

PA152

Implementace
databázových systémů

Pavel Rychlý

pary@fi.muni.cz

Laboratoř zpracování přirozeného jazyka

<http://www.fi.muni.cz/nlp/>

Technické informace

<http://www.fi.muni.cz/~pary/pa152/>

- přednáška
- ukončení písemnou zkouškou
- příklady k samostudiu
- projekty

Zkouška

- písemná
 - asi 8 převážně praktických příkladů
 - maximum 50 bodů
 - za projekt(y) je možné získat až 40 bodů
- hodnocení:
 - 43 a více bodů -- A
 - 42--37 bodů -- B
 - 36--32 bodů -- C
 - 31--28 bodů -- D
 - 27--25 bodů -- E

Projekty

- samostatná práce (příp. skupina max. 3 lidí)
- výsledkem písemná zpráva
 - (+ funkční programy)

- Techniky používané v DBMS (20)
- Sada testovacích otázek (5)
- Názorné ilustrace k přednášce (5)

Obsah přednášky

- H. Garcia–Molina, J.D. Ullman, and J. Widom.
- Database System Implementation
- Prentice Hall, New Jersey, 2000

- Sig: D89

Předpoklady:

- relační model DB
- (SQL)

Úvod

- "Databáze"
 - programuje většina programátorů
 - potřebuje každá firma
 - je součástí většiny aplikací
- Databázi obstarává DBMS
 - Database Management System
 - Database Engine
 - databázový stroj
- PA152
 - jak naprogramovat databázový stroj

Základní pojmy

Data

- jakékoliv údaje, které chceme uchovávat

Databáze

- vnitřně strukturovaný soubor dat
- v konzistentním stavu

Databázový stroj (DBMS)

- systém zajišťující správný chod databáze

Relační model databáze

□ Databáze

- sada relací a pravidel

□ Relace

- tabulka

▷ obsahuje řádky (záznamy), sloupce (atributy)

□ Schéma

- popis struktury

□ Normální formy, relační algebra/kalkul

Vlastnosti DBMS

Programming interface (API)

- aplikační rozhraní

Persistent storage

- trvalé uchování dat

Transaction management

- řízení transakcí
- současný přístup více uživatelů/procesů

Části DBMS

Storage Manager

- správa bloků na disku
- správa vyrovnávací paměti

Query Processor

- překlad dotazu
- vyhodnocení dotazu

Transaction Manager

- atomičnost, izolovanost a trvalost transakcí

Co se nedozvíte

- jak "programovat databáze"
- návrh schématu databáze
- tvorba pravidel/omezení
- tvorba dotazů
- návrh formulářů/sestav

Ale:

- vždy je dobré vědět, jak věci fungují

Uložení dat

- databáze obsahují velké množství dat
- potřebujeme trvalé uložení

kam ukládat

jak ukládat

Hierarchie pamětí

- vyrovnávací
 - procesor
- hlavní/operační
 - RAM
- sekundární
 - disk
- záložní
 - pásy, optické disky

Vlastnosti pamětí

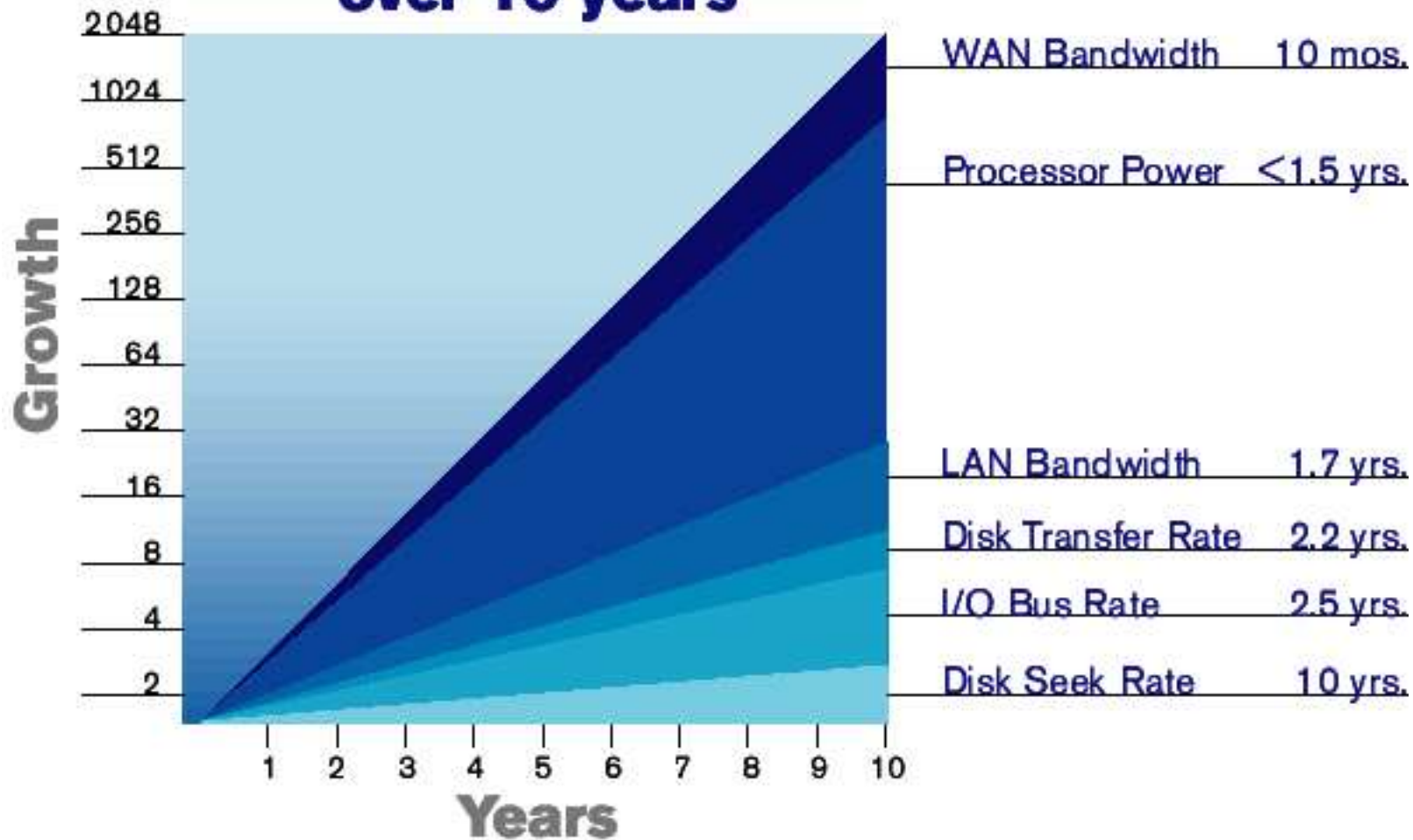
- větší kapacita → menší rychlost
- větší rychlost → menší kapacita

- Mooreův zákon
 - dvojnásobek za každých 18 měsíců
 - rychlost procesoru
 - velikost operační paměti za jednotnou cenu
 - velikost disku za jednotnou cenu
 - ale: NE rychlost disku

Mooreův zákon

Exponenciální růst

Technology Growth Rates over 10 years



Základní uložení dat

trvalé uložení dat

trvalost transakcí

→ využití trvalé paměti

disk

Disk

□ Charakteristiky

- hlavičky, povrchy, cylindry, stopy, bloky
- náhodný přístup

□ Vlastnosti disku

- náhodný přístup je řádově pomalejší než jakýkoliv výpočet i sekvenční přístup
- atomickým prvkem je blok ~ 4 kB

- → algoritmy optimalizujeme na počet náhodných přístupů na disk
- → zpracováváme celý blok

Diskové operace

- čtení bloku
- zápis bloku

- modifikace bloku
 - čtení bloku
 - modifikace v paměti
 - zápis bloku
 - (kontrola – čtení bloku)

Příklady algoritmu

Vyhledání

Třídění

- SELECT ... ORDER BY ...
- nejčastější netriviální operace DBS

Vyhledání

sekvenční

- $O(n)$

binární v uspořádaném poli

- $O(\log n)$

- ze skoro každého bloku použijeme jenom zlomek

Třídění

QuickSort (a jiné)

- předpokládají data v operační paměti

MergeSort

- třídění sléváním
 - rekurzivní dělení na menší posloupnosti
 - slévání setříděných postí
- $O(n \log n)$
- v praxi mnohem pomalejší