

Naman Kanwar

nmnknwr@gmail.com
linkedin.com/in/namankanwar/

+1-470-263-0059
naman26.github.io

Professional Summary

5+ years of experience in software engineering, data engineering, and machine learning. Extensive experience in working with multi-disciplinary teams, establishing and supporting ETL pipelines for Walmart and LexisNexis Risk. Seeking opportunities to leverage data skills in a collaborative environment.

Education

- **Georgia State University** Atlanta, GA
 - Master of Science in Computer Science Aug 2019 – Jul 2020
 - Bachelor of Science in Computer Science Jun 2016 – Jul 2019

Skills

- **Languages:** Python (Proficient), Java, SQL
- **Database and Frameworks:** Google Cloud Platform (Big Query, Dataproc, GCS, Dataflow, VMs), Azure (Databricks, Blob storage), Airflow, Pyspark, AWS Lambda, Amazon (QuickSight, SageMaker, S3), MariaDB
- **Concepts:** Object-Oriented Programming, Agile/Scrum, Multithreading, ETL, Data Analysis, Data Mining, Big Data
- **Other:** Spring Boot, Linux, Git, Jenkins, Docker, JFrog, Flask, Selenium

Experience

- **Walmart** Atlanta, GA
Walmart Global Tech - Data Engineer III August 2022 – Present
 - Orchestrated, developed, and maintained 20+ batch/streaming ETL pipelines (Airflow).
 - Integrated service now requests tickets with existing CI/CD process saving over 10 manual hours per deployment.
 - Followed industry standard and improved code coverage of major repositories to 95% (SonarQube)
 - Onboarded a new project creating batch data pipelines with SCD tables (Airflow, Dataproc, Google Dataflow, Pyspark, BigQuery, and GCS).
 - Migrated legacy Automic pipelines to Airflow with 100% code coverage (Pyspark, GCP, Databricks, and Azure).
 - Automated CI/CD flow to drain and create new dataflow jobs saving manual effort of 1-2 sprints per pom.xml update (Jenkins, Docker and Maven).
 - Implemented and maintained proper security standards for all data pipelines end to end ensuring 0% data loss (Google Secret Manager).
 - Optimized cloud cost by fine-tuning clusters and worker nodes saving over \$100K annually (Azure and GCP).
 - Contributed to internal artifactory with 70+ downloads across the division (JFrog).
 - Improved a REST API dag's performance by reducing 90% of cloud compute time (Dataproc, Pyspark).
 - Mitigated on-call failures in pipelines and supported expected outages by providing clients with data within the 1st hour of failure.
- **LexisNexis Risk Solutions** Alpharetta, GA
LNRS - Software Engineer II February 2021 – August 2022
 - Developed search functionality and an interface for clients to find matches within 2s from entire database.
 - Using ETL process, implemented phonetic and fuzzy matching to reduce search time by 5ms.
 - Analyzed data usage and optimized databases to reduce storage cost for cloud migration.
 - Created and evaluated crosswalk interfaces based on precision and recall.
 - Worked with different environments, coordinated merges and code releases using git, CI/CD (Git lab) and SQL Server.
 - Created and automated a testing system by utilizing CRON jobs saving 120 work hours.
 - Met client expectations by generating B2C and C2B reports of data sets containing 30,000,000+ records, in a timely manner.

- **Georgia State EBCS** Atlanta, GA
Evidence Based Cyber Security Group - Data Engineer Aug 2019 – Aug 2020
 - Developed web scrapers to collect data from suspicious websites. Engineered CAPTCHA breaking system with 90% accuracy rate to automate the scrapers.
 - Created automatic weekly business report, empowering institute-wide policy decisions using Python, Selenium, SQL and folium.
 - Coordinated with the associate director to maintain and develop EBCS website (ebcs.gsu.edu). Onboarded and handed project to incoming developer.
- **Fiserv** Atlanta, GA
Global Finance Solution - Software Development Engineer Intern Jun 2019 – Aug 2019
 - Worked on 35,000+ lines of C++ code with no existing documentation. Wrote documentation for existing and newly added code.
 - Automated database server shutdown and creation, saving 12% annual costs.
- **Georgia State CS Department** Atlanta, GA
Computer Science Department - Data Engineer Oct 2017 – Aug 2019
 - Implemented an algorithm that visualizes evolutionary trees using Python. Reduced the execution time of current algorithm by 99%.
 - Designed an API along with a full stack website architecture to deploy the algorithm.
 - Integrated interactive visualization of data using D3.js. Completed and deployed the website within 4 months which is being used by hundreds of biologists.
- **Verizon Connect** Atlanta, GA
Fleetmatics - Data Analyst Jun 2017 – Aug 2017
 - Directly increased userbase by 20% utilizing data mining techniques for analysis.
 - Built analytics dashboards from Hive tables using Java to depict user driving behavior.
 - Utilized Apache Spark to develop data analytics for business intelligence and product development.

Certifications

- **AWS Fundamentals Specialization** Credentials: <https://coursera.org/share/fd170135c42c76a2015a0d961a6eddbdf>
Coursera Oct 2020
- **Java Developer** Credentials: hackerrank.com/certificates/4f83321244ab
HackerRank Aug 2020
- **Python Developer** Credentials: hackerrank.com/certificates/540a2bdaaf11
HackerRank Aug 2020

Other

- **Awards**
4+ National Hackathon Awards
 - **HackGT**: 2019: Optical Character Challenge Runner-Up
 - **HackGSU**: 2019: 1st in Crypto challenge, 2017: Anthem Grand Prize
 - **HackATL**: 2018: Health and Wellness Challenge Runner-Up
 - **Georgia State University**: 2018: Outstanding Undergraduate Research Award