

Durga Prasanth Merugu

Dayton, OH — merugudurgaprasanth@gmail.com — (865) 789-4197 — linkedin.com/in/Merugu-prasanth/

Professional Summary

Data Engineer with **over 5 years** of experience designing and implementing cloud-native data solutions on **AWS and GCP**. Skilled in building scalable batch and real-time pipelines using **Airflow, Spark, and dbt**, and managing modern data platforms such as **Snowflake, BigQuery, and Databricks**. Brings hands-on expertise across healthcare and financial domains, delivering analytics-ready datasets that drive critical business decisions.

Education

University of Dayton, Dayton, OH

Aug 2022 – May 2024

Master of Science, Computer Science

MVGR College of Engineering, Vishakapatnam, Andhra Pradesh

Jul 2015 – Jun 2019

Bachelor of Technology, Electronics & Communication

Skills

Programming & Scripting: Python, SQL, PySpark, Spark SQL, Java

Cloud Platforms: AWS (S3, Glue, Lambda, EMR, Step Functions), GCP (Dataflow, Pub/Sub)

Orchestration & Workflow Management: Apache Airflow, Terraform

Data Modeling & ETL: dbt, Talend, Alteryx, SSIS

Data Warehousing & Databases: Snowflake, BigQuery, Redshift, PostgreSQL, MySQL, Oracle, SQL Server

Streaming & Real-Time Processing: Apache Kafka, AWS Kinesis, GCP Pub/Sub

Business Intelligence & Visualization: Tableau, Power BI, Looker

DevOps & Monitoring: GitHub, CI/CD, Prometheus, Grafana

Professional Experience

Lowe's, Data Engineer

Sep 2023 – Present

Remote, United States

- Built and managed ETL pipelines on Databricks, using Apache Airflow to automate workflows for multi-terabyte retail datasets. Improved data accuracy to 99% and cut pipeline run times by 35%.
- Designed dbt models that turned raw sales and inventory data into analytics-ready tables, enabling business teams to access real-time insights on supply chain performance and pricing trends.
- Deployed workflows on AWS (S3, Glue, Lambda, Redshift), supporting both batch and streaming use cases for inventory tracking and customer behavior analysis.
- Set up monitoring and alerting with Prometheus and Grafana, reducing time to detect and fix pipeline issues by 40% and boosting system reliability.
- Worked closely with engineering, analytics, and business teams to define KPIs, deliver dashboards in Power BI and Looker, and align data products with business goals.
- Ensured data security and governance by implementing IAM roles, encryption (at rest and in transit), and compliance practices to meet SOC 2 standards.
- Documented workflows, data models, and processes to support onboarding and keep the team aligned as projects scaled.

Tata Consultancy Services, Data Engineer

Jan 2019 – Jun 2022

Pune, India

- Designed and maintained ETL workflows to integrate sales and weather data using Python (Pandas, NumPy) and Apache Spark, ensuring consistent and reliable datasets for financial analytics.
 - Built and optimized data pipelines on GCP (BigQuery, Dataflow, Pub/Sub), enabling efficient processing of large datasets and reducing pipeline execution times by 25%.
 - Developed data transformation logic with Talend and SSIS, improving data reliability and cutting ETL job failures by 20%.
 - Implemented real-time data ingestion for sales trend analysis using Pub/Sub and Dataflow, supporting near real-time reporting for business operations.
 - Set up monitoring and alerting systems with Prometheus and Grafana to detect pipeline issues early, reducing average recovery time by 30%.
 - Created Tableau dashboards for key financial and demand KPIs, enabling stakeholders to make faster, data-driven decisions.
 - Applied data governance and security best practices, including IAM configurations, encryption, and compliance with internal audit requirements for financial data.
-

Projects

Real-Time Retail Insights Platform (GitHub)

- Designed and implemented a real-time data platform to process and analyze transactional data from an e-commerce application. Orchestrated workflows with Apache Airflow and managed infrastructure as code using Terraform.
- Ingested high-velocity data streams with Apache Kafka, performed transformations in Databricks (PySpark), and loaded analytics-ready data into Amazon Redshift.
- Delivered dashboards in Power BI enabling business teams to monitor customer behavior in real-time, driving a 20% improvement in conversion rates and reducing stockouts by 15%.

Healthcare Data Lakehouse Solution (GitHub)

- Built a HIPAA-compliant data lakehouse on GCP to unify and process multi-source healthcare datasets (EHR, claims).
- Developed ETL pipelines using Google Cloud Dataflow and transformed raw data into curated layers with dbt for analytics. Integrated IAM roles and encryption for data security, meeting audit requirements and improving governance.
- Empowered analytics teams with Looker dashboards for real-time patient monitoring and operational insights, reducing reporting latency by 40%.