Ramya Sri Sonar

Seattle, WA | 703-717-3087 | ramyasonar7@gmail.com | www.linkedin.com/in/rssonar

Technical Skills

Big Data: Apache Spark, Apache Kafka, Apache Airflow, PySpark, Databricks, Hive, Hadoop, Snowflake, Teradata
Languages: Python, SQL
Cloud Services: Amazon Web Services (AWS) - S3, EC2, EMR, IAM roles and permissions, Redshift, Glue, Lambda
Libraries: Pandas, NumPy, Matplotlib, Plotly, Keras, TensorFlow, Spark MLlib, SQLAlchemy, Beautiful Soup, Boto3
Miscellaneous: Data Modelling, Data warehousing, MS Excel, ETL, Linux, MS Excel, Jira, Git, Tableau, Visual studio, Pycharm

Work Experience

Data Engineer| Nike June 2021 – Present

- Operationalizing data for GSM (Global Sourcing and Manufacturing), a team focused on the management of the quality, sustainability, cost, and profitability of Nike's factory sourcing activities
- Implemented end-to-end data pipelines with **PySpark, HQL**, **Amazon Web Services (S3, EMR)** and **Apache Airflow** for seamless data processing.
- Engineered Spark applications with optimal design principles, achieving a 30% increase in EMR cluster computing resource efficiency.
- Automated **ETL** workflows with Airflow to load scheduled batch and history data to **Hive** and **Snowflake** data warehouses, resulting in 70% reduction in manual interventions.
- Improved Snowflake views retrieval time from 4 hours to 2 minutes with Spark.
- Proficient in exploiting JIRA, an agile project management tool, to streamline development processes and track project progress.
- Collaborated with cross-functional teams to understand data needs and deliver timely solutions.
- Performed data inquiries to resolve issues encountered within the data pipelines

Environment: Python, AWS S3, EMR, Spark, PySpark, Apache Airflow, Databricks, Snowflake, Teradata. Git, Hive, SQL, Tableau

Data Engineer | Apprenticeship | Accenture Federal Services January 2021 – May 2021

- Improved the efficiency of the planning and procurement process by leveraging Machine Learning, Deep learning, Big Data, and Amazon Web Services (AWS).
- Gathered and processed diverse data sets in formats such as CSV, XLSX, and TXT, totaling approximately 2.5 million records, collaborating closely with stakeholders to ensure data quality and accuracy.
- Conducted preprocessing of the dataset, involving the removal of special characters and digits, tokenization, correction of inconsistent spellings, lemmatization, stop word removal, and stemming.
- Attained 80% accuracy using LSTM Deep Learning to automate asset classification in Supply Chain procurement.
- Achieved 92.7% optimization, 32% classification performance boost by developing novel NLP algorithm for asset grouping.
- Created scalable and portable REST API and GUI using Python on AWS with Flask Framework.

Environment: Python, Pyspark, AWS EMR, S3, Redshift, Machine learning models (K Nearest Neighbor,Random forest model), Deep Learning models (Recurrent neural network (RNN), Long short-term memory (LSTM))

Graduate Teaching Assistant | George Mason University August 2020 – May 2021

- Delivered tutoring and teaching activities for graduate-level Big Data course - Hadoop, Spark, and data visualization tools.
- Tasks encompassed grading, mentoring, and supervising homework/assignments for 60+ students.

Environment: Hadoop, Spark, Python, R, Tableau, AWS, SQL, NoSQL

Education

George Mason University May 2021
Masters of Science, Data Analytics Engineering
Jawaharlal Nehru Institute of Technology May 2019
Bachelors of Technology, Computer Science Engineering