Leela Nimmagadda
DATA ENGINEER

Ph: 937-829-0315 Email: n.leelasaikiran@gmail.com LinkedIn: www.linkedin.com/in/leelasaikiran

Data Engineer with 4 years of experience in building and deploying highly distributed, fault-tolerant data pipelines across Azure, GCP and Skilled in Python, SQL, PySpark, Apache Airflow, Kafka, and relational/non-relational databases/warehouses (BigQuery, Snowflake, Azure SQL, Redshift, Delta Lake). Strong background in database, data modeling, data architecture, CDC pipelines, batch and stream processing, and ETL/ELT optimization. Skilled in automating complex pipelines for actionable analytics and seamless cloud integrations.

TECHNICAL SKILLS:

Cloud Platforms: Azure (Data Factory, Databricks, Synapse, ADLS), GCP (Big Query, GCS), AWS.
Databases: Snowflake, PostgreSQL, Oracle, DB2, Delta Lake, Hadoop, Redshift, No SQL databases.
Languages: Python (Pandas, PySpark), SQL, Apache Spark, Spark SQL.
Data Tools: Apache Kafka, Airflow, Azure DevOps, Docker, CI/CD (GitHub Actions), Jira, Unity catalog, Databricks.
Data Modeling: Medallion Architecture, SCD2, CDC, OLAP/OLTP, Data Lake/Warehouse, ETL, ELT, Data warehouse.

PROFESSIONAL EXPERIENCE:

CGI - Data Engineer
April 2023- Present

- Migrated 2TB of legacy healthcare data to GCP BigQuery via Airflow, reducing storage costs by 25% and improving scalability while ensuring HIPAA compliance.
- Built PySpark ETL pipelines processing 500K+ daily records, improving delivery speed by 40%.
- Reduced batch times from 8 hrs to 45 mins by implementing incremental (CDC) pipelines.
- Deployed 15+ real-time Kafka pipelines ingesting IoT/API health data, enabling timely clinical and operational insights.
- Boosted Power BI performance by 25% via SQL optimization and stored procedure tuning.
- Cut Power BI dashboard load times by 25% via SQL tuning and indexing, supporting real-time financial insights and decision-making while maintaining HIPAA compliance.

EPT IT Solutions – Software Engineer
September 2020 –August 2021

- Automated the ingestion of Avro, JSON, and Delta data from to Azure Data Lake using Azure Data Factory and Databricks, saving over 30 hours of manual work per month.
- Improved query performance by 35% and reduced data redundancy by 25%.
- Cut Azure Synapse and Azure Data Lake Storage costs by 15% through partitioning, compute optimization.
- Developed reusable Python scripts to automate data extraction and transformation, reducing data load times by 20% and compute costs by 15% annually.
- Involved in the migration of 3M+ sensitive customer records to Azure, deploying data masking in Databricks, reducing potential data breach exposure by 90% and increasing data trust.

EDUCATION:

University Of Dayton, OH, 2023

Master ’s in Computer Science.

Sri Satya Gurunanak University, 2019

Bachelors in Computer Science

Key Achievements:

- Achieved data retrieval speeds by 40% by optimizing PySpark jobs (parallelism tuning, broadcast joins) and SQL queries (indexing, query plan analysis) for enterprise logistics datasets, supporting faster reporting for 500+ daily users.
- Increased logistics ETA accuracy by 15% by integrating weather and traffic API data.
- Enhanced query performance by 25% by implementing clustering strategies.

CERTIFICATIONS:

- Databricks certified professional Data Engineer | 2024
- Microsoft Certified Azure|2024
- Apache Pyspark certified from Udemy|2024
- PCEP Certified Python Programmer|2023
- Certified in Data Engineering foundations from Linkedin|2023