

Divyansh Patel

Data Engineer

West Haven, CT | +1 (203) 936-7278 | divyansh9144@hotmail.com | <https://divyanshpatel.com>

Linkedin: <http://linkedin.com/in/pateldivyansh/>

Experience

JobTarget LLC April, 2022 - Present
Data Engineer

- Optimized Apache Spark processes for large data sets by implementing partitioning and caching techniques, **increasing processing speed and resource utilization efficiency by 25%**.
- Designed and maintained data warehouse using AWS Redshift, enabling over 10 data scientists to efficiently analyze large datasets, **accelerating the generation of insights by 40%**.
- Developed and **orchestrated ETL pipelines** using Apache Airflow for big data and streaming data, automating workflows and **reducing data processing costs by 30%** on AWS Cloud.

University of New Haven September, 2020 - Dec, 2021
Research Assistant

- Developed a database and backend system using MySQL and Django for a web application, **ensuring secure data storage**, quick data access, and seamless front-end integration.
- Implemented version control using Git and established **continuous integration** with Jenkins, ensuring high **code quality and smooth collaboration** among the research team.
- Architected a custom cloud infrastructure on AWS for internal use, **increasing data storage capacity by 50%** and **reducing operational costs by 25%** while ensuring scalable and secure data processing.

NewTech Engineers Jun, 2017 - Aug, 2019
Software Engineer

- Integrated payment processing and invoice generation into the system using Stripe API, streamlining financial operations and **reducing manual invoicing errors by 25%**.
- Spearheaded the migration of the parts inventory system to a microservices-based architecture, enabling the system to **handle a 50% increase in inventory data** without performance degradation.
- Implemented **real-time data analytics** using Apache Kafka within the stock management system, providing actionable insights that improved inventory management efficiency.

Education

University Of New Haven Jan, 2019 - Jan, 2022
Computer Science Masters

Gujarat Technological University Jun 2015 - May 2019
Electrical Engineering Undergrad

Projects

Real-time Data Ingestion Pipeline

Designed and implemented a scalable real-time data ingestion pipeline on AWS, handling high-velocity streaming data. The pipeline enabled seamless data ingestion, processing, and storage, ensuring the delivery of accurate and timely insights. This solution supported data-driven decision-making, improved response times, and enhanced the overall efficiency of data analysis processes.

Automated ETL Pipeline with Apache Airflow

Developed a robust ETL pipeline using Apache Airflow, automating the entire process of data extraction, transformation, and loading. The pipeline not only cleaned data but also ensured compliance by masking sensitive information, significantly improving data accuracy. This automation reduced the need for manual intervention, increased processing speed, and provided a reliable and secure data flow for downstream analytics and reporting, ultimately enhancing overall data management efficiency.

AI-Powered Data Enrichment Pipeline

Worked on an AI-powered data enrichment pipeline that enhanced raw data by integrating additional contextual information from external data sources. Leveraged machine learning models to automatically enrich and refine data, improving the value and accuracy of insights. This pipeline enabled more informed decision-making by providing enriched datasets, which enhanced the quality of analytics and reporting, and reduced the need for manual data augmentation.

Big Data Migration and Data Lakehouse Creation

Led a big data migration from MySQL to Amazon S3 using AWS DMS, and built a data lake with Apache Hudi. Employed Medallion Architecture, organizing data into Raw, Bronze, and Silver tables for efficient storage and processing. This project improved data management and analytics, making data easily accessible and scalable, while also enhancing data quality and integrity across the organization.

Skills

Technical

Python, Serverless, Django, Apache Spark, SQL, Terraform, ELK Stack, Docker, Kubernetes, DynamoDb, Apache Hudi, PostgreSQL, Apache Hadoop, AWS, MySQL, Microsoft SQL Server, Apache NiFi, Amazon Redshift, Google BigQuery, Snowflake, Power BI, Git, Jenkins, GitLab CI, Airflow

Soft

Problem-solving, Communication, Time management, Attention to Detail, Adaptability, Leadership