

Alex Joseph

(+91) 9176390201
alexsanjoseph@gmail.com
blog.alexsanjoseph.com
alexsanjoseph
alexsanjoseph

Experience

- Aug 2021 – Ongoing** **Principal Engineer, Data Architect, HBK India**
- Architect of the Data Platform, bringing disparate data products in the enterprise under the same roof.
 - Currently implementing the Batch Pipeline/Data Discovery modules using ArgoWorkflows/FastAPI service layer at the backend and a React Frontend.
- Apr 2020 – Aug 2021** **Principal Engineer, Data Architect and Team Lead, VIMANA**
- Led the largest and most senior team in the company, Data Engineering to scale the pipeline over a period of 6 weeks, enabling the product to handle 10x more data (20 billion records in a day) than previously.
 - Architected transitioning of pipeline originally hosted on AWS to a cloud agnostic model using Kubernetes eventually supporting enterprise customers in GCP and Azure.
 - Redesigned the Data Science Pipeline to use the Modern Data Stack rather than homegrown technologies, reducing the cost footprint by 90% and enabling the platform to scale automatically on demand.
 - Extended the Manufacturing Device Data Model on MTConnect standard to support Connected Vehicles.
 - Functioned as the Scrum Master of the Data Engineering team and the Product Owner of the Systems team when SAFe was adopted into the organization.
- Mar 2017 – Apr 2020** **Principal Engineer, Data Science Team Lead, VIMANA**
- Conceptualized and built the VIMANA Data Science Pipeline — a framework to enable advanced reporting in VIMANA by leveraging batch processing using R.
 - Created the proprietary configuration tool, Deviceomatic, in Python, which generates device configs in the MTConnect standard. This reduced the time to add a new machine tool into VIMANA from a few hours to less than a minute.
 - Designed the Data model for the real-time streaming analytics rules engine, Enrich which processes billions of data points per day from manufacturing devices.
 - Collaborated closely with an enterprise customer to build the Maintenance Manager product, doing E2E product development from conceptualization to validation in person on the shop floor in Japan.
 - Architected the Vimana Data Science Stack. In charge of exploring new technology to improve the product.
 - Mentored senior engineers and team leads to work them to lead their teams.
 - Collaborated with the CTO on the company's strategic roadmap.
 - Headed the tech facet of the Recruitment and Training panel, interviewing 100+ engineers.
 - Lead and Scrum Master of the 5-member Data Science team.
- Apr 2014 – Mar 2017** **Senior Engineer, Data Science Team Lead, VIMANA**
- Integrated Data Science into the VIMANA Product by working with the Application team by building production-quality pipelines to showcase reports directly in the product rather than via BI tools.
 - Digital Thread — Developed the first R MTConnect adapter on CRAN to parse machine tool data. This package also allows the user to create a simulation of machine data from G-Code and check for adherence between planned and actual operation. Worked in-person with the faculty at National Institute of Standards and Technology, Gaithersburg, Maryland.
 - TimelineR — Developed a CRAN R package to visualize time series data allowing the user to see discrete and numeric data in the context of each other on a single timeline.
 - Mentored the Data Science team and led the team in various conferences.
- Jun 2011 – Apr 2014** **R & D Engineer, Manufacturing System Insights India**
- Collaborated with a strategic enterprise partner in creating ProPEL, the flagship product of the company, from nascency. Worked with the sales team and showcased the product at the IMTS 2014, Chicago.
 - Interfaced with the client services team to understand customer requirements allowing interpretation of domain specific data to create custom reports.
 - Created algorithms using raw machine data to identify Value Added Time in a manufacturing process.

Education

- 2019–2022** **Georgia Institute of Technology, Atlanta**, Online Master of Science, Analytics
GPA – 4.0/4.0
- 2007–2011** **Indian Institute of Technology, Madras**, Bachelor of Technology, Mechanical Engineering
GPA – 8.23/10

Personal Projects

- Jan 2022* **Searching for a Riemann Hypothesis Counterexample**, Golang
- Finding a Riemann Hypothesis counterexample(ironically), by iteratively searching across integers.
 - This application follows the footsteps of a master programmer who did the same using Python.
- Sep 2021* **Github Bulk Downloader**, Golang
- Simple application to download repositories in bulk from Github using a configuration supplied as YAML.
- Oct 2020* **AWS Instance Recommender**, Python
- Webapp hosted in heroku to identify the cheapest combination of AWS nodes required to support a use case for a specified amount of CPU/RAM.
 - The problem is expressed as a linear optimization equation and solved using the PuLP package.
- Jun 2020* **Plagiarism Detection using Code Similarity Metrics**, Python
- Small official project to automatically detect candidates who plagiarized in a campus selection challenge
 - Successfully managed to find ~600 malefactors who were automatically penalized from a pool of ~2000 candidates
- Jan 2019* **BodyBuilder**, Python
- Python port of the famous BodyBuilder package on NodeJS to create Elasticsearch queries using a DSL instead of manually creating the query JSON.
- Jan 2018* **Openscale Health Data Visualization**, Python
- Simple application hosted in heroku to track personal health using exported weight data from a health app.
- Oct 2017* **Duolingo Streak Save**, Python
- Application to protect the Duolingo streak by automatically purchasing a streak saver item.
 - This application uses Cloudwatch to run an AWS Lambda function periodically.
- Jan 2017* **Scrum Agent**, Python
- Package to download user data from JIRA and identify user-defined patterns.
 - Uses webhooks in slack to notify the user when patterns are identified.
- Oct 2016* **Personal Blog**, Hugo
- Hosted in Github Pages using the Hugo Framework — containing the blog entries, portfolio and resume.
- Jan 2016* **CompareDF**, R Package
- Package to compare two datasets, published on CRAN with two major releases and many minor/patch releases.
 - Implemented state of the art paradigms in R packaging — CI/CD, code coverage and Docker deploy.
 - Has 70+ stars, with 80K downloads, currently averaging ~2000 downloads per month.
 - Supporting the package for 6+ years, with zero pending issues and semantically versioned releases.

Skills

- Languages* R (Expert), Python (Advanced), SQL (Advanced), NodeJS (Intermediate), Golang (Intermediate), Javascript (Beginner), Java (Beginner), HTML/CSS (Beginner)
- Frameworks* Tidyverse, GGplot, FastAPI, ReactJS, Hugo, Streamlit, Springboot
- Ops Tools* Docker, Kubernetes, Helm, Flux, ArgoCD, Ansible, Rundeck, Teamcity, TravisCI, ELK
- Data Tools* Grafana, Kibana, Kafka, Tableau, Argo Workflows, Prefect, Spark, Livy, PrestoDB/Trino
- Databases* Elasticsearch, TimescaleDB, PostgresDB, MongoDB, AmazonRDS, Google CloudSQL
- Cloud Tech* AWS, GCP, Azure
- Technologies* Microservices, Cloud Native, SaaS, Data Platform, MicroFrontend, Distributed Systems, GitOps
- Dev Process* Agile/Scrum, SAAFe, ISMSD, ITSM

Interests

- Languages* English | Malayalam | Tamil | Hindi | German | Spanish
- Fitness* Running | Calisthenics | Swimming | Surfing | Scuba Diving
- Travel* South India | Himalayas | Central Europe | Dubai | Georgia | USA | SE Asia | Japan
- Misc* Duolingo | WayBetter | Five-Thirty-Eight | The Ken | Codewars | Audiobooks | Seinfeld Strategy