Dhawal Negi

📞 +1(857)-390-5835　✉ negi.d@northeastern.edu　in linkedin.com/in/dhawalnegi1/　🌐 github.com/dhawalnegi1

AWS Certified Data Engineer with strong foundations in **ETL, data Warehousing, and cloud services**, skilled and experienced in building scalable solutions and enhancing data-driven decision-making processes.

WORK EXPERIENCE

AI Researcher and Teaching Assistant | *Northeastern University*　　　　　　**Sept 2024 - Present**
Snowflake, AWS, Python, LLMs, AI Agents, Langchain　　　　　　　　　　　　　　*Boston, MA*
- Built **advanced data workflows on cloud** that fused AI capabilities with data engineering for business applications.
- Conducted PoCs on building data apis using **Snowpark Container Service** and Snowflake Warehouses
- Applied AI tools (OpenAI, Perplexity, Hugging Face) to enhance **big data processing and insights**.
- Integrated advanced AI techniques into big data labs, bridging theory with **cutting-edge AI practices**.
- Developed AI-focused teaching materials, empowering students to harness big data for data-driven decisions.

Cloud Data Engineer Co-op | *Precision Planting*　　　　　　　　　　　　　　**Jan 2024 - Aug 2024**
Python, ETL Pipeline, AWS Cloud, Tableau, GitHub Action, Data Modelling　　　　　*Bloomington, IL*
- **Led** the AWS data lakehouse implementation, efficiently consolidating data and reducing retrieval times by 60%.
- Developed automated data validation scripts and cataloging, reducing error rates by 40% and **enhancing data quality**.
- Analyzed sales trends using Tableau for multiple product lines, improving forecasting accuracy and **resource planning**.
- Streamlined data pipelines with AWS Glue and Spark, **reducing data silos** and boosting efficiency across teams.
- Designed **high available and scalable setups** with AWS CDK, achieving 99.9% uptime for critical databases.
- **Optimized deployment processes** using GitHub CI/CD actions, achieving a 60% reduction in deployment time.
- Analyzed sales data trends for multiple product lines, improving forecasting accuracy and **better resource planning**.
- **Collaborated** with cross-functional teams to deliver highly available data services for business on time and within budget.

Senior Software Engineer (Data) | *Newfold Digital*　　　　　　　　　　　　**Jan 2020 - Aug 2022**
Kafka, PostgreSQL, Python, Java, Data pipeline, REST Api　　　　　　　　　　　*Bangalore, India*
- Built pipelines for **change data capture** with PostgreSQL and Kafka, enhancing sync speed by 10 times.
- Reduced deployment time by 50% by Dockerizing and Migrating legacy systems to Containerized OKD setup.
- Configured Kong Gateway with LUA plugins, ensuring **secure data transmission** and lowering service load by 25%.
- Developed **CI/CD pipelines** for OKD using **Helm**, integrating automated testing to reduce errors by 40%.
- Rebuilt legacy payment gateway module as a Spring Boot **microservice**, cutting request processing time by 30%.
- **Integrated APIs** from India's GST portals to automate invoice registration, enhancing billing compliance and accuracy.

EDUCATION

Northeastern University | Master of Science in Data Architecture and Management　**Sep 2022 - Dec 2024**
Courses - Gen AI Data Engineering, Data Modelling, Big Data Architecture and Analytics, Data Governance　*Boston, MA*
Visvesvaraya Technological University | B.E in Computer Science and Engineering　**Jul 2016 - Jul 2020**
Courses - Data Mining and Warehouse, Cloud Computing, Machine Learning, Advanced Algorithms　*Bangalore, India*

TECHNICAL SKILLS

Programming Languages	: Python, SQL, Shell script, Java, Spark, R, JavaScript
Databases	: PostgreSQL, MySQL, Oracle, Snowflake, DynamoDB, Redis, MongoDB, Neo4J
ETL Tools	: AWS Glue, Databricks, Azure Data Facory, DBT, Talend, Apache Kafka
Orchestration and Enterprise	: Apache Airflow, Docker, Amazon CDK, Jenkins, Terraform, Kubernetes, Datadog
Cloud & DevOps	: GCP(Bigquery, DataPrep), AWS(S3, Lambda, EC2, Athena, Iac), Github actions
Certifications	: **AWS Solution Architect - Associate**, GCP Professional Data Engineer (Pursuing)

ACADEMIC PROJECTS

SnowGPT: Langchain based snowflake querying | *Snowflake, Streamlit, Langchain, OpenAI*
- Loaded and **transformed datasets into Snowflake**, ensuring accurate analytics with consumer and app usage data.
- **Developed UDFs** for behavior analysis, identifying user interests and demographics for enhanced insights.
- Converted English queries to SQL using **Langchain**, enabling seamless data retrieval from Snowflake.

Health Connect: Personalized Health Digest Platform | *Databricks, Neo4j, AI Agent, Spark*
- Engineered data pipelines with **Databricks and Spark**, automating ETL processes for processing patients medical history
- Implemented knowledge graph with Neo4j, **optimizing medical history** for personalized access.
- Developed **RESTful APIs** with FastAPI for JWT authentication, ensuring secure and seamless data interaction.

Multi-city Food Inspection Analysis | *Python, Talend, PowerBI, Tableau, SQL*
- Conceptualized **dimensional models and implemented ETL pipelines** with Talend for efficient data warehousing.
- Optimized data loading processes, facilitating seamless transition from **staging to facts and dimensions.**
- Crafted visually **engaging reports**, improving stakeholder understanding of multi-city food inspection data.