Kartik Gupta

kartikgt23@gmail.com   kartikgt.github.io
linkedin.com/in/kartikgupta23   (413) 468-0353

EDUCATION

University of Massachusetts Amherst   *Sep 2022 – May 2024*
MS in Computer Science (specializing in Data Science)

Indian Institute of Technology Kharagpur (IIT KGP)   *Jul 2013 – Apr 2018*
BTech. + MTech. in Electronics Engineering (specializing in Microelectronics and VLSI design)

EXPERIENCE

Data Science Intern | *Mathematica*   *Jun 2023 – Aug 2023*

- Developed an application "Community Connector" for visualizing health outcomes of all the US counties based on certain social determinants of health scores. Performed end-to-end data collection, transformation, and modeling (RandomForest and Lasso Regression) to estimate similar counties for effective knowledge transfer opportunities.
- Designed and deployed an automated cloud ETL framework for creating a medical insurance price database in Redshift for housing a 250 TB+ data model, employing serverless architecture and orchestration in AWS stack.
- **Tech:** Python, Pandas, Sklearn, Redshift, AWS Lambdas, AWS Step Functions, Botocore, Git, SQL, Selenium, APIs

Graduate Researcher | *Adobe*   *Jan 2023 – Jun 2023*

- Researched efficient machine learning pipelines for computer vision tasks. Utilizing variational autoencoders and hyperpriors for effective latent representations, achieved a 37% and 23% reduction in latency and memory resp.
- **Tech:** Pytorch, Lightning, HuggingFace, CompressAI, Torchvision, PIL, Matplotlib, CUDA, GPU, Git

Software Engineer (Data) | *Uber*   *May 2021 – Aug 2022*

- Managed all 150PB+ Uber Eats data resources as a member of the Data Solutions team, specifically event streaming and batch analytics at scale, and liaisoned with the new acquisitions of Cornershop and Drizzly for data integration.
- Modeled a rating table for all orders including support for new and complex feedback options, and created data pipelines serving daily and monthly reports to all merchants, helping in reducing churn by 13% over a quarter.
- Developed a query monitoring tool that identified resource-intensive Spark and Presto queries, diagnosed the un-optimized code (via parsing the AST tree from the execution plan and pattern matching), and suggested improvements following Spark and SQL best practices. The solution saved $3.5M and was deployed organization-wide.
- **Tech:** Python, Java, Scala, SQL, Presto, Spark, Hadoop, Hive, Kafka, Flink, Git, Pandas, Athena, Cassandra, Grafana

Technical Staff (Big Data) | *Envestnet Yodlee*   *Oct 2020 – Apr 2021*

- Architected a new Master Data Management platform optimizing for query speed and multi-source integration by migrating from RDBMS solution to a NoSQL Neo4j graph database hosted on AWS and modeled the data via Cypher.
- **Tech:** Java, Neo4j, Cypher Query Language, Python, Boto, AWS - EMR, EC2, Data Pipelines, Lambdas, S3

Software Engineer | *SAP*   *Jul 2018 – Sep 2020*

- Developed a Java and selenium based automation test suite for periodic application monitoring and logging errors.
- Managed Splunk and ELK (ElasticSearch, Logstash, Kibana) deployments and monitoring dashboards.
- Developed an organization-wide data discovery tool for end-to-end analytics. Designed a Hadoop datalake using Spark and Hive, and built a Java web interface for ad-hoc querying and generating analytics and cost reports.
- **Tech:** Python, SQL, Java, Selenium, Hadoop, Hive, Splunk, ElasticSearch, Grafana, Zookeeper, Shell, Azure

PROJECTS

- Evaluating in-context learning of finetuned LLM agents *(Pytorch, Langchain, HuggingFace, LLama, GPT, GPTQ)*
- Implementing vector attention for long-range 3D shape generation from point clouds *(Pytorch, Git, CUDA)*
- Anomaly detection in a real-time event stream using Apache Flink *(Java, Flink)*
- Tackling data scarcity in medical domains with Transfer Learning and Few-Shot approaches *(Python, Pytorch, CUDA)*

CERTIFICATIONS

Microsoft Certified: Azure Data Scientist Associate   Credential ID: H446-0997
Neo4j Certified Professional   Credential ID: 17127043
Coursera : Operationalizing LLMs on Azure   In progress