

Venu Gopal

+91 8210230975 | venugpl1994@gmail.com
LinkedIn | **GitHub**

PROFILE

Data Engineer with 4+ years of experience in building large-scale data pipelines, ETL processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark, Airflow, AWS and Kafka** to develop multi-terabyte scalable big data solutions for Fortune 100 Pharmaceutical and E-commerce companies.

TECHNICAL SKILLS

Programming Languages: Python, SQL, Shell-scripting
Big Data Technologies: Spark, PySpark, Spark SQL, YARN, Jenkins, DynamoDB, Hive
Streaming Technology: Kafka, Spark structured streaming, Amazon Kinesis
Cloud computing: AWS Lambda, EC2, Databricks, AWS Redshift, Amazon S3, AWS Glue, AWS EMR
Data Engineering Tools: AWS CodePipeline, Data Modelling, ETL/ELT data Pipeline
Orchestration: Airflow, EventBridge Scheduler
Familiar: Apache Iceberg, Machine Learning

EXPERIENCE

- Ascendion Engineering Pvt. Ltd** Feb 2024 - Present
Senior Software Engineer *Bangalore, India*
 - Build a configuration driven extension to airflow where developers can define their workflows in YAML format and consistently produce automatic DAG code which airflow can interpret.
 - Conduct data validation on source and staging tables using DAG factory with a YAML configurational approach. Execute table transformations using Airflow and migrate data into other SQL environments.
 - Integrated and processed pharmaceutical data from multiple sources including **AWS, Azure, and on-premise database** handling various file formats like JSON, CSV, and Parquet.
 - Implemented Spark optimization techniques such as **caching, multithreading, and broadcast** joins, resulting in a 20% decrease in processing time for handling a daily load.
- LT Technology Services** Jun 2023 - Nov 2023
Senior Engineer *Bangalore, India*
 - Created an API service using Python to generate dynamic DAGs in **Apache Airflow**.
 - Designed and implemented advanced scheduling capabilities using Airflow for data pipeline orchestration, reducing manual intervention time by 80% and streamlining workflow efficiency.
 - Developed a solution leveraging Amazon MSK (Managed Streaming for Apache Kafka) to push messages for real-time tracking of processing status (success or failure).
- Hinduja Global** Dec 2021 - Apr 2023
Data Engineer *Bangalore, India*
 - Worked within the marketplace team with the primary objective of enhancing seller experience.
 - Integrated and transformed data from various digital platforms, prepared data, built ETL and ELT processes, and orchestrated and monitored pipelines utilizing AWS Glue.
 - Optimized overall process performance through **Spark performance tuning**, improving job run times by 20% and efficiently managing a **Terabyte(TB)** dataset.
 - Worked on data ingestion pipeline to ingest the flat file in the **Data lake**.
- Phasor Pvt. Ltd** Oct 2019 - Nov 2021
Business Analyst *Mumbai, India*
 - Worked extensively with large marketing datasets concerning user behavior, identifying trends, and conducting in-depth analysis to draw meaningful conclusions.
 - Utilized standard metrics such as Click-Through Rate, Conversion Rate, ROI, Cost per Click, etc., to gauge the success of campaigns..

EDUCATION

CMR Institute of Technology Bangalore, India
Bachelor of Technology in Computer Science 2018