Preston Thomas

Grand Rapids, Michigan

CELL: (248) 990-6194
EMAIL: preston510@outlook.com
GITHUB: www.github.com/preston-thomas
LINKEDIN: www.linkedin.com/in/prestonthomas6

SKILLS & CERTIFICATIONS

Languages: SQL, Python, Java

Databases: PostgreSQL, Microsoft SQL Server, SQLite

Frameworks and Tools: Apache Spark, PySpark, Databricks, SSIS, SSRS, Power BI, Pandas, Microsoft Fabric, Azure Data Factory

Certifications: Microsoft Certified: Fabric Data Engineer Associate (DP-700)

TECHNICAL WORK EXPERIENCE

Data Engineer Intern

Mutually Human, Grand Rapids, MI
July 2024 - Present

- Engineered a **multithreaded Java service** orchestrated via **SSIS** to **concurrently ingest 8 tables** across multiple source systems, **reducing runtime by 92% (over 3 hours to 15 minutes)**, while **enforcing 5-day windowing logic** to **load 50,000+ records nightly**.
- Owned development of an **end-to-end ETL pipeline** for a **$20B national retailer** using **Azure Data Factory** and **Databricks**, handling chained API dependencies, inconsistent response schemas, and 30-day retention constraints. Implemented all cleansing and transformation logic in **PySpark**, producing **Medallion-layered** datasets and **Star Schemas** for downstream analytics.
- Serve as the technical lead for **evaluating** and **implementing** out-of-the-box **LLM solutions alongside data engineering pipelines**, validating **NLP** and **predictive models** against enterprise datasets to enable context-aware analytics use cases.
- Improve data reliability and support reporting workflows for **20+ clients** by developing and maintaining efficient **ETL pipelines** using **SQL**, **Python**, and **PySpark**.

Information Technology Intern

Morrison Industries, Grand Rapids, MI
January 2024 - July 2024

- Architected a **SQL-based** reconciliation tool to integrate asset inventory with carrier usage logs, identifying **97** retired devices and **eliminating $45,000+** in annual overhead.

EDUCATION

B.S. in Computer Science

Grand Valley State University, Allendale, MI
August 2022 - May 2026 (Anticipated)

- **Minor:** Mathematics
- **GPA:** 3.84
- **Relevant Coursework:** Applied Machine Learning, Data Structures & Algorithms, Operating Systems
- **Honors:** Dean's List (all semesters), First-Year College of Computing Scholarship, Frederik Meijer Honors College

PERSONAL PROJECTS

Sickle Dodge Autonomous Agent Pipeline

- Crafted a **real-time telemetry extraction pipeline** by modifying the game's **Lua source code** to broadcast internal state data via **UDP sockets** to a custom **Python** listener.
- Constructed an **automated ETL workflow** using **Pandas** to capture, clean, and augment gameplay datasets, training a **predictive model** to achieve autonomous play.

NFL Advanced Statistics Data Pipeline and Visualization

- Designed a **PostgreSQL** schema to ingest and centralize historical NFL play-by-play data from open-source repositories.
- Developed a custom **PyQt** desktop application integrating **Pandas** and **Matplotlib** to execute optimized **SQL** queries and render interactive season-performance visualizations.