VIGNESH CHANDRA VAJJA
DATA ENGINEER
7162596302 | vignesh.c@mycvtalent.com | Buffalo, NY

SUMMARY

- 3+ years of experience as a Data Engineer with a proven track record of designing and building scalable data solutions. Skilled in leveraging Cloud Data Engineering tools (Azure, AWS) to deliver Data Warehouses, Hadoop ecosystems, Big Data analytics pipelines, and Data Visualization and Reporting solutions.
- Proficient in developing reusable Airflow DAGs (Directed Acyclic Graphs) for various data processing tasks, promoting code maintainability and reusability.
- Experience in creating interactive notebooks on Databricks for data exploration and analysis, enabling faster data discovery and decision support.
- Expert in developing interactive dashboards using AWS QuickSight, Power BI, and Tableau, visualizing key performance indicators (KPIs) from Amazon Redshift data for business stakeholders.

SKILLS

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

EDUCATION

Masters of Professional Studies in Data Science and Applications **Jan 2024**
University at Buffalo, Buffalo, New York

Bachelor of Technology in Electronics and Communication Engineering **May 2022**
RVR & JC College of Engineering, Andhra Pradesh, India

EXPERIENCE

JPMorgan Chase & Co., NY **Sep 2023 – Current**
Data Engineer

- Enhanced data ingestion pipelines using Sqoop, successfully transferring petabytes of data from relational databases to HDFS.
- Built and orchestrated complex data pipelines with Airflow, scheduling and automating daily data processing tasks for 2x reduction in manual intervention.
- Orchestrated data pipelines using AWS Pipeline, ensuring reliable and scheduled data movement between various AWS services.
- 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.
- Accomplished interactive Power BI reports leveraging advanced DAX and Power Query modeling techniques to unlock deeper data insights and facilitate informed decision-making.
- Automated data extraction, transformation, and loading (ETL) pipelines using AWS Glue, reducing manual effort by 50%.
- Designed and configured Kafka producers and consumers to handle high-volume data streams efficiently with minimal data loss.

Dixon Technology, India **Jan 2020 - July 2022**
Data Engineer

- Developed Spark SQL queries to perform complex data transformations and aggregations, reducing data processing time by 25%.
- Utilized Databricks platform to build and deploy scalable Spark applications in the cloud, reducing infrastructure management overhead by 40%.
- Established Sqoop scripts with incremental loading capabilities, reducing data redundancy and improving data pipeline efficiency.
- Optimized PySpark code using techniques like partitioning and caching, achieving a 25% reduction in processing time for complex data transformations.
- Designed and implemented MapReduce jobs to process large datasets (terabytes of log data) for data cleansing, transformation, and aggregation tasks.
- Employed Snowflake clustering and partitioning techniques, reducing data storage costs by 20%.
- Automated data pipeline deployments using Azure DevOps pipelines, achieving continuous integration and continuous delivery (CI/CD) for data pipelines.
- Migrated data pipelines from legacy ETL tools to ADF, reducing development time and improving data quality through robust error handling.
- Optimized Databricks cluster configurations to achieve 30% faster data processing times for computationally intensive tasks.

CERTIFICATIONS

- AWS Cloud Practitioner | AWS Developer Associate