SANKET DATTAKUMAR DALVI

551-775-3180 | dalvisanket2010@gmail.com | linkedin.com/in/-sanketdalvi-/ | github.com/dalvisanket

EDUCATION

Rutgers University — *New Brunswick, NJ*
Master of Science in Computer Science | GPA 3.94 / 4.0 — *May 2024*

University of Mumbai — *Mumbai, India*
Bachelor of Engineering in Information Technology | GPA 8.39 / 10.0 — *Oct. 2020*

PROFESSIONAL EXPERIENCE

Data Engineering Intern — *June 2023 – Aug. 2023*
ThredUP — *Oakland, CA*

- Developed **Python** and **Spark SQL multithreaded** backup script for **Data Lakehouse**, allowing customizable schedules, auditing, and selective data exclusion; resulting in 30% performance gain over single-threaded approach
- Optimized data wrangling routines with **Pandas** and **Numpy** for **large-scale e-commerce datasets**, accelerating processing by **20%**
- **Pioneered and tested strategic roadmap** for workspace migration to **Databricks** Unity Catalog, enhancing data governance
- Leveraged **Terraform** to configure **AWS** S3 buckets, IAM roles, policies; improving data security and resource management

Software Engineer — *Nov. 2020 – May 2022*
Xoriant — *Mumbai, India*

- Deputed in **foundational team at Morgan Stanley**, architected **scalable microservices** to facilitate Regulatory Reporting
- Engineered robust **Java Spring Boot REST APIs**, reusable **JavaScript** components, setting framework adopted across teams
- Built scalable Java pipeline for **batch processing 500k records**, utilizing **Kafka** for high-throughput data ingestion and distributed processing, **Junit** Integration and Unit test suite **ensuring 95% coverage**
- Implemented **Shell scripts** for large file processing, automated via **Airflow** job, and enabled sub-10-second querying with **Spark**
- **Automated PostgreSQL change management** for bank's largest project, incorporating **MongoDB** to handle unstructured data for increased adaptability. Authored new and optimized existing **SQL ETLs**, resulting in a **25% reduction in query execution time**
- Led service setup, **Gradle** build steps, and **Kubernetes based Jenkins CI/CD** pipeline creation; optimizing resource utilization
- Collaborated with cross-functional teams in an **Agile** environment, gathering requirements to boost project throughput

Software Intern — *June 2018 – July 2018*
Prako Technocratz — *Mumbai, India*

- Developed **Python-Django** examination portal, enabling teachers to launch timed tests and students to participate
- Designed intuitive web pages for user signup and login, integrating **robust user authentication and defining access rights**

TECHNICAL SKILLS

Languages: Java, Python, SQL, TypeScript, JavaScript, Shell script, HTML, CSS

Databases: MySQL, PostgreSQL, IBM DB2, Greenplum, Databricks, MongoDB, Liquibase, JDBC

Frameworks & Build Tools: Spring, Spring Boot, Angular, React, Kubernetes, Gradle, Node.js, Maven, Jenkins

Other Technologies: AWS, Azure, Splunk, Docker, Kafka, Pandas, Numpy, GIT, JIRA, Postman, TensorFlow, Airflow

PROJECTS

Electricity Billing Management System | *Java, Springboot, React, Kafka, Gradle, MySQL, Liquibase*

- Architected **Java** full-stack microservices application with a **React.js** dashboard for automating electricity billing tasks, including meter reading, bill generation, and customer management with analytics
- Implemented **Kafka** event store and Java Spring Boot RESTful API calls for **parallel processing** and scalability, and utilized **Spring Cloud Netflix** for load balancing and improved response times
- Designed **MySQL** schema, used Liquibase to track schema, migration; streamlining database development and deployment

No ML Question Generator | *Azure, Apache Spark, SQL*

- Architected cloud-native Java-Spark application on **Azure**, leveraging dynamic SQL for dataset-specific question generation
- **Innovated system** capable of discerning categorical and numerical data types; refining precision and relevance of questions
- Leveraged Spark SQL for performance optimization, ensuring rapid generation of high-quality questions from large datasets

Plastic Waste Profiling | *Python, Tensorflow*

- Co-authored a **Machine Learning application** to identify plastic waste and its manufacturer from image, to create a location specific plastic waste profile, with **95% accuracy**
- Calibrated training parameters to categorize plastic waste and recognize duplicate images; increasing accuracy by 5%