Joining Tables

Dataforståelse 5

NIFR 2022

Joining tables

Dataforståelse 5

- Finishing Data Cleaning project
- Primary Key / Primærnøgle
- Foreign Key / Fremmednøgle
- Joining Tables

Porteføljeopgave 1

Housekeeping

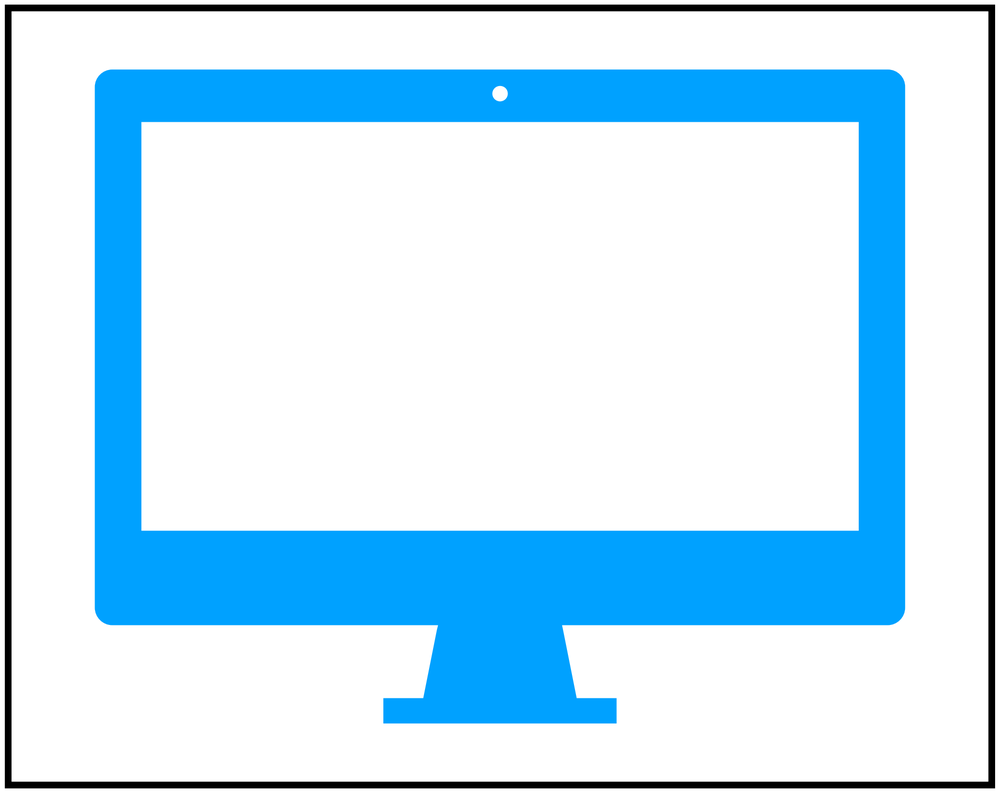
- Analysere et komplekst dataset
- Læse det ind i en database
- Besvare spørgsmål vedrørende datasettet
  - Udtræk af data fra datasættet

Observation: Datatyper

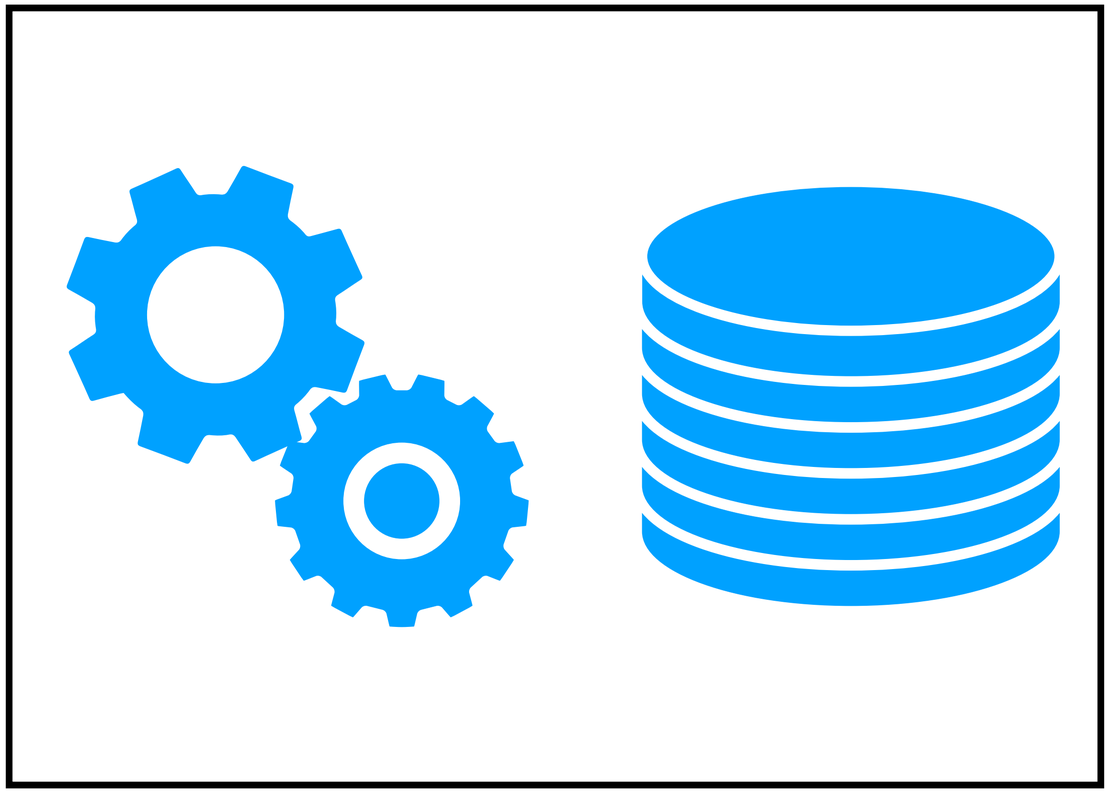
VARCHAR = “NICKLAS”
INT / NUMERIC = 123

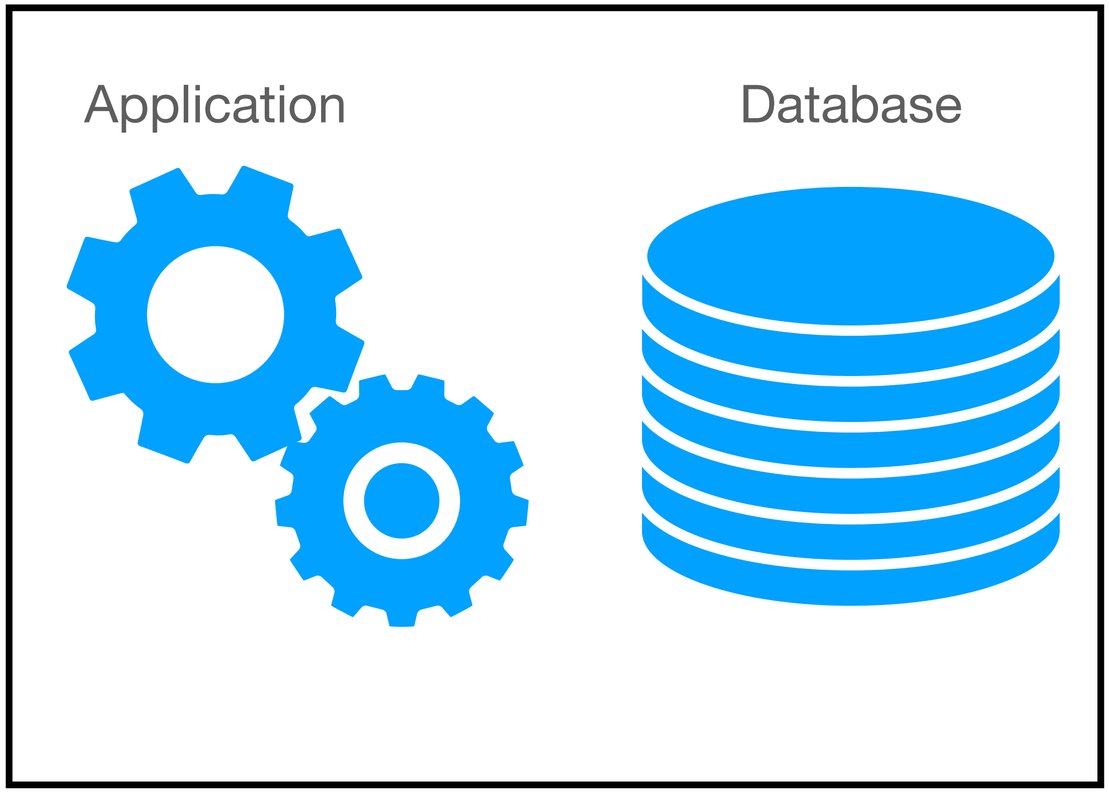
Recap

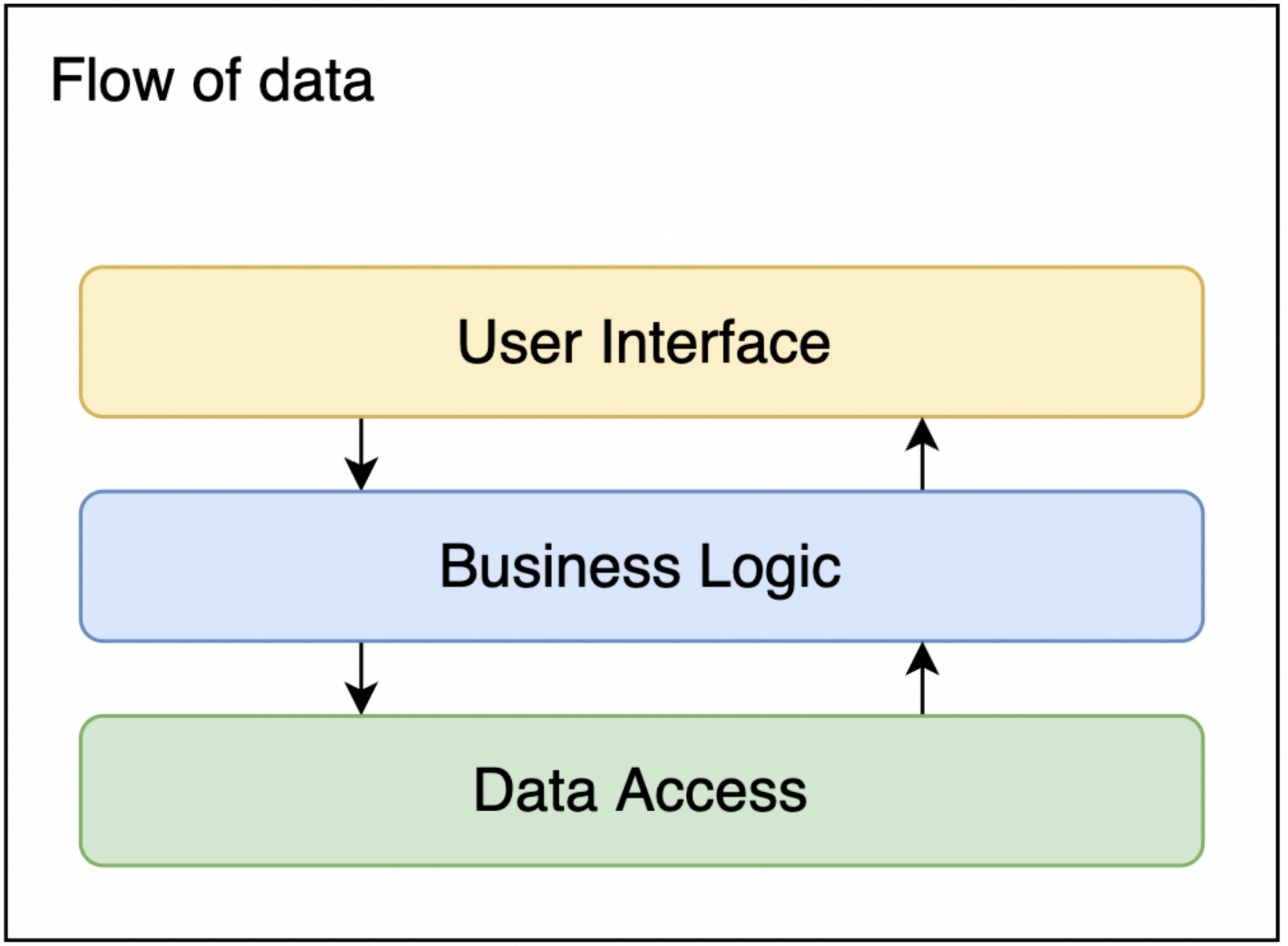
Front-end



Backend




Application
Database


Flow of data
User Interface
Business Logic
Data Access

Datas form er udtrykt gennem type

Data Types

Measurement Scales

There are four types of measurement scales used in statistics: nominal, ordinal, interval and ratio. Each scale has different properties and uses.

QUALITYGURUS

Nominal

The category labels are not ordered, so it doesn't matter which number comes first.

Ordinal

Iin the ordinal scale of data, there is an order. However, the difference between them can not be quantified.

Interval

In the interval scale, we do have an order (just like ordinal data), and we can find the exact difference between the two values.

Ratio

The ratio scale has all the features of the Interval scale, and in addition, there is an absolute or true zero as well.

Multi Dimensional Nature of Data Quality

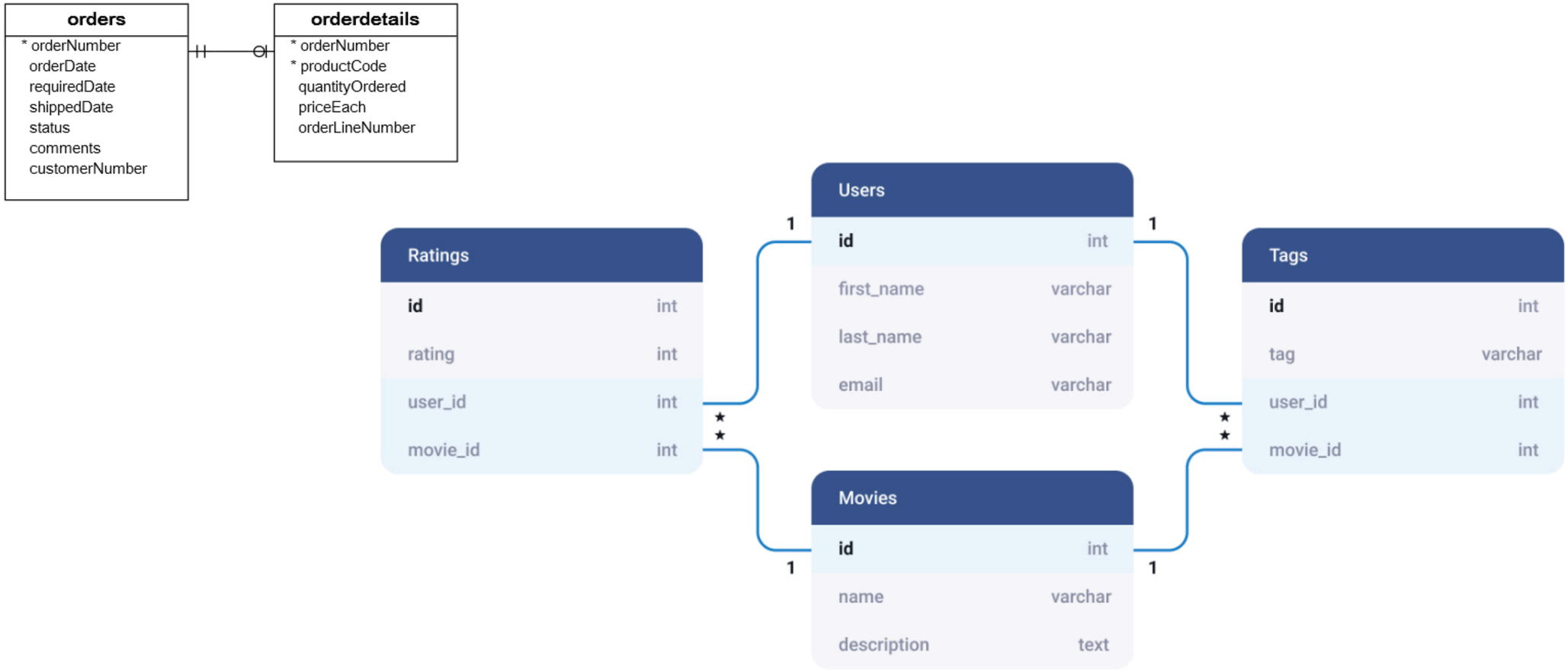
Parameters

- Accuracy
- Completeness
- Time-Related Dimensions
  - Currency, Timeliness & Volatility
- Consistency

```
SELECT [Columns]
FROM [Table Name];
```

```
SELECT [Columns]
FROM [Table Name]
WHERE [Filter Expression]
```

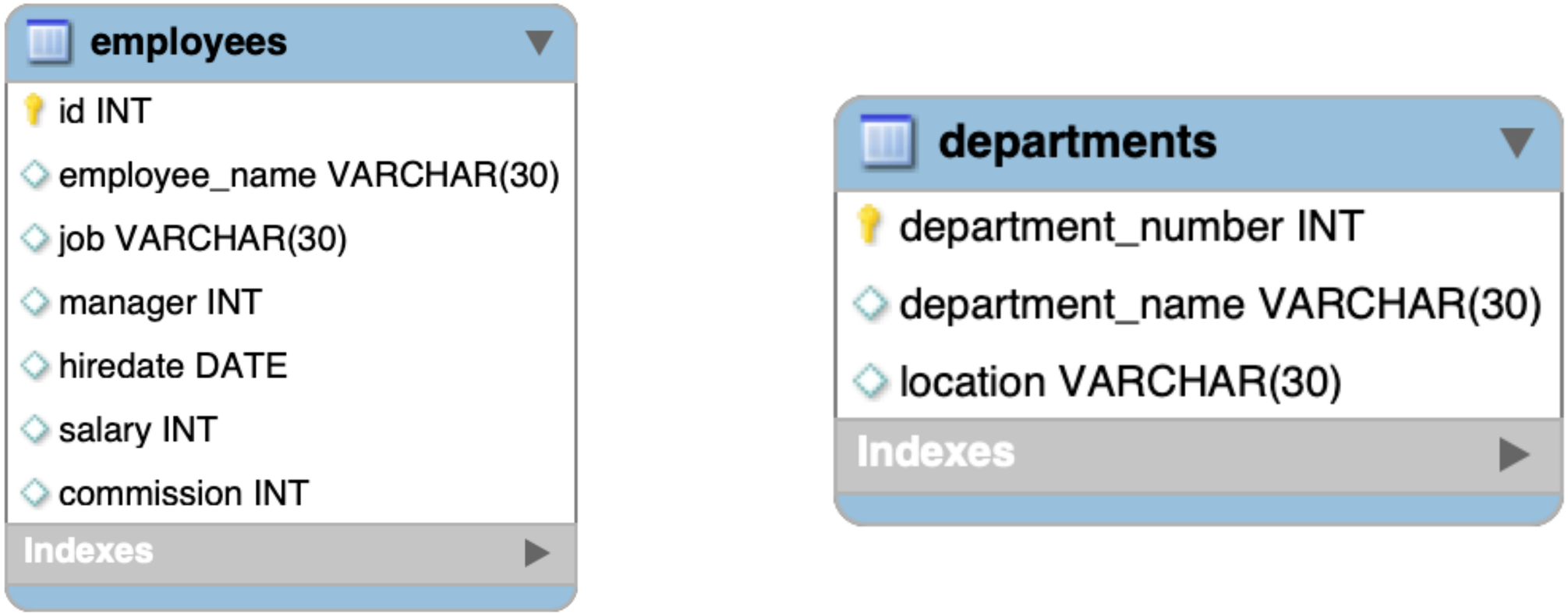
The Relational Database Model?


orders
* orderNumber
orderDate
requiredDate
shippedDate
status
comments
customerNumber
orderdetails
* orderNumber
* productCode
quantityOrdered
priceEach
orderLineNumber
Users
id int
first_name varchar
last_name varchar
email varchar
Ratings
id int
rating int
user_id int
movie_id int
Movies
id int
name varchar
description text
Tags
id int
tag varchar
user_id int
movie_id int
1
1
1
1
★
★
★
★

Case: Employees & Departments

(Enhanced) Entity Relationship Diagram

EER



How are these related?

```
SELECT *
FROM departments;
```

department_number	department_name	location
10	ACCOUNT…	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIO…	BOSTON

```
SELECT employee_name, job, department_number
FROM employees;
```

employee_name	job	department_number
SMITH	CLERK	20
ALLEN	SALESMAN	30
WARD	SALESMAN	30
JONES	MANAGER	20

End Goal: Overview of who works at what department

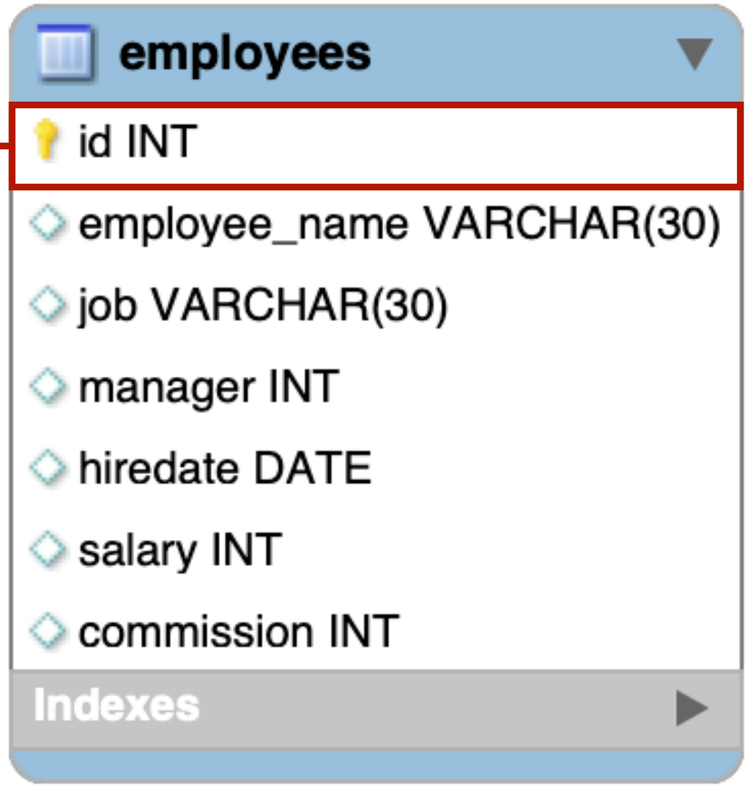
Relationship between employee and department

department_number	department_name	location
10	ACCOUNT…	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIO…	BOSTON

employee_name	job	department_number
SMITH	CLERK	20
ALLEN	SALESMAN	30
WARD	SALESMAN	30
JONES	MANAGER	20

Primary Key

department_number	department_name	location
10	ACCOUNT…	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIO…	BOSTON



A Primary Key: A unique field that identifies a row

End Goal: Who works at what department?

```
SELECT *
FROM departments
INNER JOIN employees
ON departments.department_number = employees.department_number;
```

Output:

department_number	department_name	location	id	employee_name	job	manager	hiredate	salary	commission	department_number
10	ACCOUNTING	NEW YORK	7782	CLARK	MANAGER	7839	09/06/1981	2450	NULL	10
10	ACCOUNTING	NEW YORK	7839	KING	PRESIDENT	NULL	17/11/1981	5000	NULL	10
10	ACCOUNTING	NEW YORK	7934	MILLER	CLERK	7782	23/01/1982	1300	NULL	10
20	RESEARCH	DALLAS	7369	SMITH	CLERK	7902	17/12/1980	800	NULL	20
20	RESEARCH	DALLAS	7566	JONES	MANAGER	7839	02/04/1981	2975	NULL	20
20	RESEARCH	DALLAS	7788	SCOTT	ANALYST	7566	19/04/1987	3000	NULL	20
20	RESEARCH	DALLAS	7876	ADAMS	CLERK	7788	23/05/1987	1100	NULL	20
20	RESEARCH	DALLAS	7902	FORD	ANALYST	7566	03/12/1981	3000	NULL	20
30	SALES	CHICAGO	7499	ALLEN	SALESMAN	7698	20/02/1981	1600	300	30
30	SALES	CHICAGO	7521	WARD	SALESMAN	7698	22/02/1981	1250	500	30
30	SALES	CHICAGO	7654	MARTIN	SALESMAN	7698	28/09/1981	1250	1400	30
30	SALES	CHICAGO	7698	BLAKE	MANAGER	7839	01/05/1981	2850	NULL	30
30	SALES	CHICAGO	7844	TURNER	SALESMAN	7698	08/09/1981	1500	0	30
30	SALES	CHICAGO	7900	JAMES	CLERK	7698	03/12/1981	950	NULL	30

End Goal: Overview of who works at what department

```
SELECT *
FROM departments
INNER JOIN employees
ON departments.department_number = employees.department_number;
```

End Goal: Who works in the sales department?

```
SELECT *
FROM departments
INNER JOIN employees
ON departments.department_number = employees.department_number
WHERE departments.department_name = "SALES";
```

Exercises 1

https://
pairs.austincodingacademy.com/

Pair Programming: Exercise 2