

Nayana Nakulan

Al Qusais 1 – Dubai, UAE

☎ +971 55 986 55 84 • ✉ nayananakulan1998@gmail.com

in [linkedin.com/in/nayana-nakulan](https://www.linkedin.com/in/nayana-nakulan)

Professional Summary

Results-driven Data Engineer with 2+ years of experience in designing, developing, and optimizing data pipelines and warehousing solutions. Skilled in leveraging big data technologies such as Hadoop, Spark, Snowflake, and Teradata to build scalable solutions. Proficient in ETL processes using Informatica, Databricks, and Airflow. Strong troubleshooting abilities, ensuring data integrity and availability. Familiar with cloud technologies like Azure and experienced in working in cross-functional teams. Enthusiastic about developing solutions for complex problems in the Enterprise Data & Analytics space.

Core Competencies

Big Data: Hadoop, Spark, Hive, HBase

Cloud: Microsoft Azure, Databricks, Kafka

ETL Tools: Informatica PowerCenter, Airflow

Languages: Python, SQL, PySpark

Data Warehousing: Snowflake, Teradata,

CI/CD: GIT, Jenkins, JIRA,

Reporting: Power BI, MS Excel

Professional Experience

Infosys

Data Engineer

India

Dec 2021 – Apr 2024

- Designed and implemented a comprehensive data pipeline integrating diverse data sources and leveraging technologies such as Apache Spark, Hive, Kafka, MySQL, HBase.
- Improved data processing efficiency by 30% through Spark-Hive and Spark-HBase integrations, reducing latency in real-time and batch processing scenarios.
- Automated data transfers using Sqoop and Airflow, resulting in a 2x reduction in data transfer time between Mysql and Hive.
- Developed Spark Streaming applications to enforce fraud detection rules, reducing false positives by 15% and enhancing the system's accuracy.
- Achieved a 40% increase in Kafka data ingestion speed, optimizing the streaming pipeline, and reducing overall processing time by 25% across various data pipeline components.
- Led performance tuning initiatives that optimized queries to yield a 5x improvement in data processing speed.
- Achieved and maintained an impressive 99.8% data pipeline up-time by leveraging Spark and Python to ingest streaming and transactional data from 8 primary sources; ensured seamless data flow and minimized disruptions
- Used Airflow to build ETL solutions that helped improve conversion rates by 16%.
- Orchestrated the development and continuous maintenance of a robust data pipeline, leveraging Google BigQuery, Google Cloud Storage, and Python to ingest and transform data from 5-6 disparate sources; achieved a 30% reduction in data ingestion time and enhanced data integrity.
- Communicated with business departments to understand needs and requests to build 20+ data pipelines for analyzing technical issues.
- Developed scalable data pipelines and ETL processes, using Python to process large datasets with a daily volume of 10 TB and improving ingestion and speed by 67%.
- Analysed large datasets with SQL to identify trends and draw insights, resulting in a 29% increase in operational efficiency.
- Integrated Data Proc for Spark job submission, achieving an average data loading speed of 1 GB per minute into BigQuery Data Warehouse. This optimization led to a 10% improvement in overall data loading performance.

Education

APJ Abdul Kalam Technological University

B.Tech in Computer Science and Engineering

India

2017 – 2021

Certifications

Infosys Certified: Python Associate

Infosys Certified: Teradata Developer

Coursera: Introduction to Data Engineering

Technical Skills

Languages: Python, SQL, PySpark, Java

Tools: Apache Spark, Hadoop, Informatica, Airflow, Jenkins, GIT, Azure DevOps

Cloud: Azure, Databricks, AWS S3, Kafka

Databases: Teradata, Snowflake

Languages

Languages: English, Malayalam, Hindi, Tamil