**SRI RAM VUPPALA  
 Data Engineer**

**Location: NY** | **Ph.** **929-379-7956** |**Email:** [sriramgowd3@gmail.com](mailto:sriramgowd3@gmail.com)

# SUMMARY

* + - 3 years of Data Engineering experience with expertise in Big Data, Hadoop Ecosystem, Analytics, Cloud Engineering, and Data Warehousing.
    - Expertise in building large-scale applications within the Big Data ecosystem, including Hadoop, Spark, Kafka, Hive, Pig, Airflow, and more.
    - In-depth understanding of Hadoop architecture and its components, including HDFS, MapReduce, and Spark.
    - Expertise in writing Hadoop jobs using MapReduce, Hive, Pig, and Splunk
    - Development of production-ready Spark applications using Spark SQL, Matplotlib, GraphX, Data Frames, Datasets, Spark-ML, and Spark Streaming.
    - Experience in developing end-to-end ETL data pipelines using Spark and Configuration of Spark Streaming for real-time data processing from Apache Kafka.
    - Proficiency in developing highly scalable and resilient Restful APIs, ETL solutions, and third-party integrations using Informatica.
    - Proficient in scripting with Python, Java, Scala, and Spark-SQL for data aggregation from various file formats.
    - Hands-on experience with data-centric Python and R packages such as pandas, ggplot2, NumPy, and Seaborn.
    - Proficient in implementing machine learning models and algorithms using Python's Scikit-learn library.
    - Connected seamlessly with databases using Python, streamlining data extraction, and querying processes.
    - Implementation of CRUD operations in MongoDB for data analysis.
    - Extensive experience with AWS services like EC2, S3, EMR, RDS, VPC, Elastic Load Balancing, IAM, Auto Scaling, CloudFront, CloudWatch, SNS, SES, SQS, and Lambda.
    - Developed comprehensive reports and visualizations using tools like Power BI, Power Pivot, and Power Map.
    - Expertise in crafting powerful visualizations and dashboards using tools like Tableau and Matplotlib.
    - Additional expertise in areas like Natural Language Processing (NLP), Image Detection, and MapReduce.

**SKILL&INTERESTS**

| Big Data Technologies: | Apache Hadoop 2.x/1.x, Cloudera CDP, Hortonworks HDP, HDFS, Yarn, MapReduce, Pig, Hive, Oozie, Apache Spark, |
| --- | --- |
| Programming Languages: | Python, C, SQL, HiveQL, Scala |
| Cloud Technologies | Amazon Web Services (EMR, EC2, EBS, RDS, S3, Glue, Elasticsearch, Lambda, Kinesis, SQS, DynamoDB, Redshift, ECS) Microsoft Azure (Databricks, Data Lake, Blob Storage, Data Factory, SQL DB, SQL DWH, Cosmos DB, Azure DevOps, Active Directory) |
| Databases | MS-SQL Server, Oracle, Cassandra, MySQL, PostgreSQL, MongoDB, DB2 |
| Visualization Tools: | Tableau, Power BI, Microsoft Excel |
| Machine Learning and AI | Recommendation systems, GANs, NLP, Regression, Classification, Random Forest, KNN, Clustering, Decision Trees, CNN |
| Operating Systems | Windows, Macintosh, Linux, Ubuntu, Unix |
| Methodologies | Agile/Scrum, Waterfall |
| Tools and IDE | Snowflake, Apache Airflow, Kafka, Docker, Kubernetes, Git, PyCharm, VS Code, Jupyter Notebook |

# EDUCATION

**New York University New York City, NY**

*M.S Computer Science Sep 2021 – May 2023*

**Anna University Chennai, India**

*B.E Computer Science Aug 2017 – Jun 2021*

# PROFESSIONAL EXPERIENCE

**JPMC, NY**

**Data Engineer Jan 2023 – Present**

* + - Utilized Azure cloud services, including HDInsight, Databricks, DataLake, Blob, Data Factory, Synapse, SQL DB, and SQL DWH.
    - Conducted data cleansing and applied transformations through Databricks and Spark data analysis.
    - Worked on query optimization using best practices and right parameters and using Hadoop, Python,and PySpark.
    - Automated workflows for daily incremental data loads, seamlessly transitioning data from traditional databases systems to data lakes.
    - Worked extensively with Azure Synapse analytics service, which combines enterprise data warehousing and Big Data analytics.
    - Created database objects like tables, views, stored procedures, triggers, packages, and functions using T-SQL for efficient data management and structure.
    - Developed SQL scripts for automation purposes.
    - Solid understanding of Data Ingestion, Airflow Operators for data orchestration, and other related Python libraries.
    - Analyzed SQL scripts and devised solutions for implementation using PySpark and Expertise in developing ETL processes using Spark.
    - Leveraged Databricks notebooks for interactive analytics using Spark APIs.
    - Contributed to building an Enterprise Data Lake with Data Factory and Blob storage, facilitating complex scenarios and machine learning solutions for other teams.

**Molina HealthCare, NY**

**Data Engineer Intern Jun 2022- Sep 2022**

* + - Developed and maintained scalable data pipelines and built out new API integrations to support continuing increases in data volume and complexity.
    - Collaborated with analytics and business teams to improve data models that feed business intelligence tools, increasing data accessibility and fostering data-driven decision-making across the organization.
    - Used the Spark API to analyze Hive data in conjunction with the EMR Hadoop Cluster.
    - Collaborated with analytics and business teams to enhance data models, enabling improved data accessibility and fostering data-driven decision-making throughout the organization.
    - Designed, developed, and maintained Tableau functional reports based on user requirements, and optimized several complex SQL queries for better performance and efficiency.
    - Generated analytical reports with recommendations for management to enable strategic planning, using ETL.
    - Established and monitored a scalable distributed system based on Hadoop HDFS, improving system performance by 25%
    - Utilized PowerShell scripting to automate and streamline data pipelines, improving efficiency and reducing manual effort in data processing tasks.
    - Worked closely with all business units and engineering teams to develop a strategy for long-term data platform architecture.

**Neebal Technologies, India**

**Data Engineer Oct 2019 – Jul 2021**

* + - Developed batch processing applications requiring functional pipelining using Spark APIs.
    - Created a data pipeline and conducted analytics using the AWS stack, including EMR, EC2, S3, RDS, Lambda, Glue, and Redshift.
    - Collaborated with client teams to automate data transformation and incorporate algorithms and models.
    - Leveraged Spark's in-memory capabilities to handle large datasets in the S3 Data Lake. Loaded, filtered, and integrated data into Hive external tables.
    - Extensive hands-on experience in crafting and modifying SQL stored procedures, functions, views, indexes, and triggers.
    - Execute ETL operations on terabytes of data using Python, Spark SQL, S3, and Redshift to extract valuable customer insight
    - Utilized Python programming skills to construct robust data pipelines and dynamic systems.
    - Profound understanding of various AWS services, including S3, EC2, IAM, RDS, and experience with orchestration and data pipeline tools like AWS Step Functions, Data Pipeline, and Glue.
    - Integrated data from diverse sources while ensuring data quality and accessibility standards.
    - Expertise in designing data transformation and processing solutions.
    - Demonstrated knowledge in large-scale search applications and the development of high-volume data pipelines..
    - Orchestrated ETL processes, engineered robust database systems, and innovative tools for both real-time and offline analytics
    - Applied expertise in Hadoop architecture, HDFS commands, and query optimization for building data pipelines.

# CERTIFICATIONS

[**AWS Solutions Architect - Associate**](https://www.credly.com/badges/2a96dd38-078d-4683-9737-7922b79ec498?source=linked_in_profile)*Certification*

Microsoft Certified: Azure Developer Associate

# ACADEMIC PROJECTS

# Heat marketplace

* + - Achieved a secure, scalable, and real-time application architecture, delivering a feature-rich shoe marketplace.
    - Leveraged Amazon S3 as the primary storage solution for media assets and ensured data durability and scalability.
    - Ensured robust user authentication and authorization using Amazon Cognito.
    - Developed a custom backend framework that interfaces with Amazon RDS for structured data management, including shoe listings and user profiles.

**Smart photo album**

* + - Utilized Amazon S3 for scalable and durable storage, optimizing the management of image assets.
    - Implemented Amazon SNS for real-time notifications and alerts regarding photo uploads and album interactions.
    - Employed Amazon SQS to efficiently decouple and manage message flow, particularly for image processing tasks.
    - Utilized Amazon SES for automated email notifications and updates, enhancing user engagement and communication.