Manideep Lekkala

Azure Data Engineer

Hyderabad, India | +91-9505207004 | manideep.mani25@gmail.com | linkedin

Experience Summary:

- Dedicated and results-driven Data Engineer with 3 years of experience in designing, implementing, andoptimizing data pipelines and architectures.
- Proficiency in programming languages such as Python and SQL
- Experience in implementing Data Lake with technologies like **Azure Data Factory (ADF), PySpark, Databricks, ADLS, Azure SQL Database** and file formats like JSON, CSV & Parquet.
- Development and administration of **ETL processes** in Azure. Experience in creating **ADF Pipelines** to source and process data sets.
- Good understanding of **Spark Architecture including Spark Core, Spark SQL, Spark Context, Driver Node, Worker Node, Executors and Tasks.**
- Designed and developed Spark applications using **PySpark Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data.
- Experience in managing and storing credentials in **Azure Key Vault**.
- Migrated SQL database to various Azure services such as **Azure Data Lake, Azure SQL Database, and Azure Databricks.**
- Experience in developing data pipelines to transform, aggregate or process data using **Azure Databricks** platform.
- Expertise in writing complex **SQL Queries**.
- Designed and implemented scalable, scheduled data pipelines using **Databricks** and **Azure Logic Apps** to ingest data from different sources and consolidate them into single data lake.
- Designed various ingestion and processing patterns based on use cases in **Delta Lake**.
- Implemented data **partitioning** to evenly distribute workloads and reduce data movement between nodes,improving processing efficiency.
- Leveraged **broadcast variables** to efficiently share large read-only data with all worker nodes, reducing communication overhead.
- Improved performance for large data joins by using optimization techniques like **broadcast joins** and **repartitioning** in Spark.
- Knowledge in Data Warehousing concepts.
- Experience in designing and optimizing data warehouse systems to efficiently store and manage largeamount of structured and unstructured data for analysis and reporting purposes.
- Strong knowledge in entire development lifecycle like analysis, design, development, testing, deployment,and support in **agile** methodologies.
- Monitor and troubleshoot data pipeline performance and reliability. Stay up-to-date with the latest trendsand technologies in data engineering and analytics.
- Collaborate with cross-functional teams to understand and prioritize data requirements.
- Documented Data workflows, ETL processes, and Data models to ensure maintainability and knowledgesharing within the team.

Academic Profile:

Bachelor's in Electronics and Communication Engineering from Guru Nanak Institute of Technology(JNTU,Hyderabad) (2016-2020)

Work Experience:

LTIMindtree – Software Engineer *(April 2021 - Present)*

Technical Skills:

- Programming Languages   : Python, PySpark

- Databases   : MySQL, Azure SQL
- Cloud Technologies   : Azure Data Factory, Blob Storage, Azure Databricks,
                                                      Azure Synapse Analytics, Azure Logic Apps

Project Experience:

Client: The Hartford Insurance
Description: Hartford is one of the largest Insurance organization in the UnitedStates, offering a unique model of integrated Insurance delivery and Investment. It provides Automobile Insurance , Property Insurance , General Liability Insurance services to its members.
Roles & Responsibilities:
- Extract, Transform and load data from source systems to Azure storage services using **Azure Data Factory, Data Bricks, and Azure Data Lake Storage**.
- Collaborate with source teams to extract data and load it into **Azure Data Lake Storage(ADLS)**.
- Establish **linked services** for source and target connectivity according to requirements.
- Responsible for extracting data from **OLTP** and **OLAP** systems into Data Lake using Azure Data Factory and Databricks.
- Worked on migration of data from On-Prem SQL server to Cloud databases **(Azure SQL DB)**.
- Experience in Data Factory pipeline to create and schedule data-driven workflows that can ingest datafrom disparate data stores.
- Performed data validation and quality checks to make sure accuracy, consistency, and reliability.
- Developed complex **SQL** queries to extract data from databases based on the **transformation** rules andmappings received.
- Implemented **PySpark** and **Spark SQL** for faster data processing and in-memory computation with spark RDDs to generate output responses.
- To meet specific business requirements wrote User Defined Functions (UDFs) in PySpark.
- Created documentation for data pipelines and data transformation processes for future reference andknowledge sharing.

Certifications:

AWS Certified Cloud Practitioner

Declaration:

I hereby declare that the information provided above is true to the best of my knowledge.

Place : Hyderabad   Manideep Lekkala