

Ulhas Meshram

AWS DATA ENGINEER

✉ ulhas.meshram1991@gmail.com

📞 +91-8149403829

PROFESSIONAL SUMMARY

- **Highly skilled and motivated** AWS Data Engineer with 4+ years of **professional experience** in designing, implementing, and optimizing data pipelines on the AWS cloud Using PySpark, SQL, Hive.
- **Experienced working in AWS cloud computing platform** and many dimensions of scalability including EC2, Pyspark, Redshift, Glue, RDS, S3, Athena, lambda, EMR, CloudWatch, SnowFlake, SNS, Airflow.
- **Developed** Airflow, Data pipeline to run Spark, EC2, Glue, Redshift that run **independently with time and data availability.**
- **Demonstrated expertise in data modeling**, ETL processes, and data warehousing principles, **ensuring the successful delivery of high-quality results within challenging deadlines.**
- **Familiarity with big data technologies**, such as Apache Hadoop, SnowFlake, AWS Kinesis for **processing and analyzing large volumes of data.**
- **Experienced in data quality assurance**, data validation, and performance tuning, **ensuring the accuracy, reliability, and efficiency of data pipelines.**
- **Experience with Spark RDD**, Spark Data frames, Spark SQL.
- **Demonstrated expertise in developing complex data transformations** and data enrichment using PySpark.
- **Experience in writing complex SQL queries** using joins, functions, sub queries, correlated sub queries and Regular Expression.
- **Good Experience of core python and advance Python concepts** like List Comprehensions, decorators, generators, OOPs, Functions.
- **Strong understanding of dimensional modeling principles**, enabling the creation of **efficient and scalable data warehouse structures.**
- **Extensive experience working in an Agile development environment.**
- **Strong Team player with Interpersonal and good communication skills.**

SKILLS

- *AWS Cloud* : Amazon Web Service (AWS), EC2, S3, Glue, EMR, Redshift, Lambda, Athena, RDS, SnowFlake, CloudWatch, SNS etc.
- *Versioning tools* : Git, Github.
- *Programming Language* : Python, PySpark, SQL.
- *Big Data Technologies* : Hadoop, Hive, HDFS, MapReduce.
- *Databases* : Oracle, MySql.
- *Data Warehouses* : Redshift, Snowflake, Hive.
- *Database Modeling* : Star Schema, Snowflake Schema, Galaxy Schema.
- *OS* : Linux, Windows.
- *Tools and Software* : VS Code, PyCharm, Jupyter Notebook.
- *Scheduling and Monitoring Tool* : Apache Airflow.
- *File Formats* : Parquet, JSON, CSV.

EXPERIENCE DETAILS

Working as a Software Engineer at ALLSAFE IT SERVICES PVT. LTD. from Aug-2019 to till date.

PROJECT DETAILS

Client : Ono Pharmaceutical

Role : AWS Data Engineer

Description : Ono Pharmaceutical is a Japanese company that focuses on drug discovery in areas of unmet medical needs, such as oncology, immunology and neurology. It has a collaboration with Iktos, a company specialized in artificial intelligence for new drug design.

Roles and Responsibilities:

- **Designing and implementing efficient data migration processes from on-premise servers (such as Oracle, MySql) to AWS S3 using Glue pipelines, ensuring secure and reliable transfer of data.**
- **Configuring and managing Glue pipelines to load refined data from S3 into Redshift, maintaining data consistency and integrity throughout the process.**
- **Building end-to-end data pipelines to extract, transform, and load (ETL) data from various sources into the AWS environment. Implementing data transformation and enrichment logic using AWS services such as AWS Glue, Apache Spark.**
- **Configuring and managing Glue pipelines to load refined data from S3 into Redshift, maintaining data consistency and integrity throughout the process.**
- **Automating the end-to-end data extraction, transformation, and load workflows using Airflow, enabling seamless and reliable execution.**
- **Setting up monitoring and alerting mechanisms to detect and resolve issues with data pipelines, infrastructure, or data quality. Troubleshooting and resolving data-related problems to ensure the integrity and availability of data.**
- **Active participation in daily stand-up meetings (Scrum) and requirement review meetings, collaborating with the team to analyze requirements and provide technical insights.**
- **Responsible for version control and code management using GIT and GitHub, ensuring proper code organization, version tracking, and collaboration within the team.**
- **Documenting the project's data engineering processes, data pipelines, and system architecture.**

Client : GoShop

Role : Data Engineer

Description : GOSHOP is an exciting new way to shop, hassle-free in the convenience of your home 24 hours a day. Shop for the world's most popular brands in beauty, electronics, fashion and more at unbeatable prices with Go Shop in Malaysia. Shop from the comfort of your home or anywhere on the go with Go Shop.

Roles and Responsibilities:

- **Leveraged AWS S3 as a robust storage solution to effectively store and manage large volumes of data, ensuring exceptional availability and durability.**
- **Designed and optimized data processing workflows using Apache Spark, significantly enhancing data processing speed and efficiency.**
- **Conducted comprehensive assessments of end-to-end architecture and implementation for a range of AWS services, including EMR, Redshift, S3 and Lambda.**
- **Debugged and resolved issues by analyzing and sampling the data.**
- **Utilized Redshift for implementing various data warehousing concepts such as star schema, snowflake schema, facts, dimensions, and aggregations.**
- **Demonstrated proficiency in utilizing Git for efficient code management and collaboration, ensuring effective version control.**

Client : Allegiant

Role : Data Engineer

Description : From America's favorite small cities to world-class destinations, Allegiant makes leisure travel affordable and convenient. With low-low fares, nonstop, all-jet service and premier travel partners, Allegiant provides a complete travel experience with great value and without all the hassle. The Company also has authority for charter service to Canada and Mexico.

Roles and Responsibilities:

- **Design and optimize database schemas, tables, and relationships to ensure efficient data storage and retrieval. Develop and execute SQL queries to extract, transform, and analyze data from databases.**
- **PySpark to develop data processing workflows, utilizing concepts such as RDDs (Resilient Distributed Datasets) and DataFrame APIs.**
- **Utilized Python programming skills to develop efficient and scalable data manipulation and analysis scripts, leveraging libraries.**
- **Developed optimized SQL queries and performed query tuning techniques to improve performance and efficiency of data retrieval and analysis processes.**
- **Demonstrated strong problem-solving skills by debugging and resolving issues encountered during data processing, leveraging knowledge of PySpark, SQL, and Python.**

EDUCATION

Bachelor of Science

Completed B.Sc. from Nagpur University.