Munich, Germany
linkedin.com/in/vaidyasejal

SEJAL VAIDYA

github.com/sejalv
vaid.sej@gmail.com

SUMMARY

- Data/ML Engineer, with 10 years of software development experience, and proficiency in Data Platforms, Infrastructure & Cloud architecture
- Recent experience with productionizing models, services, and workflows in ML & analytics
- Core skills: Python, AWS, Airflow, NoSQL, SQL, Spark, Flask, Docker

WORK EXPERIENCE

▉▉▉▉▉▉▉▉ | Healthcare | Munich, Germany   *June 2020 – Present*
Data Engineer

- Building a data platform for Healthcare products to research on patient journeys, drug discovery, clinical trials etc.
- Designed a centralized Data Lake for public health and behavioural data. Developed pipelines for ingestion and schema validation using AWS (Lambda, S3, Batch, ECR, RDS), Python, Flask, Postgres
- Collaborated with SRE/Infrastructure team to orchestrate deployment (CI/CD) flows using with Docker, Kubernetes, Terraform & GitHub Actions

Fineway GmbH | *Travel* | Munich, Germany   *Jan 2020 – May 2020*
Data Engineer

- Developed services/APIs for the Conversation engine to generate faster Trip Plans, redesigned them with horizontal scaling, and using Python, MongoDB, Flask, Celery, Docker
- Collaborated with Data Scientists on model deployment & serving, versioning, orchestration of training jobs, as well as pipelines for analytics & monitoring using Airflow, Python, PostgreSQL, Docker, Kubernetes

Freeletics GmbH | Fitness | Munich, Germany   *Oct 2018 – Dec 2019*
Data Engineer

- Redesigned system to provide analytics on training journeys of 35+ mil users, with event-driven microservice architecture in AWS, and a fault-tolerant infrastructure (IaC) with Terraform, Jenkins & Travis
- Developed scalable data pipelines to asynchronously ingest & process 1k+ events per minute, using Python, AWS (Lambda, Kinesis, DynamoDB, S3, Glue, Athena) & serverless (framework)
- Contributed to requirements gathering for product/marketing analytics to align with long-term data strategy
- Conducted a hands-on workshop on building a serverless app in Python & AWS for 20+ diverse engineers

Servify | Consumer Services | Mumbai, India   *Apr 2017 – Nov 2017*
Senior Data Engineer

- Architected a Data Platform to enable cross-functional analytics for post-sales & service lifecycle
- Led team to build a Data Lake and distributed pipelines to ingest & process streams using Kafka, Spark, HDFS/Hive, Tableau
- Created a demand forecasting engine for inventory management, to predict SKU usage and reorder parts on crossing MOQ. Implemented ML models (EWMA, scikit-learn, pandas, numpy) with Python & Flask APIs

LexisNexis Risk Solutions (Reed Elsevier/RELX) | Insurance | Mumbai, India   *Apr 2015 – Sep 2016*
Senior Software Engineer

- Developed data products to enable Fraud Analytics & Risk Management in Life & Health Insurance space
- Designed and developed a search engine with Information Retrieval for Identity Linkage of customers
- Facilitated Anomaly Detection to identify outliers within claims. Developed services for Named Entity Recognition, risk profiling, and user segmentation
- *On-Site Visit (Atlanta, USA)*: Collaborated on architectural decisions with business teams to develop POCs

***

WORK EXPERIENCE (cont.)

JP Morgan Chase Services | Finance | Mumbai, India   *Feb 2013 – Mar 2015*
Applications Developer
- Designed logical data models & BI reporting mechanisms for <u>Finance, Credit Risk & Marketing Analytics</u>
- Developed ETLs for Incremental & Historical data, using PL/SQL & Shell scripts, and utilities for auditing
- Redesigned DWH schemas with dimensional modelling, optimised query & database performance in Oracle

HERE Technologies (Formerly NAVTEQ) | GIS | Mumbai, India   *Dec 2009 – Feb 2013*
Geographic Technician
- Data management & Quality Analysis for Maps. Optimized SQL queries to improve data retrieval by 37%
- <u>Golden Impulse Award, Sep '12</u>: Developed test automation framework to reduce quality overheads by 45%

EDUCATION

University of Mumbai, India | 2006 –2009 | Bachelor of Science – Information Technology (IT) | First Class

---

INDEPENDENT PROJECTS & PUBLICATIONS

- [Containerized ML deployment with AWS Lambda](#) using AWS Lambda, Docker & ECR (2020-2021)
    - Workshop for [DataTalks.Club](#): [Serverless Machine Learning Deployment with AWS](#)
    - Repo: [github.com/sejalv/serverless-ml-workshop](#)
    - Medium: [sejalv.medium.com/containerized-ml-deployment-with-aws-lambda](#)
- [Data Engineering Confessions](#): Podcast interview for [Pipeline Data Engineering Academy](#) (2020)
- [SQL in a Nutshell](#): Article published in [Towards Data Science](#) (Medium)
- [Foundational projects for MLE & MLOps with PySpark](#)
- [Movie Recommendations Website](#): with CF & K-Means using Python, PostgreSQL, scikit-learn

---

PROFILE

Visa: EU / Blue Card
Links:
- [linkedin.com/in/vaidyasejal](#)
- [github.com/sejalv](#)
- [medium.com/@sejalv](#)
- [sejalvaidya.com](#)

◇◇◇◇◇◇