SENTHIL NAYAGAN

Senior Data Engineer
Email: senthil.nayagan@gmail.com | Phone: +91 9840405915
LinkedIn | GitHub | Twitter | Website

PROFESSIONAL SUMMARY

Senior Data Engineer with a decade of expertise specializing in designing and building enterprise-scale data platforms using open-source technologies and AWS cloud infrastructure. Proven track record of architecting modern data engineering solutions that serve 100+ tenants and process 100k+ daily transactions. Expert in data mesh architectures, lakehouse implementations, and establishing comprehensive data governance frameworks. Committed to technical excellence and team development, with demonstrated success in building and mentoring high-performing engineering teams.

Key Achievements:

- Architected enterprise-level data engineering platforms with end-to-end data governance, supporting seamless data flow from ingestion through landing, curation, and presentation zones across multiple domains
- Established automated metadata management ecosystems including data contracts, lineage tracking, quality monitoring, and data cataloging using tools like OpenMetadata and Collibra
- Built multi-tenant data platforms serving 100+ tenants with 100k+ daily transactions while maintaining GDPR/CCPA compliance through lakehouse architectures
- Led cross-functional teams through successful cloud migrations and platform modernizations, fostering a culture of technical excellence and continuous improvement

TECHNICAL SKILLS

Big Data Ecosystem:

AWS Lake Formation, AWS Glue, AWS EMR, Apache Hadoop, Apache Spark, Apache Hive, Amazon Athena, Apache Sqoop, Apache Oozie, Apache Pinot

Data Streaming:

Apache Kafka, Apache Spark Structured Streaming, Apache Flume

Programming Languages and Scripting:

Python, Scala, Rust, Java, Unix Bash Scripting

Cloud:

AWS: EMR, EC2, S3, ECS, DynamoDB, RDS, ElastiCache, Lambda, Beanstalk, Glue, KMS

GCP: Dataproc

DevOps and Orchestrations:
Jenkins, uDeploy, Docker, Airflow

Database (SQL and NoSQL):
MySQL, MongoDB, PostgreSQL

Operating Systems:
Linux (Red Hat, CentOS, Ubuntu), macOS, Windows

Source Control Management:
Git - GitHub, GitLab, Bitbucket

PROFESSIONAL EXPERIENCE

Tata Consultancy Services (TCS)

Data Architect (Data Engineering & Analytics) | February 2025 – Present

Project: Financial Insights

Provide personalized, timely, and actionable advice to users using data-driven insights.

Responsibilities:

- Designed and created Spark-based ETL platform
- Developed data ingestion framework to export insights from S3 to MongoDB
- Data quality enforced via Great Expectations

Outcomes:

- Provide personalized and actionable insights to end-users
- Generate various insights (80+)

Technology Stack:

Apache Spark, AWS EMR, AWS Glue, Athena, MWAA, CloudWatch, Secrets Manager, Python, Java, Scala, Great Expectations, SBT, Python Wheel, GitLab, Docker, Airflow (MWAA)

Tata Consultancy Services (TCS)

Data Architect (Data Engineering & Analytics) | February 2023 – December 2024

Project: Self-Serve Data Analytics Platform

Decentralized data platform built on top of AWS Lake Formation and AWS Glue in accordance with the Data Mesh architecture.

Responsibilities:

- Designed and created decentralized domain
- Created LF-tags-based federated data access and governance
- Implemented data quality strategy using Great Expectations
- Integrated with OpenMetadata to have a single source of truth for Data Discovery, Data Lineage, and Data Quality
- Implemented Transaction Data Lake (Lakehouse) using Apache Hudi

Team Management:

- Spearheaded initiatives to align team objectives with organizational goals, contributing to the development of a high-performing and cohesive team that shared a common vision, goals, and values

Outcomes:

- Improved agility, faster decision-making, and greater accountability as domain experts take ownership of their data
- Increased efficiency and reduced dependency on Data Engineering teams

- Granular-level access using LF-based Tags
- Improved security, compliance, and data governance
- Transactional ability with ACID properties, crucial for compliances such as GDPR and CCPA

Technology Stack:

AWS Lake Formation, AWS Glue, AWS EMR, Apache Spark, Apache Hudi (Lakehouse), Python, Scala, Great Expectations, OpenMetadata, SBT, Python Wheel, GitLab, Docker, Airflow

Accenture

Technical Architect - Data Engineering | March 2020 – January 2022

Project: myWizard (Multi-Tenant Platform)

Integrated automation across the software engineering lifecycle with multi-tenancy support.

Responsibilities:

- Designed and built data pipeline with seamless data flow from landing to presentation zones, constructed with durability and consistency capabilities borrowed from Delta Lake
- Created data ingestion for both batch and real-time ingestion (via CDC), orchestrated via Airflow with automated Data Contract
- Created ETL and data processing using Apache Spark
- Automated data catalog, data lineage, and monitoring via Collibra and Splunk tools
- Enforced data governance in every facet: IAM, RBAC, PoLP, Data at Rest, Data in Motion, CSFLE; Envelope Encryption (AWS KMS) method was used
- Exposed data using RESTful APIs and Databricks Delta Sharing to BI apps and ML models

Outcomes:

- Able to serve 100+ tenants
- Handled 100k transactions per day
- Using Data Lake, efficiently met GDPR/CCPA compliance

Technology Stack:

AWS EMR, HDP (Hortonworks), Apache Spark, Apache Hive, Databricks, Delta Lake, Python, Scala, Splunk, Collibra, AWS KMS (Envelope Encryption), SBT, Python Wheel, Azure DevOps, Docker, Airflow, TFS, GitHub, Scrum

Accenture

Lead Data Engineer | May 2016 – February 2020

Project: Cigna IBIS

Workflow creation engine that abstracts Hadoop internals of ingesting RDBMS data. Created data workflow using Hadoop stacks (Oozie, Sqoop) and was instrumental in real-time data pipeline in Kafka.

Responsibilities:

- Involved in the design and development of the ingestion-workflow engine known as IBIS
- Migrated the data streaming platform from Flume to Kafka to increase scalability and minimize latency
- Created both Kafka Producers and Consumers to move data from Hive to other data stores
- Customized Kafka Connect: Enhanced the ability of data extraction from Confluent Kafka's Registry-less Avro Converter

Outcomes:

- Built IBIS as a single-source data ingestion workflow engine used across Cigna
- Reduced data streaming latency and achieved high throughputs

Technology Stack:

Cloudera CDH, Apache Sqoop, Apache Spark, Apache Hive, Apache Impala, Python, Scala, Apache Kafka, Confluent Kafka (Schema Registry), Maven, SBT, Jenkins, uDeploy, GitLab, Scrum, Kanban

EDUCATION

Master of Computer Applications (MCA)

Bharathidasan University, Tiruchirappalli, India

1999 – 2002

CERTIFICATIONS

AWS Certified Developer - Associate

Issued by: Amazon Web Services (AWS)
Issued: April 2022 – April 2025
Credential: View Certificate

Deep Learning

Issued by: Coursera
Issued: December 2018 – No Expiration Date
Credential ID: S3A9F2SU7D2W

Reactive Architecture – Level 2

Issued by: Lightbend, Inc.
Issued: December 2022 – No Expiration Date
Credly Badge ID: b74abbd1-1896-4d5e-8b2c-48d89a2d24d1