

Základy forenzních databází

Tereza Uhlíková

verze y25

Kdo jsem

Tereza Uhlíková

Ústav analytické chemie

skupina teoretické spektroskopie

místnost A277

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

tereza.uhlikova@vscht.cz

Heslo

Tell me and I forget,
teach me and I remember,
involve me and I learn.

Sylabus v SIS

1	17.9.	Úvod do databází, jejich účel a historie
2	24.9.	Architektura databází a informatika
3	1.10.	Programování - algoritmus, kódy, výroková logika
4	8.10.	Základy databází, klíče a kardinalita
5	15.10.	ERA model, normalizace
6	22.10.	Vyhledávání dat
7	29.10.	PŘEDNÁŠKA NENÍ - SKORO STÁTNÍ SVÁTEK
8	5.11.	Strojové učení pro vyhledávání
9	12.11.	Kvantové počítače pro práci s daty
10	19.11.	SQL Standardizovaný strukturovaný dotazovací jazyk
11	26.11.	MySQL, SQL Server, MS Access, pandas python
12	3.12.	Návrh relační databáze I
13	10.12.	
14	17.12.	Zkouška

Co bude předmětem

Data v chemii a forenzní vědě,
uložiště,
přenosová rychlost,
práce s daty,
algoritmizace, složitost
počítačové ukládání, řazení, vyhledávání,
základy stojového učení,
využívání AI,
kvantové počítače jako hrozba?

...

co doplníte vy...

S jakým očekáváním přicházíte?

Jen kredity?

Co se chcete dozvědět - jaké téma?

Co bude u zkoušky

- prezentace max 5 minut
- data a informace pro vaši bakalářku
- jaká data potřebujete a jaká databáze by se hodila pro vaši práci, jak by taková databáze mohla vypadat, jak by vám ML mohlo pomoci

Jiné předměty o databázích

Databázové systémy v chemii - ing. Sedláková Zuzana, Ph.D.

Databáze v chemické a forenzní analýze - doc. Raich Ivan, Ph.D.

Databázové systémy - ing. Valenta Michal, Ph.D.

Databázové systémy v chemii - doc. Kukal Jaromír, Ph.D.

Skupinová práce

Jaké znáte databáze?

Co je to databáze?

Co databáze obsahuje?

Proč databáze?

Forenzní: Jaká témata by forenzníky (tedy vás) zajímala?

Co je to databáze?

Co je to databáze?

“Definice” Databáze neboli datová základna (Data Base)

Co je to databáze?

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

= data i software

Databáze je utříděný souhrn souvisejících informací (Univerzita Princeton). Obsahuje: co možná kompletní data pro popis dané skutečnosti pro daný účel.

Proč databáze?

*Když už se něco dozvíme,
chceme to i později použít.*

Proč databáze?

*Když už se něco dozvíme,
chceme to i později použít.*

Jsou všude

Proč databáze?

*Když už se něco dozvíme,
chceme to i později použít.*

Jsou všude

Periodická tabulka prvků...telefonní čísla....Archivy počasí...Kartotéka ve zdravotnictví...firmy (Sigma Aldrich)... filmy (Netflix)...hry (Duna)...osobní uložště...ve všech sportech....sociální media...vládní organizace....google...

Proč databáze?

- Databáze poskytuje rychlejší přístup k datům než soubory.
- Databáze umožňuje přímý přístup k datům.
- Databáze má zabudovaný mechanismus pro paralelní přístup k datům.
- Databáze má zabudovaný systém uživatelských práv.
- Databáze umožňuje pomocí dotazů snadno extrahovat množiny dat, která vyhovují zadaným kritériím.

Kdo ji tvoří?

Kdo ji tvoří?

- vědec, forezník, “šéf” - tvoří myšlenku
- “pomocník šéfa” - Ph.D. student - tvoří plán
- podezřelý, vyšetřovatel - dodává data
- programátor - vytvoří databázi podle návrhu za pomoci databázového systému
- IT architekt - dokončí navrhovanou databázi pro dostupný hardware

Správný návrh databáze je velmi abstraktní a velmi složitá záležitost.

Co je potřeba k vytvoření databáze?

Co je potřeba k vytvoření databáze?

- myšlenka - odpovědi na otázky
- programovací jazyk pro komunikaci s počítačem
- hardware k uskladnění dat
- databázové prostředí k tvorbě
- znalost vyhledávání
- informace, data
- logické propojování informací - hlubší znalost tématu
- syntaxe a sémantika - struktura relačních databází
- objekty pro práci nad databází

Od počátku

Začněme pěkně od začátku.

Kdysi dávno

Kdysi dávno

12st.př.nl. snaha o sběr dat a
vytvoření abecedy - Ugarit
(Sírie) ~ 1500 tabulek
(dopisy, diplomacie, básně,
hebrejská bible...)



Kdysi dávno

12st.př.nl. snaha o sběr dat a vytvoření abecedy - Ugarit (Sírie) ~ 1500 tabulek (dopisy, diplomacie, básně, hebrejská bible...)



475př.n.l. stal Řím republikou - snaha o třídění společnosti (Dějiny psané Římem)

...



První přírodovědecká databáze

První přírodovědecká databáze

1729 první kartotéka - Carl Linnaeus



První mechanizování

- 1890 prvním elektromechanické zpracování dat - Herman Hollerith -> International Business Machines (IBM)
- 1935 v USA uzákoněna (Social Security Act) nutnost vedení informací o cca 26 milionech zaměstnancích



Rychlý vývoj elektroniky

- 1951 UNIVAC - první digitální počítač pro komerční využití - Electronic Discrete Variable Automatic Computer (EDVAC)
- 1959 měl Pentagon již více než 200 počítačů
- 1960 Common business-oriented language (COBOL-60) - programovací jazyk pro databáze
- 1961 magnetické disky - nesériový přístup k datům



```

Command ==>
=COLS> -----1-----2-----3-----+-----
***** ***** Top of
000100 000100 IDENTIFICATION DIVISION.
000200 000200 PROGRAM-ID. HELOWOLD.
000300 000300 ENVIRONMENT DIVISION.
000400 000400 PROCEDURE DIVISION.
000500 000500 DISPLAY 'HELLO WORLD'.
000600 000600 STOP RUN.
***** ***** Bottom o
  
```

Relační databáze

1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data
- 1980 Sybase, Informix, dBASE

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data
- 1980 Sybase, Informix, dBASE
- 1980 První SQL databáze (Oracle pro počítače VAX)

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data
- 1980 Sybase, Informix, dBASE
- 1980 První SQL databáze (Oracle pro počítače VAX)
- 1985 Ingres -> Postgres (relačně-objektová DB)

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data
- 1980 Sybase, Informix, dBASE
- 1980 První SQL databáze (Oracle pro počítače VAX)
- 1985 Ingres -> Postgres (relačně-objektová DB)
- 1996 studenti Jolly Chen a Andrew Yu -> PostgreSQL

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data
- 1980 Sybase, Informix, dBASE
- 1980 První SQL databáze (Oracle pro počítače VAX)
- 1985 Ingres -> Postgres (relačně-objektová DB)
- 1996 studenti Jolly Chen a Andrew Yu -> PostgreSQL
- 1997 David Axmark, Allan Larsson and Finnish Michael "Monty" Widenius - MySQL

Relační databáze

- 1970 Edgar Frank „Ted“ Codd publikoval článek „A Relational Model of Data for Large Shared Data Banks“ -> použít relačního kalkulu a algebry i pro netechnické uživatele při ukládání a manipulaci s daty.
- 1971 zavedeny pojmy - integrita dat, datový model, schéma databáze, atomita, entita ... tedy relační model
- 1976 Donald D. Chamberlin a Raymond F. Boyce - Structured English Query Language (SEQUEL)
- 1977 Interactive Graphics and Retrieval System (Ingres) - systém pro geografická data
- 1980 Sybase, Informix, dBASE
- 1980 První SQL databáze (Oracle pro počítače VAX)
- 1985 Ingres -> Postgres (relačně-objektová DB)
- 1996 studenti Jolly Chen a Andrew Yu -> PostgreSQL
- 1997 David Axmark, Allan Larsson and Finnish Michael "Monty" Widenius - MySQL

Používané dnes

1985 objektové databáze

Používané dnes

1985 objektové databáze

2000 Extensible Markup Language (XML) - pro velké objemy dat,
NoSQL - nestrukturované databáze (BigTable, Dynamo)

Používané dnes

1985 objektové databáze

2000 Extensible Markup Language (XML) - pro velké objemy dat,
NoSQL - nestrukturované databáze (BigTable, Dynamo)

2020 velký opravdu velmi velký objem dat + autorství, GDPR

Používané dnes

1985 objektové databáze

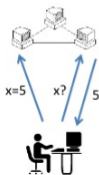
2000 Extensible Markup Language (XML) - pro velké objemy dat,
NoSQL - nestrukturované databáze (BigTable, Dynamo)

2020 velký opravdu velmi velký objem dat + autorství, GDPR

problém s CAP teorémem - Consistency, Availability, a Partition tolerance

CAP Theorem

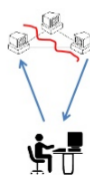
Consistency



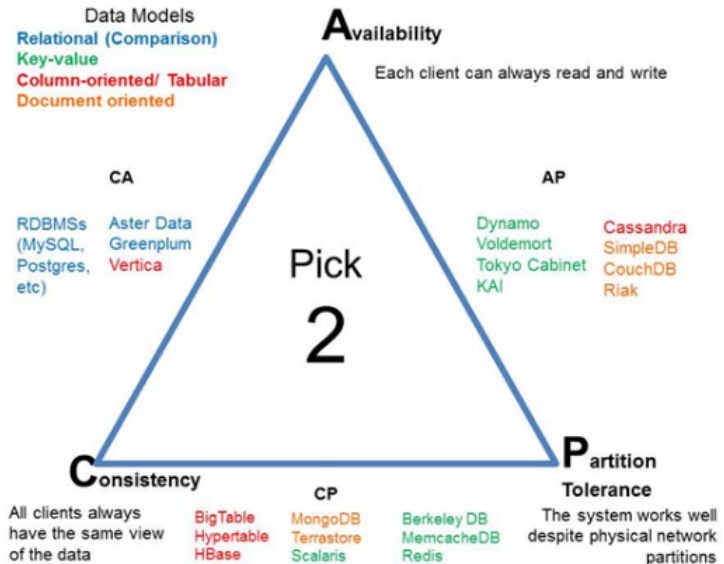
Availability



Partition tolerance



CAP teorem



Skripta, stránky a materiály

Berg, Kristi & Seymour, Dr. Tom & Goel, Richa. (2012). "History Of Databases". International Journal of Management & Information Systems (IJMIS). 17. 29. 10.19030/ijmis.v17i1.7587.

https://en.wikipedia.org/wiki/Carl_Linnaeus

Gilbert, Seth; Lynch, Nancy (2002). "Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services". ACM SIGACT News. Association for Computing Machinery (ACM). 33 (2): 51–59. doi:10.1145/564585.564601

obrázky převzaty z:

https://en.wikipedia.org/wiki/Ugaritic_texts

https://en.m.wikipedia.org/wiki/File:Roman_Empire_Trajan_117AD.png

<https://en.m.wikipedia.org/wiki/File:HollerithMachine.CHM.jpg>

https://cs.wikipedia.org/wiki/Soubor:Grace_Hopper_and_UNIVAC.jpg

<https://www.ibmmainframer.com/cobol-tutorial/cobol-hello-world/>

<https://www.mysoftkey.com/architecture/understanding-of-cap-theorem/>

<https://java-questions.com/cap-theorem.html>