

Databáze v chemické a forenzní analýze

Tereza Uhlíková

March 11, 2025

doc. Mgr. Tereza Uhlíková, Ph.D.

Ústav analytické chemie

skupina teoretické spektroskopie

místnost A277

<https://web.vscht.cz/~uhlikovt/>

tereza.uhlikova@vscht.cz

Pár otázek pro začátek

Co je to normalizace? Druhá věta termodynamiky. Co je to Primární klíč? Kde leží město Slunce? Na co se používá Cizí klíč? Jak se tvoří Relace? Co si představíte pod pojmem umělá inteligence? Co je to Entita? Jak hluboko se dostaneme, když používáme Hlubkové učení? Co znamená Tabulka v relačních databázích? Co znamená atomická hodnota? Čeho je zkratka NIST? Kdo všechno se podílí na tvorbě databáze? Jak často používáte ChatGPT? Co je to Systém Řízení Báze Dat? Co znamená zkratka SQL? Co je to Acces? Jaký je princip RSA šifry? Jaký je princip Starkova jevu? Jak se jinak nazývá sloupec tabulky? V čem je hrozba kvantových počítačů? Jaké znáte datové typy? Co zapisujeme do řádku tabulky? Co znamená ERA? Jaké máme druhy integrity? Druhá sloka písně "Kde domov můj". Co znamená kardinalita vztahu? Jaké znáte Forezní databáze? Jaké znáte Chemické databáze? Jaké znáte Spektroskopické databáze? Co je to Perceptron? Kdo to byl Alan Turing? Co je to Qbit? Jak postupujete při hledání odpovědí? Jak postupuje počítač při hledání?

O čem to bylo

1	19.3.	Opakování databází, normalizace, ERA
2	26.3.	Spektroskopické databáze - atmosférické, vesmírné
3	2.4.	NIST
4	9.4.	MS, NMR, krystalografické
5	16.4.	AI a Kvantové počítače pro práci s daty
6	23.4.	Barvy, laky, sklo, textilie
7	30.4.	Půda, hořlaviny, nukleární materiály
8	7.5.	STŘEDEČNÍ ROZVRH = NENÍ PŘEDNÁŠKA
9	14.5.	Toxikologické
10	21.5.	Zkouška

cvičení: společná databáze volně prodejných psychotropních látek

O čem to bude?

		Téma přednášky	Téma cvičení
1	11.3.	Opakování a seznamování	Hledání odpovědí pomocí AI
2	18.3.	Princip AI	AI databázové nástroje
3	25.3.	Chemické databáze	Analýza spekter pomocí databází
4	1.4.	Forenzní databáze	zkoušková témata
5	8.4.	společná databáze volně prodejných psych. látek ERA	
6	15.4.	společná databáze volně prodejných psych. látek VY a AI	
7	22.4.		
8	29.4.		
9	6.5.	ČTVRTEČNÍ ROZVRH = NENÍ PŘEDNÁŠKA	
10	...	Zkouška	

Pár otázek pro začátek



- Co je to databáze?
- Kdo ji tvoří?
- Proč ji tvoří?
- Komu slouží?

Pár otázek pro začátek

- Jaké znáte?
- Jaké používáte?
- Kolik stojí?
- Kolik pracujících lidí?
- Jací lidé jsou potřeba?
- Kolik času?

Databáze neboli datová základna (Data Base) je místo, kam se ukládají **určitým** způsobem organizované a strukturované údaje.

= data i software

DBMS - DataBase Management System

SŘBD - Systém Řízení Báze Dat

je softwarové vybavení, které zajišťuje práci s databází.

Databázový systém musí být schopen efektivně pracovat s velkým množstvím dat a spravovat je.

- Definovat bázi
- Konstruovat databázi
- Manipulovat s databází

databázový systém x programovací jazyk

- Oracle
- DB2
- Sybase Adaptive Server Enterprise
- FileMaker
- Informix
- Microsoft Access
- Microsoft SQL Server
- Microsoft Visual FoxPro
- MySQL
- PostgreSQL
- Progress
- SQLite
- CSQL
- OpenLink Virtuoso

Relace, Tabulka, Entita

Co je to relace?

- spojení, souvislost, vztah

- v tabulce jsou data v relaci mezi sloupcem a řádkem

⇒ tabulka

Objekt reality, který je zapsán pomocí jedné tabulky ⇒ entita

Relace \approx Tabulka \approx Entita

relace (tabulka)	sloupec			
	Název sloupce 1	Název sloupce 2	...	Název sloupce N
řádek	pole			

- Každá tabulka má jednoznačné jméno.
- Každý sloupec v tabulce má jednoznačné jméno.
- Všechny hodnoty daného sloupce musí být stejného typu.
- Nezáleží na pořadí sloupců.
- Nezáleží na pořadí řádků.
- Tabulka nemůže mít duplicitní řádky.
- Všechny hodnoty jsou atomické.
- Každá tabulka musí mít primární klíč.

Atribut: sloupec

- určují vlastnosti objektů (příjmení, fakulta,...)
- jedinečný název a - nepoužívat diakritiku a mezery, tedy např.:
navez_odberatele
- určený datový typ (číslo, text, logická hodnota, ...)

Záznam: řádek

- každý záznam je jednoznačně rozlišitelný - jeden řádek reprezentuje
např. jednoho zaměstnance s hodnotami daných atributů
- mít svůj jedinečný identifikátor

id_zamestnance	prijmeni	jmeno	plat	datum_nastupu
1	Novák	Adam	32000	2013-05-02
2	Nová	Jana	45000	2013-06-11

vícehodnotná položka

špatně

správně

strelec	zasahy
Pavel	4, 9, 2
Michal	1, 5, 7

strelec	zasah1	zasah2	zasah3
Pavel	4	9	2
Michal	1	5	7

Návrh tabulek

opakující se hodnoty → plýtvání místem → hromadné upravení
špatně **správně**



vypočítaná položka

- sloupec, který se dá vytvořit výpočtem z ostatních položek v záznamu, by neměl být součástí tabulky
- výpočet se provede v dotazu

špatně

obdelnik	sirka	vyska	obvod	obsah
malý	2	4	10	8
velký	30	15	90	450

Normalizace

Slouží k jasnému vyřešení problémů s nadbytečností.

odstranění redundance - opakovaných výskytů stejného údaje v databázi

- eliminovat duplicitní (opakující se) sloupce v jednotlivých tabulkách;
- vytvořit pro každou skupinu dat s jasně vymezeným významem samostatnou tabulku a každý řádek opatřit jedinečným primárním klíčem;
- obsahem jednotlivých sloupců tabulky by měla být jednoduchá, dále nedělitelná informace;
- podmnožinu dat se shodnou hodnotou pro určitý sloupec tabulky převést do samostatné tabulky a spojit s původní tabulkou cizím klíčem;
- odstranit z tabulky sloupce, které jsou přímo závislé na jiné skupině sloupců tabulky než pouze na primárním klíči

1NF

všechna data v relaci musí být atomická

2NF

data závisí na celém klíči

3NF

neklíčová data jsou závislá jen na klíči a ne mezi sebou

BCNF

všechna data jsou závislá jen na klíči a ne mezi sebou

4NF

složený primární klíč nesmí být tvořen z nezávislých dat

5NF

trojný a vícový primární klíč nesmí obsahovat párové cyklické závislosti

Úprava databázové tabulky dle normálních forem

Záznamy o výpůjčkách						
Číslo záznamu	Jméno zákazníka	Bydliště	Vypůjčil si			Vrátil
			DVD	Název filmu	Dne	
1	Lukáš Pavel	Studentská 25, Ostrava Přívoz, 708 31	21	Piráti z Karibiku I.	5.1.2012	9.1.2012
2	Lukáš Pavel	Studentská 25, Ostrava Přívoz, 708 31	5	Honza málem králem	5.1.2012	8.1.2012
3	Pavel Kvapil	Karvinská 4, Karviná-Ráj, 735 64	21	Piráti z Karibiku I.	10.1.2012	11.1.2012
4	Pavel Kvapil	Karvinská 4, Karviná-Ráj, 735 64	22	Piráti z Karibiku II.	10.1.2012	11.1.2012
5	Pavel Kvapil	Karvinská 4, Karviná-Ráj, 735 64	23	Piráti z Karibiku III.	10.1.2012	
6	Levá Pavla	Hornosušská 24, Horní Suchá, 735 35	51	Popelka	10.1.2012	10.1.2012
7	Pavel Lukáš	Studentská 25, Ostrava-Přívoz, 708 31	51	Popelka	15.2.2012	19.2.2012
8	Levá Pavla	Studentská 1, Ostrava-Přívoz, 708 31	5	Honza málem králem	15.2.2012	16.2.2012
9	Karel Novák	Vnitřní 1151, Těrlicko, 735 65	22	Piráti z Karibiku II.	16.2.2012	16.2.2012
...						

- nejednoznačnost** údajů – vyhledávací anomálie – není jednoznačně určeno, co je jméno a co příjmení zákazníka (viz záznamy 1, 2 a 7), název města v záznamech 1, 2 a 7 je odlišný i přesto, že jde o stejné město;
- redundance** záznamů – když si zákazník vypůjčí více filmů najednou, pak se pro každý záznam opakují pole bydliště a datum vypůjčení, totéž při dalších výpůjčkách stejného zákazníka;
- nekonzistence** záznamů – pokud zákazník změní bydliště (viz záznamy č. 6 a 8), pak je nutné provést změnu ve všech záznamech.

Relace je v první normální formě, pokud každý její atribut obsahuje jen atomické hodnoty.

Řešením je rozklad polí Jméno zákazníka a Bydliště na několik dalších polí

Záznamy o výpůjčkách										
Číslo záznamu	Jméno zákazníka	Příjmení zákazníka	Ulice	Číslo	Město	PSČ	Vypůjčil si			Vrátil
							DVD	Název filmu	Dne	
1	Lukáš	Pavel	Studentská	25	Ostrava Přívoz	708 31	21	Piráti z Karibiku I.	5.1.2012	9.1.2012
2	Lukáš	Pavel	Studentská	25	Ostrava Přívoz	708 31	5	Honza málem králem	5.1.2012	8.1.2012
3	Pavel	Kvapil	Karvinská	4	Karviná-Ráj	735 64	21	Piráti z Karibiku I.	10.1.2012	11.1.2012
4	Pavel	Kvapil	Karvinská	4	Karviná-Ráj	735 64	22	Piráti z Karibiku II.	10.1.2012	11.1.2012
5	Pavel	Kvapil	Karvinská	4	Karviná-Ráj	735 64	23	Piráti z Karibiku III.	10.1.2012	
6	Pavla	Levá	Hornosušská	24	Horní Suchá	735 35	51	Popelka	10.1.2012	10.1.2012
7	Lukáš	Pavel	Studentská	25	Ostrava-Přívoz	708 31	51	Popelka	15.2.2012	19.2.2012
8	Pavla	Levá	Studentská	1	Ostrava-Přívoz	708 31	5	Honza málem králem	15.2.2012	16.2.2012
9	Karel	Novák	Vnitřní	1151	Těrlicko	735 65	22	Piráti z Karibiku II.	16.2.2012	16.2.2012
...										

Každý neklíčový atribut je plně závislý na primárním klíči

Rozložení tabulky pro data s jasně vymezeným významem

Zákazník						
Číslo záznamu	Jméno zákazníka	Příjmení zákazníka	Ulice	Číslo	Město	PSČ
1	Lukáš	Pavel	Studentská	25	Ostrava-Přivoz	708 31
2	Pavel	Kvapil	Karvinská	4	Karviná-Ráj	735 64
3	Pavla	Levá	Studentská	1	Ostrava-Přivoz	708 31
4	Karel	Novák	Vnitřní	1151	Těrlicko	735 65
5	...					

Výpůjčka		
ID výpůjčky	Zákazník	Dne
1	1	5.1.2012
2	2	10.1.2012
3	3	10.1.2012
4	1	15.2.2012
5	3	15.2.2012
6	4	16.2.2012
...		

Položka				
Číslo položek	DVD	Název filmu	Výpůjčka	Vráceno
1	21	Piráti z Karibiku I.	1	9.1.2012
2	5	Honza málem králem	1	8.1.2012
3	21	Piráti z Karibiku I.	2	11.1.2012
4	22	Piráti z Karibiku II.	2	11.1.2012
5	23	Piráti z Karibiku III.	2	
6	51	Popelka	3	10.1.2012
7	51	Popelka	4	19.2.2012
8	5	Honza málem králem	5	16.2.2012
9	22	Piráti z Karibiku II.	6	16.2.2012
10	...			

Žádný z jejich atributů nevykazuje tranzitivní závislost, tzn., že všechny neklíčové atributy jsou navzájem nezávislé
vytvoření nových tabulek *Město* a *DVD*

Zákazník					
Číslo záznamu	Jméno zákazníka	Příjmení zákazníka	Ulice	Číslo	PSČ
1	Lukáš	Pavel	Studentská	25	708 31
2	Pavel	Kvapil	Karvinská	4	735 64
3	Pavla	Levá	Studentská	1	708 31
4	Karel	Novák	Vnitřní	1151	735 65
5	...				

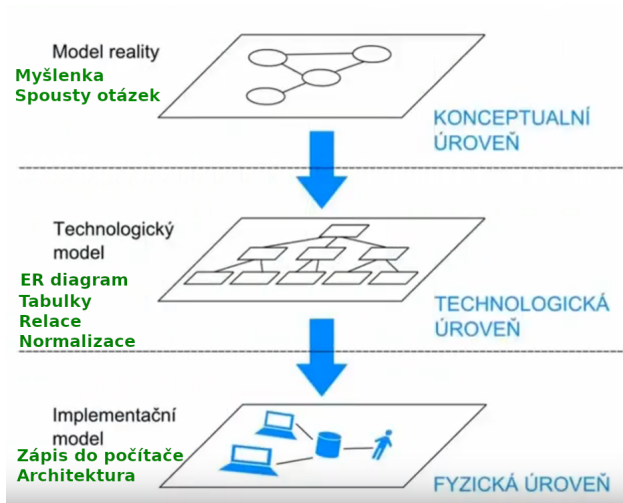
Město	
PSČ	Město
708 31	Ostrava Přívoz
735 64	Karviná-Ráj
735 35	Horní Suchá
735 65	Těrlicko
...	

Výpůjčka		
ID výpůjčky	Zákazník	Dne
1	1	5.1.2012
2	2	10.1.2012
3	3	10.1.2012
4	1	15.2.2012
5	3	15.2.2012
6	4	16.2.2012
...		

DVD		
Číslo DVD	Název filmu	Kategorie
...
5	Honza málem králem	Pohádka
...
21	Piráti z Karibiku I.	Komedie
22	Piráti z Karibiku II.	Komedie
23	Pirátki z karibiku III.	Komedie
...
51	Popelka	Pohádka
...

Položka			
Číslo položek	DVD	Výpůjčka	Vráceno
1	21	1	9.1.2012
2	5	1	8.1.2012
3	21	2	11.1.2012
4	22	2	11.1.2012
5	23	2	
6	51	3	10.1.2012
7	51	4	19.2.2012
8	5	5	16.2.2012
9	22	6	16.2.2012
10

Návrh relační databáze



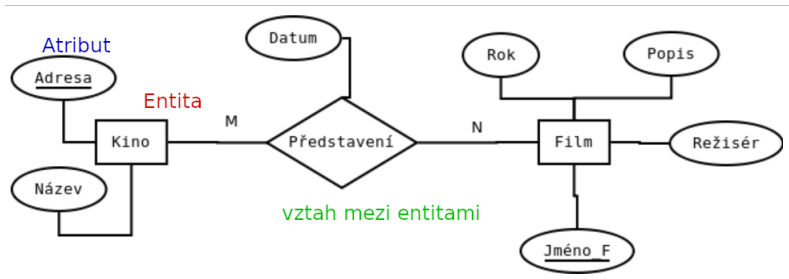
- **Entity** (objekt z reálného světa) → tabulka
- **Relationship** (vztah mezi entitami) → vztahy mezi tabulkami
- **Attribute** (vlastnosti entit) → sloupeček tabulky

Modeluje „oblast zájmu“ (např. podnik) jako kolekci entit a vztahů (relationships) mezi nimi

- grafická reprezentace struktury databáze

dělá se před vytvářením databáze!! a slouží k rozmyšlení, jak bude databáze vypadat

Klasický formát



Integrita Entitní – jedinečná identifikace každé entity - pomocí primárního klíče (PK)

Primární klíč

- zajistí jednoznačnost záznamů v tabulce,
- AUTO INCREMENT - zajistí automatické číslování => hodnota primárního klíče se v tabulce nesmí opakovat
- data nesmí obsahovat duplicitu nebo hodnotu Null.
- teoreticky může být libovolného typu
- může být složený z více sloupců – unikátnost je pak brána ze všech najednou, v jednotlivých sloupcích se tedy mohou opakovat stejné hodnoty

Integrita Referenční – pomocí cizího klíče (FK)

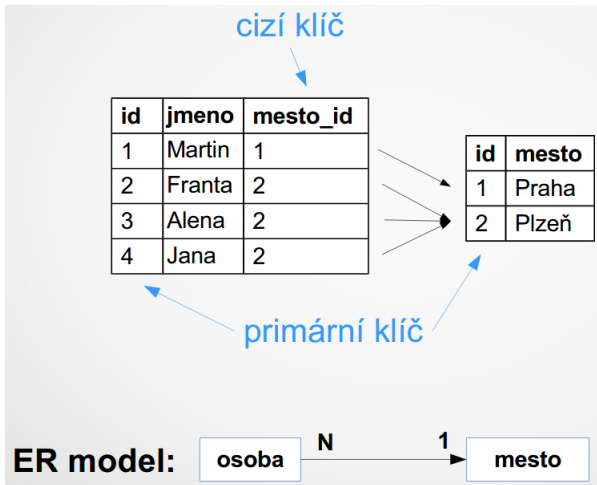
Cizí klíč

- odkazuje se na primární klíč v jiné tabulce – odkazuje se na právě 1 tabulku
- dvojice primární a cizí klíč tvoří vazbu (relaci) mezi tabulkami
- cizí klíč musí být stejného typu a velikosti jako primární

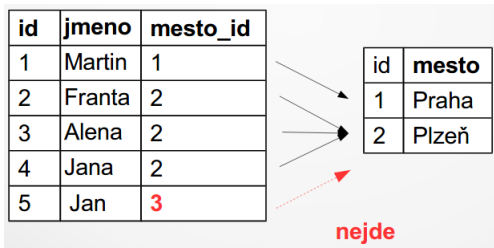
if primární klíč AUTO INCREMENT => cizí klíč dlouhé celé číslo

if primární klíč složený => struktura cizího klíče stejná

Do sloupce s definovaným cizím klíčem nelze vložit hodnotu, která není vložena v nadřazeném sloupci primární tabulky.



databáze nedovolí do cizího klíče vložit hodnotu, která neexistuje v primárním klíči v cílové tabulce



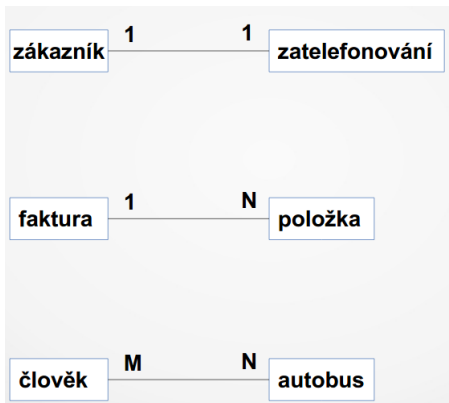
mocnost vztahu mezi entitami
neboli

Kolik položek z tabulky A se váže na kolik položek z tabulky B?

3 typy

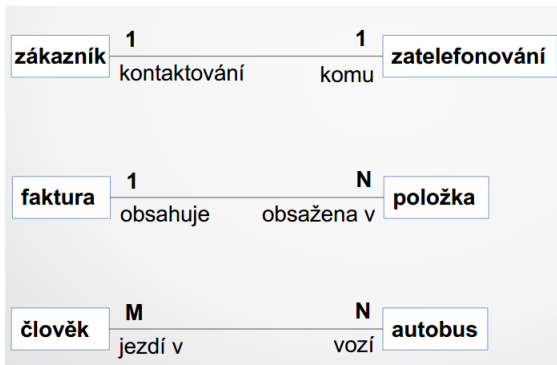
- 1:1 - na 1 položku z tabulky A připadá maximálně 1 položka z tabulky B, ta patří jen té jediné z A, žena-muž (ČR)
- 1:N, N:1 - na 1 položku z tabulky A připadá libovolné množství položek z tabulky B, na 1 položku z tabulky B připadá maximálně 1 položka z tabulky A
- M:N - bez omezení množství vazeb v obou směrech, žena-muž (ČR) v čase

Kardinalita vztahu



Kardinalita vztahu

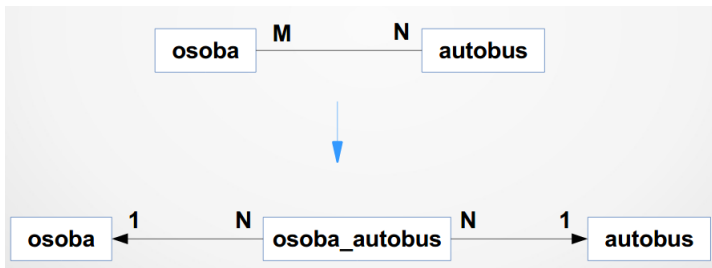
pomůže lepšímu porozumění ER(A) modelu



Kardinalita M:N

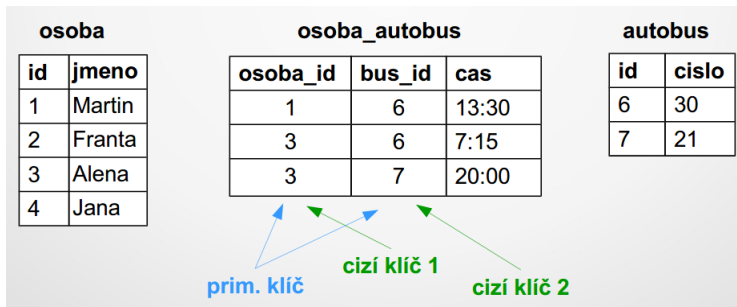
- několika (M) entitám odpovídá několik (N) jiných entit
- např. vztah student - předmět
 - jeden student studuje více předmětů
 - jeden předmět navštěvuje více studentů
 - nejsložitější případ vazby

vazbu M:N nelze vytvořit přímo mezi tabulkami
je potřeba rozložit pomocí nové tabulky na dvě vazby 1:N



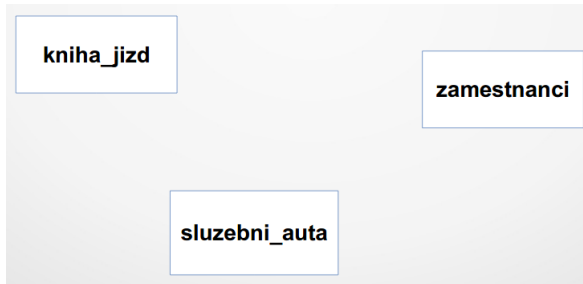
rozkladová tabulka nemusí, ale může obsahovat další informace
primární klíč v rozkladové tabulce se často skládá z obou cizích klíčů (ale může být i vlastní sloupec id)

- např. v kolik kdo jakým autobusem jezdí
- není nutné kvůli tomu vytvářet další tabulku



Kardinalita vztahu

doplňte vazby a jejich kardinality do ER modelu – zaměstnanci používají služební auta a zapisují do knihy jízd (každý zaměstnanec pro každé auto svou knihu jízd)



výsledek - Pro relaci M:N je rozkladová tabulka kniha jízd

