Shivam Agarwal

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

TECHNICAL SKILLS

Languages: Python, SQL, PySpark, C++
Libraries: Pandas, NumPy, Matplotlib, PyTorch, Tensorflow, SciKit-Learn, SciPy, Seaborn, Keras
Technologies: Data Engineering, ETL/ELT, Cloud Computing (**AWS, GCP & Azure.**), Machine Learning, NLP, A/B Testing, Statistical Analysis, Hypothesis Testing, Generative AI, Model Tuning, Feature Engineering
Developer Tools: Kafka, Git, dbt, Snowflake, SAS, Linux (Ubuntu), PowerBI, Tableau, AWS SageMaker, Huggingface, Databricks
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

Amazon Inc   Seattle, WA
Data Engineering Intern   *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**. and internal tools **Andes** and **Ganges**.
- 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%**.
- Built data pipeline from **Smartsheets** to Data Lake with **AWS Lambda**, **Step Functions**, and **CloudFormation** to support **ROI Predictive Analytics**.
- Secured dashboard access through **IAM roles**, **Lake Formation permissions**, and **KMS** while integrating with in-house UI.
- Built data pipelines for **ML use cases** and reporting solutions for sister teams improving model optimization by **6%**.

Hashedin by Deloitte   Bengaluru, India
Software Engineer - III (Data Science and Engineering)   *Sept 2020 - Sept 2023*

- Architected high-efficiency data pipelines, optimizing data ingestion, transformation, and validation processes to improve pipeline efficiency by **30%**.
- Developed a **switch framework** for need-based **validation checks**, reducing processing times by **73% (7 hours)** across workflows.
- Built a **real-time**, event-driven pipeline using **Spring Cloud Stream** and **Kafka** to process **Guidewire** events, automating tasks and reducing workflow latency by **45%**.
- 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**.
- Improved system reliability by **15%** via **AWS** and **Airflow** integration and built a **SFTP/FTPS** ingestion modules with **CI/CD** for automated deployment, saving **2 months** of manual effort.
- Architected an LLM-powered code correction application, integrating **SonarQube** to identify and resolve Python code issues, achieving **93%** correction accuracy using **GPT-3.5 Turbo**.

Eaglytics Co.   Remote
Consultant (Data Science and Engineering)(Part-time)   *Sept 2021 - Sept 2023*

- Integrated **HubSpot, Google Ads,** and **Airtable APIs** with GCP-based data lakes, managing **15TB+** in cumulative data volume and optimized 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%.**
- Improved data readability by **60%** using **Google Data Studio** across multiple sources by designing high-impact data visuals.

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   Seattle, WA, USA
*Master of Science in Data Science; **GPA: 3.9/4.0***   *2023 – March'25*

Nitte Meenakshi Institute of Technology   Bengaluru, India
*Bachelor of Engineering in Computer Science; **GPA: 9.45/10.0***   *2016 – 2020*