Tanish Sakate

LinkedIn | GitHub | Leetcode

Location: Pune,Maharashtra,India
Email: tanishsakatewrk@gmail.com | Mobile: 9359237642

DATA ANALYST/ENGINEER

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. Proficient in **Python, SQL optimization, query execution plans, ETL/ELT, and Azure technologies**, with experience in DevOps practices like software configuration, **CI/CD pipelines**, and deployments.
As leading the data engineering side of the entire project, 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 providing excellence.

TECHNICAL SKILLS

Languages	:	Python, SQL, Pyspark, SparkSQL, Rust, NoSQL
Frameworks	:	Apache Spark, Apache Airflow, Pandas, PostgreSQL
Libraries	:	BeautifulSoup, Selenium, OpenAI, Socket, Fuzzy Logic, SQLAlchemy, Pyodbc
Data Storage	:	SQL Server, PostgreSQL, MongoDB, CosmosDB, Elastic Search, Cognitive Search, Azure Data Lake Storage (ADLS), AWS S3
Dev Tools	:	Visual Studio, Visual Studio Code, Git, Databricks, Azure Data Factory, Jupyter Notebook
APIs & Services	:	RESTful APIs, ETL Pipelines
Cloud & DevOps	:	Azure, CI/CD pipelines

EXPERIENCE

Consultant — Mar 2024 – Present
Ellicium Solutions — *OnSite*

- **GTO** (Consultant) Mar 2024 –
  - * **Summary :**
    A leader managing and mentoring a team of 10+ data engineers while overseeing the entire data engineering aspect of projects. Skilled in delivering scalable data solutions, optimizing workflows, and automating processes to exceed client expectations. Proficient in maintaining high-quality outputs and providing strategic guidance to resolve complex technical challenges.
  - * **Roles and Responsibilities :**
    - · Led a cross-functional team of data engineers and business analysts to develop a robust and complex RFP analysis tool using **Python, SQL, Databricks** and **Azure Data Factory (ADF)**.
    - · Led the redesign of **data models** for a critical module, improving query performance and enhancing overall system efficiency.
    - · Designed and guided the development of a **Excel formula-based analysis tool and Dynamic cell formatting** using **Openpyxl**, collaborating with clients and mentoring data engineers to deliver a robust and interactive solution.
    - · Directed the implementation of a **geo-spatial tool** for route optimization using **Azure Maps service** and **GeoPy**, enabling accurate and efficient destination mapping.
    - · Developed an **API-based pipeline** using **Azure WebApp service** to automate should-cost calculations for the machinery industry, reducing manual effort by **90%**.
    - · Developed and Supervised the development of multiple **ETL pipelines** to ingest data from external APIs into a **central database** and ensured seamless data synchronization with client instances.
    - · Created dynamic **stored procedures** to streamline **Infrastructure as Code (IaC)** deployment scripts, improving deployment efficiency and consistency.
    - · Led the migration of **Python code** to **FunctionApp services**, ensuring code quality, adherence to standards, and improving system scalability and maintainability.
    - · Led and guided the team in migrating **supplier intelligence** from **SQL** to **NoSQL CosmosDB**, updating **API endpoint codes** in **WebApp service** to align with CosmosDB integration.
    - · Led the design and optimization of a chat-bot data model in **CosmosDB**, ensuring fast data retrieval, efficient insertion, and effective management of storage constraints.

- **GTO** (Associate Consultant)   Apr 2023 – Mar 2024
  - * **Summary :**
    - · I focused on **Python**, **SQL**, **Azure services**, **PySpark**, **Azure Data Factory (ADF)**, and **Prompt Engineering**. My role involved leading the data engineering team and overseeing development projects. I worked directly with client consultants to solve data-related problems, developed ETL pipelines, and created robust algorithms. Additionally, I developed data models for applications and prepared data for client demonstrations and pitches. I also trained a group of interns in Python.
  - * **Roles and Responsibilities :**
    - · Led the development of data pipelines for multiple client demos, showcasing business impact and technical feasibility.
    - · Migrated deduplication backend to **Cognitive Search Azure Service**, improving vendor deduplication accuracy to **80%** and significantly reducing execution time for each run.
    - · Developed the **Big Module data model** by applying optimal **normalization** and **indexing techniques**, enhancing query and database performance.
    - · Created Python **APIs** using **FastAPI** to parse Excel files with **Openpyxl**, validate and separate invalid files, update erroneous cells, store results in **ADLS**, and log data into an SQL database.
    - · Developed a feedback generation pipeline by integrating SQL, categorizing vendor bids, generating comments, and creating feedback documents using the **python-docx** library to send to vendors.
    - · Designed an algorithm to calculate savings using vendor input files and purchase order data, performing validations, standardizing currencies to USD, and ranking the best vendors based on the article.
    - · Guided the data team in designing an optimized **SQL data model** for supplier data by applying normalization techniques and indexing methods, improving query performance by **80%**.
    - · Preprocessed, transformed, and validated client data according to requirements using **Python** and **PySpark**, ensuring high-quality, analytics-ready datasets.
    - · Developed and Guided multiple robust **ETL pipelines** using **ADF** and **Databricks** for multiple clients, storing data into **SQL**, **ADLS**, **AWS S3**, and implementing mail notifications via **Azure SendGrid Service**.
    - · Implemented supplier search functionality using **OpenAI**, **BingSearch**, and **FastAPI**, enhancing user experience with live responses using **Socket**.
    - · Built a dynamic web scraping tool to extract information from supplier websites using **Bing Search API**, **OpenAI**, and **Azure Maps Service** for processing raw address text.
    - · Developed a versatile Excel consolidation tool leveraging **Openpyxl** and **OpenAI**, significantly reducing redevelopment efforts.
    - · Designed and implemented an **ETL process** for a should-cost use case, creating complex **stored procedures** and **views**, and integrating them into a **Power BI dashboard**.
    - · Developed a **cost breakdown generator** using **OpenAI** and **WebApp**, streamlining the product analysis process.
    - · Migrated supplier intelligence functionality to **Cognitive Search**, optimizing search performance by **98%**.
    - · Developed an automated email notifier using **Python** and **Pywin32** to alert clients on resource utilization, eliminating manual work. Created a portable Python package that installs dependencies on first use.

-

GTO PRIME (Software Engineer)   March 2022 – Apr 2023
  - * **Summary :**
    - · At GTO PRIME, Primarily focused as a data engineer/analyst to extract data from various sources, develop pipelines, and extract insights from data. Research and implementation of risk analytics. Integrating external APIs.
  - * **Roles and Responsibilities :**
    - · Developed and enhanced web scrapers using **BeautifulSoup**, **Selenium**, **AsyncIO**, and **multiprocessing** to extract data from multiple sources, reducing extraction time by **90%**, while researching and implementing algorithms for supplier risk KPIs.
    - · Developed multiple algorithms for AD-hoc tasks using **Regex**, **fuzzy** algorithms, and other methods to transform data.
    - · Developed an algorithm to find important attributes of a company's address from its raw extracts using **CosmosDB**.

- Developed a **Rule engine** to identify the level of risk and mitigation integrating external API for weather alerts.
- Migrated whole GTO PRIME database and queries to **MongoDB** as POC.
- Optimized SQL queries to reduce application loading time between **60**% - **90**%.
- Developed risk analytics pipeline using **Apache Airflow** integrating machine learning models and risk score algorithm and storing it normalized database maintaining data integrating.
- I contributed to the development of a pipeline that downloads articles from different web browsers using **Python**, **Redis**, **multiprocessing**, and **advanced Google search** techniques.
- Developed a vendor deduplication algorithm to increase data quality by 75% using **Elastic Search** database, **Hamming distance, Jaro-Winkler distance, Damerau-Levenshtein distance, Cosine similarity** and data cleaning methods.
- Developed address deduplication algorithm using **Python** and **SQL** to increase data quality by **70**%.
- Migrated and developed Risk analytics pipeline to **PySpark** on Azure **Databricks** and **ADF** pipelines to trigger them timely.

-

GTO PRIME (Data Science Intern)   Jan 2022 – March 2022

- * **Summary :**
    Gained practical experience in various data science technologies, including data warehousing, SQL, Python, Tableau, Power BI, and communication training. Deployed to assist the senior BI developer in creating dashboards.
- * **Roles and Responsibilities :**
    - Performed data cleaning and integration tasks using SQL, ensuring high data quality and reliability.
    - Designed and developed interactive Tableau dashboards, utilizing advanced DAX calculations to derive actionable insights.

EDUCATION

Fergusson College   Pune, Maharashtra, India
Master of Science in Data Science   *Nov 2022*
GPA: 9.20/10

Marathwada Mitra Mandal's College of Commerce   Pune, Maharashtra, India
Bachelor of Science in Computer Science   *May 2020*
GPA: 7.0/10