Lakkaraju Anjaneya Vara Prasad

Programmer Analyst

lakkarajuanjaneyavaraprasad@gmail.com 8519829962 hyderabad,India

https://www.linkedin.com/in/lakkaraju-anjaneya-vara-prasad-00070b235 04/08/2001 Indian

Headline

Azure Data Engineer | ADF - Databricks (PySpark) - Power BI - Data Quality & Optimization

Profile

Highly motivated Azure Data Engineer with 1.3 years of experience designing and developing scalable cloud-based data pipelines using Azure Data Factory, Databricks (PySpark), ADLS Gen2, and Synapse Analytics. Proven ability to automate ETL processes, optimize PySpark workloads, and ensure high data quality for analytics and reporting in banking domains. Microsoft Certified Azure Data Engineer Associate (DP-203).

Education

Bachelor of Technology (B.Tech), *Vasireddy Venkatadri Institute of Technology* 06/2019 – 05/2023
CGPA : 8.0 Guntur, India

Professional Experience

Programmer Analyst, *Cognizant Technology Solutions* 12/2023 – Present
hyderabad, India

- Designed and automated end-to-end ETL pipelines ingesting 1–2 million banking transactions/day from SQL Server, MySQL, and flat files into ADLS Gen2 and Synapse Analytics using Azure Data Factory and Databricks PySpark.
- Enhanced PySpark job performance by 15–20% through partitioning, caching, and optimized joins.
- Developed and maintained 8–10 ETL pipelines with automated error handling and scheduling, reducing manual monitoring by ~20% and lowering failure incidents by ~15%.
- Authored SQL validation queries improving reporting data accuracy by ~10% for Power BI dashboards used by 10–15 stakeholders in fraud detection and customer analytics.
- Assisted in monitoring Databricks clusters for compute cost and resource optimization.
- Collaborated closely with business analysts, QA, and senior engineers to deliver compliant, high-quality data solutions.

Skills

Technical Skills:

- Cloud Platforms: AWS (Glue, EMR, Redshift, Lambda, S3), Azure (Data Factory, Synapse), GCP (BigQuery, Dataflow)
- Big Data & Distributed Processing: EMR (Spark), Hadoop, Hive, Sqoop, HDFS
- Azure Data Factory (ADF), Azure Databricks, Azure Data Lake Storage Gen2 (ADLS)
- Azure Synapse Analytics, Azure SQL Database, MySQL
- PySpark tuning, SQL (Advanced), ETL pipeline automation
- Power BI dashboards and reporting
- Data validation and governance
- Python and Data ware house concepts

Soft Skills:

- Communication
- Teamwork
- Problem-solving
- Attention to detail
- Time management

Languages

- English
- Telugu
- Hindi

Certificates

Microsoft Certified: Azure Data Engineer Associate (DP-203)
(AWS certification in progress – targeting AWS Certified Data Analytics or Solutions Architect)

Mini Projects

Streaming state management for user session tracking using pyspark:
This project leverages PySpark Structured Streaming to implement real-time session tracking and stateful analytics. It processes a continuous stream of employee activity events—such as logins, clicks, and logouts—to dynamically manage session states. By applying techniques like sessionization, stateful aggregation, and windowed operations, the system provides live insights into session durations, user engagement, and cumulative actions. The project showcases the power of streaming state management in operational analytics, enabling responsive dashboards and real-time decision-making across domains like HR, security, and product usage.

Cloud Data Orchestration: From Buckets To BigQuery With Dataflow and Cloud functions
This project automates the end-to-end pipeline for ingesting, processing, and loading data from Google Cloud Storage into BigQuery using Cloud Functions and Dataflow. The workflow is event-driven: when a new file is uploaded to a bucket, a Cloud Function is triggered to launch a Dataflow job that transforms and loads the data into BigQuery. The solution demonstrates scalable and efficient cloud-native orchestration, highlighting seamless integration across Google Cloud Platform (GCP) services. It emphasizes real-time responsiveness, modular architecture, and cost-effective data movement.