# Sai Sridhar Reddy Palli
## Data Engineer
**+1 716-907-4297 | sridhar.p@mycvtalent.com | Buffalo, NY | [LinkedIn](#)**

## SUMMARY

- Data Engineer with 3+ years of experience designing data-intensive applications leveraging Cloud Data Engineering, Data Warehousing, Hadoop Ecosystem, Big Data Analytics, Data Visualization, and Data Quality solutions.
- Leveraged Spark SQL and PySpark APIs to perform large-scale data analytics on distributed datasets, achieving faster processing compared to traditional MapReduce jobs.
- Ability to Databricks cloud platform to deploy and manage Spark applications, ensuring scalability, elasticity, and cost-effectiveness for big data workloads.
- Expert in developing interactive dashboards using AWS QuickSight, Power BI, and Tableau, visualizing key performance indicators (KPIs) from Amazon Redshift data for business stakeholders.
- Experience in Implementing Airflow DAGs (Directed Acyclic Graphs) with custom operators and error handling mechanisms, ensuring data pipeline robustness and easy troubleshooting.

## 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) |
| **Visualizations:** | Tableau, Power BI, Excel |
| **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.
- Used Airflow monitoring and alerting capabilities to proactively identify and troubleshoot data pipeline issues, minimizing data downtime and maximizing data quality.
- Integrate Kafka for real-time data streaming, achieving a measurable reduction in data latency and improving the processing speed of customer transactions.
- Accomplished interactive Power BI reports leveraging advanced DAX and Power Query modeling techniques to unlock deeper data insights and facilitate informed decision-making.
- Designed and implemented serverless data pipelines using AWS Lambda to process real-time data streams, achieving a 10% reduction in processing latency compared to traditional batch processing methods.

**Kpit Technologies, India | Data Engineer**                                                   **Jan 2020 – July 2022**

- Developed and maintained HiveQL queries to analyze large datasets stored in HDFS, enabling efficient data exploration and reporting.
- Utilized Databricks platform to build and deploy scalable Spark applications in the cloud, reducing infrastructure management overhead by 40%.
- Implemented AWS Glue to automate data catalog creation and schema management for our data lake on Snowflake, ensuring data consistency and discoverability for data consumers.
- 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.
- Empowered data analysts with self-service capabilities by setting up Amazon Athena for interactive querying of data stored in S3, reducing reliance on data engineers for ad-hoc analysis.
- Integrated Apache Kafka with Airflow for real-time data ingestion from streaming sources, enabling near-real-time data analysis and decision-making.
- Improved Power BI report performance through efficient data modeling techniques, reducing query execution time and improving user experience.
- Developed 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.

## CERTIFICATIONS

- AWS Solution Architect Associate
- AWS Cloud Practitioner