Sitanshu Kushwaha

sak9813@nyu.edu   929-643-4480   sitanshu.tech   New York, NY

linkedin.com/in/sitanshukushwaha/   github.com/Sitanshuk

EDUCATION

New York University, MS in Computer Science — Sep 2023 – May 2025 | New York
Big Data, Cloud Computing, Data Science, Machine Learning, Data Management & Strategy, Computer Vision

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

WORK 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.
- Reengineered legacy processes by centralizing data in **Databricks' Unity Catalog** and introducing a **self-service tool**, enhancing **data governance** and improving **stakeholder transparency**. This streamlined **forecast extrapolation** across demographics, reducing communication overhead.
- **Optimized** data workflows, accelerating **report delivery time** by **2 days** and eliminating **20 hours** of **manual intervention** per quarter. This improved **efficiency** and **accuracy** of forecasting processes.

Data Engineer, Enterprise Data Management - NYU IT — Oct 2023 – present | New York

- **Orchestrated an ETL pipeline using Airflow and dbt to collect, transform, and store metadata in Snowflake**, enabling a **RAG system** for **LLM-powered impact analysis**, allowing developers to assess schema changes, track dependencies, and retrieve insights on jobs, tables, and stored procedures.
- **Developed a GenAI-powered chatbot in Streamlit 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

- Optimized **Databricks Spark** code, achieving a **30% reduction in execution time** for 50% of transformation jobs, enhancing data timeliness and **scalability** for **multi-TB datasets**.
- Implemented **event-based triggers** in **Azure Data Factory** for ETL pipelines, enhancing efficiency in handling **Big Data** from diverse sources and reducing **cloud costs by 25%**.
- Designed a **Tableau Dashboard** for **monitoring real-time** data flow architecture, enabling early identification of bottlenecks and reducing system outages by **40%**.
- Implemented a **CI/CD** pipeline in **Azure DevOps**, automating build validation, testing, and deployments across environments, reducing manual effort by 70% and ensuring code reliability.

SKILLS

Big Data — PySpark, Kafka, Databricks, BigQuery, Snowflake, AWS, GCP, Data Lake, ETL, NoSQL, Airflow, **Machine Learning** — Scikit Learn, Tensorflow, NLP, Neural Networks, Deep Learning, **Data Analytics** — SQL, Pandas, Numpy, Matplotlib, Web Scraping, Tableau, **Languages** — Python, JAVA, R., **Tools** — Git, Docker

PROJECTS

DineSync - Real-Time Culinary Exploration in NYC, (Big Data, Spark, Kafka, MongoDB, Django)

- Engineered DineSync, a **real-time restaurant recommendation** system leveraging **Kafka** for **processing live** user check-ins, ensuring accurate seat availability data with 95% accuracy.
- Engineered a solution that automatically recommended alternative restaurants when primary choices were fully booked, resulting in a **20% decrease in drop-off rates** during peak reservation times.

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, LLM capable of handling millions of users.

AWARDS

1st place - Innovation Business Case Project, NBCUniversal — Aug 2024

2nd place - Build with AI, Google Developer Group NYC — Apr 2024