

DEVARATA OZA

188 Tonnele Ave, Jersey City, New Jersey, USA, 07306

📞 732-522-6179 ✉ do309@rutgers.edu 🌐 <https://www.linkedin.com/in/devarata/> 🐙 <https://github.com/devarata>

Summary : *Software Engineer with 3+ years of experience skilled in Python, SQL, Spark, Hive, Big Data & Machine Learning. Proven track record in building end to end data systems, integrating data pipelines, and ensuring data accuracy.*

TECHNICAL SKILLS

Languages & Databases: Python, C++, TypeScript, Shell, SQL, Apache Hive, HiveQL, Oracle, MySQL, MongoDB

Big Data & Cloud: PySpark, Apache Spark, SparkSQL, Apache Hadoop, HDFS, MapReduce, AWS, GCP

ML & Analysis: Spark MLlib, TensorFlow, Pytorch, Pandas, NumPy, Matplotlib, scikit-learn, NLTK, spaCy, OpenCV

DevOps & Tools: BeautifulSoup, Requests, Git, Airflow, CI/CD, MLFlow, Control-M, Argo workflows, Docker

WORK EXPERIENCE

Paramount Global, Advanced Advertising Team

July 2024 – Present

Software Engineer (Data Engineer)

New York, NY

Project: Vantage | **Skillset:** Python, Snowflake, AWS S3, SQL, Airflow, Argo workflows, TypeScript, Pydantic, Docker

- **Migrated Comscore ETL processes** to Vantage ETLs, replacing Comscore API with RAPLib API, and executed schema migrations while validating reach data accuracy against production data in **Snowflake**.
- Developed and integrated end to end pipelines to transfer data from **Redshift** to **Snowflake** using **AWS S3** for intermediate storage, ensuring seamless data movement and integrity.

Center for Advanced Biotechnology and Medicine, Rutgers University

Jun 2023 – Sep 2023

Software Engineer (Research Assistant)

New Brunswick, NJ

Project: Mice Monitoring System | **Skillset:** Python, ReactJS, RESTful API, SQL, bs4, Raspberry Pi, MLFlow

- Led development of **scalable full-stack** prototype mice monitoring system using YOLOv8 model.
- Trained ML models on dataset of **10,000** self-annotated **images** and operated through a **multi-node** Pi network.
- Engineered **ReactJS**-based web app for mice monitoring system, integrating **RESTful APIs** for data integration.

Deutsche Bank, Legal Management Team

Jul 2019 – Jul 2022

Software Engineer (Senior Analyst)

Pune, India

Project: ELM Reporting | **Skillset:** PySpark, scikit-learn, Yarn, Hive, SQL, Tableau, CI/CD, XGBoost, Control-M, GCP

- Designed machine learning-based **multiclass classification system** for automated invoice classification, processing approx. 4 million invoices in minutes and saving **3 man-months of effort**.
- **Led migration** of over 3 TB of legal historical data from **Oracle to Big Data** for enhanced processing.
- Collaborated with cross-functional teams to design, implement, and maintain a **33-table Star Schema** data warehouse, ensuring data governance and quality through established policies, standards, and regular validation checks
- Utilized **PySpark, Hadoop, and Hive** for **ELT / ETL** pipelines to handle **5 GB** of data in **Parquet** format daily.
- **Designed and optimized 200+ SQL queries** for data consolidation across data marts, adhering to business logic.
- **Implemented CI/CD** pipeline, for streamlined development workflows, reducing code deployment issues by 60%.
- **Optimized report generation** using **40+ SQL scripts** and data extracts in **Tableau** and **SAP BO**.
- **Scheduled 10+ data pipelines** using **Control-M**, ensuring reliable and timely execution of ETL processes.
- Constructed prototype using Google's Cloud Platform (**GCP**) to deliver 5 banking services to people with disabilities.

EDUCATION

Master of Science in Computer Science

Rutgers University, New Brunswick

GPA: 3.83/4.0

Bachelor of Engineering in Computer Engineering

Thadomal Shahani Engineering College, Mumbai, India

CGPA: 8.34/10

ACHIEVEMENTS & AWARDS

- Recognized as department's best-performing employee, receiving **Deutsche Bank Recognition** Award for H2 2020.
- Secured a position among the top 8 teams in the **Deutsche Bank Global Hackathon 2021**
- Awarded 1st place at TSEC Project-Exp'o'19 for **Smart Waste Management** (Internet of Things)