Sai Shreya Peyyala

Email: saishreya.peyyala@gmail.com **Phone:** +91-7981849988 **GitHub:** GitHub **LinkedIn:** LinkedIn

Summary

Data Engineer with 4+ years of experience in designing, building, and optimizing scalable data pipelines and ETL workflows across cloud platforms like GCP, AWS, and Azure. Proficient in Big Data technologies including PySpark, Kafka, and Hadoop, with hands-on expertise in orchestrating data workflows using Apache Airflow and CI/CD tools like Jenkins. Proven ability to drive performance improvements, reduce costs, and deliver data-driven solutions in fast-paced, enterprise environments.

Skills

Programming Languages: C, Python, Shell Scripting

Databases: Oracle, MongoDB

Frameworks: PySpark, Django

Developer Tools: Git, GitHub, CI/CD, Jenkins, Apache Airflow

Cloud Platforms: GCP (GCS, BigQuery, Dataproc, Dataflow), Azure (Data Factory, Databricks), AWS (S3, Lambda, Glue, EMR)

Concepts: ETL, Data Migration, Data Warehousing, Real-Time Data Processing, Machine Learning, Generative AI, HDFS, NLP

Work Experience

Cognizant, Hyderabad Mar 2025 – Present
Data Engineer

- Led the migration of 1000+ ETL workflows from Informatica to Google Cloud Dataflow, resulting in 60–85% cost savings and elimination of legacy license dependencies.
- Built and maintained end-to-end data pipelines using Apache Airflow for orchestration and monitoring.
- **Tools:** SQL, Python, Dataflow, Airflow, GitHub, Kubernetes, Docker, Shell Scripting, Google BigQuery

Tata Consultancy Services, Hyderabad June 2021 – Mar 2025
Data Engineer

- Developed a scalable ETL framework for data migration at PayPal using Python, Google Cloud Storage (GCS), and BigQuery, reducing migration time by 20% and improving scalability by 30%.
- Automated data validation processes by integrating SQL queries into a Rule Execution Framework.
- Migrated legacy Hive-based ETL to PySpark, reducing runtime by up to 70% and cutting execution costs by over 50%.
- Optimized Spark queries, resulting in a performance improvement of up to 70%.
- Designed and deployed CI/CD pipelines with Jenkins, and managed Docker image builds and registries.
- **Tools:** Python, SQL, Dataproc, Google BigQuery, Google Cloud Storage, Jenkins, Docker, PySpark, Machine Learning

Education

Institute of Aeronautical Engineering Aug 2017 – Jun 2021
B.Tech. in Electrical and Electronics Engineering *CGPA: 8.49/10*
Relevant Courses: Data Structures, Algorithms, DBMS, Machine Learning

Project Work

- **Sentiment Analysis on Product Reviews:** Developed a machine learning model to classify product reviews as positive or negative using NLP techniques. Preprocessed text using tokenization, stopword removal, and TF-IDF. Built and compared Naive Bayes and LSTM models. *Tools: Python, scikit-learn, TensorFlow, NLTK; Dataset: Amazon Reviews (Kaggle)*
- **Real-Time Vote Counting and Winner Prediction System:** Built a streaming data pipeline using Apache Kafka and Apache Spark to process real-time vote data. Implemented dynamic winner prediction logic with fault tolerance and scalability. *Tools: Apache Kafka, Apache Spark Streaming, Python*

Awards and Certifications

- **Google Cloud Certified:** Associate Cloud Engineer, Professional Data Engineer
- **Microsoft Certified:** Azure Data Engineer Associate
- Cleared Round 1 of the **TCS CodeVita 2020** global coding contest organized by Tata Consultancy Services