ABHISHEK NIGAM

9140022978 ◇ **nigama43@gmail.com** ◇ Bangalore, India
Linkedin

PROFESSIONAL SUMMARY

Dedicated and skilled Data Engineer with 3 years of hands-onexperience in designing, implementing, and optimizing data streaming solutions and Proven expertise in Big Data ecosystem. Also Worked on Pepsico Hands-on project using services from Azure to provide big data solutions as part of Incubation Lab. Able to interact with clients and team members to gather business requirements and implement the same.

EDUCATION

Pranveer Singh Institute of Technology Kanpur
B. Tech, Computer Science and Engineering 2017-2021

SKILLS

Prog. Languages:	Pyspark, Python, SQL
Cloud Technology	Azure Cloud, Azure Data Factory, Azure Databricks
Databases	MySQL, PostgreSQL, ADLS, Blob storage
Tools	Git, Github, Docker, Kafka, Airflow

EXPERIENCE

Data Engineer Nov 2021 - Present
Tiger Analytics India Private Limited
PROJECTS *Bangalore ,India*

- Developed a streaming pipeline for capturing change data records using **Kafka Debezium**. Built a consumer pipeline in **Databricks** using Spark API for reading change data records. Automated status checks for **Kafka connectors** using Python and set up alerts in Event Hub. Used **Docker** for starting the Kafka environment.
- Orchestrated workflows in an **Airflow POC**. Created DAGs for various operators, including Data Factory and Databricks run-now operators. Worked on dynamic DAGs and managed data passing between tasks.
- Worked on Sales database Pilot project as part of Incubation lab using Azure Services mainly **ADLS gen 2** for storage, **Azure databricks** for transformations, **Synapse SQL pool** for storing data, **Power BI** for dashboarding, **Logic app** and **Function app** for mail notification also created fake data using faker library for filebased ingestion. All resources were organized under a single pipeline using **Azure Data Factory**.
- **Optimized** an existing **ETL pipeline** in **Azure Databricks** to enhance performance and reduce processing time..
- Also Developed **ETL pipeline** in **Azure DataFactory** and Azure Databricks which copies data from source and integrated Azure Databricks Notebooks in **Pyspark** for transformation and ingest data to sink in parquet format for further use.
- Implemented **Kafka consumers** in Azure Databricks to retrieve, transform, and load data into a target data warehouse.
- Developed and deployed ETL pipelines using Azure Data Factory and Azure Databricks to process and transform large datasets. The project involved creating data ingestion workflows in Azure Data Factory to extract data from various sources, followed by transformation processes in Azure Databricks. Leveraged PySpark for data transformation and cleansing, and orchestrated the entire workflow using Azure Data Factory.
- Implemented **Delta Lake** for optimized data storage, enabling **ACID transactions**, scalable metadata handling, and efficient data versioning.
- Improved overall pipeline reliability and scalability, handling growing data volumes efficiently.

ACHIEVEMENTS

- Reduced ETL processing time by 50% through strategic optimizations and resource management.
- Achieved a 30% reduction in data processing latency by optimizing Kafka producer and consumer configurations in real time streaming pipeline.