

Akshan Soman

Delhi • 7838023036 • akshansoman98@gmail.com • linkedin.com/in/akshan-soman-0a8807166

Azure Data Engineer

Certified Azure, Databricks, and Pyspark data engineer with 4 years of experience in leveraging big data technologies. Experienced in designing and developing data warehouses, ETL processes and other data engineering projects in banking ,BI and data analytics domain. Demonstrated ability to quickly adapt to new technologies and take ownership of data engineering tasks. Seeking a challenging role as a data engineer to utilize existing skills and continue to learn.

WORK EXPERIENCE

KPMG Global Services • Gurgaon

Dec 2023 - Present

Associate Consultant • Full-time

- Integrated data related to retail pricing, market spend, workforce analytics, human capital, and Sector health across 1000+ companies into a centralized data warehouse from various SFTP locations, ensuring high data accuracy and timeliness.
- Ingested 500+ GB of data from multiple SFTP locations into Azure Data Lake Storage (ADLS) using Azure Synapse, optimizing the data pipeline for efficiency and scalability.
- Processed large datasets using PySpark, extracting valuable insights that facilitated informed decision-making and strategic analysis for stakeholders.
- Collaborated with the Power BI team to thoroughly understand and translate requirements, resulting in the delivery of tailored data solutions that met business needs.

Accenture Solutions Pvt Ltd • Noida

Jan 2021 - Nov 2023

Data Engineering Analyst • Full-time

- Successfully obtained diverse datasets from multiple sources, including Microsoft SharePoint, Excel, and CSV files, into Azure Data Lake Storage Gen2 (ADLS Gen2) via Azure Data Factory (ADF), averaging over 10,000 records daily.
- Leveraged the power of Databricks and Pyspark to transform raw data into valuable insights, implementing advanced cleaning, filtering, and preprocessing techniques for improved data quality.
- Drastically improved data processing efficiency, resulting in a notable 30% enhancement in query performance and a significant acceleration in data retrieval speed.
- Automated data ingestion workflows with Azure Data Factory pipelines, reducing manual intervention by 30% and establishing an efficient end-to-end data management process.

Data Engineering Associate • Full-time

- Created and managed Flink streaming jobs to process over 300 million data points in real-time, including data transformations, filtering, aggregations, and windowing.
- Formulated over 50 successful ETL processes using IBM Infosphere Datastage from multiple sources such as Oracle, Teradata, Hive, delimited files, and MongoDB.
- Scheduled 200+ IBM Infosphere Datastage jobs with the Tivoli Workload Scheduler (JADE) tool to enhance workflow productivity.
- Single-handedly streamlined the process of DDL creation, validation, and execution, leveraging Python (Pandas and NumPy), leading to a 70% reduction in manual effort.

EDUCATION

Bachelor of Technology in Computer Science

GL Bajaj Institute of Technology and Management • Greater Noida • 76%

Jul 2016 - Jul 2020

Class 12th

Evergreen Public School

Apr 2015 - May 2016

Class 10th

Kerala Education Senior Secondary Society School • Delhi • 9

May 2014

SKILLS

- Azure Databricks
- Azure Data Factory
- Azure Synapse Analytics
- Pyspark
- IBM Datastage
- Apache Kafka
- Apache Flink
- Python
- Java
- SQL

CERTIFICATIONS

DP 203 - Azure Data Engineer Associate

Microsoft

Databricks Certified Associate Developer for Apache Spark

Databricks

Databricks Certified Associate Data Engineer

Databricks

Google Cloud Platform Associate Cloud Engineer

Google