Harsh Wadhawe

📞 +1 (979) 721-0991 ✉ harsh.wadhawe@tamu.edu in LinkedIn ○ GitHub 💼 Portfolio

Education

Texas A&M University, Mays Business School   **College Station, Texas**
Master of Science in Management Information Systems   **June 2026**

COEP Technological University   **Pune, India**
Bachelor of Technology in Computer Engineering   **June 2022**

Work Experience

Graduate Teaching Assistant, INFO Department - Mays Business School   **Jan 2025 – Present**

- Developed interactive modules in Excel, SQL, and web development, integrating real-world problem solving with Agile methodologies.
- Automated grading processes and optimized database queries to streamline evaluations for 100+ students.

Big Data Engineer, CITI   **Aug 2023 – June 2024**

- Engineered ETL pipelines to transform raw financial data from a Data Lake into actionable insights, collaborating with cross-functional teams.
- Designed and optimized distributed data models using PySpark, Java, Scala, and Hadoop, resulting in a 30% increase in operational efficiency.
- Leveraged data mining techniques to extract valuable patterns, and enhanced infrastructure with Cloudera, Apache HBase, and Hue, improving processing speeds by 35%.

Software Developer, CITI   **July 2022 – July 2023**

- Architected and implemented scalable web services using Java and Spring Boot within a microservices framework, applying object-oriented design principles.
- Led the migration from Tibco EMS to a multi-broker Apache Kafka cluster, establishing robust distributed storage and real-time data streaming, which improved throughput by 40%.
- Developed and maintained RESTful APIs with comprehensive testing (JUnit, Pytest, Swagger) to ensure high-quality code and system reliability.
- Employed advanced shell scripting on Linux systems to optimize runtime configurations, reducing manual intervention by 50%.

Certifications

- **AWS Certified Data Engineer – Associate**   View Certificate ⬈
- **AWS Certified Solutions Architect – Associate**   View Certificate ⬈
- **Machine Learning Specialization** — Stanford University & DeepLearning.AI   View Certificate ⬈
- **Professional Scrum Product Owner I (PSPO I)**   View Certificate ⬈

Projects

Real-Time Data Pipeline on AWS using Kafka and Spark   **View Project ⬈**

- Developed a real-time data ingestion and processing pipeline on AWS EC2 using Apache Kafka and Spark Streaming.
- Employed Python for data cleaning and transformation, ensuring high data quality and reliability.
- Productionized the pipeline to integrate with existing systems, thereby enhancing real-time analytics and distributed storage capabilities.

Document RAG Chatbot   **View Project ⬈**

- Created an end-to-end Retrieval-Augmented Generation (RAG) model using ChromaDB and OllamaLLM to handle document-based queries.
- Processed PDF uploads and generated document embeddings, delivering contextualized responses via Flask.
- Deployed the solution on AWS Elastic Beanstalk, ensuring scalability and robust performance.

Leadership Activities

Product Owner - WalkMe Application Integration
- Took charge of the WalkMe application integration at CITI following a leadership transition, ensuring project continuity through Agile sprint planning.
- Facilitated cross-team collaborations that led to a 20% boost in plugin functionality efficiency.

Lead Developer - Data Pipeline Development
- Led a team of five developers in designing and deploying a distributed Data Lake pipeline, improving data processing speeds by 35%.
- Coordinated with analysts and subject matter experts to optimize workflows and resolve technical challenges.

Technical Skills

Programming Languages: Python, Java, C++, Go, Scala, SQL   (*familiar: C#, Perl*)
Core Competencies: Object-Oriented Design, Data Structures, Algorithms, Operating Systems (Linux, Windows)
Distributed Systems: Distributed Storage, Web Services, Microservices Architecture, Apache Kafka, Hadoop
Big Data & Data Mining: Apache Spark, Hive, HBase, Cloudera, Hue, Flink, Data Mining Techniques
ETL & Data Warehousing: Data Modeling, ETL Development, Apache Airflow, MapReduce
DevOps & Cloud: Git, Docker, Kubernetes, Jenkins, AWS, Databricks, CI/CD
Data Visualization: Tableau, Power BI, Pandas
Methodologies: Agile, Scrum
Additional Skills: Shell Scripting, UNIX, Project Management

Achievements

Publication: Electronic Health Record Management using Hyperledger Fabric
- Presented at the 2022 ICCUBEA Conference and published in an IEEE journal.
- Developed a blockchain-based system for secure, decentralized management of electronic health records, enhancing data privacy.