SRINADH DIVAKARUNI

DATA ENGINEER

✉ divakarunisrinadh@gmail.com   📱 9133736301   📍 India

in https://www.linkedin.com/in/srinadh-divakaruni-519940143/

SUMMARY

A seasoned Data Engineer boasting 3.9 years of expertise, adept in crafting scalable data pipelines and proficient across multiple programming languages. Possessing a mastery of cloud platforms, with a specialization in data modeling and warehousing. Demonstrates excellence in ETL processes and ensures smooth data integration. Recognized for collaborative teamwork, strong problem-solving abilities, and effective communication skills.

EDUCATION

Bachelor of Technology

SRKR Engineering College *Jun 2017 - Jun 2021*

Graduated in the stream of Computer Science and Engineering with GPA 8.74

EXPERIENCE

Data Engineer

Payoda Technologies *May 2024 - Present*
Client: Apple (Maps Platform)
Role: Data Engineer

Project: Address Automation and Maintenance

- Developed and implemented a streamlined data pipeline to publish address and place of interest data in Protocol Buffers format for multiple countries, ensuring efficient integration across regional vendors.
- Ingested input data from Cassandra into HDFS, processed it using Spark, and subsequently published the output data back to HDFS, ensuring efficient data flow across the pipeline.
- Optimized Spark jobs, reducing processing time from 12 hours to 2 hours, significantly improving the overall system performance and contributing to enhanced data processing efficiency.
- Designed and implemented comprehensive validation checks on incoming data, ensuring high data quality and consistency before publishing to production-standard maps.
- Automated regional data processing by performing iterative testing and deployment cycles, resulting in increased automation and reduced manual intervention across multiple regions.
- Collaborated with cross-functional teams to ensure seamless data flow from input to output, facilitating effective data analysis using in-house tools.

Data Engineer

Modak Analytics *Apr 2021 - Apr 2024*
Role: Data Engineer

Project: Unstructured Automation

- Designed, constructed, and operationalized ETL pipelines for diverse data sources, incorporating end-to-end monitoring and error handling, deployed them to the on-prem environment, and orchestrated them via Airflow.
- Collaborated and worked alongside clients to comprehend business requirements and crafted curation data pipelines to meet their needs.
- Containerized all data pipelines with Docker and seamlessly deployed them to Kubernetes.
- Enhanced data pipeline efficiency by introducing multi-threading, significantly reducing runtime from a month to 2 days.

- Successfully developed and migrated all data pipelines from on-premises environments to Google Cloud Platform (GCP). Developed a crawler that helps find new files from the GCS bucket and was involved in designing a data model in GCP.

Project: Structured Curation

- Created Pyspark code to extract data from the Azure SQL database, apply necessary transformations according to business requirements, and then transfer the curated data to ADLS.
- Developed an Azure Data Factory pipeline for user input, directing it to Databricks notebooks to execute curation processes. The resulting metrics were loaded into the Azure SQL database, minimizing manual intervention.
- Employed Azure Key Vault to securely store credentials, ensuring that Databricks notebooks could access them for establishing connections to the SQL database.

Project: Automated Metrics Monitoring

- Developed Pyspark code to automate metrics monitoring of multiple structured and unstructured sources data pipelines along with log metrics and query metrics based on custom input by the consumers, developers and the data scientists.
- Reducing monitoring time and saved 30% revenue and 50% of the monitoring time by orchestrating using Airflow by enabling task/DAG level email/teams alerts.

TECHNICAL SKILL SET

- **Relational Databases:** PostgreSQL, MySQL
- **CI & CD:** Azure DevOps, Unified Build
- **Version Control:** GitHub
- **Big Data Tools:** Hadoop, Hive, Spark, Kafka
- **Unified Analytics Platform:** Databricks
- **Languages:** Java, Python, Pyspark, Scala, Shell, SQL, Data Structures and Algorithms
- **Cloud Platform:** Microsoft Azure, AWS(Glue, EMR, Lambda), GCP(BigQuery, GCS, Workflows, Google Batch, GKE)
- **Container Orchestration tools:** Docker, Kubernetes
- **Operating System:** Linux, Windows
- **Pipeline Orchestration tools:** Airflow
- **Data Modelling**

CERTIFICATIONS

DP 900: Microsoft Certified: Azure Data Fundamentals
Microsoft   *2023-02-11*

DP 203: Microsoft Certified: Azure Data Engineer Associate
Microsoft   *2023-05-27*

Apache Airflow: Airflow Fundamentals
Astronomer   *2022-02-11*

Apache Airflow DAG Authoring
Astronomer   *2023-01-25*

Databricks Lakehouse Platform Fundamentals
Databricks   *2023-01-15*

Databricks Certified Data Engineer Associate
Databricks   *2024-02-15*