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
University of Mumbai, BE in Computer Engineering	Aug 2016 – Nov 2020 │ Mumbai

💼 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 distribute computational load, reducing end-of-year query times from **hours to minutes** for **35M**+ subscribers, and optimizing resource utilization and costs.
- Spearheaded the creation of a self-service data application utilizing **JavaScript**, Python, and SQL; **automated** data updates that **eliminated 20 hours** of **manual** intervention per cycle, **reducing** communication **overhead** and **enhancing accuracy**, **scalability**, and **reliability**.

Software Engineer, 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%**.

Data Engineer, LTIMindtree   Jan 2021 – Jun 2023 │ Mumbai
Technical Lead, Visioncare MFF Data Engineering team - Johnson and Johnson
- Spearheaded **optimization** efforts in Databricks Spark code, resulting in a **30% reduction** in **execution time** for 50% 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%.**
- Designed **Monitoring Tableau Dashboard** for the data flow architecture, improving visibility for early identification and resolution of **bottlenecks** and other issues, and **mitigated unforeseen outages by 40%.**
- Optimized deployment processes by implementing **CI/CD** using **Git** and **ARM** templates (**IaC**) in Azure, resulting in a **70% reduction** in **deployment time**.

Intern Machine learning Engineer, Toflo Fintech Consulting   Dec 2019 – Mar 2020 │ Mumbai
- **Developed a scalable Recommender System** for a global **import-export e-commerce platform**, leveraging **Big Data technologies** to boost the **click-through rate by 33%** on recommended products.
- Designed the system using a **microservices architecture** to efficiently process large-scale data within a **distributed environment**, ensuring seamless integration with other core platform services.

Python Developer Intern, Innolearn Solutions pvt. ltd.   Dec 2018 – Jun 2019 │ Mumbai
- **Deployed an interactive, scalable web application** using **Django**, featuring **real-time stock updates**, **personalized email alerts**, and **calendar integrations**, resulting in adoption by **1,000+ daily active users** within the first three months.

🧠 SKILLS

Big Data & Cloud — PySpark, Kafka, Databricks, BigQuery, Snowflake, AWS, GCP, Data Lake, ETL, NoSQL, Airflow,
Programming Languages — Python, Java, JavaScript, SQL, **Web Technologies** — Django, Flask, React, HTML, CSS, **Others** — Git, Docker, Kubernetes, Jenkins, RESTful APIs, Microservices, MVC Architecture, CI/CD

📁 PROJECTS

Talk2Doc - (GCP, RAG, LLM, APIs) 🔗
- Architected a centralized ecosystem for students using Google Cloud Platform (**GCP**), integrating Retrieval-Augmented Generation (**RAG**) for personalized note searching and automated job application tracking.
- Architected **modular**, event-driven highly scalable system utilizing **serverless** functions, queues and LLM, capable of handling millions of users.