Arjun Srivastava

+91-877-084-9368 | hey@arjunsriva.com

linkedin.com/in/arjunsriva | arjunsriva.com | github.com/arjunbazinga

EDUCATION

Indian Institute of Technology Indore — 8.6/10
Bachelors of Technology in Computer Science and Engineering — *June 2016 – June 2020*

EXPERIENCE

BookMyShow — Software Developer
Catalog ETL — *July 2020 – Present*
- Wrote 3 kafka consumers which ingest transform and store data pushed by multiple teams to expose information for our front end by storing it in our hot caches (aerospike), search engine (Elasticsearch) and CDN layers (minio + Cloudflare) across two data centers. allowing us to scale to 1.1 million concurrent users.
- Added Kibana-APM tooling to measure latency, tracing to identify and solve bottlenecks, made critical event ingestion flow process go from 4 hours to 20 minutes

Masters API
- Rewrote a legacy CRUD API in Go (using Gin in the process) fixing major bugs, making it testable, while maintaining service for multiple teams.
- Acts like a central metadata store that is used across all teams for tagging every event inside bookmyshow.
- The heaviest request take less than 120ms on 99th percentile, down from 700 ms
- Wrote OpenAPI specs for consumers, which brought time taken for integration from couple of days to approximately 3 hours

Vabhra Entertainments — Data Scientist
Modelling and Optimization — *Aug 2018 – Feb 2019*
- Used A/B testing infrastructure to pick between multiple email campaigns
- Prototyped multi-armed bandit approaches to better run A/B tests by automatically selecting the best options and running test only for as long as necessary, minimizing cost.
- Collaborated with product team to define KPIs and to assess the progress thereof; also, to propose and build product analytics projects such as user segmentation.
- Created Dashboard for location, device types, email opening rates, for high level overview using plotly and dash.
- Mentored engineering, product, and business development teams in building a data-fluent culture.

Analytics Infrastructure
- Built the company's reporting and analytics functions almost from scratch, including the schema design and ETL of a data warehouse and an A/B testing platform.
- Created tooling to keep in house tracking of mails opened, link click data, with meta-data like, time, IP (for approx location), device used.
- Wrote a web scraper to track of tickets sold overtime from our ticketing partners.

Zebi Data — Distributed Systems / Blockchain Engineer
Highly Reliable Database Layer — *June 2017 – June 2018*
- Currently running on 10 nodes, held by 6 independent stake-holders, 100 txn/s load tested.
- Wrote a distributed, reliable, byzantine fault tolerant, strongly consistent key-value store .
- Used in the back-end for multiple services like restaurant check-ins, property and land records.
- Identified Bottlenecks in the back-end by measuring throughput,latency of each subsystem, to figure out what to optimise, increasing overall system throughput.
- Written in Go, using LevelDB, gRPC, Protocol Buffers, Tendermint.

IIT Indore — Software Engineer
Leave Management System — *Dec 2016 - Jan 2017*
- Wrote and deployed a leave management system for PhD grads for requesting and granting leaves.

- Currently in use inside the Computer Science department. Made the whole process paperless, reduced average turnaround time from 3/4 days to 4 hours

What Should I Study?
- Formulated scoring a good GPA as an optimization problem
- Used previous years marks data to create models to approximate future marks distributions.
- Using model, grading policy , marks obtained till date made personalized recommendations for the best subjects to study to maximise GPA

PROJECTS

Michael Burry - Discord Bot | *Typescript, Node, GCP, Postgres, Elasticsearch, Sentry*
- Wrote a discord bot, with features like play music from YouTube on voice channels, Covid-19 updates, kelly optimal returns calculator, reminders, message archiving/export; Currently running on 7 servers, 500+ messages per day.

Coronavirus Trends | *Python, Numpy, Scipy, Observable*
- Fit an SEIR Model to model Coronavirus growth trends, using live data from multiple sources
- Created visualizations to display the massive uncertainty in predictions due to noisy data, and the folly of trying to predict the expected final value using a noisy signal

NBC - Multimodal Optimization - Paper | *Python*
- Created a Novel genetic differential algorithm to achieve, state of the art results on the CEC2013 benchmark for multimodal optimization challenge
- Paper on track to be published in Applied Intelligence as : "Improved Differential Evolution based on Multi-Armed Bandit for Multimodal Optimization Problems"

YouTube Data Analysis | *Python, Pandas, Matplotlib, Multiprocessing*
- Analysed YouTube watch time data, from google takeout.
- Visualized interesting questions, like amount of time watched by channel, month, genre
- Parallelized YouTube video meta-data downloading using python multi-processing
- Discovered patterns like YouTube's monetization algorithms lead creators to make videos 10 mins long.

Learning To Walk using Reinforcement Learning | *Python, OpenAI Gym, Tensorflow*
- Trained bipedal agent using reinforcement learning (PPO) to walk/run.

TECHNICAL SKILLS

Languages: Python, Matlab, Go, Kotlin, Typescript, Javascript, C/C++, SQL, Bash
Databases: Postgres, MariaDB, MySQL, Aerospike, Elasticsearch, MongoDB, leveldb
Data Science/ Machine Learning: Jupyter, Pandas, NumPy, Matplotlib, Plotly, Scipy, Sk-learn, Tensorflow, Pytorch
Machine Learning : linear and logistic regression, random forest, boosted decision trees, naive Bayes, SVM, k-means clustering, Gaussian mixture model, SVD/PCA, tf-idf, LDA, word2vec, Anomaly detection

WORKSHOPS/CONFERENCES

Presented: Thesis Presentation "Multi-Strategy Differential Evolution for Multimodal Optimization Problems"
Conducted: Workshop "An Introduction to Arduino" , ~100 participants
Conducted: Workshop "An Introduction to Linux, Git, and Open Source", ~100 participants
Presented: Talk "Distributed Consensus : How to open your own bank", ~60 attendees
Attended: International Symposium on Computational Mathematics, Optimization, and Computational Intelligence

COURSES/MOOCS

Astrostatistics — IIT Indore: Only undergrad to attend a graduate level course on statistical methods used to model physical systems. Learned and implemented Metropolis–Hastings, Hamiltonian Monte Carlo, and time series methods like MA, AR, ARMA and ARIMA

Statistical Methods in civil engineering — IIT Indore: Implemented probability distributions like Gaussian, Weibull, Bernoulli, LogNormal, Poisson, methods like LSE, MLE to estimate parameters of distributions in given data.

Coursera Deep Learning Specialization — Standford: Learned face recognition, deep style transfer, and YOLO for object Localization in TensorFlow; Language modeling, Implemented seq2seq, Neural machine translation with attention