NAGA MANOHAR YELUBANDI

LinkedIn | nagamanohar07@gmail.com | 8125433156| Bloomington, IN

OBJECTIVE

Data engineer specializing in AWS, GCP, and big data technologies with data science master's degree. Seeking a data engineer role to leverage expertise in building scalable data pipelines and innovative data solutions in data-driven organization.

SKILLS

Programming & Database System	Python, SQL, R, Java, C, MySQL, PostgreSQL, Amazon RDS, MongoDB, DynamoDB, Cassandra.
Big Data Technologies	Spark, PySpark, Kafka, Amazon Kinesis, Athena, EMR, Hadoop, Hive, HBase, NiFi, Druid.
Data Engineering & Cloud Services	Data Modeling, Data Warehousing, ETL, Snowflake, AWS - S3, Lambda, Glue, Step Functions Redshift.
CI/CD	GitHub, Jenkins, AWS CodeBuild, Docker, Kubernetes, Terraform, DBT, Airflow, MAGE.
Visualization Tools	Tableau, Microsoft Power BI, Streamlit, D3, Looker, AWS Quicksight, Google Data studio.
Machine Learning	NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow, Statistical Modeling, Amazon SageMaker.

PROFESSIONAL EXPERIENCE

O'Neil School of Public and Environmental Affairs, Indiana University | Bloomington, IN

Data Analytics Engineer   **May 2023 – Aug 2023**

- Researched in collaboration with Professor Joanna and integrated data from various sources, uncovering that the arts and cultural sector constitutes **4.5%** of the United States' GDP.
- Implemented standardized relational data modeling in **PostgreSQL**, conducted data cleansing and transformations using **PySpark**, resulting in **80%** acceleration in data retrieval speed with **100%** enhancement in data accuracy.
- Developed an interactive sunburst visualization using the **D3.js JavaScript** library and deployed it on the ArtsAnalytics.org website, resulting in notable **16%** increase in user interaction.

Accenture | Hyderabad, India

Associate Data Engineer   **Aug 2021 - July 2022**

- Created and deployed human resource data model in **AWS**, resulting in **25%** decrease in HR operational costs.
- Led the data migration process from the HCM Fusion Model, successfully transferring and validating data of employees.
- Enhanced HR efficiency by introducing an automated system for managing employee data, reducing data errors by **50%** and saving over **10 hours** weekly in manual data entry.
- Generated HR reports using **Power BI**, leading to **20%** reduction in time-to-fill job vacancies and **15%** enhancement in quality of new hires.

PROJECTS

Reddit Data Engineering Pipeline

- Led the development of a robust data pipeline using **Apache Airflow**, **Docker containers**, **AWS S3**, and **AWS Redshift**.
- Leveraged **Terraform** for AWS resource deployment, employed Python scripts with **PRAW API** for Reddit data extraction and S3 ingestion, utilized **DBT** for data transformation, data quality checks and data reliability.
- Achieved a significant **25%** improvement in process efficiency and empowered stakeholders with actionable insights through intuitive data visualization using **Google Data Studio**.

Financial Data Streaming Pipeline

- Orchestrated the development of a real-time financial data streaming pipeline using **Docker, Kubernetes, and Terraform,** resulting in a **20%** reduction in data processing time.
- Executed data ingestion with **FinnhubProducer**, connecting to **Finnhub.io API** for message retrieval and ingestion into **Kafka**.
- Established a robust message broker layer with Kafka, monitored with **Kafdrop**, and set up **Zookeeper** for metadata management, leading to a **15%** improvement in message processing efficiency.
- Built a stream processing layer with **Spark Scala** application for message transformation and data loading into **Cassandra,** achieving **40%** decrease in data transformation time.
- Constructed visualizations using **Grafana** to serve data through a dashboard, ensuring timely updates.

Airline Data Ingestion

- Devised the event-driven data pipeline using **AWS S3**, **EventBridge**, and **Step Function**, automating data processing upon file arrival and achieving high operational efficiency.
- Integrated **AWS Glue** for data transformation and **Redshift** as the centralized data warehouse, ensuring seamless data integration and enhancing operational reliability for airline data analytics.

Real-Time Food Delivery Data Analytics with AWS

- Engineered and implemented a real-time data processing pipeline using **AWS Kinesis Data Stream** and **PySpark** Streaming on **EMR**, achieving streamlined data ingestion and processing of food delivery orders.
- Orchestrated end-to-end data workflows with **MWAA**, scheduling and monitoring data transformations for timely loading into **Redshift**.
- Designed and deployed a **star schema** data model on **Redshift**, optimizing data storage and query performance to facilitate insightful analysis for stakeholders using **QuickSight** dashboards.
- Established the robust **CI/CD** pipeline with **AWS CodeBuild**, automating deployment of dimension tables, DAG files, and PySpark scripts to **S3**, enhancing development efficiency and ensuring reliability of the data pipeline.

EDUCATION

Indiana University, Bloomington, IN| *Master of Science in Data Science*   **Aug 2022 – May 2024**

Relevant Coursework: Big Data with Hadoop and AWS, Data Mining, Advanced Database Concepts.

Vishnu Institute of Technology, AP, India| *Bachelor of Technology in Electronics & Communications Engineering*   **Jun 2017 - May 2021**

Relevant Coursework: Data Structures, Computer Architecture and Organization, DBMS.