

NATARAJAN ANNADURAI

Big Data Engineer

☎ +91 8508929010 | ✉ natarajan.annadurai.tech@gmail.com

🌐 natarajblog1.wordpress.com | github.com/NatarajRaj | [LinkedIn](#)

📍 Rasipuram, Namakkal - 637406

PROFESSIONAL SUMMARY

Big Data Engineer with around 2 years of experience in designing, implementing, and optimizing real-time data pipelines and large-scale data solutions. Proficient in Apache Kafka, Spark, Databricks with expertise in data ingestion, stream processing, and data integration. Skilled in leveraging Java, Python, and SQL to deliver scalable, high-performance solutions for Big Data environments. Passionate about solving complex data challenges and driving data-driven decision-making.

TECHNICAL SKILLS

- **Big Data Tools:** Spark, PySpark, Azure Databricks, Kafka
- **Programming Languages:** Java, Python
- **Databases:** MySQL, PostgreSQL, Cassandra, Snowflake
- **Cloud Platforms:** Microsoft Azure (ADF, ADLS, Azure Synapse), AWS (S3, Lambda, Glue, RDS)
- **Tools & Frameworks:** Apache Airflow, Shell Scripting, Power BI, GitHub

PROFESSIONAL EXPERIENCE

Software Engineer II

MahilTechLab Solutions – Namakkal, Tamil Nadu

Feb 2025 – Present

- Architected real-time data pipelines using **Apache Kafka** and **Spark Structured Streaming** on **Azure Databricks** to process streaming data with low latency.
- Designed and implemented ETL workflows on Microsoft Azure using **Azure Data Lake Storage (ADLS)**, **Azure Data Factory (ADF)**, and **Azure Synapse**, enabling efficient ingestion, transformation, and storage of large datasets.
- Built automated data validation workflows in **Apache Airflow**, including schema checks, null value detection, and duplicate handling, integrated with **PostgreSQL** and **Azure Data Lake Storage**.

Senior Associate

Fiserv – Bengaluru, India

July 2023 – November 2024

- Designed and implemented **real-time data pipelines** using **Apache Kafka** and **Azure Databricks (Apache Spark)** to process **high-velocity data streams**, reducing latency by 30%.
 - Assisted in **Azure cloud data migration** projects using **Azure Data Factory**, optimizing data flow and ensuring zero data loss during the transition.
 - Conducted rigorous testing of **Java applications** using **JUnit** and **Mockito**, improving software reliability by 20%.
-

PROJECTS

Real-Time Data Processing Pipeline

Tech Stack: Apache Kafka, Spark Structured Streaming, Cassandra, Grafana

- Developed a real-time data ingestion and processing system to handle high-velocity event streams using Kafka and Spark.
 - Persisted processed data into **Apache Cassandra** for scalable, low-latency querying.
 - Integrated with **Grafana** for real-time metrics visualization and system monitoring.
- Impact:** Reduced data processing latency by 30% and improved system throughput by 25%.

Data Migration and Optimization Project

Tech Stack: Apache Airflow, PostgreSQL, AWS S3, Azure Data Factory

- Supported migration of data from on-premise legacy systems to **cloud-based modern platforms**.
 - Automated data validation tasks using **Apache Airflow**, including schema checks, row count matching, and type validation.
 - Coordinated with teams to ensure a seamless, disruption-free migration.
- Impact:** Achieved 99.9% system uptime and completed migration with zero data loss.

EDUCATION & CERTIFICATIONS

- **JAVA Fullstack Development Course** - *Payilagam Institute, Chennai*
Relevant Coursework: Big Data Technologies, Data Engineering, and Software Testing.
 - **Complete Guide to Databricks for Data Engineering** - *LinkedIn Learning* – Apr 2025
Skills: PySpark, Data Engineering
 - **End-to-End Real-World Data Engineering Project with Snowflake** - *LinkedIn Learning* – Apr 2025
Skills: Snowflake, ETL Pipelines, Real-Time Data Engineering
-

LANGUAGES

- **English:** Professional Proficiency
 - **Tamil:** Native Proficiency
-