# RISHI VARAHAGIRI

rishi.varahagiri@outlook.com | San Antonio, TX | (210) 812-1176 | linkedin.com/in/rishi- varahagiri-417729102

## SUMMARY

Data enthusiast loves playing with data. Team player, quick learner with over 6 years of professional experience as Data Engineer and proficient in Python, Spark, SQL, DBT, ETL Modelling, Dimensional Modelling, Tableau and Cloud technologies

## TECHNICAL SKILLS

| | |
|---|---|
| **Programming Languages:** | Python, Shell Scripting, JavaScript |
| **Databases:** | Snowflake, MS SQL Server, MySQL, Teradata, PostgreSQL |
| **ETL:** | Matillion, Stream Sets, Microsoft SSIS, Informatica, Apache Spark |
| **Tools/Libraries:** | DBT, Tableau, Azure, AWS, Git, NumPy, Pandas, Scikit, UiPath, Docker, Gen AI |

## WORK EXPERIENCE

**Senior Data Engineer, Zscaler, San Jose, CA**                                          **Jan'23 – Present**

- Engineered an end-to-end framework in Python that can run scripts on an EC2 instance. Utilized this framework to modify long-running jobs, resulting in reduced runtime by at least 25% and consequently reducing Snowflake credit consumption, leading to cost savings
- Integrated an automated PR review process in GitHub, leveraging the OpenAI API to provide developers with actionable feedback on optimized and efficient code, resulting in a 20% reduction in review time and a significant improvement in overall code quality
- Developed an automated solution for correcting erroneous data and validating data hygiene in Salesforce, replacing the expensive third-party tool, achieving a 10x improvement in data processing times and a 3x increase in data availability, leading to 230k in annual savings
- Implemented an efficient AES 256 encryption and decryption procedures in Python, significantly enhancing data security for sensitive information, reducing runtime by 4X compared to AWS lambda solution, and achieving a 2x increase in data availability on Snowflake
- Modelled DBT workflows for streamlining the process of mapping winning territories for Accounts in Salesforce, encompassing data extraction, logical computations, and reverse updates, resulting in a 50% decrease in man-hours spent and 80% reduction in system validation errors

**Data Engineer, Elevate Credit, Dallas, TX**                                          **Feb'20 – Dec'22**

- Spearheaded the management of a large-scale Data Lake, optimizing data flows across multiple products and successfully migrating Elevate Data Warehouse from Teradata to Snowflake resulting in 20% decrease in overall data management overhead
- Designed and implemented scalable data pipelines utilizing Stream Sets ETL tool, successfully ingesting over 10 million records daily into Snowflake and leveraged Stream Sets Control Hub for job orchestration, resulting in a 40% reduction in manual intervention
- Implemented a near real-time data ingestion solution using Snowflake streams and tasks architecture, automating the batch data ingestion process and improving data freshness by 80% and improved data ingestion efficiency by 90%

**Data Engineer, TATA Consultancy Services, Hyderabad, India**                                          **Jan'16 – Feb'18**

- Architected and delivered a high-performance, scalable data warehouse solution using MS SQL Server, and visualized telecom assets reports on Tableau and Power BI, providing actionable business intelligence, driving data-driven decision-making
- Developed data integration workflows using SSIS, SQL stored procedures, and data pipelines to extract, transform, and load data from a variety of sources, including Oracle, MySQL, flat files, and others.
- Awarded Star of the Month for innovatively automating repetitive data loads, resulting in a reduction of 8 hours of manual efforts per month

## ACADEMIC PROJECTS

**Analyzing Common Media Sense Reviews**                                          **Jan'19 – May'19**

- Programmed python script to consume Common Media Sense reviews using requests module across different media.
- Configured Pyspark, Pandas docker images on AWS and analyzed 100k+ reviews to study user review patterns
- Developed data pipelines to draw insights from user reviews and performed aggregations on the data
- Visualized data over Power Bi to identify user review pattern across media categories

**Parallelizing Social Score Computations**                                          **Mar'19 – May'19**

- Computed social scores for 100K+ tweets to analyze and study the toxicity of Twitter tweets and Facebook comments
- Increased processing speed of computations by 10x using Python's Open MPI and collective communication framework
- Configured virtual machines on OpenStack to perform calculations across nodes to achieve maximum throughput

## EDUCATION

**The University of Texas at San Antonio, San Antonio, TX**                                          **Aug'18 – Dec'19**

Masters in Computer Science (MS)

**M.V.G.R. College of Engineering (JNTUK), India**                                          **Oct'11 – May'15**

Bachelors in Computer Science & Engineering (B.Tech)

## CERTIFICATIONS

SNOWFLAKE SNOW PRO CORE                                          **Mar'24**
MICROSOFT CERTIFIED PROFESSIONAL AZURE FUNDAMENTALS                                          **Sep'20**