Usman Ghani Mughal

Data Enthusiast & Software Engineer

I'm an enthusiastic and diligent software technology professional. Having worked on a variety of projects around various domains equips me with the ability to adapt, enhance and optimize towards quality solution delivery. I'm an equally exceptional performer under pressure with core focus at adhering timelines.
https://usman-ghani-mughal.000webhostapp.com

osmanghanimughal@gmail.com

92-314-5270917

usman-ghani-mughal.000webhostapp.com

linkedin.com/in/usman-ghani-a50488142

github.com/Usman-Ghani-Mughal

stackoverflow.com/users/10096169/usman-ghani-mughal

WORK EXPERIENCE

Senior Data Engineer
Afiniti
11/2022 - Present

Achievements/Tasks
- Creates Data Pipelines for prepareing US Demographics Data (Python | Airflow | Parallel Processing | REST API | MongoDB)
- Trainning Junior resources in order to understand the existing processes

Big Data Engineer
The Entertainer
05/2022 - 10/2022

Achievements/Tasks
- Build ETL Pipelines using (Data Factory | Python | MySQL | Azure Synapse Analytics | Apache Spark)
- Build Data Lake in AWS S3 and wrote ELT processes to integrate data from multiple sources. (Apache Nifi | Kafka | Hadoop | Hive)
- Build Analytics Pipelines using PySpark.(AWS DataBricks | Spark | Amazon EMR | SalesForce)
- Perform Data Modeling for DWH on Azure Synapse Analytics.
- Build scripts to automate the process of data quality checks.(Python)
- Build AWS Lambda function which runs when another Data Bricks job pushed csv file into S3 bucket. (Python, AWS lambda)

Data Engineer
Afiniti
04/2021 - 04/2022   *Islamabad*

Achievements/Tasks
- Develop and Deploy Data Pipeline for multiple clients e.g. AT&T, T-Mobile, virgin media in US and UK.(AirFlow | MySQL | Mongo | AWS RedShift | Nifi)
- ETL processes | Stored procedure | Feature engineering | Data Modeling.(MongoDB, MySQL | DWH | Nifi | Spark)
- Build and optimize processes based on Multiprocess and Multithread in order to improve the efficiency of a process in controlled environment so we can manage QPS and use resources efficiently. (python | OS)
- Web Scraping Engines for acquiring publicly available data from multiple sources without getting blocked with acceptable QPS. (python)
- Data Modeling and Data warehouse pipelines (MySQL | Redshift)

EDUCATION

Bachelors in Computer Science
COMSATS University Islamabad
02/2017 - 02/2021   *GPA - 3.14*

SKILLS

- Apache (Nifi | Kaka | Druid | Airflow | Spark | Hadoop | Hive)
- Python | Java | Scala | Java Script | C++
- Avro | Parquet
- MySQL | PostgreSQL | SQL Server
- MongoDB | Cassandra
- Data Modeling(OLTP | OLAP)
- DataWareHouse | DataLake
- Microsoft Azure (Data Factory | Data Bricks)
- Docker | GIT
- AWS (Redshift | S3 | EC2)
- SQL (UDF | Stored Procedures)
- Data Manipulation (Numpy | Pandas | Matplotlib | Seaborn)
- ML | DL (Mlxtend | Scikit learn | Keras)
- Data Pipelines
- Rest API (Node | Python)
- Scraping (Scrapy | requests)

PROJECTS

Academic
- News Classification & Recommendation Application(python | Scrapy | Aprori | Random Forest | JavaScript | MERN)
- Data Modeling (PostgreSQL | Apache Cassandra | Redshift)
- Data Warehouse(ETL pipeline| S3 | Redshift)
- Data Lake(Spark | S3)
- ETL pipeline & Data Modeling (Nifi | Airflow)
- Data Mining Tool (Python/ML)

Personal
- Pakistan E-Market website (MERN)
- Game of Life(C++)
- Network Traffic Analyzer (Python)
- Product Affinity Analysis using Apriori algorithm
- Zameen.com Scraping Engine

Free Lance
- Stock (Time Series) forecasting using deep learning (LSTM)
- Heart Disease Classification using ML (Random Forest)
- Retail Churn Prediction using ML (Random Forest)
- Extract data from Pdf (retail invoices) using python
- C++ program optimization based on memory management

CERTIFICATES
- **Data Engineering (Nanodegree - udacity)**
- **Linkedin Assessments:** C | C++ | OOP | Java | Python | JavaScript | Machine Learning | MySQL | MongoDB | Node.js| REST APIs | PHP
- **Hackerrank:** SQL(Intermediate) | Problem Solving (Basic | 3 Stars)
- *https://usman-ghani-mughal.000webhostapp.com*
- **Introduction to Big Data with Spark and Hadoop** (Coursera | IBM)
- **ETL and Data Pipelines with Shell, Airflow and Kafka** (Coursera | IBM)
- **Advanced Python** (Linkedin)