




# Tanish Sakate

 [linkedin.com/in/tanishsakate](https://www.linkedin.com/in/tanishsakate) |  [tanishsakate@gmail.com](mailto:tanishsakate@gmail.com) |  +91 9359237642

## EDUCATION

---

### Fergusson College

*M.S. Data Science*

May 2022

GPA: 9.2/10

### Marathwada Mitra Mandal College of Commerce

*B.S. Computer Science*

May 2020

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 2.7 years of experience in the tech industry, including 2.4 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 5 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**.
- Migrated Python code 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.
- 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%.
- Developed a supplier research module API using FastAPI, OpenAI, and Bing API to fetch supplier names based on industry and all contact information.
- 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 notebooks and database model updates.
- 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 cleaning, and uploads the data into an Amazon S3 bucket.
- Contributed to developing a supplier intelligence database with data modeling, creating supplier search API, uploading, and validating the input data.
- Led a team of 4 data engineers to migrate the supplier intelligence SQL database to the **NoSQL CosmosDB** database.

- Should Cost Modeling: Interacted with clients to understand and gather requirements to develop should cost models for products, handle databases, 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.