Tanish Sakate

Home : Flat 201 D-wing, Kville Ravet, Pune, 412101, Ravet, India

Email: tanishsakatewrk@gmail.com **Phone:** (+91) 9359237642

Date of birth: 27/06/1998 **Nationality:** Indian

WORK EXPERIENCE

[01/03/2024 – Current]

Consultant

Ellicium Solution

City: Pune | **Country:** India

- 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 an 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, ensuring 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 chatbot data model in CosmosDB, ensuring fast data retrieval, efficient insertion, and effective management of storage constraints.

[01/01/2022 – 01/03/2022]

Data science Intern

Ellicium Solution

City: PUNE | **Country:** India

- 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.

[01/03/2022 – 01/04/2023]

Software Engineer

Ellicium Solution

City: PUNE | **Country:** India

- 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 by integrating an external API for weather alerts.
- Migrated the entire GTO PRIME database and queries to MongoDB as a proof of concept (POC).
- Optimized SQL queries to reduce application loading time by 60%–90%.
- Developed a risk analytics pipeline using Apache Airflow, integrating machine learning models and a risk score algorithm, storing data in a normalized database while maintaining data integration.
- 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 an address deduplication algorithm using Python and SQL to increase data quality by 70%.
- Migrated and developed a risk analytics pipeline to PySpark on Azure Databricks and ADF pipelines to trigger them timely.

[01/04/2023 – 01/03/2024] **Associate consultant**

Ellicium Solution

City: PUNE | **Country:** India

- Led the development of data pipelines for multiple client demos, showcasing business impact and technical feasibility.
- Migrated the 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.

EDUCATION AND TRAINING

[01/01/2021 – 01/03/2022]

Msc Data Science

Fergusson College https://www.fergusson.edu/

City: Pune | **Country:** India | | **Level in EQF:** EQF level 7

LANGUAGE SKILLS

Mother tongue(s): Marathi , Hindi

Other language(s):

English

LISTENING B2 **READING** C1 **WRITING** B2

SPOKEN PRODUCTION B2 **SPOKEN INTERACTION** B2

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

DIGITAL SKILLS

My Digital Skills

Data warehousing | Power BI | ETL | Rust | WebAPP | Pyspark | Tableau | Data Modelling | Pysql | Python | PSQL | SQL | FunctionApp | Databricks | Data Factory | Microsoft Excel | Azure | ELT | CosmosDB | Elastic Search | Azure cognitive search | Python Beautiful Soup, Selenium | Apache Airflow. | Socket e Websocket | MongoDBCompass | OpenAI GPT2 | Git, Git Hub, SVN | RESTful APIS | visual studio, visual Basic | Git | Databricks/Pyspark | azure datafactory | Data Analytics | Relational databases