SRAVYA PAMULA

Data Engineer

📞 +14702830393   @ psravya2019@gmail.com   🔗 **https://www.linkedin.com/in/sravyapamula/**

📍 **Columbia, MD, 21045**

EXPERIENCE

Data Engineer
Humana

📅 06/2023 - 09/2024   📍 Remote

- Implemented data integration processes to consolidate data from various sources, including member records, claims data, EHR
- Developed Spark-based ETL pipelines to process and analyze 3-5 GB of physicians, claims, accounts, payments data daily
- Ensured integration of structured and unstructured data sources such as CSV, JSON, and Parquet formats
- Processed large datasets in a parallelized environment using Spark, resulting in a 20% reduction in data processing time and accelerating insights delivery by 15%
- Integrated AWS S3 with Snowflake to seamlessly transfer over 5 TB of processed data into the data warehouse
- Managed Snowflake database objects such as Streams, Tasks, Stages, Views, Stored Procedures, and Snow pipes, enabling bulk and continuous data integration with zero data loss
- Optimized SQL queries, reducing execution time by 25% and improving data retrieval performance

Data Engineer
Next phase solutions and services

📅 05/2022 - 05/2023   📍 Columbia, Maryland

- Built scalable data pipelines using AWS Glue, S3, and Spark to process over 10 TB of data daily, improving data processing capacity
- Implemented machine learning models using PySpark for fraud detection, improving detection accuracy by 20% and reducing fraudulent claims processing by 15%
- Collaborated with data science teams to deploy predictive models that improved customer retention and policy lapse predictions by 15%, leveraging MLlib for feature engineering and model training
- Integrated Databricks for advanced analytics and large-scale data processing using Python and Spark
- Participated in Agile sprints and tracked progress in JIRA, which helped ensure timely delivery of ETL tasks and transparent communication with project managers and stakeholders

Data Engineer
Cyient technologies

📅 07/2019 - 08/2021   📍 Pune, India

- Automated workflows using Informatica and SSIS, cutting ETL time by 30%
- Optimized complex SQL queries across PostgreSQL and SQL Server databases, reducing execution time by 25%
- Extracted data from various source systems, such as databases, flat files, APIs, and other structured or unstructured sources
- Built OLAP and OLTP systems with relational and multidimensional databases
- Created SSIS packages using data transformation tasks like Lookups, Derived Column, Conditional Splits, Error Handlers
- Leveraged Linux command-line tools to optimize data processing jobs, contributing to 10% faster ETL operations
- Worked in Agile project teams, using JIRA to track tasks and ensure alignment on project timelines and deliverables
- Built Tableau dashboards to present weekly data insights on ETL performance and reporting metrics, making data actionable for business teams

EDUCATION

Master's in data science
University of Maryland Baltimore County

📅 Date period   📍 Baltimore, MD

- Relevant skills: Power BI, Excel, NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, ML, PySpark, Time series analysis, Kafka, Airflow
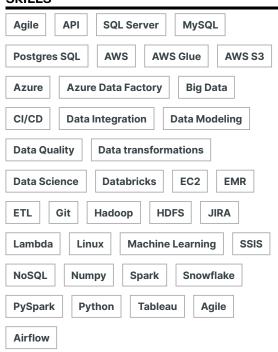
BS-MS Physics
Indian Institute of Science Education and Research

📅 Date period   📍 Pune, India

SUMMARY

Over 4 years of experience as a Data Engineer with a strong technical background. Skilled in Python, SQL, Big Data technologies (Hadoop, Spark), and relational (RDBMS) & NoSQL databases. Good experience in developing and optimizing ETL data pipelines, SQL queries, and Spark jobs. Hands-on experience with cloud platforms including AWS, Azure, Snowflake, and Databricks Skilled in data ingestion, data modeling, data transformation and data migration processes. Good experience in performance tuning, production support, and managing data quality checks.Proven ability to collaborate with cross-functional teams, demonstrating strong interpersonal skills.

SKILLS

`Agile` `API` `SQL Server` `MySQL`

`Postgres SQL` `AWS` `AWS Glue` `AWS S3`

`Azure` `Azure Data Factory` `Big Data`

`CI/CD` `Data Integration` `Data Modeling`

`Data Quality` `Data transformations`

`Data Science` `Databricks` `EC2` `EMR`

`ETL` `Git` `Hadoop` `HDFS` `JIRA`

`Lambda` `Linux` `Machine Learning` `SSIS`

`NoSQL` `Numpy` `Spark` `Snowflake`

`PySpark` `Python` `Tableau` `Agile`

`Airflow`

FIND ME ONLINE

🌐 **Linkedin**
https://www.linkedin.com/in/sravyapamula/

🌐 **Github**
https://github.com/Pam2020

🌐 **Certifications**
https://credentials.databricks.com/1053116b-2aa8-46cc-9568-322284ec8ce9