Tanish Sakate

 linkedin.com/in/tanishsakate | ✉ tanishsakate@gmail.com | 📞 +91 9359237642

EDUCATION

Fergusson College — May 2022
M.S. Data Science — *GPA: 9.2/10*

Marathwada Mitra Mandal College of Commerce — May 2020
B.S. Computer Science — *GPA: 7.0/10*

SKILLS

Programming Languages: Python, Rust, Pyspark, SparkSQL, PL/SQL, Apache Airflow, Prompt Engineering
Dev Tools: Git/GitHub, Microsoft Azure, Azure Data Factory, Databricks, VS Code, SSMS, Visual Studio, Azure Data Studio, SQLProfiler, DBeaver, Jira, DevOps Boards, Snowflake
Databases: SQLServer, MySQL, MongoDB, CosmosDB, Cognitive Search, ElasticSearch

EXPERIENCE

Ellicium Solutions | *Software Developer* — Jan. 2022 – Present

With 3 years of experience in the tech industry, including 2.7 years specializing in supply chain industry analytics, I am proficient in continuous learning, acquiring new skills, communication, initiative, and customer focus. My technical expertise includes **Python, SQL/PSQL, PySpark, ADF, ETL tools, and Azure**. As a leader of a team of over 10 data engineers, I contribute to effective data engineering while prioritizing customer needs, ensuring operational compliance, and fostering a collaborative team environment. I am dedicated to further strengthening my career and expanding my skill set by delivering excellence.

PROJECTS

GTO — March 2022 – Present

- Developed an optimized, normalized data model to store supply chain information.
- Developed asynchronous Python web scraper pipelines using **asyncio**, reducing the time by 60% to extract supplier information and supplier network using **Beautiful Soup** and **Selenium**.
- Developed and later Migrated Python codebase to **PySpark** and developed **Data Factory** pipelines to trigger **Databricks** notebooks to fetch, transform, and load data into the database.
- Developed Data Factory pipelines integrated with SQL databases to track pipeline stages and failures, with email notifications to stakeholders and owners.
- Migrated whole Application SQL database to MongoDB as POC which included developing NoSql database and converting SQL queries to MongoDB syntax.
- Developed a Python supplier name deduplication algorithm based on **regex** and various **fuzzy logic** methods using **NoSQL** databases **Elasticsearch**/**Cognitive Search**, reducing consultant work by 55% and achieving an algorithm accuracy of 70% and storing resultant for future reference to reduce time and optimize the process.
- Developed a supplier research module API using FastAPI, OpenAI, and Bing API to fetch supplier names and all contact information from supplier websites using dynamic web scrapping based on industry or product names.
- Developed a Python web scraping API to extract supplier contact information from official websites.
- Developed an end-to-end RFP analysis and validation tool that fetches data from **ADLS**, processes the file, performs validations, creates analyses, uploads data into the database, and generates feedback based on the analysis.
- Developed a generic Excel parser using Python and created a dynamic SQL data model to store all information regardless of input, reducing redevelopment by 70% or more.
- Led a team of 3 data engineers to build the RFP analysis tool, create the ADF pipeline, and develop Databricks scripts and database model updates to keep it optimized.
- Developed an ETL tool where ADF is triggered when a file is uploaded to the SFTP portal. Databricks reads data into a notebook, performs data transformation generates input file machine learning prediction model, uploads the data into an Amazon S3 bucket, and notifies stakeholders with email notifications using SendGrid API integrated with ADF.

- Contributed to developing a supplier intelligence database with data modeling and optimizing SQL queries, creating supplier search API, validating and uploading the data into Cognitive search.
- Led a team of 4 data engineers to migrate the supplier intelligence SQL database and python codebase to the **CosmosDB** database over Azure WebApp.
- Should Cost Modeling: Interacted with clients to understand and gather requirements to develop should cost models for products, handle databases, and stored procedures, and present analyses on PowerBI dashboards.
- Developed a SQL procedure to dynamically identify missing columns and constraints in the production database during deployment from IAC scripts. Dynamically generated alter/update SQL statements and upgraded the database by referring to IAC, reducing the occurrence of redundant code errors.
- Leading group of data engineers to build RFP analysis tool considering extensive file validations, integrating external API to fetch currency, freight, duty, and tariff data. developed mapping algorithms and formula-based Excel sheets using Python.
- Leading group of data engineers to build a geospatial algorithm to find the most optimized path to transport vendor-to-vendor items considering sea routes and roadways.
- Lead data engineers in developing to migrate databricks code to Azure function app to reduce cold start time and better user experience.
- Developed a live socket python code to stream live data fetched from web crawler to avoid application backend failure.
- Led data engineering team to develop and integrate multiple external APIs to fetch data using ADF, applying multiple data sanity checks and storing it in the central database.
- Led data engineering team to develop Excel parser and should cost builder pipeline. Algorithm to map the best commodity to a product using regex and OpenAI and creating Complex uploadable template which reduced manual BA work by around 85%.