

# Nikhil Mhatre

mhatrenikhil36@gmail.com | 6823139102 | GitHub/MhatreNikhil36 | LinkedIn/nikhil-nandkumar-mhatre

## SUMMARY

**Software Engineer** with 2+ yrs building **cloud-native solutions** and secure API backends using Python, Java and TypeScript. Designed, deployed, and tuned high-throughput Spark & Kafka workflows for significant efficiency gains. Graduating MS- CS program focused on distributed systems and cloud computing with a 4.0 GPA.

## EDUCATION

**University of Texas at Arlington | TX, USA**

**August 2023 – May 2025**

Master of Science | Computer Science

**GPA: 4.0/4.0**

*Relevant Courses: Data Analysis and Modeling Techniques, Cloud Computing & Big Data, Machine Learning, Distributed Systems, Secure Programming*

**Mumbai University | Navi Mumbai, India**

**August 2017 – July 2021**

Bachelor of Engineering | Computer Science & Engineering

*Relevant Courses: Database Management System, Object-oriented Programming, Artificial Intelligence, Operational Research*

## PROFESSIONAL EXPERIENCE

**Data Engineer, LTIMindtree**

**June 2021 – July 2023**

- Designed Java-based **serverless microservices** on **AWS (Lambda, MSK)** to ingest and process **real-time** employee lifecycle **events** to stream validated data to downstream audit systems
- Steered a cross-functional investigation of **HRIS** data anomalies across 40,000+ global records, analyzing **Spark** scripts and **Informatica** workflows, collaborating with Business Analysts and Data Scientists to implement a fix that **improved data accuracy**
- Developed workflows using **Python**, **stored procedures**, and **materialized views** that **eliminated 5+ hours** of manual effort weekly while enhancing data visibility and business insights through interactive Power BI dashboards
- Engineered and optimized **event driven ETL data** pipelines using **Snowflake**, **Spark**, **Kafka**, **Scala**, and **SQL**, processing data from 20+ sources, improving pipeline efficiency by 15%, and implementing data validation checks to maintain accuracy
- **Documented** software architecture, APIs, and runbooks to improve debugging and troubleshooting for **production incidents**, **reducing Mean Time to Recovery by 40%**

**Research Assistant, Ramrao Adik Institute of Technology**

**May 2019 – June 2020**

- Led a **team of five** to reduce research paper processing time by developing a solution with **version control** and **discussion boards** using **PHP**, **MySQL**, and **JavaScript**
- Digitized voting for Mumbai University student election system, resulting in automation for 750+ affiliated institutions
- Built a **scalable attendance management** platform capable of handling 15k+ concurrent requests using React and Spring Boot
- Recognized with **Best Summer Internship Project** for delivering **high-impact automation solutions**

## SKILLS

- **Programming Languages:** Java, Python, SQL, JavaScript (React/Node.js), Scala, Bash
- **Cloud & DevOps:** AWS (Lambda, S3, SQS, MSK, CDK), CI/CD Pipelines (GitHub Actions, Jenkins), Docker, Kubernetes, Snowflake
- **Data Engineering:** Apache Spark (Core, SQL), Kafka, Airflow, Flink, Hadoop, HiveQL, GCP BigQuery, PySpark, DataBricks, DBT
- **Databases:** DynamoDB, MongoDB, Azure SQL, MySQL, Redis
- **Frameworks/Tools:** Spring Boot, gRPC, TensorFlow, REST APIs, Informatica, Power BI

## CERTIFICATIONS

- AWS Certified Developer-Associate
- SnowPro Core: Snowflake
- Meta Front-End Developer Certificate
- JavaScript: Microsoft

## PROJECTS

**AI Enhanced Fitness Tracking Platform**

**2025**

- Led a 5-person team to develop a full-stack AI-powered fitness tracking application using React, Next.js, and AWS, managing end-to-end development from architecture to production deployment
- Architected backend (repository pattern/dependency injection) and React frontend (atomic design/Redux), integrating OAuth security and an ML recommendation engine to improve user engagement

**Distributed Key Value Storage using Raft Algorithm**

**2024**

- Implemented distributed consensus using Raft Algorithm Java, Python, gRPC ensuring fault tolerance with leader election recovery in under 5 seconds and simulated network failures and optimized RPC communication for minimal latency
- Enabled robust handling of permanent failures with minimum data overhead by using Merkel Tree for state synchronization
- Simulated network partitions and automated recovery for more than five failure scenarios, ensuring high system availability

**No-Code Data Pipeline Creator (a cloud-native web application)**

**2024**

- Developed no-code ETL platform by integrating PySpark with GCP services(Dataflow, BigQuery), validating 20+ transformation workflows across structured and unstructured test datasets
- Engineered a secure React frontend with OAuth 2.0 authentication and chunked multipart uploads to cloud storage, automating CI/CD via GitHub Actions to achieve high test coverage and slash deployment time from 30 to 5 minutes