

NOORUL HUDA KHAN

☎ +919990734395 ✉ noorulhuda101099@gmail.com [in linkedin.com/noorul](https://www.linkedin.com/in/noorul) github.com/THENHKHAN

Experience

Fluree — Java Developer (Full-time)

Aug 2023 – present

Java Spring Boot Developer

Bangalore, India

- Engineered custom queries, algorithms, and APIs for an internal JSON-LD database using Java and Apache Jena, optimizing data retrieval processes. Engineered custom prompts and trained ChatGPT to understand and execute FlureeDB queries, improving efficiency and user experience.
- Added functionalities and tools to convert TTL, RDF, and OWL files into organized formats, enhancing overall data management efficiency.
- Implemented a Spring Boot microservices architecture with JanusGraph, ScyllaDB, keycloak and PostgreSQL, utilizing Docker Compose for seamless orchestration and deployment of multi-container applications.
- Developed a web scraping solution with Python, BeautifulSoup, and pandas to automate extraction and structuring of large-scale product data from multiple websites into Excel, significantly improving data accessibility and analysis.
- I built a PySpark pipeline that uses SentenceTransformers to batch-encode millions of records, stores them in a FAISS index, and performs fast nearest-neighbor searches for deduplication. I optimized it with batch processing, distance thresholds and Spark SQL post-processing to handle large datasets in minutes instead of hours.

Fluree — Data Engineering Intern

Feb 2023 – Aug 2023

Internship

Bangalore, India

- Designed and implemented address parsers in Julia and C, optimizing data extraction and contributing to improved geospatial processing within the system.
- Additionally, spearheaded the construction of a comprehensive catalog and implemented improvements in the existing JSON-LD database, demonstrating a commitment to data organization and system refinement.

Technical Skills

Languages: Java, Python, Scala, Spark, PySpark, Sparql, Javascript

Backend: Spring Boot, Fast-API, Familiar with (Flask, Django), SQL, PostgreSQL, MongoDB

Technologies: Docker, Databricks, Microsoft Azure, Git, Github, Spring Boot, GraphQL, Postman

AI Models: all-MiniLM-L6-v2 (To Generate Embeddings), local LLaMA model (Q & A), mistral-7b-instruct-v0.1.Q4_K_M Hugging Face model, MCP-Server

Education

Aligarh Muslim University

June. 2021 – July 2023

Master of Computer Applications (MCA): 8.4 CGPA

AMU, India

Jamia Millia Islamia

June. 2017 – July 2020

B.sc(Hons) Physics : 8.7 CGPA

JMI, India

Projects

Production Multi-Format QA Bot API Service

[Github](#)

- Built a production-ready Retrieval-Augmented Generation (RAG) QA API supporting PDF, DOCX, XLSX, PPTX, CSV, and TXT ingestion with semantic search using ChromaDB.
- Implemented LangChain pipeline with `sentence-transformers/all-MiniLM-L6-v2` embeddings and a local Mistral-7B model for context-aware question answering.
- Optimized for Linux VM deployment with API key authentication, concurrency control, health checks, Dockerized services, and persistent storage.
- Tech:** Python, FastAPI, LangChain, Mistral-7B (GGUF), SentenceTransformers, ChromaDB, Docker

Build Real-Time REST APIs with Spring Boot -Blog App | *Java, Hibernate, PostgreSQL, Docker*

[Github](#)

- Build Rest API's for Blog App Using Spring Boot, Spring Security, JWT , Spring Data JPA (Hibernate), PostgreSQL Database.
- Implemented Pagination and Sorting, Login and Sign-up, REST API Validation, Handled exceptions globally as well as specific.

Task Reminder Application | *fastapi, oauth2, react.js, mongodb, javascript*

[Github](#)

- Developed a Task Reminder Application with FastAPI, OAuth2, and MongoDB, demonstrating proficiency in backend development.
- Implemented a responsive front end using React.js, enhancing user experience and interaction in the Task Reminder Application.