

ANEESH KOKA

koka.a@northeastern.edu | [LinkedIn](#) | 7573866334 | Boston, MA

EDUCATION

Northeastern University

Master of Science | Information Systems

May 2025

Boston, MA

Data Science and Engineering Tools, Parallel Machine Learning and AI, Big Data and Intelligence Analytics

Gitam Institute of Technology

Bachelor of Technology | Computer Science and Engineering

May 2022

Telangana, India

TECHNICAL SKILLS

- **Programming Languages & Concepts:** C/C++, Python, R, Java, SQL, Linux, GPU, RDMA network
- **Machine Learning/Deep Learning:** PyTorch, TensorFlow(TensorFlow certified developer), Deep Neural Networks(CNN,RNN), LLM, Natural Language Processing, Distributed Systems
- **Big Data/Cloud:** PySpark(spark), AWS, Airflow, Kafka, Snowflake, Dbt, FastAPI, Django, Redshift, Docker, Git
- **Data Visualization:** Tableau(Tableau certified associate), PowerBI, Excel, Matplotlib, Seaborn, Plotly

WORK EXPERIENCE

Accenture | Data Science Practitioner

August 2022 - July 2023

- Implemented large-scale data integration and data warehouse projects using Apache Airflow and Snowflake, processing over 1TB of data daily to ensure efficient data management and retrieval.
- Built Customer Data Platforms (CDP) and marketing data warehouses from scratch using AWS Redshift and dbt, resulting in a 30% increase in marketing campaign effectiveness through enhanced customer insights.
- Developed robust data engineering pipelines using Apache Spark, on-premise solutions, and AWS services (S3, Glue), reducing data processing time by 40% and improving system scalability to handle up to 5 million records per day.
- Developed 20+ performance dashboards in Tableau, integrating KPIs for comprehensive system performance insights

Salesforce | Data Engineer intern

April 2022 – July 2022

- Designed and implemented data models aligning with Process Builder and Flow Builder needs, boosting automation efficiency by 10% during my Salesforce tenure
- Led data transformation initiatives using Salesforce DTL, resulting in a 12% improvement in data suitability for automation workflows and a 15% reduction in data errors at Salesforce
- Optimized data streaming and ETL processes using Apache Spark orchestrated using Airflow, leading to a 20% increase in data retrieval speed and a 25% improvement in reporting performance at Salesforce
- Deployed scalable ML pipelines using Docker containers, orchestrated with Kubernetes, facilitating consistent development and production environments.

Child Society | Co-Founder

March 2022 – Present

- Successfully conducted over 75 technical workshops, benefiting 800+ students within the first operation year
- Raised funds and formed key partnerships, resulting in a 30% annual funding increase for program expansion

ACADEMIC & PERSONAL PROJECT

Data Pipeline Automation with Snowflake and Airflow

SQLAlchemy | Airflow | MYSQL

- Designed and implemented an end-to-end ETL data pipeline solution to automate data ingestion, transformation, and loading processes using SQLAlchemy and Apache Airflow
- Enabled dynamic Snowflake data warehouse integration with Airflow DAGs for seamless scalability and real-time processing, enhancing agility in meeting evolving business requirements
- Streamlined Airflow DAGs, harnessing Snowflake's parallel processing, slashing data processing time by 25%

Anomaly Detection of Industrial Control Systems Based on Transfer learning

CNN | NLP | OpenCV | GPU

- Built an 8-layer RNN and used fine-tuning technology for transfer learning to detect abnormal datasets of ICSs
- Enhanced neural network inference times by 20% through GPU optimization, efficient memory management, and custom CUDA kernel development
- Fused NLP-driven insights with a deep residual CNN to enable comprehensive multimodal anomaly detection

AI Enhanced Healthcare Response System

Airflow | Selenium | Pinecone | FastAPI | LLM

- Developed an AI-driven healthcare application using Streamlit and WhisperAI for voice-based user interactions, enabling immediate diagnostic support and personalized medication recommendations based on user medical histories.
- Orchestrated a comprehensive ETLT pipeline with Apache Airflow, automating Selenium web driver for medical data scraping, cleaning, validation, and loading.
- Integrated advanced symptom extraction using Llm and ensured data quality and security by managing data flow into Snowflake and Pinecone vector databases for efficient querying and retrieval.