Sura Kishore

Data Engineer

+91-9177955377 | surrakishore@gmail.com |
https://www.linkedin.com/in/sura-kishore-62964131b/

PROFILESUMMARY:

- Having 3+ year experience in IT INDUSTRY And designing and developing end-to-end ETL pipelines using Azure Data Factory (ADF), integrating data from on-premise and cloud sources into Azure Data Lake Storage (Gen2) and Azure SQL Database.
- Expertise in parameterizing ADF components such as Linked Services, Datasets, Pipelines, and Activities to create reusable and dynamic workflows.
- Strong working experience with Azure BI services including ADF, ADLS Gen2, Azure SQL Database, Key Vault in real-time enterprise data platforms.
- Good hands-on experience in implementing secure key management through Azure Key Vault, leveraging secrets within ADF pipelines for credential management.
- experience in building automated data workflows using ADF triggers, and developing dynamic pipelines capable of handling multiple source-to-target mappings through a single pipeline design.
- Strong knowledge of Delta Extraction techniques and incremental load strategies to bring delta records from source systems to ADLS efficiently.
- Hands-on expertise in building and optimizing PySpark notebooks in Azure Databricks for large-scale data processing, transformation, and business logic implementation.
- Experience in working with Delta Lake tables and schema evolution in Databricks, ensuring ACID compliance for real-time data ingestion.
- Experience in monitoring and debugging ADF pipelines, identifying root causes of failures, and implementing retry and alert strategies.
- Proficient in SQL Server – strong in DML/DDL operations, joins, stored procedures, views, UDFs, and index optimization.
- Worked in the MEGA LT COE team using **HCL's Advisor Tool** to assess and modernize legacy applications.
- Collected and validated application details such as technology stack, database, and integrations for assessment.
- Uploaded data into the Advisor Tool and executed analysis runs to evaluate **complexity, reusability, and automation potential**.
- Supported client discussions and created summary presentations for modernization planning.

WORK EXPERIENCE:

Aug, 2022 - Oct, 2024 **HCL Techonologies**. as GET, India.

Aug,2025 – Till Date **DATARISE SOLUTIONS PVT LTD** as Data engineer ,India.

Type equation here.

TECHNICAL SKILS:

- Azure Databricks
- Azure Data Lake Storage
- Azure Data Factory

- Azure SQL database
- Pyspark
- Spark SQL
- Delta Lake
- SQL Server
- Spark SQL
- Azure Logic Apps
- Spark Streaming

EDUCATION

B. Tech from **Jawaharlal Nehru Technological University**, **Anantapur**.

RELEVANT PROJECTS:

Client: ST MicroElectronic, **USA**
Environment: Azure Databricks, Delta Lake, PySpark, Azure Data Lake Storage, Azure Data Factory, SQL Server, Spark SQL
Project: Manufacturing
Responsibilities:

- Built scalable data pipelines by implementing Medallion Architecture (Bronze, Silver, Gold) using Delta Lake on Azure Databricks.
- Created PySpark modules to validate incoming data (e.g., null checks, date format, email structure) and enforce schema consistency before ingestion.
- Ingested raw data into the Bronze layer from various source systems using parameterized notebooks.
- Processed Bronze data by applying cleansing, deduplication, filtering, and join operations, and loaded refined data into the Silver layer.
- Developed logic for full and incremental loads using Delta MERGE statements based on primary key comparison.
- Applied SCD Type 2 transformation to maintain historical versions of records in Dimension tables.
- Designed and created Fact and Dimension tables in the Gold layer to support business reporting and Power BI dashboards.
- Utilized Delta Lake capabilities like ACID transactions, schema evolution, and time travel for safe updates and rollback scenarios.
- Optimized Spark performance by addressing small file problems (auto-compaction), applying salting to handle data skew, and using partition pruning.
- Tuned Spark jobs using the Catalyst Optimizer and reviewed execution plans to enhance transformation performance.
- Implemented try-catch blocks, checkpointing, and audit logging to ensure fault tolerance and traceability in pipeline execution.
- Coordinated with BI developers and business stakeholders to finalize data requirements and deliver accurate sales and manufacturing insights.

Client: Best Buy USA
Environment: Microsoft Azure, Azure Data Factory, Azure Data Bricks, PySpark, Azure data lake Storage, SQL server, Azure SQL Database
Project: E-commerce
Responsibilities:

- Developed and maintained end-to-end ETL pipelines in Azure Data Factory (ADF) to process and orchestrate large-scale data workflows from various source systems.
- Set up and configured Self-Hosted Integration Runtime (SHIR) to securely connect with on-premise source systems.
- Created and managed Linked Services, Datasets, and Data Flows to enable seamless connectivity between diverse source and sink systems.
- Designed and implemented dynamic and reusable ADF pipelines to ingest and transform data from multiple structured and unstructured sources.

- Maintained audit logging mechanisms by writing ADF pipeline run metadata into Azure SQL Database to support monitoring and troubleshooting.
- Implemented full and incremental loading strategies based on business logic, leveraging watermarking and last modified timestamps.
- Automated error notification workflows using Azure Logic Apps to send email alerts for failed pipeline executions.
- Achieved end-to-end pipeline automation by utilizing ADF triggers, parameterization, and dependency conditions.
- Developed data reconciliation mechanisms to validate data completeness, consistency, and alignment across systems.
- Performed comprehensive data quality checks ensuring accuracy, completeness, and integrity in alignment with business and technical requirements.
- Integrated Azure Key Vault to securely manage secrets, and configured secure linked services for authenticated data access.
- Collaborated cross-functionally with infrastructure, network, database, application, and BI teams to ensure data reliability and availability.
- Established connectivity and enabled secure data movement between Azure Data Factory and AWS S3, supporting hybrid cloud architecture.
- Integrated Snowflake Data Warehouse with ADF and orchestrated data ingestion into Snowflake tables, enabling downstream analytics and reporting.