India for Academic Merit	May 2003