Priyadharshan Sengutuvan

sengutuvan.p@northeastern.edu | (857) 397-8847 | linkedin.com/in/priyadharshan-sengutuvan

EDUCATION

Northeastern University, Khoury College of Computer Sciences — Boston, MA
Master of Science in Data Science — May 2026
Coursework: Data Management, Cloud Computing, Distributed Systems, MLOps, Big Data Analytics, Algorithms

Sri Krishna College of Engineering and Technology — Coimbatore, India
Bachelor of Technology in Information Technology — May 2023
Coursework: Big Data Analytics, DBMS, Cloud Computing, Distributed Systems, Data Mining

SKILLS

Programming: Python, SQL, C#
Data Engineering: Apache Spark, Apache Airflow, PostgreSQL, SQL Server, ETL Pipelines, Data Modeling
Cloud & Infrastructure: GCP (BigQuery, Cloud Storage), Azure DevOps, Docker, CI/CD Pipelines
ML Infrastructure: MLOps, Model Deployment, Vector Databases, LangChain, FastAPI
Tools & Frameworks: Git, REST APIs, Microservices Architecture

EXPERIENCE

AriesView — Remote (Boston, MA)
AI/ML Engineer Intern — Sept 2025 – Dec 2025

- Scaled document processing throughput by architecting RAG pipeline infrastructure with PostgreSQL vector storage and Docker containerization for OCR-based document ingestion.
- Reduced data retrieval latency by 45% by designing optimized chunk-based indexing systems for real estate PDFs, enabling sub-second semantic search across 100K+ documents.
- Improved system reliability by implementing automated data validation pipelines and monitoring workflows for PostgreSQL databases handling financial modeling data.
- Accelerated feature deployment by establishing Docker-based development environments and CI/CD automation for data pipeline integration.

Psiog Digital Private Limited — Chennai, India
Software Engineer Intern — Feb 2023 – Dec 2023

- Increased data processing efficiency by 25% by building automated ETL pipelines in C#/.NET that extracted, transformed, and loaded operational data from multiple sources.
- Reduced dashboard query latency by redesigning SQL Server database schemas with optimized indexing strategies and implementing materialized views for frequent queries.
- Enhanced system integration speed by developing REST APIs that centralized data access across siloed backend systems, enabling real-time data synchronization.
- Improved deployment reliability by implementing CI/CD pipelines in Azure DevOps with automated testing and rollback capabilities for data infrastructure.

ACADEMIC PROJECTS

AI-Powered Operations Assistant with RAG (GitHub)

- Processed 500K+ seller records daily by orchestrating Apache Airflow ETL pipelines with Docker containerization, ingesting data into GCP BigQuery for real-time analytics.
- Reduced data pipeline runtime by 60% by implementing parallel processing workflows and optimizing data partitioning strategies in BigQuery.
- Improved data availability by 50% by building automated data quality monitoring and alerting systems that detected and resolved pipeline failures within 15 minutes.

BERT-based Sentiment Analyzer on Amazon Reviews (GitHub)

- Processed 2M+ Amazon reviews by building scalable data ingestion pipelines with distributed preprocessing across multiple workers using Apache Spark.
- Reduced model training time by 35% by implementing efficient data loading pipelines with optimized tokenization batching and multi-threaded data preprocessing.
- Enhanced data pipeline reliability by implementing checkpoint mechanisms and retry logic for handling large-scale dataset processing failures.

Credit Card Fraud Detection (GitHub)

- Enabled real-time fraud detection by building streaming data pipelines that processed 10K+ transactions per second with sub-second latency using optimized feature engineering.
- Reduced false positive rate by 15% by designing feature stores with automated data quality checks and anomaly detection on high-dimensional transaction logs.
- Improved model serving reliability by implementing model versioning and A/B testing infrastructure for seamless deployment of fraud detection updates.