

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, Hive.

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

Tools/Platform/Frameworks: Git, Github, Excel, Visualstudio, Data Structures, 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, integrating edge computing solutions to reduce latency and increase data processing efficiency. Refined data models to achieve a 25% boost in batch data processing efficiency and a 20% improvement in scalability, supporting capacity planning and future growth.
- Optimized ETL data pipelines, reducing operational costs by 30%, while maintaining data accuracy and integrity. Implemented Python-based machine learning algorithms and predictive models using libraries like Scikit-learn or TensorFlow, contributing to revenue optimization and reducing seller contacts.
- Improved existing data architecture to include new metric capabilities for customer data, streamlining capacity planning and ensuring scalability for future projects.
- Optimized NoSQL queries in Azure Cosmos DB to enhance data storage and retrieval, utilizing edge computing techniques to further reduce data retrieval time and improve real-time decision-making.
- Managed 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. Leveraged project management best practices to ensure on-time delivery and maintain code quality.
- Built Tableau dashboards to improve decision-making for inventory planning, utilizing Agile methodologies for project management and ensuring effective communication and collaboration with cross-functional teams. This supported capacity planning and resource allocation across the business.

AMAZON.COM, INC

Sep 2022 – Oct 2023

Data Engineer

Seattle, WA

- Designed a cloud-computing data pipeline to process new data from external systems and routinely validated production data by performing data wrangling using AWS Glue connectors, while leveraging services like Lambda, S3, Redshift, and EC2 to ensure compliance with industry standards.
- Coordinated with cross-functional and Business Intelligence teams to make data-driven decisions for optimizing revenue and business processes. Developed machine learning algorithms and predictive models, resulting in a 10M annual reduction in seller contacts. This enhanced the product lifecycle by improving creative features and troubleshooting performance issues in the process, leading to a more efficient querying of data for reporting and growth strategy formulation.
- Performed operations involving data relay through reading and writing into S3, querying Athena, and Glue catalog, loading data into DynamoDB, and executing ETL transformations using customized EMR clusters. This ensured compliance, supported large-scale data solutions, and facilitated communication of business objectives through the product lifecycle.
- Optimized data-driven models to support large-scale operations, allowing effective querying on log data for backend business objectives. Troubleshooting performance bottlenecks throughout the pipeline helped refine the growth strategy.

- Developed automation and processes to simplify the CI/CD pipeline (maintenance) for cloud computing environments. This allowed data analysts to deliver solutions to production quicker, ensuring compliance and reducing retooling time to support the growth strategy.
- Collaborated with AWS Cloud Services to render auto-generated, dynamically optimized display ads across platforms, including Amazon DSP, IMDb, and Amazon Video. The implementation of React for front-end development, combined with .NET for backend services, significantly enhanced efficiency, drove advertiser adoption with innovative features, and contributed to the product lifecycle's continuous improvement.

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 using C# and ASP.NET for backend services.
- Developed single-page applications (SPAs) with HTML5, CSS3, C#, JavaScript, and 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 with C# for backend logic.
- Documented learnings from on-call tickets for continuous customer service, merchandising, and established root cause analysis with C#-based automation solutions to improve troubleshooting.

Education

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

Aug 2022