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

Cloud Data Engineer Co-op | *Precision Planting*   **Jan 2024 - Aug 2024**
Python, ETL Pipeline, AWS Cloud, GitHub Action, ETL, 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**.
- 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 | *Newfold Digital*   **June 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.
- **Migrated** legacy systems to a dockerized setup on OKD, enhancing scalability and reducing deployment times by 50%.
- 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.

Software Engineering Intern | *Newfold Digital*   **Jan 2020 - May 2020**
Python, Java, Jenkins, Pytest, RPA   *Bangalore, India*
- Implemented RPA framework to **automate deployment with testing**, reducing manual testing time by 70%.
- **Designed** and integrated database models for the persistent cart, ensuring data integrity and platform reliability.
- Enhanced e-commerce platforms by integrating session data, **improving user experience** and retention.

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, 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(Lambda, EC2, Athena, Iac), Microservices, Github actions
Certifications	: **AWS Solution Architect - Associate**, GCP Professional Data Engineer (Pursuing)

ACADEMIC PROJECTS

NewsSphere: Personalized News Digest Platform | *Airflow, OpenAI, Docker, Pinecone, Google Cloud, Fast API*
- Engineered scalable data pipelines with **Airflow on GCP**, automating ETL processes for news data every 8 minutes.
- Implemented data storage systems with Azure SQL and Pinecone, **optimizing news indexing** for personalized access.
- Developed **RESTful APIs** with FastAPI for JWT authentication, ensuring secure and seamless data interaction.

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.

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.