



PROFILE

Results-driven Data Engineer with a compelling record of consistently achieving and surpassing data-driven objectives. With over 6 years of professional experience, I have made significant contributions in open-source projects like Shikshalokam in developing & optimizing robust data plumbing systems and ETL processes. Most notably, I have automated and optimized batch processes to reduce manual workload while increasing runtime efficiency by 34%. Seeking to improve my skills over cloud technologies while contributing towards efficient and optimized data pipelines.

SKILLS

MAJOR

- Python
- BASH
- SQL
- Scala
- GCP
- Oracle
- Apache Druid
- Apache Kafka
- Apache Flink
- NoSQL
- Hadoop

SECONDARY

- PySpark
- PyTorch
- Flask
- Data Wrangling
- Data Visualisation
- Git / GitHub
- LLM
- Docker
- Data Scraping
- CI/CD

WORK EXPERIENCE

DATA ENGINEER

Cloudcraftz Solutions • 2022 - Present

Shikshalokam

- Maintained data pipeline up-time of 97.8% while ingesting streaming and transactional data across 8 different data sources using **PySpark**, **Druid**, **Azure Storage**, and **Python**.
- Orchestrated the data migration from **Apache Druid** to **CassandraDB**, transitioning to event-based data capture increasing performance by 17%.
- Utilised **Apache Flink**, and **Kafka** for real-time event pre-processing, enabling dynamic Cassandra-based reporting to be used in BI tools like **Apache Superset**.
- Developed cloud-agnostic data storage architecture, automated data-pipeline logs, and proposed **MongoDB** schema improvements for enhanced user data validation for support team usability.

Talent Spot (Product)

- Handled a team of three(3) interns, to design and create a Q/A model for screening candidate resumes using LLMs, where I acted as the maintainer of the GitHub project.
- Engineered text extraction using Python, leveraged **Pinecone** for text embedding storage, employed GPT-3.5-turbo for response generation, verifying metadata, applied top K algorithm to rank results according to job descriptions, created APIs for it to be interacted with and deployed the application on **AWS Lambda** for client demonstrations.
- Implemented memory in LLMs using LangChain, enabling contextual inquiries based on preceding questions for advanced conversation capabilities. Based on 122 unique resumes, the application was able to find relevant resumes at a approximation of 68%.

Udaan

- Analyzed transactional data to identify Profit & Loss trends for specific commodities. Developed a predictive **time-series** model, leading to a 28% improvement in vendor procurement KPIs.
- Scraped data from reference websites detailing daily commodities prices to facilitate price prediction for a day and a week to ensure inventory model as well as JIT model.
- Designed the database schema for vendor deals allowing for a systemic commodity procurement based on price, quantity and vendor relationship.

DATA SERVICES

Concentrix • 2019 - 2021

Intuit

- Handled complicated data issues by restoring damaged corporate files for Quick-books Enterprise & Premier products utilising **MySQL** commands (foreign keys, inner joins) with a satisfaction score of 89.3%.
- Engaged in technical support to address user-reported database challenges, employing a customer-focused approach. Effectively communicated solutions to non-technical users, showcasing the ability to translate complex data issues into user-friendly explanations.
- Implemented rigorous quality assurance measures in the analysis and repair of Quick-books database files. Ensured the accuracy and integrity of financial data crucial for small business operations including data migration and harmful entry identification.

CUSTOMER OPERATIONS

Amazon • 2016 - 2019

- Managed customer relations, participated in the pilot team for Concession Abuse Prevention, helping create business protocols and policies based on live customer queries.
- Served as a Subject Matter Expert (SME) in helping new hires understand Concession Abuse, along with helping team managers with sales queries data collection and analysis allowing for a streamline process for better business oriented results.
- Participated in monitoring and assigning tickets regarding payment failures and cancellations along with handling professional mail escalations reducing the payment failures by 15%.

CERTIFICATIONS

- Data Structures and Algorithms - **Jovian**
- A Workshop in Data & Analytics - **Sporjo**

EDUCATION

BACHELOR OF TECHNOLOGY

JIS College of Engineering : **2012 -2016**

PROJECTS

Audible Scraper

- This is a project that scraped Audio-books data for all possible categories present in audible.in. This is done using both **Selenium** and **Beautiful-soup** libraries of Python. I have used Selenium to manually navigate pages and BeautifulSoup to scrape the data from the navigated web-pages.