

# Ankit Soni

✉ ankitsoni9@gmail.com

🏠 Vancouver, Canada

🏠 Perm. Resident

🌐 github.com/ankitson

## SUMMARY

I love building infrastructure & backend systems, doing a mix of design, software engineering and ops as needed. I'm particularly experienced in building large scale data pipelines, performance/profiling/monitoring on the JVM, distributed systems and math. However, I love most aspects of technology, and have the aptitude & training of a generalist, and believe in learning & applying the right tools for the task.

## EDUCATION

**Carnegie Mellon University** – B.S, Computer Science, Math Minor.

Pittsburgh, PA

TA for Great Practical Ideas in Computing

June 2014

Favorite Courses – Distributed Systems, Natural Language Processing, Programming Languages

## EXPERIENCE

### Applied Cloud Computing

Mumbai, India

*Senior Software Developer*

Mar '17 – Jun '18

- Architected and built a cloud system on AWS to analyze payment information, performing real-time streaming aggregation & analysis used to generate personalized discounts/promotions for customers.
- Built a clickstream analysis platform on AWS to understand customer behavior and drive changes on a web application.
- Built applications & infrastructure on AWS on a client-specific basis.

**Technologies: Scala, Java, Spark, SQL, AWS Kinesis, AWS EMR**

### LogicHub

Mountain View, CA

*Principal Software Engineer*

Oct '16 – Mar '17

- Did a mix of engineering + ops to build a pre-launch Scala/Play/Spark web app that helps security teams (SOCs) formalize, automate & scale their threat detection workflows.
- Optimized a compiler from a SQL-like DSL to Apache Spark, reducing execution times by 3x-5x for the most common cases. Built tooling & monitoring to benchmark Spark based operations, ensuring performance improvement/parity as we added functionality. Owner of product performance.
- Implemented algorithms for data provenance, streaming export of records & other features.

**Technologies: Scala, Spark, SQL, AWS, Play**

### Uber

San Francisco, CA

*Software Engineer - Driver Growth*

Jan '16 – May '16

- Built, launched & iterated on several experiments & features in the driver onboarding flow to grow driver signup rates (5-10%) and trip rates.
- Designed & implemented new features on the driver onboarding platform, such as phone number signup. Designed the data models, wrote API endpoints and code in a microservice environment.
- Migrated part of driver onboarding backend from querying a monolith to a newer microservice based architecture.

**Technologies: Python, Flask, React.js**

### AppLovin

Palo Alto, CA

*Software Engineer – Platform Team*

Oct '14 – Oct '15

- Built, operated and evolved a data analysis platform using Apache Spark & HDFS to enable both ad-hoc analysis and ETL to train predictive models to improve ad targeting.
- Built the infrastructure from zero to regularly training production models in 6 months. Ran on a cluster of >100 machines, analyzing ~100TB of data per job. Built APIs, data ingestion/pre-processing tools, notebooks for ad-hoc analysis & solved performance problems.
- Worked closely with the data science team to implement, optimize models & evolve the platform as per their needs. Solved problems around data ingestion, schema evolution, building an understanding of what data means and scaling/performance tuning jobs.
- Wrote a parallelized program to create database backups ~10x faster, automated to run periodically. Enabled the platform team to iterate quickly with confidence.

**Technologies: Scala, Java, SQL, Spark, Zookeeper, Parquet, Zeppelin**

**Amazon**

Seattle, WA

*Software Development Engineer Intern*

Jun '13 – Aug '13

- Built an MVP product to generate product recommendations for Amazon.com customers based on aggregated web browsing data.
- Built a data pipeline using AWS EMR, EC2 & S3 to generate training data & leveraged a collaborative filtering machine learning model to generate recommendations.
- Defined a concrete measure of URL-Product relevance and used it to evaluate the precision & recall of the model.
- **Technologies: Java, AWS EC2, S3**

**Avere Systems**

Pittsburgh, PA

*Engineering Intern*

May '12 – Aug '12

- Built an automated suite of tests for a web application using Python, Selenium & Javascript. Tested config for NAS storage accelerators, testing functionality across multiple browsers & operating systems.
- Deployed & operated the testing infrastructure (VMs & browsers), wrote a DSL to generate tests quickly and tooling to ensure consistent behavior across different browsers & environments.
- Built a web application using Django to display & filter the results of tests.
- **Technologies: Python, Javascript, Django, Selenium**

**PROJECTS & OSS**

---

**spark-jobserver**

- Implemented high-availability/multi-tenancy - #1 most requested feature at the time. Involved re-architecting and rewriting major parts of an application built using Akka actors using Akka clustering.

**Scala & Dotty**

- Decoupled the rendering & evaluation phases of the Dotty REPL to fix an issue relating to uninitialized values being accessible. Diagnosed the problem, designed a fix and implemented it independently.
- Implemented minor bugfixes around integer-parsing in scala-js.

**SKILLS**

---

**Programming Languages** – Java (8/10), Python (7/10), Scala (7/10), OOP & functional programming**Distributed Systems** – Fault Tolerance, Concurrency, Performance & Memory Tuning. HA, consistency, replication, reliability.**Data Analysis & Engineering** – Spark, Akka, Parquet, Hadoop, SQL**Web Development** – HTML/CSS/Javascript, React.js, web architecture – HTTP/caching/load balancing/queuing etc.**Cloud Platforms** – Amazon AWS