

Mohamed Kashifuddin

📞 93630-99491 — ✉ mohamedkashifuddin24@gmail.com — 🔗 linkedin.com/in/mohamedkashifuddin — 🌐 github.com/mohamedkashifuddin

Summary — Data Engineer focused on cloud data infrastructure. Building scalable data pipelines and advanced analytics solutions. Proficient in cloud-native architectures, adept at leading cross-functional teams to deliver high-impact data solutions that drive business growth..

Experience

DXC Technology

May 2022 – Present

Data Engineer (Analyst II Data Engineering)

- Optimized and maintained scalable data pipelines to process upstream data from Azure Blob Storage and store it in Snowflake, improving data accessibility for analytics.
 - Designed and implemented new ETL workflows and enhanced existing pipelines to meet evolving business requirements, ensuring efficient and reliable data processing.
 - Managed end-to-end data processing across multiple architectures, including Snowflake-based pipelines and Databricks-powered workflows for data transformations.
- Challenges :
- Reduced data corruption incidents by isolating overlapping write operations into staging areas and enabling atomic moves to final destinations only after validation.
 - Enhanced performance of large-scale data transformations by optimizing partition strategies, caching intermediate results, and tuning resource allocation for stable execution.
 - Strengthened pipeline observability by implementing proactive alerting, structured log management, and health checks, enabling early detection and resolution of recurring failures before impact of SLA.
 - Resolved schema evolution challenges by implementing resilient strategies that supported new data attributes while ensuring historical compatibility across analytical datasets.
 - Enhanced collaboration across teams by documenting troubleshooting guides, and recovery procedures on Confluence for faster onboarding and knowledge sharing .

Kloudone

Jan 2020 – May 2022

Software Engineer

- Played a key role in transitioning large-scale data workloads to a new cloud platform, ensuring minimal disruption to downstream processes.
- Enhanced data validation processes and implemented improvements to ensure higher data integrity, consistency, and reliability across workflows.
- Fine-tuned data ingestion and transformation processes, reducing overall processing time and improving system efficiency.

Projects

GCP E-commerce Data Pipeline

- Built an end-to-end batch data pipeline on GCP using medallion architecture (Bronze, Silver, Gold) with PySpark, BigQuery, and dbt.
- Designed schema utility for consistent schema enforcement across Bronze and Silver layers, ensuring data quality and type integrity.
- Optimized Spark jobs with caching and timezone configurations, improving pipeline performance and consistency.
- Implemented Airflow orchestration with retries, concurrency control, and dbt integration for automated data transformations and testing
- Configured service accounts for cross-service authentication; recognized trade-offs and best practices for IAM in production.
- Connected BigQuery to Metabase for KPI validation and dashboards, enabling fast business insights.

Skills & Education

Skills

- **Cloud:** GCP, Microsoft Azure
- **Data Processing:** PySpark
- **Orchestration:** Azure Data Factory, Apache Airflow
- **Object Storage & Warehousing:** Snowflake, BigQuery
- **Languages:** Python, SQL
- **CI/CD:** Jenkins, GitHub

Education

- **Bharath Institute of Higher Education and Research**
B.Tech. in Electronics and Communication Engineering

Certifications

- GCP - Certified Google Cloud Professional Data Engineer
- Azure - Certified Azure Data Engineer Associate
- Databricks - Databricks Certified Data Engineer Professional