

# Sohith A.

Houston, TX | +1-346-770-9578 | [sohith0708@gmail.com](mailto:sohith0708@gmail.com) | <https://www.linkedin.com/in/asohith0209/>

## PROFESSIONAL SUMMARY

Experienced Data Engineer designing scalable ETL pipelines with Python, SQL, and Apache Spark. Proficient in cloud platforms (AWS, GCP) and data warehousing with Snowflake. Skilled in data modeling, query optimization, and automation using AWS Glue, Lambda, and Kubernetes. Strong collaborator with data scientists, delivering data structures for analytics and machine learning. Adept at Git version control and Agile methodologies for efficient project delivery.

## EDUCATION

### University of Houston

*Masters, Computer Science*

**Aug 2023 - May 2025**

### SRM University

*Bachelor of Technology, Computer Science*

**Jul 2019 - Jun 2023**

## SKILLS SUMMARY

- **Languages:** Python, SQL, R
- **Frameworks and Tools:** Apache Spark, Hadoop, Kafka
- **Databases:** Snowflake, MongoDB, PostgreSQL, MySQL, Oracle
- **Cloud Platforms:** Amazon AWS (Glue, Lambda, Redshift, SageMaker), Azure, GCP
- **DevOps and Version Control:** Git, Docker, Kubernetes, Terraform
- **Concepts:** ETL Pipelines, Data Modeling, Big Data Processing, Agile Development

## EXPERIENCE

### Health Catalyst

*Data Engineer*

**May 2024 - Present**

*Houston, TX*

- **Data Pipeline Development:** Designed and implemented scalable ETL pipelines using Python, SQL, and Apache Spark to process large datasets into Snowflake, improving data availability by 25%.
- **Cloud Automation:** Developed Python scripts for AWS Glue and Lambda to automate data workflows, reducing manual processing time by 30%.
- **Data Modeling:** Built and optimized data models and schemas in Snowflake and MongoDB, enhancing analytics performance for cross-functional teams.
- **Machine Learning Integration:** Created Amazon SageMaker notebooks with Python to automate data correction, improving data quality by 20%.
- **Query Optimization:** Engineered complex SQL queries and stored procedures in Snowflake, enabling faster reporting and analytics for business units.
- **Processed large-scale data** using Databricks and Hadoop, enhancing support for advanced analytics and machine learning workflows.
- **Container Orchestration:** Utilized Kubernetes to manage containerized data applications, ensuring scalability and reliability in production environments.
- **Collaboration:** Worked with data scientists to deliver optimized data structures for machine learning models and reporting dashboards.

### Accenture (Gatorade)

*Data Engineer*

**May 2021 - Jul 2023**

*Houston, TX*

- **ETL Pipeline Engineering:** Architected end-to-end ETL pipelines using Python, PySpark, and SQL to ingest and transform data into Snowflake, increasing pipeline efficiency by 15%.
- **Data Quality Automation:** Developed reusable Python modules for data cleansing and validation, improving data integrity for downstream analytics.
- **Cloud Workflow Automation:** Automated data workflows with AWS Glue and Lambda, reducing operational overhead by 20%.
- **Snowflake Optimization:** Implemented SQL-based views and stored procedures in Snowflake to support advanced analytics, reducing query execution time by 10%.
- **Containerized Deployments:** Used Kubernetes to orchestrate containerized applications, enhancing deployment scalability and reliability.
- **Version Control:** Maintained codebases using Git, ensuring traceability and efficient collaboration across teams.
- **Data Ingestion:** Built robust ingestion workflows with Python and cloud ETL tools to process large-scale datasets, improving data readiness.

## CERTIFICATIONS

- **AWS Certified Solutions Architect - Associate:** Amazon Web Services (AWS), Issued: Jan 2025
- **Google Certified: Associate Cloud Engineer:** Google Cloud Platform (GCP), Issued: Jun 2025