Rishabh Hedaoo
|+91 7841919546 | thefast490@gmail.com

EDUCATION

IIT BHU, Varanasi, India.   **June 2024**

PROFESSIONAL EXPERIENCE

Overstand Labs   **Feb 2025- Present**
Software Engineer - Data
- Built **LLM Gateway** with **Redis caching** and **PII scrubbing,** achieving 27% cost reduction
- Designed **data pipelines for ingestion and analytics** using **Python, PostgreSQL, and Palantir Countour**
- Integrated **multi-provider LLM APIs** (Claude, OpenAI, Vertex AI) with fallback routing and token optimization
- Accelerated development using **AI-assisted tools** (Claude Code) for feature delivery.

Tredence Analytics   **August 2024- Feb2025**
Analyst - Data Engineering
- Experience in designing and implementing ETL pipelines using **Pyspark, Azure Databricks, and Apache Spark**.
- Expertise in **optimizing ETL pipeline**s for performance and scalability.
- Experience in migrating data from a data warehouse to a **medallion architecture** for analytical purposes.

Newton School   **March 2024-July2024**
Subject Matter Expert - Data Science
- Designed projects in **Python, SQL, and Power BI** to enhance students' hands-on learning.
- Orchestrated and Streamlined evaluation processes, delivering constructive feedback to boost student performance.
- Created reports and dashboards using **Google Looker Studio, Excel, and PowerPoint** to track progress
- Collaborated with teams to align curriculum with industry standards and student needs.

ACADEMIC PROJECT

E-Commerce Data Pipeline   **Mar 2025 – Aug 2025**
- Built **end-to-end ETL** pipeline using **Databricks**, **PySpark**, and **Delta Lake** following Medallion Architecture
- Processed e-commerce data from **Amazon S3** using **Pyspark** and SQL queries for analytics
- Implemented **Lakehouse architecture** with Delta Lake for data consistency and optimized query performance

Audiophile e2e Pipeline,   **Aug 2024 – Sept 2024**
- **Scraped and ingested raw data** from Crinacle's website, storing bronze datasets in **AWS S3.**
- **Parsed and validated data** using Pydantic to produce silver datasets, then loaded into **AWS S3, Redshift, and RDS** for varied downstream use.
- **Transformed and tested data** in Redshift using dbt, enabling reliable analytics and future project integration.
- **Provisioned all cloud infrastructure** using Terraform, ensuring reproducibility and best practices

CERTIFIED COURSES
- Databricks Certified Data Engineer Associate | **Databricks**
- Advanced Business Analytics Specialization | **Coursera**
- Mathematics for machine learning and data science Specialization | **Coursera**

SKILLS
Analytical Tools: Databricks, Power BI, MS Azure, Palantir Foundry

Programming Languages: SQL, Python - Numpy, Pandas, Pyspark, FastAPI

Software used: Databricks Palantir Foundry, Apache Flink, JIRA, Kanban board