NAVAS SHERIF I

✉ navassherif2023@gmail.com   🔗 Linkedin : https://www.linkedin.com/in/navassherif/

📞 +91 8921954946   🖥 Protfolio : https://navassherif98.github.io/Portfolio/

📍 Bangalore , India   👤 Github   : https://github.com/navassherif98/

PROFESSIONAL SUMMARY

Experienced Data Engineer with 3 years of expertise in Big Data. Proficient in designing scalable data solutions, optimizing pipelines, and leveraging advanced analytics. Skilled in **Spark, Scala, Python, SQL and Azure Databricks** for efficient data processing. Successfully orchestrated the migration from a Mainframe legacy system to a Big Data platform, achieving a 40% reduction in data processing time and a 30% increase in overall system efficiency.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, Scala

- **Big Data Technologies:** Hadoop, Apache Spark, Hive, Impala, PySpark, Spark-SQL, HDFS, Kafka, Sqoop

- **Cloud Technologies:** Azure, Azure Databricks, Azure Storage, Azure Data Lake

- **Database Management:** MySQL

- **Pipeline Management:** Apache Airflow

- **Data Analysis:** Microsoft Excel, Data Science

- **Data Warehousing and ETL Processes**

- **Project Management Tools:** JIRA, Agile Methodology

- **Scripting:** Bash Scripts

- **Version Control:** Git

- **Continues Integration/Continues Deployment (CI/CD):** Jenkins

WORK EXPERIENCE

Associate Software Developer   Aug 2021 – Current

IQVIA – Bengaluru, India

APAC - Legacy System Migration to Modern Big Data Platform

This project involved creating an end-to-end data pipeline for processing pharmaceutical data, ensuring efficient batch processing, timely reporting, and improved data accuracy.

This initiative resulted in significant operation cost reductions, and enhanced data accuracy and enhanced data processing speed.

Roles and Responsibilities :-

- **Databricks & Spark Application Development :** Developed and maintained high-performance, scalable application using Databricks and Spark Scala, ensuring clean, maintainable, and efficient code.
- **End-to-End Pipeline :** Implemented a comprehensive data pipeline using Airflow DAGs, ensuring seamless data processing from data ingestion to reporting.
- **Spark Job and Code Optimization :** Enhanced Scala code performance and optimized Spark applications by tuning parameters, resulting in a 40% reduction in processing time.
- **Data Warehousing and Modeling :** Designed and implemented efficient data warehousing solutions and optimized data models using dimensional modeling techniques.
- **Job Monitoring and Issue Resolution :** Monitored and resolved issues in scheduled jobs to ensure smooth data processing and reducing downtime by 25%.
- **Report Analysis and Root Cause Identification :** Analyzed reports to identify and fix root causes of issues, improving data quality by 20%.
- **Cloud Migration :** Managed the migration from on-premises Cloudera Distribution Hadoop (CDH) to Cloudera Data Platform (CDP) on the cloud, enhancing scalability and reliability.
- **Collaborative Problem Solving :** Partnered with stakeholders, data teams, and engineers to define requirements, troubleshoot, optimize, and deliver data-driven insights.
- **Version Control Management :** Handled code movement and version control using Git, ensuring smooth and organized code deployments.
- **ETL Process Management :** Managed ETL processes achieving a 30% improvement in data integration efficiency.

CERTIFICATIONS

IBM Data Science Professional Certificate   [View Credentials]

Microsoft Azure AZ-900 Azure Fundamentals [View Credentials]

Microsoft Azure DP-900 Azure Data Fundamentals [View Credentials]

Azure Databricks & Spark for Data Engineers (Pyspark , SQL) [View Credentials]

EDUCATION

Bachelor of Technology (BTech) : Computer Science   2017 - 2021

Ahalia School of Engineering and Technology,

 Palakkad, Kerala, India

SOFT SKILLS

- **Analytical Reasoning**
- **Creative Troubleshooting**
- **Collaborative Influence**
- **Strategic Decision Making**