Ankur.

Data Engineer

ankur3103verma@gmail.com/ ankur.sv@yahoo.com
9676948240/6304368475
Hyderabad, Telangana

GitHub: https://github.com/Ankur-ipynb | LinkedIn: https://www.linkedin.com/in/ankur-verma-87047ba3/

Proactive and detail-oriented Data Engineer in transforming raw data into actionable insights through data ingestion, transformations, visualization. Previously, as a Systems Engineer to optimize system performance, workflows, and enhance operational efficiency. Proficient in Python, PySpark with ability to bridge technical expertise and business acumen to drive strategic decision-making and process improvements, having total exp. 6+ years with 3+ as Data Engineer.

Project Experience

Inventory Anomaly Detection System (PySpark)

Version 1.0: Dashboard for Anomaly Detection and Notifications

GitHub : https://github.com/Ankur-ipynb/Inventory_AnomalyDetectionV1.git

- Created an interactive dashboard to visualize live inventory metrics, ML predictions, and anomalies, boosting decision-making for managers across multiple stores.
- Engineered a pipeline using Pulsar for real-time ingestion and Spark for distributed processing ~20 batched records/run, delivering key metrics like stock value.
- Implemented ARIMA, XGBoost, Prophet, and Isolation Forest models predicting stock levels and detect anomalies, enhancing inventory accuracy and efficiency.
- Developed app.py to seamlessly integrate data ingestion, processing, ML execution, and anomaly detection, accessible at http://localhost:<port>.

Flight Delay Prediction (PySpark)

Version 2.0: Delay analysis and Prediction

GitHub : https://github.com/Ankur-ipynb/Flight_Delay_PredictionV2.git

- Developed a V2 of scalable data pipeline using PySpark to analyze flight patterns, airport performance, and plane efficiency.
- Implemented Spark ML for predictive modeling, optimizing flight predictions improvement in accuracy.
- Utilized feature engineering techniques to enhance model performance, improving real-time decision-making for airline operations.
- Integrated Optimized data processing speed by leveraging distributed computing in Spark, reducing query times.

Metro Network Analysis (Python)

Version 1.0: Geospatial based analysis

GitHub : https://github.com/Ankur-ipynb/Metro_Network_Analysis.git

- Created and analyzed geospatial data for Metro Network along with Metro stations using the python based visualizations.

Professional Experience

Formerly System Engineer (01/2019 – 07/2024), currently IT Analyst (08/2024 – Present)

- Over 6 years of total experience in the IT industry, with the last 3+ years specializing as a **Data Engineer**.
- Proficient in **Python**, **PySpark**, and distributed data processing, with a proven ability to bridge technical expertise and business acumen.
- Led and contributed to data pipeline projects, enabling strategic decision-making and driving process improvements across teams.
- Strong grasp of data engineering best practices, data modeling, ETL workflows, and real-time data processing.
- Experience collaborating with cross-functional stakeholders to translate business requirements into scalable technical solutions.

Technical Skills

- **Data Analytics**: Python, PySpark and DB
- **Big Data**: Apache Spark, Apache Pulsar, Apache Flink(basic) and Apache Kafka
- **Cloud (Basic)**: AWS, APIM, DataBricks
- **Visualization:** Python based libraries
- **Programming:** Python, Jupyter Notebook, CMD(basics)
- **Other:** Distributed computing, Data Modeling, ETL
- **DBs:** SQLite (lite computational)

Education

- **Mechanical Engineer, 2018**
  B.Tech
  JNTUH

Certifications

- Scientific Computing with Python
- Data Analyst using Python

Additional Skills

- Problem-Solving
- Analytical Thinking
- Communication and Collaboration
- Time Management
- Attention to Detail