

NAISHAL SHAH

+1(930) 904-3950 | shahnai@iu.edu | [LinkedIn](#)

PROFESSIONAL SUMMARY

Experienced Data Engineer and ETL Developer with 4 years in Banking and Finance, specializing in scalable data solutions. Pursuing a Master's in Data Science at Indiana University, I apply advanced machine learning techniques to drive impactful business insights.

EDUCATION

Indiana University – Bloomington, M.S. in Data Science – 3.9/4 GPA **Aug 2024 – May 2026**
Coursework: Statistics, Applied Machine Learning, Computer Vision, Cloud Computing
Vellore Institute of Technology – Vellore, B.Tech in Computer Science – 8.28/10 CGPA **July 2016 – June 2020**

TECHNICAL SKILLS

Programming & Libraries: Python (Pandas, NumPy, PySpark, Scikit-Learn, Tensorflow) SQL, PL/SQL, R, Hive, JavaScript, Presto
Cloud & Data Platforms: AWS (Redshift, S3, EMR, SQS, SNS, MWAA, CloudWatch, Lambda, Athena), Azure Databricks, Snowflake, Hadoop, Apache Spark, Apache Hive, MySQL Workbench, Oracle SQL Server
Machine Learning & MLOps: Regression, Clustering, Deep Learning, Neural Networks, Computer Vision, Data Forecasting, Reinforcement Learning, Feature Engineering, PCA, Model Deployment
Data Visualization Tools: Tableau, Power BI, Amazon QuickSight, MS Excel
ETL & Workflow Automation: Airflow, Git, CI/CD Pipelines, Jenkins, RLM, BitBucket
Certifications: CCA175 (Cloudera Certified Hadoop and Spark Engineer), LTI Shoshin School AWS Cloud Practitioner

EXPERIENCE

Data Engineer II, EXL **Feb 2024 – Aug 2024**
Client – Visa Inc.

- Designed and managed **Issuer Processor and Acquirer Processor monthly risk reports** for the regions - USA, Canada, Cemea and **created client-friendly dashboards** using Tableau managing over Tableau Server.
- Performed **ETL operations on Fraud and Settlement transaction data** using Hive and PySpark, streamlining data flow, enabling the generation of critical downstream reports **25% faster to improve reporting accuracy and timeliness**.
- Developed an **end-to-end ETL system on Apache Airflow** to automate data ingestion, transformation, and the generation of final risk reports and visualizations for the Canada and CEMEA regions.

Associate Data Engineer, TresVista Analytics **Oct 2023 – Feb 2024**
Client – CPP Investments

- Developed and oversaw over **20 ETL pipelines** for multiple vendors using Apache Airflow, Databricks, PySpark, and Python, ensuring streamlined data processing and **reducing data latency by 40%**, resulting in a **15% increase in reporting efficiency**.
- Utilized **AWS SQS** and **SNS** services to streamline data ingestion from vendor buckets and **APIs** into client **S3** buckets, reducing **ingestion time by 60%**, leading to faster and more reliable downstream processing.
- Performed **DQ** checks, built **Tableau dashboards**, and mentored junior analysts to ensure project success and team proficiency.

Senior Data Engineer, LTIMindtree LTD **Jul 2020 – Oct 2023**
Client – Citibank N.A.

- Led a team in developing a data solution that migrated legacy SWIFT MT messages to ISO 20022-compliant MX formats—leveraging the XML-based, globally adopted standard (established in 2014) to deliver richer, structured financial data that improves straight-through processing, compliance, and operational efficiency.
- Implemented an **alert system using scenario modeling** to classify high-risk accounts (HRAC) in the NAM, LATAM, and MX regions, reducing false positives by 18% and improving identification accuracy by 25% for more effective risk mitigation.
- Migrated the Compliance Data Warehouse framework from legacy systems to EAP servers for banking and financial data in 15 LATAM countries, reducing data retrieval time by over 50% and enhancing scalability to support triple data growth.
- Worked with clients to understand requirements, turning them into appropriate data models and mapping documents for **transactional data pipeline – Cash, Wire and MI Transactions**.

PROJECTS

Fantasy Premier League Optimization | Random Forest, XGBoost, Reinforcement Learning **Aug 2024 – Dec 2024**

- Developed a machine learning system for Fantasy Premier League decision-making using **Random Forest, Gradient Boosting, and Reinforcement Learning**, achieving **83.87% accuracy** in captaincy selection, fixture difficulty ratings, and team optimization.
- Leveraged advanced **feature engineering** and **real-time data analysis** to deliver actionable insights, demonstrating expertise in data preprocessing, model evaluation, and result visualization for sports analytics.

Detection of Phishing Websites | Python, Anaconda Jupyter, ML algorithms **Dec 2019 – May 2020**

- Led a project team to develop a GUI-based phishing detection web application enabling users to analyze URLs with four established algorithms plus a custom one—reducing time complexity by 90% while maintaining 95% accuracy and identifying malicious sites.