

Prednáška 5

Dátové modelovanie

Obsah prednášky

- Organizácia
- Dátové modelovanie
 - Úroveň abstrakcie dátových modelov
 - Entitno Relačný diagram
- Zadanie 3

Organizácia

- Posunutie zadania 2 v dôsledku IO operácií v rámci Azure DB
 - Môžete sa vypnúť DB v rámci Azure

Motivácia

- Ako navrhnuť správne DB?
 - Môžeme vytvoriť jednu tabuľku/súbor alebo viacero tabuliek
 - Redundancia dát, možnosť porušenia integrity
 - Ako sa budú dáta meniť v čase, čo všetko chcem mať uchované
- Ako postupovať pri návrhu DB?

Dátové modelovanie

- **Dátové modelovanie**

- Proces, pri ktorom sa definujú a analyzujú požiadavky na štruktúru dát pre informačný systém
 - Výsledok - **dátový model**

- **Dátový model**

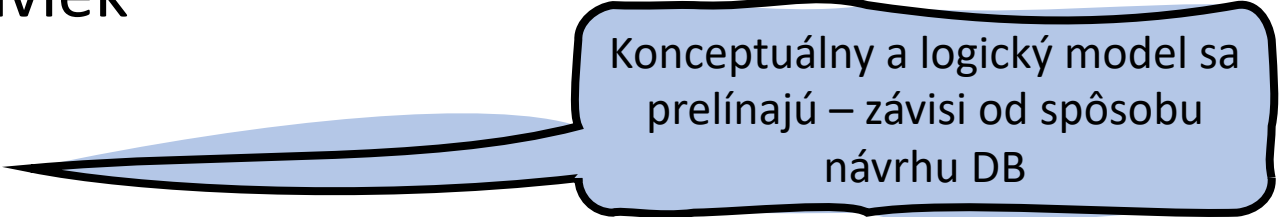
- Popisuje formát, štruktúru a vzťahy dát informačného systému

- V kontexte relačných DB

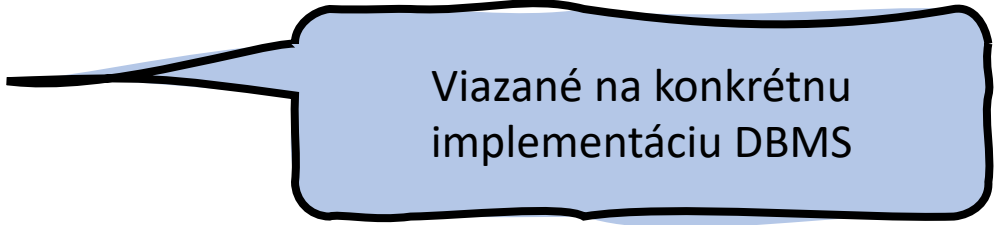
- Definovanie tabuliek, ich prepojení, obmedzení
 - Všetko čo sme si ukazovali doteraz v rámci DDL

Dizajn DB - fázy

- Analyzovanie požiadaviek
- Konceptuálny dizajn
- Logický dizajn
- Fyzický dizajn
 - Súvisí aj s optimalizáciou query processing
 - Vstupujú špecifiká DBMS



Konceptuálny a logický model sa prelínajú – závisí od spôsobu návrhu DB



Viazané na konkrétnu implementáciu DBMS

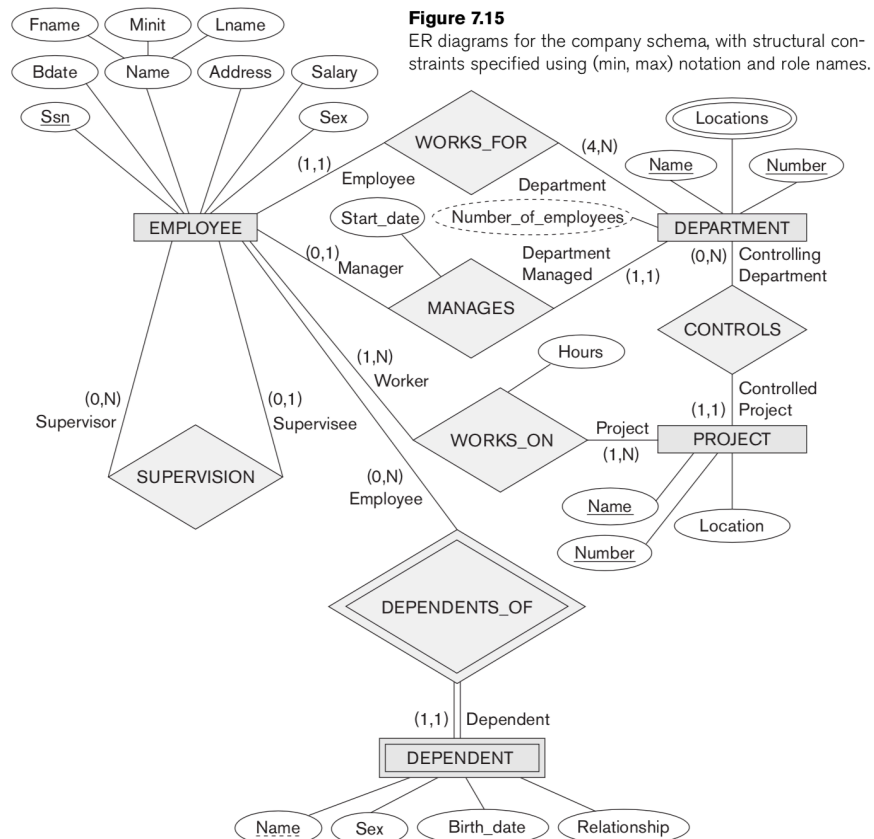
Konceptuálny model

- Možnosť reprezentovaním Entitno-Relačný (ER) modelom
 - Relation – relácia
 - Relationship – vzťah medzi reláciami
- Entitno-relačný model
 - Transformácia reálneho sveta na entity a ich vzájomné vzťahy
 - Grafické znázornenie
 - Závisí od vybranej grafickej reprezentácie
 - Ľahká transformácia na relačný model

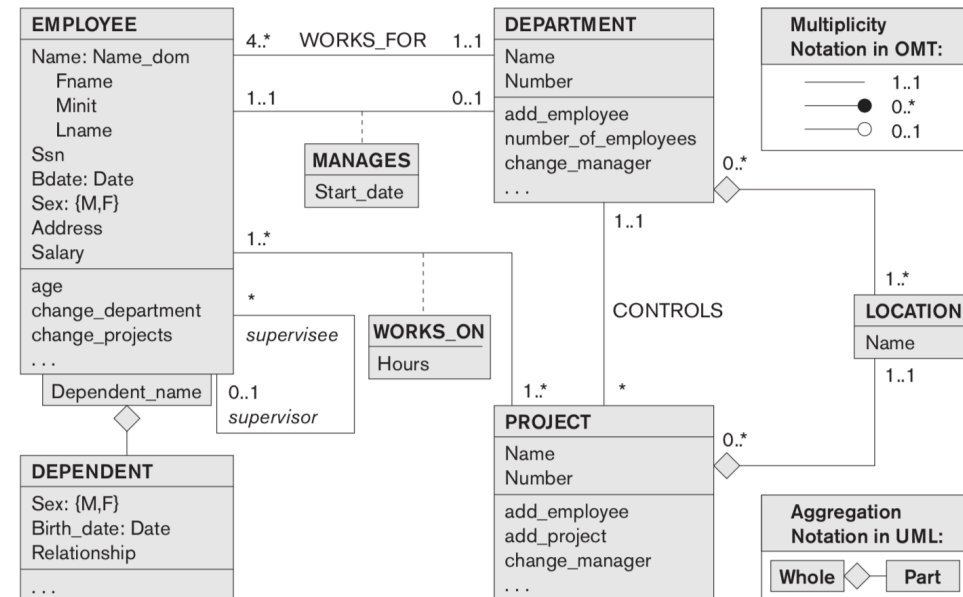
Entitno-relačný model

- Základné pojmy
 - Entita - konkrétny objekt rovnakého charakteru
 - Množina entít – zoskupenie entít rovnakého typu
 - Tabuľka/Relácia v relačnom modeli
 - Atribút
 - Vzťah
 - Vzťahová množina – matematická relácia nad množinami entít
 - Kľúčový atribút (key attribute)

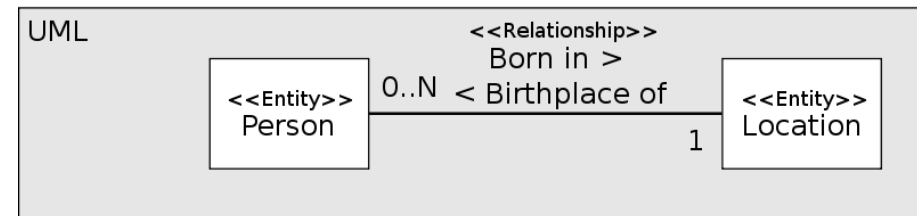
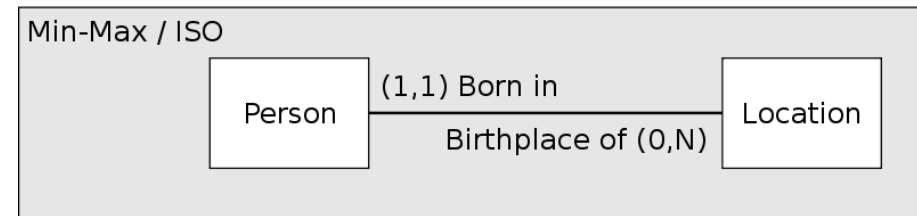
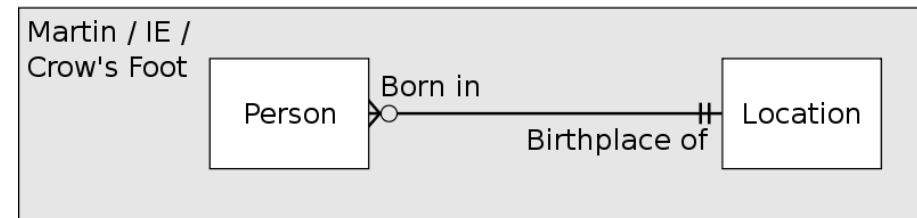
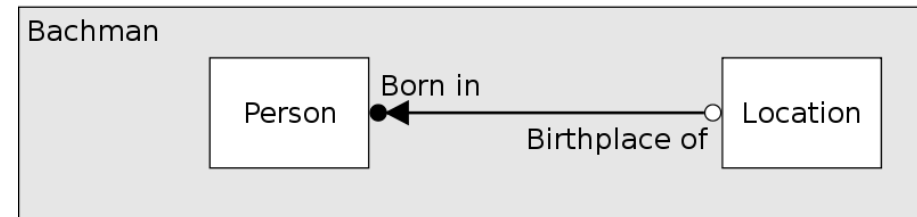
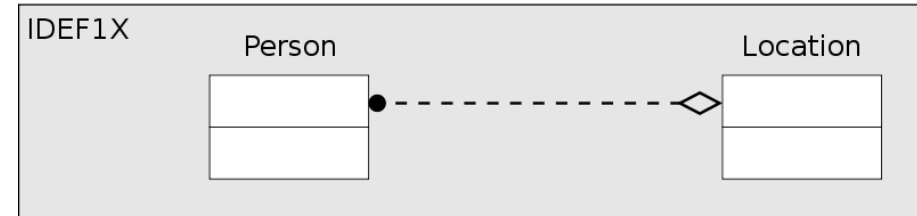
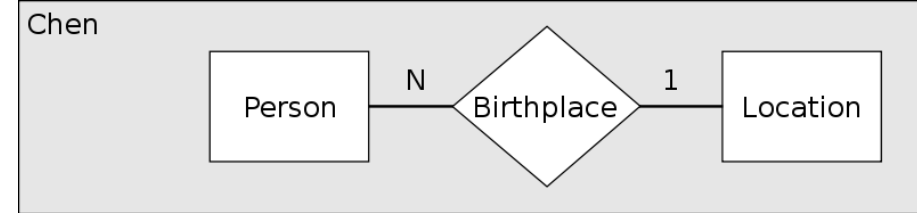
E-R model -vizualizácia



UML



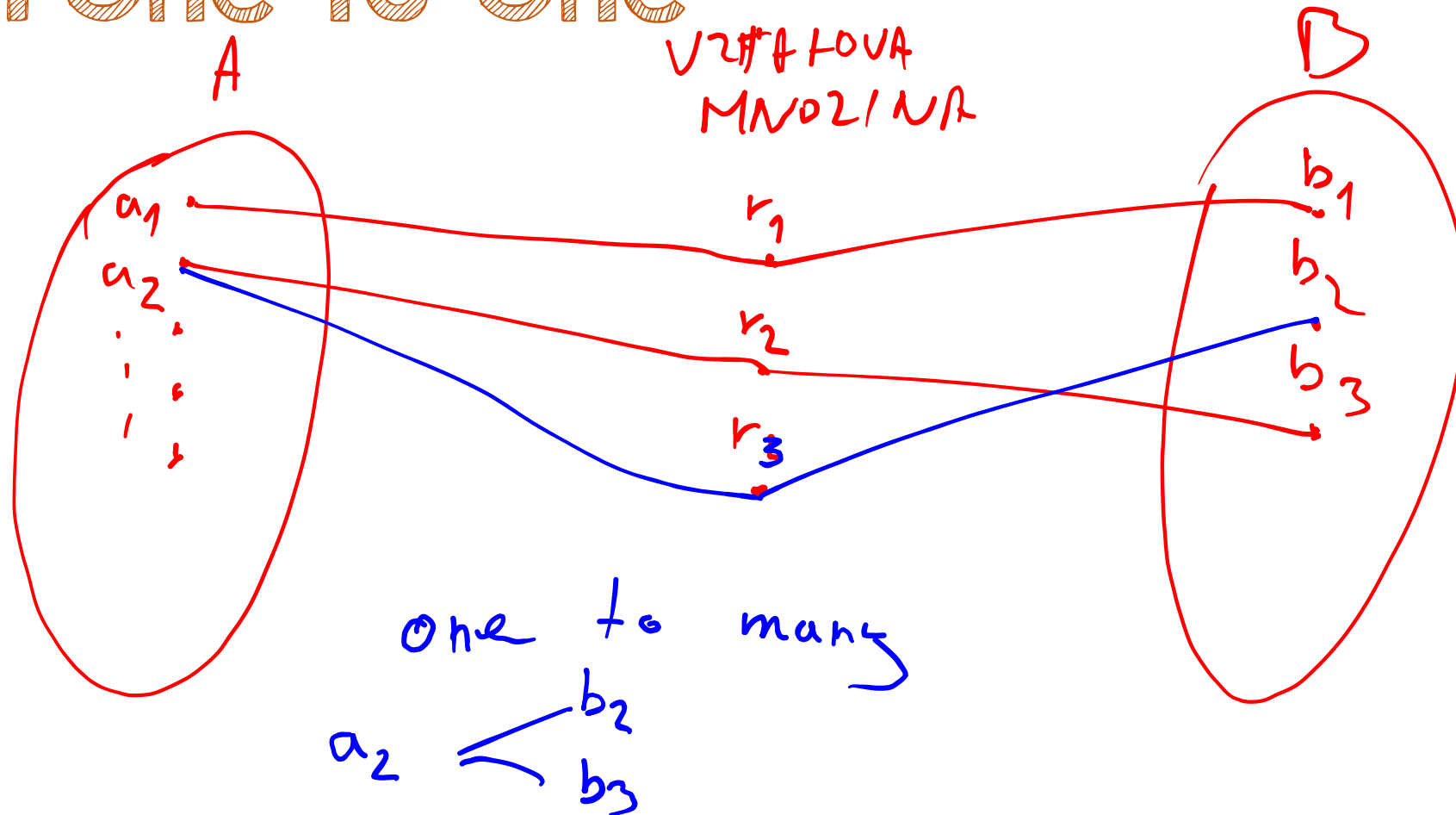
Vizualizácia



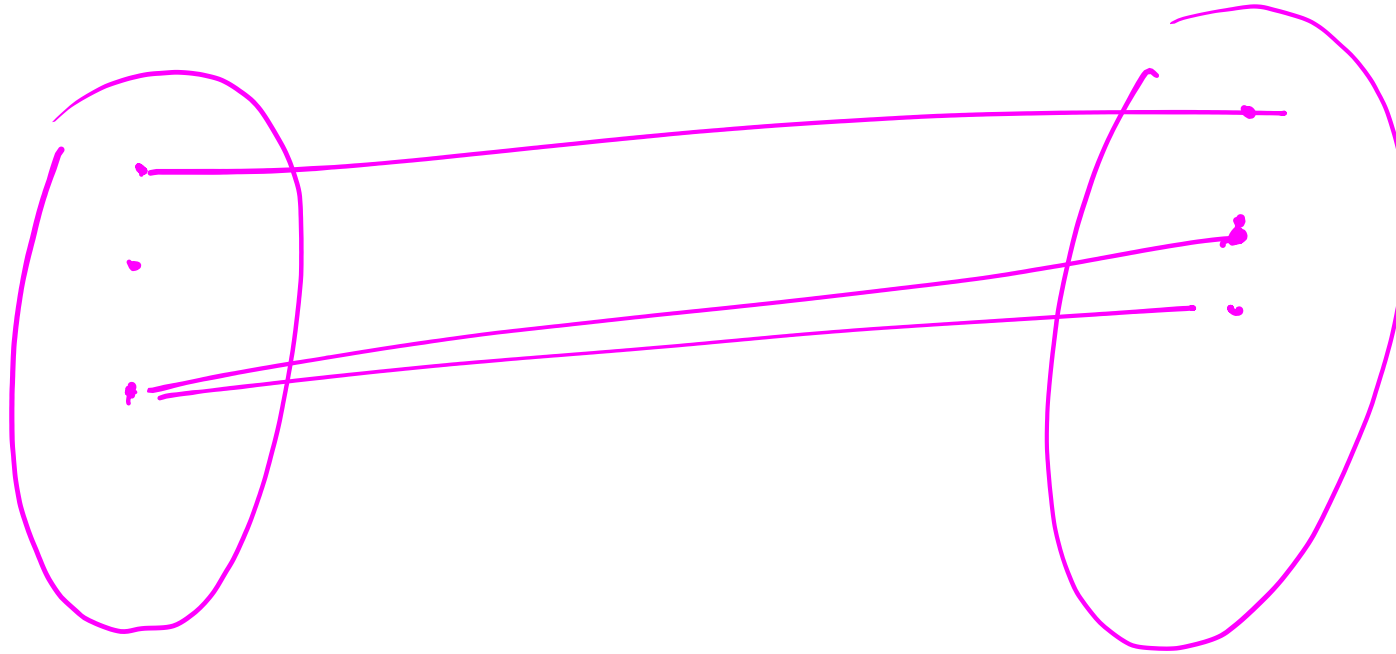
Kardinalita

- Označované aj ako násobnosť (multiplicity)
- Súvisí so vzťahmi medzi entitami
- V rámci relačného modelu súvisí s tým ako sú vytvorené referencie medzi jednotlivými tabuľkami
- Tri typy
 - **1:1 (One-to-One)**
 - prezident riadi štát
 - **1:N (One-to-Many)** alebo **N:1 (Many-to-One)**
 - oddelenie zamestnáva zamestnanca , zamestnanec je zamestnaný na oddelení
 - **M:N (Many-to-Many)**
 - zamestnanec pracuje na viacerých projektoch a na projekte môže pracovať viacero zamestnancov

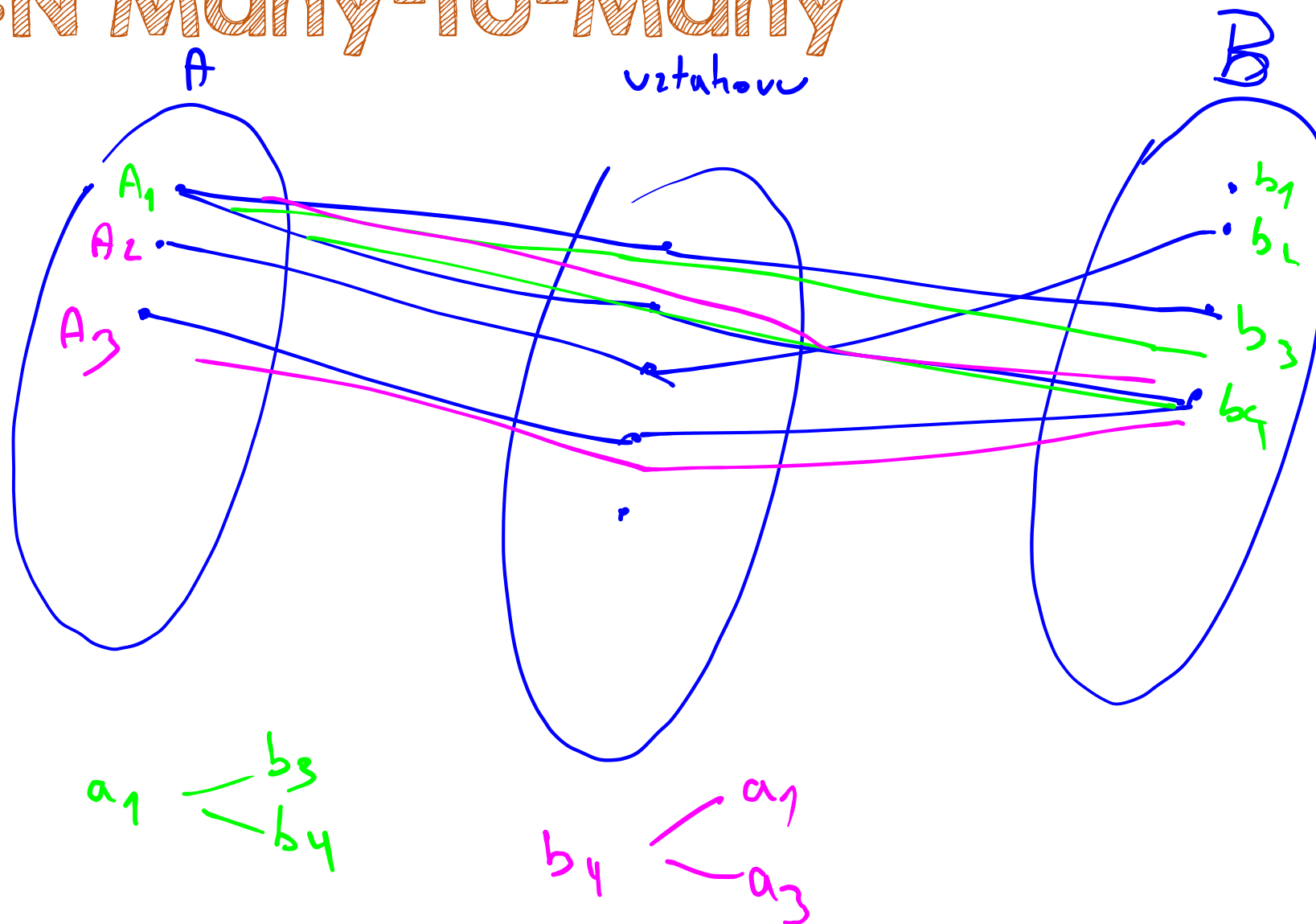
One-to-One



1:N One-to-Many

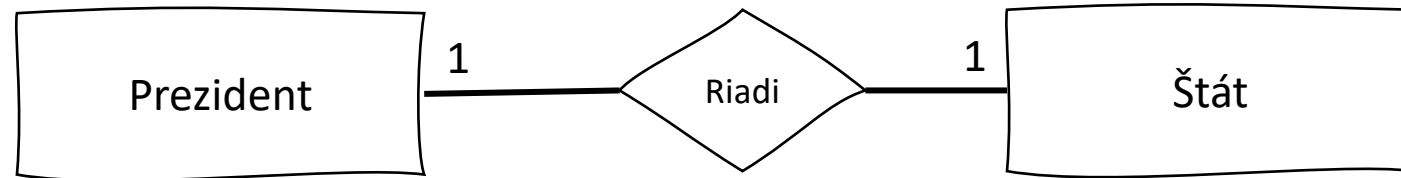


M:N Many-to-Many

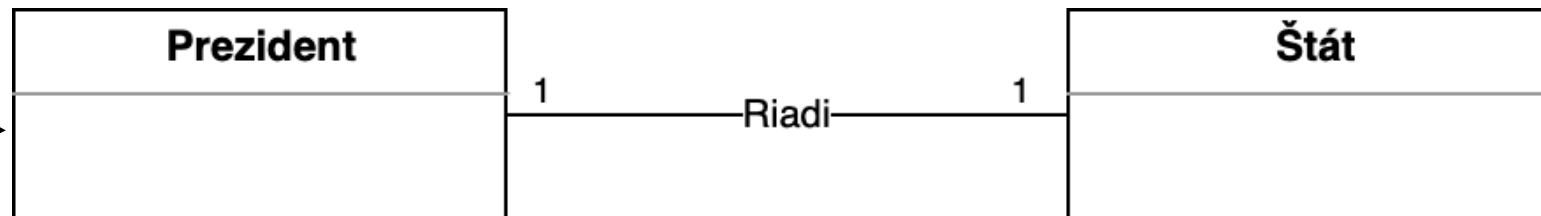


1:1 One-to-One

- *Príklad:* prezident riadi štát

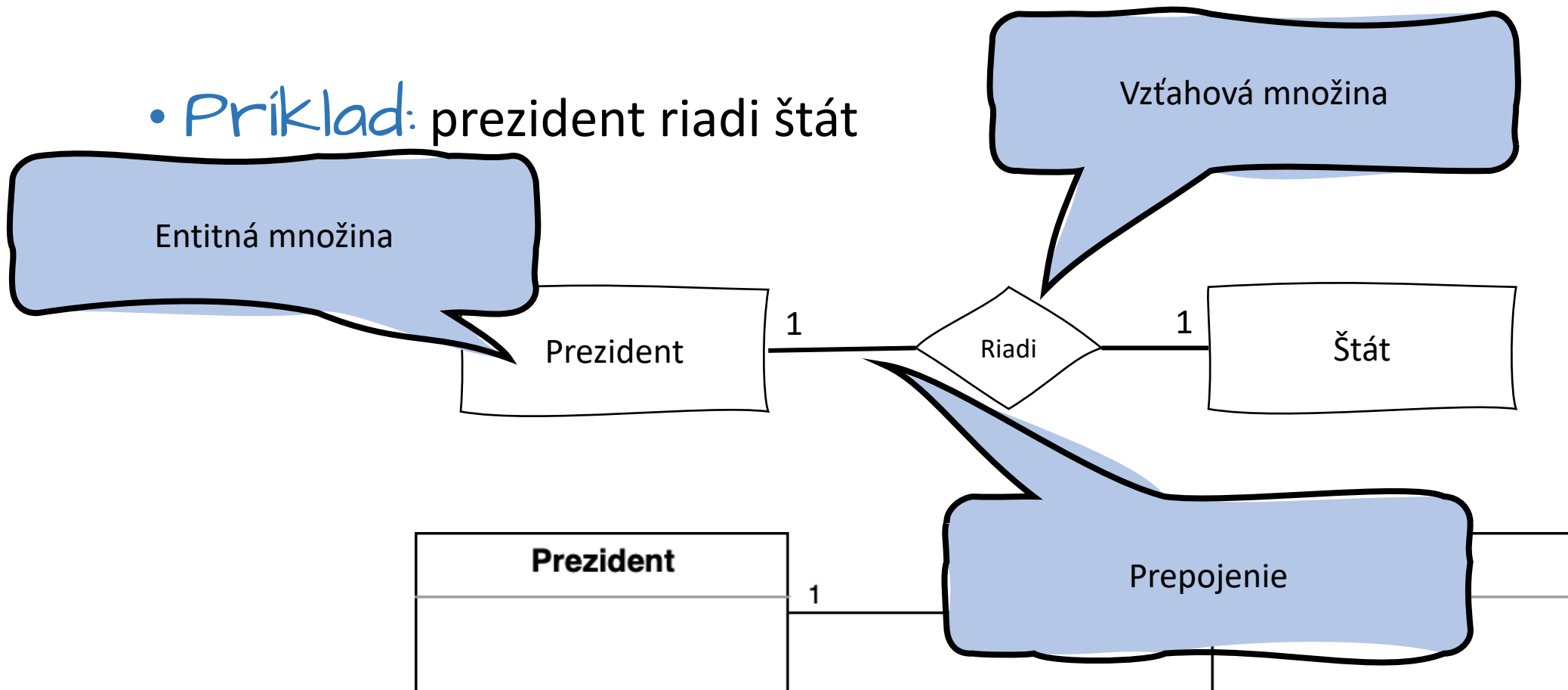


UML
reprezentácia
vzťahov



1:1 One-to-One

- *Príklad:* prezident riadi štát



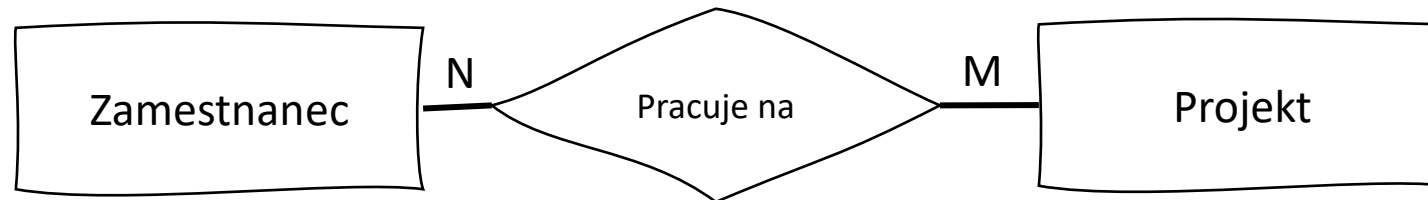
1:N One-to-Many

- **Príklad:** oddelenie zamestnáva zamestnanca , zamestnanec je zamestnaný na oddelení

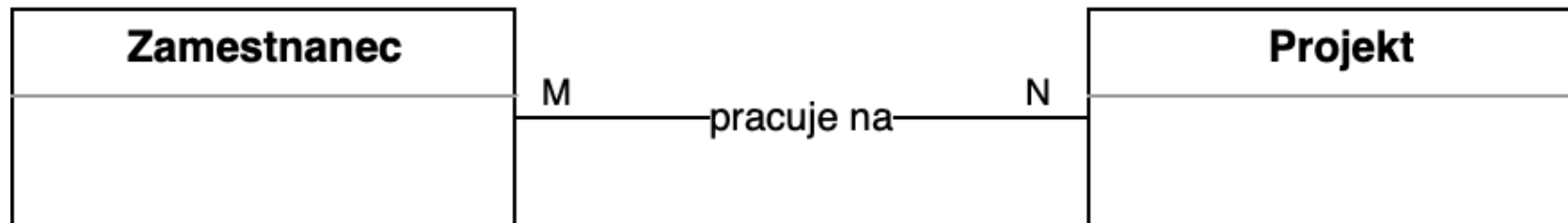


M:N Many-to-Many

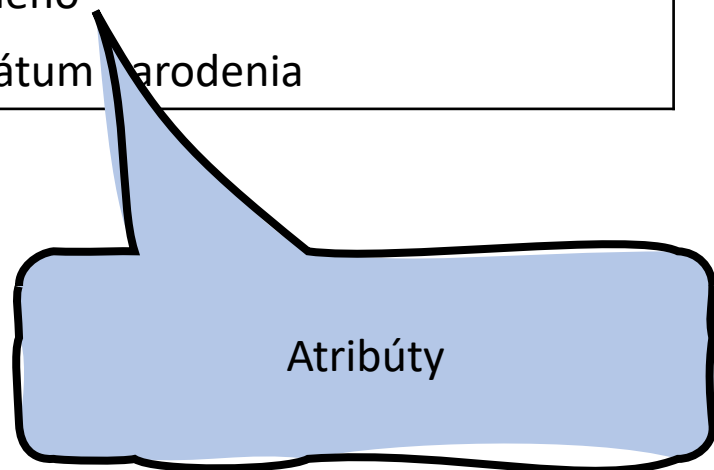
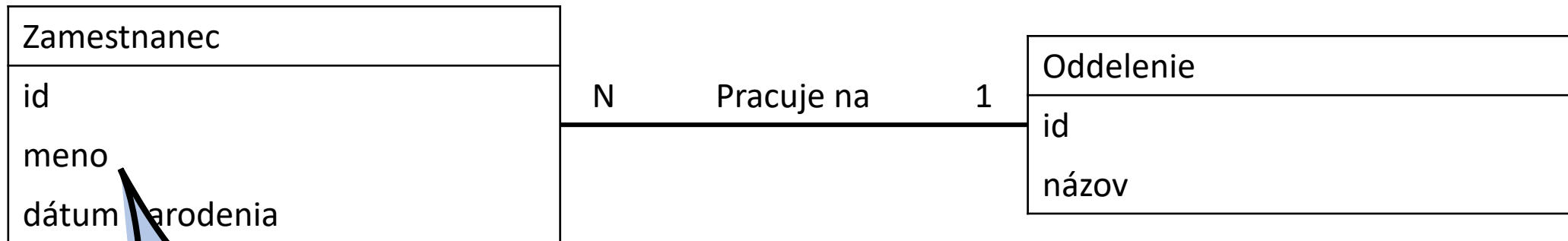
- **Príklad:** zamestnanec pracuje na viacerých projektoch a na projekte môže pracovať viacero zamestnancov



Zdroj: [2]



Atribúty



Atribúty

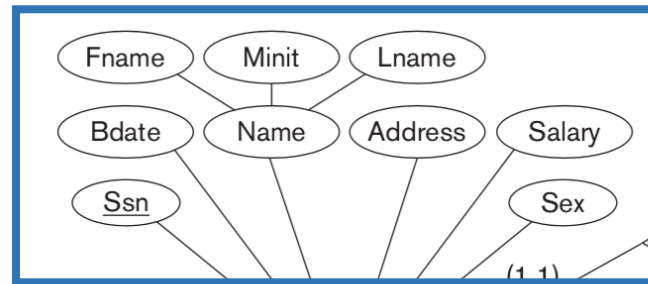
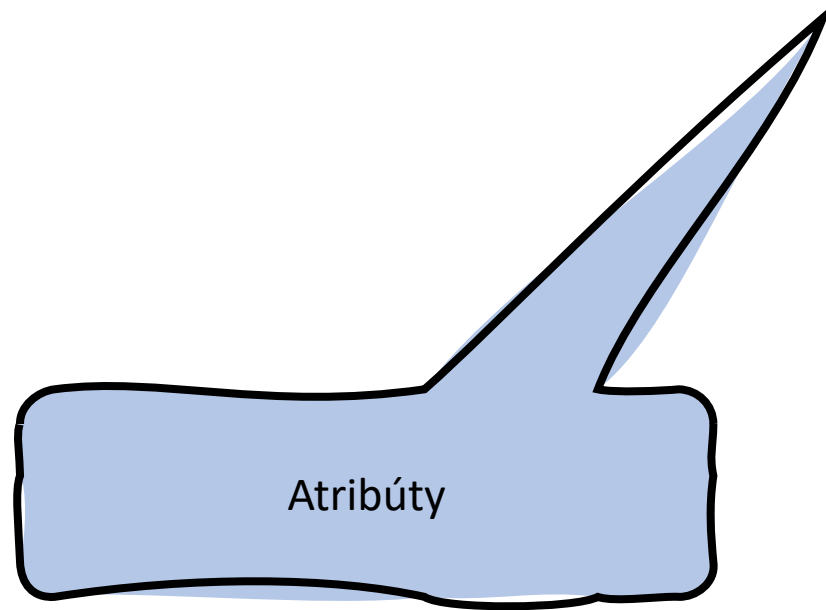
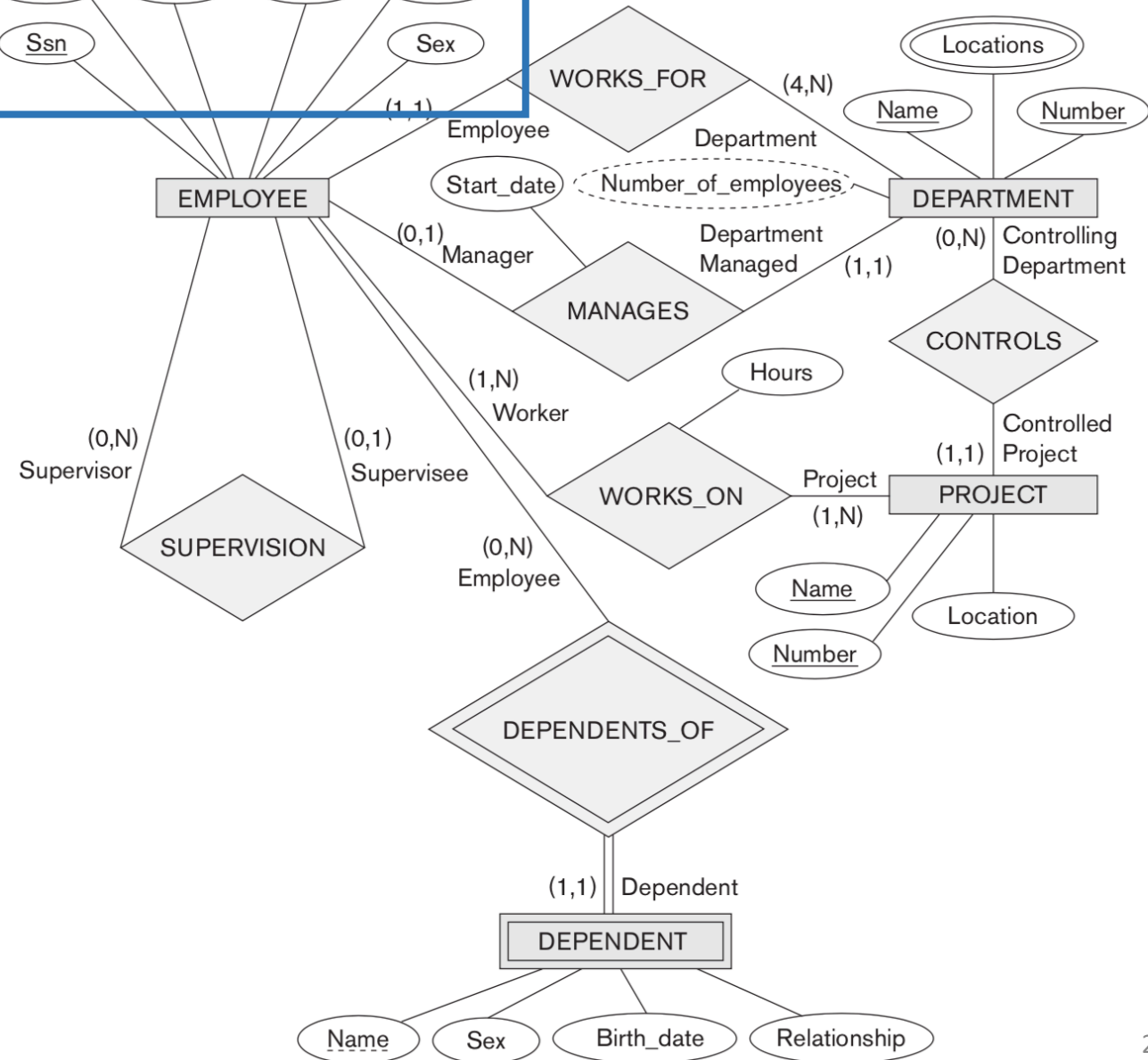


Figure 7.15

ER diagrams for the company schema, with structural constraints specified using (min, max) notation and role names.

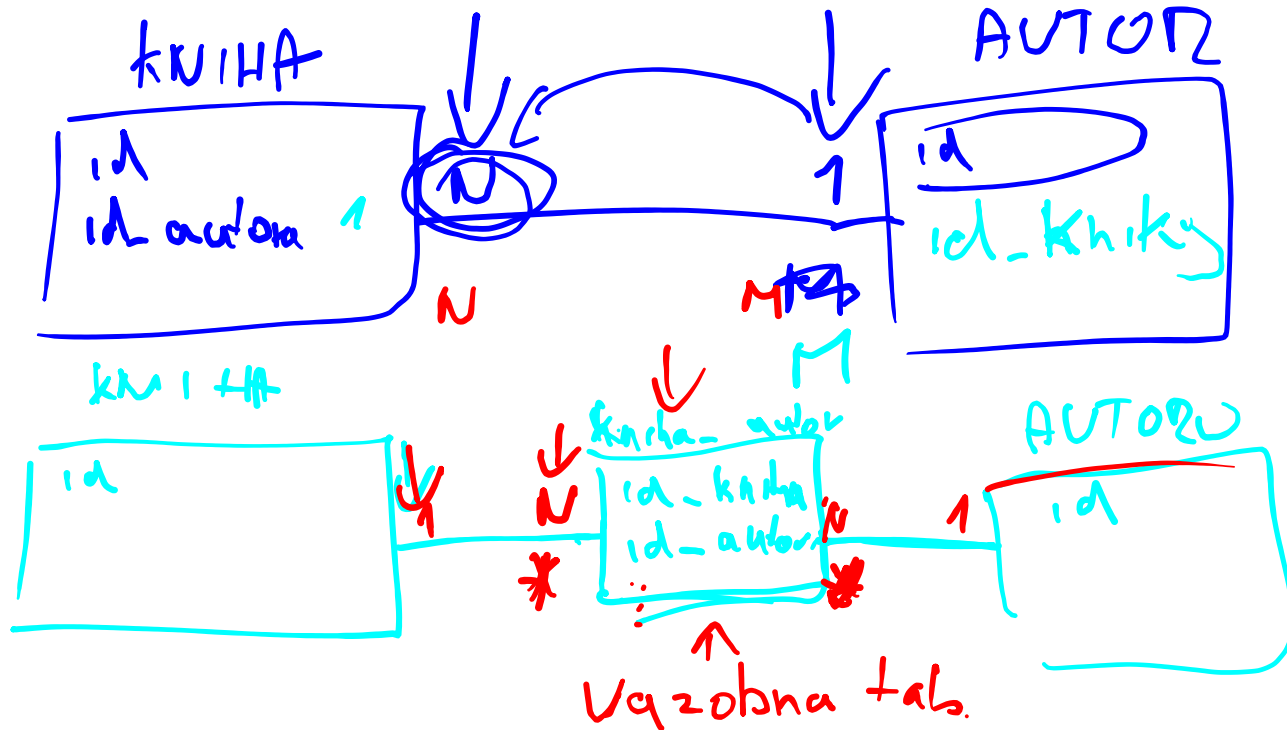
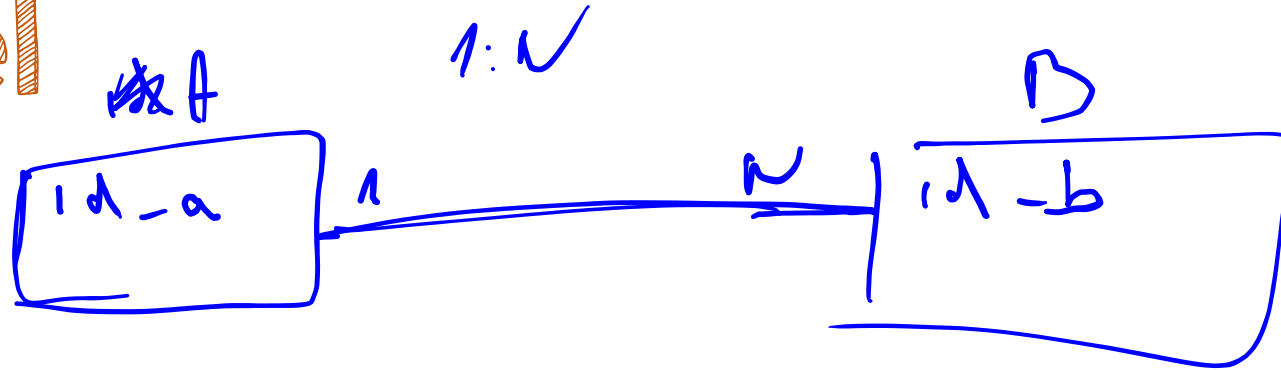


ER vs relačný model

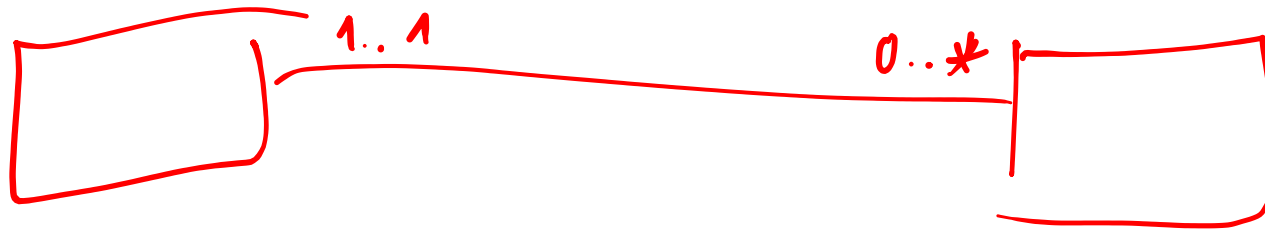
ER model	Relačný model
Entitná množina	Relácia/tabuľka
1:1 alebo 1:N vzťah	Cudzí kľúč alebo dodatočná relácia vzťahu
M:N vzťah	Relácia/tabuľka vzťahu s 2 cudzími kľúčmi
n-ary vzťah	Relácia/tabuľka vzťahu s n cudzími kľúčmi
Jednoduchý atribút	Atribút
Zložený atribút	Viacero jednoduchých atribútov
Viachodnotové atribúty	Relácia/tabuľka a cudzí kľúč
Množina hodnôt	Doména
Kľúčový atribút	Primárny kľúč

ER transformácia na relačný model

PK
FK



Relačný model - kardinality



$0..1$

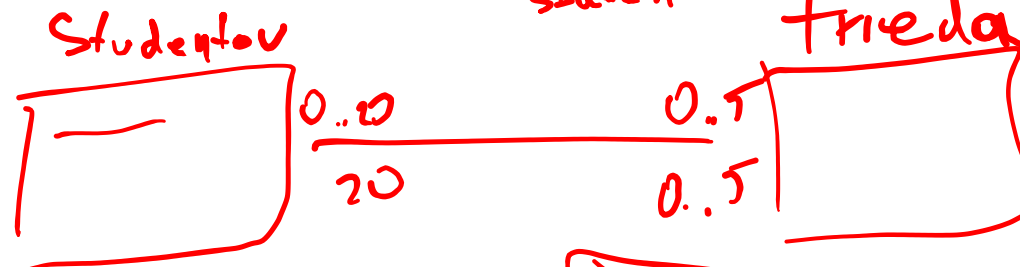
$1..*$

$0..*$

$1..3$

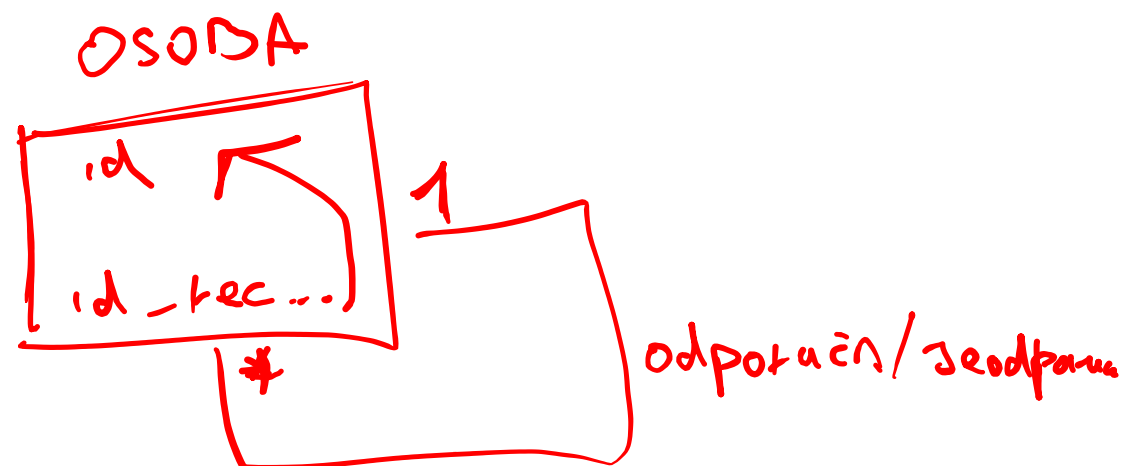
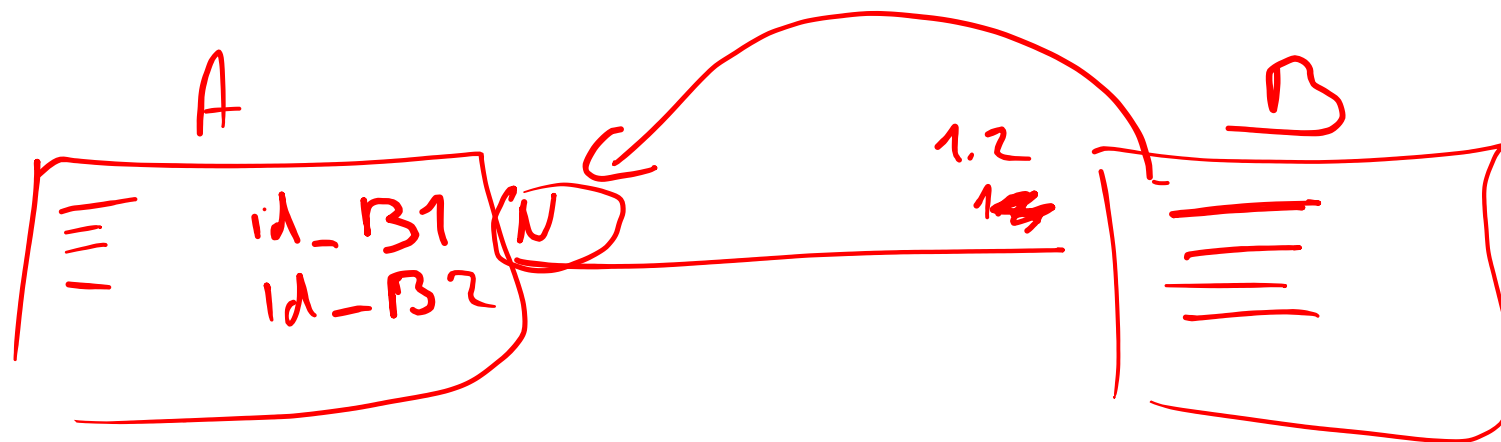
$1..4$

počet študentov max 20
 počet študentov max 5 trieda

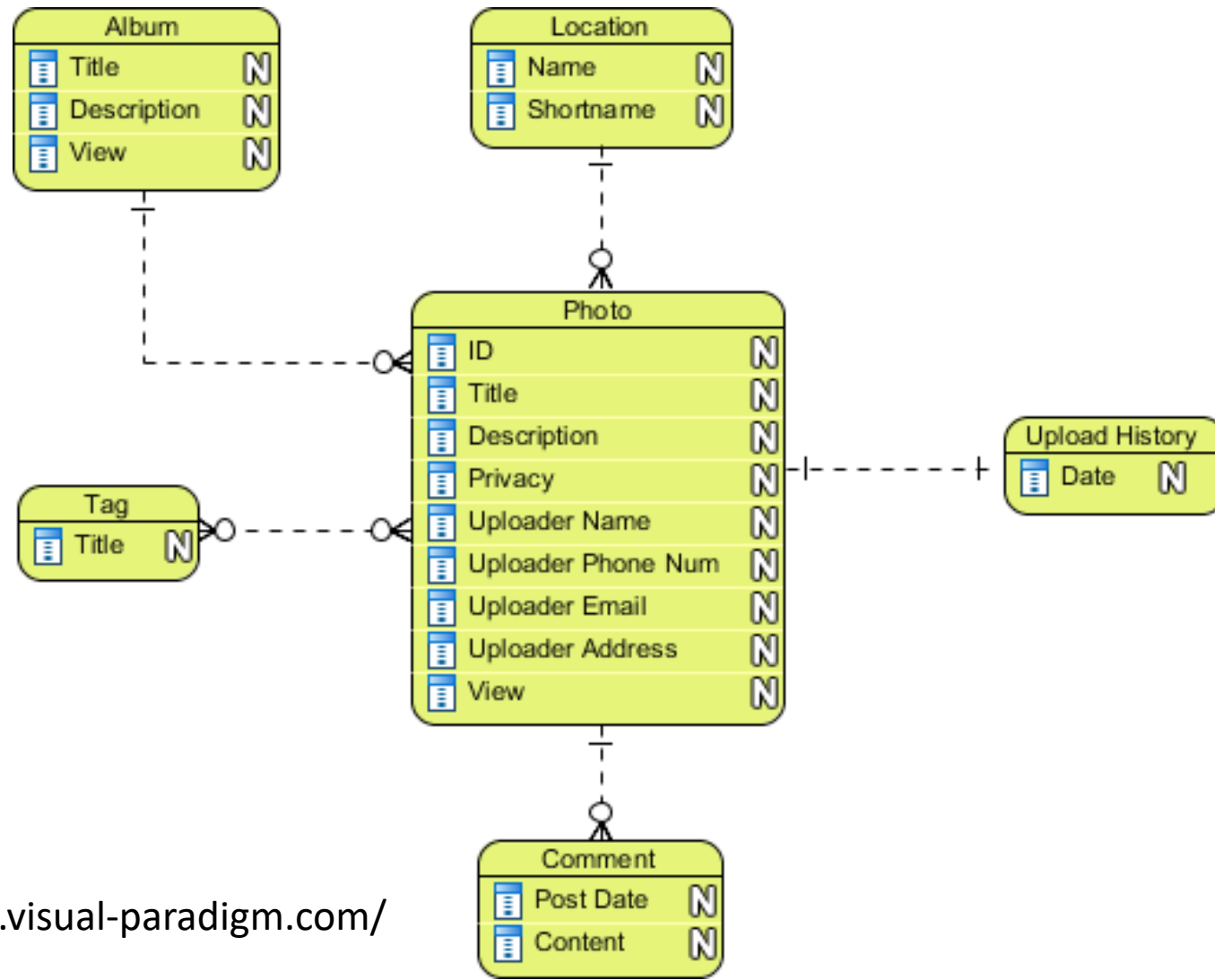


$1 = 1..1$
 $* = 0..*$

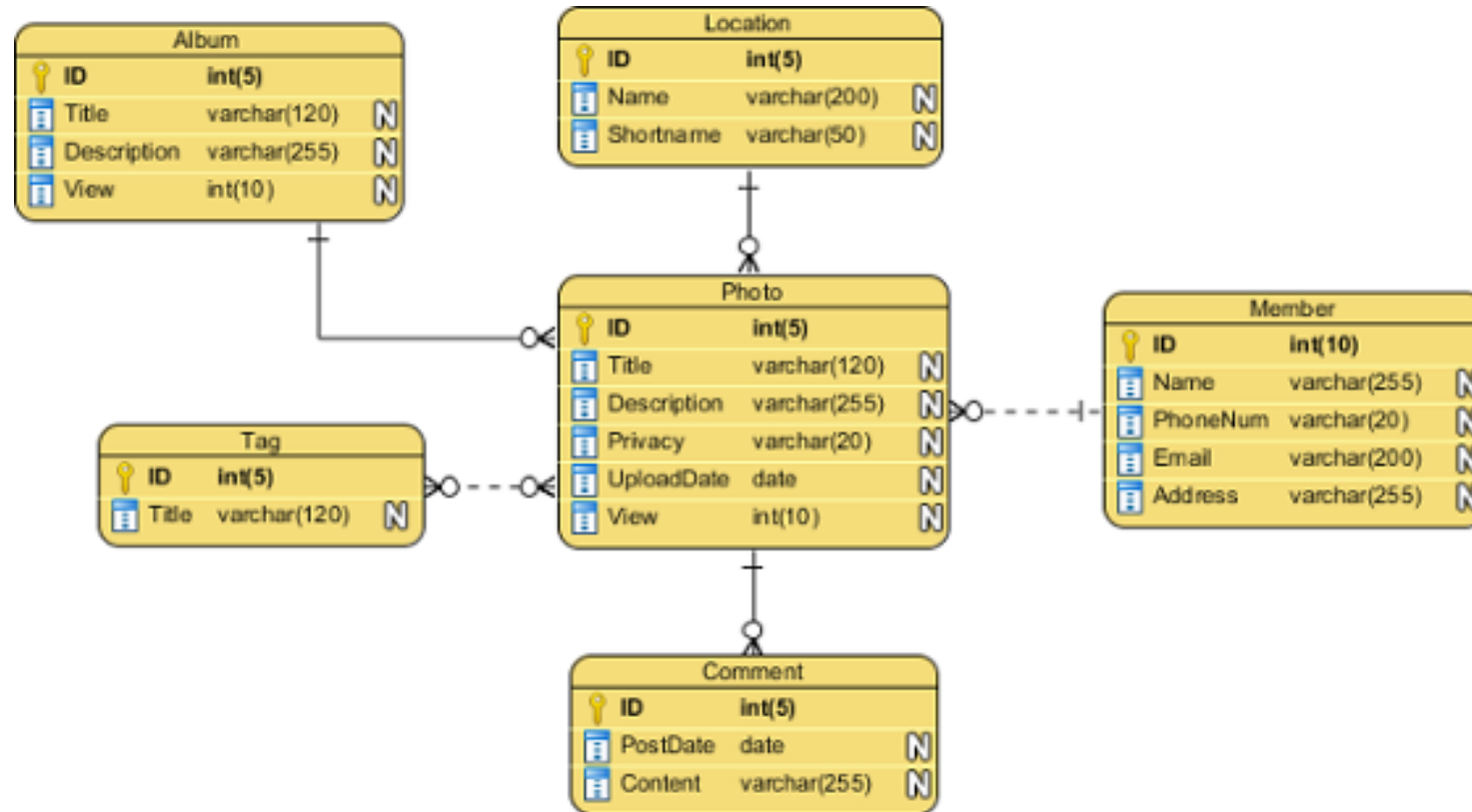
1 študent
 1 trieda



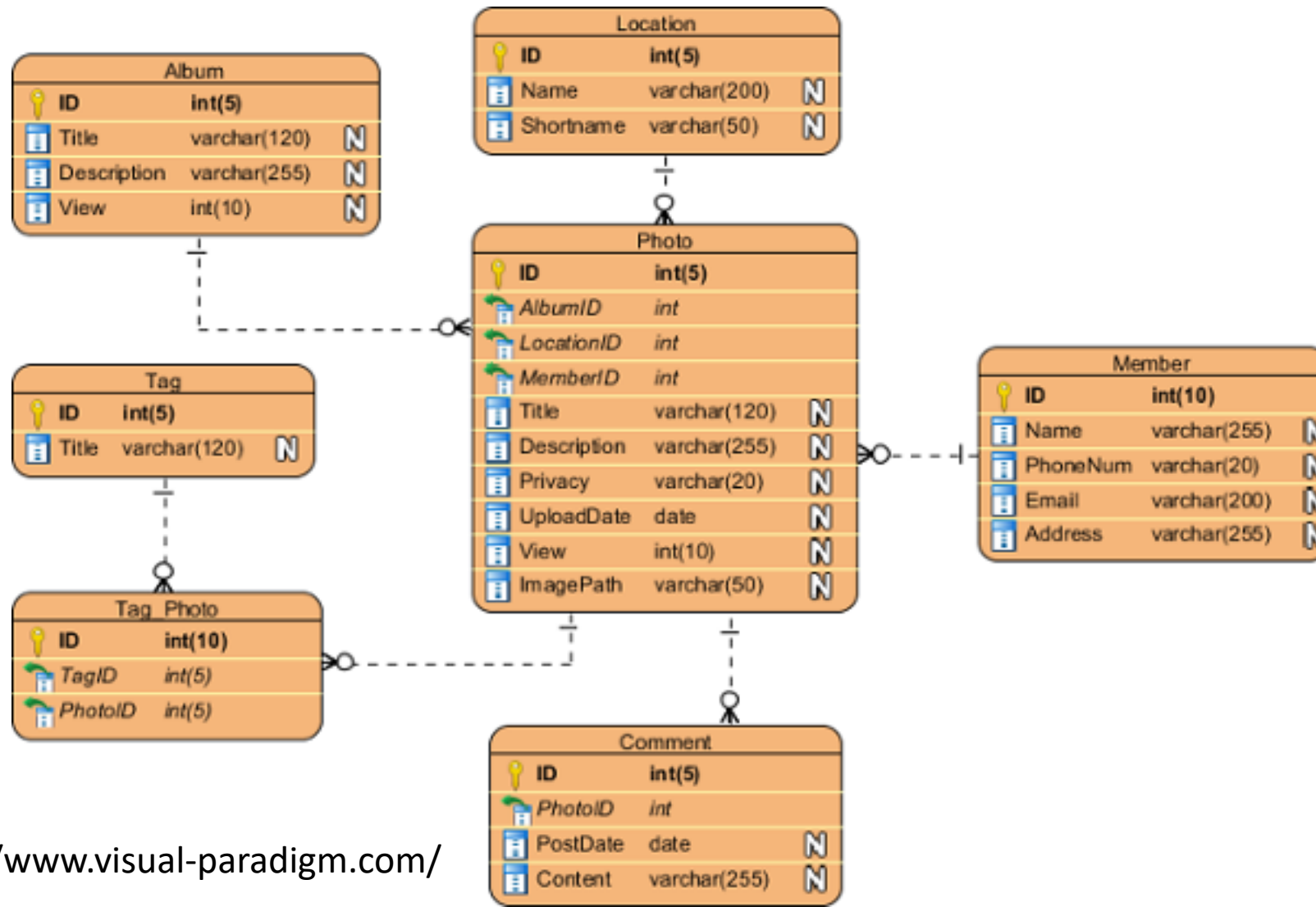
Príklad - konceptuálny dátový model



Príklad - logický dátový model



Príklad - fyzický dátový model



Normalizácia

Motivácia

- Uvažujeme nad DB, ktorá obsahuje študentov, predmety, ktoré si zapisujú a tiež miestnosť, kde sa uskutočňuje daný predmet.

Študent	Predmet	Miestnosť
Janko	DBS	Google-meet
Ferko	DBS	Google-meet
Jožko	DBS	Google-meet
..

- Čo je na danom dizajne nedostatočné ?

Anomália pri vkladani

- Ako vložíme nový predmet, ktorý nemá žiadnych študentov?
 - Nutnosť použitia null hodnoty

Študent	Predmet	Miestnosť
Janko	DBS	Google-meet
Ferko	DBS	Google-meet
Jožko	DBS	Google-meet
..

Anomália pri mazaní

- Keď chceme odstrániť študenta z daného predmetu tak ako realizovať danú zmenu v tabuľke?
 - Ak zmažeme celý záznam, tak môžeme prísť o celú informáciu o tom, že nejaký predmet máme
 - Ak nahradíme študenta NULL hodnotami, tak nám začnú vznikať rovnaké záznamy
 - Aj keď nie su to rovnaké záznamy, keďže NULL hodnoty sú navzájom neporovnateľné

Študent	Predmet	Miestnosť
Janko	DBS	Google-meet
Ferko	DBS	Google-meet
Jožko	DBS	Google-meet
..

Anomália pri aktualizácii

- Ak chceme zmeniť miestnosť predmetu napr. v dôsledku preklepu alebo z akéhoto iného dôvodu
 - Potreba aktualizovania veľkého množstva riadkov
 - Niekedy je potrebné dať pozor na to, či je možné urobiť danú zmenu
 - Napr. Zmenenie preklepu PU42 na PU41 v dôsledku preklepu je potrebné robiť spolu s atribútom Predmet inak by sme prepísali miestnosť PKS, ktorý však je správny


Študent	Predmet	Miestnosť
Janko	DBS	PU41
Ferko	DBS	PU42
Jožko	PKS	PU42
Anicka	DBS	PU42
..

Dekompozícia

- Správny návrh relačnej DB pre predchádzajúci scenár by mohol vyzeráť nasledovne
 - Vieme dospieť pomocou normálizácie

Študent	Predmet
Janko	DBS
Ferko	DBS
Jožko	DBS
..	..

Predmet	Miestnosť
DBS	Google-meet
PKS	Aula
..	..



Normalizácia

- Slúži nám na overenie kvality návrhu
- V rámci normálizácie
- Dekompozícia - slúži nám na to ako fixnúť daný problém
- Normálové formy – slúžia na odhalenie nedostatkov v dizajne.
- Každú relačnú DB je možné reprezentovať jednou veľkou tabuľkou
 - Prečo sa to nerobí ?
 - Redundancia
 - Riziko nekonzistencie
 - Anomálie pri vkladaní, mazaní, a modifikácií dát
 - Potreba NULL hodnôt
 - Plýtvanie pamäťou

Zadanie 3