Saurabh Takle

Jersey City, NJ | 5512806021 | saurabh.takle@gmail.com |linkedin.com/in/saurabhtakle | GitHub

EDUCATION

Pace University, NYC, NY | Master of Science (MS) in Computer Science, Concentration: Data Science 05/2023
Mumbai University, Mumbai, MH | Bachelor of Engineering (BEng), Major in Computer Science 06/2021

SKILLS

Database Management: SQL Server Management Studio, Oracle, MySQL, PostgreSQL, NoSQL (MongoDB), Dynamo DB
Data Visualization Tools: Tableau, Power BI (DAX), SAS, QlikView, Google Analytics, Looker, Excel
Programming Languages: Python (NumPy, Pandas, Scikit-learn), Scala, Bash, MATLAB, R, C++, Java, Perl, Javascript
Big Data & Cloud Services: AWS (S3, Lambda, Redshift), Azure (Databricks, Synapse), Apache (Spark, Airflow)
Version Control & DevOps: git, GitLab, Docker, Kubernetes

EXPERIENCE

Data Engineer, Product and Data Analytics | Vaandu LLC, Newark, NJ 08/2023 – Present

- Managed **Snowflake-based data warehouses**, optimizing schema design and query performance to **reduce data retrieval time by 25%**
- Worked in an **Agile Scrum** environment, utilizing **JIRA for sprint tracking** and **Bitbucket/Bamboo** for automated build and deployment, reducing deployment time by **40%**
- Developed **data workflows using Apache Airflow and DBT**, increasing ETL process efficiency by 40%
- Conceptualized systematic workflows on Azure using **Azure Data Factory** (ADF) to monitor real-time financial database updates, reducing error rates to **below 1%** when processing large datasets
- Engineered ETL pipelines using **Hive SQL** and **Spark**, streamlining **data ingestion** and **transformation** for largescale datasets, reducing processing time by 70%
- Leveraged **Azure Databricks** for **financial data modeling**, accelerating trend evaluations by 35% and delivering actionable insights for stakeholders
- Developed and maintained ETL processes using **SQL Server Integration Services (SSIS)**, enhancing data flow efficiency by 30%
- **Designed and executed A/B tests** to evaluate the impact of new features on user behavior, resulting in data-driven decisions that enhanced product usability
- Developed monitoring dashboards in **Tableau and Looker** to track data pipeline performance, reducing failure resolution time by 40%
- Collaborated with and supported the data science team by preparing tailored fact and dimension tables, improving their ability to build **data models** and generate actionable insights by 20%

Data Analyst, Sales and Business Intelligence | SAID Technologies, Pune, MH 04/2020 – 12/2020

- Orchestrated ETL workflows using **AWS Glue Scheduler** and **AWS Lambda** to streamline data warehousing in **AWS S3**, reducing weekly workforce hours and improving data delivery efficiency by 70%
- Analyzed large datasets using **Python (Pandas, Numpy)** and **SQL**, creating impactful visualizations to uncover product trends, leading to strategic bundling initiatives that increased total sales by 25%
- Enhanced analytics capabilities with **AWS Redshift** by optimizing **schema designs** and implementing indexing strategies, reducing query response times from 15 seconds to under 3 seconds
- Designed and generated quarterly statistical reports using **SQL queries** and **Power BI Dashboards**, improving report accuracy and reducing report preparation time by 17%
- Led cross-functional project teams to deliver data solutions on time, improving project completion rates by 15%
- Worked closely with stakeholders to gather business requirements, ensuring that analytics projects and visualizations addressed key performance metrics effectively

PROJECTS

Event-Driven Financial Data Processing System

- Developed a **Kafka-based** event-driven financial data processing system, enabling **real-time stock price tracking** and **anomaly detection** with a 40% improvement in data ingestion speed
- Integrated **SQL-based quality checks and Informatica Data Validation (DVO)**, improving data integrity and reducing processing errors by **30%**

Personalized Product Recommendation System: Scalable Data Pipeline on GCP

- Processed large-scale user activity data using Google Cloud Storage and Dataflow to identify patterns, increasing data processing efficiency by 30%
- **Applied statistical analysis** to optimize recommendation algorithms, resulting in a 25% increase in user satisfaction

CERTIFICATIONS & PUBLICATIONS

Certifications: Machine Learning, Tableau | **Publications:** DOI: https://doi.org//10.32628/CSEIT217287