

Ishan Makkar

Kurukshetra, Haryana, India | ishanmakkar651@gmail.com | +91 9053688078
[linkedin.com/in/iishanmakkar](https://www.linkedin.com/in/iishanmakkar) | github.com/iishanmakkar | leetcode.com/u/s4oehfbewb/

PROFESSIONAL SUMMARY

Third-year B.E. Computer Science Engineering student (CGPA 9.47/10) building batch and real-time data pipelines with Apache Kafka, Spark (PySpark), and SQL. Built and deployed end-to-end ETL on AWS and Azure using medallion architecture, automated data-quality checks, and Terraform-managed infrastructure; built a live public MCP data-cleaning server exposing REST APIs. Solved 100+ LeetCode problems; solid grounding in data structures, algorithms, DBMS, and operating systems.

TECHNICAL SKILLS

Data Engineering: Apache Kafka, Apache Spark / PySpark, Hadoop, HDFS, Hive, Big Data, Apache Airflow, ETL Pipeline Design, Data Modeling, Medallion Architecture, Data Warehousing

Cloud & Data Platforms: AWS (S3, EC2, Lambda, Glue, Athena, IAM), Azure (Data Lake Storage Gen2, Databricks), Snowflake

Programming & Databases: Python (pandas), SQL, Java, C++, JavaScript, TypeScript, HTML, CSS, MySQL, MongoDB

Web Development: React, Node.js, Express, RESTful APIs, Frontend

DevOps & IaC: Docker, Terraform, CI/CD

Fundamentals: Data Structures & Algorithms, OOP, Object-Oriented Design, Design Patterns, DBMS, Operating Systems, Computer Networks, Linux, SDLC, Agile, Microservices

Tools: Git, GitHub, Power BI, MCP (Model Context Protocol), Pytest, Jest, Unit Testing

PROJECTS

RetailSync: Real-Time E-Commerce Streaming Pipeline

Kafka, PySpark, Spark SQL, AWS, Terraform, Delta Lake

- Engineered a real-time streaming data pipeline on Apache Kafka moving ~10K synthetic e-commerce events across 3 domains (orders, payments, shipping) through PySpark ETL jobs into a 3-layer medallion architecture (Bronze/Silver/Gold) on Delta Lake.
- Resolved nulls, duplicates, late arrivals, and schema evolution in the Silver layer; computed business KPIs (RFM segmentation with Spark window functions) in Gold and served them through SQL views and dashboards.
- Provisioned AWS infrastructure with Terraform for reproducible deployments across a 7-stage pipeline, from synthetic data generation to dashboard validation.

Tokyo Olympics: Azure Data Engineering Pipeline

Azure Data Lake Storage Gen2, Azure Databricks, PySpark, SQL

- Developed a 5-stage batch ETL pipeline (load, clean, transform, analyze, write) on Azure Databricks over ~12K Tokyo Olympics records (11K+ athletes) stored in ADLS Gen2.
- Cleaned and transformed raw CSVs with PySpark and SQL, applied data-quality checks, and published curated datasets back to ADLS Gen2 for downstream analytics.

Agentic Data Cleaner: Public MCP Server

Python, FastAPI, REST APIs, Model Context Protocol (MCP), Pytest — Live: <https://dc-mcp-26zy.onrender.com>

- Created an agentic data-cleaning engine for tabular data across 4 formats (CSV, Excel, JSON, Parquet) detecting 6 defect classes: duplicates, placeholder nulls, type mismatches, spelling variants, outliers, and PII. Safe fixes apply automatically while destructive actions require human approval.
- Deployed a public, keyless MCP server (10 tools) over Streamable HTTP for 4 coding clients (Muse, Claude, Cursor, VS Code), backed by REST APIs. Hardened with auth, SSRF guards, per-IP dataset isolation, and a fixed action registry with no code execution.

Notflix: Study-Material Sharing Platform

React, Node.js, Express, MongoDB, Redis, Socket.IO

- Launched a MERN study-material platform with REST APIs where students discover, upload, and download lecture notes (PDF/DOCX/TXT up to 10 MB) through an admin approval workflow, with JWT and Google OAuth auth and Socket.IO real-time chat rooms.
- Integrated OCR.space image-to-text (5 MB limit), Redis-backed caching and rate limiting, and Cloudinary uploads, plus endpoint hardening (Helmet, CORS allow-list, health checks).

EDUCATION

Chitkara University — B.E., Computer Science and Engineering (2024 – Expected 2028)

CGPA: 9.47/10

Relevant Coursework: Data Structures & Algorithms, Database Management Systems, Operating Systems, Computer Networks, Object-Oriented Programming, Linux

CERTIFICATIONS

- Cisco Networking Academy — Cybersecurity
- Infosys Springboard — Explore Machine Learning using Python
- Infosys Springboard — Design Thinking