Fahmida Mahmood

DATA ENGINEER

Denton, TX | Mobile: 202-688-1593**| Email:** fahmidamahmood801@gmail.com

SUMMARY

- Accomplished Data Engineer with over 4 years of experience in designing, developing, and optimizing data pipelines and architectures for diverse industry verticals.
- Proficient in implementing data engineering concepts and methodologies, including SDLC, Agile, and Waterfall, to drive data-driven solutions and enhance business processes.
- Expertise in programming languages such as Python, SQL, C, and C++, leveraging them to extract, transform, and load data efficiently.
- Expert in designing efficient and reliable ETL data pipelines using Hadoop and Spark.
- Skilled in utilizing popular data science libraries and frameworks like NumPy, Pandas, Matplotlib, SciPy, Scikitlearn, and TensorFlow for advanced data analysis and machine learning applications.
- Proficient in data visualization using Tableau, Power BI, and Advanced Excel, providing insightful and actionable visual representations of complex data.
- Experienced with cloud platforms like Amazon Web Services (AWS) and Google Cloud Platform (GCP), employing them to design scalable and cost-effective data solutions.
- Demonstrated proficiency in managing databases like MySQL, PostgreSQL, and MongoDB, ensuring data accuracy and accessibility.
- In-depth understanding of data engineering tools like Apache Spark, Apache Hadoop, Apache Kafka, and AWS Glue for handling big data and enabling real-time data processing.
- Skilled in deploying and managing containerized applications using Docker, ensuring seamless development and deployment processes.

SKILLS

Methodologies:	SDLC, Agile, Waterfall
Programming Language:	Python, SQL, C, C++
Packages:	NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, TensorFlow
Visualization Tools:	Tableau, Power BI, Advanced Excel (Pivot Tables, VLOOKUP)
IDEs:	Visual Studio Code, PyCharm, Jupyter Notebook
Cloud Platforms:	Amazon Web Services, Google Cloud Platform (GCP)
Database:	MySQL, PostgreSQL, MySQL, MongoDB
Data Engineering Concept	Apache Spark, Apache Hadoop, Apache Kafka, AWS Glue, ETL/ELT
Other Technical Skills:	Docker, Amazon Redshift, Snowflake, Data Quality and Governance, Machine Learning Algorithms, Natural Process Language, Big Data, Advance Analytics, Data Mining, Data Visualization, Data warehousing, Data transformation, Critical Thinking, Communication Skills, Presentation Skills, Problem-Solving
Version Control Tools:	Git, GitHub
Operating Systems:	Windows, Linux

EDUCATION

Master of Science in Computer Science - University of North Texas, Denton, Texas

Bachelor of Engineering in Information Technology – Chaitanya Bharathi Institute of Technology, India

EXPERIENCE

Data Engineer | Freddie Mac, TX | May 2022-Present

- Spearheaded the modernization and optimization of loan origination data infrastructure at Freddie Mac, implementing a cloud-based data lake architecture on AWS.
- Successfully utilized Amazon S3 for efficient storage of raw and processed data, ensuring improved data accessibility and availability.
- Prepare detailed documentation for Ab-initio used applications like Design Documents, Run Books, Code Checklists, Test plans and so on.
- Design, Develop, deploy and support Teradata procedure to encrypt the data in database layer and created used defined functions to be called from ETL layer
- Led the implementation of AWS Glue for automated ETL processing, resulting in a remarkable 40% reduction in infrastructure costs.
- Developed and deployed data pipelines with Apache Spark, enabling parallel processing and achieving a 30% increase in data processing speed.
- Implemented rigorous data validation routines and checks, leading to a 25% reduction in data errors and enhanced data accuracy.
- Implemented data lineage tracking and cataloguing using Apache Atlas, ensuring robust data governance and compliance with regulatory requirements.
- Employed Python and scikit-learn to automate mortgage default risk prediction, significantly improving risk assessment model accuracy by 20%.
- Streamlined ETL processes through Apache Kafka, enabling real-time data ingestion and reducing data ingestion latency by 50%.

Data Engineer| Exert Infotech, India | Jun 2019-July 2021

- Led the development of a recommendation engine for an e-commerce platform, using collaborative filtering and content-based filtering algorithms with Python and Apache Spark.
- Built a robust cloud-based data pipeline on AWS, integrating S3 for data storage, AWS Glue for ETL processing, and Apache Kafka for real-time data streaming.
- Achieved a 30% reduction in data processing latency by leveraging Apache Kafka, enabling near real-time product recommendations.
- Data Migration from existing Teradata Systems to Hortonworks HDInsight cluster on Azure
- Design and Implemented ETL for data load from heterogeneous Sources to SQL Server and Oracle as target databases and for Fact and Slowly Changing Dimensions SCD-Type1 and SCD-Type2.
- Implemented machine learning models with 85% accuracy for personalized product recommendations, leading to a 25% boost in sales revenue.
- Developed APIs using Flask for seamless integration of the recommendation engine with the e-commerce platform.
- Successfully increased customer engagement by 20%, resulting in improved user interactions with the platform.
- Improved average order value by 15% through effective and relevant product recommendations.
- Utilized AWS services (S3, Glue, Kafka) and Python for scalable and cost-effective data processing.