Sandi Poojitha
Azure Data Engineer

SUMMARY

- Working as an **Azure Data Engineer** with exclusive hands-on experience in implementing cloud services
- Expertise in creation of **ETL data** pipelines using **Azure Data factory**
- Experience on Data bricks with Pyspark for Data migration and ETL Lake house project implementation using Data bricks Notebooks and scheduling jobs
- Good understanding of spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks
- Analyzing the existing mappings to understand the project flow and new design
- First-rate analytical and problem-solving skills dedicated to maintaining high quality standards
- Analyze and make the Code Changes as per the business requirement.

Technical Expertise

- Python(Pyspark, Pandas)
- Spark SQL, Spark interactive, HDFS, Hive
- SQL Server(MS SQL)
- Azure Data Factory,
- Azure Storage Explorer
- Azure Synapse Analytics
- Azure Data bricks
- Azure Logic apps
- Azure Blob Storage
- Azure Data Lake
- Oracle 12c.
- Big Data
- Spark
- Data Warehousing
- IntelliJ, PyCharm
- HTML
- C - Language
- Basic JAVA
- MS-Office

EXPERIENCE

Accion Technologies Pvt Ltd - Azure Data Engineer

Project   : Halcyon
Technology : Python, Azure SQL, Azure Data Factory, Azure blob storage, Azure data bricks, Azure logic apps, Azure Data Lake.

Description : Halcyon's mission is to help financial institutions, tax professionals, and the people they serve work better together.

Roles & Responsibilities:

- Hands-on experience with Databricks Tables, using Delta Lake/Lake House architecture, creating dashboards to visualize the job outputs
- Hands-on experience with Spark notebooks, notebooks with parameters and creating notebook workflows

- Having good experience on Creating Data Factory Pipelines, datasets, data flows and triggers in Azure data factory
- Integrating Delta Lake tables with PowerBI and Tableau
- Created multiple pipelines for automating end to end data engineering processing using copy activities and data bricks notebooks
- Improved system performance by troubleshooting and fixing data-related problems
- Designed and maintained large-scale data systems using Spark
- Improved data quality by implementing data validation and cleansing   procedures
- Creating standardized cluster policies, defining dedicated cluster pools and setup of processes for monitoring clusters
- Creating workable strategy for cost chargeback
- Established strategy for use of Azure Key vault with Azure Storage Accounts (ADLS g2)
- Staying up-to-date: Keeping up-to-date with new technologies, tools, and best practices in the field of data engineering
- Performance optimization: Monitoring system performance and troubleshooting data-related issues to improve system efficiency and data accuracy

EDUCATION

- **B.Sc.**, Computer Science **Sree Vidyanikethan Degree College**