

Temporální databáze

Z FITwiki

(Přesměrováno z Temporální DB)

Základní pojmy

Relační databáze

$(D, =, r_1, \dots, r_n)$

- (D, ρ) nad $(D, =)$ a schématem ρ
- **D - doména neinterpretovaných konstant**
- **konečná instance ρ nad D** (množina n-tic, tedy tabulka)

Relační algebra

- standardní množinové operace (průnik, sjednocení, rozdíl, kartézský součin)
- relační operace
 - selekce (výběr řádků)
 - projekce (výběr sloupců)
 - spojení (join tabulek)
 - dělení (všechny prvky A, které jsou v relaci s některým prvkem B)

Relační kalkul

- dotazovací jazyk na bázi logiky
- skládá se z
 - proměnných, konstant
 - odkazů na atributy
 - predikátových symbolů

Doporučuju přečíst [1]

(http://www.fit.vutbr.cz/study/courses/DSI/public/pdf/nove/4_2.pdf)

Obsah

- 1 Základní pojmy
- 2 Čas
 - 2.1 Časová doména
 - 2.2 Modely času
 - 2.2.1 Snímkový model času (Snapshot)
 - 2.2.2 Model času platnosti (časová razítka)
- 3 Spojení času s databází
 - 3.1 Snímková tabulka (snapshot)
 - 3.2 Tabulka s platnými časy
 - 3.3 Transakční tabulka
 - 3.4 Tabulka obojího času (bitemporální)
- 4 Techniky
 - 4.1 Kódování intervalů
 - 4.2 Shlukování (Coalescing)
 - 4.3 Rozdělování (partitioning)
 - 4.4 Odsávání dat
 - 4.5 Indexování
 - 4.6 Temporální integritní omezení
- 5 Dotazovací jazyky
 - 5.1 Výroková temporální logika (prvního řádu)
 - 5.2 Temporální relační kalkul
 - 5.3 Jazyky temporálních DB
 - 5.4 Jazyky s větší silou
- 6 TSQL2
 - 6.1 Časový model
 - 6.2 Druhy tabulek
 - 6.3 Dotazy
 - 6.4 Verzování schématu
- 7 Problémy

Temporální databáze

- **Databáze s časovou dimenzí.**
- Typické použití: bankovníctví, pojišťovnictví, účetnictví, medicína, katastrální

Čas

- **Čas platnosti** udává, po který interval jsou data platná (pravdivá v modelu)

- **Čas transakce** udává, kdy byla data přítomna v DB (vložena)

Časový okamžik

bod na časové ose

Časový úsek (interval)

doba mezi dvěma okamžiky, orientované časové trvání

Trvání

časový úsek se známou délkou, ale neznámým začátkem či koncem

- dopředné (pozitivní)
- zpětné (negativní)

Časový element

konečné sjednocení časových úseků

Časová doména

$(T, <)$ (lineárně uspořádaná, neomezená množina)

- \mathbb{N}, \mathbb{Z} - diskretní čas, každý bod v čase má jednoho předchůdce
- \mathbb{Q} - čas s "hustotou", mezi každými dvěma body existuje další
- \mathbb{R} - spojitý čas

Nelineární časové domény

jsou pouze částečně uspořádané, podporují větvení času (verzovací systém)

Modely času

Snímkový model času (Snapshot)

- Je blíže výrokové TL
- každá DB (nie každý snapshot??) popisuje stav světa v konkrétním časovém okamžiku. Relace uspořádání potom definuje tok času.
- Temporální DB je potom funkce typu
 - $T \rightarrow DB(D, \rho)$ (každý čas se zobrazí na některý snímek databáze ve kterém je stav DB k tomuto času)
 - datové typy $T \rightarrow (D_n \rightarrow bool)$
- Snapshot model není moc vhodný na dotazy typu

$$\{t : DB \models \varphi(t)\}$$

(tj. všechny okamžiky, kdy byla podmínka v rámci DB platná - musel by se kontrolovat každý snímek databáze).

Model času platnosti (časová razítka)

Časové razítko

- atomické hodnoty bez vnitřní struktury
- základní jednotka **Chronon** (kvantum času)
- pro každé $r_i \in \rho$ definujeme relaci R_i tak, že

$$R_i = \{(t, a_1, \dots, a_k) \mid (a_1, \dots, a_k) \in r_i \text{ in } D\} \text{ (prosté řádky s časovým razítkem)}$$

Temporální databáze s časovými razítky je pak

$$(D, =, T, <, r_1, \dots, r_n)$$

- **konečná instance ρ nad D**
- **časová doména**
- **konečná instance ρ nad D, T**
- **Pozn.: Intervalová reprezentace (od-do) je jen otázkou kódování, nikoliv formální definice**
- Výhoda modelu s razítky – **dotaz na časové okamžiky je realizovatelný**
 - čas je “jen” jedním z typů
 - ale je součástí daného uspořádání
 - datový typ: $(T \times D^n) \rightarrow \text{bool}$ (tj. ke každému řádku je přidáno ještě časové razítko z T)

Pozn.: Oba modely (razítka, snímky) jsou ekvivalentní!

Spojení času s databází

Sémantika temporální DB je nezávislá na implementaci

Snímková tabulka (snapshot)

- zaznamenává stav dat v jistém okamžiku
- Relace uspořádání nad těmito snímky tvoří tok času... **historii** (posloupnost stavů).
- Historie databáze je také snímkový model.
- Problém, pokud se dotazuje na "všechny okamžiky, kdy platilo, že ...".
- Příklad:
 - 1990 Asistent
 - 1995 Odborný asistent
 - 2000 Docent

Tabulka s platnými časy

- obsahuje kromě datových sloupců ještě sloupec pro platnost (možnost i nekonečno)
- možno modifikovat
- Příklad: Databáze obsahuje jeden záznam na každou změnu:
 - <1990 - 1995) Asistent
 - <1995 - 2000) Odborný asistent
 - <2000 - forever) Docent

Transakční tabulka

- obsahuje pouze časy transakcí
- data vypadají stejně jako pro tabulku platného času, ale rozdíl je v tom, že se nejedná o dobu, kdy data platila, ale kdy se vložila do DB
- Pouze připojuje data na konec

Tabulka obojího času (bitemporální)

- ukládá **čas platnosti i čas transakce**
- pouze připojuje -> Nemění údaje od/do, ale přidá nový záznam s opravenými hodnotami.

- k historii navíc obsahuje historii změn (tabulka platného času + 2x tabulka transakce (čas vytvoření, čas zneplatnění))

Techniky

Kódování intervalů

- Necht' T_p je časová doména.
- Definujeme množinu intervalů

$$I(T) = \{(a, b) | a \leq b, a \in T \cup \{-\infty\}, b \in T \cup \{\infty\}\}$$

- Relace na množině $I(T)$
 - $([a, b] <_{--} [a', b']) \Leftrightarrow a < a'$ (tj. průnik je od a' **doprava** teoreticky až do nekonečna)
 - $([a, b] <_{-+} [a', b']) \Leftrightarrow a < b'$ (tj. průnik je **úsečka** $a - b'$)
 - $([a, b] <_{+-} [a', b']) \Leftrightarrow b < a'$ (tj. není **žádný průnik**)
 - $([a, b] <_{++} [a', b']) \Leftrightarrow b < b'$ (tj. průnik je od b **doleva** teoreticky až do nekonečna)

Intervalová časová doména vzhledem k T_p

$$T_I = (I(T), <_{--}, <_{-+}, <_{+-}, <_{++})$$

Shlukování (Coalescing)

- jednorozměrná temporální relace obsahuje shluky pokud je každý fakt spojovaný s nejvýše konečným počtem nepřekrývajících se intervalů
 - je potřebné zaručit, pokud se nad relacemi vykonávají ne-logické operace
- Někdy je potřeba spojit intervaly kvůli dalšímu zpracování (například při selekci sloupců)
- Intervaly, které se překrývají (nebo na sebe navazují bez mezery) a nesoucí stejná data je možné spojit (např. každoroční výkazy při omezení pouze na sloupec adresa).
- zjednodušují selekci, projekci, spojení
- nezjednodušují temporální operace
- Tj místo abychom měli dva řádky:

(Hodnota1, [1,6])

(Hodnota1, [5,7])

dostaneme jediný shluknutý řádek:

(Hodnota1, [1,7])

- Pozn.: pro více rozměrů času je shlukování nejednoznačné -> problém
- Generičnost dotazů = dotaz je generický, pokud jeho výsledek nezávisí na způsobu uložení dat v DB:

pokud je v db uložena fakta $(a, [0,3])$ (a platí od 0 do 3) nebo $(a, [0,2]), (a, [1,3])$, tak v první případě se jedná jen o shluknutí intervalů pro fakt a
 ale pro dotaz: $\exists i, j. \exists x (R(i, x) \ \&\& \ R(j, x)) \ \&\& \ i \neq j$ platí v druhém případě, ale v prvním jen díky shluknutí ne

Rozdělování (partitioning)

- rozdělení intervalů na několik intervalů o maximální délce x
- otázkou je v kterých bodech dělit

- v podstatě opak seskupování

Odsávání dat

- Data s časem transakce pořád rostou a mohou přerůst datový prostor (objem dat versus potřeby archivace)
- Data platná v okamžiku odsávání se pochopitelně neruší
- Odsávání odstraňuje stará, pravděpodobně zastaralá a nekorektní data

Indexování

Problém indexace

je problém, chceme-li například najít intervaly, které mají neprázdný průnik s jiným

- R-Tree (viz prostorové DB)
- AP-Tree (index jen pro přidávání dat, ISAM a B+-tree)
- Time Index
 - na začátku a konci intervalu uloží seznam dalších intervalů, které obsahují tento okamžik
 - nad těmito body postav B-Tree

Temporální integritní omezení

Temporální integritní omezení

uzavřené formule prvního řádu temporálního dotazovacího jazyka

- **Použití integritních obmezení**

zachycení sémantiky DB aplikace

důvod zavedení je ukládání len "významných" dat

návrh DB - normální formy - dobrá schémata bez anomálií a dobrá dekompozice

- Historie H splňuje omezení O jestliže je O pravdivé v každém stavu H
- Konečná historie H potenciálně splňuje omezení O jestliže může být rozšířena do nekonečné historie tak, že splňuje O
- Důvod zavedení: ukládání pouze "významných" dat

Dotazovací jazyky

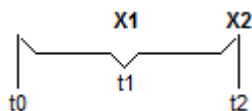
Nejsou standardizovány.

Výroková temporální logika (prvního řádu)

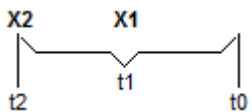
- výroková logika prvního řádu rozšířená o
 - časové spojky
 - časové proměnné a kvantifikátory
- Temporální logika TL(since, until) je shodná s TL(until) nad úplným lin. uspořádáním omezeným v minulosti
- Nemá vlastnost oddělení z výrokové logiky

Časové/temporální spojky

- **X1 until X2** (X1, X2 - predikáty, tX - časy)



■ X1 since X2



- $\diamond X$ - někdy v budoucnosti
- $\blacklozenge X$ - někdy v minulosti
- $\square X$ - vždy v budoucnosti
- $\blacksquare X$ - vždy v minulosti
- $\circ X$ - v příštím kroku (pro diskrétní čas)
- $\bullet X$ - v předchozím kroku (pro diskrétní čas)

Temporální relační kalkul

- **relační kalkul** rozšířený o výrokovou temporální logiku

Jazyky temporálních DB

- je možná textová reprezentace (na rozdíl od např. prostorových DB)
- **TQUEL** - rozšíření jazyka QUEL o čas platnosti, čas transakce, kódování intervalů a shluky
- **TSQL, TSQL2** - rozšíření SQL o čas platnosti, čas transakce, časová razítka (není formální sémantika) - více níže
- **SQL/TP** - rozšíření SQL o nový datový typ, intervaly
- HRDM, IXRM, ATSQL, ...

Jazyky s větší silou

- podporují více časových dimenzí (ETL, μ TL, TempLog)
- časové spojky vyššího řádu (množiny a podobně)

TSQL2

- rozšíření SQL
- Typy DATE, TIME, DATETIME, INTERVAL (z SQL92), **PERIOD** (doba trvání, přidává TSQL2)
- **Systém obojího času** (platnost a transakce), podporuje všechny typy tabulek.
- Časová neurčitost (souvisí se zrnitostí) -- asi na konci září.
- **Odsávání dat**
 - probíhá asynchronně po příkazu ALTER
 - Nechcené, ale stále platné hodnoty mohou být odstráněny pomocí příkazu DELETE
- **Zástupci (surrogates)** - klíče jednoznačně identifikující jeden časový záznam (datový typ sloupce)
 - jedinečná hodnota, která může být testovaná na rovnost, ale jinak nie je viditeľná
 - vhodné tam, kde je třeba identifikovat objekty a přitom klíč je časovo závislý
 - jedná se o datový typ pro sloupec tabulky
- **Agregační funkce** a seskupování může být použito na časové údaje
 - aplikované na každý snímek v mezivýsledku
 - je možná optimalizácia vyhodnotenia, lebo viac riadkov môže niesť rovnaké hodnoty
- Podporuje partitioning

Doporučuju projít [2] (<http://aov.webzdarma.cz/tsql2.ppt>)

Časový model

- Lineární časový model oboustranně omezený
- Nerozlišuje mezi diskrétními, hustotními a spojitými modely
- Diskrétní časová osa, nejmenší jednotka je **chronon** (tik hodin).
- Možné nastavit různou granularitu času (a přecházet).

Časová zrnitost (granularita)

- Hrubost časového dělení (minuty, hodiny, dny, ...) - tj s jakou tolerancí se časy považují za stejné
- operandy predikátů musí mít stejnou zrnitost
- zrnitosti se sdružují do kalendářů

Druhy tabulek

- **Snímek** (snapshot relation) - neukládá žádný čas
- **Platný čas(stav)** - AS VALID [STATE] <zrnitost> (ukládají čas od-do)
- **Čas události** - AS VALID EVENT <zrnitost> (ukládají mžikové události)
- **Čas transakce** - AS TRANSACTION
- **Oba druhy času (stav)** - AS VALID [STATE] <zrnitost> AND TRANSACTION
- **Oba druhy času (událostní)** - AS VALID EVENT <zrnitost> AND TRANSACTION
 - události dostávají časové razítka v podobě množin okamžiků
 - každý řádek udává určitou okamžikovou událost', časové razítko přiřazené řádku hovorí, kedy daná událost' prebehla
 - tieto tabuľky môžu byť spojené s časom transakcie
- Konvenční tabulky obdržíme z časových uvedením klíčového slova SNAPSHOT (vynechají se časové sloupce)

Dotazy

- SELECT SNAPSHOT ... - vše nyní i v minulosti (bez časů)
- SELECT ... - záznamy spojené s jedním nebo více časovými úseky
- SELECT VALID(A) ... - součástí budou sloupce s časovými razítky s časem platnosti pro A

Verzování schématu

- TSQL narozdíl od SQL92 podporuje
- schéma se stane množinou tabulek s časem transakce
- Návrat k datu SET SCHEMA DATE <čas>

Problémy

Problémy s ukládáním

- **intervaly a skutečné intervaly**
 - interval ako kódovanie množín časových okamžiků
 - vs.
 - interval ako body v 2-rozmernom priestore
- **zhluky** zjednodušujú selekciu, projekciu a spojenie ale nezjednodušujú temporálne operácie

Zlyhanie zhlučovania

- nejednoznačné zhlučovanie pre viac rozmerov
- FO-úplný dotazovací jazyk sa neobíde bez n-rozmernosti v dotazoch
- dotazy sú závislé na reprezentácii - negenerické dotazy

Další problémy

- dělení intervalu (jaký bod vybrat)
- dotazy jsou optimalizovány odhadem, paralelizace není jednoduchá
- problém integrity, když se historie nemůže měnit
- databáze jsou velmi prostorově náročné

Citováno z „[http://wiki.fituska.eu/index.php?](http://wiki.fituska.eu/index.php?title=Tempor%C3%A1ln%C3%AD_datab%C3%A1ze&oldid=13104)

[title=Tempor%C3%A1ln%C3%AD_datab%C3%A1ze&oldid=13104](http://wiki.fituska.eu/index.php?title=Tempor%C3%A1ln%C3%AD_datab%C3%A1ze&oldid=13104)“

Kategorie: Pokročilé databázové systémy | Státnice 2011

- Stránka byla naposledy editována 25. 12. 2015 v 20:46.