

# Syeda Marium

## Faheem

### Contact

Email: [mariums82@gmail.com](mailto:mariums82@gmail.com)

Github: @[mariumfaheem](https://github.com/mariumfaheem)

Medium: @[MariumFaheem](https://medium.com/@MariumFaheem)

LinkedIn : @[syeda-marium-faheem](https://www.linkedin.com/in/syeda-marium-faheem)

Phone No : 03131102920

### Skill Highlights

- Creative design
- Strong decision maker
- Complex problem solver
- Data Structures
- Database
- System Analysis & Design
- Networking

### Cloud Skills

- AWS RDS
- AWS Redshift
- AWS Data Migration Service
- AWS Simple Storage Service (S3)
- AWS Glue Data Catalog
- AWS Athena
- AWS Lake Formation
- AWS Elastic MapReduce (EMR)
- AWS CodeCommit & CodePipeline.
- AWS EC2
- DynamoDB
- Amazon Quicksight
- AWS EKS
- EMR on EKS
- AWS IAM
- AWS ECR
- AWS SSO
- AWS VPC & Networking
- Set up & configure cloud env on GCP
- Perform Foundational Infrastructure task in GCP
- Create and manage cloud resources on GCP
- Automating Infrastructure with Terraform

### Summary

Detail-oriented Data Engineer Providing Flawless Programming solutions to the IT Industry. Having a degree in computer science and two years of hands-on coding experience I love working on big data pipelines to provide key insights to the engineering stakeholders for keeping a check on their customers' performance. Dealing with large data sets in complex data warehouse environments is my passion. I believe in helping business and engineering stakeholders determine what performance metrics they should be tracking.

### Work Experience

#### Bazaar Technologies

01-01-2022 - present

#### Data Engineer II

My responsibilities includes

- Work on EMR on EKS with Apache Hudi
- Create airflow DAGs using python to schedule tasks and submit on Apache Spark using Apache Livy for data ingestion and data transformations.
- Work on building/operating highly available, distributed systems for data extraction, ingestion, and processing of large data sets.
- Manage Multi node spark cluster which is setup on multiple AWS EC2 machine using docker containers.
- Maintain dockerized airflow and spark container on AWS machine
- Work on continuous integration and continuous deployment CI/CD using CodePipeline.
- Work with stakeholders including the Executive, Product, Data and Design teams to assist with data-related technical issues and support their data infrastructure needs. Keep our data separated and secure across national boundaries through multiple data centers and AWS regions.
- Design, implement, and maintain a big data pipeline and analytical environment using third-party reporting tools, modeling metadata, and building reports and dashboards.

#### Data Engineer I

26-08-2020 - 30-12-2021

My responsibilities includes

- Designing of data road map
- Development of product strategy using valuable data
- Development of data ingestion pipelines and ETL processing, including low latency data acquisition and stream processing
- Implementation of Kafka for event streaming
- Designing and End to End Implementation of Bazaar Data Lakes & Meshes
- End to End Data warehouse on AWS Redshift & AWS RDS (MySQL).
- ETL data pipelines with Glue for Data Warehouses
- End to End Lake-house Architecture for data availability
- Kubernetes cluster and various application deployments with helm charts
- Data transformation with Python and PySpark
- Building Bazaar Analytical Platform with k8s , Hadoop and open sources tools
- Internal tools Deployments , management & configuration such as tableau.
- Gathering datasets, Preprocessing, Data Analysis and Features co-relation between data points
- Optimisation of data delivery and re-designing infrastructure for greater scalability
- Assist DBA for database-level performance analysis and troubleshooting
- Delivering training, technical support, and functional support

#### Pakistan Meteorological Department

19-08-19 – 20-09-19

#### System Engineer Intern

Knowledge about function of Pakistan metrological department  
Hands on PMD various unit pertaining to the subject

## Big Data Skills

- Elastic Map Reduce - Hadoop.
- Apache hudi
- Apache Arvo
- Apache Kafka
- Apache Trino
- Apache Airbyte
- Apache Airflow
- Apache Superset
- Apache Spark (PySpark)
- MySQL
- Postgres
- Hive
- HBase
- Cassandra
- MongoDB
- Spark Streaming

## Languages & Frameworks

- Python
- Scala
- C++
- JavaScript
- Django Web Framework
- Flask Web Framework
- DS and ML libs (Pandas ,Numpy , SKlearn,Seaborn )

## DevOps Skills

- Terraform
- Helm
- Kubernetes
- CI Pipelines

## Activities & Achievements

- PM Kamiyab Jawan Scholarship
- PM Laptop for high GPA
- Managing Pakistan's Data Team Community

## Languages

- Urdu
- English

## Education

### BE in Computer and Information System

Oct 2020

NED University of Engineering and Technology, Karachi  
Department of Computer Engineering  
CGPA: 3.747

### Intermediate in Pre-Engineering

July 2016

BAMM PECHS Government College, Karachi.  
Grade: A+

### Matriculation in Science

July 2014

Metropolis Education System  
Grade: A+

## Semester and Freelance Projects

- AI based Packets Inspection for Anomaly Detection -7th Semester
- Integration of MTD in SDN Architecture Final Year Project
- Royal Rental Car Website Using Flask Framework with Stripe and Twilio API's - Freelance Project

## Trainings and Certifications

- Data Engineering Foundations Specialization from Coursera
- IBM Data Engineering Professional Certificate from Coursera
- AWS Fundamentals Specialization from Coursera
- Data Engineer Career Track from Datacamp
- Database Management Essentials from Coursera
- Big Data on Amazon web services (AWS) from Udemy.
- Big Data on AWS: The Big Picture from pluralsight
- Data Warehousing with Amazon Redshift from pluralsight
- Prime Minister Kamiyab Jawan CCNA
- CCNA from Cisco Network Academy
- Programming Essentials in Python by CISCO
- Linux Essentials Professional Development Certificate from Linux Professional Institute (LPI).
- SDN crash course (Openflow,Mininet) from Udemy.
- Introduction to SDN and Openflow from Udemy.
- Python Django Full Stack Developer from Udemy.
- Python Flask Framework from Udemy.
- IBM Data Science Professional In PROGRESS from Coursera
- Big Data Specialization In PROGRESS from Coursera
- Modern Big Data Analysis IN PROGRESS from courser