

Software Engineer - Hadoop Platform

Location: Palo Alto / San Francisco

How to apply: Email your resume to hadoop-platform-jobs@uber.com

About The Role

As a Hadoop Platform developer, you will build libraries and distributed services to support Uber's analytical data infrastructure. This includes libraries at the very core of Hadoop (e.g. file format for optimal real-time storage of analytical data with support of update/deletes) as well as platformy products on top of Hadoop (e.g. Schema Management service, Ingestion/Dispersal libraries,...). Regardless of whether you are a core Hadoop infrastructure developer or a platform developer, you'll find a challenging project within Hadoop Platform team to continue your career.

What You'll Do

- You will expand on a solid data infrastructure, that collects, stores and serves 100+ petabytes of information, services & frameworks that makes data more accessible to Uber's developers, services & teams.
- For example, you will be building out a storage file format for Hadoop that supports semi-real-time updates/deletes for analytical columnar data (See [Hudi](#) or [blog](#)).
- You will be tasked with implementing SDKs for accessing the vast datasets we house and managing periodic computations, to keep our products and services up to date.
- You will be creating several gateway services, that offer low latency access to information in our Hadoop data lake.
- You will be working on creating a sophisticated data access platform, that can help our developers quickly build value on top of our troves of geo data.
- You will utilize the latest breakthroughs in data processing technologies, to extract enormous value out of Uber's data.

What You'll Need

- Deep engineering background, ideally experienced with large scale data/distributed systems, housing several TBs of data
- Dedication to moving fast in the short-term, while simultaneously building for long term
- Deep & Hands on experience building MapReduce/Spark or similar such big data pipelines/frameworks/services (Hadoop, Hive, Kafka, Presto, Parquet, Avro,...)
- Can quickly learn and be productive at new data technologies
- Can produce quality code fast, preferably Java on *nix
- Excellent foundation in distributed systems concepts

Bonus Points If

- Prior experience developing foundational services at scale (eg: ML Pipelines, Relevance systems, Search services)

About the Team

Hadoop Platform team is responsible for building out the required reliable/scalable data platform that serves petabytes of data. We heavily leverage Hadoop/Hive/Kafka/Spark/... to build distributed systems or libraries required to serve data to the rest of the company.

Fundamentally, you will be working with a team of stellar colleagues, and contribute to the next generation of intelligent data services that will have a profound impact on how we march towards the vision of of Uber's data driven, real-time marketplace. Working on data at Uber now, is an exciting opportunity as we need to balance the company's needs to move fast today while continuing to create platforms that will be leveraged for the products of tomorrow.

You can learn more about our projects by checking our recent open-sourced projects: [Hudi](#), read our engineering blog post about it [here](#) , or watch the presentation about Uber's Big Data Platform journey over the past few years [here](#) .