SENTHIL NAYAGAN

Data Engineering and Strategy

LinkedIn: SenthilNayagan

Twitter: @SenthilNayagan

GitHub: SenthilNayagan

Website: SenthilNayagan.net

Email:
senthil.nayagan@gmail.com

Phone:
+91 9840405915

PROFESSIONAL SUMMARY

I am a seasoned IT professional with a decade of expertise in designing and building complex business solutions with open-source technologies. I've been working in the field of data engineering and cloud migration (AWS) for the past few years. Through effective leadership and mentorship, I contribute to building high-performing teams that deliver exceptional results.

- Designed and built enterprise-level data engineering platforms that collect data from a variety of sources and optimize them along the **"modern" data pipeline**'s journey. These data pipelines were created to support seamless data flow from landing to presentation zones, and **data governance** is enforced at every step along the journey.
- Also, established and automated all metadata and other activities such as **data contract**, **data lineage**, **data catalogue**, and **monitoring**.
- **Mentorship and Team Development:** I'm passionate about team building—actively managing and mentoring my team to transform them into a high-performing team.
- **Personal Interest:** Passionate about reactive systems, functional programming, and modern programming languages like Rust, which is well known for memory safety and zero-cost abstraction.

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

SCMs
Git - GitHub, GitLab, Bitbucket

EXPERIENCE

Tata Consultancy Services (TCS) -> 02/2023 – Present

Data Architect (Data Engineering & Analytics)

Self-Serve Data Analytics Platform
It's a decentralized data platform built on top of AWS Lake Formation and AWS Glue in accordance with the Data Mesh architecture.

Responsibilities
Designed and built the following:

- 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** aka 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, which is crucial for compliances such as GDPR and CCPA.

Technology Stacks

- **Hadoop Ecosystems:** AWS Lake Formation, AWS Glue, AWS EMR, Apache Spark, Apache Hudi (Lakehouse)
- **Programming Languages:** Python, Scala
- **Data Quality:** Great Expectations
- **Metadata Discovery:** OpenMetadata
- **Build Tools & SCM:** SBT, Python Wheel; GitLab
- **DevOps, Orchestrations:** Docker, Airflow

Accenture -> 03/2020 – 01/2022

Technical Architect - Data Engineering

myWizard (Multi-Tenant Platform)
myWizard is an integrated automation across the software engineering lifecycle with muti-tenancy support.

Responsibilities
Designed and built the following components:

- **Data Pipeline** with seamless data flow from landing to presentation zones, constructed with the durability and consistency capabilities borrowed from Delta Lake.
- **Data Ingestion** – Created for both batch and real-time ingestion (via CDC); Orchestrated via Airflow; Created an automated Data Contract.
- **ETL and Data Processing** - Created using Apache Spark.
- **Data Catalogue**, **Data Lineage and Monitoring** – Automated via Collibra and Splunk tools.
- **Data Governance** is enforced in every facet: IAM, RBAC, PoLP, Data at Rest, Data in Motion, CSFLE; Envelop Encryption (AWS KMS) method was used.
- **Data Expose** - RESTful APIs and Databricks Delta Sharing are used to expose data to BI apps and ML models.

Outcomes

- Able to served 100+ tenants
- Handled 100k transactions per day
- Using Data Lake, we were able to efficiently meet GDPR/CCPA compliance.

Technology Stacks

- **Hadoop Ecosystems:** AWS EMR, HDP (Hortonworks), Apache Spark, Apache Hive, Databricks, Delta Lake
- **Programming Languages:** Python, Scala
- **Monitoring, Meta Operations:** Splunk, Collibra
- **Security:** AWS KMS (Envelop Encryption)
- **Build Tools:** SBT, Python Wheel
- **DevOps, Orchestrations:** Azure DevOps, Docker, Airflow
- **SCM:** TFS, GitHub
- **Agile Methodology:** Scrum

EXPERIENCE

Accenture -> 03/2020 – 01/2022

Technical Architect - Data Engineering

myWizard (Multi-Tenant Platform)
myWizard is an integrated automation across the software engineering lifecycle with muti-tenancy support.

Responsibilities
Designed and built the following components:

- **Data Pipeline** with seamless data flow from landing to presentation zones, constructed with the durability and consistency capabilities borrowed from Delta Lake.
- **Data Ingestion** – Created for both batch and real-time ingestion (via CDC); Orchestrated via Airflow; Created an automated Data Contract.
- **ETL and Data Processing** - Created using Apache Spark.
- **Data Catalogue**, **Data Lineage and Monitoring** – Automated via Collibra and Splunk tools.
- **Data Governance** is enforced in every facet: IAM, RBAC, PoLP, Data at Rest, Data in Motion, CSFLE; Envelop Encryption (AWS KMS) method was used.
- **Data Expose** - RESTful APIs and Databricks Delta Sharing are used to expose data to BI apps and ML models.

Outcomes

- Able to served 100+ tenants
- Handled 100k transactions per day
- Using Data Lake, we were able to efficiently meet GDPR/CCPA compliance.

Technology Stacks

- **Hadoop Ecosystems:** AWS EMR, HDP (Hortonworks), Apache Spark, Apache Hive, Databricks, Delta Lake
- **Programming Languages:** Python, Scala
- **Monitoring, Meta Operations:** Splunk, Collibra
- **Security:** AWS KMS (Envelop Encryption)
- **Build Tools:** SBT, Python Wheel
- **DevOps, Orchestrations:** Azure DevOps, Docker, Airflow
- **SCM:** TFS, GitHub
- **Agile Methodology:** Scrum

Accenture -> 05/2016 – 02/2020

Lead Data Engineer

Cigna IBIS
IBIS is a workflow creation-engine that abstracts the Hadoop internals of ingesting RDBMS data. Created data workflow using Hadoop stacks (Oozie, Sqoop). Instrumental in real-time data pipeline in Kafka. Trained resources into Hadoop stacks.

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: Customized and 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 Stacks

- **Hadoop Ecosystems:** Cloudera CDH, Apache Sqoop, Apache Spark, Apache Hive, Apache Impala
- **Programming Languages:** Python, Scala
- **Data Streaming:** Apache Kafka, Confluent Kafka (specially for *Schema Registry*)
- **Build Tools:** Maven, SBT
- **DevOps Tools:** Jenkins, uDeploy
- **SCM:** GitLab
- **Agile Methodology:** Scrum, Kanban

EXPERIENCE

Accenture -> 11/2004 – 04/2016

Backend Engineer (Junior->Senior->Lead)

Responsibilities

- Collaborated with the product owners to comprehend end-user needs, create use cases, and then convert them into a technically proficient solution.
- Designed and developed ORM-based layer data access layer.
- Developed messaging queue services.
- Developed CI/CD pipelines to enable faster feedback while also improving quality and testability.
- Optimized code quality through writing unit tests, automation, implementing code principles, and performing code reviews.
- Taken part SQL query optimization.
- Directly responsible for managing and developing junior team member by coaching, enhancing skills, and offering constructive feedback.

Technology Stacks

- **Programming Languages:** Java, Python
- **Rule Engines:** Drools
- **Web and App Servers:** Tomcat, JBoss
- **ORM and DI:** Hibernate, Spring
- **MQ Tools:** RabbitMQ
- **Build Tools:** Ant, Maven
- **Database:** DB2, MySQL
- **DevOps Tools:** Jenkins, uDeploy
- **SCM:** SVN, CVS, GitHub

EDUCATION

Bharathidasan University, Tiruchirappalli, India

Master of Computer Applications – MCA
1999 - 2002

CERTIFICATIONS

AWS Certified Developer - Associate

By: Amazon Web Services (AWS)
Issued: April 2022 – April 2025
Credential:
https://www.credly.com/badges/b2148907-fefd-4df3-912d-ae0326bc253f

Deep Learning

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

Reactive Architecture – Level 2

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