IDUPULAPATI GAYATHRI SAI YASASWI

Richardson, TX | (945) 527-4904 | yasaswiigsms@gmail.com | LinkedIn |

Skilled in designing scalable ETL pipelines, leveraging cloud platforms, and implementing big data solutions to enable data-driven decision-making. Proven expertise in GCP, AWS, Python, and data warehousing

EDUCATION

The University of Texas at Dallas, Richardson, TX — Aug 2023 - May 25
Master of Science in Computer Science GPA: **3.8/4.0**

Ramaiah Institute of Technology, Bangalore, India — Aug 2017 - May 21
Bachelor of Engineering in Computer Science GPA: **3.8/4.0**

SKILLS

Programming: Python, Java, JavaScript, C, C++, TypeScript
OS: Unix, Linux, Windows
Full-Stack Development: Angular, React, Node.js, Spring Boot, Bootstrap, Flask, HTML, CSS, Restful API
Frameworks/Libraries: Hadoop, Apache, Pyspark, Scrapy, Scikit Learn, PyTorch, Pandas, Numpy, Keras, Tensorflow
Database Management: MySQL, MongoDB, PostgressSQL, SQL, NoSQL
Additional Skills: Git, GCP, Jira, Microservices, Azure, Big Query, UNIX, Docker, Kubernetes, DevOps, AWS

EXPERIENCE

Data Engineer Intern,
Tenet Health Corporation, Dallas, United States — Jun 2024 – Present

- Built and optimized ETL pipelines using Google Cloud Platform (GCP) services, including Big Query, Dataflow, and Cloud Storage.
- Automated file ingestion frameworks in Python to streamline integration of data from multiple sources into Big Query.
- Migrated complex data processes from T-SQL to Big Query, improving data accessibility and analytics for Power BI dashboards.
- Assisted data science teams by developing Python-based tools to ensure seamless integration with GCP analytics solutions.

Data Engineer
Lam Research, Bangalore, India — Jan 2021 - Aug 2023

- Designed and implemented scalable ETL pipelines to process unstructured data, utilizing Python, SQL, and Apache Hadoop.
- Led the migration of legacy systems to cloud-based solutions on AWS and GCP, gaining a 35% improvement in processing speed.
- Built automated data quality monitoring scripts to ensure accuracy and reliability of large-scale data pipelines.
- Developed data ingestion frameworks integrating diverse sources into centralized warehouses using AWS services like Redshift, S3, and Glue.
- Deployed distributed data processing systems with Spark on AWS EMR, significantly enhancing computational efficiency.
- Developed a real-time data visualization platform to monitor system health using Python, Flask, and Angular.
- Engineered RESTful APIs for data integration, reducing backend latency by 20%.
- Designed scalable ETL workflows for data ingestion and transformation, delivering actionable operational reports.
- Created dynamic dashboards in Angular and JavaScript to visualize production system performance metrics.

PROJECTS

NIKE Stock Price Prediction using LSTM, UT Dallas, Richardson, TX — Mar-May 2024

- Implemented LSTM using pyspark and mapreduce to predict the stock prices with efficiency of 96% accuracy.
- Performed a comparative study with GBT regressor and highlighted the model's strength and weakness in terms of efficacy.

Argon DB Engine, UT Dallas, Richardson, TX — Aug - Dec 2023

- Implemented a DB Engine in Python, using file-per-table approach for efficient storage of table files, index files at the OS level.
- The system effortlessly manages DDL, DML, DQL, and SDL command lines, alongside API calls, with an accuracy rate of 99%.

PUBLICATIONS

Football Player Move Prediction, RIT, Bangalore, India

Developed a predictive system using computer vision and CNN frameworks to forecast player movements in video data.

Hybrid Decision Support System to Enhance crop productivity

Designed a hybrid ML system integrating ARIMA and LSTM to forecast crop prices and yields with a 20% performance improvement.

Document Retrieval Through Cover Density Ranking

Created an enhanced information retrieval method, CDR, surpassing VSM with 85% better computational efficiency.