# Education

Subhranil Das

812-671-5394 | [dassubhranil1998@gmail.com](mailto:%20dassubhranil1998@gmail.com)| [linkedin.com/in/subhranil-das](https://www.linkedin.com/in/subhranil-das)

**Indiana University, Bloomington August 2022 – May 2024**

*Master of Science in Data Science*

**Relevant Coursework**: *Data Mining, Statistics, Deep Learning, Machine Learning, Database Concepts, Data Visualization, Applied Algorithms, Artificial Intelligence, Big Data Applications and Database Design.*

**Future Institute of Engineering and Management, Kolkata August 2017 – June 2021**

*Bachelor of Technology in Applied Electronics and Instrumentation Engineering*

# Technical Skills

**Programming/Scripting:** Python, SQL, R, Java, C, C++

**Databases:** MySQL, PostgreSQL, MongoDB, Cassandra, Snowflake, NoSQL, PL/SQL, AWS Redshift, Spark/SQL

**Frameworks:** Spark, PyTorch, TensorFlow, Keras, Django, Flask (RestAPI)

**Data Management and Analytics tools:** dBT, Apache Airflow, Power BI, Tableau, Excel, Google Analytics

**CI/CD and DevOps:** Kubernetes, Docker, Linux, JIRA, Git, GitLab, Jenkins, Bitbucket

**Cloud Platforms:** Microsoft Azure (Blob Storage, Functions, Data Factory, Databricks, Event Hubs, Synapse Analytics, Data Lake), AWS (S3, Glue, Kinesis, EMR, Athena, Lambda), Google Cloud Platform (Bucket, BigQuery)

# Work Experience

**O’Neill School of Public and Environmental Affairs, Indiana University** *March 2024 – Present*

***Data Engineer*** *Bloomington, IN*

* Collaborated with cross-functional teams to gather and preprocess datasets using **Azure Data Factory** and **Apache Spark**, **increasing data accuracy by 20%.**
* Designed and **implemented data pipelines** with **Azure Databricks** and **Airflow**, optimizing **data processing** time by 15%.
* Developed **scalable solutions** for data storage and retrieval with **Azure Data Lake** and **Azure SQL Database**, improving data infrastructure efficiency.

**eGain Corporation** *September 2021 – July 2022*

***Solution Success Engineer*** *Pune, India*

* Investigated and resolved **100+ data pipeline issues** using **Azure Log Analytics** and **Azure Functions**.
* Collaborated with the engineering team to identify and **fix bugs in ETL processes**.
* Analyzed data pipeline failures with **Apache Airflow**, implementing enhancements that reduced DAG s’ errors by 30%.
* Led cross-functional teams during data outages, utilizing **Azure Data Lake** and **Databricks**, reducing downtime by 30%.

**eGain Corporation** *June 2021 – August 2021*

***Data Engineer Intern*** *Pune, India*

* Implemented debugging strategies for **Azure Data Factory** pipelines using **Azure Monitor**, decreasing bug resolution time.
* Collaborated with cross-functional teams on **20+ projects**, gaining insights into the product cycle and support processes.
* Supported data infrastructure on **Azure** and **J2EE** platforms, managing servers and applications effectively.

# Projects

**Generative AI for Pathology Datasets** *| Python, Tensorflow, Scikit-learn, Keras, OpenCV, GANs* [[ Link ]](https://github.com/Atharva-Shah-2298/DLS)

* Directed a project focused on **Generative AI** for **Pathology Datasets**, specializing in applying Generative Adversarial Networks (GANs) for **nuclear detection** in **medical imaging**.
* Overcame **HIPAA-related data** access restrictions to **create synthetic datasets** for nuclei detection model training.
* Utilized advanced **GAN** architectures such as DCGAN, Variational Autoencoders, and StyleGAN3 to enhance model performance by **40%.**
* Trained a **YOLOv8** nuclei detection architecture, achieving an **accuracy of 85%** in **nuclei detection models**.

**Turbocharge Retail Insights** *| Python, Apache Airflow,dBT, Soda, Docker, GCP, Big Query,* *Metabase* [[ Link ]](https://github.com/dassubhranil/Turbocharging-Retail-Insights)

* Led the development of a comprehensive Apache Airflow **ETL pipeline** integrating GCP Bucket, BigQuery, Soda, and dbt, achieving 25% enhancement in process efficiency.
* Optimized over **20 SQL scripts** and **dbt** models, and implemented automated testing procedures (**DAGs**) for **financial data ingestion**, quality checks, and transformations, significantly improving data quality.
* Developed a **real-time dashboard** in Metabase for data visualization to enhance **reporting and analysis** capabilities, providing actionable insights to stakeholders.