Maruthi Manoj Kolli

📞 716-292-5717 ✉ kollimmanoj@gmail.com 🌐 Portfolio

Summary

Master's graduate in Data Science with experience in Data Engineering, Software Development, and Business Analysis. Experienced in building scalable ETL/ELT solutions across various domains using cloud platforms (Azure, GCP, AWS). Skilled in developing end-to-end data pipelines, integrating CI/CD and software systems. Strong in stakeholder collaboration and driving insights for strategic decisions.

Technical Skills

Programming Languages: Python, Java, R, C#, .Net, Scala
Databases: MySQL, SQL Server, PostGreSQL, Couch Base, Oracle, NoSQL, TSQL
Cloud Technologies: Azure Databricks, Azure Data Factory, Data Lake Storage, AWS
Visualization tools: Tableau, QuickSight , Power BI(including DAX)
Tools/Platform/Frameworks: Git, Excel, Visualstudio, Data Structures, NLP, WYSWIG, Jenkins, Docker,Jupyter Notebook, Flask, Data Analytics, data mining.
ML and Big Data: Regression, Classification, Supervised ML, Apache Spark, Pyspark, Databricks

Certifications

AWS Certified Solutions Architect	May 2024

Professional Experience

State Of Missouri - ITSD Dec 2024 – Present

Enterprise Architect *Missouri*

- Designed and implemented scalable **cloud-based data architectures** using **AWS and Azure**, ensuring high availability, robust security, and **seamless integration** with existing systems to support data-driven decision-making.
- Developed and maintained **database pipeline architectures** to streamline data ingestion, transformation, and accessibility, employing advanced SQL, low-code development, and API integrations for efficient data workflows.
- Translated complex business requirements into optimized data models, **data warehouses**, and **data streams** by leveraging architecture methodologies and **master data management** practices to enhance **data accuracy and accessibility**.
- **Supervised and mentored a team of 6 data professionals**, fostering knowledge-sharing and collaboration to build robust algorithms and develop best practices for transforming raw data into actionable insights.
- Created detailed technical documentation and conducted knowledge-sharing sessions to establish standardized procedures and **ensure compliance with organizational data governance** and application dependency requirements.
- Mentored junior engineers on AWS best practices, CI/CD methodologies, and machine learning operations, fostering a culture of continuous learning and improvement.
- Implemented data validation and monitoring practices using CloudWatch to ensure accuracy and consistency in data processing workflows.

R1RCM Nov 2023 – Oct 2024

Data Engineer *Remote*

- Developed data pipelines in **Azure Data Factory** for **SQL, CSV, JSON**, and **REST API** inputs, integrating edge computing to reduce latency. Enhanced data models, boosting **processing efficiency** by 15% and **scalability** by 8%.
- Optimized **ETL pipelines**, cutting **operational costs** by 10% while ensuring **data accuracy**. Used **Python-based machine learning** models to boost **revenue** and lower **seller contacts**.
- Enhanced **data architecture** with new **metric capabilities** for customer data, streamlining capacity planning and scalability, while applying **project management** practices to ensure efficient implementation.
- Optimized **NoSQL** queries in **Azure Cosmos DB** with **edge computing**, reducing retrieval time and enhancing real-time decision-making.
- Managed code reviews, version control, and deployment using GitBash, ensuring efficient branching, merging, and staging. Applied project management best practices for timely delivery and code quality.
- Built **Tableau dashboards** for inventory planning, using **Agile** to enhance **cross-functional collaboration**, supporting capacity planning and resource allocation.

AMAZON.COM, INC Sep 2022 – Oct 2023

Data Engineer *Seattle, WA*

- Designed a cloud data pipeline to process and validate external data using **AWS Glue, Lambda, S3, Redshift**, and **EC2**, ensuring compliance with industry standards.

- Coordinated with **BI** and **cross-functional teams** to drive data-based decisions, reducing **seller contacts** by **10M** annually, and enhancing features for better product lifecycle and data-driven growth strategies.
- Managed **data operations** across **S3, Athena, Glue, DynamoDB**, and EMR clusters, ensuring compliance and supporting large-scale data solutions throughout the product lifecycle.
- Optimized **data models** for efficient log querying and resolved **on-call** ticket issues, addressing bottlenecks to improve backend operations and refine growth strategies.
- Documentation of learning from On-call tickets for continuous learning and establish root cause analysis.
- Developed **CI/CD automation** to streamline deployments, enabling faster solution delivery, compliance, and reduced retooling time.
- Collaborated with AWS Cloud Services to render auto-generated, dynamically optimized display ads across platforms, including Amazon DSP, IMDb, and Amazon Video.
- Implemented serverless data pipelines using AWS Lambda and Glue, integrating with the Glue Catalog for querying data with Athena and ensuring efficient data management.
- Implemented data transformation tasks using Scala within Spark, assisting with batch processing and data validation tasks. Collaborated with senior engineers to integrate Scala scripts into existing ETL pipelines, enhancing processing efficiency.
- Collaborated with data engineering teams to optimize Scala-based applications, ensuring high-quality data pipelines and reducing processing times by 15%.

TATA CONSULTANCY SERVICES - State Farm Insurance, USA   May 2018 – Sep 2021
Data Engineer/Full Stack Developer   *Hyderabad, India*
- Developed end-to-end **ETL pipelines** in **Azure Data Factory**, handling 300GB daily data, and optimized **Snowflake** for efficient storage and retrieval.
- Boosted **data efficiency** by 20% with **PySpark**, automating workflows via **Azure Databricks** for seamless Python script management.
- Utilized Spark to perform advanced data analytics and processing on large datasets, implementing machine learning algorithms and ensuring high-performance distributed computing.
- Built dynamic web features with **C#, ASP.NET, JavaScript**, and **C# automation** for root cause analysis, improving engagement and troubleshooting.
- Enhanced data processing efficiency by 20% using PySpark for data transformations, ensuring high data quality and integrity with Python libraries.
- Automated data workflows using Azure Databricks, enabling seamless deployment and management of Python scripts for efficient data processing.

Logic Info Tech-systems   May 2017 – May 2018
Data Engineer   *Hyderabad, India*
- Developed a predictive maintenance system leveraging machine learning models to forecast equipment failures, reducing unplanned downtime in manufacturing operations.
- Implemented data pipelines using Azure Data Factory to ingest IoT sensor data from industrial machines into Azure SQL Database and Azure Data Lake for further analysis.
- Utilized Python, PySpark, and Scikit-Learn for data preprocessing, feature engineering, and training machine learning models, including Random Forest and LSTMs for time-series forecasting.
- Deployed AI models using Azure ML and Dockerized APIs for real-time failure predictions, integrating insights into Power BI dashboards for maintenance teams.
- Automated workflows and model retraining pipelines using Databricks and Azure DevOps, ensuring scalable and efficient predictive maintenance operations.

Education

MS in Engineering Science - Data Science, State University of New York at Buffalo.   Aug 2022