HARSHA VARDHAN SAI MACHINENI

Data Engineer

+1 (313) 529-4839 |harsha@jobittech.com|LinkedIn|GitHub|Website

SUMMARY

Data Engineer specializing in designing and building data lakes and data warehouses. With 3+ years of experiencein data transformation and data modeling. Providing data architectures that improvise the organizations with actionable insights.Data Engineer with expertise in data architecture and data warehouse design. Experienced in optimizing data storage, data transformation, and creating data solutions that drive business insights. Experienced Data Engineer with a focus on machine learning and AI data pipelines. Proficient in feature engineering, model deployment, and supporting data-driven AI initiatives.

SKILLS

Methodologies: SDLC, Agile, Waterfall

Language: Python, R, SQL, SAS

Cloud Platforms: AWS (Glue, Redshift, EMR, RDS, DynamoDB, Lambda, Data Pipeline, Kinesis, Lake Formation), Azure(Data Factory, Synapse Analytics, Databricks, HDInsight, Cosmos DB, SQL Database, Stream Analytics, Blob Storage) , GCP(AI Platform, AutoML Vision, Vision API, Vertex AI)

Database: SQL Server, MySQL

Big Data Tools : Apache Hadoop, Apache Spark, Apache Hive, Apache HBase, Apache Kafka, Apache NiFi, Apache Flink, MongoDB, Cassandra, Snowflake, Google BigQuery, Apache Airflow, Databricks.

ML Algorithm: Linear Regression, Logistic Regression, Decision Trees, Supervised Learning, Unsupervised Learning, Classification, SVM, Random Forests, Naive Bayes, KNN, K Means

Packages: NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, TensorFlow, Flask, Apache Hadoop, Apache Hive, Apache Airflow,

Al Technology: Generative Al (including GANs, VAEs), Natural Language Processing (NLP), Reinforcement Learning, Transfer Learning, Computer Vision (object detection, image recognition)

Visualization Tools: Tableau, Power BI

Other Tools: Git, MS Excel, AWS

Operating System: Windows, Linux

EXPERIENCE

Data Engineer, Freddie Mac, USA **01/2024-Present.**

- Collaborated with data engineers and operation team to implement the ETL process, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.
- Explored and analyzed the customer-specific features by using Matplotlib in Python and ggplot2 in R.
- Incorporated Python (NumPy, SciPy, pandas, Scikit-learn, seaborn) and R to develop models and algorithms for analytic purposes.
- Developed a scalable infrastructure for processing and analyzing social media data, uncovering insights into consumer behavior, reflecting in a 20% increase in data processing speed.
- Implemented AWS Redshift, extracted, transformed, and loaded data from heterogeneous data sources and destinations, leading to a 25% reduction in data processing time.
- Implemented a disaster recovery strategy for big data applications, ensuring 99.99% uptime and data integrity during unplanned outages.
- Created and applied Natural Language Processing (NLP) models on AWS for automating text classification tasks, leading to a 30% boost in productivity.
- Designed a scalable and fault-tolerant Kafka messaging system, processing 10,000 messages per second with zero data loss.
- Engaged with cross-functional teams of data analysts to gather insights, refine data models, and implement robust data solutions tailored to meet diverse analytical needs and business objectives.

Data Engineer, Trigyn Technologies, India **05/2019-07/2022.**

- Designed a data encryption and security strategy, securing sensitive information and complying with industry security standards.
- Optimized a NoSQL database solution (Cassandra), resulting in a 70% improvement in response times for high-traffic web applications.
- Established a detailed project plan and helped manage the data conversion migration from the legacy system to the target Snowflake database, bringing 30% increment in data accuracy post-migration.
- Advanced a robust error handling and logging mechanism for data pipelines, reducing downtime and improving reliability.
- Prepared AWS Data pipeline to configure data loads from S3 to into Redshift.
- Experience in developing with RDBMS such as SQL Server, MySQL, and NoSQL databases such as MongoDB.
- Executed version control for data models, enhancing collaboration and reducing conflicts in a team of 20 data engineers.
- Settled machine learning models such as Logistic regression, KNN, and Gradient Boosting with Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn in Python, following a 22% improvement in model accuracy.
- Proficient in leveraging Azure services and tools to design, deploy, and manage scalable data solutions, ensuring optimal performance, reliability, and security in cloud-based environments, heading to a 18% increase in system uptime.

PROJECTS

Splitwise Data Analysis , Michigan, USA - March 2024

- Analyzed shared expenses, enhancing financial transparency and equitable distribution among roommates using Python and Pandas.
- Identified discrepancies between total spending and individual contributions, improving accuracy in financial reporting.
- Improved data recording practices and expense-sharing mechanisms through stakeholder discussions and data validation techniques.
- Demonstrated data manipulation and visualization proficiency, fostering a culture of accountability in shared financial management.

End-to-End ML Model Monitoring using Airflow and Docker Michigan, USA – January 2024

- Led an MLOps initiative for loan eligibility ML models, enhancing accuracy by 25% and data integrity.
- Boosted data processing and model training efficiency by 30% using Docker and Airflow.
- Implemented robust Random Forest and Gradient Boosting models for dynamic production environments.

MLOps Deployment in Digital Banking Transformation Michigan, USA - December 2023

- Directed AWS-based ML model deployment, impacting over 200,000 customers' marketing strategies.
- Developed Flask apps, reducing deployment time by 40% via AWS CodePipeline.
- Enhanced deployment efficiency by 35% with automated cluster creation, improving targeted marketing.

Real-Time IoT Data Analysis and Storage Using AWS IoT Device Simulator Michigan, USA - September 2023

- Utilized AWS IoT Core and Kinesis Firehose, improving data flow efficiency by 20%.
- Simulated IoT data management from The O2 Arena, London, demonstrating data handling expertise.
- Employed AWS CDK for effective cloud infrastructure management, showcasing Python skills.

A/B Testing for Machine Learning in Production Environment Michigan, USA - March 2023

- Developed a framework for A/B testing, capable of analyzing 20,000 simulated interactions to potentially enhance application performance by 20%.
- Gained insights on strategic decision-making impact via A/B and A/A testing.
- Utilized Python, Pandas, Stats models, and Matplotlib, enhancing data analytics skills.

Churn Prediction Deployment on AWS Michigan, USA - December 2022

- Created a churn prediction model on AWS, aimed at improving customer retention by 20%.
- Built a Flask and Docker-based scalable cloud solution, managing 500,000+ customer data points.
- Focused on AWS ECS, EC2, and CodePipeline for a 25% more efficient model deployment process.

PowerBI Road Accident Analysis in the UK Michigan, USA - September 2022

- Analyzed UK road accidents using PowerBI, creating interactive dashboards for data-driven safety insights.
- Enhanced data integrity and report accuracy by integrating multiple sources and employing advanced visualization techniques.

EDUCATION

Master of Science in Data Science,
University of Michigan, Dearborn, MI
Bachelor of Technology in Electronics and Communication Engineering,
BMLMunjalUniversity, India

CERTIFICATIONS

AWS Certified Cloud Practitioner, Amazon Web Services **10/2023, Michigan, USA**
AWS Certified Machine Learning - Specialty, Amazon Web Services **11/2023, Michigan, USA**
Azure Data Engineering Associate: Microsoft **03/2024, Michigan, USA**