Registry (hardware); primární (operační) paměť (software); sekundární paměť (disk); magnetické pásk, CD, …
Posbíraná data -> vyčištěná data -> použitelná data
Data – logický pohled (jaké atributy...); fyzický pohled (jak jsou řešeny); implementace (jak jsou umístěny)
Délka záznamu R, velikost bloku B; $B/R>1$ snaha o celočíselné, =0 ideální, <1 přerostlé záznamy – problém

Magnetické pásky – 9 stop z 1 strany (8 data, 1 parita) mezi bloky prázdno, abych se nemusel vracet na konci
Drahé, dlouho vydrží, rychlé na sekvenční přístup, proudová data, zálohování; desítky GB za hodinu
Magnetický disk – hlavičky v zákrytu, aktivní jen 1 (posuny při zahřívání); tisíce ot./ min.; až 1TB
parametry s(eek) – vystavení hlavy; r(otation delay) – polovina otáčky disku; b(lock)t(ransfer)t(ime) přenos
Diskové pole RAID – redundantní pole pevných disků; ke zvýšení kapacity/ rychlosti/ bezpečnosti...
RAID 0 – střídavý zapis na jednotlivé disky – při zničení jednoho ztratím vše; rychlý zápis (do obou naráz)
RAID 1 – stejná data na oba disky – pomalý zápis, rychlé hledání; nejvíc plýtvání/ nejbezpečnější
RAID 0+1 – levné souborové systémy; kombinace 0, 1; alespoň 4 disky
RAID 1+0 – databáze; kombinace 0, 1; alespoň 4 disky
RAID 3 – pracuje po bitech, jeden vyhrazen pro paritu; rychlé IO pro sekvenční data; dražší řadič...
RAID 4 – data zapisována po blocích, jeden blok pro paritu; rychlé IO pro sekvenční data; pomalý zápis
RAID 5 – parita postupně na různé disky; rozumný kompromis mezi bezpečností, kapacitou a výkonem
RAID 6 – 2 paritní disky – horizontální a vertikální; složité a drahé/ bezpečné – pro kritické aplikace
Optické disky – CD, DVD, Blue-ray...; Jukeboxy – data uložená na více médiích, přístup až sekundy
Flash paměti – na všechna data stejně rychle; zápis pomalejší než čtení; nevydrží moc zápisů; větší spotřeba
Připojování – přímo DAS; přes síť NAS; bloková úložiště SAN

Soubor – identifikovaná kolekce logicky souvisejících záznamů; jsou-li stejného typu a velikosti – homogenní s.
Klíč – množina atributů jednoznačně identifikujících záznam – hodnotové, hashované, relativní (od zač. soub.)
Operace nad soubory – insert, delete, update, fetch
Schema organizace souboru – abstrakce zobecňující soubor s určitými vlastnostmi – struktura, algoritmy, data
dynamické – rovnoměrná naplněnost stránek; omezená délka cesty; jinak statické
Operace nad schematy organizace souboru – build, reorganization, stav souboru, close
Čtení zadaného souboru – sériové (dle zadaného uspořádání), sekvenční (dle uložení)

Hromada – nehomogenní soubor; k atributům i jména;
Sekvenční – nejjednodušší, dávkové systémy; nesetříděný pro sesbíraná data/ uložený – pro práci se všemi daty
primární soubor + s. aktualizací, reorganizace = setřídění a slití, hledání půlením intervalu $\log_2(N/|b|)(s+r+btt)$
Index-sekvenční – přístup sekvenčně i přímo; data, index, oblast přetečení; účetní systémy platebních karet.
Kapsy – indexované množiny bloků ~ spojový seznam; bitová mapa co obsazeno $|\log_{|B/R|}N/|B/R||$
Indexový – vyhledávání dle různých klíčů; primární soubor a indexy – setříděné klíče + odkazy; $|\log_{|B/R|}N|$
Částečná shoda – vyváření průniků adres; kombinovaný index pro více atributů zároveň
Nepřímé adresování – primární a sekundární klíč – mohu reorganizovat soubor bez změny indexů
Bitové mapy – pro malé domény hodnot (až 1000); booleovské dotazy; združování procesů – rychlejší
Indexy: Statické, seznamy adres, bitové mapy, B stromy, kombinované indexy, vícerozměrné indexy
S přímým přístupem – velmi rychlé přístupy dle konkrétního klíče; adresa dána klíčem; nevhodné pro řídká data
přímo (logická/ fyzická adresa); nepřímo (transformovaná adresa musí být jednoznačná)
1vyhl. klíč -> indexový s.; jen sekvenční přístup -> sekvenční s.; jen přímý přístup -> s. s přímým p.; index-s. s.

Hashování – řešení kolizí – prostá fce, první volné místo za, sekundární fce, perfektní hashování
Cormack – dvoufázové – do adresáře a primárního souboru; sekundární fce $(k>>i)\bmod r$
$h(k, s)$ do tabulky délky s; p počátek oblasti, i o kolik jsem shiftoval; p počet prvků
Larson+Kalja – 2 sady fcí; data do m stránek, každá zámek; $h_i(k)=(k+i)\bmod s$; test zámku $s_i(k)=(k\text{XOR}i)\bmod z$
Fagin (rozšiřitelné) – hlavičky po kolika bitech se liší co je ve stránce; dělení – rozpůlím, inkrementace hlavičky
Litwin (lineární) – vždy po L záznamech rozštěpím stránku, která je na řadě; oblast přetečení, vždy pak projít
Skupinové štěpení – dělím po L záznamech G stránek do G+1, jsou-li všechny, přerovnám a znovu...
přístup simuluje tvorbu – $h()$, $h_1()$, $h_2()$,... $k \bmod s$, $k \bmod (G+1)$, $(k \operatorname{div} (G+1)) \bmod (G+1)$

B-strom – redundantní (všechna data v listech), neredundantní (v každém uzlu, data unikátní)
kořen 2-m potomků, ostatní uzly $|'m/2'|$-m potomků; všechny větve stejně dlouhé, v uzlu setříděné
B+strom – poslední člen odkazuje na další list; jen u redundantních – pro rychlé sekvenční čtení

*B**strom – kouknu k sousedům, mají-li místo, podstrčím jim prvek, nemají-li místo, vezmu je a rozdělím do n+1
Proměnná délka záznamu – řetězce...; mít znak pro konce slov...; u redundantních netřeba v nelistech celá slova
B-strom s odkládaným štěpením – nerozdělím uzel hned, jen stránka přetečení a teprve až přeteče i ta...
Více rozměrný *B*-strom – v jednotlivých úrovních různé klíče – problém když ty horní neznám...
Vícerozměrná mřížka – obdélníčky na tutéž stránku, nad tím indexy, když přeteče rozdělím na obdélníčky
při dělení 1x1 musím rozdělit matici a přestavět to...; velmi náročné na údržbu

Spojené klíče – jak je seřadit, kolik dát kterému bitů, neznáme-li vždy všechny
d # bitů, p_i pravděpodobnosti; $d_i=(d-\sum\log_2 p_j)/n+\log_2 p_i$; $d_i<0$ -> eliminuj A_i přepočti; $d_i>d$ -> $d_i:=d$, zbytek :=0; $d_i>\log_2|A_i|$ -> $d_i:=\log_2|A_i|$, $d-=d_i$, eliminuj A_i, přepočti; zaokrouhlit
problém – musíme prohledat hodně stránek, zda tam něco není... -> deskriptory, grayovy kódy...
deskriptory – 2. hašovací fce nastaví při vložení bit, nejdřív testuji, zda tam má cenu hledat
rozdělit na dvě části a testovat nejdřív první pak druhý... - víc úrovní...; pro větší domény hodnot
Grayovy kódy – upravené číslování stránek – změna vždy jen v jednom bitě

Prohledávání kolekcí dokumentů – procházet (slovo, dokument, pozice) – je třeba mít dost místa na disku
setřídíme primárně dle slov -> slovo, dokument, pozice, pozice..., dokument,... a v tom hledám...
hledání „slovo1 slovo2“ vs „slovo1. Slovo2“ - členit raději ještě na odstavec, věta; ale Karel IV. …
potřeba program na základní tvary, problém „travička“...

(*a*,*b*) stromy – počet potomků mezi *a* a *b*; když chci vkládat a po cestě narazím na plný uzel, tak ho rozštěpím
méně časté štěpení vs. zhoršení naplnění
Spirálová paměť - ???
Bitemporální databáze – občas si musím pamatovat i změny, kdo ji provedl.../ odkdy do kdy platí (jízdní řády)