

Ashish Agarwal

(203) 626-2315 : agarwal.ashish.singhal@gmail.com || <https://CRLannister.github.io/>
<https://www.linkedin.com/in/agarwalashishsinghal/> || <https://github.com/CRLannister>

SUMMARY

Data science professional with expertise in building scalable data pipelines, deploying machine learning models in production, and managing cloud-based infrastructure. Proficient in Python, AWS, Docker, Kubernetes, and deep learning frameworks. Experienced in ETL/ELT processes, MLOps practices, and leveraging LLMs for data-driven solutions. Strong background in statistical analysis, A/B testing, causal inference, time series analysis, machine learning, and cloud-native technologies with a passion for optimizing workflows and enhancing operational efficiency.

EDUCATION

Master of Science, Data Science	(University of New Haven)	GPA: 3.94	[West Haven, CT :- Aug 2023 – Dec 2024]
Bachelor's in computer engineering	(Institute of Engineering)	GPA: 3.20	[Pulchowk, Lalitpur, Nepal :- Nov 2016 – Sep 2021]

RELATED EXPERIENCE

Data Engineer Intern	(North East Scientific)	[Waterbury, CT :- June 2024 – Aug 2024]
<ul style="list-style-type: none">Led the development of ETL pipelines using MasterControl and Netsuite APIs, automating data extraction, transformation, and storage in MySQL, reducing external dependencies by 30%.Developed dashboards and reports for Inventory, Sales and Production insights using QLIK Sense Cloud.Designed a Retrieval-Augmented Generation (RAG) system for document processing, improving knowledge retrieval efficiency and streamlining employee training.		
Graduate Research Assistant	(Secure and Assured Intelligent Learning Lab)	[West Haven, CT :- March 2024 – Dec 2024]
<ul style="list-style-type: none">Engineered multi-tenant GPU environments for brain-computer interface (BCI) research using Docker and Kubernetes, improving computational resource allocation by 40%.Applied advanced statistical techniques (wavelet transforms, ICA, SVM) for EEG signal processing and classification, enhancing model accuracy by 12%.Automated workflows for data management, web development, and deployment with Datalore and WordPress.Guided students in developing and refining their capstone projects, focusing on innovative applications of machine learning and data science.		
Data Engineer	(University of New Haven)	[West Haven, CT :- Oct 2023 – Dec 2024]
<ul style="list-style-type: none">Developed and optimized ETL/ELT pipelines on AWS (Lambda, EC2, S3, Redshift), improving data integration performance by 66%.Implemented OCR for feature extraction from image data and fine-tuned deep learning models using Hugging Face transformers, reducing model inference time by 15%.Collaborated with stakeholders to develop Power BI dashboards, delivering actionable insights to senior management		
Associate Software Engineer	(LIS Yomari)	[Lalitpur, Nepal :- April 2021 – May 2022]
<ul style="list-style-type: none">Migrated on-premises data warehouses to AWS, implementing secure data transfer and scalable ETL pipelines using Apache Spark, Kafka, and AWS services (Batch, Lambda, Kinesis).Enhanced data models with star/snowflake schemas, boosting query performance by 30%.Developed MicroStrategy dashboards and automated reporting.		
Data Science Intern	(Tootle)	[Lalitpur, Nepal :- Jan 2021 – March 2021]
<ul style="list-style-type: none">Developed customer segmentation models using LRFM algorithms, identifying high-value user segments and increasing customer retention by 4%.Built real-time dashboards for data visualization using MongoDB, Django, and Python, improving decision-making efficiency.		

RELEVANT PROJECTS

Multi-Tenant GPU Cluster	West Haven, CT
Secure and Assured Intelligent Learning Lab	Sep 2024 – Dec 2024
<ul style="list-style-type: none">Built a Kubernetes-based GPU cluster with JupyterHub integration, enabling multi-user access and efficient GPU resource sharing.Configured multi-tenant resource profiles using Kubernetes and Helm, supporting customized resource allocations, which improved utilization by 40%.Developed a secure access framework through Kubernetes Dashboard and JupyterHub authentication, ensuring isolated and reliable user access.	

- Authored detailed documentation covering setup, deployment, troubleshooting, and maintenance steps, streamlining cluster management for research and high-compute workloads.

RAG-based Knowledge Management System for Training and Information Retrieval

Waterbury, CT

North East Scientific

June 2024 – Aug 2024

- Developed a Retrieval-Augmented Generation (RAG) system by scraping and cleaning data from the company's website and specification documents, converting them into markdown format.
- Built a Chroma database incorporating LLM models such as Llama3, Phi3, and Qwen2 with quantization techniques to optimize GPU usage.
- Employed advanced prompt engineering to ensure accurate, non-hallucinated responses by reranking documents and interfacing with the database.
- Designed a user-friendly web interface, similar to ChatGPT or OpenWebUI, enabling user context tracking and delivering reliable, query-specific responses for training and operational purposes.

Retail Sales Inventory and Traffic ETL and Reporting

Lalitpur, Nepal

LIS Yomari [Client- Ralph Lauren]

July 2021 – March 2022

- Architected a robust ETL/ELT pipeline on AWS, ingesting data from TrueVUE APIs, S3 buckets, GCP, and FTP servers, leveraging EC2, EMR, Kinesis, Glue, Lambda, S3, Airflow, Redshift, CloudWatch, CloudFormation, and IAM.
- Orchestrated workflows using AWS (EC2, EMR, Lambda, Airflow) and implemented automated monitoring with CloudWatch, ensuring 99.9% uptime.
- Developed metadata objects, metrics, and attributes based on specifications in MicroStrategy, conducted data validation and performance testing, designed dashboards and dossiers tailored to client requirements.

TECHNICAL SKILLS

Languages & Tools: Python, SQL, NoSQL, R, Pytorch, TensorFlow, Scikit-Learn, Spark, Hugging Face, Docker, Kubernetes, Git, Bash, Linux, Dask, Excel, Polars

Cloud & DevOps: AWS (Lambda, EMR, S3, IAM, EC2, Redshift, Kinesis, Glue, Sagemaker, ECR, SNS), Azure, Kafka, Airflow, Jenkins, GitHub Actions, Terraform

Data Engineering & MLOps: ETL/ELT Pipelines, CI/CD, Spark, Data Lakes, CloudFormation, Helm, MLflow, OCR

Visualization Tools: Power BI, QLIK, Tableau, MicroStrategy

Machine Learning: Supervised Learning, Unsupervised Learning, Reinforcement Learning, Deep Learning, Hyperparameter Tuning

EXTRACURRICULAR ACTIVITIES

Helping Hands Nepal, HENN

Kathmandu, Nepal

Volunteer and Technical Supervisor

May 2017 – July 2019

- Led technical initiatives to implement Linux-based educational systems in remote schools, providing resources for underprivileged students.
- Developed locally hosted educational materials in Nepali, improving access to learning resources in remote regions.