

# Sai Rohith Tulasi

Masters in Computer Science  
University of Texas at Arlington

+1-682-560-9473

sairohithtulasi24@gmail.com

LinkedIn Profile

## PROFESSIONAL SUMMARY

- Data Engineer with 5+ years of experience in designing, building, and optimizing large-scale, distributed systems. Proven expertise in creating scalable and fault-tolerant data pipelines using Python, Scala, SQL, and Apache Spark. Skilled in data warehousing, ETL processes, and cloud-based solutions, with a strong focus on delivering data-driven insights and automating workflows.

## TECHNICAL SKILLS

**Programming Languages:** Python, Scala, Java, SQL, Shell Scripting

**Data Technologies:** Apache Spark, Spark SQL, Hadoop, Hive, Snowflake, S3

**Data Serialization Formats:** Parquet, Avro, ORC

**Data Warehousing & ETL:** Data Modeling, ETL Pipelines, Data Analytics, Business Intelligence

**DevOps & Automation:** Jenkins, Git, CI/CD, Apache Airflow

**Databases:** Oracle, MySQL, HBase, NoSQL Databases

**Development Methodologies:** Agile, Scrum, Waterfall

## PROFESSIONAL EXPERIENCE

### •Data Engineer

August 2024 - Present

Ven Solutions, Irving, Texas

- Designed and implemented scalable data pipelines using Apache Spark and Scala, processing over 1TB of data daily for real-time analytics and reporting.
- Optimized Spark SQL queries, reducing query execution time by 25% and improving overall data processing efficiency.
- Collaborated with teams to integrate Snowflake with AWS S3 for seamless data ingestion, enhancing data availability for analytics and reporting.
- Developed and maintained ETL workflows using Hive and Snowflake, enabling seamless data integration and transformation across multiple data sources.
- Utilized Parquet and Avro formats for efficient data storage and retrieval, reducing storage costs by 15%.

### •Big Data Engineer

June 2021 - July 2022

Optum Global Solutions, Hyderabad, India

- Migrated 50+ ETL jobs from Talend to Apache Airflow, cutting licensing costs by 30% and enhancing job scheduling efficiency.
- Processed 10+ GB of patient data daily using Python and Spark, transforming raw data into structured formats for analysis.
- Designed and managed Hive tables for data warehousing, enabling efficient querying and analysis of large datasets.
- Utilized HBase for metadata management and optimized Spark jobs, achieving a 25% reduction in processing time.
- Collaborated with DevOps teams to implement CI/CD pipelines, ensuring smooth deployment of data solutions.

### •Data Engineer

November 2017 - June 2021

ADP Private Limited, Hyderabad, India

- Developed and maintained SQL and SQR scripts for Oracle databases, automating data updates and processing 500+ client-specific reports.
- Executed complex SQL queries for payroll and HR data, improving data accuracy and efficiency for 500+ clients.
- Implemented data validation scripts to resolve 100+ critical production issues related to data mismatches.
- Utilized Python for automating repetitive tasks and improving data extraction processes, reducing manual effort by 25%.
- Collaborated with cross-functional teams to integrate data from multiple sources, enhancing data consistency and reliability.

## EDUCATION

### •Master of Science in Computer Science

August 2022 - May 2024

University of Texas at Arlington, Arlington, TX. GPA: 3.66

### •Bachelor of Technology in Computer Science and Engineering

August 2013 - May 2017

Kakatiya University College of Engineering and Technology, India. GPA: 3.4