

UMAIR MAQSOOD

Phone: (+91) 7014859789

Email: syedumair9574@gmail.com

3 years | Data Engineer | Immediate Joiner

Nationality: Indian

Age: 26 | Gender: Male

Language : English, Hindi, Urdu

PROFILE

Experienced Data Engineer with over 3+ years specializing in Big Data and cloud data warehousing. Proficient in designing and optimizing data pipelines using Hadoop (HDFS, Hive, Sqoop, Spark) and cloud platforms like AWS (S3, Glue, Redshift) and Azure (Data Factory, Synapse, Databricks). Skilled in ETL processes, PySpark, and Databricks, with extensive knowledge of diverse file formats such as JSON, PARQUET, and AVRO. Hands-on experience with Apache Kafka for real-time streaming and Spark memory tuning. A quick learner and effective problem solver who thrives in Agile environments, leveraging GitHub for version control.

SKILLS

- **Cloud Services:** Amazon Web Services (AWS), Microsoft Azure.
- **Data Ecosystem:** Hadoop, Sqoop, Hive, Apache Spark, Airflow, Databricks.
- **Databases:** MySQL, Teradata, PostgreSQL, DB2, MongoDB, HBase
- **Languages:** Python, SQL, Spark, Scala, Unix Shell scripts
- **Data Warehouse:** Hive, AWS Redshift, Snowflake
- **Job Scheduling:** Airflow, AWS Step Functions
- **IDE:** PyCharm, IntelliJ, Jupyter Notebooks
- **Version Control:** Git, BitBucket
- **Software Tools:** WinSCP, VMware Workstation, PuTTY
- **Operating Systems:** Linux, Windows
- **Methodologies:** Agile/Scrum, Kanban

PROFESSIONAL EXPERIENCE

Intellisense Software Private Ltd | Data Engineer | April 2023 – November 24

Projects:

Project 1: Data Migration - AWS

Project Role: Data Engineer

Client: Adidas

Geo Location: US-based

Environment: Python | Rest-API | Airflow | AWS – (S3, Glue, Redshift) | Github | Scrum.

Description: The use case is to Migrate data from on-premises to AWS cloud and on top of that business requirement is to perform ETL operation using Glue on data available in S3 and to store Transformed data in Redshift according to their business requirement.

Responsibilities:

- Built an ETL by integrating AWS - S3, Glue, Redshift services.
- Extracted the data from client's server through API and imported raw data into AWS-S3.
- Automated the extraction process through Airflow.
- Utilized AWS-Glue service to create tables in Glue Data Catalog for the data present in S3 using Crawlers.
- Created Glue jobs to summarize, transform and load the data into Redshift by utilizing the Glue Catalog Tables.
- Integrated Github to Glue jobs to maintain the script changes effectively.
- Implemented event based triggers and scheduled triggers for effective and continuous migration of data from S3 to Redshift
- Performed Analysis on data available in Redshift using Redshift Query Editor.
- Participated in Business Requirement gatherings to analyze and transform the data as per their requirement.

Project 2: Data Migration - Microsoft Azure

Client: Nike

Geo Location: US-based

Environment: Azure Data Factory | Azure Blob Storage | Azure Synapse Analytics | Azure Databricks | Azure Data Catalog | Azure Purview | Python | PySpark | HDFS | Hive | Airflow | Sqoop | MySQL | Git | JIRA | VM | Agile methodology

Description:

The project involved migrating data from on-premises infrastructure to the Azure cloud, ensuring the transformation of data in alignment with business requirements. ETL operations were conducted using Azure Data Factory and Databricks, with the transformed data stored in Azure Synapse Analytics according to specific business needs.

Responsibilities:

Extracted data securely from the client's server using API calls and imported it into Azure Data Lake Storage.

- Developed ingestion pipelines with PySpark in Azure Databricks for efficient data processing.
- Created and implemented extraction pipelines using Azure Data Factory, AzCopy, and custom scripts as needed.
- Designed data loading pipelines to transfer data efficiently to Azure storage and databases.
- Established logging and monitoring using Azure Log Analytics for real-time tracking and error handling.
- Tuned PySpark jobs for optimal performance, utilizing partitioning and bucketing strategies to enhance processing efficiency.
- Ensured accessible metadata management using Azure Purview to maintain data lineage and compliance.
- Implemented security measures for sensitive data throughout the migration process.
- Followed data governance policies and ensured regulatory compliance throughout the project lifecycle.

TVS Credit | Software Developer | November 2021 – March 2023

Project Name: Vendor Management System (VMS) Data Integration and Analytics

Project Role: Software Developer

Environment: MySQL | Sqoop | Hadoop | Spark (Python, Scala) | Hive | HDFS | GitHub | Agile

Description:

The project focused on integrating vendor data from the Vendor Management System (VMS) into a Big Data ecosystem. Data from MySQL databases was imported into HDFS, transformed, and stored in Hive for processing, analysis, and reporting. The system supported vendor registration, invoice processing, and Capex/Revex management while ensuring compliance with SLAs and TATs.

Responsibilities:

- Utilized Sqoop to import vendor and invoice data from MySQL into HDFS.
- Developed Spark-SQL queries to transform and process data from HDFS to Hive tables using DataFrames.
- Created Spark jobs to aggregate and transform vendor and financial data stored in Hive.
- Implemented dynamic partitioning and bucketing in Hive to enhance data access efficiency.
- Processed and analyzed large datasets of structured and semi-structured vendor and invoice data using Big Data techniques in a multi-node Hadoop cluster.
- Performed advanced analytics on the data using Spark-SQL, DataFrames, and RDDs.
- Optimized Spark jobs for performance and enhanced the execution of Spark SQL queries.
- Collaborated with business stakeholders to gather requirements and stored processed data in HDFS using appropriate file formats.
- Supported the business team in data validation, ensuring data quality and accuracy throughout the system.

Internship | Accend System | July 2021 – November 2021

Roles and Responsibilities:

- Gained hands-on experience in SQL and Python for data extraction and manipulation.
- Assisted in developing basic SQL queries to retrieve and update data from relational databases.
- Assisted in documenting processes and database schemas for internal use.

EDUCATION

Amity School of Engineering & Technology (Amity University)

B. Tech in Computer Science & Engineering

Noida, India

2017-2021