

UNIVERSITETET
I OSLO

IN2090
Databases and data modelling

Today's topics:

- Relational model
- Relational algebra

Relational model

Relational model

- **Data model**
  Set of concepts to describe the structure of a database
- **Relational model**
  The database can be considered as a collection of tables

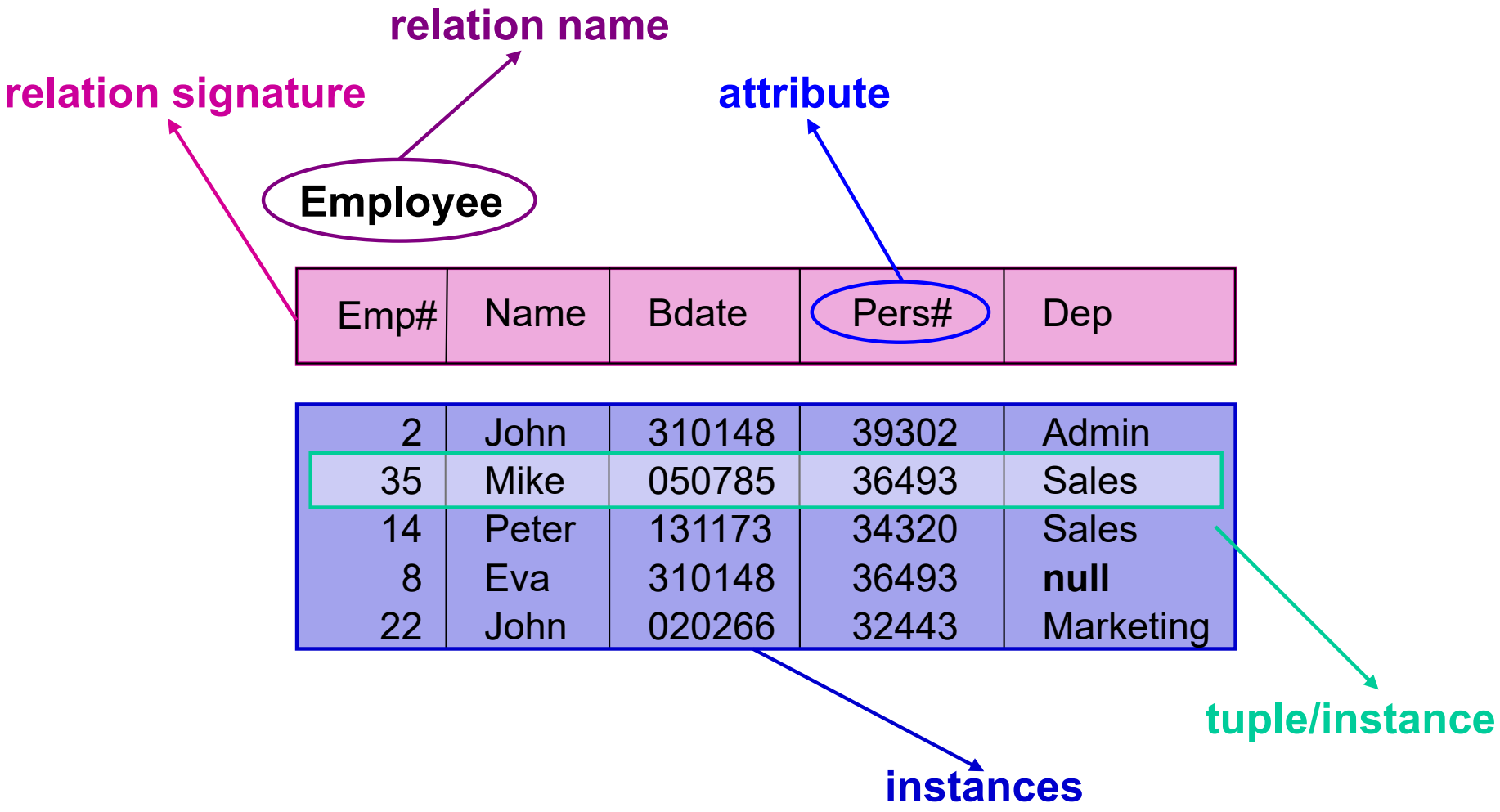
Relations and relational databases

Employee

Emp#	Name	Bdate	Pers#	Dep
2	John	310148	39302	Admin
35	Mike	050785	36493	Sales
14	Peter	131173	34320	Sales
8	Eva	310148	36493	**null**
22	John	020266	32443	Marketing

- **Relation**: A mathematical concept which can be interpreted as a table containing values
- **Relational database**: A set of relations
- **null** indicates that no value is stored in the position it appears

Relations – terminology



Relations – terminology

dom(Bdate) =
{six-digit numbers
with restrictions
on what numbers
are allowed}

dom(Dep) =
{Admin,
Sales
Marketing}

Employee

Emp#	Name	Bdate	Pers#	Dep
2	John	310148	39302	Admin
35	Mike	050785	36493	Sales
14	Peter	131173	34320	Sales
8	Eva	310148	36493	**null**
22	John	020266	32443	Marketing

Formal definitions

- **Domain**: A set of *atomic* values
  (The fact that the elements in a domain are atomic means that the elements themselves cannot be sets)
- **Attribute**: A name for a role played by a domain ("*column name*")
  If A is an attribute, we write dom(A) = D to express the fact that A is a role played by the domain D
- **Relation signature** $R(A_1, A_2, \ldots, A_n)$: A named set of attributes $R = \{A_1, A_2, \ldots, A_n\}$ where R is **relation's name**
  n is called relation's *degree* or *arity*

Formal definitions

- **Instance** of a relation signature $R(A_1, A_2, \ldots, A_n)$: A set $\{t_1, t_2, \ldots, t_m\}$ ("*rows*") where each t_k is an n-tuple of values from the domains of $A_1, A_2, \ldots, A_n$
  I.e., for all rows $t = \langle v_1, v_2, \ldots, v_n \rangle$, $v_i \in dom(A_i) \cup \{\textbf{null}\}$
  (Some values can be **null**)

- Notation: $t[A_i, A_j, \ldots]$
  - If t is a tuple in an instance of $R(A_1, A_2, \ldots, A_n)$ and $t = \langle v_1, v_2, \ldots, v_n \rangle$, then, e.g., $t[A_2] = \langle v_2 \rangle$ and $t[A_3, A_1, A_5] = \langle v_3, v_1, v_5 \rangle$

Formal definitions

- **Relation**: Relation signature + corresponding instances

  The relation signature is called relation's **intention**
  The set of instances is called relation's **extention**

- We often say "**relation**" even though we mean **relation signature**

Remarks

- The order of the tuples in an instance is arbitrary
- The order of the values in a tuple is basically not arbitrary (in order to make the notation easier)
- No identical tuples can exist in the same instance
- A domain can be finite or infinite
- Two attributes in a relation signature can have the same domain, but not the same name

Relational data vs. Tabular data

(mathematical constructs vs. visual constructs)

Relational Model	Tabular Data
Relation	Table
Tuple	Row
Attribute	Column
Domain	Column data type
Arity	Number of columns
Signature (i.e., intension)	Table header
Instance (i.e., extension)	All rows (except header)