

Sahil Agarwal

Software Engineer

(+91) 9818308563 | sahil.agarwal94@gmail.com
in.linkedin.com/in/sahilagarwal94
github.com/SahilAgarwal321

EDUCATION

Institution	Subject	Grade
Apeejay Stya University	B.Tech. (Computer Science & Engineering)	3 .1 / 4
APJ School	Class XII (PCM & Computers)	90%

EXPERIENCE

Senior Data Engineer, H1 (www.h1.co)

Dec 2022 – Present

H1 provides insights for life sciences, academic medical institutions, healthcare systems, and payors.

- Data Delivery – Created multiple compute intensive ETL pipelines (Spark, Scala, Airflow, EMR, EKS) that deliver data to client via AWS S3, Azure Blob Storage, SFTP etc.
- Datafeeds (Spark, Scala, Python, Docker) – Migrated critical pipelines from Databricks to AWS EKS.
- Datafeeds – Created documentation and tests, automated all manual tasks, added yaml configurations. This led to reduced team workload and quicker onboarding of new clients. Also made code based and spark config optimizations to reduce runtime from 24-36 hours to 6 hours.

Senior Data Engineer, GroundTruth (www.groundtruth.com)

Dec 2019 – Dec 2022

GroundTruth delivers location-based insights across 100 million points of interest spanning 21 countries globally.

- Status Log – Created an API service, python client to access RDS DB (AWS Lambda, API Gateway, AWS SAM).
- IP accumulator – Created a job that extracts and processes user IPs from Athena table (Airflow, Spark).
- Dragonglass – Built a highly-available data pipeline (AWS SAM) to consume external data, where an S3 event triggers AWS Lambda which runs an AWS Glue Job (Apache Spark). The job populates an AWS Elasticache (Redis) cluster, which has a traffic of 75,000 QPS and a response time of 40 ms.
- GTID – Built a pipeline (Spark, AWS EMR) that reconciles multiple external and internal datasets and stores mappings into AWS Elasticache (Redis). The API over this database has a traffic of 40,000 QPS.
- Inventory Forecasting – Real time forecasting of campaign delivery based on user filters. Built a framework to fetch, analyze Theta Sketches and compare cardinalities with source data (AWS Athena).
- Brand Insights – Developed the backend for a client portal using GraphQL, Spring Boot, Java.

Software Engineer, The Smart Cube (www.thesmartcube.com)

May 2017 – Nov 2019

The Smart Cube provides research, analytics consulting for sectors like CPG, Financial Services, Industrials and Retail.

- Revamped architecture of a news aggregator. Re-designed MSSQL DB to MySQL and Django ORM.
- Built and deployed APIs and services to streamline access to internal tools (DRF, MySQL).
- Developed a Load Balancer using Flask, RabbitMQ.
- Built a time series forecasting engine which ranked 4th in M4 Competition (www.m4.unic.ac.cy).
- Developed a crawling framework (that crawled 1500 websites) using Scrapy, Selenium, Python, Django.
- Automated the deployment process of existing projects by creating CI/CD scripts (Bash Shell).

Technology Intern, Search Team, Snapdeal (www.snapdeal.com)

Jan 2017 – May 2017

- Implemented sorting of documents during index creation, early search termination of queries (Lucene, Solr).

Technology Intern, Craftsvilla (www.craftsvilla.com)

June 2016 – Aug 2016

- Created a WSGI Request Logger Middleware using Werkzeug library (Python, Flask).

TECHNOLOGY STACK

Python, Scala, Spark, Bash Shell, Java, Docker, Airflow, Jenkins, Kubernetes, Databricks, Flask, Django, MySQL.
AWS – Lambda, Glue, EMR, EKS, API Gateway, SNS, Elasticache, CloudFormation, SAM, RDS, Athena.

INTERESTS

Gaming. Movies. Reading. Writing.