

MARUTHI MANOJ KOLLI

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

Summary

I am a Master's graduate in Data Science with over five years of experience as a Data Engineer and Software Development 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, Hadoop, Pytorch, Apache spark, PySpark, Kafka, Scala

Web Technologies: JavaScript, HTML, CSS, Bootstrap, Typescript, React

Database Technologies: Microsoft SQL Server, postgres, MySQL, NoSQL, TSQL, Oracle, Redshift.

Cloud Technologies: AWS, Azure, Google Cloud Platform, Snowflake, Databricks.

Tools/Platform/Frameworks: Git, Github, Excel, Visualstudio, Data visualization - Tableau, Power BI, NLP, WYSIWIG, Jenkins, Docker, Numpy, Pandas, Matplotlib, Seaborn, Plotly, Tensorflow, Scikit-Learn, ggplot, Data Analytics, data mining.

Certifications

AWS Certified Solutions Architect - Associate

May 2024

Professional Experience

R1RCM

Nov 2023 – Present

Data Engineer

Remote

- Developed robust data pipelines using Azure Data Factory to handle SQL, CSV, JSON, and REST API inputs; refined data models, achieving a 25% boost in data processing efficiency and 20% improvement in scalability
- Optimized ETL data pipelines, reducing operational costs by 30%, while maintaining data accuracy and integrity. Implement Python-based machine learning algorithms and predictive models using libraries like Scikit-learn or TensorFlow to optimize revenue and reduce seller contacts.
- 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.
- Managing code reviews, version control, and deployment pipelines using GitBash, including branching, merging, and staging, to ensure that changes to code are properly tracked and managed.
- 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 with cross-functional and Business Intelligence 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 to facilitate effective querying on data to perform business Objectives.
- 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.
- Optimizing Data-driven models to support large-scale, to facilitate effective querying on the log data to perform business Objectives at the backend.
- 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.
- Collaborated with AWS Cloud Services to render auto-generated and dynamically optimized display ads on various platforms, including Amazon DSP, IMDb, and Amazon Video, using React for front-end development, significantly increasing efficiency and driving advertiser adoption with new creative features.

TATA CONSULTANCY SERVICES - State Farm Insurance, USA

May 2018 – Sep 2021

Data Engineer/Full Stack

Hyderabad, India

- Developed end-to-end ETL pipelines in Azure Data Factory, handling 300GB daily data, and optimized Snowflake data warehouse for efficient storage and retrieval.
- Enhanced data processing efficiency by 20% using PySpark for data transformations, ensuring high data quality and integrity with Python libraries.

- Transformed raw data into user-friendly formats using PySpark, leading to a 20% increase in data processing efficiency and enabling more effective data analysis.
- Automated data workflows using Azure Databricks, enabling seamless deployment and management of Python scripts for efficient data processing.
- Implemented dynamic web content features, including real-time location-based content delivery and customer interaction analytics, improving user engagement.
- Developed single page applications (SPAs) with HTML5, CSS3, C, JavaScript, Ajax, enhancing user experience and data retrieval efficiency.
- Designed rich user interfaces and in-page navigations using CSS, Bootstrap, and JavaScript, integrating CMS core services for dynamic content management.

Education

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

Aug 2022