ABHISHEK DAS

+91-8249698971 ⋄ abhishekdas69597@gmail.com ⋄ Hyderabad, India
Github/dabhishek316 ⋄ in/abhishek-das/ ⋄ medium/@abhishekdas69597

OBJECTIVE

Driven Data Engineer with 3+ years of experience in Python, SQL, Spark, and advanced big data technologies. Proficient in Azure, Databricks, Data Lake, and Delta Lake, specializing in building and optimizing robust data pipelines. Demonstrated success in enhancing data processing efficiency and reducing computation time. Eager to leverage skills to drive impactful data solutions and support organizational growth through efficient data management and analysis. .

EDUCATION

B.Tech Computer Science, C.V. Raman College of Engineering   2017 - 2021
CGPA : 8.7

EXPERIENCE

Consultant Data Engineer   Jul 2023 - Present
Deloitte In   *Hyderabad, India*

- Engineered and deployed a Data Exchange framework using PySpark and Python, enhancing data processing by 20% and reducing data processing time by 15% by integrating ServiceNow ticket creation via API from Azure Data Factory.
- Initiated and optimized a workflow to trigger jobs in Azure Databricks from Azure Data Factory using service principal authentication, implementing parallelism with multi-threading to reduce computation time by 20%.
- Engineered and deployed the Query Generator framework, transforming XML inputs into SQL queries; ensured secure, efficient extract delivery, resulting in 40% faster data retrieval and a 25% reduction in manual query errors.

Data Engineer   Jul 2021 - Jul 2023
Cognizant   *Hyderabad, India*

- Processed data using Spark DataFrames, including transformations (select, filter, map, groupBy, agg, join) and actions (show, write.save), improving data operational accuracy by 15%.
- Architected and executed an event-triggered ADF pipeline, reducing data ingestion latency by 25%.
- Developed bronze and silver layer scripts to process data to different layers using PySpark and store in s3 buckets, resulting in a 30% improvement in data computation efficiency.

SKILLS

Programming Language	: Python,Pyspark
Data Engineering	: Delta Lake, Spark SQL, Lakehouse, Cloud Optimization, Delta Live Table (DLT)
Cloud Technologies	: Azure (ADF, ADLS, ASA), AWS (Resdhift, S3 Buckets)
Databases/DWH	: SQL, Teradata SQL ,Postgres, T- SQL
Tools	: GIT, Postman, PyCharm, VsCode, VM, SQL Server Management Studio

PROJECTS

Amazon Data Analysis Project in Pyspark — *Pyspark, Databricks, Python, Git* — Github

- The Project leverages PySpark to analyze sales data and uncover key business insights. It identifies the **region and country with the highest profits**, determines the **maximum orders placed within 72 hours**, and finds the **month with the highest average units sold**. Additionally, it calculates the **lowest and highest margin percentages** and highlights **the top 3 highest-value orders for specific item types**. This project showcases **PySpark's capabilities** in handling and analyzing large datasets for e-commerce optimization