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, Hadoop, Pytorch, Apache spark, PySpark, Kafka, Scala
Web Technologies:JavaScript, HTML, CSS, Bootstrap, Typescript
Database Technologies:Microsoft SQL Server, postgres, MySQL, NoSQL, PLSQL, TSQL, Oracle, Redshift, Tensorflow.
Cloud Technologies: AWS, Azure, Google Cloud Platform, Snowflake, Databricks.
Tools/Platform/Frameworks: Git,Github, Excel, Visualstudio, Data visualization - Tableau, Power BI, NLP, WYSWIG, Jenkins, Docker,Numpy, Pandas, Matplotlib, Seaborn, Plotly, Scikit-Learn, ggplot, Data Analytics, data mining.

Certifications

AWS Certified Solutions Architect - Associate **May 2024**

Professional Experience

WORKMATE INC **Nov 2023 – Present**
Data Engineer *Remote*

- Developed robust data pipelines using AWS to handle SQL, CSV, JSON, and REST API inputs; refined data models, achieving a 45% boost in data processing efficiency and 60% 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 NoSQL queries in DyanmoDB 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 Amazon QuickSight 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 clusters.
- Optimizing Data-driven models to support large-scale, to facilitate effective querying on the log data to perform business Objectives at the backend.
- Documentation of learning from On-call tickets for continuous learning and establish root cause analysis.
- 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 ETL data pipeline operations, processing approximately 300GB of daily data on AWS. Utilized Python libraries within PySpark transformations to effectively manipulate and cleanse data, significantly enhancing data quality and integrity.
- Implemented a cloud infrastructure-based data warehouse solution using Amazon Redshift Spectrum, optimizing schema, tables, and views for efficient data storage and retrieval.

- Transformed raw data into user-friendly formats using PySpark, leading to a 20% increase in data processing efficiency and enabling more effective data analysis.
- Utilized Python libraries within PySpark transformations to manipulate and cleanse data effectively, enhancing data quality and integrity.
- Engineered ELT/ETL pipelines for data ingestion and transfer within the Snowflake data store on AWS, utilizing Python and Snowflake SQL for data modeling. Leveraged Snowflake's performance and scalability to meet complex data processing requirements.

Education

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