

Kashif Sohail

kashifsohailus@gmail.com | +1 (203) 435-4669 | [LinkedIn](#)

Work Experience:

Azure Cloud & Infrastructure Engineer | LTIMINDTREE (MICROSOFT)

JAN 2023 – PRESENT

- Improved system performance and data flow efficiency for Microsoft clients by identifying and resolving issues in Azure messaging services, including Service Bus, Event Hub, Event Grid and AI Search, resulting in enhanced reliability and customer satisfaction.
- Identified and resolved issues by analyzing backend logs across various dashboards (Namespace, scale unit, entity, QoS etc.) and inspecting publisher and consumer end applications/services.
- Built a config-driven ingestion framework using Azure Databricks Auto Loader and Spark to process both batch and real-time data from ADLS Gen2.
- Designed and implemented Medallion Architecture (Bronze, Silver, Gold layers) using Delta Lake, improving data reliability, traceability, and transformation efficiency.
- Developed scalable ETL/ELT pipelines using PySpark and Spark SQL, handling structured and semi-structured datasets.
- Implemented Unity Catalog for centralized governance, enabling fine-grained access control (RLS/CLS), schema enforcement, and data lineage tracking.
- Created optimized Delta tables and curated views to support downstream analytics and reporting workloads.
- Improved pipeline performance through Spark optimization techniques such as partitioning, caching, and query tuning.
- Documented data pipelines and workflows, enhancing maintainability, operational efficiency, and knowledge sharing across teams.
- Diagnosed latency, throttling, authentication, and server errors by analyzing ARM logs, KQL queries in Azure Data Explorer, and ETW logs to identify root causes.
- Collaborated across services like Function App, Entra ID, Blob Storage, ADLS Gen2, and logic app to resolve the issue and identify root causes, improving issue resolution efficiency and customer satisfaction.
- Managed high-pressure CritSit and Severity A cases by coordinating with product teams and handling multiple issues simultaneously, ensuring timely resolutions and maintaining SLA compliance.
- Teamed up in triages with SMEs to review and identify errors in Microsoft documentation and TSGs, implementing appropriate corrections.

Azure Data Engineer | LTIMINDTREE (MICROSOFT)

June 2022 – December 2022

- Developed custom data ingestion pipelines using PySpark on Azure Databricks, loading data from APIs, Azure SQL, and Blob Storage into ADLS Gen2.
- Orchestrated end-to-end data workflows using Azure Data Factory (ADF) and Databricks Workflows, ensuring reliable and automated pipeline execution.
- Implemented data security and governance using Unity Catalog, including table-level, row-level, and column-level security.
- Optimized Spark jobs and Databricks clusters, achieving ~30% improvement in processing time through efficient resource utilization and query optimization.
- Leveraged Delta Lake features to improve data reading performance and reduce latency for large datasets.
- Managed schema and metadata organization using Unity Catalog metastore, ensuring consistency across environments.
- Integrated CI/CD pipelines using Azure DevOps to automate deployment and version control of data engineering workflows.
- Analyzed business requirements and provided technical guidance on optimizing analytics workflows using Azure Databricks, Synapse, Blob Storage, Data Lake Storage, Cosmos DB, and SQL Data Warehouse.
- Resolved service unavailability, UI issues, networking/connectivity problems, Databricks Runtime version and instance errors, and optimized cluster cost management in Azure Databricks.
- Managed product and bug-related issues, provided guidance on code defects, resolved gateway and deployment issues, and assisted in implementing secure services and real-time solutions using Azure Databricks.
- Investigated customer issues using advanced Kusto Language Queries and Jarvis logs with Microsoft telemetry tools, troubleshooting under agreed SLAs.
- Collaborated with Data Factory, Networking, Azure Key Vault, and Log Analytics teams to streamline data workflows and enhance security and monitoring capabilities.
- Configured Azure and Databricks environments, including mounting storage accounts, setting up VNET injected Databricks workspaces, creating virtual networks, and reviewing firewall, cluster, and Docker configurations.

- Experienced loading raw data into Spark Data Frames and validating the data.
- Experienced in converting the validated RDDs into Data frames for further processing
- Implemented the Spark SQL code logic to join multiple data frames to generate application specific aggregated results.
- Experienced in fine tuning the jobs for better performance in the production cluster space.
- Extensively worked on distributed File Systems like Hadoop (HDFS) and Distributed Processing
- Worked with Log4j framework for logging debug, info & error data.
- Implemented the Spark Scala code for Data Validation in Hive.
- Imported data from AWS S3 into Spark RDD, Performed transformations and actions on RDDs.
- Experience in using Avro, Parquet and JSON file formats, developed UDFs in Hive.
- Designed appropriate partitioning/bucketing schema to allow faster data retrieval during analysis.
- Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
- Experience working with various File-Formats (Parquet, Avro & JSON) and Compressions (Snappy & Gzip).
- Responsible for generating actionable insights from complex data to drive real business results for various applications teams and worked in Agile Methodology projects extensively.
- Worked with SCRUM team in delivering agreed user stories on time for every Sprint.

Projects:

MAR 2022- 2024

- Migrated client from on-premises to cloud using Amazon EKS and containerization, enhancing scalability and productivity.
- Developed a responsive website using HTML5, CSS, JavaScript, MySQL, and PHP on Helio Host.
- Built a relational database for cricket roles, tactics, and formations using ER diagrams, MS SQL Server, and MySQL.
- Designed and implemented a Data Warehouse with Star-Schema using MS SQL Server and Azure Data Studio, improving data retrieval and analytics efficiency.

Education:

MS Data Science | University of Cumberlands

December 2024

MS Computer and Information Systems | CALUMS

March 2022

Technical Skills:

Programming Languages: Python, SQL, PySpark, and Scala.

Other Tools: Eclipse, PyCharm, GitHub, Jira, Confluence
Gitlab, Visual Studio, MS Project, MS Teams, Postman,
MongoDB, MySQL.

Azure Cloud: Azure Databricks, Azure Data Factory, Azure SQL Database, Azure Synapse Analytics, Azure Cosmos DB, Azure Data Lake Storage (ADLS), Service Bus, Event Hub, Event Grid, AI Search, Notification Hub, Relay, DevOps, CI/CD.

Certifications:

AWS Solution Architect Associate Certification,

Microsoft Azure Technology Certifications: Cloud Academy

Azure Developer – Compute - Compute- Cloud Services, APIM, Service Bus (Azure Search, Event Grid, Event Hub, Notification Hub, Premium Messaging-Specialist)