

PA152: Uložení dat

Implementace databázových systémů

Pavel Rychlý

pary@fi.muni.cz

3. října 2008

Obsah

- 1 Datové elementy
- 2 Záznamy
- 3 Organizace bloků
- 4 Příklady z praxe

Formy uložení dat

- Co chceme ukládat?
 - položky
 - záznamy/objekty
 - bloky
 - soubory
- Kam ukládat?
 - disk
 - (operační paměť)
- Co je důležité
 - rychlost čtení z disku do paměti

Obsah

- 1 Datové elementy
- 2 Záznamy
- 3 Organizace bloků
- 4 Příklady z praxe

Uložení datových elementů

- Co chceme ukládat?
 - jméno
 - plat
 - datum
 - obrázek
- Kam ukládat?
 - posloupnost bajtů

Typy datových elementů

- celé číslo
 - 2/4 bajty
 - proměnlivá bitová délka (kódy Elias, Golomb)
- reálné číslo
 - plovoucí čárka
mantisa + exponent
 - pevná čárka
- znaky
 - kódová stránky ASCII/UTF

Typy datových elementů

- datum
 - počet dní od "počátku"
 - řetězec YYYYMMDD
- čas
 - počet sekund od půlnoci
 - řetězec HHMMSSFF
- výčtové typy
 - očíslování hodnot
 - red → 1, green → 2, blue → 3, yellow → 4, ...

Typy datových elementů

Pravdivostní hodnota

- True 11111111
- False 00000000
- Použití méně než 1 bajtu?
 - s rozmyslem
 - (velké) pole pravdivostních hodnot → bitové pole

Typy datových elementů

Řetězec znaků

- pevná délka
 - omezení velikosti
- proměnlivá délka
 - uložená délka
 - ukončení nulou
 - nutnost číst celé
 - nelze uložit nulu v textu

Uložení datových elementů

- většinou pevná délka
 - každý typ má svoji bitovou reprezentaci
- proměnlivá délka
 - velikost na začátku
- každý element "má" svůj typ
 - interpretace bitů
 - velikost
 - speciální hodnota "neobsazeno" (NULL)

Obsah

1 Datové elementy

2 Záznamy

3 Organizace bloků

4 Příklady z praxe

Uložení záznamů

- řádkové uložení tabulky
(záznam = řádek)
- pevný formát
 - schéma je uloženo mimo záznamy
 - každý záznam na stejný počet bajtů
- proměnlivý formát
 - každý záznam obsahuje svoje schéma

Proměnlivý formát záznamu

- použití
 - "řídké" záznamy – většina položek s hodnotou NULL
 - opakování položek stejného typu
 - vyvíjející se formát – změny schématu během života databáze
- varianty kombinující pevný a proměnlivý formát

Uložení objektů

- Objekty/třídy
 - základní vlastností je zapouzdření (nevidíme do vnitřní struktury)
- databázové algoritmy pracují se strukturou DB
 - víme jakých typů jsou v tabulkách položky
- objekty nepatří do databáze

Sloupcové uložení tabulky

Místo záznamů (řádků) ukládáme sloupce

- zvláštní soubor (pole hodnot) pro každý atribut
- provázáno pomocí ID/pořadí
- velice výhodné pro zpracování omezené na několik atributů
- složitější aktualizace (mazání, vkládání)
- lze řešit na logické úrovni rozdělením tabulky na několi menších

Obsah

- 1 Datové elementy
- 2 Záznamy
- 3 Organizace bloků
- 4 Příklady z praxe

Uložení záznamů do bloků

- záznamy
 - pevné délky
 - proměnné délky
- bloky pevné velikosti

Uložení záznamů do bloků - možnosti

- oddělení záznamů
- souvislé bloky
- bloky se záznamy různých typu
- rozdělení záznamu
- uspořádání záznamů
- odkazy na záznamy

Oddělení záznamů

- identifikace začátku a konce záznamu
 - značky
 - adresy
- záznamy pevné délky
 - začátek pole záznamů
 - počet obsazených prvků pole

Souvislé/nesouvislé bloky

- nesouvislé
 - každý záznam součástí jednoho bloku
 - jednodušší, ale může plýtvat místem
- souvislé
 - záznam může začínat na konci jednoho bloku a pokračovat na začátku dalšího bloku
 - nutné pokud velikost záznamu je větší než velikost bloku
 - musíme udržovat dané pořadí bloků (alespoň logicky)

Míchání záznamů

- blok obsahuje záznamy různých typů
sdužování několika tabulek
- záznamy, ke kterým často přistupujeme dohromady,
jsou uloženy ve stejném bloku
 - Příklad: student + studium
- hlavní nevýhoda: složitější struktura
- kompromis: bez míchání, ale bloky se souvisejícími
záznamy jsou na stejném stopě

Rozdělení záznamů

- kombinace položek s pevnou a proměnlivou délkou
- rozdělení do různých bloků
 - blok s pevnou částí
 - blok s proměnlivou částí
 - obrázky
 - velká data, která DBMS neinterpretuje
(neumí indexovat, vyhledávat)

Uspořádání záznamů

- záznamy jsou v souboru (a bloku) setříděny podle hodnot klíče
- efektivní čtení záznamů v daném pořadí
 - např. pro merge-join
- způsob uspořádání
 - záznamy v daném pořadí
 - seznam propojený ukazateli
 - oblast přetečení

Odkazy na záznamy

- Jak ukládat odkazy (ukazatele) na záznamy?
- fyzická adresa
 - ID bloku
 - (č. zařízení, cylindru, stopy, bloku)
 - offset v bloku
- nepřímo
 - převod: ID záznamu → fyzické adresa
 - ID záznamu
 - libovolná posloupnost bitů

Odkazy na záznamy - výhody

- fyzická adresa
 - cena přístupu
- nepřímo
 - jednoduchost použití

Modifikace záznamů

- vkládání
- mazání
 - správa volného místa
- aktualizace
 - stejná velikost – na místě
 - jiná velikost – smazání + vložení

Vkládání záznamů

- bez uspořádání
 - vkládáme na konec nebo do volného místa po smazaném záznamu
 - problémy alokace místa pro záznamy s proměnlivou velikostí
- uspořádané záznamy
 - pokud je "blízko" volné místo, přeuspořádáme
 - jinak – oblast přetečení

Mazání a odkazy

- mazání záznamů
 - visící ukazatele ukazují na neplatné místo
- náhrobky
 - ponecháme značku v převodní tabulce nebo na staré fyzické adrese
- ID záznamu v záznamu
 - při přechodu přes ukazatel testujeme shodu ukazatele a ID

Správa vyrovnávací paměti

- obecné strategie uvolňování vyrovnávací paměti
 - nejdéle nepoužitý
- přípíchnuté bloky
 - trvale umístěny ve vyrovnávací paměti

Prohazování odkazů

- blok ve vyrovnávací paměti
 - odkazy ukazují do paměti místo na disk
- automatické
 - při načtení do paměti se všechny ukazatele prohodí
- na žádost
 - prohození při prvním použití odkazu
- žádné

Obsah

- 1 Datové elementy
- 2 Záznamy
- 3 Organizace bloků
- 4 Příklady z praxe

Komprese

- vhodně zvolená komprese může zvýšit rychlost přístupu
- neexistuje univerzální metoda/algorithmus
- hodně záleží na způsobu použití (náhodný/sekvenční přístup)
- musíme znát strukturu dat a vhodně zvolit kompresní metody

Bigtable

Distribuovaný systém pro uložení a správu obrovských dat

- PB (=1000 TB) dat
- tisíce serverů
- místo tabulky: mapování (řádek, sloupec, čas) → řetězec
- bez omezení počtu sloupců/řádků
- identifikátory řádků/sloupců = řetězce (do 64kB)
- rodiny sloupců – optimalizace přístupu
- časové značky a automatické mazání "starých" dat

Hadoop Distributed File System

- Distribuovaný systém souborů na běžných strojích
- předpoklady a cíle:
 - výpadky hardware
 - sekvenční přístup, write-once-read-many
 - NameNode, DataNodes
 - volitelný stupeň replikace