Course Intro

Elements of Data Integration (CS 333E)

Jan 16, 2026

The traditional data warehouse architecture

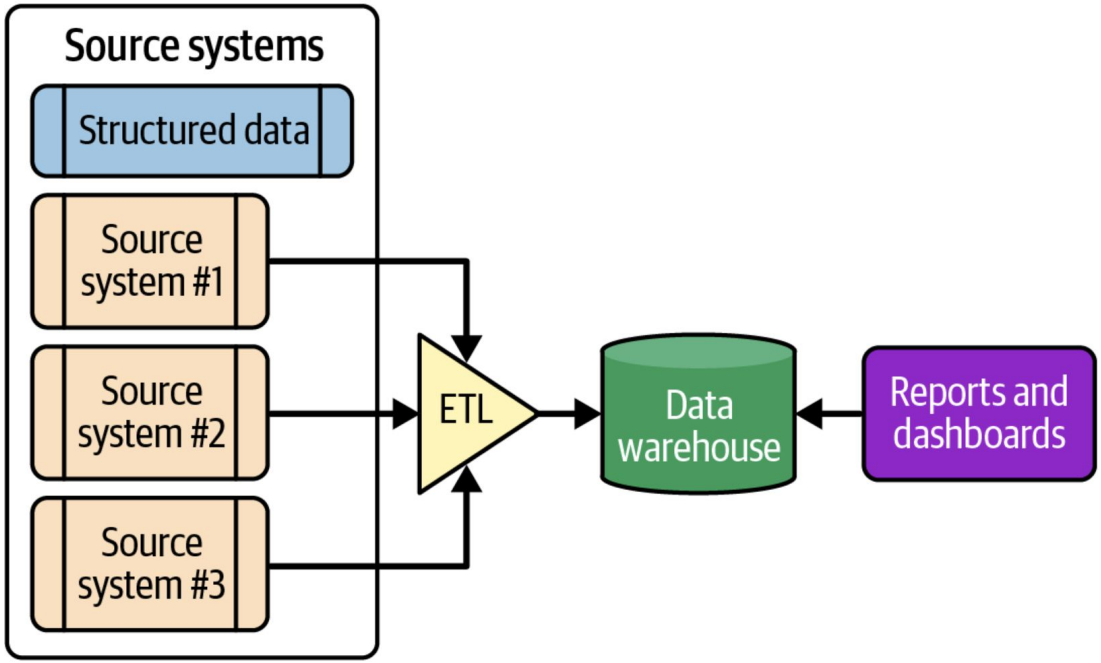

Figure 2-1. Data warehouse architecture

Source: Gaurav Ashok Thalpati, Practical Lakehouse Architecture, First Edition, O'Reilly, 2024.

The traditional data lake architecture

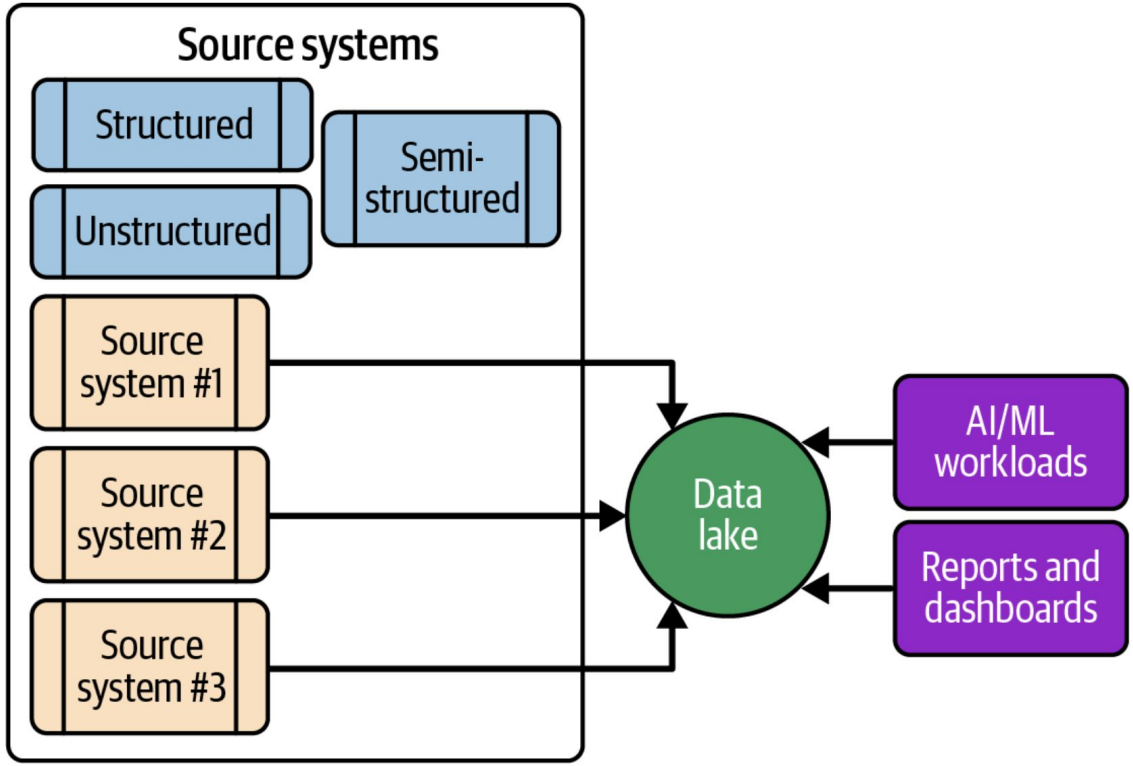

Figure 2-2. Data lake architecture

Source: Gaurav Ashok Thalpati, Practical Lakehouse Architecture, First Edition, O'Reilly, 2024.

Convergence of warehouse and lake architectures

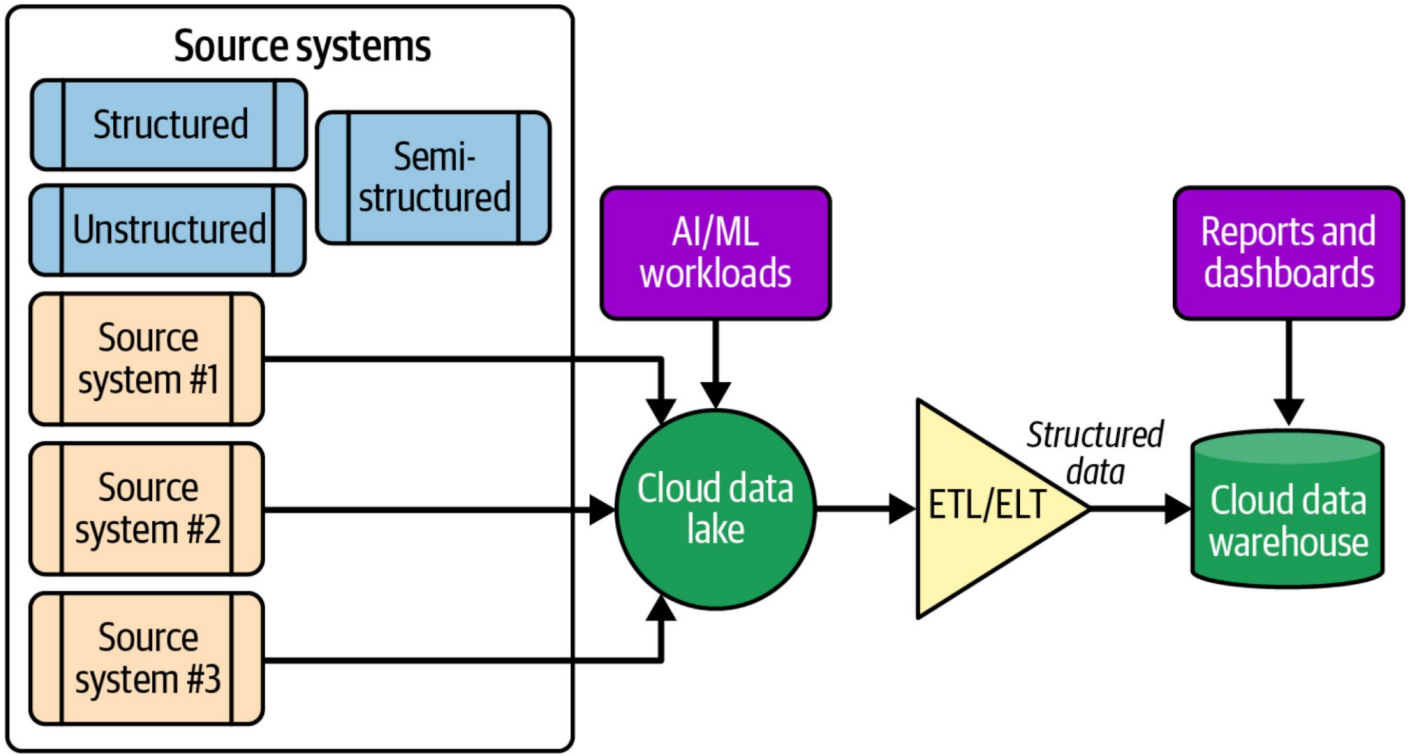

Figure 2-3. Data platform based on a combined architecture (cloud data lake + cloud data warehouse)

Source: Gaurav Ashok Thalpati, Practical Lakehouse Architecture, First Edition, O'Reilly, 2024.

How this class will work & what I expect from you

- Major portions of our class time will be spent on project work. Be prepared for hands-on keyboard work in class.
- Be prepared to finish off your project outside of class.
- Work collaboratively with your partner on all the projects, which means honoring your commitments.
- If you have to miss class, email me, the TA, and your partner in advance.
- If you are having some collaboration challenges with your partner, email me ASAP and come to office hours to discuss your options.
- Assignments will come with code samples to help you get started.
- Assignments won't come with a complete playbook, you are expected to figure some things out on your own.
- We will offer you a resubmission window on most assignments.
- Our week-by-week schedule is tentative. It may be adjusted based on the majority needs of the class.
- Please use Ed for questions and issues that you run into, just but be careful not to overshare your work.
- Share your feedback with me early and often, especially if you feel that you are falling behind or if you feel that we are not moving fast enough.