

Shivam Agarwal

+1-206-739-4093 | agarwal.shivam07@gmail.com | [LinkedIn](#) | [Github](#)

TECHNICAL SKILLS

Languages: Python, SQL, PySpark, C++

Libraries: Pandas, NumPy, Matplotlib, PyTorch, Tensorflow, SciKit-Learn, SciPy, Seaborn, Keras

Technologies: Machine Learning, NLP, A/B Testing, Statistical Analysis, Hypothesis Testing, Generative AI, Model Tuning, Feature Engineering, Data Engineering, ETL/ELT, Cloud Computing (AWS, GCP & Azure.)

Developer Tools: AWS SageMaker, Huggingface, Databricks, Kafka, Git, dbt, Snowflake, SAS, Linux (Ubuntu), PowerBI, Tableau

Certifications: dbt Fundamentals issued by **dbt Labs** (2022), **AWS Data Analytics Specialty Certification** issued by **AWS** (2020),

Academy Accreditation - Generative AI Fundamentals issued by Databricks(2023).

EXPERIENCE

University of Washington - eScience Institute

Graduate Student Research Assistant

Seattle, WA

Apr 2024 – Present

- Preprocessed **100+ GB geospatial data** using **Xarray**, engineering critical features to improve model accuracy by **1.4%**.
- Designed and deployed **LSTM models** on **AWS**, **Google Colab** and **Azure** to predict snow flow in the **Skagit River basin**.

Amazon Inc

Data Engineering Intern

Seattle, WA

June 2024 – Sept 2024

- Managed data ingestion and transformation pipelines for on-boarding recruitment data consisting of **680+ million** records to the Datalake using **AWS services** such as **Lake Formation, Glue, and Athena**.
- Developed and implemented **AWS Quicksight** visualizations to predict application trends and monitor candidates, benefiting **1000+ users** and enabling effective allocation of **\$200,000 annually**.
- Optimized **SQL queries** to enhance data processing efficiency, improving query performance and reducing execution time by **21%**.
- Designed and built data pipeline to onboard data from **10+ Smartsheets** onto the Data Lake using **AWS Lambda, Step Functions, and CloudFormation**, enabling **Predictive Analytics** of project ROI.
- Built data pipelines for **ML use cases** and reporting solutions for sister teams improving **model optimization by 6%**.

Hashedin by Deloitte

Software Engineer - III (Data Science and Engineering)

Bengaluru, India

Sept 2020 – Sept 2023

- Executed **10+ GenAI projects** delivering capabilities in **code generation, content moderation, and Text-to-SQL** through models such as **GPT, Claude and DALL-e**.
- Architected an LLM-powered code correction application, integrating **SonarQube** to identify and resolve Python code issues, achieving **93%** correction accuracy using **OpenAI GPT-3.5 Turbo**.
- Developed a **switch framework** for need-based **validation checks**, reducing processing times by **73% (7 hours)** across workflows.
- Constructed a custom **Change Data Capture (CDC)** system using **PostgreSQL on AWS RDS**, tracking daily changes for **70+ files** and implementing an **error notification mechanism via AWS SES**.
- Enhanced system reliability by **15%** through **AWS** integration and **Airflow** orchestration, designing **SFTP/FTPS** modules that saved **two months** of manual effort.

Eaglytics Co.

Consultant (Data Science and Engineering)(Part-time)

Remote

Sept 2021 – Sept 2023

- Integrated **HubSpot, Google Ads, and Airtable APIs** with GCP-based data lakes, managing **15TB+** in cumulative data volume.
- Designed scalable data pipelines on GCP and optimize SQL queries saving over **\$18,000** in annual run costs.
- Crafted data schemas for **BigQuery**, optimized handling of nested **JSON API** responses and improved data retrieval speed by **35%**.
- Optimized free shipping threshold for an e-Commerce client using **A/B testing**, boosting Average Order Value by **12.6%**

OPEN SOURCE CONTRIBUTIONS

TorchFix by Meta: Analyzed **over 100 GitHub** repositories, categorized by stars, forks, and size, to generate insights on common errors, resolution patterns, and error resolution rates.

Airbyte: Migrated source connectors from **Python CDK to Low-code CDK**, including **OpenWeather API** and **Todoist API**.

ACHIEVEMENTS

Served as a **Subject Matter Expert (SME)** for **AWS** and curated questions for **AWS Associate Data Engineering** certification.

Secured **Runner-up** position in **MSDS Hackathon-2024** at the University of Washington.

Achieved **Runner-up** position in **Airbyte's Written Tutorial Hackathon** and the **Connectors Development Hackathon**.

Awarded **"Top Impactor"** and **"Tech Champion"** for making pivotal contributions to technical vertical of Deloitte.

EDUCATION

University of Washington

Master of Science in Data Science; **GPA: 3.9/4.0**

Seattle, WA, USA

2023 – 2025

Nitte Meenakshi Institute of Technology

Bachelor of Engineering in Computer Science; **GPA: 9.45/10.0**

Bengaluru, India

2016 – 2020