DIVYANSH PATEL

Data Engineer | Software Engineer

divyansh9144@hotmail.com | linkedin.com/in/pateldivyansh | github.com/divyansh8866 | divyanshpatel.com | medium.com/@divyansh9144

PROFESSIONAL SUMMARY

Results-driven Data Engineer and Software Engineer with 7+ years of experience designing, building, and optimizing scalable data pipelines and cloud-native applications. Expertise in AWS ecosystem (Glue, Lambda, S3, DynamoDB, Redshift), Apache Spark, Apache Hudi, and serverless architectures. Published author of books on data engineering, agentic AI, and Apache Hudi. Proven track record of delivering high-impact data lakehouse implementations, ETL automation, and infrastructure modernization across fast-paced environments.

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, Shell/Bash

Cloud & Infrastructure: AWS (Glue, Lambda, S3, DynamoDB, Redshift, SQS, SNS, Kinesis, DMS, CodeWhisperer, CloudFormation), Serverless Framework

Data Engineering: Apache Spark, Apache Hudi, Apache Airflow, PySpark, dbt, ETL/ELT Pipelines, Data Lakehouse Architecture

Databases & Storage: DynamoDB, Redshift, PostgreSQL, S3, Data Warehousing

Tools & Frameworks: Git, Docker, CI/CD, REST APIs, Terraform, LangFlow, MCP Servers

Concepts: Data Pipeline Design, Incremental Processing, Schema Evolution, Data Quality, RAG Systems, Agentic AI

PROFESSIONAL EXPERIENCE

Data Engineer — *JobTarget*

Stamford, Connecticut | April 2022 — Present

- Design, implement, and maintain scalable data pipelines using AWS Glue, Lambda, and S3, processing data across a comprehensive data lakehouse framework.
- Build robust ETL workflows with Apache Spark and PySpark, enabling efficient extraction, transformation, and loading from diverse data sources into target systems.
- Integrate Apache Hudi for large-scale incremental data processing, ensuring data freshness, schema evolution, and optimized storage within the lakehouse.
- Implement serverless architectures using AWS Lambda and the Serverless Framework, driving cost optimization and elastic scalability in production environments.
- Optimize data storage and retrieval strategies on AWS S3, reducing query latency and storage costs for analytics workloads.
- Collaborate with cross-functional teams to align data infrastructure with business requirements, delivering reliable data products for downstream consumers.

Research Assistant — *University of New Haven*

Connecticut | September 2020 — December 2021

- Collaborated with engineering teams to improve system consistency and user experience by integrating feedback loops and optimizing application features.
- Led the application lifecycle, guiding coding, debugging, and code review processes to maintain high standards of quality and project integrity.
- Developed web applications with clean, maintainable code using modern frameworks, delivering responsive and accessible interfaces.
- Applied scientific and statistical coding techniques to process research data, extracting meaningful insights and identifying patterns across datasets.

- Generated detailed research reports summarizing methodologies, key findings, and actionable implications for stakeholders.

Software Engineer — *NewTech Engineers*

Gujarat, India | June 2017 — August 2019

- Collaborated with cross-functional teams to understand data requirements and deliver software solutions aligned with business objectives.
- Designed and implemented data extraction, transformation, and loading pipelines from diverse sources into target systems.
- Implemented validation measures to ensure accuracy and integrity of data throughout the software development lifecycle.
- Optimized existing processes through systematic troubleshooting, improving system performance and reducing development cycle times.

EDUCATION

MS in IT Project Management — *New England College*

Henniker, New Hampshire, USA

MS in Computer Science — *University of New Haven*

West Haven, Connecticut, USA

BE in Electrical Engineering — *Gujarat Technological University*

Gujarat, India

PROJECTS

- **DynamoDB Sync Pipeline —** Real-time DynamoDB-to-DynamoDB synchronization pipeline for cross-region data replication.
- **Langflow Deployment Boilerplate —** Production-ready deployment template for LangFlow-based AI agent workflows.
- **MCP Server DataHub —** Dockerized MCP server with HTTP SSE transport for DataHub metadata integration.
- **PDF RAG System —** Retrieval-Augmented Generation system for intelligent document querying and knowledge extraction.

PUBLICATIONS & BOOKS

- **Agentic AI —** Autonomous AI Agents for Data Engineering
- **Data Engineering Algorithms —** Blueprint for Algorithm-Driven Data Engineering
- **Apache Hudi Configuration Handbook —** Complete Reference for Configuring Apache Hudi
- **Data Pipelines Handbook —** Design, Build, and Manage Efficient Data Pipelines

Author of DataDecode — a newsletter on data engineering and cloud (datadecode.divyanshpatel.com)

CERTIFICATIONS

- AWS Data Services
- AWS Database Migration Service (DMS)
- AWS Edge Storage, Data Transfer & File Transfer Services
- AWS Redshift Essentials
- AWS Storage Data Protection Services
- AWS for Developers: DynamoDB
- AWS for Developers: SNS, SQS & SWF
- Advanced Python
- Amazon CodeWhisperer
- Apache PySpark by Example
- Data Engineering with dbt
- Introduction to AWS Automation Tools
- Introduction to Data Warehouses
- Learning Apache Airflow
- Python Advanced Design Patterns
- Serverless Computing with AWS Lambda