

ABHISHEK DAS

+91-8249698971 ◇ abhishekdas69597@gmail.com ◇ Hyderabad, India
[github/dabhishek316](#) ◇ [in/abhishek-das/](#) ◇ [medium/@abhishekdas69597](#)

OBJECTIVE

Microsoft Certified: Azure Data Fundamentals (DP-900) Data Engineer with 3+ years of hands-on experience in Python, SQL, Spark, and cutting-edge big data technologies. Proficient in Azure, Databricks Lakehouse, Data Lake, and Delta Lake, I specialize in building and optimizing robust data pipelines. Winner of SIH 2020. Eager to leverage my skills in a dynamic environment 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
CGPA : 8.7

2017 - 2021

EXPERIENCE

Consultant Data Engineer
Deloitte In

Jul 2023 - Present
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 processing 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
Cognizant

Jul 2021 - Jul 2023
Hyderabad, India

- **Processed** data using Spark DataFrames, including transformations (select, filter, map, groupBy, agg, join) and actions (show, write.save), improving data processing 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 processing efficiency.

SKILLS

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

PROJECTS

Amazon Data Analysis Project in Pyspark — *Pyspark, Databricks, Python, Git* — [Github](#)

- The Amazon Data Analysis Project in PySpark 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 processing and analyzing large datasets for e-commerce optimization.