RUPENDRA SAI LANKE
DATA ENGINEER
rupendra.l@mycvtalent.com | 9725565924 | Denton, TX

SUMMARY

- 4 years of experience as a Data Engineer building robust data pipelines for large-scale structured and unstructured datasets. Expertise in data acquisition, validation, modeling (predictive, statistical, and data), and visualization, driving actionable insights
- Ability to develop efficient dimensional models in Snowflake, optimizing query performance for various data analysis scenarios.
- Processed and transformed large datasets (petabytes/terabytes) on Databricks using Spark SQL and Spark DataFrames, enabling advanced data analytics for projects with users.
- Experience in RDBMS concepts, Data Modeling (Facts and Dimensions, Star/Snowflake schemes), Data Migration, Data Cleansing and ETL Processes.
- Engineered enterprise solutions by employing batch processing with DataBricks and integrating streaming frameworks, including Spark Streaming, Apache Kafka, and Apache Airflow.
- Excellent knowledge of AWS cloud services, including EC2, S3 Bucket, Amazon Redshift, Glue, Lambda, and Athena, with expertise in infrastructure management, storage, data warehousing, serverless computing, and automated deployment.

EXPERIENCE

Allstate, TX ***Jan 2024 – Current***

Data Engineer

- Developed reusable Spark libraries for common data processing tasks, leading to a decrease in development time for future pipelines.
- Established data lineage tracking within Databricks to trace data origin and transformations, improving data governance and auditability.
- Improved data decoupling by decoupling data producers and consumers using Kafka, allowing for independent development and scalability of data pipelines.
- Automated data ingestion and transformation pipelines using AWS Glue and AWS Pipeline, reducing manual effort and improving data processing speed.
- Implement Airflow pipelines that orchestrate batch processing of financial data at specific intervals for regulatory reporting, ensuring adherence to reporting deadlines.
- Build and maintain various dashboards and reports to monitor data quality, usage, and performance metrics using Tableau.

Genius SoftTech, India ***Jan 2019 – July 2022***

Data Engineer

- Designed and implemented highly scalable data pipelines on Databricks using PySpark, ingesting terabytes of data per day from various sources (relational databases, log files, APIs).
- Developed and maintained reusable Airflow DAGs (Directed Acyclic Graphs) for various data processing tasks, promoting code maintainability and scalability.
- Created a Power BI dashboard to visualize key business KPIs, resulting in a weekly time savings of 2 hours on manual reporting.
- Employed a robust data streaming architecture using Kafka, efficiently handling the ingestion and distribution of real-time data across diverse applications.
- Established serverless functions using Lambda to trigger data processing tasks upon specific events in S3, achieving 20% cost savings compared to traditional compute instances.
- Utilized Amazon Redshift for advanced analytics, enabling the implementation of complex analytical queries and providing valuable insights for strategic decision-making.
- Collaborated closely with data scientists and analysts to understand their data requirements and ensure the availability of high-quality, relevant datasets.
- Implemented ETL workflows using Apache Spark and Python (Pandas, NumPy, BeautifulSoup), resulting in a 30% reduction in data processing time and improved data accuracy.

SKILLS

Programming Language:	Python, R, SQL
IDE's:	PyCharm, Jupyter Notebook
Big Data Ecosystem:	Hadoop, Hive, Apache Airflow, Apache Kafka, Apache Spark, Apache Flink, DataBricks
Cloud Technologies:	AWS (EC2, S3, RDS, Lambda, Glue, Athena, AWS Pipeline, Redshift)
Visualizations:	Tableau, Power BI, Excel
Packages & Data Processing:	NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PySpark, Data Pipelines, Jenkins
Version Control & Database:	GitHub, Git, SQL Server, PostgreSQL, MongoDB, DynamoDB, MySQL, Snowflake
Operating Systems:	Windows, MacOS

EDUCATION

Master of Science in **Advanced Data Analytics** ***May 2024***

University of North Texas, Texas