AANANDHENE M

aanandhene26@gmail.com · +91 6369675002
aanandhene · Aanandhene

PROFILE

Proactive **Data Engineer** with **2+ years** of experience building scalable, cloud-ready data platforms across streaming, batch, warehouse, and lakehouse systems. Skilled in designing pipelines using Kafka, Apache Flink, Python, SQL, and Postgres. Experienced in automating ETL/ELT workflows, optimizing data models, and integrating complex datasets across environments. Highly adaptable to new tools, technologies, and data architectures, with a strong focus on real-time analytics and intelligent automation.

EDUCATION

Bachelor of Computer Science & Engineering, Kongu Engineering College CGPA - 8.83 (2020 - 2024)

SKILLS

- **Tools & Technologies:** Apache Kafka, Apache Flink, Streamlit, Apache Doris, MinIO, Airflow, Docker, GitHub, Jenkins, Superset, Microsoft Power BI, Microsoft Excel, OpenMetadata, Postman, n8n Automation
- **Backend:** Python, REST APIs
- **Databases:** SQL Server, PostgreSQL, MongoDB, Supabase
- **ETL & Data Engineering:** ETL/ELT Pipelines, Data Migration, Data Warehousing, Data Modeling, Streaming Pipelines, Data Governance, Lineage
- **Languages:** Python, SQL

EXPERIENCE

SPAN TECHNOLOGY SERVICES PRIVATE LTD. **December 2023 – Present**
A product-based technology company specializing in logistics, taxation, compliance, and enterprise data solutions across the US market.
Junior Data Engineer Erode, India

- Managed end-to-end data engineering workflows across multiple environments (Sprint, Staging, UAT, Live) involving SQL Server, Kafka, Flink, Postgres, and Redis.
- Designed and optimized high-performance ETL/ELT pipelines for streaming and batch workloads, ensuring consistent data availability for applications and analytics teams.
- Supported data warehousing initiatives by modeling schemas, defining relationships, and standardizing data flow across business domains.
- Ensured data governance and lineage using OpenMetadata by maintaining schema consistency, validation rules, and automated quality checks.
- Collaborated with application, DevOps, BI, and QA teams to troubleshoot production issues, optimize SQL queries, and improve system reliability during peak traffic seasons.
- Assisted business teams by enabling faster reporting and analytical capabilities through dashboard enhancement, optimized data sources, and performance tuning.

PROJECTS

- **Enterprise Data Lakehouse Modernization (SQL → Kafka → Flink → Doris)**
  Designed and built a scalable lakehouse integrating SQL Server, MySQL, MongoDB, and Postgres into Apache Doris using Kafka ingestion, Flink processing, MinIO storage, and Airflow orchestration. Improved analytical query speed by 70%, reduced API latency by 55%, and enabled strong data governance with lineage and schema enforcement.
- **Real-Time Payment Data Reconciliation (Stripe & Cybersource)**
  Implemented a unified reconciliation pipeline by ingesting gateway data, transactional logs, and application events into Postgres through Kafka and Flink. Added rule-based matching, anomaly detection, and late-arrival handling. Automated refund/void alerts via Slack, reducing manual checks by 85% and lowering duplicate payments by 60%.
- **IRS Data Ingestion & Standardization Pipeline**
  Built an automated ingestion process for IRS datasets with validation, deduplication, schema consistency checks, and standardized loading into Postgres. Improved data readiness by 3× and ensured cleaner inputs for compliance reporting and BI analytics.

- **Real-Time Monitoring & Observability Framework**
  Developed system-wide monitoring for Flink jobs, connector failures, event-lag spikes, skew issues, and server health (CPU, RAM, Disk). Delivered real-time Slack notifications using Python-based alerting, boosting incident detection by 80% and cutting recovery time by 50%.
- **Distributed BI Data Models & Performance Optimization**
  Designed star-schema models and optimized materialized views in Apache Doris/Postgres for revenue, tax, and user analytics. Applied indexing, partition pruning, and storage optimizations, reducing dashboard refresh times by 70% and improving query performance during peak seasons.
- **Streamlit-Based ETL Automation Studio (PyFlink Generator)**
  Created a Streamlit application that auto-generates PyFlink ETL scripts using dynamic metadata and mapping rules. Reduced ETL development time by 60%, standardized ingestion patterns, and eliminated repetitive manual coding effort.

CERTIFICATIONS

Microsoft Azure Fundamentals (Az-900) **Microsoft**

Google Data Analytics **Coursera**

AWARDS

- **Span Super Star Award** — Awarded for delivering critical projects within short timelines, independently managing complex tasks, and effectively collaborating with cross-functional teams to ensure successful delivery.