BARKAAT ALI

Empowering Data with Microsoft Certified Expertise in AI and Analytics

@ Barkaatali199@gmail.com Mobile: 0308-5616873 in barkaat-ali

🕠 barkaat-ali

Lahore

EXPERIENCE

Data & Analytics Engineer Ascend Analytics

December 2024 - Present

Lahore, Hybrid

- Streamlined the modernization of data warehouse infrastructure (AS/400) to Azure Cloud, shifting infrastructure from CAPEX to OPEX and reducing costs by 80%.
- Re-architected the Enterprise Data Model (EDM), transitioning from a complex fact constellation schema to a star schema, significantly improving query performance and simplifying analytics.
- Designed and implemented an audit logging system in Azure SQL to monitor and analyze activity, enhancing data governance and compliance.
- Consolidated 400+ datasets in Azure Data Factory into 37 reusable, parameterized datasets, streamlining the ETL architecture and eliminating redundant data sources.
- Reduced the number of pipelines from 700 to 250 by developing dynamic, metadata-driven pipelines, greatly improving maintainability and scalability.

Data Engineer

Dotlabs

☐ June 2024 - Present

Lahore, Onsite

Project: Hopi Housing Service

- Developed a scalable data pipeline to capture real-time changes for the Hopi Housing Service, enabling efficient data processing and insightful reporting.
- Transformed JSON streaming data to a structured format using Kinesis Firehose, storing the results in Parquet, reducing query costs by **70%**.
- Implemented a DynamoDB trigger using a Lambda function to capture changes and store them in S3, sending real-time updates to Slack via SNS.
- Built an interactive KPI dashboard in Amazon QuickSight to present data insights to stakeholders for better decision-making.
- Engineered complex data transformations using AWS Glue, optimizing the processing logic and storing results in Parquet, reducing memory usage by 40%.
- Automated weekly crawlers in AWS Glue to handle evolving schemas, ensuring the data catalog remains agile and up-to-date with the latest data
- Leveraged Athena for querying the results in the data catalog and integrated with QuickSight for advanced data visualizations and reporting.

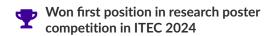
Project: Data Warehousing for Sunderstorm Cannabis Company

- Developed a scalable data pipeline for Sunderstorm, using AWS Glue to process data from APIs and store it in Amazon S3, reducing manual effort by 40%.
- Implemented SCD Type 2 to track historical changes, improving reporting accuracy by 25%.
- Designed a Redshift data warehouse, optimizing the data model for faster querying and reporting, improving query performance by 30%.

SKILLS

Python/C++/C# SQL/NoSQL **AWS Redshift** AWS Glue AWS Lambda **AWS Kinesis** AWS Athena AWS EC2 Azure Data Factory Azure SQL Database Azure Synapse Analytics Azure Databricks **Azure Stream Analytics** Azure HDInsight Azure ML IAM Azure Data Lake Storage Azure Blob Storage Fabric Analytics Engineer Power BI Data Warehousing Airflow ETL Machine Learning Deep Learning Spark Hadoop Hive **MLOps** Tableau **Feature Engineering** Tableau Prep Excel **EDA** LLM Fine-Tuning **Hypothesis Testing** NoSQL Databases **SQL** Server Data Integration Data Modeling PowerShell Snowflake Hubspot **Databricks** Matplotlib/Seaborn Grafana

ACHIEVEMENTS



Among the top 5 students in the class to receive a laptop under Youth Scheme.

CERTIFICATION

Certified Data Scientist Associate click here to view

Microsoft Certified: Azure Data Engineer Associate click here to view

Microsoft Certified: Azure Al Engineer Associate

click here to view

Microsoft Certified: Fabric Analytics
Engineer Associate
click here to view

- Built and optimized stored procedures for data management in Redshift, reducing manual intervention by 30%.
- Created a mechanism for data management in Redshift, improving data accuracy and reducing inconsistencies by 20%.
- Engineered data transformations using AWS Glue and Parquet format, cutting query costs by 70% and boosting efficiency by 35%.
- Automated data ingestion and transformation, reducing manual handling by 40% and enabling real-time updates.
- Leveraged AWS Redshift for data-driven reports, enabling stakeholders to make decisions 50% faster.

Data Scientist

VaporVM

☐ July 2023 - June 2024

Lahore, Onsite

- Automated daily Excel reporting tasks using Python, resulting in a significant reduction in processing time from 1 hour to just 2 minutes.
- Implemented and executed machine learning models in a distributed environment, substantially decreasing the time required for model training.
- Established and managed a Cloudera cluster, effectively assigning tasks and responsibilities across multiple nodes for optimal performance.
- Conducted ETL/ELT operations for data warehousing, enabling efficient data transformation and loading, ultimately contributing to the creation of an OLAP system.

Data Scientist PACRA

J June 2022 – August 2022

Lahore, Onsite

- Worked on credit risk modeling using Python, Google Data Studio, Azure and Tableau Prep.
- Financial report data was extracted using Azure Form Recognizer, cleaned with Python and Tableau Prep, and utilized to create a predictive model with deep learning algorithms. This streamlined process reduced manual data punching by 60%, allowing analysts to concentrate on extracting meaningful insights.
- Evaluated its performance. Created interactive dashboards in Google Data Studio to visualize the insights.

Data Engineer

Contract.PK

☐ August 2022 – September 2022 Lahore, Remote

- Engineered robust ETL pipelines using Python, optimizing extraction from SQL and NoSQL databases and ensuring precise data transformation for downstream analytics.
- Architected an OLAP system with Python and SQL by designing optimized schemas, tables, and indexes to boost query performance and support scalable analytics.
- Established rigorous data consistency and concurrency controls, enabling reliable transaction management and high-volume data processing.



GCP Certified: Professional Data Engineer

click here to view



Data Management in Databricks click here to view



Improving Query Performance in SQL Server

click here to view



Data Science with Tableau click here to view

VOLUNTEER WORK



General Secretary - ACM UET NC 05/2023 - Present

> Secured a partnership with DataCamp to provide 50+ students with \$500 worth of subscriptions.

PROJECTS

Dynamic Malware Analysis Using ML

Inventory Analysis in Tableau

HR Analytics in Powe BI

Intelligent Agent Deployment with Reasoning in Vertex Al

Synapse-to-Fabric Modernization Project

Lakehouse Architecture with AWS Glue, S3, and Athena

Credit Risk Data Engineering Prediction Pipeline

Middilion Data Architecture in Azure Synapse