SON T. D. LE

(269) 207-5878 | linkedin.com/in/son-le-3b3313268 | github.com/CLAY3106 | Son.LeDinhTruong24@kzoo.edu

---

EDUCATION:

Kalamazoo College   **Kalamazoo, MI**
Bachelor of Arts in Computer Science and Mathematics   *Expected Graduation: June 2028*
GPA: 3.806

SKILLS:
- **Languages:** Python, Java, SQL
- **Frameworks:** Apache Spark, Apache Airflow, scikit-learn
- **Data / Cloud:** PostgreSQL
- **Tools: Docker,** Git, Linux, Kaggle, RStudio, Google Workspace, Microsoft Office

---

RELEVANT COURSES:

Data Structures, Object-Oriented Programming, Calculus III, Design and Analysis of Algorithms, Machine Learning, Probability, Cryptography

---

EXPERIENCE:

FPT IS Company Limited   **Hanoi, Vietnam**
Data Engineering Intern   *July 2025 - September 2025*
- Migrated selected Hive/HDFS datasets to PostgreSQL by redesigning schemas and indexing strategies; validated row-level consistency against source tables, reducing query latency by ~3×.
- Designed batch ingestion workflows to pull data from internal RESTful services and scheduled FTP exports, implementing controlled batch loads (~100 records per request) to ensure stable downstream processing.
- Built containerized ETL pipelines using Airflow, Spark, and Docker to orchestrate extraction, transformation, and loading of ~1.2M records per week into PostgreSQL.
- Executed schema normalization, rule-based deduplication, validation checks, structured logging, and idempotent rerun mechanisms, reducing duplicate records by ~80% and cutting failure recovery time by ~40%.

Department of Computer Science & Mathematics - Kalamazoo College   **Kalamazoo, MI**
Teaching Assistant   *September 2025 - Present*
- Applied standardized grading rubrics to assignments and exams across sections, ensuring consistent evaluation while handling ~100 papers per week.
- Managed high-volume grading using structured checklists to maintain accuracy and timely turnaround under workload constraints.
- Provided concise, rubric-aligned written feedback on coding and math assignments to clarify errors and expected solutions.
- Identified recurring misconceptions in student work and addressed them through targeted comments, contributing to improved performance in later submissions.

Collaboration Center - Kalamazoo College   **Kalamazoo, MI**
Collaboration Center Associate   *September 2025 - Present*
- Assisted ~25 students per week with debugging and algorithm design during tutoring sessions, helping them identify errors and improve solution approaches.
- Guided students through systematic debugging techniques and common algorithm patterns to address recurring mistakes and improve problem completion.
- Facilitated labs and peer study groups with structured activities, supporting collaborative learning for groups of ~20 students per session.
- Explained course concepts and solution strategies aligned with class assignments, contributing to improved engagement and understanding in subsequent coursework.

Football Data Pipeline   **Kalamazoo, MI**
Personal Project   *January 2026 - Present*
- Implemented a football data pipeline in Python using Docker and Apache Airflow to ingest match data, apply deterministic cleaning and transformations, and load results into a relational database, enabling repeatable refreshes of analytics-ready tables.
- Structured ETL code into separate ingestion, transformation, and load steps with explicit schemas and simple validation checks, which reduced debugging time and made it straightforward to add new seasons or competitions.
- Led the project independently, handling pipeline design, Airflow DAG configuration, containerization, and documentation, resulting in a working, well-documented system suitable for experiments and downstream analysis (e.g., modeling and simulations).

LEADERSHIP EXPERIENCE:

The STEAM Project   **Hatinh, Vietnam**
Co-founder/Head of Organization/Program Sponsor Mentor   *April 2024 - July 2025*
- Co-founded and directed a student-run STEM camp, coordinating a 30-member team to deliver hands-on programs to ~120 campers.
- Designed and facilitated debates, experiments, and team-based challenges, driving high participation and positive participant feedback.
- Secured sustainability by leading fundraising and budgeting mentorship, raising approximately $1,500 USD to support continued operations.