

Sourabh Bajaj

Software Engineer

THÔNG TIN

Email: mail@website.com

Phone: +1-123-456-7890

KỸ NĂNG

- TensorFlow
- Apache Beam
- Scala
- Python
- Javascript
- C++
- SQL
- Java
- AWS
- Play
- React
- Kafka
- GCE
- Hadoop
- Cassandra
- Docker
- Amazon Redshift
- D3

KINH NGHIỆM

Software Engineer

Google | Oct 2016 - Present

TensorFlow: TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models. **Apache Beam:** Apache Beam is a unified model for defining both batch and streaming data-parallel processing pipelines, as well as a set of language-specific SDKs for constructing pipelines and runners.

Senior Software Engineer

Coursera | Jan 2014 - Oct 2016

Notifications: Service for sending email, push and in-app notifications. Involved in features such as delivery time optimization, tracking, queuing and A/B testing. Built an internal app to run batch campaigns for marketing etc. **Nostos:** Bulk data processing and injection service from Hadoop to Cassandra and provides a thin REST layer on top for serving offline computed data online. **Workflows:** Dataduct an open source workflow framework to create and manage data pipelines leveraging reusable patterns to expedite developer productivity. **Data Collection:** Designed the internal survey and crowd sourcing platform which allowed for creating various tasks for crowd sourcing or embedding surveys across the Coursera platform. **Dev Environment:** Analytics environment based on docker and AWS, standardized the python and R dependencies. Wrote the core libraries that are shared by all data scientists. **Data Warehousing:** Setup, schema design and management of Amazon Redshift. Built an internal app for access to the data using a web interface. **Dataduct integration** for daily ETL injection into Redshift. **Recommendations:** Core service for all recommendation systems at Coursera, currently

used on the homepage and throughout the content discovery process. Worked on both offline training and online serving. Content Discovery: Improved content discovery by building a new onboarding experience on coursera. Using this to personalize the search and browse experience. Also worked on ranking and indexing improvements. Course Dashboards: Instructor dashboards and learner surveying tools, which helped instructors run their class better by providing data on Assignments and Learner Activity.

Data Scientist

Lucena Research | Summer 2012 and 2013

Portfolio Management: Created models for portfolio hedging, portfolio optimization and price forecasting. Also creating a strategy backtesting engine used for simulating and backtesting strategies. QuantDesk: Python backend for a web application used by hedge fund managers for portfolio management.

DỰ ÁN

QuantSoftware Toolkit

Vai trò:

Open source python library for financial data analysis and machine learning for finance.

Github Visualization

Vai trò:

Data Visualization of Git Log data using D3 to analyze project trends over time.

Recommendation System

Vai trò:

Music and Movie recommender systems using collaborative filtering on public datasets.

