Rohit Prasad

Technology Specialist - Big Data, Azure

Highly Motivated, enthusiastic developer experienced in Big Data and its technologies such as Hadoop, Spark, Hive as well as ETL & DW concepts and Azure Cloud.

Rohit.prasad95@outlook.com ✉

9007398029 📱

Bangalore, India 📍

rohit-blogs.herokuapp.com/ 🌐

linkedin.com/in/rohit-pr 🔗

github.com/RohitPr ⭘

EDUCATION

Kuvempu University
B.Sc(Information Technology)
03/2012 - 09/2015

PROJECTS

Covid19 World Data Visualization(Batch)

- Real World project create to visualize the impact of Covid in most countries.
- At first, data was collected as a CSV data and uploaded to HDFS Cluster. Then the data was imported to Spark Notebook and data was aggregated to transform the data for our downstream source i.e. MySQL DB. Finally, the datasource was visualized using Grafana through MySQL database.
- GitHub : https://github.com/RohitPr/Data-Engineering-Projects/tree/main/Covid19_SparkBatch

Marvel Characters(Spark SQL)

- Project based on Marvel's dataset of their character and their appearances in a comic. The objective of the program was to locate the character who appeared the most and the least in all the Marvel's comic.
- Initially, we load a txt file with Marvel's Dataset of characters who appeared with other characters in a comic. Then we aggregate and filter the character's ID with most count and least count in the file.
- GitHub: https://github.com/RohitPr/Data-Engineering-Projects/tree/main/Marvel_Connection_SparkSQL

Spotify ETL Project

- Real world project to collect User's spotify songs and store them in SQL DB.
- At first, we make an API request to collect the User's Songs data from the Spotify account using Spotify's Token. Then the data is transformed to collect the song details and store only the unique songs in SQLite DB. Also, the project has been automated using Apache Airflow to run once everyday.
- GitHub: https://github.com/RohitPr/Data-Engineering-Projects/tree/main/Spotify_ETL_Airflow

SKILLS

Python · SQL · Hadoop · Apache Spark · Linux · Apache Hive · Git · DBMS · DSA · Apache Airflow · ETL & DW Modeling · AWS · MS Azure · MySQL · Cassandra · Grafana · Pandas & Numpy · REST API · Docker · OOP

ACHIEVEMENTS

Apache Databricks Spark Developer certification
Apache Databricks in-house Spark and Databricks learning Curriculum for their Lakehouse and Spark 3.0 features

Microsoft Azure Fundamentals & Azure Data Fundamentals Certification
MS Certification for knowledge of all Azure services and relational and non-relational data, and different types of data workloads such as transactional or analytical

100 days of Python
Python Course on Udemy by Dr. Angela Yu based on the 100 days of Code challenge

WORK HISTORY

Telligent Support LLP (09/2020 - Present)
Customer Support Representative

Concentrix (12/2017 - 06/2020)
Customer Support Representative

Wipro BPS (12/2015 - 07/2017)
Customer Support Representative