+1 (682) 247-7728
San Francisco CA 94538

KASI VANDANAPU

https://kasivisu4.github.io/
kasivisu3109@gmail.com

PROFESSIONAL SUMMARY

I'm a versatile Data Engineer with 2+ years of work experience, skilled in Pyspark, dbt, Python, Airflow and Google cloud services. My hands-on experience extends to web technologies, including React, D3.js, Node.js.

SKILLS

Languages: Python; JavaScript; Java; Bash; C.
Frameworks: Apache Spark; Apache Airflow; dbt; Delta Lake; D3.js; Databricks; ASG-Zena.
Databases: Firestore; MongoDB; DuckDB; MySQL.
Cloud Technologies: S3; Amazon Athena; BigQuery; Cloud Run; Pub/Sub; Cloud Functions; Kubernetes.
Certifications: Building Batch Data Pipelines on GCP; Smart Analytics, Machine Learning, and AI on GCP**.**

WORK EXPERIENCE

Data Engineer **Infosys, India** **May 2019 – Dec 2021**

- Designed a data transformation pipeline using **dbt**, leveraging **Google Cloud Services** for batch processing.
- Built complex data pipelines for Demand Forecast Application and Virtual Inventory Reservation prediction.
- Scheduled and monitored diverse data pipelines on weekly basis using **Apache Airflow.**
- Modified **Spark transformations** to meet the changing daily business requirements in **Databricks**.
- Enhanced data processing efficiency using window functions, reducing **Spark job runtime by 30%.**
- Established a streaming data pipeline through Google Cloud Service **Cloud Pub/Sub**.
- Incorporated a **PostgreSQL** based Metadata-driven framework for dynamic data transformations.
- Collaborated with clients including **Levi Strauss & Co, Kraft Heinz**, and **HSCS** as a data engineer.

Software Developer, Intern **Health Edge Inc, India** **Feb 2019 – May 2019**

- Actively developed microservices using frameworks **Spring Boot, Hibernate** and **Apache Maven.**
- Worked on the HealthRules platform, optimizing logic to align with organizational objectives.
- Utilized Liquibase to manage schema, ensuring consistency across development environments.

Intern Technology **Virtusa, India** **May 2018 – Jun 2018**

- Created an effective schema generator based on data using **Hadoop (MapReduce)** and **Python**.
- Designed a user interface using the Python GI library to showcase dataset schema and attributes.
- Gained valuable hands-on experience with **Apache Kafka** for enhanced real-time data streaming.

EDUCATION/ACADEMIC ROLES

- Northeastern University - **Master's in computer science** (Jan'22 – Dec'23)
- Amrita school of Engineering - **Bachelor of technology in computer science** (Aug'15 – May'19)
- **Teaching Assistant** - Human Computer Interaction (Summer'23); DBMS(Fall'23)
- **Research Assistant** - List Curator (Summer'22, Fall'23)

PROJECTS/RESEARCH

Benchmarking In-Browser Libraries (**Technologies: Arquero, DuckDB-WASM)** **Spring'23**

- Developed benchmarking platform to aid users in choosing data processing libraries.
- Built a **MongoDB client** that integrates with Observable, utilizing **Node.js** as intermediate server.

Hate Speech Detection (**Technologies: Pyodide, Natural Language Processing)** **Summer'23**

- Developed a dashboard using **XGBoost** and **TF-IDF** to classify hate and offensive speech.
- Published on the Observable platform, utilizing **Pyodide**, a browser-based Python distribution.

PUBLICATION

Author: Vandanapu kasi. Hadoop and Natural Language Processing Based Analysis on Kisan Call Center (KCC) Data; 2018 International Conference on Advances in Computing, Communication, and Informatics.