Mehroos Ali

Github: github.com/mehroosali

Email: mehroos.ali@utdallas.com
Mobile: +1-214-940-7050

EDUCATION

University of Texas at Dallas   Richardson, Texas
Masters of Science - Computer Science   *August 2021 - Present*
Courses: *Operating Systems, Data Structures, Computer Architecture, Discrete Structures*

Motilal Nehru National Institute of Technology Allahabad   Prayagraj, India
Bachelors of Technology - Mechanical Engineering   *July 2014 - July 2018*
Courses: *Software Engineering, Computer Programming, Numerical methods and Statistical Techniques*

SKILLS SUMMARY

- **Languages**: SQL, JAVA, Python
- **Big Data**: Hadoop, PySpark, Hive, Sqoop, Nifi
- **CI Tools**: Maven, GIT, Jenkins, Oozie, Airflow
- **Cloud Services**: GCP - BigQuery, GCS, Dataproc, Cloud Composer, AWS - S3,Lambda, SNS, SQS, Cloudformation
- **Databases**: Oracle, MySQL
- **Platforms**: Linux, Windows

EXPERIENCE

Onward Technologies Limited   Chennai, India
Data Engineer   *Jan 2021 - Aug 2021*

- Migrated existing spark jobs from on-premise HDP to GCP.
- Developed pySpark code with custom business logic to transform data from google cloud storage to persist in bigQuery tables for ad-hoc analysis and visualization by the business intelligence team.
- Designed ETL pipeline on Cloud Composer (Airfow) for daily data ingestion from google cloud storage to BigQuery.
- Improved runtime of slow-running spark jobs by 60% by optimizing joins.

Cognizant Technology Solutions India Pvt Ltd   Chennai, India
Data Engineer   *Nov 2018 - Jan 2021*

- Used Sqoop to extract and load incremental and non-incremental data from RDBMS systems into Hadoop.
- Converted some existing sqoop, hive jobs to SparkSQL applications to read data from Oracle using JDBC and write it to hive tables.
- Written hive queries to parse the raw data and store the refined data in partitioned and bucketed tables.
- Designed Nifi workflows for data ingestion from various sources such as RDMS, REST API, Kafka topic, etc.
- Developed shell scripts for dynamic partitions adding to hive stage table, verifying JSON schema change of source files, and verifying duplicate files in the source location.
- Worked with different file formats like JSON, AVRO and parquet along with different compression techniques.
- Developed Lambda function in NodeJS and automated its deployment to AWS via Jenkins.
- Written CloudFormation scripts to automate the creation and configuration of AWS resources such as SNS, SQS, Lambda, CloudWatch, DynamoDB, KMS, and IAM.

PROJECTS

- **Twitter Data Sentiment Analysis (Nifi, Kafka Spark, Elastic Search, Kibana, NLP)**: pulls live Twitter data using Nifi getTwitter processor and pushes to Kafka topic which is then consumed by a Spark Streaming application where basic sentiment analysis is performed and the final result is stored in elastic search for visualization using Kibana.
- **Realtime Customer Viewership Analysis (Spark, Hive, Kafka, Nifi)**: data pipeline using the lambda architecture is created for unification and consolidation of real-time customer web events, weblogs, and profile data into a hive warehouse for adhoc analysis.

CERTIFICATIONS

- Microsoft Certified: Azure Fundamentals (AZ-900)
- Microsoft Certified: Azure Data Fundamentals (DP-900)
- Microsoft Certified: Azure AI Fundamentals (AI-900)
- Astronomer Certified: Airflow Fundamentals

HONORS AND AWARDS

- Received the Star Team Performer award in a project team of 20 employees in Cognizant for my annual performance.