

Atish Dhamala

Vermillion, SD (Open to relocation)

(605)-910-2476 | atish.dhamala@gmail.com | [LinkedIn.com/in/atishdhamala](https://www.linkedin.com/in/atishdhamala)
github.com/Adh101

SUMMARY

Data Engineer with 3+ years of experience designing scalable data pipelines and robust data models for analytics. Proficient in Snowflake, dbt, SQL, and Airflow, with hands-on experience using AWS services to support both real-time and batch processing. Demonstrated success in enhancing data quality and reducing costs through end-to-end pipeline ownership and collaborative cross-functional initiatives.

EDUCATION

University of South Dakota

Aug 2024 - Dec 2025

Masters, Artificial Intelligence

- **Coursework:** Machine Learning, Advanced Data Mining, Computer Vision, Artificial Intelligence, Information Storage and Retrieval, Design and Analysis of Algorithms, Reinforcement Learning

Pokhara University

Sep 2016 - Aug 2020

Bachelor of Engineering, Information Technology

EXPERIENCE

SuperAI Brain Inc.

Feb 2022 - Aug 2024

Data Engineer - Remote

San Francisco, United States

- Designed the data model for client datasets for financial and pharmaceutical companies and orchestrated large-scale ETL and ML data pipelines using AWS, Databricks, Apache Spark, and Talend, processing 2M+ merchant transactions daily; optimizing the runtime by 40%.
- Built a Python-based data validation framework that detected 95% of anomalies, enhancing merchant data quality and reducing manual QA efforts by over 20 hours per week.
- Managed Snowflake data warehouses and implemented comprehensive unit, integration, and data quality tests, reducing production issues by 85% and enabling 3× faster analytics reporting across 20+ data sources.
- Led the team to build the automated Exploratory Data Analysis (EDA) and data cleaning features and integrated them into the SuperAI platform(self-service data analytics platform).
- Integrated BI tools (Power BI, Tableau) with an in-house LLM to facilitate natural language ad hoc analysis, thereby improving non-technical user access to data insights.
- Collaborated with cross-functional data scientists, modelers, and stakeholders to align data infrastructure with ML modeling needs, ensuring robust pipeline development and data governance while achieving 100% client satisfaction on 5 major projects.

Ingrails Pvt. Ltd.

Mar 2021 - Feb 2022

Research Analyst - Data

Lalitpur, Nepal

- Developed Python/SQL pipelines to ingest and clean multi-source marketing and CRM data, enhancing downstream model accuracy and adhering to data governance principles.
- Built a customer segmentation and churn prediction model using K-Means clustering and logistic regression on 100K+ customer behavior records, identifying high-risk segments and enabling targeted campaigns that reduced churn by 30%.
- Created Tableau dashboards and automated reporting tools for cross-functional teams, reducing manual reporting time by 40%.

TechAxis Nepal

Aug 2023 - Dec 2023

Teaching Instructor

Lalitpur, Nepal

- Taught Data Analysis with Python and Machine Learning with Python, mentoring students on analytics, ETL/ELT processes, predictive modeling, and pipeline deployment.

Freshktm Pvt. Ltd.

Apr 2020 - Feb 2021

Data Analyst

Bhaktapur, Nepal

- Built SQL-based supply chain pipelines integrating ERP inventory systems, improving operational efficiency by 20%.
- Developed ARIMA and Prophet-based forecasting models, improving prediction accuracy by 25% and reducing stockouts by 15%.
- Developed and maintained Tableau dashboards with weekly refreshes, reducing executive reporting time by 30%.

SKILLS

- **Programming Languages:** Python, SQL, R, Javascript
- **Databases:** PostgreSQL, MySQL, Snowflake, DuckDB
- **Data Engineering:** Snowflake, dbt, Databricks, Apache Spark, Apache Flink, Apache Kafka, Talend, Big Data Platforms
- **Cloud & DevOps:** AWS (EC2, S3, Glue, Athena), GCP, Azure, Docker, GitHub Actions (CI/CD), Airflow, cron
- **Analytics & ML:** Pandas, Numpy, scikit learn, PyTorch, Tensorflow
- **BI & Visualization:** Tableau, Power BI, Plotly, Seaborn, Matplotlib
- **Knowledge Domains:** Data Modeling, ETL/ELT Pipelines, Software Engineering Principles, Data Warehousing, Machine Learning, Deep Learning, Data Governance

PROJECTS

Real-Time Stock Market Data Pipeline for Scalable Analytics

2025

- Developed a real-time data pipeline to support stock market analytics use cases such as price movement monitoring, anomaly detection, and trade signal generation.
- Enabled scalable ingestion of live stock data using Kafka producers and consumers, reducing latency in data delivery and eliminating batch processing delays.
- Streamlined data storage by automating structured JSON exports to AWS S3, ensuring durability and availability for downstream analytics and machine learning workflows.
- Solved infrastructure-level challenges (e.g., IAM permission conflicts, broker connection failures, and EC2 service misconfigs) to ensure reliable data streaming and fault tolerance.

Trail-Trekker Analytics Pipeline SQLMesh + DuckDB

2025

- Delivered an end-to-end data pipeline tracking user subscription plan switch paths (from to, monthly to annual) and cancellations using DuckDB for warehousing and SQLMesh for seeding, staging, and modeling the data.
- Built dimensions (with SCD-1 and 2 types) and fact data model to enrich facts (price/limits) for downstream churn, upgrade/downgrade, and MRR/ARR analyses, and orchestrated hourly updates, removing manual runs and keeping dashboards current.

End-to-end Modern Data Warehouse & BI Automation

2025

- Designed Medallion-structured Data Warehouse (Bronze-Silver-Gold) integrating ERP & CRM data into PostgreSQL.
- Developed a fully automated analytics pipeline using Python and Google Sheets API to feed a dynamic, interactive Tableau Public dashboard from PostgreSQL data warehouse, visualizing KPIs such as Total Sales, AOV, and regional/category performance.

IPL Data Analytics Pipeline using AWS and Databricks

2025

- Designed and implemented a cloud-native ETL pipeline to extract raw IPL cricket data from the data.world, ingest into Amazon S3, and orchestrate transformations within Databricks.
- Built Spark-based transformations in Databricks notebooks to clean, standardize, and analyze large datasets, enabling advanced insights on player and match performance.
- Delivered exploratory data analysis (EDA) and visualizations within Databricks, improving accessibility of cricket analytics for downstream reporting and BI use cases.

Brain MRI 3D Image Segmentation using Deep Learning

2024

- Implemented 3D brain MRI segmentation models using U-Net and Attention U-Net architectures to segment brain tumors and structures from the BraTS 2020 dataset, achieving high precision in identifying tumor cores, edema, and enhancing regions with Dice Coefficients and IoU metrics.
- Developed end-to-end training pipeline leveraging TensorFlow and Keras to preprocess and augment 3D MRI data, fine-tune models using Adam optimizer with Total Loss, and evaluate segmentation performance, resulting in improved segmentation accuracy using IoU and Dice Coefficient, achieving accurate segmentation (Final IoU: 0.72 training, 0.63 validation).

CERTIFICATIONS

- **Data Scientist with Python:**2020
- **SQL Fundamentals Track:**2021
- **Associate Data Engineer Track:**2025
- **Datacamp Data Fellowship, Code For Nepal:**2020-2025
- **AWS Data Engineer Associate Certification:**In Progress