NARENDRAN MUDADI

Technology Specialist
Big Data, AWS.

✉ narendas10@gmail.com

📞 +91 9884863650

🔗 narendran-mudadi

🔘 NarenBot

🪪 OVERVIEW

Data Engineer with 4 years of experience in designing, developing and optimizing data pipelines and ETL workflows in cloud environments. Skilled in building scalable and cost-efficient data solutions to drive business insights and automation.

🧠 TECHNICAL SKILLS

Programming Languages: Python, SQL.

Python Libraries: PySpark, Pandas.

Cloud Services: AWS (S3, Glue, Redshift, Lambda, Athena, EC2, EMR, DMS)

Data Engineering: ETL Pipelines, Data Warehousing, Data Modeling, Data Lakes, PowerBI.

Big Data Technologies: Apache Spark, Kafka, Flink, Apache Nifi.

Databases: MySQL, PostgreSQL, MongoDB, DynamoDB, CassandraDB.

🖐 CERTIFICATES

SQL For Data Professionals ⧉

Full Stack Data Science ⧉

🔧 TOOLS & TECHNOLOGIES

Version Control: Git, GitHub.

Workflow Orchestration: Apache Airflow, AWS Step Functions.

DevOps & CI/CD: Git, Docker.

Machine Learning: Scikit-Learn.

Monitoring & Logging: CloudWatch, Prometheus and Grafana.

🎓 EDUCATION

**Information Technology, B.Tech
- Rajalakshmi Engineering College**
Sep 2005 – Apr 2009

💼 PROFESSIONAL EXPERIENCE

**AWS Data Engineer | June 2023 – Mar 2025
Physics Wallah - iNeuron.ai | Bangalore**
- Designed and implemented scalable ETL pipelines using AWS Glue and PySpark to process 50–60 GB of data daily, saving the output in Parquet format.
- Implemented partitioning and caching strategies in PySpark to manage large datasets effectively, improving query performance and minimizing memory consumption.
- Collaborated with analytics and business teams to deliver actionable insights from structured and semi-structured data.

**Python Data Engineer | Mar 2020 – June 2022 | RRD |
Senior Graphic Designer | June 2012 – Mar 2020**
- Implemented data cleansing and validation pipelines using PySpark to standardize and enrich structured and unstructured data before loading it into AWS Redshift and S3.
- Generated PDF files from extracted API data using Python, stored them in S3 for scalable access and managed references in PostgreSQL for efficient tracking.
- Implemented API monitoring and auditing by logging response_status, extraction_time and metadata in CassandraDB, ensuring operational visibility and compliance.
- Integrated Kafka with Flink for real-time event processing and streamed the processed data into S3.
- Reduced storage costs by 50-60% through optimized data partitioning and compression in S3 and Redshift.

**Web Designer | Feb 2012 – June 2012
Concentrix | Chennai**

**Programmer | Oct 2009 – Jan 2011
Intellectual Info Technologies | Chennai**

📂 RECENT PROJECTS AND USECASES

Real-Time Data Streaming for E-commerce Analytics
- Designed a real-time data pipeline using AWS Kinesis to ingest and process user activity logs.
- Store data like session_id, user_id, product_views and time_spent in CassandraDB for personalized recommendations.
- Enabled fraud detection alerts using AWS Lambda and SNS notifications.

Automated ETL Pipeline for Financial Data Processing
- Developed an ETL pipeline using AWS Glue to extract data from RDS, transform it using PySpark and load it into Redshift.
- Applied PySpark DataFrame APIs for efficient data transformations, reducing processing time by 30%.
- Implemented AWS Step Functions to orchestrate multi-step data processing tasks.

Customer Churn Prediction using AWS SageMaker
- Collected and preprocessed customer transaction data using AWS Glue.
- Built and deployed an ML model on SageMaker to predict churn probabilities.
- Integrated predictions with AWS Lambda to trigger retention campaigns via AWS SES.