Sitanshu Kushwaha

✉ sak9813@nyu.edu   in linkedin.com/in/sitanshukushwaha   🔗 sitanshu.tech   ⬡ github.com/Sitanshuk

🎓 EDUCATION

New York University, MS in Computer Science   Sep 2023 – May 2025 | New York
Relevant Courses: **Design and Analysis of Algorithms**, **Cloud Computing, Big Data, Data Management and Strategy**

University of Mumbai, BE in Computer Engineering   Aug 2016 – Nov 2020 | Mumbai
Relevant Courses: **Software Engineering, Object Oriented Programming in Java & C++, Operating Systems, Distributed Systems**

💼 PROFESSIONAL EXPERIENCE

Data Engineering Intern, NBCUniversal   Jun 2024 – present | New York
- Implemented a **year-round aggregation strategy** in BigQuery for Peacock's annual user insights, processing **petabyte-scale data** incrementally to **reduce end-of-year query times** from hours to minutes for **35M+ subscribers**, optimizing resource utilization and costs.
- Spearheaded a **self-service data application** using JavaScript, Python, and SQL; **automated data updates**, eliminating **20 hours of manual intervention per cycle**, improving accuracy, scalability, and reliability.

Software Engineer Intern, NYU IT   Oct 2023 – May 2024 | New York
- Developed a **Django** web application for real-time monitoring and analysis of 100+ processes running on **AWS** and **Snowflake**, providing users with **actionable insights** that improved job performance by 30%.
- **Developed a GenAI-powered chatbot with conversational memory**, enabling developers to query Snowflake metadata using **natural language**, receive contextual follow-ups, and analyze schema modifications, leveraging **LLMs and Cortex AI** for intelligent retrieval and automated impact assessment.

Data Engineer, LTIMindtree   Jan 2021 – Jun 2023 | Mumbai
Technical Lead, Visioncare MFF Data Engineering team - Johnson and Johnson
- Spearheaded **optimization** efforts in **pyspark**, resulting in a **30% reduction** in execution time of Transformation Jobs, improving data timeliness.
- Adopted **event-based triggers** within Azure Data Factory for **ETL** pipelines, strategically **enhancing** the **efficiency** of handling Big Data from diverse sources, **reducing costs by 25%.**
- **Enhanced workflow observability** by embedding **timestamp-based tracking** across **pipeline stages**, enabling **precise latency measurement** and **bottleneck identification**. Developed a **Tableau dashboard** to monitor **system performance**, reducing **unforeseen outages by 40%.**

Machine Learning Engineer Intern, Toflo Fintech Consulting   Dec 2019 – Mar 2020 | Mumbai
- Developed a **scalable Recommender System** for a global B2B **e-commerce platform**, leveraging **Big Data technologies** to boost the **click-through rate by 33%** on recommended products.

Python Developer Intern, Innolearn Solutions pvt. ltd.   Dec 2018 – Jun 2019 | Mumbai
- Built and deployed a **full-stack** web application using **JavaScript, Python, HTML, CSS**, integrating real-time stock updates and automated personalized notifications, resulting in adoption by 1,000+ daily active users within the first three months.

🧠 SKILLS

Programming Languages — Python, Java, JavaScript, SQL | **Software Development** — Design patterns, Multi-threading, Concurrency, Data Structures & Algorithms | **Web Technologies** — Django, Flask, React, HTML, CSS, REST APIs |

Big Data & Cloud — PySpark, Kafka, Databricks, BigQuery, Snowflake, AWS, GCP, Data Lake, ETL, NoSQL, Airflow |

DevOps & Deployment — Docker, Kubernetes, Jenkins, CI/CD Pipelines

📁 PROJECTS

Talk2Doc - (GCP, RAG, LLM, APIs)
- *Designed and deployed a **scalable web service** on **GCP**, utilizing **real-time data processing**. Integrated Retrieval-Augmented Generation (RAG) and **multi-threaded job scheduling** to support personalized note searching and job tracking for thousands of users.*
- Architected **modular**, event-driven highly scalable system utilizing **serverless** functions, queues and LLM, capable of handling millions of users.