



Data a relační databáze

Úvod do ICT, 25. listopadu 2024



Životní cyklus dat


1. Plánování správy dat
2. Organizace a dokumentace (příprava před zpracováním)
3. Zpracování dat
4. Uložení dat
5. Ochrana dat
6. Archivace a zveřejnění dat
7. (Znovu)objevování dat



Zdroj: <https://dmeg.cessda.eu>

/

Fáze I až III

- Plánování správy dat:
 - aspekty technické, organizační, strukturální, právní, etické a udržitelné
 - Organizace dat:
 - způsob uložení a pojmenování, zajištění integrity, metadata k datům
 - Zpracování dat:
 - Kódování vstupu, kontrola, čištění, transformace, propojování, analýzy, vizualizace
- 

Fáze IV až VII

- Uložení dat
 - soubory, databáze / disk, cloud / zálohování / zodpovědnost za data
- Ochrana dat:
 - autentizace a autorizace uživatele, šifrování dat, anonymizace, GDPR
- Archivace a zveřejnění dat:
 - Spolehlivost média, licence, interoperabilita, persistentní identifikátor
- Objevování dat:
 - Datové repozitáře, veřejné rejstříky, indexace, otázka důvěryhodnosti

Relační databáze

- Patrně nejpoužívanější způsob ukládání (propojených) dat
- Má základy v matematice, tzv. relačním modelu (kalkulu)
 - selekce, projekce, kartézský součin, rozdíl, sjednocení, spojení
- Data jsou uložena v tabulkách, které lze propojovat
 - Co řádek, to údaje o jedné entitě (osoba, předmět, studijní program, ...)
 - Co sloupec (atribut), to konkrétní údaj (jméno, rodné číslo, pohlaví, e-mail, ...)
 - K propojování slouží sloupce s funkcí klíče – obvykle jednoznačného identifikátoru

Ukázka – knihy v katalogu

barcode	call_no	author	title
4200402293	FIN-241	Vlčková, Libuše	222 otázek a odpovědí k zákonu č. 222/19
4200402427	FIN-242	Benda, Václav	222 otázek a odpovědí k zákonu č. 588/19
4200436823	SOC-460	Buchanan, Patrick J.	Smrt Západu :
4200435151	PRA-848	Pohl, Tomáš	Abeceda obchodního práva pro podnikatele
4200410412	VŠO-4		Abeceda obchodníka.
4200417135	SOC-236	Kuchařová, Věra	Aktuální otázky postavení žen v ČR /
4200401694	VÝCH-1	Motyka, Gustav	Aktuální otázky státní správy a samosprá
4200425053	HPO-378		Aktuální problémy hospodářské politiky /
4200429365	ŘÍZ-350	Hopfenbeck, Waldemar	Allgemeine Betriebswirtschafts- und Manag
4200418344	18344	Hermann, Zoltán	Allocation of local government functions
4200400687	MAK-3		Alokačné mechanizmy globálnej investične
4200462993	ÚČT-618		Abeceda mzdové účetní ... /
4200427643	BAK-1161	Dvorník, Petr	Analýza pojistného produktu a kalkulace
4200424606	DIP-771	Krul, Jiří	Analýza pojistných produktů Nationale-Ne
4200424424	DIP-660	Huječková, Jana	Analýza pojištění motorových vozidel pod
4200426049	BAK-1134	Pavelcová, Jitka	Analýza pojištění odpovědnosti za škodu
4200430677	SOC-417	Keller, Jan	Abeceda prosperity /
4200436275	T-64	Pease, Allan	Řeč těla :

Ukázka – rezervace boxů v SVI

id	user	b	from	to	note	jmeno
1637	392425	5	2013-11-20 14:00:00	2013-11-20 16:00:00	mikro 2	Mička Adam
1638	391285	5	2013-11-12 15:00:00	2013-11-12 16:30:00	Manahra	Hýbela Jakub
1639	391501	1	2013-11-12 17:30:00	2013-11-12 19:00:00	manahra	Zamborský Juraj
1640	376038	3	2013-11-14 10:35:00	2013-11-14 12:00:00	doučování AJ	Vašková Veronika
1641	376038	3	2013-11-14 12:00:00	2013-11-14 13:50:00	doučování AJ	Vašková Veronika
1642	349183	3	2013-11-13 13:10:00	2013-11-13 14:00:00	Príprava prezentácie POKO	Nosál Tomáš
1643	323084	2	2013-11-12 08:00:00	2013-11-12 10:00:00	diplomová práce	Anděra Jindřich
1644	323084	2	2013-11-12 10:00:00	2013-11-12 12:00:00	diplomová práce	Anděra Jindřich
1645	405383	5	2013-11-13 11:00:00	2013-11-13 12:30:00	Anglictina	Široká Anna - Marie
1646	405511	4	2013-11-13 11:15:00	2013-11-13 12:00:00	Skupinová seminární práce - AJ	Pechová Monika
1647	369837	1	2013-11-14 12:45:00	2013-11-14 14:00:00	výuka Ekonometrie	Nešleha Josef
1648	358395	4	2013-11-15 13:45:00	2013-11-15 15:15:00	projekt do STRP	Sekanina Jan
1649	391164	1	2013-11-26 13:30:00	2013-11-26 15:30:00	manahra	Švecová Jana
1650	mirkaw	4	2013-11-19 12:30:00	2013-11-19 15:00:00	ESZS	Mirka Wildmannová
1651	385240	2	2013-11-18 12:00:00	2013-11-18 14:00:00	příprava mikro2	Samková Lenka
1652	405390	3	2013-11-13 08:00:00	2013-11-13 08:30:00	Prezentace AJ	Stehlíková Dita
1653	425113	1	2013-11-13 15:00:00	2013-11-13 16:00:00	English	Dýšek Miroslav
1654	390563	3	2013-11-18 13:00:00	2013-11-18 15:00:00	ZAEK - skupinová práce	Lipková Andrea

Ukázka informací o tabulce

```
mysql> describe bloky;
```

Field	Type	Null	Key	Default	Extra
id	int(10) unsigned	NO	PRI	NULL	auto_increment
cas	timestamp	NO		CURRENT_TIMESTAMP	
username	varchar(10)	NO			
box	tinyint(3)	NO		0	
from	datetime	NO		0000-00-00 00:00:00	
to	datetime	NO		0000-00-00 00:00:00	
note	varchar(255)	NO			
flag	varchar(2)	NO			
jmeno	varchar(100)	NO			

- int, tinyint – čísla; id = číselný identifikátor s automatickým počítáním s funkcí primárního klíče
- varchar – řetězec proměnlivé délky (v závorce definována jeho maximální délka)
- timestamp – časové razítko, při změně položky se uloží aktuální čas; datetime – datum a čas
- Null – zda v rámci daného atributu může být uložena „prázdná“ („neznámá“) hodnota

Proč více tabulek?

- Za první – šetříme místo tím, že neopakujeme data v rámci jedné tabulky
 - Např. pokud máme tabulku objednávek zákazníků, pak nemusíme ke každé objednávce jednoho zákazníka opakovaně vypisovat jeho kontaktní údaje – ty máme zvlášť v jiné tabulce
- Za druhé – urychlíme práci s daty (vyhledávání, řazení)
 - Už jenom tím, že nepracujeme s tabulkami, které zrovna nepotřebujeme
- Normalizace databáze – hledání ideální struktury propojených tabulek

Jazyk SQL

- Standardizovaný jazyk pro práci s relačními databázemi
- Zkratka od Structured Query Language
- Do značné míry univerzální napříč systémy různých výrobců
- Navržen tak, aby co nejvíce připomínal běžné anglické věty:
 - `SELECT něco FROM tabulka;`
 - `INSERT INTO tabulka (atribut1, atribut2) VALUES(hodnota1, hodnota2);`

Příkaz SELECT – projekce

- Nejkomplexnější příkaz s nespočtem variací
 - Na velikosti písmen nezáleží, každý příkaz je ale nutno ukončit středníkem
- Jednoduchá projekce:
 - `SELECT id, username, jmeno FROM bloky; -- vyber konkrétní sloupce`
 - `SELECT * FROM bloky; -- vyber vše`
 - `SELECT count(*) FROM bloky; -- spočítej počet řádků`
 - `SELECT min(id), max(id) FROM bloky; -- zjistí nejmenší a největší id`
 - `SELECT DISTINCT jmeno FROM bloky; -- vypiš jména bez duplicit`

Příkaz SELECT – selekce

- Jednoduchá selekce:

- `SELECT id FROM bloky LIMIT 10; -- výpis id omez na prvních 10 řádků`
- `SELECT id FROM bloky WHERE box=1; -- vypiš id rezervací boxu č. 1`
- `SELECT * FROM bloky WHERE jmeno LIKE '%Petra%';`
`-- jen záznamy, které si rezervovala jakákoliv Petra`
- `SELECT id, username FROM bloky WHERE `from` BETWEEN '2015-01-01' and '2015-02-01';`
`-- jen záznamy s rezervací v lednu 2015`
- `SELECT * FROM bloky WHERE TIMEDIFF(`to`, `from`)>'2:00:00'`
`-- jen záznamy s délkou rezervace přes dvě hodiny`



Příkaz SELECT – řazení a agregace

- Ukázky řazení:

- `SELECT * FROM bloky ORDER BY `from`;` `-- dle data rezervace`
- `SELECT * FROM bloky ORDER BY box, `from`;` `-- dle boxu a data rezervace`
- `SELECT * FROM bloky ORDER BY `from` DESC;` `-- v obráceném pořadí`
- `SELECT * FROM bloky ORDER BY RAND();` `-- v náhodném pořadí`

- Ukázky agregace

- `SELECT box, count(*) FROM bloky GROUP BY box;` `-- počty dle boxů`
- `SELECT DAYOFWEEK(`from`) as Den, count(*) FROM bloky GROUP BY DAYOFWEEK(`from`);`
`-- dle dnů v týdnu, kdy 1=neděle`

Příkaz SELECT – spojování tabulek

- Propojení dvou tabulek

- `SELECT jmeno, tema FROM texty, student WHERE texty.username=student.username;`

- Propojení tří tabulek

- `SELECT zajimavost, jmeno, tema FROM hodnoceni, student, texty
WHERE hodnoceni.username=student.username AND hodnoceni.txt=texty.id;`

- Propojení čtyř tabulek

- `SELECT zajimavost, X.jmeno, Y.jmeno, tema FROM hodnoceni, student X, student Y,
texty WHERE hodnoceni.username=X.username AND texty.username=Y.username and
hodnoceni.txt=texty.id;`

Další možnosti příkazu SELECT

- Je možné zapisovat vnořené poddotazy
 - `SELECT count(*) from (SELECT ...) X;`
- Je možné upřesňovat podmínky agregací (klauzule HAVING)
- Dotazy je možné sjednocovat (UNION) a odečítat (MINUS) či dělat jejich průnik (INTERSECT)
 - Nemusí podporovat všechny databázové systémy
- Je možné používat funkce a zjednodušené řídicí struktury (IF ... ELSE ...)
- Je možné výsledek dotazu přímo uložit do jiné tabulky

Příkazy pro manipulaci s daty

- `INSERT INTO tabulka VALUES ()` – vložení dat
- `UPDATE tabulka SET ... WHERE ...` – aktualizace dat
- `DELETE FROM tabulka WHERE ...` – smazání dat
- `CREATE` – vytvoření tabulky nebo databáze
- `DROP` – zahození tabulky nebo databáze
- `LOAD DATA` – hromadný import dat ze souboru
- `GRANT` (a `REVOKE`) – nastavování (a odebrání) oprávnění k datům

Databázové systémy

- Oracle
- MSSQL
- PostgreSQL
- MySQL
- MariaDB
- SQLite

Možnost testování na <https://sqliteonline.com/>

- Pozor, relační databáze nejsou jedinou možností uchování dat
 - V současné době zažívají rozmach tzv. NoSQL databáze, které na to „jdou jinak“

Domácí úkol

Naformulujte dotazy v jazyce SQL

- Ve cvičné databázi na w3schools.com
- Spočítejte, kolik zákazníků je z USA
- Vypište měsíce, ve kterých přepravoval dopravce č. 3 (Federal Shipping)
- Vypište tabulku se jmény zákazníků a počtem jejich objednávek seřazený dle počtu objednávek od největšího
- Vypište produkty, kterých bylo objednáno od 40 do 60 kusů

SQL můžete zkusit např. také na <https://www.sql-practice.com/>