# Saumil Shah

http://linkedin.com/in/saumil-shah-0257 | saumil0257@gmail.com | https://github.com/saumilsjsu
+1(510) 402-9585 | San Jose, CA

## Summary

Proficient in Data Structures & Algorithms and an avid problem solver actively participating in competitions.
Industrial exposure in creating automated python scripts to build ETL (Extract-Transform-Load) pipeline.
Experienced in Data Analysis and creating Visual Dashboards on SAS and Tableau.
Skilled in writing complex SQL queries and working with relational and NoSQL databases.
Hands-on experience with Containerization technologies like Docker and Kubernetes and most of Amazon Web Services.

## Education

**Master of Science in Software Engineering**                                                          Jan 2019 – Dec 2020

San Jose State University, San Jose, CA, USA. GPA: 3.54
Relevant Coursework: Software Systems Engineering, Enterprise Software Platforms, Cloud Technologies, Data Mining, Large Scale Analytics, Machine Learning, Big Data, Human Computer Interaction.

**Bachelor of Technology in Information Technology**                                            Aug 2014 – May 2018

NMIMS University, Mumbai, MH, India.
Relevant Coursework: Data Structures, Designing Algorithms, Object-Oriented Programming, Operating Systems, Data warehousing, Data mining, SAS, Database management systems, Cloud Computing.

## Skills and Tools

| | |
|---|---|
| **Programming Languages:** | Python, Java, Golang, HTML5, CSS3. |
| **Databases:** | Relational (MySQL, PostgreSQL) Non-relational (MongoDB, Firebase, Cassandra, Redis). |
| **Tools & Platforms:** | Git, AWS, Docker, Kubernetes, Tableau, SAS, Google Analytics, REST. |
| **Data Science:** | NumPy, Pandas, Seaborn, Matplotlib, TensorFlow, Pytorch, Keras, Sklearn, NLTK, Spacy, Gensim. |
| **Frameworks:** | Django, Flask, Kafka, Hadoop, Spark, Kafka, RabbitMQ. |

## Work Experience

**Data Engineer Intern**, DeveloperDB, Sunnyvale, CA                                            July 2020 – Present

- Design, develop and launch scalable data ETL pipelines to curate databases streaming gigabytes of data on a daily basis.
- Automate and optimize existing database design to aid in building a sustainable architecture.
- Build generic data scrapers to crawl data from various search engines performed extensive cleansing and preprocessing of raw data to procure analyzed actionable data to be used for modeling.
- Compute business-specific algorithmic solutions for feature implementation.
- Generate test rules to identify malformed or incorrect data points and troubleshooting data related issues.

## Project Work

**Quora Replica |AWS – NodeJS – MongoDB – MySQL| -** *(See Project)*                    Feb 2020 – Apr 2020

- Constructed in a team of 4 a distributed, robust, multi-tier application similar to quora.com capable of managing 10k questions, answers and users. Programmed all major use cases and delegated frontend responsibilities.
- Designed and implemented system architecture, REST APIs, Kafka as a message queueing service and hosted application on AWS as an EC2 instance.
- Achieved reliability and scalability on JMeter, Jest and Mocha and improved throughput by adopting caching strategies through Redis.

**Credit Card Fraud Detection (IEEE-CIS) |Python| -** *(See Project)*                        Oct 2019 – Dec 2019

- Led a team of 4 to train Naïve Bayes, Logistic Regression, Random Forest, SVM to detect fraudulent transactions using dataset provided by Vesta Corporation with 433 features and 1.9 GB of skewed and masked data.
- Enforced memory reduction and dimensionality reduction steps for optimal data preprocessing.
- Constructed a comparative analysis between classifiers based on five different metrics.

**Online Food Ordering Application |AWS - Go - NodeJS- Heroku| -** *(See Project)*      Feb 2019 – Apr 2019

- Configured containerized microservices by building 6 dockerized Go APIs running parallelly deployed on AWS as EC2 instances over multiple regions.
- Created 6 MongoDB clusters on AWS as EC2 instances to store data for each microservice.
- Ensured fault tolerance and scalability of architecture, where each service was running independently.

**Go Mini-Stack Project |AWS - Go - NodeJS| -** *(See Project)*                              Feb 2019 – Mar 2019

- Configured Amazon RDS and Elasticache in a private network to connect to 2 Go APIs under an internal load balancer.
- Deployed dockerized APIs on AWS instances and established Kong Gateway for URL cleaning in a public network.
- Connected dockerized NodeJS application to Kong for a GUI display.

## Activities

- Attended workshops in emerging technologies, hackathons and coding challenges including Google Hash Code 2020.
- Volunteered to collect old newspapers/books and represented Indian Development Foundation's W2K mission.