

NATARAJAN ANNADURAI

Big Data Engineer

+91 8508929010 | anataraj95@gmail.com

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

Rasipuram, Namakkal - 637406

PROFESSIONAL SUMMARY

Big Data Engineer with 2 years of experience in building and optimizing real-time data pipelines. Skilled in Apache Kafka, Spark, Azure Databricks, and cloud platforms (Azure, AWS). Proficient in Java, Python, and SQL, with expertise in stream processing and data integration to enable 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 Databricks, Azure Synapse), AWS (S3, Lambda, Glue, RDS)
- **Tools & Frameworks:** Apache Airflow, Shell Scripting, Power BI, GitHub, Jenkins

PROFESSIONAL EXPERIENCE

Freelance Data Engineer (Remote)

Dec 2023 – Present

- Built real-time data pipelines using **Apache Kafka** and **Azure Databricks (Spark Structured Streaming)** for low-latency processing of high-velocity data.
- Designed and implemented ETL workflows using **Azure Data Factory**, **Azure Data Lake Storage**, and **Databricks** for scalable data transformation and loading.
- Automated data validation using **Apache Airflow**, including schema checks, null/duplicate validations integrated with **Azure Blob Storage** and **Azure PostgreSQL**.
- Containerized applications using **Docker**, ensuring consistent, isolated environments for development and testing.
- Delivered modular and cloud-optimized data engineering solutions aligned with client requirements.

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%.
 - Provided end-to-end support for **Talend ETL job execution**, ensuring seamless data ingestion and integration across **Azure-based systems**.
 - 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

- Built a **real-time data processing pipeline** using **Apache Kafka** for data ingestion and **Spark Streaming** for processing **high-velocity data streams**.
- Stored processed data in **Cassandra** for **low-latency querying** and **scalability**.
- Visualized real-time metrics and insights using **Grafana**, enabling stakeholders to monitor system performance and make **data-driven decisions**.
- **Achievements:** Reduced data processing latency by **30%** and improved system performance by **25%**.

Data Migration and Optimization

- Supported the team in migrating data from legacy systems to modern platforms, ensuring seamless transfer with minimal disruption.
 - Automated data validation and transformation using Apache Airflow, improving efficiency by 25%.
 - **Achievements:** Successfully completed migration with zero data loss and 99.9% uptime.
-

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

ACHIEVEMENTS

- Optimized **real-time data pipelines**, reducing latency by **30%** and improving system performance by **25%**.
- Improved software reliability by **20%** through rigorous testing using **JUnit** and **Mockito**.

LANGUAGES

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