

OBLIKOVANJE BAZA PODATAKA

Vježba 6

Princip komunikacije s diskom

- Osnovni princip rada RDBMS-a s diskom:
 1. RDBMS dohvati podatke s diska u radnu memoriju
 2. RDBMS odrađuje zadane naredbe **nad podacima u radnoj memoriji**
 - SELECT
 - INSERT
 - UPDATE
 - DELETE
 3. RDBMS vraća promijenjene podatke (ako ih ima) natrag na disk

Smještaj na disku

- SQL Server podatke na disku smješta u:
 - **Primarna podatkovna datoteka (mdf)**
 - Obavezna, sadržava sve objekte i podatke iz baze podataka
 - **Zapisnička datoteka (engl. *database log*) (ldf)**
 - Obavezna, koristi se za podršku transakcijama
 - Često se naziva i transakcijska datoteka (engl. *transaction log*)
 - Opcionalne **sekundarne podatkovne datoteke (ndf)**
 - Omogućavaju razmještanje podataka na više diskova i optimizaciju izrade sigurnosnih kopija

Organizacija podataka (1/2)

- Podatkovna datoteka je podijeljena na **stranice** (engl. *pages*)
- Dio radne memorije (engl. *buffer pool*) RDBMS-a se sastoji od **odsječaka** (engl. *slot*)
 - Svaki odsječak je toliko velik da u njega stane jedna stranica
- Stranica je osnovna količina podataka u RDBMS-u
 - **IO operacije** (engl. *input output*)
 - Čitanje kopira jednu ili više stranica iz podatkovne datoteke u odsječke u radnoj memoriji
 - Pisanje kopira jednu ili više stranica iz odsječaka iz radne memorije u podatkovnu datoteku

Stranice i skupovi stranica

- Na SQL Serveru u 1 MiB memorije stane 128 stranica
 - $1024 \text{ KiB} / 8 \text{ KiB} = 128$
- Na svakoj stranici se tih 8.192 bajta dijeli na:
 - Zaglavlje stranice (132 bajta)
 - **Podaci na stranici (8.060 bajtova)**
- Podaci koje stranica sadržava ovise o tipu stranice
 - Mogu biti retci iz tablice, dijelovi indeksa, ...
- 8 stranica koje se nalaze jedna iza druge u memoriji nazivamo **skup stranica** (engl. *extent*)

Tipovi stranica

- SQL Server stranica po tipu može biti:
 - **Data Page** (type 1) – stranica koja čuva retke iz tablice
 - **Index Page** (type 2) – stranica koja čuva ključeve indeksa
 - Text/Image (type 3, 4) - Large Objects (LOB) podaci
 - Sort Page (type 7) – privremeni page za potrebe sortiranja
 - Global Allocation Map (type 8) (GAM) – info o zauzetosti extentova
 - Shared Global Allocation Map (type 9) (SGAM) – info o zauzetosti dijeljenih extentova
 - Index Allocation Map (type 10) (IAM) – info o extentima koje koriste indeksi
 - Page Free Space (type 11) (PFS) – info o slobodnim prostorima u stranicama
 - Differential Changed Map (type 16) - GAM extenti koji su mijenjani nakon full ili differential backupa
 - Bulk Changed Map (type 17) – GAM extenti koji su mijenjani bulk naredbama nakon backupa

Podjela tipova podataka

- Podjela tipova podataka:
 - 1. LOB:** nvarchar(max), varchar(max), varbinary(max), xml
 - Čuvaju podatke do veličine od **2 GiB**
 - Uvijek se čuvaju na LOB stranicama
 - Podatkovna stranica čuva pokazivače na LOB stranice
 - 2. Varijabilne duljine:** nvarchar(*n*), varchar(*n*) i varbinary(*n*)
 - Ako stanu, čuvaju se na podatkovnoj stranici
 - Inače se čuvaju na LOB stranici
 - 3. Poznate duljine** (svi ostali tipovi podataka): int, datetime, ...
 - Uvijek se čuvaju na podatkovnoj stranici

Smještaj redaka na stranicama

- Pravilo: **svaki redak mora cijeli stati na jednu stranicu**
=> duljina svakog retka ograničena na najviše 8.060 bajtova
- Duljinu retka **određuju upisani podaci** u redak, a ne definicije stupaca
- Primjer: tablica s tri nvarchar(4000) stupca
 - Ako umetnemo jedan redak s vrijednostima 'a', 'b' i 'c'
 - Ukupna duljina retka je 6 bajtova (svako slovo po 2 bajta)
 - Sve vrijednosti ostaju na podatkovnoj stranici
 - Ako umetnemo jedan redak s vrijednostima 'aaa...a', 'bbb...b' i 'ccc...c' (svako slovo 4.000 puta)
 - Ukupna duljina retka je 24.000 bajtova
 - Sve vrijednosti se izmještaju na LOB stranice

Naredbe za analizu stranica

- Za analizu ćemo koristiti **DBCC** naredbe (proširenje T-SQL-a):
 - Potrebno uključiti ispis svih DBCC poruka na ekran:
DBCC TRACEON(3604)
 - Prikaz stranica koje pripadaju tablici:
DBCC IND('naziv_baze', 'naziv_tablice', -1)
 - Vrijednost -1 označava podatkovne i indeksne stranice
 - Prikaz detalja stranice:
DBCC PAGE('naziv_baze', FID, PID, 3) WITH TABLERESULTS
 - Vrijednost 3 označava najviše detalja

Korištenje naredbi

- **DBCC TRACEON(3604)** vrijedi za cijelu konekciju
- **DBCC IND** će nam vratiti sljedeće bitne podatke:
 - **PageFID** – rbr podatkovne datoteke (kod nas uvijek 1)
 - **PagePID** – rbr stranice unutar podatkovne datoteke
 - **PageType** – tip stranice
 - 1 = podatkovna stranica
 - 2 = indeksna stranica
- Za **DBCC PAGE** koristimo gornje podatke
 - Vraća zaglavlje i niz *slotova* koji sadržavaju retke

Vježbe 1

1. Odgovorite na pitanja:

- a. Na koliko stranica su smješteni podaci iz tablice **Drzava** (PageType jednak 1)?
- b. Napišite redne brojeve stranica (PagePID) i redne brojeve datoteke u kojoj su smještene stranice (PageFID).
- c. Kolika je duljina prvog, drugog i trećeg retka u bajtovima?
- d. Po kojem stupcu su poslagani reci na stranicama?
- e. Pronađite redak za "Njemačka". Koji su reci ispred i iza njega?

Vježbe 2

2. Odgovorite na pitanja:

- a. Na koliko stranica su smješteni podaci iz tablice **Proizvod**?
- b. Napišite redne brojeve stranica (PagePID) i redne brojeve datoteke u kojoj su smještene stranice (PageFID).
- c. Kolika je duljina prvog, drugog i trećeg retka u bajtovima?
- d. Koji je ProizvodID zadnjeg retka na prvoj stranici? Koji je ProizvodID prvog retka na drugoj stranici?
- e. Koji je ProizvodID zadnjeg retka na drugoj stranici? Koji je ProizvodID prvog retka na trećoj stranici?
- f. Po kojem stupcu su poslagani retci na stranicama?

Vježbe 3

3. Odgovorite na pitanja:

- a. Na koliko stranica su smješteni podaci iz tablice **Kupac**?
- b. Napišite redne brojeve prvih 5 stranica (PagePID) i redne brojeve datoteke u kojoj su smještene stranice (PageFID).
- c. Koliko redaka je smješteno na prvoj stranici?
- d. Kolika je duljina prvog, drugog i trećeg retka u bajtovima?