

Girish Manjunatha Saraswathipura

gsaraswathipura@gmail.com | +1 (703)-400-9756 | [LinkedIn](#) | [GitHub](#) | Bloomington IN

Education

Master of Science in Data Science - ML, Data Mining, Advance Database, AI Research, Applied Algorithm

08/2024 - 04/2026

Indiana University Bloomington, USA

Bachelor of Engineering in Computer Science - Mathematics, DevOps, Software Engineering, MongoDB

08/2018 - 06/2022

Visvesvaraya Technological University, India

Work Experience

IU Innovates

06/2025 - 08/2025

Machine Learning Research Intern - Deep Learning, HPC, Fine Tune LLM, Statistics, Data Profiling, CNN, ANN, GPU, CUDA

- Automated profiling for 50M+ records using Python and SQL, **reducing data prep time by 45% and accelerating time-to-insight for analytics and user growth teams.**
- Built reusable ingestion and preprocessing modules ML pipeline onboarding, accelerating readiness and reducing hand-off time by 30%.

Informatica Inc.

01/2024 - 07/2024

Associate Consultant, ETL Developer - ETL optimization, Data Quality, Synapse SQL, AWS EC2, Azure, Snowflake

- Migrated 1,100+ ETL workflows to IICS, **reducing infrastructure cost, improving data availability, and enabling 21% faster analytics for cross-functional business and product stakeholders.**
- Architected star and snowflake schemas in Snowflake and Oracle to support high-volume product analytics queries; improved dimensional data lookup performance by 30% and cut reporting pipeline storage by 30%.
- Implemented data lineage automation for 300+ dimensions using IICS metadata APIs; improved root-cause analysis time by 15%
- Designed reusable transformation templates for migration, cutting future migration efforts by 31% across additional business units.
- Integrated data quality profiling rules using Informatica DQ and produced standardized quality dashboards consumed by operations teams, resulting in a 20% drop in pipeline failures.

Informatica Inc.

08/2022 - 12/2023

Migration Factory Associate Specialist, Data Engineer - Data Warehouse, Data Wrangling, Oracle, PySpark, Hadoop

- Partnered with engineering teams at Volvo and Charles Schwab to migrate 370+ ETL pipelines, representing 8% of their enterprise workflows, with <2% failure rate after production cutover.
 - Developed a Python + AWS EC2 automation tool that reduced internal certificate validation time from 10 hours to 3 hours, lowering cloud compute costs by ~\$4,000 per month.
 - Delivered automated pre- and post-migration validation scripts in Python and SQL to surface data anomalies early, increasing migration accuracy and **boosting confidence for downstream product analytics teams by preventing 100+ defects.**
 - Provided hands-on troubleshooting of session/performance failures using **IICS Monitor, session logs, SQL explain plans, and error categorization**, reducing post-migration defect cycles by **~50%**.
 - Integrated Snowflake, Hadoop, and Redshift into scalable analytics pipelines, leveraging distributed SQL engines (Hive, AWS Athena) for terabyte-scale data modeling and analysis.
-

Projects

Cloud-Native ETL Pipeline with AWS Glue, S3, and Redshift - PySpark, AWS Glue, S3, Redshift, Data Pipeline

07/2025 - 09/2025

- Designed and deployed a scalable Glue ETL pipeline to ingest and transform 10M+ records daily from S3 into AWS Redshift; enforced schema evolution, partitioning, and job bookmarks for incremental load efficiency.

Hybrid Workflow Orchestration with IICS Application Integration - Informatica, Application Integration, Scheduling

03/2024 - 07/2024

- Orchestrated multi-phase ETL workflows across Oracle, Snowflake, and Salesforce using CAI service connectors and schedulers, reducing orchestration logic by 50% and stabilizing cross-platform data loads.
-

Skills

Data Engineering: Informatica Cloud (IICS), AWS Glue, AWS Redshift, PySpark, Hadoop Ecosystem (HDFS, Hive), Data Warehousing, Data Governance, Master Data Management, Azure, Docker, Git, Linux ; **ETL:** Informatica PowerCenter, Application Integration (CAI), ELT/ETL Pipelines, Airflow, Data Pipeline Scheduling, Metadata Management, Schema Evolution, CDC Patterns ; **Databases & Querying:** SQL, Snowflake, Oracle, PostgreSQL, AWS Athena, Azure Synapse, JDBC/ODBC ; **Data Modeling:** Snowflake Schema, Metadata Management, Data Quality Checks, Data Lineage, Schema Evolution, CDC Patterns ; **Programming:** Python, R, C++, C, Java Basics

Certifications and Credentials: AWS AI Practitioner, Informatica Certified Data Integration Professional, PowerCenter Data Integration 10 Administrator Informatica Certified Professional, MongoDB M001, Informatica Cloud Application Integration Practitioner