

VIRAJ RATHOD

virajrathod99@gmail.com • (623) 396-9127 • [linkedin.com/in/virajrathod](https://www.linkedin.com/in/virajrathod)

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL
- **SQL Databases:** Oracle 12+, MySQL, Maria Db
- **Tools:** Debezium
- **ETL Tools:** DOMO, Tableau, Dataiku
- **Big Data:** Kafka, Zookeeper
- **In-Memory:** Apache Ignite, Time-series: Influx DB, Postgres, Prometheus
- **DevOps:** ELK & TICK stack, CICD (GitLab Runner, Jenkins, Ansible), Grafana, Sonarqube

EXPERIENCE

DigiCert, Inc.

Salt Lake City, UT

Data Engineer

May 2020 - Present

- Designed, planned, developed data pipelines to perform automated ETL between data sources, flat files and API-call based tables and created a Data Warehouse (TBs+ range), for strategic reporting and improved efficiency by 45%.
- Designed and implemented a real-time scalable data pipeline to process semi-structured data by integrating 10T+ million raw records from different data sources using Kafka and Python based CDC pipeline, and store processed data in MySQL based Data Warehouse, improving the data migration by 33%, used Kafka connect, Debezium connector, Cassandra and MariaDB MPP architecture.
- Designed a Python framework to automate dynamic certificate generation in MariaDB using class methods, static methods.
- Developed complex business queries for database management that fuel reporting and advanced analytics.
- Created customer-focused QA Audit scripts to identify data anomalies, outliers, spikes at each stage of the workflow, codebase being homegrown.
- Create real-time dashboards on Grafana, to track the machines, by exporting metrics from Prometheus for alerting purposes through OpsGenie.

University of Utah, Business Intelligence Services

Salt Lake City, UT

Data Science Intern

Sep. 2019 – Apr. 2020

- Devised a data lineage model using Scala and Spark's GraphX for centralized availability of Data for the University and visualized it using Neo4j.
- Implemented multi-user security for Dataiku, using Sentry and Active Directory for the isolation of active projects in the team.
- Developed two predictive models using student login data, student surveys, and financial status of the student using Dataiku, Tableau, and Cloudera cluster with a predictive accuracy of 65%

Reliance Jio Infocomm Limited

Mumbai, India

Data Engineer

Jul. 2018 - Jul. 2019

- Built a data ingestion pipeline of over 100M records using Apache Kafka, HDFS, and Apache ignite with complete data integrity and consistency.
- Developed, maintained, and supported the company-wide Continuous Integration (CI-CD) framework for code deployment based on Gitlab, which reduced the deployment time by 90%.
- Maintained 12+ Servers using Ansible, ELK Stack, and maintained using Grafana, Kibana dashboards, and TICK stack.
- Actively managed 10+ Azure servers, improved, and monitored cloud infrastructure on Microsoft Azure.

Data Science Trainee

Jan. 2018 - Jul. 2018

- Trained Deep speech Mozilla on 350M customer data records. Models and libraries used: CNN, RNN, Wordnet, NLP.
- Sentiment Analysis of Jio products: Automated the process of mining and analyzing feedback using SVM.

EDUCATION

University of Utah, David Eccles School of Business

Salt Lake City, UT

Master of Science in Information Systems

Aug. 2020

University of Pune

Pune, India

Bachelor of Science in Computer Engineering

Jun. 2017

ACADEMIC PROJECTS

Earthquake Damage Visualization (Final Project) [Project link](#)

Aug. 2019 – Dec 2019

- Effectively visualized the damage incurred to 20K buildings in San Francisco using HTML, CSS, D3.js and used Dataiku for predicting the damage percentage. Achieved a predictive accuracy of 79%.

Tag-based Image Search [Publication](#)

Jan. 2017 – Jun. 2017

- Created a flat-file containing tags from two big data sets and ranked them according to the image properties and return the searched image using the tag score.