**Divya Sri Hanumandla**

Data Engineer

Irving, Texas - 75234

9452491740 / divyasri.h2511@gmail.com

**LinkedIn:** [linkedin.com/in/divya-sri-484015247](https://www.linkedin.com/in/divya-sri-484015247)

**Summary**



Experienced professional with 5 years in designing, developing, and implementing data pipelines, machine learning workflows, and big data solutions. Strong background in migrating data pipelines to Google Cloud Platform (GCP) and integrating various advertising platforms through custom API packages. Proficient in using Apache Airflow for orchestrating complex ETL workflows, ensuring scalable, flexible, and cost-effective data processing solutions. Experienced in orchestrating pipelines and training and inferencing model pipelines on Azure and using Kubernetes for orchestrating model pipelines. Collaborative team player with expertise in data preprocessing, feature engineering, model training, and deployment. Proficient in Java, Python, SQL, and ML libraries like TensorFlow and scikit-learn. Skilled in Agile methodologies and continuous improvement, with experience as a Java Full Stack Developer. Passionate about staying updated on emerging technologies. Skilled in data visualization, data modeling, and data governance.

**Technical Skills**



**Programming Languages:** Java, Python, SQL

**Data Engineering:**  Data preprocessing, Feature engineering, Data cleaning, Data transformation, Data integration

**Machine Learning:** Supervised learning, Unsupervised learning, Deep learning, Natural Language Processing (NLP), Computer Vision

**ML Libraries/Frameworks:** TensorFlow, scikit-learn, Py-Torch, Keras, Pandas, NumPy

**Cloud Platforms:** GCP, Azure

**Data Storage and Processing:** Hadoop, Spark, Hive, BigQuery, Dataflow, ADLS

**DevOps Tools:** Jenkins, Docker, Kubernetes

**Version Control:** Git, GitHub, Bitbucket

**Web Technologies:** HTML, CSS, JavaScript

**Frameworks and Libraries:** Spring Framework, Flask, Django

**Data Analysis and Reporting:** SQL queries, Excel, Jupyter Notebook,

**Operating Systems:** Windows, Linux, macOS

**Data Governance:** Data quality, Data lineage, Data cataloging, Compliance

**Professional Experience**



**Kiranam Technologies Inc-** Farmers Branch, Texas**December 2022 – Present**

**Client: CVS Health**

**Data Engineer** – Irving, Texas (Contract)                      **February 26, 2024 - Present**

* Seamlessly migrated the company's data pipeline infrastructure from Info-Works to GCP, improving data processing speed and scalability.
* Developed and deployed custom API integration packages for Google Ad Manager, Google AdWords, The Trade Desk (TTD), Campaign Manager, and Pinterest, ensuring accurate and timely data extraction.
* Designed and implemented optimized ETL workflows using Apache Airflow, significantly reducing data processing time and increasing reliability.
* Achieved cost savings by reducing dependency on third-party tools and leveraging GCP’s cost-efficient services.
* Transitioned from Info-Works pipeline to a custom-built solution on Google Cloud Platform, enabling more flexible and scalable data processing capabilities.
* Created custom packages to make API calls to multiple advertising platforms, including Google Ad Manager, Google AdWords, The Trade Desk (TTD), Campaign Manager, and Pinterest, ensuring reliable and efficient data extraction.
* Increased data extraction efficiency by 22% through the development of custom API integration packages for advertising platforms, ensuring accurate and timely data extraction.
* Utilized Airflow to orchestrate and automate ETL workflows, configuring Airflow DAGs (Directed Acyclic Graphs) to manage task sequences and dependencies.
* Leveraged Airflow’s scheduling capabilities for timely execution of data pipelines, using Airflow's monitoring tools to track workflow status and handle errors effectively.
* Utilized GCP’s infrastructure to handle large volumes of data efficiently, with custom packages allowing for tailored data extraction and transformation processes.
* Reduced reliance on third-party tools, lowering operational costs, and improved performance optimization and error handling through custom solutions and Airflow’s robust monitoring.

**Environment:** Google Cloud Platform (GCP), Apache Airflow, Kubernetes, TensorFlow, scikit-learn, Python, Java, SQL, Agile methodologies.

**Client: CVS Health**

**Data/ML-OPS Engineer – Irving, Texas (Contract)               December 2022 – February 2024**

* Collaborated with data scientists to understand data requirements and modeling objectives, designing, and developing end-to-end data pipelines supporting the data science workflow.
* Developed preprocessing steps to clean, normalize, and transform raw data into a format suitable for analysis, handling missing data, outliers, and other anomalies.
* Implemented feature engineering techniques for meaningful insights extraction.
* Implemented continuous improvement strategies, resulting in 15% reduction in pipeline errors and improved overall data processing efficiency.
* Implemented infrastructure and tools for model training, including distributed computing and GPU resources, integrating machine learning libraries like TensorFlow and scikit-learn.
* Improved data reliability and accuracy by implementing automated workflows for model evaluation and performance measurement, resulting in 7% reduction in data errors.
* Established monitoring systems to track model performance, detect anomalies, and gather feedback for continuous improvement, optimizing pipelines for efficient data processing and inference.
* Streamlined pipeline operations by 11% by orchestrating pipeline tasks using Kubernetes, automating data ingestion, preprocessing, model training, and deployment processes.
* Optimized pipeline performance up to 13% by parallelizing computations and leveraging cloud computing resources, addressing performance bottlenecks to handle increasing data volumes.

**Environment:** Snowflake, Azure, ADLS, Kubernetes, TensorFlow, scikit-learn, Jupyter-notebook, Python, Pytorch, SQL, Agile methodologies.

**Vista Applied Solutions -** Herndon, Virginia (Internship) **August 2022 – December 2022**

**Big Data Developer**

* Provided technical and functional support to business users, ensuring efficient data solutions met their requirements.
* Designed solution architectures for multiple high-priority projects, aligning with business requirements and leveraging tools like Oracle, Hadoop, Spark, NiFi, and other data analytics platforms.
* Developed proposals for scalable big data architecture, gaining exposure to technical aspects of data engineering projects.
* Developed proof of concepts for technology areas to warrant development work ahead of formal projects, ensuring feasibility and alignment with business goals.
* Designed and prototyped Oracle and Hadoop big data models to capture data as per functional requirements, ensuring data integrity and usability for analytics.
* Designed architectures and applications incorporating innovative data analytic technologies such as Hadoop, Spark, NiFi, and other open-source and proprietary platforms, enhancing data processing and analysis capabilities.

**Environment:** Big data, Pyspark, Hadoop, Big query, Nifi, power BI

**Tata Consultancy Services January 2019 – September 2021**

**Systems Engineer (Java Developer)** Hyderabad, India (Full Time)

* Developed functionalities using Agile Scrum Methodology and participated in Agile Scrum methodology for application development, including analysis, design, coding, unit, and integration testing.
* Extensively used Agile (Scrum) methodologies with sprint planning every two weeks and daily stand-up meetings to monitor the project status.
* Implemented Test Driven Development (TDD) and developed business services using JSON, XML, and Restful Web Services technologies.
* Used Spring Boot and Maven for application development and employed SOAP and REST web services for transmitting large blocks of XML data over HTTP.
* Developed UI using Spring MVC framework, JSP, JSTL, HTML, CSS, and JavaScript, including creating Reusable Modules and Custom Directives in AngularJS.
* Implemented AngularJS Controllers and Service calls, extensively tested application controllers with AngularJS built-in testing facilities.
* Utilized core Java concepts like Multithreading, Collections Framework, File I/O, and concurrency, and created POJO's and DAO's using Spring JDBC annotation mappings.
* Integrated with Spring Cloud projects for distributed tracing and other features, developed microservices with Spring and tested applications using Spring Boot, and used Maven for build scripts and GIT repositories for version control.

**Environment**: Spring Boot, Maven, Spring MVC, AngularJS, RESTful and SOAP web services, XML, HTML, CSS, JavaScript, Java core concepts.

**Education**



**Master’s computer science- GPA 3.8 /4                                              August 2021 – December 2022**

**University of Central Missouri** – Lees summit, Missouri

**Bachelor of Technology – Electrical and Electronics – GPA 4.8/5 June 2014 – July 2018**

**Malla Reddy Engineering college** – Hyderabad, Telangana, India