

NOTETAKERS NEEDED

We are looking
for students

Recording ...

NOTETAKER

class. If you would
like to volunteer
please email
notetaker@uwindSOR.ca



NOTETAKERS NEEDED

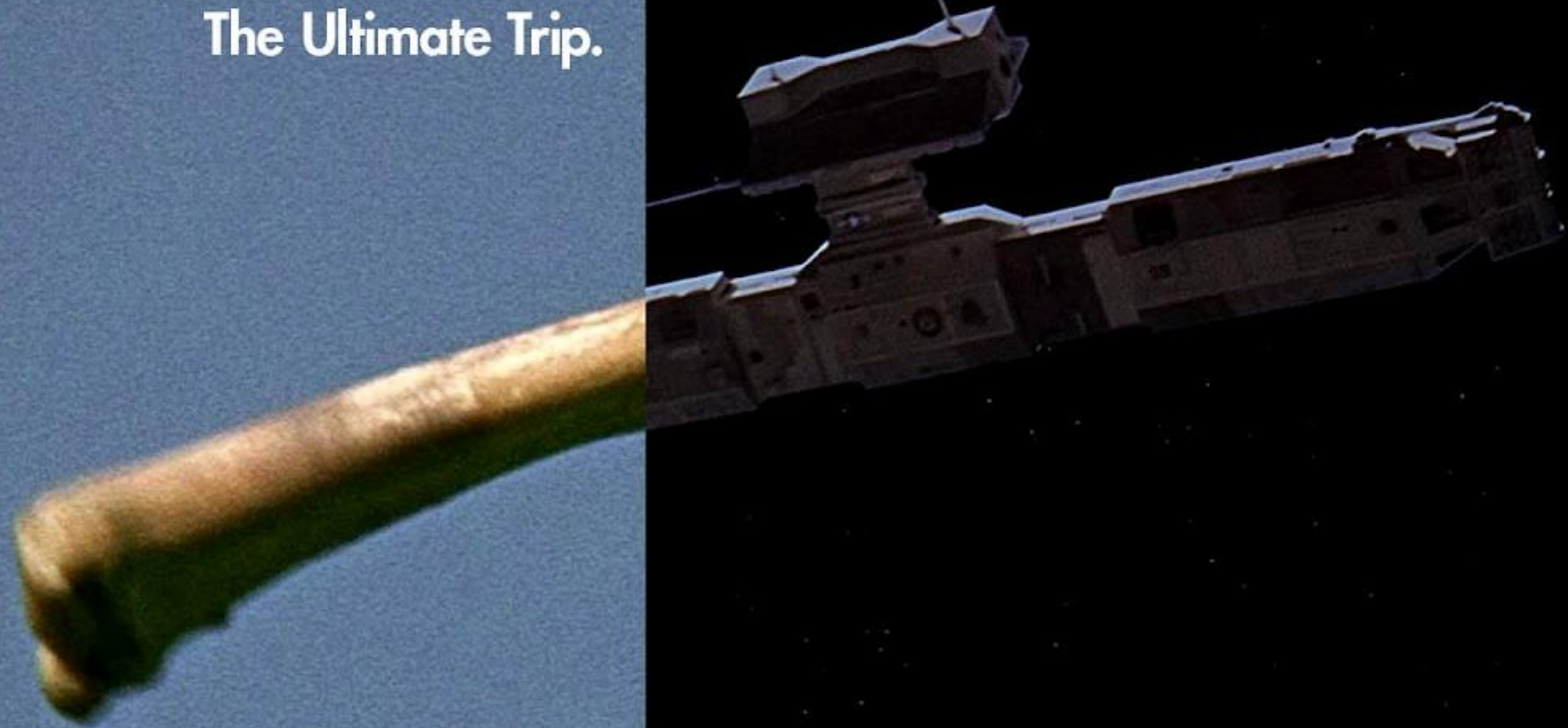
**VOLUNTEER
TO BE A
NOTETAKER**

We are looking
for a volunteer
notetaker for this
class. If you would
like to volunteer
please email
notetaker@uwindSOR.ca



2022: A Data Odyssey

The Ultimate Trip.



From apelike ancestors to human being: 6 million years
From bones to spaceships
2001: A Space Odyssey (1968), Stanley Kubrick

This Week

4

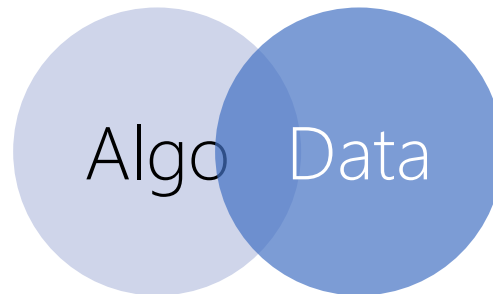
Welcome | Course Information | Data Modeling | Memory | File | Database
Welcome | Course Information | Data Modeling | Memory | File | Database
Welcome | Course Information | Data Modeling | Memory | File | Database
Welcome | Course Information | Data Modeling | Memory | File | Database
Welcome | Course Information | Data Modeling | Memory | File | Database
Welcome | Course Information | Data Modeling | Memory | File | Database

Welcome | Course Information | Data Modeling | Memory | File | Database

2022: A Data Odyssey × Academy

5

Algorithm Design
Algorithm Analysis
Artificial Intelligence (AI)
Machine Learning
Data Mining



Data Structure (Memory)
File Structure (File Systems)
[Database Management Systems](#)
Data Warehouse
Big Data
Cloud

2022: A Data Odyssey × Real World

6

Data Modeling: Real World Entity

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Computable Entity

2022: A Data Odyssey × Real World

7

Data Modeling: ~~Real World Entity~~

Conceptual Level | Logical Level | Physical Level

~~Conceptual Level~~ | Logical Level | Physical Level

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Computable Entity

Data Modeling × Conceptual Level

8

1. Identify Real World Entities, Attributes, Relationships
 2. Create Schema
- 

Data Modeling × Conceptual Level

9

Movie **Schema**

Title (char[])

Genre (char[])

Language (char[])

RunningTime (int)

Poster (char[])

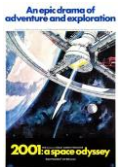
Movie Data **Instance**

2001: A Space Odyssey

Sci-fi

English

142



Data Modeling × Conceptual Level

10

Data **Schema**

Data about Data: Meta-data

Defined at Setup Time

Rarely Change

Data **Instance**

Actual Data

Inserted at Running Time

Rapidly Change

Must Conform to Schema

Data Modeling × Conceptual Level

11

Movie × Director × Company Relationships:

Director makes Movie

| Movie is made by Director

Company distributes Movie

| Movie is distributed by Company

Company budgets Movie

| Movie is budgeted by Company

Data Modeling × Conceptual Level

12

Practice2: Data Modeling for the [UWindsor](#) at Conceptual Level

Entities: ?

Attributes: ?

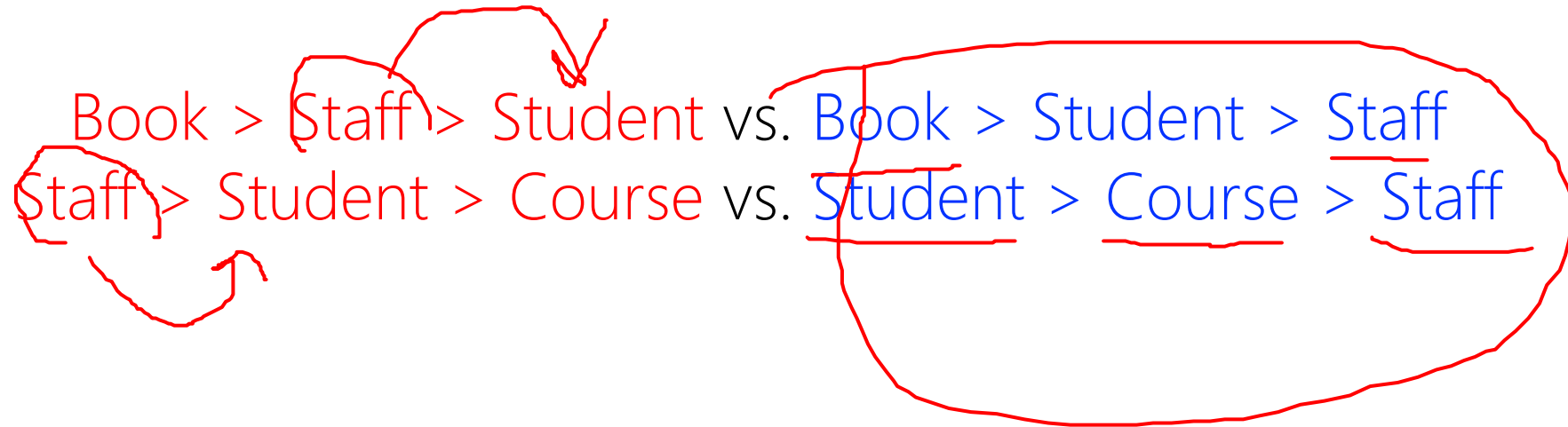
Relationships: ?

Data Modeling × Conceptual Level

13

Priorities: Time & Money

Book > ~~Staff~~ > Student vs. Book > Student > Staff
~~Staff~~ > Student > Course vs. Student > Course > Staff



2019: A Data Odyssey × Real World

14

Data Modeling: Real World Entity

Conceptual Level | Logical Level | Physical Level

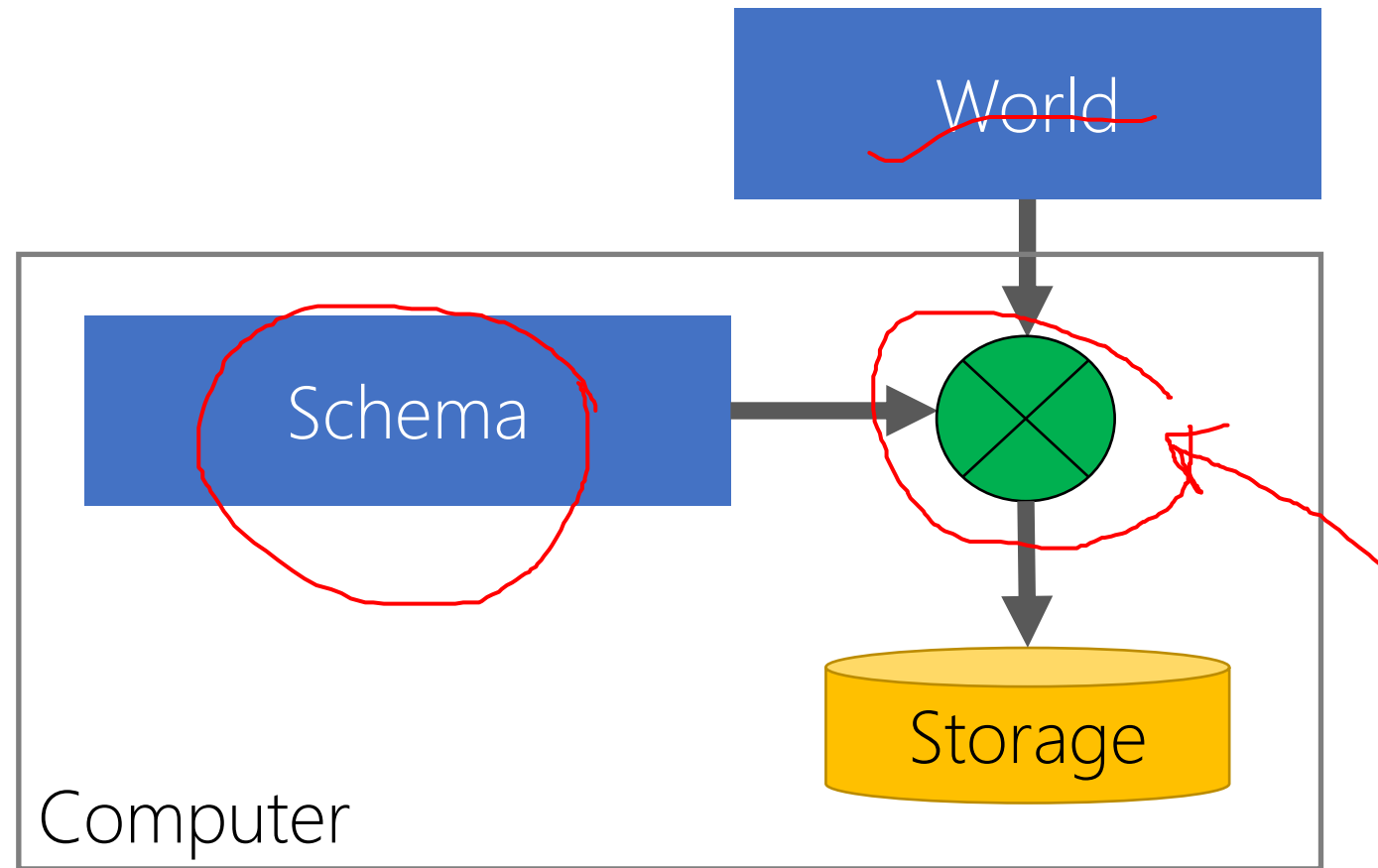
Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Computable Entity

Data Modeling × Logical Level

15



Data Modeling × Logical Level

16



How

1. Instances of entities
 2. Their values of attributes
 3. Relationships
- should be represented.

Data Modeling × Logical Level



17

Tabular!

Movie Schema	Title (string)	Genre (string)	Language (string)	RunningTime (decimal)	Poster (byte[])
Movie Instance 1	<i>2001: A Space Odyssey</i>	<i>Sci-fi</i>	<i>English</i>	142	
Movie Instance 2	<i>Rosemary's Baby</i>	<i>Horror</i>	<i>English</i>	136	



Data Modeling × Logical Level

18

Director Schema	Name (<code>char[]</code>)	DateOfBirth (<code>int</code>)	PlaceOfBirth (<code>char[]</code>)	Photo (<code>char[]</code>)
Director Instance 1	<i>Stanley Kubrick</i>	<i>1949</i>	<i>USA</i>	
Director Instance 2	<i>Roman Polanski</i>	<i>1933</i>	<i>France</i>	



Data Modeling × Logical Level



19



Company Schema	Name (<code>char[]</code>)	Address (<code>char[]</code>)	DateOfEstablishment (<code>int</code>)	Logo (<code>char[]</code>)
Company Instance 1	<i>Metro-Goldwyn-Mayer Studios Inc.</i>	<i>Beverly Hills, CA, USA</i>	1924	
Company Instance 2	<i>Warner Bros</i>	NULL	1923	

Data Modeling × Logical Level

20

Movie Schema	Title (char[])	Genre (char[])	Language (char[])	RunningTime (int)	Poster (char[])
Movie Instance 1	<i>2001: A Space Odyssey</i>	<i>Sci-fi</i>	<i>English</i>	142	
Movie Instance 2	<i>Rosemary's Baby</i>	<i>Horror</i>	<i>English</i>	136	

Director Schema	Name (char[])	DateOfBirth (int)	PlaceOfBirth (char[])	Photo (char[])
Director Instance 1	<i>Stanley Kubrick</i>	1949	USA	
Director Instance 2	<i>Roman Polanski</i>	1933	France	

Company Schema	Name (char[])	Address (char[])	DateOfEstablishment (int)	Logo (char[])
Company Instance 1	<i>Metro-Goldwyn-Mayer Studios Inc.</i>	<i>Beverly Hills, CA, USA</i>	1924	
Company Instance 2	<i>Warner Bros</i>	NULL	1923	

Data Modeling × Logical Level

21

Movie Schema	Title (string)	Genre (string)	Language (string)	RunningTime (decimal)	Poster (byte[])
Movie Instance 1	2001: A Space Odyssey	Sci-fi	English	142	
Movie Instance 2	Rosemary's Baby	Horror	English	136	

No Relationship!

Movie × Director × Company

Director Schema	Name (string)	Birth (date)	Death (date)	Photo (byte[])
Director Instance 1	Stanley Kubrick	1949	USA	





Conceptual level identifies the entities, attributes, and relationships only.

Does not say how to represent them!

Company Schema	Name (string)	Address (string)	Establishment (date)	Logo (byte[])
Company Instance 1	Metro-Goldwyn-Mayer Studios Inc.	Beverly Hills, CA, USA	1924	
Company Instance 2	Warner Bros	null	1923	







Data Modeling × Logical Level

22

Movie Schema	Title (char[])	Genre (char[])	Language (char[])	RunningTime (int)	Poster (char[])	DirectedBy (Director Schema)			
						Name (char[])	DateOfBirth (int)	PlaceOfBirth (char[])	Photo (char[])
Movie Instance 1	2001: A Space Odyssey	Sci-fi	English	142		Stanley Kubrick	1949	USA	
Movie Instance 2	Rosemary's Baby	Horror	English	136		Roman Polanski	1933	France	

Data Modeling × Logical Level

23

Movie Schema	Title (char[])	Genre (char[])	Language (char[])	RunningTime (int)	Poster (char[])	DirectedBy (Director Schema)			
						Name (char[])	DateOfBirth (int)	PlaceOfBirth (char[])	Photo (char[])
Movie Instance 1	2001: A Space Odyssey	Sci-fi	English	142		Stanley Kubrick	1949	USA	
Movie Instance 2	Rosemary's Baby	Horror	English	136		Roman Polanski	1933	France	
Movie Instance 3	A Clockwork Orange	Sci-fi	English	136		Stanley Kubrick	1949	USA	







Data Modeling × Logical Level

24

Redundancy! Wasting space. Not important tho!

Data Modeling × Logical Level

25

Movie Schema	Title (char[])	Genre (char[])	Language (char[])	RunningTime (int)	Poster (char[])	DirectedBy (Director Schema)			
						Name (char[])	DateOfBirth (int)	PlaceOfBirth (char[])	Photo (char[])
Movie Instance 1	2001: A Space Odyssey	Sci-fi	English	142		Stanley Kubrick	1949	USA	
Movie Instance 2	Rosemary's Baby	Horror	English	136		Roman Polanski	1933	France	
Movie Instance 3	A Clockwork Orange	Sci-fi	English	136		Stanley Kubrick	1949	America	







Data Modeling × Logical Level

26

Inconsistency! Incoherent pieces of information.

Data Modeling × Logical Level

27

Movie Schema	Title (char[])	Genre (char[])	Language (char[])	RunningTime (int)	Poster (char[])	DirectedBy (Director Schema)			
						Name (char[])	DateOfBirth (int)	PlaceOfBirth (char[])	Photo (char[])
Movie Instance 1	2001: A Space Odyssey	Sci-fi	English	142		Stanley Kubrick	1949	America	
Movie Instance 2	Rosemary's Baby	Horror	English	136		Roman Polanski	1933	France	
Movie Instance 3	A Clockwork Orange	Sci-fi	English	136		Stanley Kubrick	1949	America	

Data Modeling × Logical Level

28

Speed! Cascading of updates/deletions.

Data Modeling × Logical Level

29

Not All Data Models at Logical Level are Efficient.

Not All Logical Models are Efficient.

Not All Data Models are Efficient.

Data Modeling × Logical Level

30

There are other **representations** as well.

There are other **logical models** as well.

There are other **data models at logical level** as well.

~1960: Object Oriented

~~~1969: Relational: Mathematical Relationships~~

~~~1996: XML~~

Data Modeling × Logical Level

31

There are other representations as well.

There are other logical models as well.

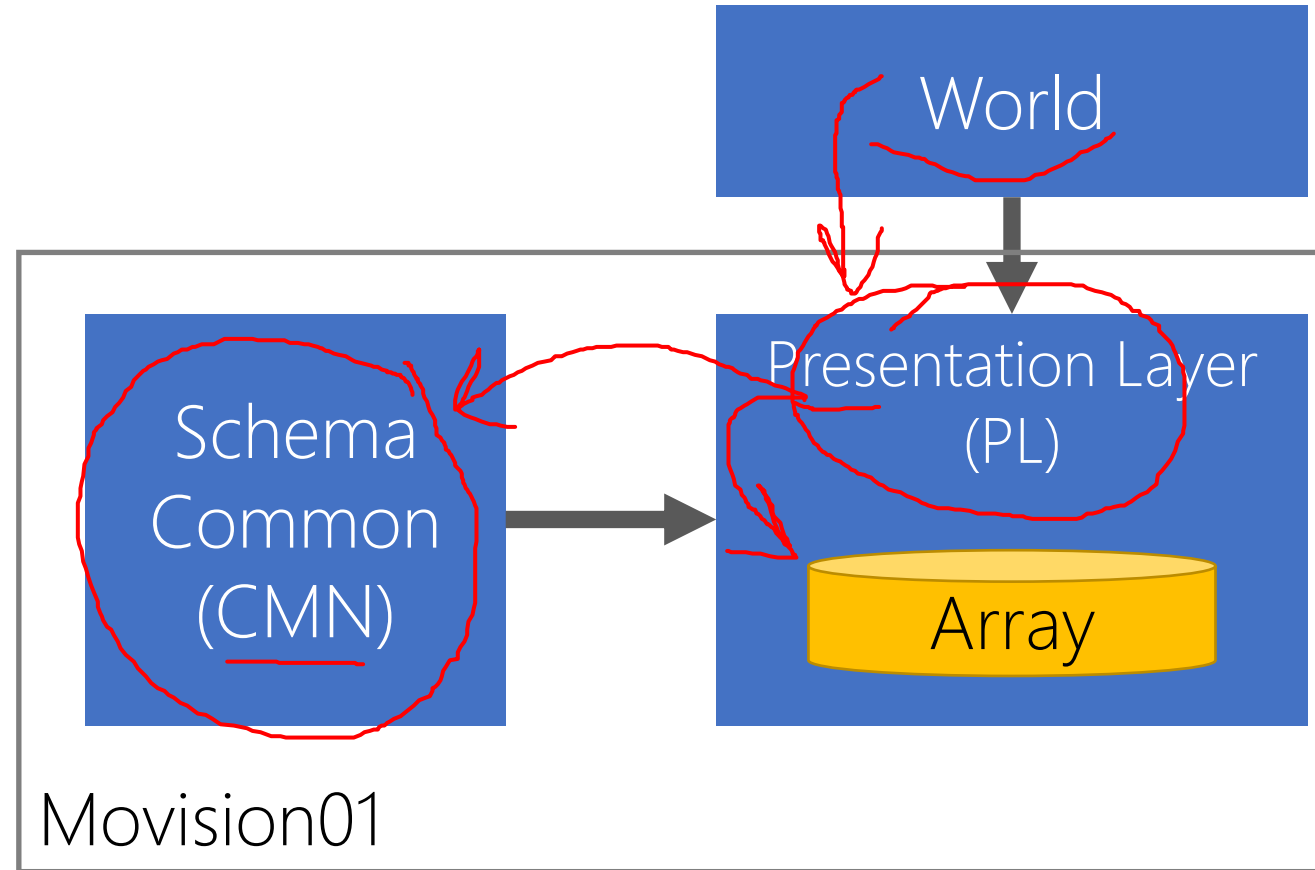
There are other data models at logical level as well.

~1960: Object Oriented

COMP-2120: Object-Oriented Programming Using Java

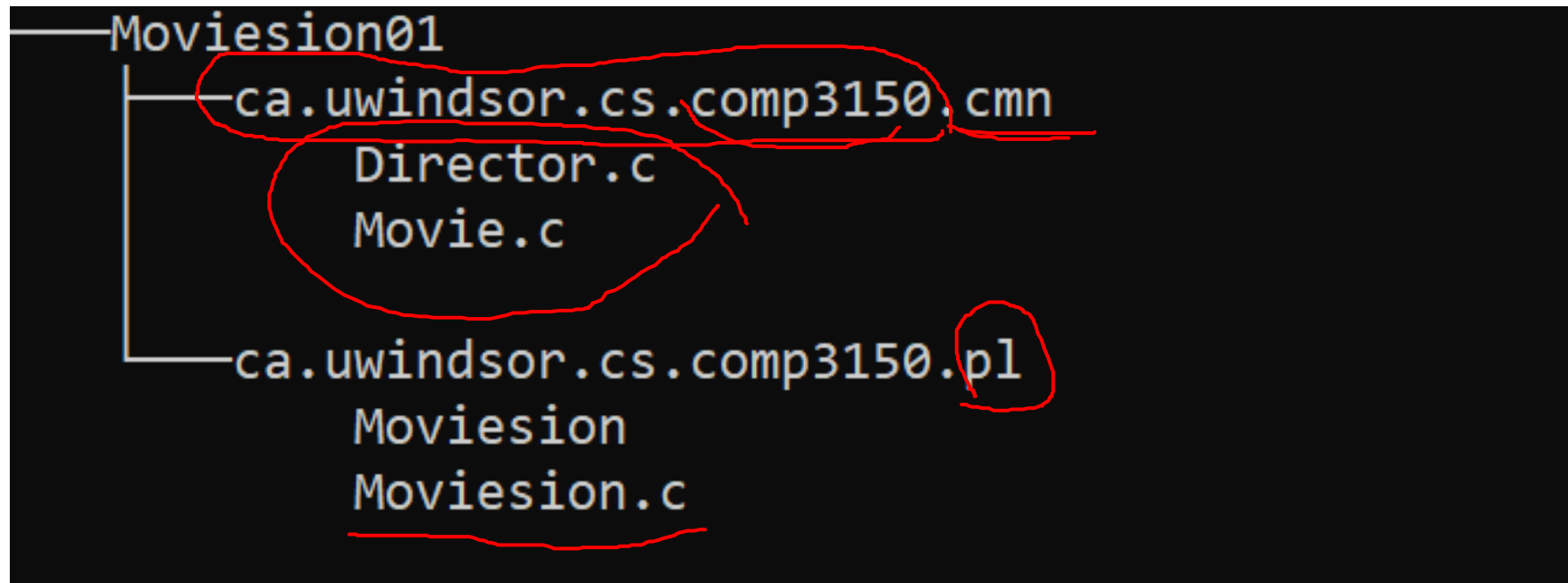
Logical Level × Object Oriented

32



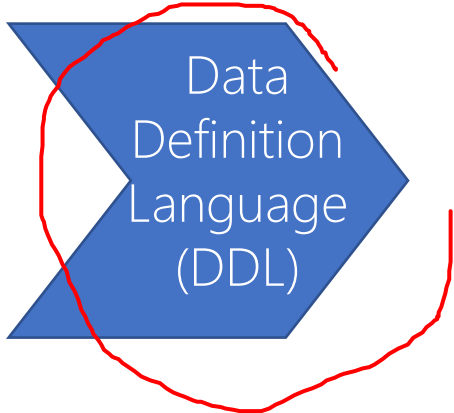
Logical Level × Object Oriented

33



Logical Level × Object Oriented

34


| Conceptual Model | | Logical Model by OO |
|---|---|--|
| Movie <u>Schema</u>
Title
Genre
Language
RunningTime
Poster
... |  | <u>struct</u> Movie {
<u>char</u> <u>title</u> [100];
char <u>genre</u> [10];
char language[2];
int runningTime;
char poster[1024];
} |

Logical Level × Object Oriented

35

Conceptual Model

Movie *Instance*

| Title | Genre | Language | RunningTime | Poster |
|-----------------------|--------|----------|-------------|---|
| 2001: A Space Odyssey | Sci-fi | English | 142 |  |

Logical Model by OO




Movie *Object*

```
struct Movie newMovie;  
strcpy(newMovie.name, "2001: A Space Odyssey");  
strcpy(newMovie.genre, "Sci-fi");  
strcpy(newMovie.language, "English");  
newMovie.runningTime = 142;  
newMovie.poster = ?;
```



Logical Level × Object Oriented

36

| Conceptual Model | Transformation | Logical Model by OO |
|--|----------------|---|
| Data Schema  | DDL | Class Struct  |
| Data Instance  | DML | Object |

Data Definition Language (DDL): Commands for setting up schema of data

Data Manipulation Language (DML): Commands to manipulate data, also called "query language"

Logical Level × Object Oriented

37

Pros

Time | Random Access Memory (RAM) | Fast

Variety of Data Structures

Array, Linked List, Stack, Queue, Tree, HashTable, ...

```
struct Movie movies[100]
```

Easy DML in Memory

```
SELECT: Movie m = movies[int index];  
INSERT: movies[int index] = m;  
DELETE: movies[int index] = NULL;  
UPDATE: movies[int index].title = "new title";
```

Any Computation

Sorting, Searching, ...

Cons

Space Far Too Small

Expensive

NOT DURABLE

Volatility | Transient | Short-term Retention

Portability

2019: A Data Odyssey × Real World

38

Data Modeling: Real World Entity

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | Physical Level

Conceptual Level | Logical Level | **Physical Level**

Conceptual Level | Logical Level | Computable Entity

Data Modeling × Physical Level

39

1. Where and how are the [data instances](#) stored in [physical storage](#).

Data Modeling × Physical Level

40

We already have seen one physical storage, haven't we?

Data Modeling × Physical Level

41

There are other [physical storages](#) as well.

There are other [physical models](#) as well.

There are other [data models at physical level](#) as well.

File

Database

Cloud

Data Modeling × Physical Level

42

There are other physical storages as well.

There are other physical models as well.

There are other data models at physical level as well.

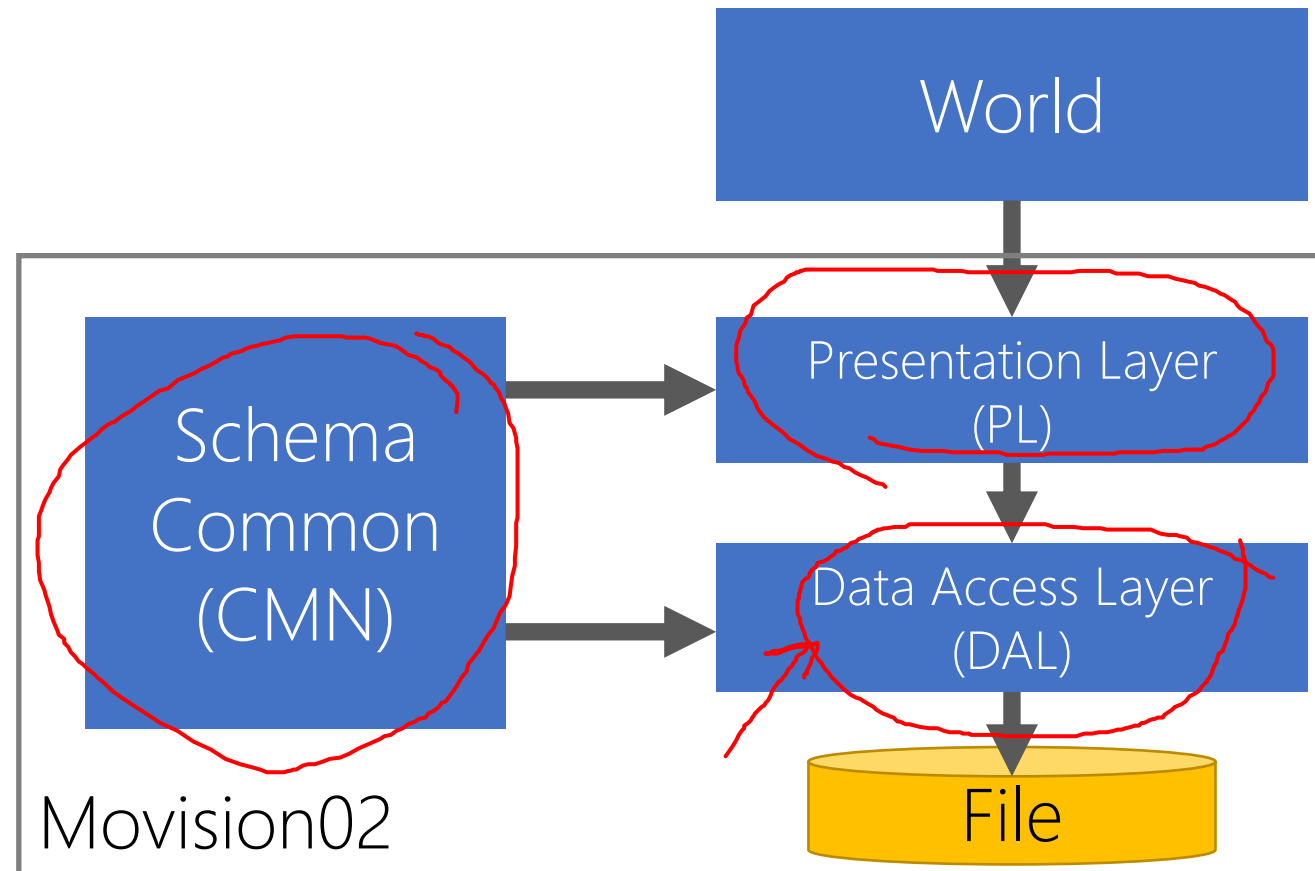
File

Database

Cloud

Physical Level × File

43



JUST THE WAY
YOU ARE



Play (k)

0:24 / 3:56



Physical Level × File

45


1,Dunkirk,Christopher Nolan
2,Get Out,Jordan Peele
3,Lady Bird,Greta Gerwig
4,Phantom Thread,Paul Thomas Anderson
5,The Shape of Water,Guillermo del Toro

2D vs. 1D

1,Dunkirk,Christopher Nolan\n2,Get Out,Jordan Peele\n3,Lady Bird,Greta Gerwig\n..

Physical Level × File

46

1,Dunkirk,Christopher Nolan
2,Get Out,Jordan Peele
3,Lady Bird,Greta Gerwig
4,Phantom Thread,Paul Thomas Anderson
5,The Shape of Water,Guillermo del Toro

2D vs. 1D

1,Dunkirk,Christopher Nolan\n2,Get Out,Jordan Peele\n3,Lady Bird,Greta Gerwig\n..

Physical Level × File × .csv

47

Comma Separated **Values**

1,Dunkirk,Christopher Nolan\n2,Get Out,Jordan Peele\n3,Lady Bird,Greta Gerwig\n..

Line Separated **Records**

Physical Level × File × .tsv

48

Tab Separated Values

1\tDunkirk\tChristopher Nolan\r\n2\tGet Out\tJordan Peele\r\n3\tLady Bird\tGreta Gerwig\r\n..

Line Separated Records

Physical Level × File × DDL

49

DDL = File Structure

Text csv, tsv, json, xml, ...

Binary pdf, jpg, pkl, ...

Portability: *Love me, Love me not!*

If only self-explanatory, *we can say it's portable.*

Physical Level × File × DML

50

Not Easy DML

DELETE

Strategy#1: Create a new file without the deleted record.

Strategy#2: Shift all following over the deleted record.

Strategy#3: Logical Deletion | Invalidate Record

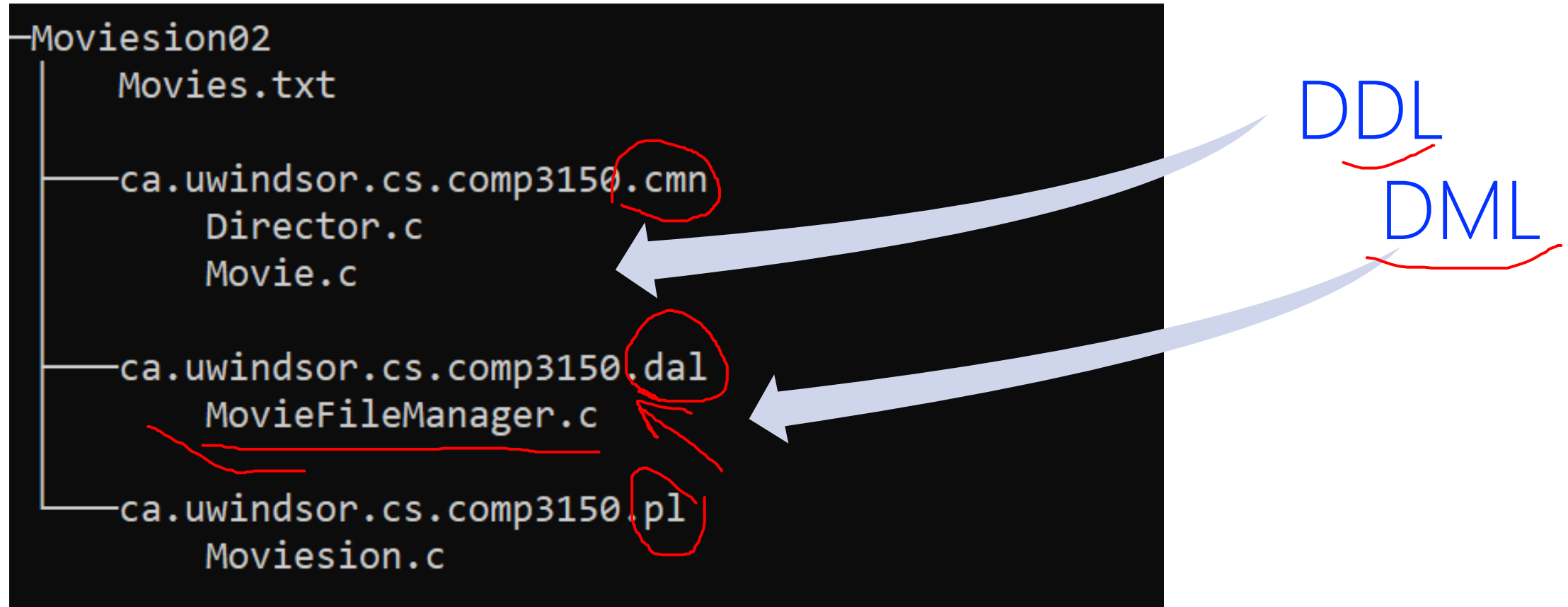
UPDATE

Strategy#1: Replace bytes of old value with bytes of new value

Strategy#2: Delete Record + Insert Record

Physical Level × File × DAL

51



Physical Level × File

52

| Conceptual Model | Logical Model by OO | Physical Model by File |
|------------------|---------------------|---|
| Data Schema | Class Struct | CSV File
JSON File
<u>XML File: XSD</u> |
| Data Instance | Object | Record |

2019: A Data Odyssey × Real World

53

Real World Entity

Conceptual Level | Instance Level | Physical Level

Conceptual Level | Logical Level | Object Level

Conceptual Level | Logical Level | Physical Level | Record

Conceptual Level | Logical Level | Computable Entity

Physical Level × File

54

Pros

Space

DURABLE

Persistency | Long-term Retention

Portable (self-explanatory)

Cons

Time

Hard Disk Drive (HDD), Sequential Access, Electromechanical
Solid State Drive (SSD), Random Access, Expensive

Not Easy DML

DELETE | UPDATE

Not Portable

No ACID Properties