Charan Lokku

📞 469-558-0098 ✉ cxl220029@utdallas.edu in LinkedIn 🐙 Github 🌐 Portfolio

Education

The University of Texas at Dallas — Jan 2023 - Dec 2024
Master of Science in Computer Science — *Richardson, Texas*
- **Relevant Coursework:** Big Data Management & Analytics, Design & Analysis of Algorithms, Database Design, Machine Learning, Web Programming Languages, NLP, Statistical Methods for Data Science

Experience

WinWin Labs — Feb 2025 – Present
Data Engineer — *Remote*
- Developed a Kafka-based Python pipeline processing 1K+ sensor events/sec, improving reliability and reducing false alerts using Faust stream processing.
- Optimized TimescaleDB with hypertable partitioning for 50% faster queries and deployed the system using Docker, cutting setup time by 60% for edge and cloud environments.
- Built Grafana dashboards with Slack/email alerts for real-time monitoring of water quality, pressure, and temperature anomalies.

KPMG (Forage) — Apr 2021 – Jun 2021
Data Analytics Consulting Intern — *Remote*
- Analyzed 10,000+ bicycle retail transactions using Python and SQL to identify customer segments and purchasing trends, improving targeted marketing strategies
- Developed a customer churn prediction model with 70% accuracy and created RFM segments to enhance retention efforts
- Designed interactive dashboards in Tableau to visualize sales performance and product preferences across regions

Projects

IMDb Data Integration Pipeline | *Python, MySQL, Metabase, Docker* — Jan 2025 – Feb 2025
- Extracted IMDb Top 250 movies data via API using Python and transformed it into a structured format with Pandas.
- Loaded cleaned data into MySQL using SQLAlchemy, ensuring automated table updates.
- Deployed Metabase via Docker for data visualization, enabling interactive querying and reporting.

Stock Price Prediction | *Python, PySpark, ARIMA, Matplotlib, Seaborn* — Apr 2024 – May 2024
- Processed large-scale stock data with PySpark, performing data cleaning, stationarity checks (ADF test), and time-series decomposition for ARIMA modeling.
- Built and optimized ARIMA models by tuning p, d, q parameters, validated accuracy using MAE and RMSE, and visualized trends and predictions with Matplotlib and Seaborn, leveraging Spark's distributed computing.

Real-Time News Data Pipeline | *Kafka, PySpark, Elasticsearch, Logstash, Kibana* — Mar 2024 – April 2024
- Built a Kafka pipeline to ingest and filter real-time news data from NewsAPI by keywords, sources, and language, enabling immediate processing.
- Processed streaming data with PySpark and Spark Streaming, performing aggregation and Named Entity Recognition to extract key entities and trends.
- Stored results in Elasticsearch via Logstash and created Kibana dashboards to visualize entity mentions and trends at 15-60 minute intervals.

NLP Chatbot with Web Crawling | *Python, Beautiful Soup, NLTK, Scikit-learn, Pandas* — Feb 2024 – Mar 2024
- Developed a Python-based web crawler using Beautiful Soup and Requests to extract structured data, then processed text with NLTK and TF-IDF to build a searchable knowledge base stored as a pickle file.
- Built a rule-based chatbot with personalized responses using the processed data, implementing NLP tokenization and fallback mechanisms for unrecognized queries to improve user engagement.

Technical Skills

Programming Languages: Python, SQL, R
Tools & Technologies: Apache Spark, Hadoop, Kafka, Docker, NumPy, Pandas, Databricks, AWS, MySQL, MongoDB, Tableau, Power BI, Git, Apache Airflow, Snowflake, Matplotlib, Seaborn, Metabase, Apache Superset
Concepts: Version Control, NoSQL Databases, Big Data Skills, Real-Time Processing, ETL Workflows, Data Warehousing