

ABDUL RAFAY AHMED KHAN

+1 (303) 9615167 rafay.ahmedkhan@colorado.edu portfolio

Education

University of Colorado Boulder

2024 – 2026

Masters of Science in Computer Science - **cGPA: 4**

Relevant Courses: Data Center Scale Computing, Data Mining, Database Systems, Big Data Architecture

Lahore University of Management Sciences

2020 – 2024

Bachelors of Science in Computer Science - **Dean's Honor List**

Technical Skills

- **Programming Languages:** Python, JavaScript, SQL, Cipher, C++, Latex
- **Databases:** PostGreSQL, Neo4J, Snowflake, MySQL, Google Cloud Storage
- **Development:** React.js, Flask, Django, RESTful APIs, Microservices Architecture, JWT-based authentication
- **DevOPS:** Infrastructure Automation, Deployment on Cloud VMs
- **Testing:** unittest, Pytest, Locust

Work Experience

Meta

May 2025 – August 2025

Data Engineer Intern - Menlo Park, California, United States

- Built and deployed scalable batch pipelines using **dbt**, **Apache Spark**, **Apache Airflow**, and **Apache Hive** to **process over 2B records daily**, supporting high-throughput ML feature engineering workflows.
- Collaborated with research scientists to productionize LLM features, improving model recall by **9%** and reducing query latency by **35%** using optimized **Presto** and **Spark** transformations.
- Developed interactive dashboards using **Tableau** and advanced **SQL (Presto, Hive, SparkSQL)** to expose model outputs and metrics, decreasing manual query time by **70%** for **50+ internal users**.
- Built and launched an **AI-powered onboarding assistant** for an internal platform, reducing onboarding time by **60%** across **7+ teams** and eliminating over **40** recurring meetings through automated, self-serve documentation guidance.

Volga Partners

June 2023 – August 2023

Software Engineer - Kirkland, Washington, United States (Remote)

- Engineered an AI SaaS product, **Amethyst**, utilizing **LangChain** and **OpenAI LLMs**, efficiently processing up to **10,000+ user queries daily**.
- Collaborated with a cross-functional team of developers and product managers, following **Agile methodologies** (Scrum) to deliver modular data import and Q&A features, ensuring seamless data integration.
- Solved performance bottlenecks by implementing **caching mechanisms** and optimizing database queries, leading to a **25% reduction in UI load times** on a scalable Flask backend with a responsive ReactJS frontend.
- Automated **performance monitoring and error logging**, improving system reliability and reducing downtime by **30%**.

Projects

DataBuff | Big Data Architecture

Jan 2025 – May 2025

- Built a **full-stack AI web app** with **React.js**, **Flask**, and **Mistral 14B** for automated dataset transformations and **Pandas** code generation for data cleaning, visualization, and ML model training.
- Implemented **contextual chat memory** via Firestore and JWT auth, enabling personalized AI interactions.
- Leveraged **Google Cloud Storage** for dataset and visualization management; deployed on **GKE** with optimized load balancing and security.
- Achieved **100% pytest pass rate** with extensive mocking, attaining **85%+** backend code coverage.

Raasta Trip Booking Website | Independent Project

Jan 2023 – Jul 2023

- Developed a dynamic web app for trip booking with **Django** and **PostgreSQL**, featuring efficient user interactions.
- Implemented **search/filter functionality** enabling users to book based on preferences such as location and price.
- Designed optimized database schema with **10 tables**, handling **10,000+ records** effectively.
- Built scalable **RESTful APIs** for efficient frontend-backend communication.

Relevant Community Experience

AI Community of Pakistan

Jun 2023 – Aug 2024

President Academics

- Led **30+** volunteers across multiple academic teams.
- Organized **5 seminars** with prominent AI industry leaders.
- Conducted **3 coding competitions** on HackerRank attracting **250+** nationwide participants.