

Suma U Data Engineer

✉ suma@jobemails.net ☎ (913) 909 7785 📍 Frisco, TX 🌐 LinkedIn

Profile

As a Data Engineer, I have 4+ years of experience in designing, implementing, and maintaining robust data infrastructure and pipelines. Proficient in Python, SQL, and Spark and Machine Learning, I specialize in ETL/ELT processes, leveraging tools like Informatica, Talend, Apache Spark and Hadoop for distributed data processing and analysis and experience creating dashboards for data analysis in BI tools like Tableau and Alteryx.

Skills

Programming Languages: Python, R, Java, SQL, Unix/Shell Scripting

Big data Tools: Hadoop architecture, HDFS, MapReduce, Hive, Sqoop, Oozie, EMR, Spark, Pig, lambda functions

Frameworks & Tools: NumPy, Scikit-learn, Pandas, Seaborn, Power BI, Tableau, MS Excel, MS Access, PySpark, Airflow, ETL, JIRA

Datastores and Cloud: Amazon Web Services (AWS), Microsoft Azure, MySQL, Oracle, MongoDB, PostgreSQL, Apache Kafka, Snowflake, Data Bricks, Scrum, Confluence, Data Warehouse

Professional Experience

Data Engineer, S&P Global 01/2023 – present | Remote, USA

- Designed and implemented ETL pipelines leveraging AWS Glue and AWS Data Pipeline for seamless data extraction, transformation, and loading from diverse sources including SQL databases and RESTful APIs.
- Automated data workflows using Apache Airflow DAGs on AWS, achieving a 50% reduction in manual data ingestion effort, and utilized AWS S3 for efficient storage and management of raw and processed data.
- Optimized data warehouse performance by designing schemas and implementing queries on Snowflake and AWS Redshift, ensuring optimal data query performance for analytics and reporting needs.
- Implemented data preprocessing and cleansing workflows using AWS Lambda functions and Python libraries(Pandas,NumPy), reducing data anomalies by 25% and enhancing data quality.
- Developed predictive models with Scikit-Learn and TensorFlow on AWS SageMaker, leveraging AWS services for scalable model training and deployment within cloud environments.

Data Engineer, Lyve Tech LLC 08/2022 – 12/2022 | Ohio, USA

- Designed and implemented robust data integration solutions on Microsoft Azure, leveraging Azure Data Factory for orchestration of complex data workflows and pipelines.
- Engineered and optimized ETL/ELT workflows using Python, Apache Spark, and Azure Functions, processing over 500GB of data daily with a 99% accuracy rate, ensuring seamless data ingestion and transformation.
- Architected scalable enterprise data platform on Azure using Azure Synapse Analytics for high-performance data warehousing and Azure HDInsight for distributed processing, enabling real-time analytics and insights.
- Implemented Azure Blob Storage and Azure Data Lake Storage for secure and scalable storage of diverse data types, implementing advanced data management strategies and access controls.
- Deployed Azure Stream Analytics for real-time data processing and utilized Azure Databricks for advanced analytics and machine learning, driving cutting-edge data processing capabilities and supporting data-driven decision-making.

Senior Data Engineer, Infosys 10/2020 – 08/2021 | Hyderabad, India

- Engineered ETL data pipelines on AWS using Glue Crawler, Glue Job, PySpark, and Spark SQL, transforming raw data into structured formats stored in the Glue Data Catalog.
- Implemented Apache Airflow with AWS CloudFormation to schedule and automate daily execution of ETL workflows, resulting in a 50% reduction in manual effort and significant cost savings.
- Utilized Athena for data analysis, enabling actionable insights from datasets, and established CI/CD pipelines for automated ETL job testing, deployment, and monitoring to ensure data reliability and accuracy.
- Optimized data warehouse and data mart integration, enhancing data integrity and reliability for improved decision-making.

Data Engineer, Infosys 09/2018 – 10/2020 | Hyderabad, India



- Managed 20 Banking & Financial data governance projects in Agile with a team of 8. Deployed data models, Oracle SQL/PL/SQL procedures, Unix Shell Scripts, and Informatica mappings for data integrity, security, and compliance in finance sector.
- Designed data models and databases using Erwin to support automated processes for OLTP and OLAP systems, integrating BMC Control-M to create 180 new jobs and enhance operational efficiency.
- Addressed 35 defect reports over the course of 6 months, implementing timely fixes and improvements to the data governance processes. This effort led to a 68% improvement in overall efficiency.

Education

Masters in Computer Science, University of Central Missouri 08/2021 – 12/2022 | Kansas, USA

Bachelors of Technology in Electronics and Communication, Gokaraju Rangaraju Institute of Engineering and Technology 09/2014 – 05/2018 | Hyderabad, India

Certificates

- MTA: Introduction to Programming Using Python - Certified 2020 
- Microsoft Certified: Azure Fundamentals 
- AWS Certified Solutions Architect – Associate 