

# VIGNESH CHANDRA VAJJA

7162596302 | [vajjavignesh@gmail.com](mailto:vajjavignesh@gmail.com) | [linkedin.com/in/vigneshvajja/](https://www.linkedin.com/in/vigneshvajja/) | [github.com/vignesh6666](https://github.com/vignesh6666)

## SUMMARY

Experienced Data Engineer with a Master's in Data Science and over 3+ years of experience, specializing in healthcare and Finance domains. Proactive problem solver skilled in ensuring compliance, security, and efficiency, collaborating with cross functional teams to address real-world healthcare and financial data management challenges.

## WORK EXPERIENCE

### Data Engineer, University at Buffalo, New York, USA

Jan 2024 - Present

- Orchestrated migration of **100,000+ healthcare** records from on-premises **SQL Server** to **Snowflake** using **Azure Data Factory**, enhancing data analytics capabilities with 20% faster data retrieval, resulting in **20%** faster data retrieval.
- Developed and Enhanced ETL pipelines using **Python**, **SQL**, **Apache Spark**, and **Hadoop**, ensuring high data quality and minimal downtime while migrating **critical** healthcare data.
- Leveraged **Azure Databricks** for scalable data processing and analytics tasks, honing performance.
- Created **Power BI** reports with advanced **DAX** and **Power Query** modeling, facilitating insightful visualization.

### Data Engineer, Vaaraahi Tech Solutions, Hyderabad, India

Mar 2020 - Jul 2022

- Led finance data transformations with **AWS Glue**, **PySpark**, **Spark SQL**, and **Kafka**, achieving a 30% efficiency boost. Integrated **AWS Step Functions** for streamlined workflow management.
- Integrated **AWS Athena** and **EMR** for enhanced data querying boosting data analytics efficiency by **25%**.
- Utilized **T-SQL** extensively for database development, including data manipulation, querying, and stored procedures, enhancing data retrieval and analysis capabilities.
- Managed security with **AWS IAM** and **AWS Key Management Service**, reducing unauthorized access by 30%.
- Leveraged **AWS Glue notebooks** for data processing, implementing advanced data modeling techniques such as Star and Snowflake schemas to enhance the structure and analysis of financial data.
- Implemented **Docker** containers to containerize data processing and deployment pipelines, enhancing portability, scalability, and consistency across environments.
- Implemented **CI/CD** pipelines using GitHub Actions and AWS CodePipeline for automated testing and deployment.

## PROJECTS

### Scalable Data Warehouse for Financial Analytics

- Led the design and deployment of a cloud-based data warehouse solution using Amazon Redshift. Created ETL pipelines leveraging Python and Apache Airflow to ingest and transform data from Alpha Vantage API. Improved query performance by 40% and reduced data processing time by 60%.

### Real-time Data Pipeline for Weather Prediction System

- Implemented a real-time data pipeline using Apache Kafka, AWS EC2, S3, AWS Glue, Crawler, and AWS Athena. Developed data ingestion scripts to collect data from Open Weather API, loaded data using Apache Kafka into AWS S3, and then crawled data into Amazon Athena. Utilized Spark MLlib to make predictions for the weather.

## RESEARCH PAPER and CERTIFICATIONS

### An Efficient Framework for Load Balancing Map Reduce Algorithm for Big Data

- This paper suggests a method for managing load balancing that involves only sending the name node to data nodes under the lightest load. [ieeexplore.ieee.org/document/9792840](https://ieeexplore.ieee.org/document/9792840)
- AWS Certified Developer - Associate** and **AWS Certified Cloud Practitioner**.

## EDUCATION

### University at Buffalo, The State University of New York

**Masters in Data Science and Applications** CGPA- 3.8/4.0

Buffalo, NY, USA

### R.V.R & J.C College of Engineering

**Bachelors in Computer Science Engineering** CGPA- 3.3/4.0

Guntur, India

## TECHNICAL SKILLS

**Programming Languages:** Python, Scala, R, C, SQL, HTML, CSS, JavaScript.

**Database and Frameworks:** MySQL, SQLite, Postgresql, Snowflake, Cassandra, Flask, Django, Fast API.

**Data Engineer Tools:** Hadoop, HDFS, YARN, Map Reduce, Spark, Sqoop, Flink, Airflow, Kafka, H-Base, PIG.

**Data Science:** TensorFlow, Keras, PyTorch, Scikit-learn, Matplotlib, Plotly, Pandas, NLP(NLTK, Spacy, BERT).

**Cloud Services:** AWS(EC2, Lambda, S3, DynamoDB, RedShift, Athena), Microsoft Azure, Google Cloud Platform.

**Developer & BI Tools:** Power BI, Tableau, Google Looker Studio, VS Code, MS Excel, Docker, Kubernetes.