Sai Sridhar Reddy Palli
Data Engineer
+1 716-907-4297 | sridhar.p@mycvtalent.com | New York | LinkedIn | Portfolio

SUMMARY

- **Data Engineer** with **around 4 years** of experience designing data-intensive applications leveraging **Cloud Data Engineering, Data Warehousing, Hadoop Ecosystem, Big Data Analytics, Data Visualization, and Data Quality solutions.**
- Expert in developing interactive dashboards using **AWS QuickSight, Power BI, and Tableau**, visualizing key performance indicators (KPIs) from **Amazon Redshift** data for business stakeholders.
- Leveraged **Spark SQL** and **PySpark APIs** to perform large-scale data analytics on distributed datasets, achieving faster processing compared to traditional **MapReduce jobs**.
- Implementing **Airflow DAGs** (Directed Acyclic Graphs) with custom operators and error handling mechanisms, ensuring data pipeline robustness and easy troubleshooting.
- Ability to use **Databricks** cloud platform to deploy and manage **Spark applications**, ensuring scalability, elasticity, and cost-effectiveness for big data workloads.

EDUCATION

Master of Science in Computer Science | University at Buffalo, NY **Dec 2023**
Bachelor of Technology in Computer Science | Sastra Deemed University, Thanjavur, India **Jun 2021**

SKILLS

Programming Language:	Python, R, SQL, Scala
IDE's:	PyCharm, Jupyter Notebook
Big Data Ecosystem:	Hadoop, HDFS, MapReduce, HBase, Sqoop, Apache Airflow, Apache Kafka, Apache Spark, Flink, DataBricks
Cloud Technologies:	AWS (EC2, S3, RDS, Lambda, Glue, Athena, AWS Pipeline, Redshift), Azure (Azure Data Factory, Data Lake, Blob Storage, Azure DevOps, Databricks)
Visualizations:	Tableau, Power BI, Excel, Google Looker Studio
Packages & Data Processing:	NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, Plotly, PySpark, Data Pipelines, Jenkins
Version Control & Database:	GitHub, Git, SQL Server, PostgreSQL, MongoDB, DynamoDB, MySQL, Snowflake

EXPERIENCE

Principal Financial, NY | Data Engineer **Sep 2023 – Current**

- Led the development and optimization of **ETL pipelines** using **Python, SQL, Apache Spark**, and **Hadoop** to ensure high-fidelity migration of critical financial data with minimal disruption.
- Integrate **Kafka** for real-time data streaming, achieving a measurable reduction in data latency and improving the processing speed of customer transactions.
- Leveraged **AWS Glue** to automate **ETL processes** for daily sales data feeds, reducing processing time by approximately 50%.
- Used **Airflow** monitoring and alerting capabilities to identify and troubleshoot data pipeline issues proactively, minimizing data downtime and maximizing data quality.
- Designed and implemented serverless data pipelines using **AWS Lambda** to process real-time data streams, achieving a 20% reduction in processing latency compared to traditional batch processing methods.

Kpit Technologies, India | Data Engineer **Aug 2019 – July 2022**

- Accomplished interactive **Power BI reports** leveraging advanced **DAX** and **Power Query** modeling techniques to unlock deeper data insights and facilitate informed decision-making.
- Leveraged **Snowflake's** cloud-based architecture to facilitate collaboration between data analysts and data scientists, leading to a 30% increase in data-driven project completion rate.
- Developed and maintained **HiveQL queries** to analyze large datasets stored in **HDFS**, enabling efficient data exploration and reporting.
- Established and deployed a high-performing **ETL pipeline** in **Databricks** to process terabytes of data daily, enabling efficient data integration and transformation for the data warehouse.
- Automated **ETL pipelines** using **Azure Data Factory** to ingest, transform, and load data across various sources and destinations.
- Configured and managed **Azure DevOps** for continuous integration and continuous delivery **(CI/CD)** of data pipelines, ensuring efficient deployment and version control.
- Designed and implemented an end-to-end data pipeline using **Azure Data Factory (ADF)** to orchestrate data movement from various sources **(databases, APIs) to ADL**.
- Integrated **Apache Kafka** with **Airflow** for real-time data ingestion from streaming sources, enabling near-real-time data analysis and decision-making.

CERTIFICATIONS

- AWS Solution Architect Associate
- AWS Cloud Practitioner