

Faraz Mazhar

csfarazmazhar@gmail.com

[LinkedIn](#) — [Github](#)

Experience Summary

With more than half a decade of professional experience, I've been designing, developing, and delivering Data Engineering solutions ranging from creating standalone ETL scripts to complex frameworks on clouds like AWS and Azure. Writing clean, reusable, and well-documented code one project at a time.

Technical Skills

Cloud Platform:	AWS Cloud IAM, VPC, CLI, Config, S3, KMS, EC2, Route53, Lambda, API Gateway, System Manager, Secrets Manager, CloudWatch, CodeCommit, CodeBuild, CodePipeline, Elastic MapReduce (EMR), AWS Glue, Athena, QuickSight, RDS, SNS, Step Functions, Transfer for SFTP, Cloudformation, DynamoDB, DMS, SCT, CloudEndure
	Azure Cloud CLI, Active Directory, Resource Groups, Virtual Machines, Functions, Virtual Networks, Storage, Azure SQL Database, App Service, EventHub (Kafka), Data Factory
Big Data:	Databricks (Lakehouse), Apache Spark (Delta Lake), Kafka, Hadoop, Hive, Yarn, HDFS, Zeppelin, Sqoop, Airflow, Trino/Presto.
Languages:	Bash, SQL, PowerShell, and Python
Databases:	MSSQL (Azure SQL Server), PostgreSQL (RDS and Redshift), MongoDB, DynamoDB
OS:	Linux, Windows

Experience

Systems Limited – Lahore, Pakistan (March 2021 - Present)

Role *Principal Consultant*

Technologies Python (Django, Pandas), Bash, Airflow, PySpark (Delta Lake), Azure Cloud, Databricks

- Data Framework (Python, Pandas, PySpark, Delta Lake)
 - Responsible as a Lead Developer for the framework which can perform trivial and nontrivial data engineering tasks like acquisition, ingestion as well as executing custom scripts, and analytical processes.
 - Designed and developed the framework to be modular.
 - Used the framework to streamline the creation of Data Pipelines.
 - Data Modeling (Delta Lake and SQL Server)
 - Worked on data modeling the data/delta lake and created Canonical Model.
 - Designed and optimized the metadata stores and data stores for the SQL Server.
 - Orchestration
 - Performed orchestration via Databricks Workflows, Airflow, and Azure Data Factory.
 - Designed and orchestrated data pipelines.
 - System Design
 - Responsible for solving design challenges for the system.
 - Designed and developed REST API for Integration services that allow outside applications to talk to the system.
 - Designed and developed GraphQL API for Semantic Integration.
 - Developed a query builder to build SQL queries for a variety of dialects based on a grid model.
-

Northbay Solutions – Lahore, Pakistan (August 2018 - March 2021)

Role *Senior Software Engineer*

Technologies Python (Flask, Pandas), Powershell, Bash, PySpark (Delta Lake), AWS Cloud

- Data Engineering
 - Worked on a project of migrating [SAS](#) code into PySpark code.
 - Developed a dependency package for AWS Glue ETL jobs.
 - Create ETL scripts using PySpark for ingesting and processing big data.
 - Worked on creating a Data Lake framework.
 - Created utilities for accessing DynamoDB.
 - Worked a microservices based project creating a platform for reviewing certain documents.
 - Created a service responsible for triggering the process when a file is placed on S3 using Lambda and Cloudwatch.
 - Created API using API Gateway for Kinesis Data Stream.
- Cloud Migration
 - Created a framework to automate cloud migration.
- Orchestration
 - Created dynamically parallel State Machines.
- Infrastructure-as-Code
 - Created Cloudformation and deployment scripts.
- Others
 - Created a customizable XML parser.
 - Created multiple APIs using Flask.

Mentor Graphics – Lahore, Pakistan (August 2018 - March 2021)

Role *Intern*

Technologies C/C++, Python, Bash, Nucleus RTOS

- Interned as QA for the Nucleus Real-time Operating System.

Academics

BS(Hons.) Computer Science, Punjab University College of Information Technology – (2014 - 2018)

Certifications

- AWS Certified Developer - Associate – (2018 - 2021)