

DATA ENGINEER

Professional Synopsis

A Results-driven data engineer with 2.9 years of experience in designing, developing, and optimizing data solutions using Azure cloud technologies. Proficient in relational databases and skilled in writing complex SQL queries for data analysis, development, and validation. Knowledgeable in the big data ecosystem, with hands-on experience in Apache Hadoop, Spark, SparkSQL, and PySpark. Experienced in collaborating with cross-functional teams and delivering projects on schedule and within budget.

Area Of Expertise

Technical Skills

- **Cloud Platforms:** Familiarity with cloud platforms like Microsoft Azure (e.g., Azure SQL, Blob Storage, Azure Data Factory, Azure Synapse Analytics).
- **Azure Data Factory :** Orchestrated and automated scalable data pipelines using Azure Data Factory, improving data accessibility, efficiency, and reducing processing times.
- **Databases:** Expertise in designing and querying relational databases (e.g., MySQL, MS SQL Server).
- **Hadoop Ecosystem:** Experience with Hadoop, including HDFS, MapReduce and Hive for large-scale data processing.
- **Apache Spark:** Expertise in using Apache Spark for large-scale data processing.
- **Apache PySpark:** Proficiency in using PySpark for distributed data processing, including data frames, RDDs (Resilient Distributed Datasets), transformations, and actions.
- **Spark SQL:** Expertise in leveraging Spark SQL to query structured data and optimize performance (using DataFrames, Datasets, and Spark SQL functions).
- **Programming Languages:** Strong knowledge of programming languages like SQL to manipulate and query large datasets, perform aggregations, joins and subqueries.
- **ETL/ELT Pipelines:** Building, managing, and optimizing ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) pipelines.
- **Data Lakes:** Expertise in creating and managing data lakes using tools like Apache Hadoop HDFS and Azure Data Lake Storage.

Functional Skills

- Ability to collaborate with business analysts, data scientists, and product teams to understand key metrics, KPIs, and the data requirements that drive business decisions.
- Expertise in translating business requirements into data models and designing solutions that address organizational needs.
- Ability to collaborate with data scientists to ensure data pipelines are tailored to their needs, particularly in terms of data preprocessing and feature engineering.
- Ability to work with DevOps and cloud engineers to ensure that the infrastructure and deployment pipelines for data systems are reliable and scalable.
- Understand the role of data engineering in enabling reporting and analytics. Although not directly responsible for generating reports, data engineers should know how to enable BI tools like Power BI to access and visualize data.
- Worked with teams to define data quality standards and ensuring that data pipelines handle data cleaning, validation, and transformation to meet those standards.
- Troubleshooting and resolving issues in PySpark jobs or data pipelines in Azure.
- Identifying and addressing performance issues in data workflows or distributed processing, ensuring efficient resource usage in Spark and Azure.
- Ability to manage multiple data engineering tasks and projects concurrently, ensuring deadlines are met.

Organisational Experience

02/2022 to Current : System Engineer at Infosys Ltd.

- Developed and maintained data pipelines using Azure Data Factory and Azure Synapse analytics.
- Developed and maintained stored procedures, views and functions in SQL server to optimize data ETL processes.
- Analyzed data using SQL to identify trends, patterns and insights.
- Collaborated with data scientists and analysts to provide data insights and support data-driven decision making.
- Troubleshoot and resolve data processing and storage issues.
- Developed and maintained documentation for data storage and processing solutions.
- Communicated findings and recommendations to stakeholders in a clear and concise manner, gained valuable teamwork experience.

Work Experience

Delivery Service – DHL :

Leverage data analyst to enhance logistics efficiency and optimize delivery routes. Designed and implement data pipelines using Azure Data Factory to streamline data collection from multiple logistic systems. Analysed delivery performance metrics using SQL to identify bottlenecks and areas for improvement. Developed predictive models in Azure machine Learning to forecast delivery times and optimize routing strategies.

Ecommerce Website - Talbots :

Analysed retail sales data to optimize Inventory management and improve customer insights. Collected and cleaned large datasets from various sources using Azure Data Lake. Developed SQL queries to extract meaningful insights from sales and inventory data. Created interactive dashboards in Power BI to visualize trends and performance metrics.

Academic Qualification

- Master in Computer Science from Sambalpur University.
- Bachelor in Computer Science from Sambalpur University.

Certifications

- Infosys Certified Big Data Developer.
- Infosys Certified Spark and PySpark professional.
- Infosys Certified Mysql Associate.
- Certified in Azure Data fundamental (AZ – 900 & DP - 900)
- Certified in Data Engineering (DP -203)

Personal Minutiae

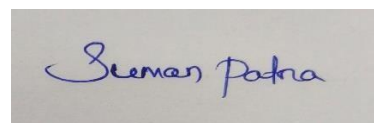
- Date of Birth : 22-05-1996
- Email : sumanpatra688@gmail.com
- Address : Pune, Maharastra

Declaration

I affirm that all the information furnished above are true to the best of my knowledge and belief.

Place: Pune, Maharastra

Date: 16/12/2024



(Signature)