

ABHISHEK KUMAR MAURYA

EMAIL abhishek2f24@gmail.com
MOBILE +91- 9506212886



Dynamic and **result-driven Data Engineer** keen to pursue roles preferably in **IT** and **Consulting** sectors with a focus to leverage expertise in **Data Engineering, MLOps** and **Cloud Computing** to drive efficiency and success for the right career prospects.

PROFILE SUMMARY

- Accumulated **over 4 years** of **extensive experience in data engineering** and **MLOps**, specializing in the Technology sector, with a proven track record of delivering impactful data-driven solutions.
- Currently **excelling as a Data Engineer at Tata Consultancy Services Limited**, architecting and implementing advanced data processing frameworks that facilitate real-time analytics and reporting.
- Developed a comprehensive **skill set encompassing data integration, machine learning model deployment, and data pipeline optimization**, contributing to a robust understanding of data lifecycle management and analytics.
- Attained significant **improvements in data processing efficiency by implementing a real-time data processing architecture**, resulting in enhanced decision-making capabilities for stakeholders.
- Showcased **proficiency in knowledge of cloud-based data solutions**, particularly within **Microsoft Azure**, enabling the design and implementation of scalable and secure data architectures.
- Spearheaded a **project that processed over 5 TB of data daily**, optimizing workflows and reducing latency, which directly contributed to improved operational efficiency and business agility.
- Well-versed in understanding of data governance and quality assurance practices**, ensuring the integrity and reliability of data used for analytics and reporting.

CORE COMPETENCIES

- Model Deployment & Data Pipeline
- Data Warehousing & ETL
- Monitoring and Logging
- Natural Language Processing
- Cloud Data Platforms & CI/CD
- Serverless Computing
- Feature Engineering
- Real-Time Processing
- Infrastructure as a Service

SOFT SKILLS



EDUCATION DETAILS

2020: Integrated Master of Science, Applied Mathematics, SV National Institute of Technology | Surat, Gujarat

TECHNICAL SKILLS

- Programming & Tools:** Python, SQL, Pyspark, Flask, Docker, Kubernetes, GitHub CI/CD
- Big Data & Streaming Tools:** Apache Kafka, Delta Lake, Azure Stream Analytics, Event Hub
- Cloud Platforms:** Microsoft Azure, AWS, Google Cloud
- ETL/ELT Tools:** Power BI, Databricks, MLOps Pipelines
- Data Analytics & Modeling:** Statistics, Feature Engineering, Dimensionality Reduction

CERTIFICATIONS

- Databricks Certified Data Engineer Associate (Planned)
- Designing and implementing a Microsoft Azure AI Solution, Microsoft | Oct, 2024
- Microsoft Azure Data Scientist Associate, Microsoft | Dec, 2022
- Machine Learning on Amazon Web Services, Amazon | May 2021

- ML with TensorFlow on Google Cloud Platform Specialization, Google | Aug, 2019

WORK EXPERIENCES

Tata Consultancy Services Limited, Pune

Data Engineer

Jan 2023 – Present

- Key Result Areas:
- Designing and implementing a robust real-time data processing architecture leveraging Azure Stream Analytics, Event Hub, Apache Kafka, and Databricks with Delta Lake, enhancing data flow efficiency and responsiveness.
 - Integrating various data warehouses to facilitate the cleaning, consolidation, and presentation of unified datasets, thereby improving data quality and accessibility for analytics.
 - Streamlining the ingestion, processing, and distribution of streaming data, enabling timely and actionable insights for informed decision-making.
 - Contributing to enhanced operational performance by ensuring the reliability and scalability of data processing systems, ultimately supporting strategic analytics and reporting initiatives.

MLOps Engineer

Apr 2021 – Dec 2022

- Key Result Areas:
- Designed & deployed an advanced chatbot featuring Natural Language Processing (NLP) capabilities, including Named Entity Recognition (NER), sentiment analysis & dialogue management, successfully integrating it across multiple platforms to enhance user interaction & support.
 - Engineered a high-performance semantic search engine for enterprise data, efficiently processing approximately 1 million entries with an impressive response time of around 100 milliseconds, significantly improving data retrieval and accessibility.
 - Developed a data-cleansing component for sensor data streams, utilizing exploratory data analysis (EDA), ETL/ELT processes, normalization, and dimensionality reduction techniques, ensuring data integrity and enhancing analytical capabilities.
 - Created and deployed robust end-to-end machine learning models and pipelines tailored for real-world applications, driving actionable insights and facilitating user engagement through seamless integration in operational workflows.

PROJECTS UNDERTAKEN

Scalable Data Pipeline for Enterprise Analytics

- Key Result Areas:
- Designed and implemented a scalable ETL pipeline capable of processing over 5TB of data daily from diverse sources, seamlessly integrating it into a Snowflake data warehouse.
 - Enhanced data workflows to achieve a 40% reduction in latency, facilitating near real-time analytics that support critical business decisions.
 - Ensured data integrity through automated validation, imputation, and deduplication checks, significantly improving overall data quality.
 - Leveraged Apache Airflow for efficient orchestration of data workflows, promoting high efficiency and reusability across processes.

Real-Time Streaming Analytics Platform

- Key Result Areas:
- Designed and implemented a robust real-time data streaming architecture utilizing Apache Kafka, Azure Event Hub, and Delta Lake to effectively handle approximately 1 million IoT events per hour.
 - Developed resilient, fault-tolerant data pipelines that ensured seamless data ingestion and transformation, even under high-load scenarios.
 - Integrated cleansed and enriched datasets with Power BI dashboards, facilitating real-time visualization and enabling actionable insights for stakeholders.
 - Applied strategic data partitioning and compression techniques, achieving a 30% reduction in storage costs while preserving optimal query performance.