

Achal Shah

Seattle – WA – 98101

✉ m@achals.com • linkedin.com/in/achalnshah

Staff Engineer with 10+ years building high-scale realtime ML Platforms at Reddit, Tecton, and Uber

Experience

Reddit, Inc.

Staff Engineer, ML Platform

Seattle, WA

April 2025–Current

- Launched the next generation of Reddit's Prediction Logging framework, handling upwards of 10 Million predictions per second in aggregate with lower unit costs than previous iterations.
- Leading the development of our next-gen Feature Engine, allowing users to express features declaratively without worrying about managing infrastructure.

Tecton

Tech Lead Manager, Realtime Compute

Seattle, WA

April 2020–April 2025

- Leading a rearchitecture of Tecton's realtime compute stack. Drove down latency by at least an order of magnitude across the board, and lowered infra costs by 20x. This architecture forms the basis for all realtime services touching user data.
- Led development of the Stream Ingest API to support high scale data ingestion into Tecton. In use by multiple customers at 1000s of QPS.
- Also worked on various platform-level challenges. I ported all of Tecton from Databricks to EMR as my starter project. I also evolved the Tecton declarative framework through two major versions, and worked on realtime feature serving.

Uber

Sr. Software Engineer

San Francisco + Seattle

Sept 2015–April 2020

- Michelangelo: Scaled Uber's Machine Learning platform. Worked on sharding the model serving layer. Supported growth in number of models from the hundreds to tens of thousands.
- On Michelangelo's Deep Learning subteam, I added deep learning model serving capabilities. This included integration with neuropod, building an automatic request batcher to increase throughput, and experimenting with fractional GPU usage for better cost savings. Worked with internal customers to support the first use case in production.
- Reliability Platform: Built and operated Hailstorm, an high-throughput, distributed, load generation platform. Helped Uber make peak days "officially boring".

Amazon.com

Software Development Engineer

Seattle, WA

July 2013–Sept 2015

- Amazon Fulfillment Technologies - Outbound Flow: I owned the labor forecasting and demand forecasting services within AFT. These forecasts were used to make automated labor planning & staffing decisions at Amazon Fulfillment Centers, to ensure customer promise.

Education

University of Pennsylvania

Philadelphia

2011–2013

Master of Science and Engineering, Computer and Information Science

Relevant Coursework: Internet and Web Systems, Database and Information Systems, Mathematical Statistics, Machine Learning

Institute of Technology, Nirma University

Ahmedabad

2007–2011

Bachelor of Technology, Computer Engineering

Skills

Fluent in: Java/Kotlin, Python, Go

Familiar with: C/C++, Rust, Javascript, Scala

Execution Frameworks & Technologies: Flink, Spark, Ray, Dask, Kafka, Kubernetes, gRPC, Spring, Temporal

Databases and Data stores: Postgres, MySQL, DynamoDB, Redis, Cassandra

Open Source

Feast

Maintainer

2022–Current

- Served as the Tech Lead for the Feast project up until 2023.
- Created new abstractions in Feast to support pluggable compute engine abstractions and led the development of the high-performance feature server.
- Led our feature registry modernization effort by migrating it to a database instead of flat files.
- Modernized feast's testing infrastructure using test containers to improve reliability and consistency across components.