DHYAN CHANDRA PANDEY

Data Engineer

📞 6386185248
🔗 https://www.linkedin.com/in/dhyan123/
📍 Hyderabad
@ dhyanchandrapandey75@gmail.com
🔗 https://github.com/dhyan-chandra
⭐ https://leetcode.com/Dhyan_Chandra_Pandey/

SUMMARY

Passionate Data Engineer with **2.7 years** specializing in BigData, ETL, PySpark, and SQL. Expertise in designing and optimizing data pipelines, transforming raw data into actionable insights.

EXPERIENCE

Data Engineer
Optum - Futurense Client
📅 09/2022 - Present   📍 Hyderabad

Tools used: **SQL, Python, Apache PySpark, Spark SQL, Airflow**

- Improved the performance by automating the job running process by **53%** by using efficient **python scripting** skills.
- Experience in developing cloud-based solution using Python, Shell Scripting, **Apache PySpark, Apache Kafka, Apache Airflow** etc.
- Optimized existing Spark scripts to run faster.
- Worked on **SQL Stored procedures**, window function, Sub queries.
- Worked on python to solve scenarios and to analyze data using Pandas, Matplotlib.
- Generated insights and visualizations utilizing **Power BI**.

Data Specialist
Futurense Technology
📅 07/2022 - Present   📍 Bangalore

Tools used: **Python, SQL, Shell Scripting, Apache Hadoop, PySpark**

- Got hands-on training in Python, SQL, Shell Scripting and Big Data tools and technology with hands on **Data Pipeline** Flow.
- Studied Data Stage, **Hadoop**, Hive, PySpark and Spark SQL.
- Data Structure & Algorithms.

Support Data Engineer
Wipro Enterprises - PMS Client
📅 08/2021 - 07/2022   📍 Bangalore

Tools used: **Python, SQL, Excel, Power BI**

- Increased **revenue by 20%** by successful implementation of **Salesforce** for Wipro lighting.
- Strong knowledge of SQL and proficient in writing SQL queries.
- Experience in developing **python** code blocks required for data analysis.
- Generated insights and visualizations utilizing Power BI.

PROJECTS

Twitter Data Pipeline Using Airflow
📅 11/2023
- GitHub Link
- Developed Airflow dag for Twitter Data Pipeline Flow.
- Connecting to the **Twitter API** to collect real-time Twitter data.
- **Amazon EC2** instances utilize for running the Apache Kafka clusters.
- Amazon S3 is used as the storage solution for the processed Twitter data.

Real-Time Stock Market Data using Kafka
📅 08/2022
- GitHub Link
- Implemented **Apache Kafka** for real-time data streaming, contributing to a responsive and low-latency stock market data processing system.
- Developed end-to-end data pipeline utilizing **Amazon S3** for storage.

SKILLS

Python   SQL   Apache Pyspark

Apache Hadoop   Apache Kafka

Apache Airflow   Shell Scripting

MongoDB   Data Stage (ETL)

Snowflake   Git   Postman   Jenkins

DSA   AWS S3   Amazon EC2

EDUCATION

PG Diploma in Data Engineering
IIT Jodhpur, Rajasthan
📅 09/2023 - 12/2023

M.Tech - Computer Science & Engineering
SHEAT Institute of Technology, Babatpur
📅 2020 - 2022

CGPA **8.5** / 10

B.Tech - Computer Science & Engineering
Ashoka Institute Of Technology, Varanasi
📅 2016 - 2020

CGPA **8.0** / 10

PUBLICATIONS

Product Recommendation Engine Using K Means Clustering & Collaborative Filtering
IJSART
🔗 https://bit.ly/4993BTJ

AWARDS

💎 **Achieved Team Award for Successful Implementation of Salesforce**

CERTIFICATION

ChatGPT Prompt Engineering for Developers
DeepLearning.AI, July 2023

Data Science - Internship Certificate
Internshala - July 2019