

Judhajit Roy

+1(872)-218-5124 | jroy13n96@gmail.com | github.com/Judhajit-Roy | linkedin.com/in/judhajit-roy

4.5+ years' experience in designing and implementing data-driven solutions, building scalable Big Data pipelines, and deploying advanced predictive models to drive business insights and innovation.

EDUCATION

University of Illinois, MS in Computer Science | Chicago, USA GPA: 4.0 / 4.0 May 2023
University of Mumbai, B.Tech in Electronics and Telecommunications Engineering | Mumbai, India June 2018

Courses: Machine Learning, Visualization, Data Structures, Algorithms, Databases, Big Data, Deep Learning, Cloud Computing, Natural Language Processing, Causal Inference and Deep Generative Models

EXPERIENCE

Egen, Data Engineer | Chicago, USA Sep 2024 - Present

- Designed a solution to implement modern data pipelines on Azure Cloud, migrating legacy processes to a scalable cloud-based architecture
- Implemented and deployed ELT data pipelines on Azure Cloud while reducing pipeline run times and achieving enhanced performance
- Built dimensional data models in Snowflake using dbt to enable efficient sales reporting and analytics across large datasets

Johnson Controls, Data Engineer | Chicago, USA July 2023 - Aug 2024

- Designed and developed scalable ETL data pipelines using Python, Spark, SQL on Google Cloud Platform (GCP), integrating data from SQL and NoSQL databases
- Created dimensional data models (star and snowflake schemas) in a modern data warehouse environment, enabling efficient querying and reporting across large datasets
- Deployed pipelines using Apache Airflow DAGs to automate, schedule, and monitor data workflows, ensuring reliable pipeline execution
- Devised BigQuery Data warehouse data backup pipeline by upgrading code in cloud functions, solving backup issues, and reducing storage needs by 10x through data compression

Encyclopedia Britannica, Machine Learning Intern | Chicago, USA May 2022 - Aug 2022

- Experimented several deep learning and generative AI models and developed techniques for named entity recognition and key-pharse extraction from over 130k articles
- Deployed an AWS pipeline to automate transforming XML data into text and extract important key phrases by ranking method using custom-built AI model

Reliance Jio Platforms Ltd , Data Engineer | Mumbai, India Aug 2018 - March 2021

- Developed a SparkML-based sentiment analysis pipeline with 86% accuracy in classifying customer reviews
- Built ETL pipelines on Spark to migrate and aggregate Big Data from legacy data warehouse and cut down the run time by 60%
- Helped in onboarding and worked closely with Business team to understand requirements and translate them to technical solutions

SKILLS

Languages and Tools SQL, Python, Java, Hadoop, PostgreSQL, MongoDB, GCS, Pub/Sub, Lambda, Docker, Agile/Scrum
Frameworks and Cloud AWS, GCP, Azure, Spark, Hive, Kafka, Presto, Airflow, React, Flask, Databricks, Snowflake, BigQuery

PROJECTS

Old Photo Restoration using De-noising Probabilistic Diffusion Generative Model

- Developed and fine-tuned a state-of-the-art diffusion model to restore and enhance degraded images, producing high-quality outputs
- Trained the model using augmented images with blur and film grain noise to achieve exceptionally clear restorations

Cryptocurrency Real-Time Sentiment Analysis Web Dashboard

- Built and deployed a real-time dashboard on AWS EC2 to visualize global crypto sentiment for Twitter data using NLP and D3.js choropleth maps
- Designed a data pipeline with Python and Flask to ingest, process, and geolocate tweets for country-level sentiment aggregation

CERTIFICATES

Google Cloud Certified Professional Data Engineer 2024
Google Cloud Certified Professional Machine Learning Engineer 2025