

Course Intro

Foundations of Data Warehousing

August 24, 2026

Data Warehouse Architecture 1.0

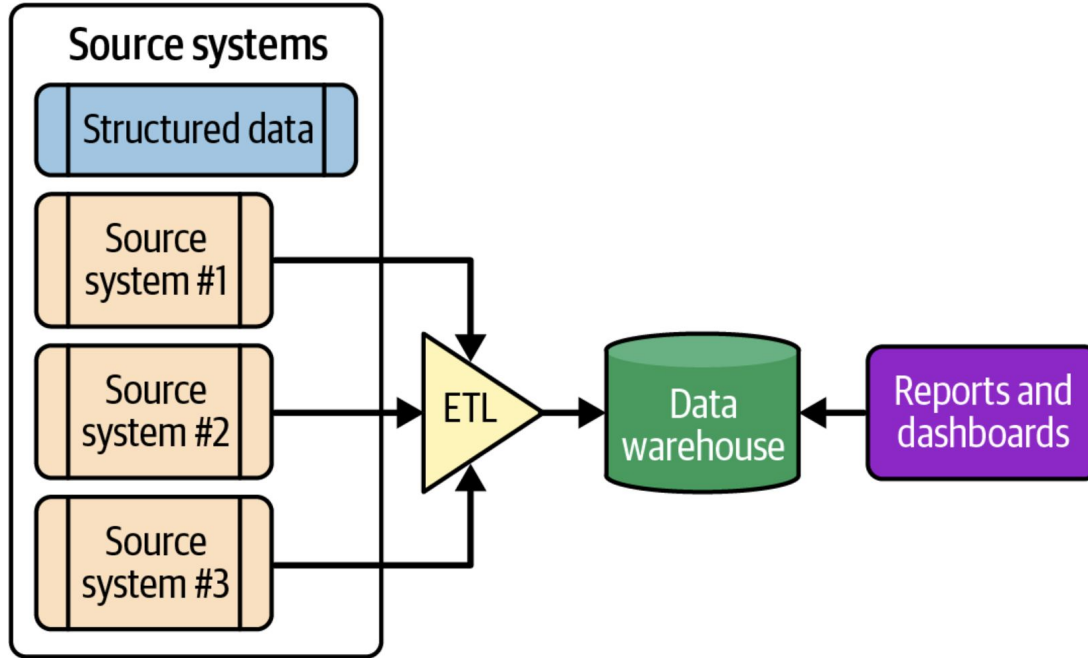


Figure 2-1. Data warehouse architecture

Data Warehouse Architecture 2.0

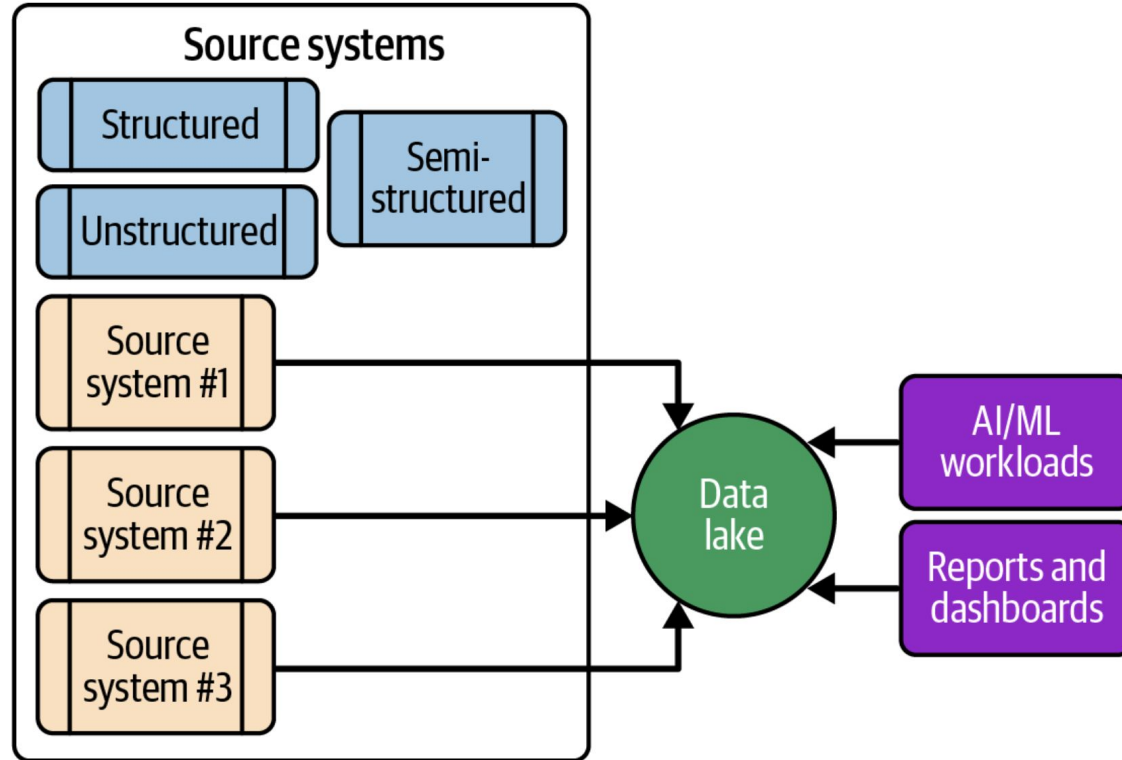


Figure 2-2. Data lake architecture

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

Data Warehouse Architecture 3.0

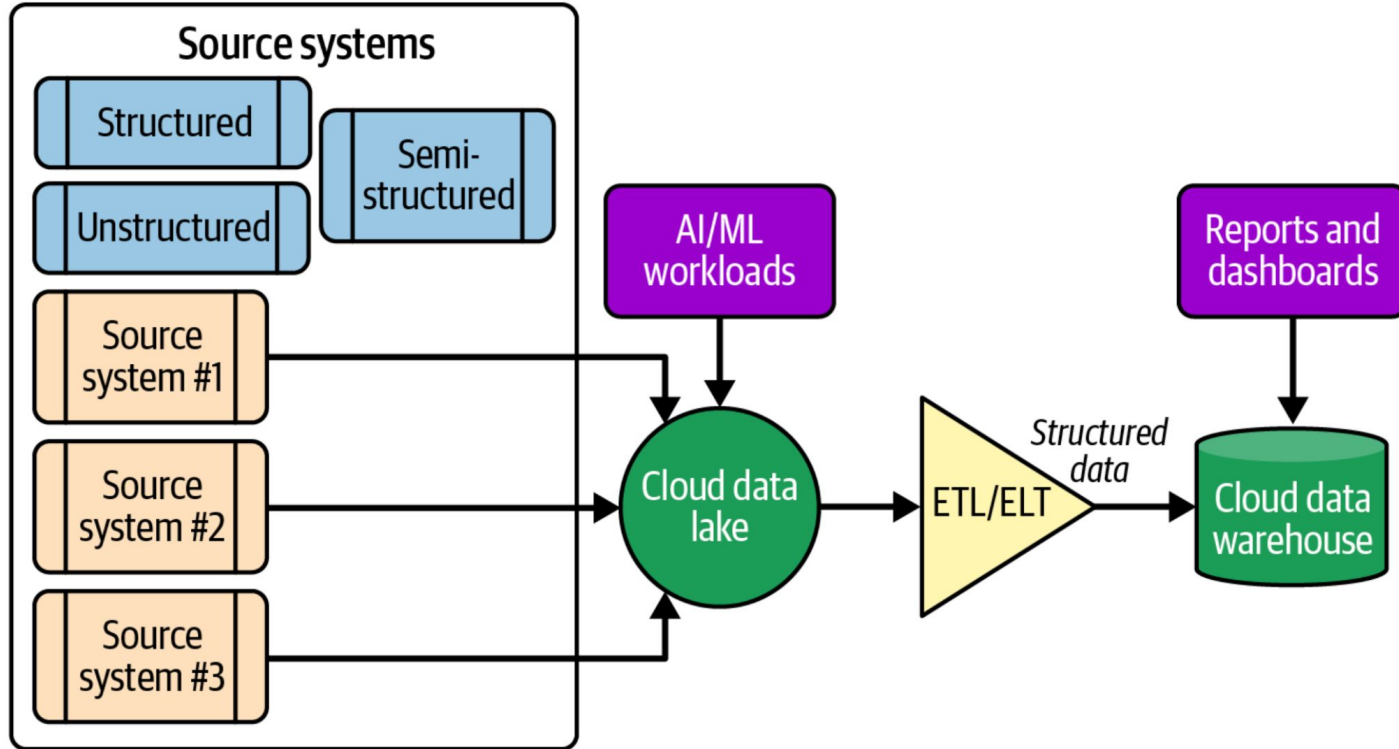
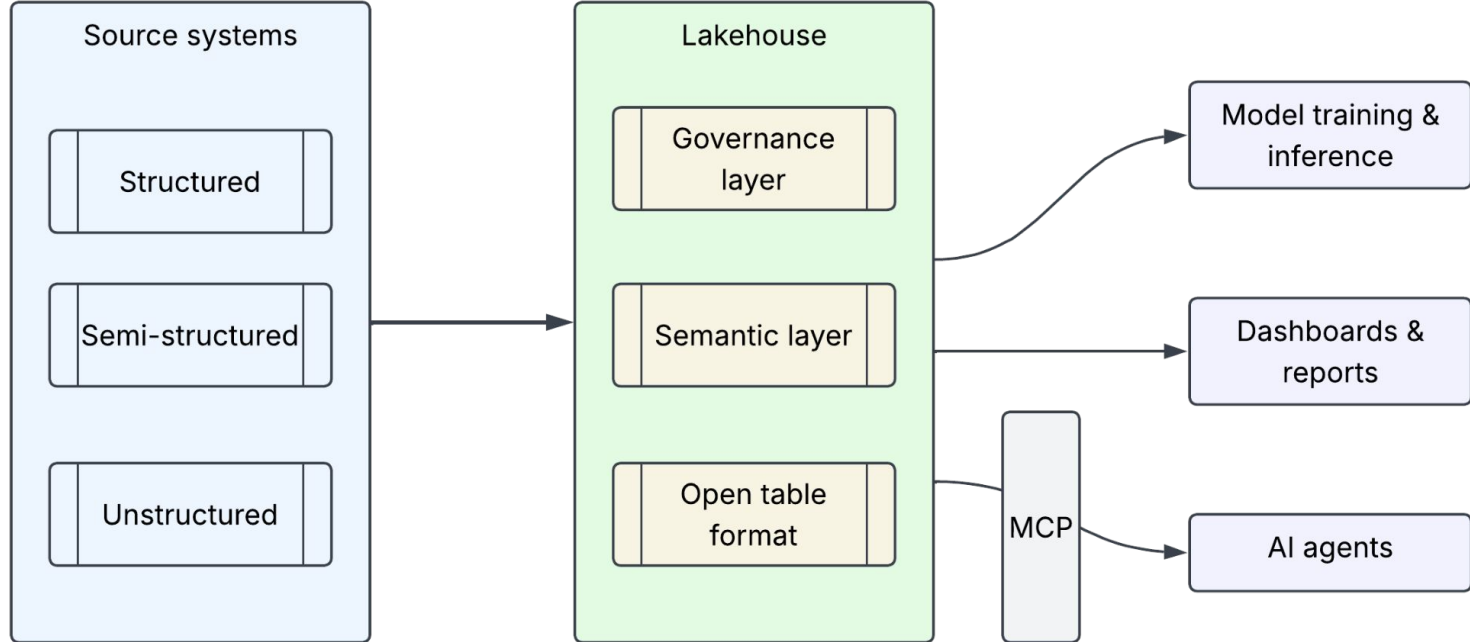


Figure 2-3. Data platform based on a combined architecture (cloud data lake + cloud data warehouse)

Data Warehouse Architecture 4.0



Claude Test

Step 1: Verify access to Claude:

- Go to claude.ai
- Log in with your EID@eid.utexas.edu

Step 2: Claude Code Install on MacOS:

- `brew install --cask claude-code`
- `claude`
- `/login`
- `/exit`

Step 2: Claude Code Install on Windows:

- Open a Windows PowerShell
- `irm https://claude.ai/install.ps1 | iex`
- `claude`
- `/exit`

Tips & tricks and what to keep in mind

- The spine of this course is the Term Project. Many of the activities we will do in class are in support of the project.
- Find a compatible project partner. Divide up the work evenly and honor your commitments.
- If you are having collaboration challenges, email me or come to office hours.
- If you have to miss class, email me, the TA, and your partner in advance.
- Assignments will generally come with code samples, not a complete playbook.
- Start the assignment as early as possible because the tech we're using changes so fast and there are some unknown unknowns.
- Our week-by-week schedule is tentative. It will be adjusted as we go based on reality.
- Ask questions and share your feedback as early and often as possible.

Lecture Two

Foundations of Data Warehousing

August 26, 2026

What is an Entity?

An entity is not a table or a file.

Most entities represent real-world things. Examples:



Places



Movies



Products

They represent distinct, identifiable concepts in the world.

In a DW setting:

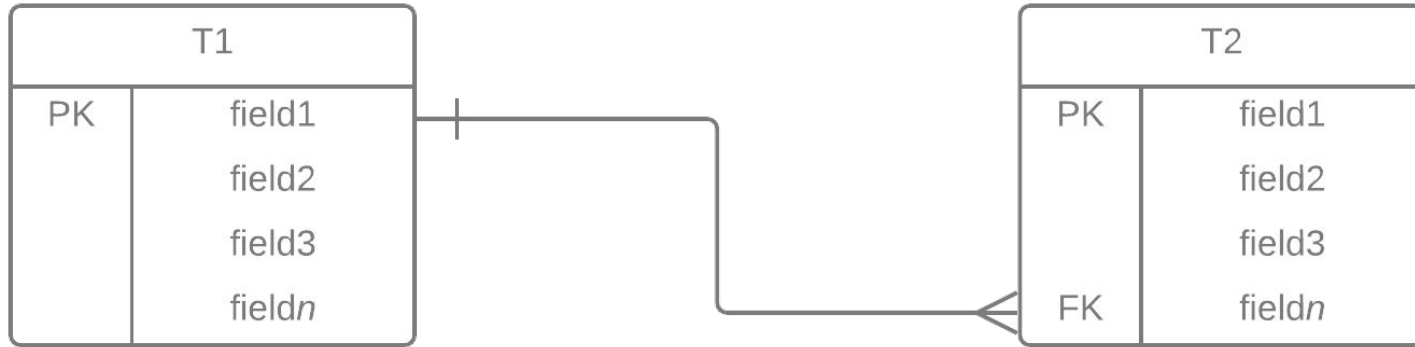
Each entity is represented by a schema.

Each entity has a unique identifier called a Primary Key (abbreviated to PK).

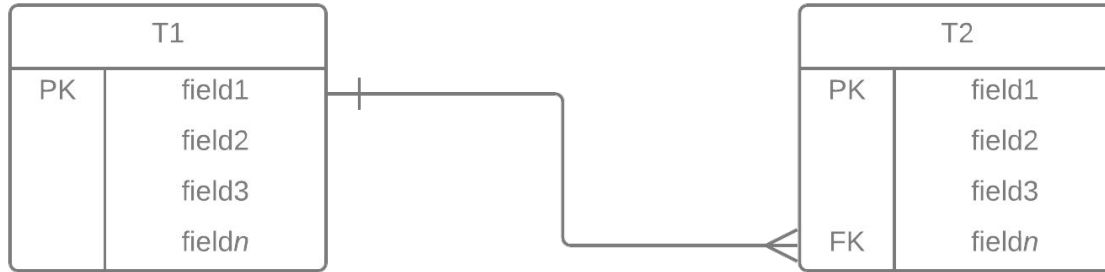
Entities do not contain other entities.

Entities link to other entities. These links are called relationships.

One-to-Many (1:*m*) Relationships



One-to-Many (1:m) Relationships



Author

<u>id</u>	name	section
1	Mary Tuma	news
2	Michael King	arts
3	Nina Hernandez	news
4	Sunil Kumar	music

Article

<u>id</u>	title	date	authid
1	Turmoil at the Zoo	2019-01-26	1
2	CodeNEXT's New Friend	2019-01-27	1
3	Quote of the Week	2019-01-27	3
4	SXSW News	2019-01-28	2
5	More from Steve Adler	2019-01-28	1

One-to-One (1:1) Relationships



One-to-One (1:1) Relationships



Article

<u>id</u>	title	date	authid
1	Turmoil at the Zoo	2019-01-26	1
2	CodeNEXT's New Friend	2019-01-27	1
3	Quote of the Week	2019-01-27	3
4	SXSW News	2019-01-28	2
5	More from Steve Adler	2019-01-28	1

Article_Stats

<u>id</u>	clicks	likes	dislikes	comments
1	120	45	9	13
2	0	0	0	0
3	8	0	0	2
4	30	4	0	1
5	9	1	3	3

Many-to-Many ($m:n$) Relationships



Many-to-Many (m:n) Relationships



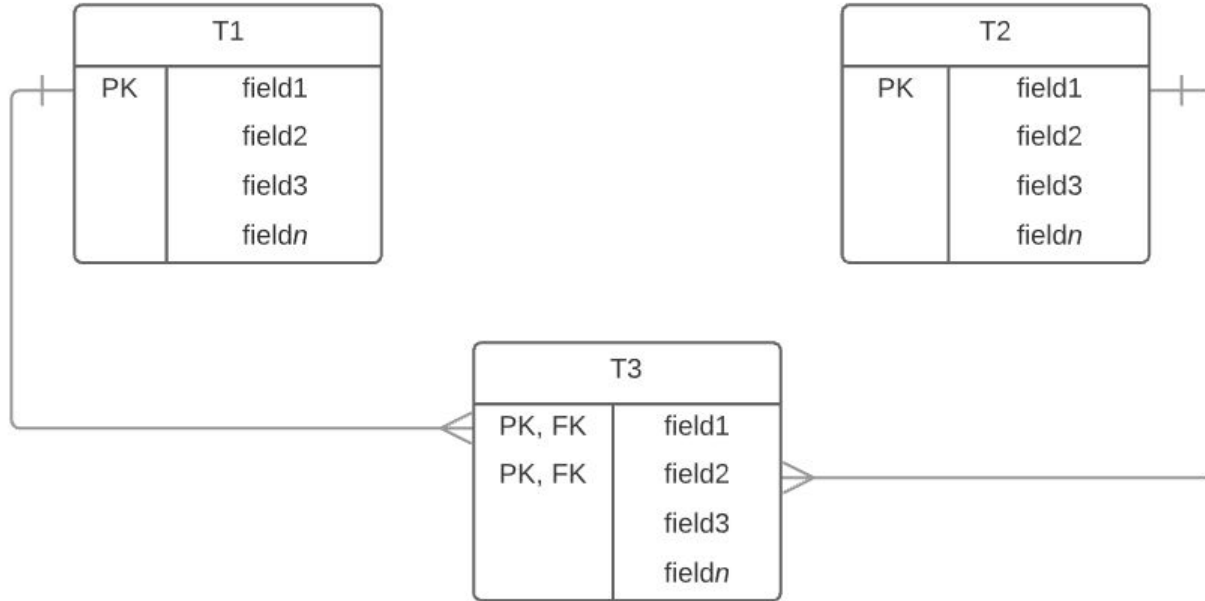
Tag

<u>id</u>	tag	<i>aids</i>
1	Politics	1, 2, 5
2	Austin	1, 2, 3, 4, 5
3	Mayor	3, 5
4	Business	1, 2, 5
6	Land Development	2
37	Animals	1

Article

<u>id</u>	title	date	<i>authid</i>	<i>tids</i>
1	Turmoil at the Zoo	2019-01-26	1	4, 37, 2
2	CodeNEXT's New Friend	2019-01-27	1	2, 6
3	Quote of the Week	2019-01-27	3	2, 3
4	SXSW News	2019-01-28	2	2, 40, 7
5	More from Steve Adler	2019-01-28	1	2, 3, 4

Representing Many-to-Many ($m:n$) Relationships



Many-to-Many ($m:n$) Relationships

Tag

<u>id</u>	tag
1	Politics
2	Austin
3	Mayor
4	Business
6	Land Development
37	Animals

Article

<u>id</u>	title	date	<i>authid</i>
1	Turmoil at the Zoo	2019-01-26	1
2	CodeNEXT's New Friend	2019-01-27	1
3	Quote of the Week	2019-01-27	3
4	SXSW News	2019-01-28	2
5	More from Steve Adler	2019-01-28	1

Tagged_Article

<u>tid</u>	<u>aid</u>
4	1
37	1
2	1
2	2
6	2