Basar Sohail Chowdhury

✉ sohail.chd0202@gmail.com | ☎ +1-925-789-8473 | **Github**: sohailchd | **LinkedIn**: linkedin.com/in/sohailchd/
Website: http://sohailchd.github.io/ | Location: San Francisco, CA

Software engineer with over 5 years of experience in design and development of backend systems and microservices for large scale applications in healthcare, finance, gaming and data storage domain. Created and designed services around cloud native solutions for large scale distributed services. Good at system design and architecture patterns. Recent course work in machine learning projects, advance statistics and data pipeline for bioinformatics analysis. Focused on distributed (map reduce) systems required for computational biology and analysis.

Skills and Technologies

Languages/Scripting: Python, Java, Golang, JavaScript (basic)
Frameworks: Flask, Angular/React, Java Spring Boot, golang web frameworks, Tensorflow/Keras
Tools/DB: Docker, Redis, Spark, Hadoop, Elasticsearch, AWS stack, MySQL, MongoDB, Airflow, K8s
Skills: Domain driven design, Hexagonal architecture, Designing/porting microservice services, Map Reduce, REST APIs, Linux, cloud native solutions and scaling, AWS stack
Course work: Statistics, Genome sequencing, Advance Algorithms, Machine Learning, Distributed computing

Education

1. Master of Science, *Bioinformatics,* (Northeastern University, Boston) *May, 2020*
2. Bachelor Of Engineering, *Information Technology, (*University Of Pune, Pune) *May, 2013*

Work Experience

Phil, Inc (SDE / Healthcare/SAAS/Prescription Hub**)**- *California, USA Sep **2020** to Present*
- As a backend engineer implemented core prescription management services and business rules
- Integrated external APIs for patient insurance, automated/optimised the insurance verification process
- Creating POCs for new frameworks and services, integrating frameworks and code improvement
- Optimizations and performance improvement for scaling the platform around DB partitions and load balancing based on the prescription profile.
- Lead the major client integrations and deliver the SLA on hard deadlines for stakeholders.

Technologies: *Golang, AngularJS, MongoDB, K8s/Containers, AirFlow, AWS EC2, AWS SNS/SQS,*

Verys (**SDE** / Royal Caribbean Cruises) *California, USA June **2019** to Dec 2019*
- As a backend engineer, I contributed to the core system for the flight reservation system to be integrated with the cruise booking system. It was a new service to be implemented within the 6 months deadline.
- Help create the flight search logic using Elasticsearch as primary search engine for real time flights
- Middleware components for processing of flight booking objects and creating adapters for cruise booking system, code cleanup and optimization by reducing/consolidating API calls

Technologies: *Java, Spring , AWS Services , MYSQL, ReactJS, Solr search, Elasticsearch*

Druva (**SDE** / Data backup and discovery / Unicorn startup) *India, Jan-**2017** to Aug-2018*
- Developed APIs for backup endpoint(PCs) and storage related services (AWS Glacier/ S3 / DynamoDB)
- Implemented cloudapps integration (Exchange mail/GDrive/Box) backup and restore APIs
- Created AWS cloud native services for compliance (HIPAA/PHI/Legal Hold) and post backup dedupe which improved the compression upto 10x, global dedupe strategy using S3/DynamoDB block hash
- Cloud services for parsing and syncing with block storage, creating snapshots and performance improvement

Achievements: Optimization of custom parser, reducing the data parsing job from 4 hrs to 30 mins
Technologies: *Python/Golang, Kubernetes, DynamoDB, Flask, AWS Cloud Services*

GSLabs (**SDE** / Chicago mercantile exchange) *India, July-**2015** to Jan-2017*
- Design and implement microservice at scale for the stream processing of finance data, primarily using Java spring boot. Lead the team with initial POC and set up the project with design and arch.
- Cloud migration of the legacy service to spark for real time data processing(40k messages/sec)
- Talking to clients and stakeholder, managing a team of 5 people and growing the project
- Created a container based distributed scheduler for automated jobs. Used round robin strategy to schedule jobs, automate many manual processes for reporting, Saved hours human effort.
Technologies: *Java, Spark, Redis, MongoDB, Linux, AWS EMR, EC2, S3, GIT*

Ubisoft (**SDET** / Gaming and Entertainment) *India, July-**2013** to July-2015*
- Ubisoft connect is a platform for gamers. Contributed to the payment system and game store implementation.
- Improve the player experience for PC, Implement the third party integrations for the online game store
- Creating custom tools for the dev version of SonyPlayStation (PS4) and contributing to the game logging and performance metrics for the game under test.
Technologies: *Python, Flask, C# .NET, window desktop forms, REST API*

Academic and Other projects

Pneumonia Detection using radiograph images:
https://github.com/sohailchd/PneumoniaDetection
- Goal was to detect pneumonia infection in patients using convolutional neural network (CNN) of radiograph images and create a prediction model to correctly identify the infection in patients
- Achieved about 96% accuracy on augmented datasets of 20GB, used google colab and python for implementation, training and predictions

Hadoop MapReduce and spark distributed processing:
- Implemented PageRank algorithm on 40GB of data using hadoop map reduce paradigm in AWS EMR clusters
- Goal was to understand distributed computation using map reduce paradigm and its limitations
- Created a data processing pipeline using spark to understand the basics of distributed realtime computation using spark on big data (10GB of followers data) from Twitter
- Goal was to understand the real time processing and understanding of spark computations and RDD concepts for in memory computation

Histogram Equalization using GPU CUDA:
- implemented Histogram Equalization using optimized GPU code in CUDA that provides the same functionality of the histogram equalization from OpenCV but can perform the algorithm faster

NGS sequencing of genomics data and RNA-Seq Analysis (R/Python) :
- Created a NGS genomics pipeline (Airflow) for sequencing of genomic data and annotation using BioPython.
- Projects on RNA-Seq analysis using DESeq2 on dataset from recount2