NITHYASHREE BALAJI

Dallas, TX | (945) 243-3494| NXB220014@utdallas.edu | LinkedIn| Portfolio| GitHub

EDUCATION

University of Texas at Dallas, *MS, **Business Analytics and Artificial Intelligence** | GPA: 3.43/4.00*   **May 2025**
Big Data, Database Foundations, Applied NLP, Machine Learning, Predictive Analytics

SKILLS

Big Data Technologies: Hadoop, Spark (PySpark, Scala Spark), Kafka
Cloud Computing: GCP (Google Cloud Platform), Azure
Programming Languages: Python, Java
Data Engineering & ETL: Data transformation, Data pipelines, Apache Airflow
Data Modeling & Warehousing: Logical & Physical Data Models, Data Marts, SQL
Backend Development: Microservices, API Development, System Integration
Business Intelligence & Analytics: Data Governance, Applied Business Acumen

PROJECTS

Scalable ETL Pipeline for Real-Time Transaction Processing (*Jan 2025***)**
- Developed a real-time ETL pipeline using Kafka, PySpark, and Redshift to process high-velocity financial transactions.
- Automated data ingestion and transformation using Apache Airflow, reducing processing time by 40%.
- Built SQL-based validation scripts, ensuring 99.9% data accuracy and compliance with data governance policies.

Big Data Analytics Platform for Customer Insights (*Dec 2024***)**
- Designed data lake architecture on Google Cloud Platform (GCP) using BigQuery for customer behavior analytics.
- Developed optimized Spark-based data pipelines, reducing data processing time by 50%.
- Enabled data-driven decision-making by building BI dashboards that improved customer engagement by 25%.

Intelligent Data Integration for E-Commerce Personalization (*Nov 2024***)**
- Engineered a real-time data streaming solution using Kafka and Azure Data Factory to consolidate customer data from multiple sources.
- Implemented machine learning-driven recommendations, increasing click-through rates (CTR) by 20%.
- Developed Airflow workflows to automate data ingestion and transformation, reducing latency by 35%.

Data Governance and Compliance Automation (*Oct 2024***)**
- Designed and implemented a data governance framework to ensure compliance with company-wide policies.
- Automated SQL-based data quality checks, reducing manual audit time by 50%.
- Integrated compliance workflows with Apache Airflow, ensuring data integrity and security.

LEADERSHIP AND VOLUNTEERING

Data-Driven Decision Making – Industry Conference (*Aug 2024***)**
- Presented data insights and analytics strategies at an industry conference with 150+ professionals, showcasing expertise in big data and business intelligence.
- Led discussions on real-time data processing, ETL pipelines, and cloud-based analytics, aligning with enterprise data-driven strategies.

Operational Efficiency & BI Analytics – DataViz Hackathon (*Apr 2024***)**
- Led a team in analyzing 200,000+ call center records using Power BI and Python, identifying inefficiencies and improving operational performance by 15%.
- Developed interactive dashboards that provided real-time insights for data-driven decision-making.

Real-Time Data Engineering for Healthcare – Axxess Hackathon *(Mar 2024)*
- Designed a real-time data processing solution for patient health tracking, leveraging Kafka and Apache Airflow.
- Built an automated analytics dashboard, improving predictive decision-making in healthcare operations.

PROFESSIONAL EXPERIENCE

Software Engineer | Cognizant Technology Solutions | *Chennai, India*   *Jan 2022 - Jul 2023*
- Built and optimized scalable ETL pipelines for real-time transaction data processing.
- Developed BI dashboards in Power BI and Tableau, enhancing real-time analytics for senior leadership.
- Automated data workflows using Apache Airflow, reducing processing time from 3 hours to 15 minutes.
- Designed predictive models for strategic decision-making, improving forecasting accuracy by 25%.
- Led training workshops for 50+ employees, promoting a data-driven culture.

Programmer Analyst | Cognizant Technology Solutions | *Chennai, India*   *Aug 2020 - Dec 2021*
- Developed and maintained ETL pipelines using SQL and Python, streamlining data processing.
- Collaborated with IT leaders to migrate on-premises data solutions to AWS (Redshift, S3), improving scalability.
- Designed data models and warehousing solutions, enabling faster analytics and reporting.