

MARUTHI MANOJ KOLLI

📞 716-292-5717 ✉ kollimmanoj@gmail.com <https://manoj-kolli.github.io>

Summary

I am a Master's graduate in Computer Science/Data Science with over five years of experience as a Data Engineer, boasting end-to-end expertise in designing, implementing, and automating data pipelines. My experience spans across health, retail, and financial domains, where I have honed strong skills in ETL/ELT, Python, PySpark, SQL, and NoSQL databases. Additionally, I possess expertise in data orchestration tools across Azure, GCP, and AWS, integrating them with CI/CD pipelines, and working with data warehousing tools.

Technical Skills

Programming Languages: Python, Java, C,C#, HTML/CSS, JavaScript, Hadoop, Pytorch, Apache spark, kafka
Web Technologies: JavaScript, HTML, CSS3, Bootstrap, Typescript
Database Technologies: Microsoft SQL Server, postgres, MySQL, NoSQL, PLSQL, TSQL, Oracle, Redshift, Tensorflow.
Cloud Technologies: AWS, Azure, Google Cloud Platform.
Tools/Platform: Git, Github, Excel, Visualstudio, Data visualization - Tableau, Power BI, NLP, WYSWIG, Jenkins, Docker.
Packages and Frameworks: Numpy, Pandas, Matplotlib, Seaborn, Plotly, Scikit-Learn, ggplot, Data Analytics, data mining.

Professional Experience

WORKMATE INC

Nov 2023 – Present

Data Engineer

Remote

- Implemented scalable data ingestion pipelines using Azure Data Factory for diverse data sources, including SQL databases, CSV, JSON files, and REST API, while also developing and refining data modeling to ensure optimal performance and scalability of data storage and data solutions.
- Optimized ETL data pipelines, reducing operational costs by 30%, while maintaining data accuracy and integrity.
- Improve existing data architecture to include new metric capabilities for customers data.
- Optimizing T-SQL, NoSQL queries in Azure Cosmos DB to enhance data storage and data retrieval.
- Building Tableau dashboards to improve decision making for inventory planning using Agile methodologies.

AMAZON.COM, INC

Sep 2022 – Oct 2023

Data Engineer

Seattle, WA

- Design data pipeline to process new data from external systems and Routinely validate production data by performing data wrangling using AWS Glue connector and leveraging services like Lambda, S3, Redshift, EC2.
- Coordinating and developing with cross-functional teams on increasing the revenue for business processes/needs and to create Machine Learning algorithms/predictive models that result in 10M seller contact reduction annually for creative product features, optimizing creative performance.
- Performed operations where data relay through reading and writing into S3, Athena queries on S3 and Glue catalog, loading into DynamoDB, and executing ETL transformations using customized EMR clusters.
- Developing automation and processes to simplify the CI/CD pipeline (maintenance) and allow data analysts to deliver solutions quicker to production and reduce retooling time.

TATA CONSULTANCY SERVICES - State Farm Insurance, USA

May 2018 – Sep 2021

Data Engineer

Hyderabad, India

- Developed and maintained end-to-end operations of ETL data pipeline and worked with about 300GB of daily data in Azure data factory.
- Performed ETL data pipelines using PySpark, transforming raw data into user-friendly formats, leading to a 20% increase in data processing efficiency and efficient data analysis.
- Developed ELT/ETL pipelines to move data to and from the Snowflake data store using a combination of Python and Snowflake SQL. Moved Data from Azure Data Lake to Snowflake internal stage using various copy options.

Analyzing Climate Change using Earth Surface Temperature Data

June 2022 - Sep 2022

Research Project

Buffalo, NY

- Employed ARIMA model for the data of eight decades to predict the surface temperature visualization of the earth.
- Developed a web interface using R-shiny to visualize the predicted temperature using graphs and to analyze the changes over time.
- Performed Exploratory Data Analysis, revealing the consistent temperature rise over the years through parameter estimation, incorporating geospatial analysis.

Education

MS in Engineering Science - Data Science, State University of New York at Buffalo.

Aug 2022