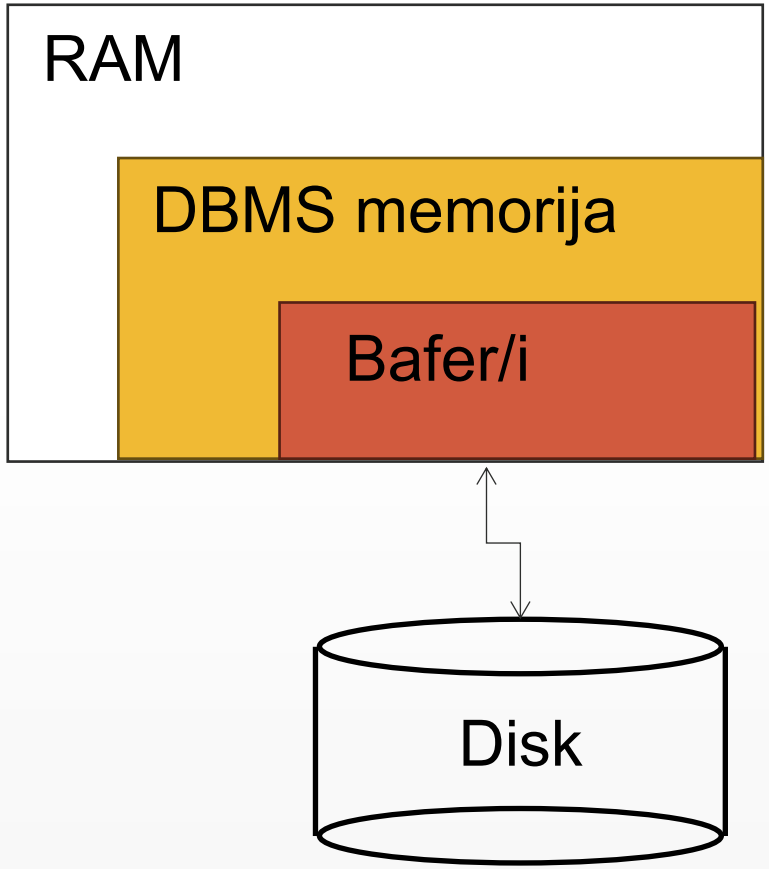
Upravljanje baferom

Buffer Management

Bafer

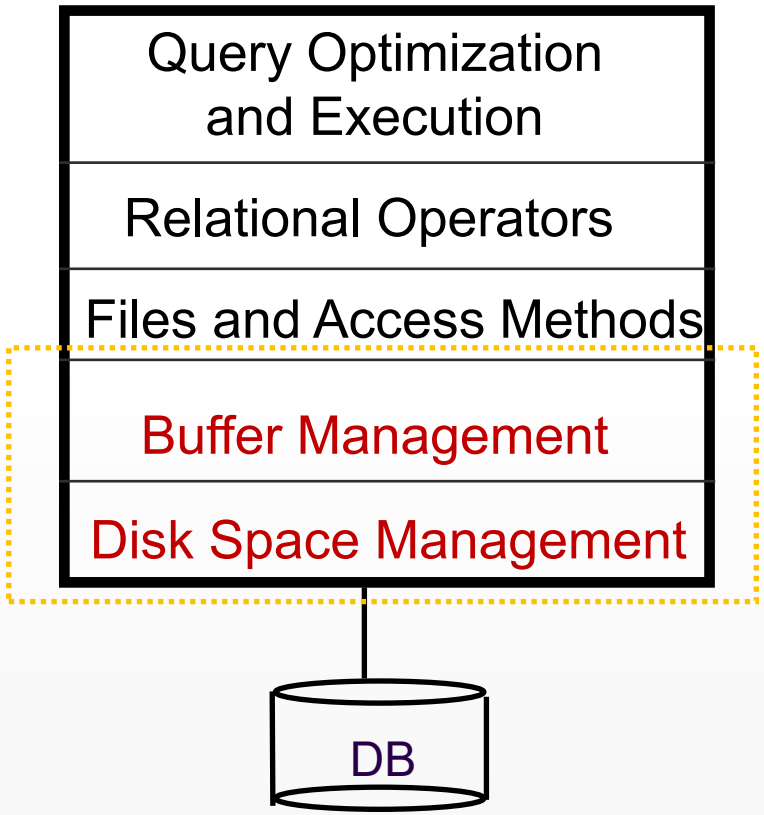
- Da bi DBMS izvršavao operacije nad podacima podaci moraju biti u radnoj memoriji
  - Problem: količina podataka koji se obrađuju najčešće prevazilazi kapacitet radne memorije dodeljene DBMS-a
- Bafer je region radne memorije koji se koristi za privremeno smeštanje podataka
  - U slučaju DBMS-a koristi se za smeštanje podatka pročitanih sa diska, a traženih od strane viših slojeva DBMS-a (npr, file menadžera)
  - Služi za keširanje podataka sa diska

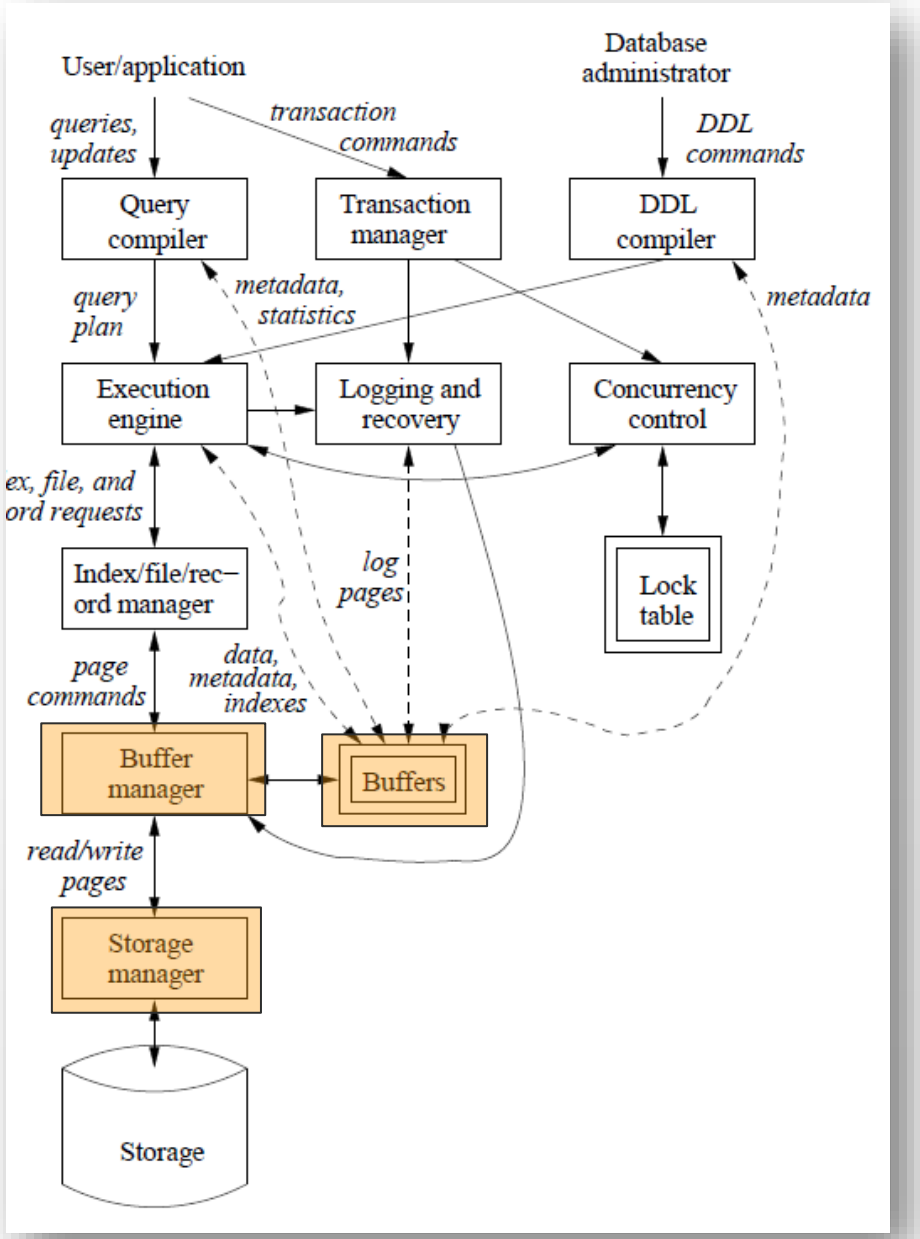


Upravljanje radnom memorijom

- Bafer menadžer – posebna komponenta DBMS-a koja je zadužena za upravljanje radnom memorijom DBMS-a
- Cilj -> minimizacija čekanja na učitavanje podataka sa diska.
  - Kada čitati, a kada pisati na disk?

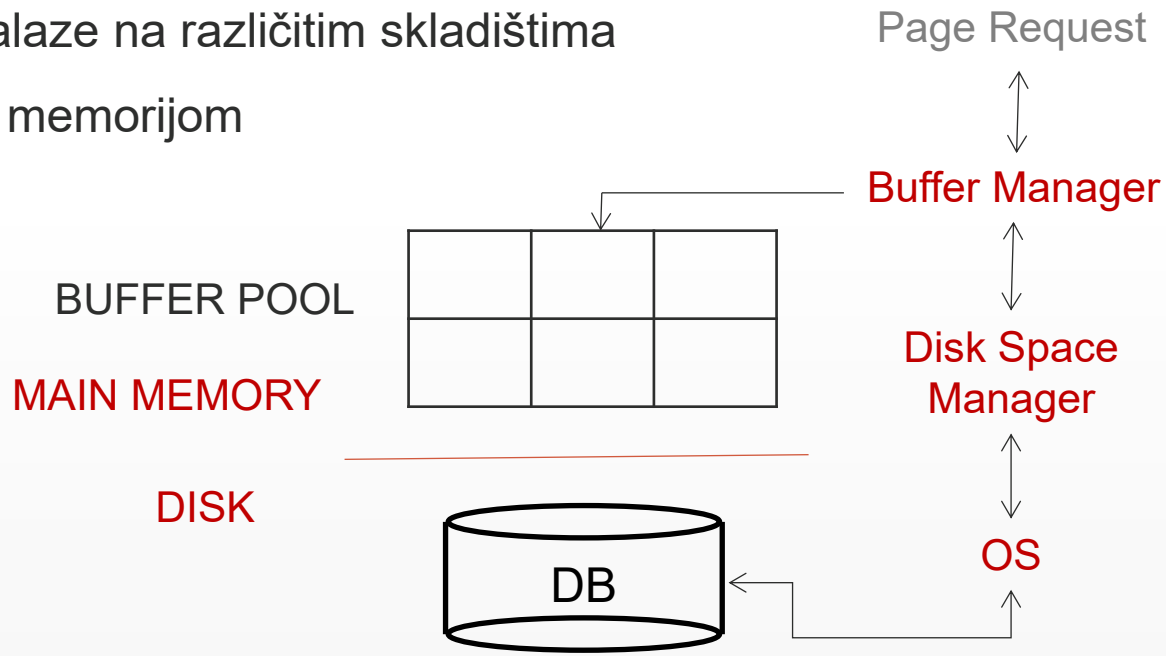
DBMS kontekst



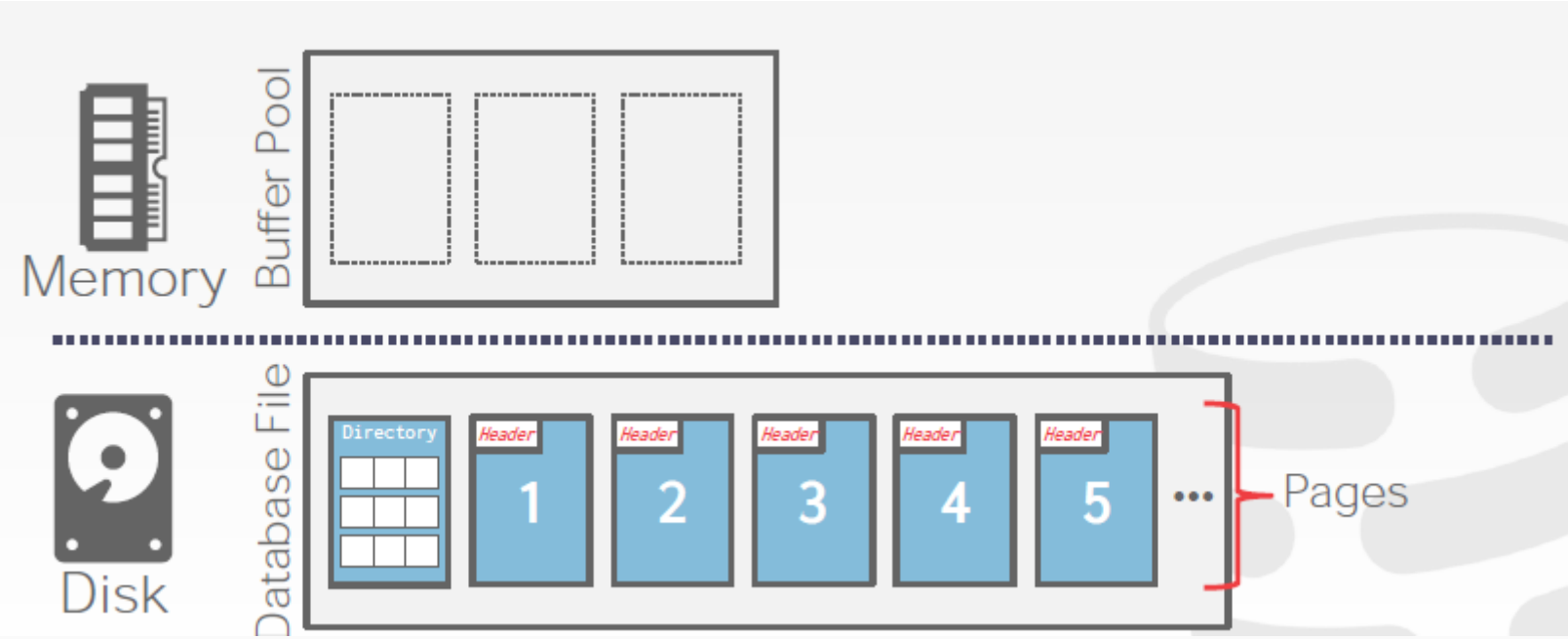


Bafer menadžer

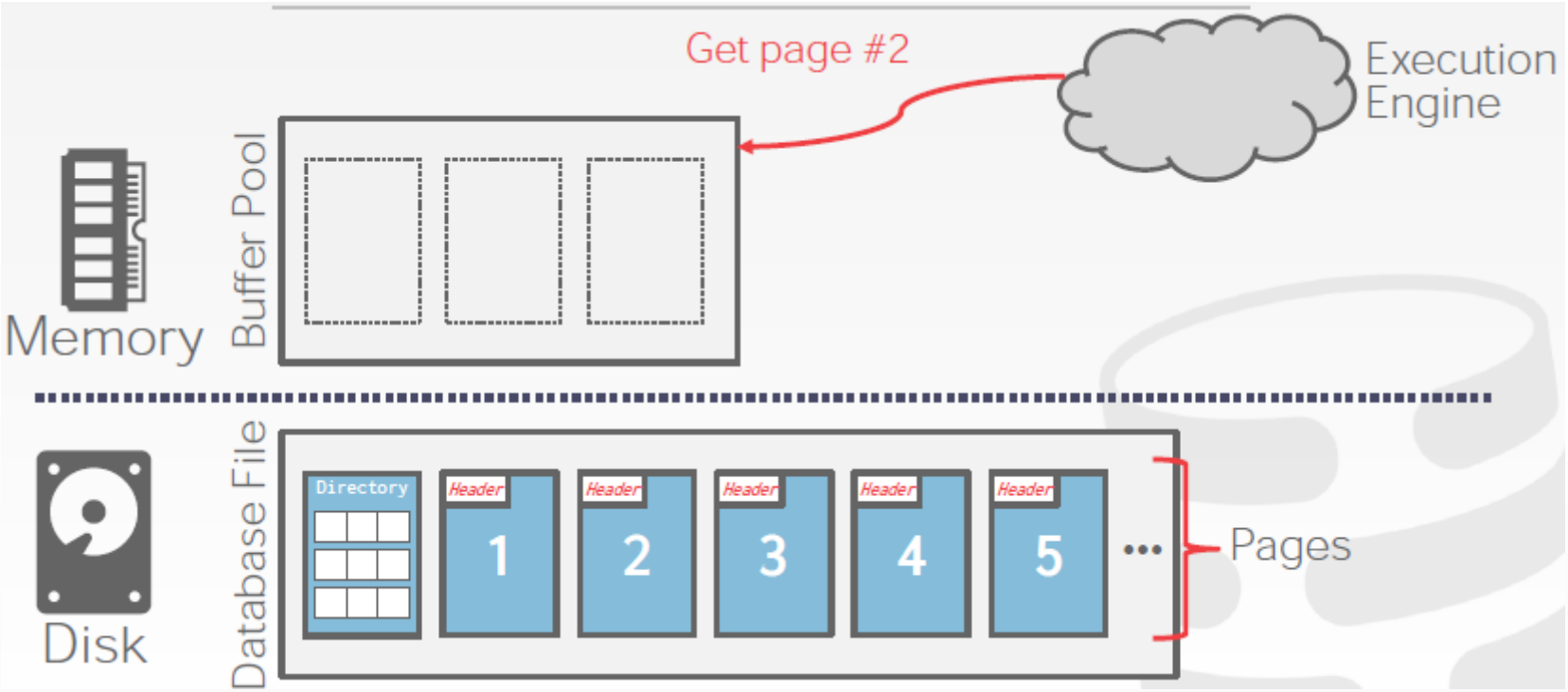
- Sloj iznad disk menadžera
  - Zadužen za dobavljanje podataka u radnu memoriju
  - Sakriva činjenicu o tome da se podaci nalaze na različitim skladištima
- Bafer menadžer upravlja dostupnom radnom memorijom
- Memoriju uređuje kao kolekciju strana, koja se naziva bafer pul (buffer pool).



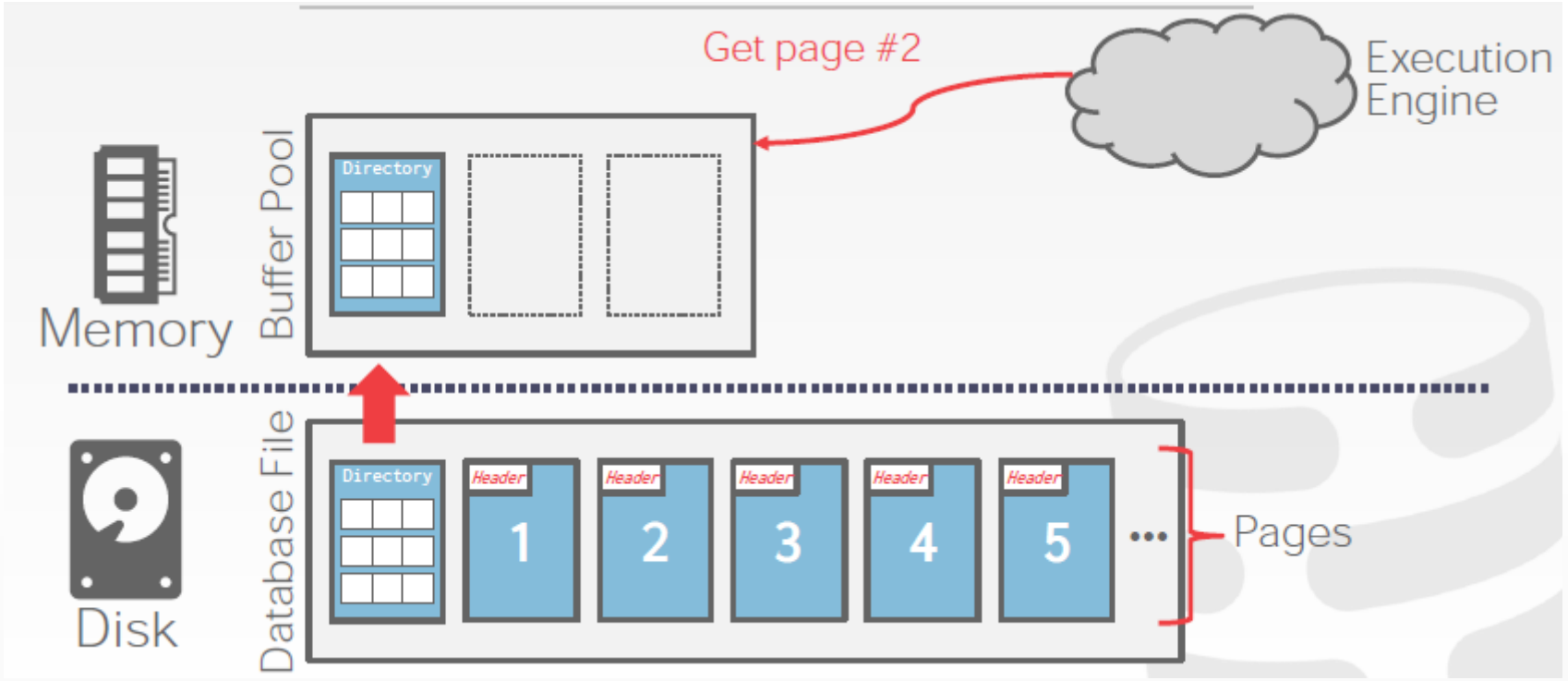
Bafer menadžer



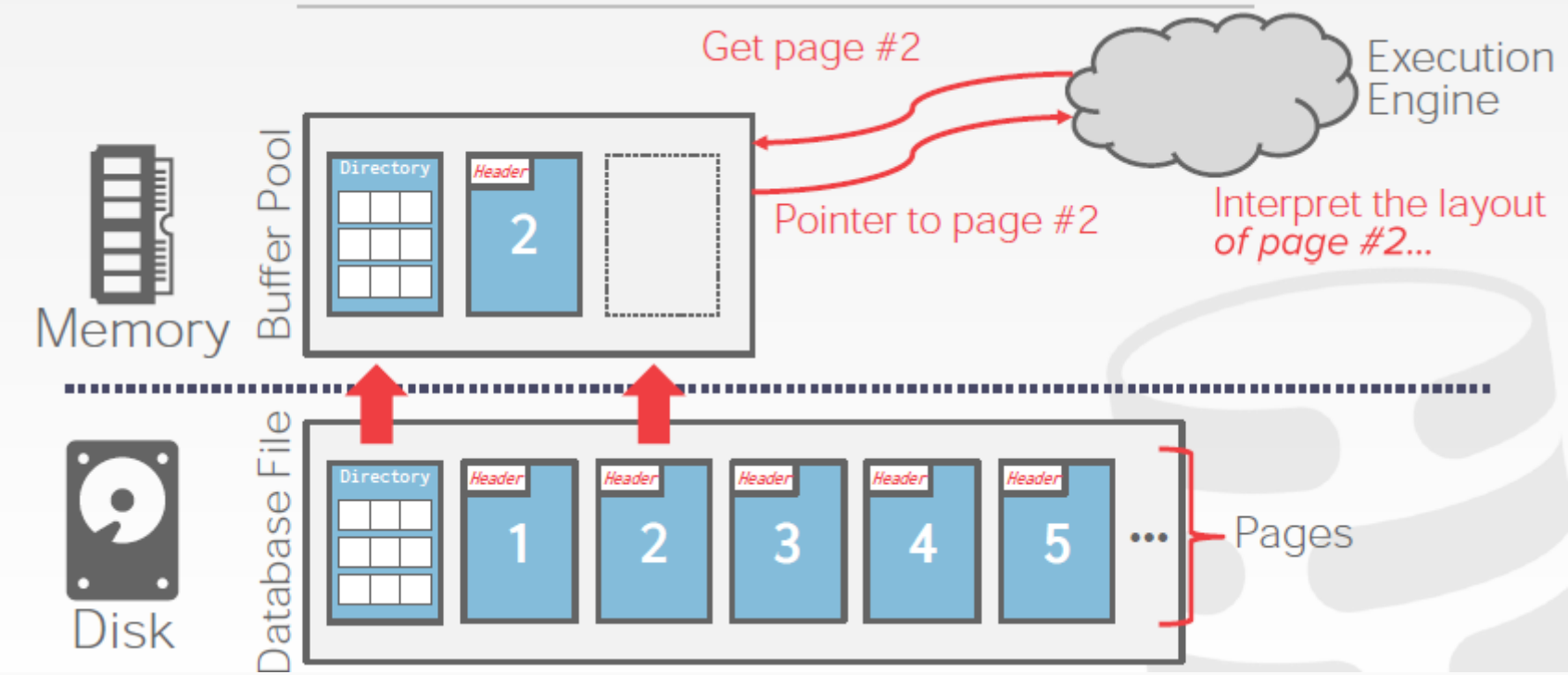
Bafer menadžer



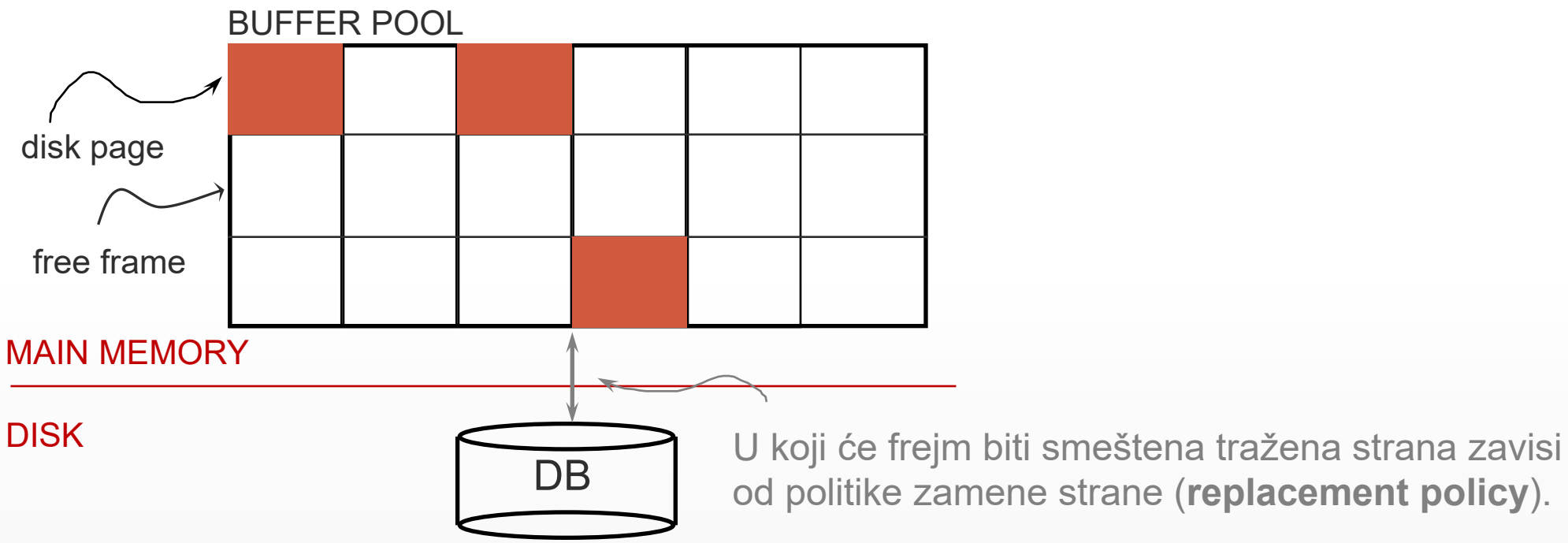
Bafer menadžer



Bafer menadžer



Bafer pul

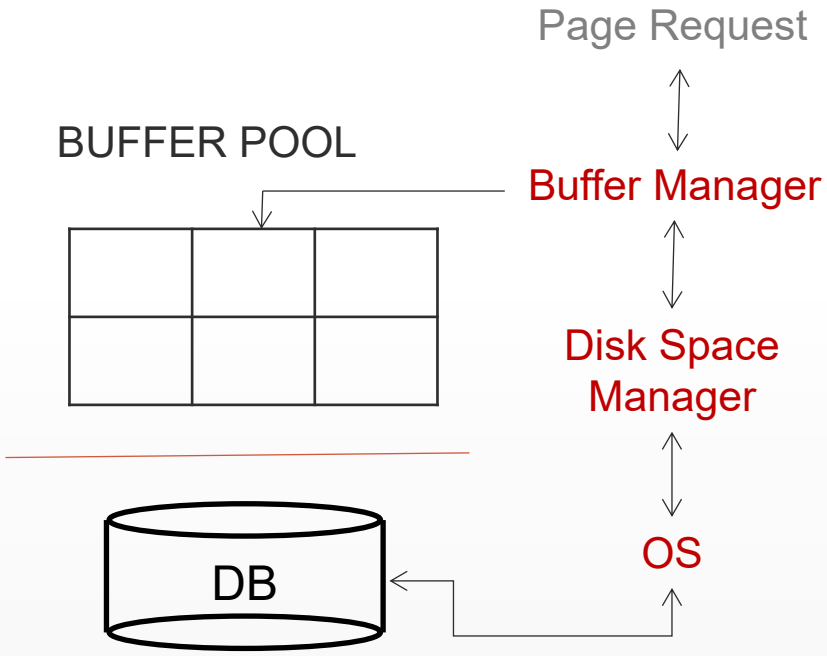


U koji će frejm biti smeštena tražena strana zavisi od politike zamene strane (**replacement policy**).

Frame – slot/mesto za stranu u baferu

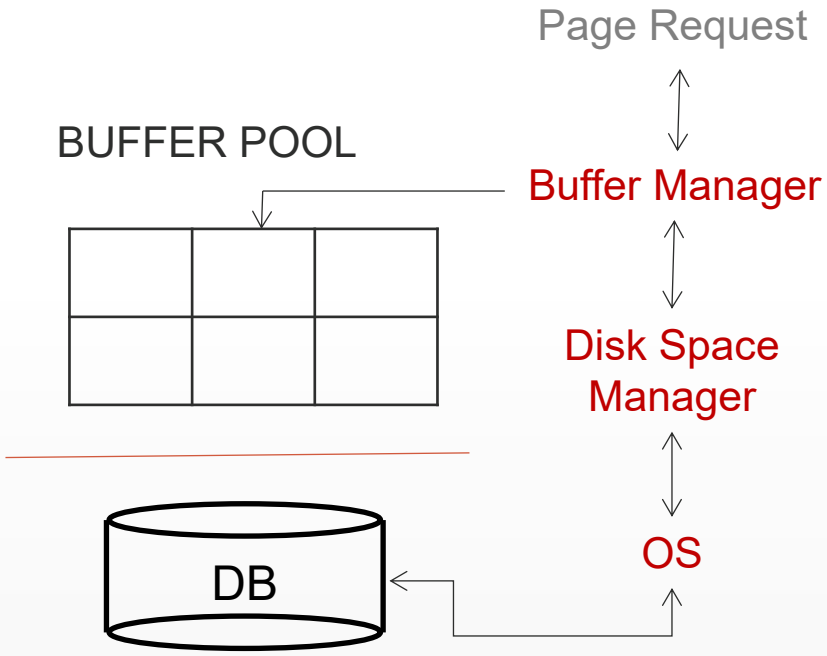
Šta se dešava po preuzimanju zahteva?

1. Zahtev za stranom
2. Strana u pulu - BM vraća referencu
3. Strana nije u pulu i ima slobodnih frejmova – BM šalje zahtev DSM, preuzima i vraća referencu
4. Strana nije u pulu i nema slobodnih frejmova
   - Određuje stranu koja se može ukloniti iz pula jer je niko ne koristi (sa ili bez pisanja na disk)
   - Preuzima traženu i smešta u oslobođeni frejm



Šta bafer menadžer mora da zna?

- Viši slojevi koji zahtevaju stranu, moraju da obaveštavaju bafer menadžer:
  - da žele da oslobode stranu koja im više nije potrebna – da bi BM znao da mesto zauzetom stranom može da oslobodi
  - da su izmenili zahtevanu stranu - da bi bafer menadžer obezbedio beleženje izmenjene kopije na disk pre njenog uklanjanja iz bafer pula



Page Table

Buffer Management

Meta podaci bafer menadžera

Page Table

- Bafer menadžer održava tabelu strana (**Page Table)**, tabelu koja sadrži:

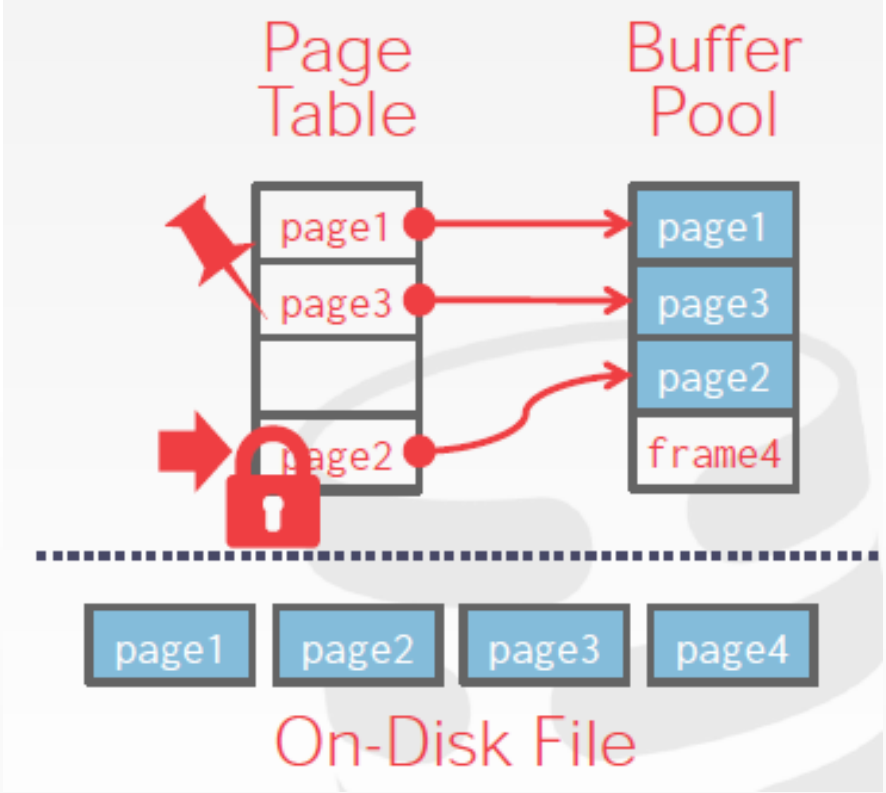
<frame#, pageid, pin_count, dirty_bit>

pin_count - brojač referenciranja, njegova vrednost govori koliko puta je strana koja se nalazi u frejmu zahtevana, a nije oslobođena, tj. broj trenutnih korisnika

dirty_bit - boolean promenljiva koja je postavljena na true ukoliko je strana koja se nalazi u frejmu menjana od trenutka kada je dopremljena u pul.

Bafer pul tabela informacija

- Kada je bafer menadžeru upućen zahtev za nekom stranom, tada:
  - Ako je strana već u pulu tada se njen brojač referenciranja uvećava za jedan - pin_count++
  - Ako strana nije u pulu tada se
    - Odabira frejm čiji će sadržaj biti zamenjen
      Samo strane sa pin_count=0
    - Ako je dirty=true, pisanje na disk
    - Postavljanje reze (Latch - Mutex) na slog u tabeli strana (Page Table) čime se rezerviše frejm u koji će biti upisana nova stranica
    - Učitavanje tražene strane u markirani frejm



Page table vs. Page Directory

- Page Directory
  - Mapira PID na lokacije stranica u db fajlovima.
  - Mora da bude čuvana kao i sve ostale stranice.
- Page Table
  - Mapira PID na frejmove, tj. kopije stranica (učitane iz DB fajla) koje se nalaze u nekom frejmu
  - Unutrašnja struktura koja se ne čuna na disku.

Lock ≠ Latch *

- Latch (reza)
  - Sinhronizaciona primitiva (mehanizam sinhronizacije) niskog novoa
  - Štiti interne DBMS strukture od uticaja drugih niti (npr. heš tabela)
  - Važeća samo tokom trajanja operacije zbog koje postavljena
  - Ne mora da podržava poništavanje promena (rollback)
  - Najčešće se implementira jezičkim primitivama kao što je mutex
- Lock (brava)
  - Sinhronizaciona primitiva višeg nivoa
  - Štiti sadržaj (torke, tabele, …) baze podataka od neregularnog uticaja od strane konkurentnih transakcija i važi tokom trajanja transakcije koja ga je postavila.
  - Mora da podržava poništavanje promena

Napomena.
Prevod termina Latch je moj lični izbor, jer nema odgovarajućeg u literaturi na srpskom jeziku

* Nije obavezno za studente matematike

Politike zamene

Buffer Management

Replacement policy

Politika zamene

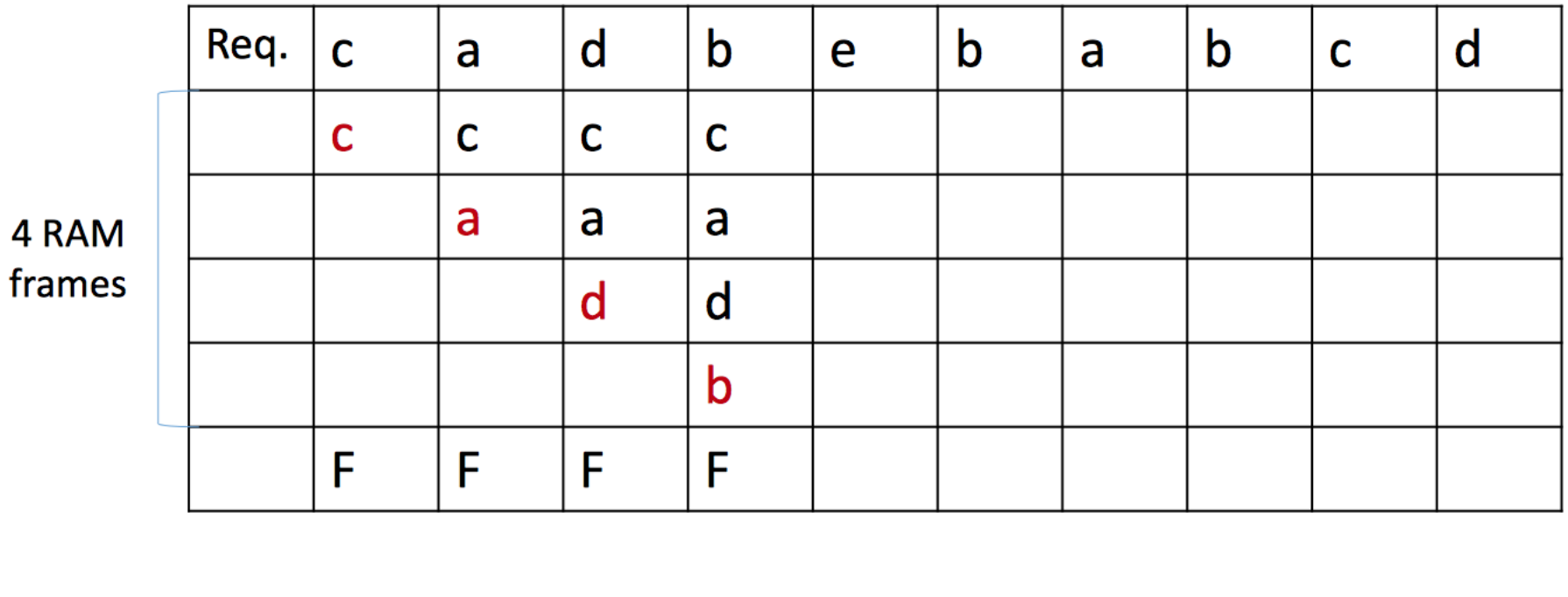
- Politikom zamene se definiše način na koji se određuje koji će slot u pulu biti popunjen novim sadržajem.
- Ciljevi:
  - Korektnost – ne dozvoliti prepisivanje pinovane strane
  - Preciznost – dobra procena toga koja stranica verovatno neće biti ponovo tražena u naradnom periodu
  - Brzo odlučivanje
  - Minimalni overhead koji nastaje uvođenjem meta podataka
- Dobro razvijeni DBMSovi koriste sofisticirane politike zamene.
  - Vode statistike o korišćenosti strana i na osnovu njih procenjuju šta će biti korišćeno.

Politika zamene

- LRU – least recently used
- Clock
- LRU-k

LRU

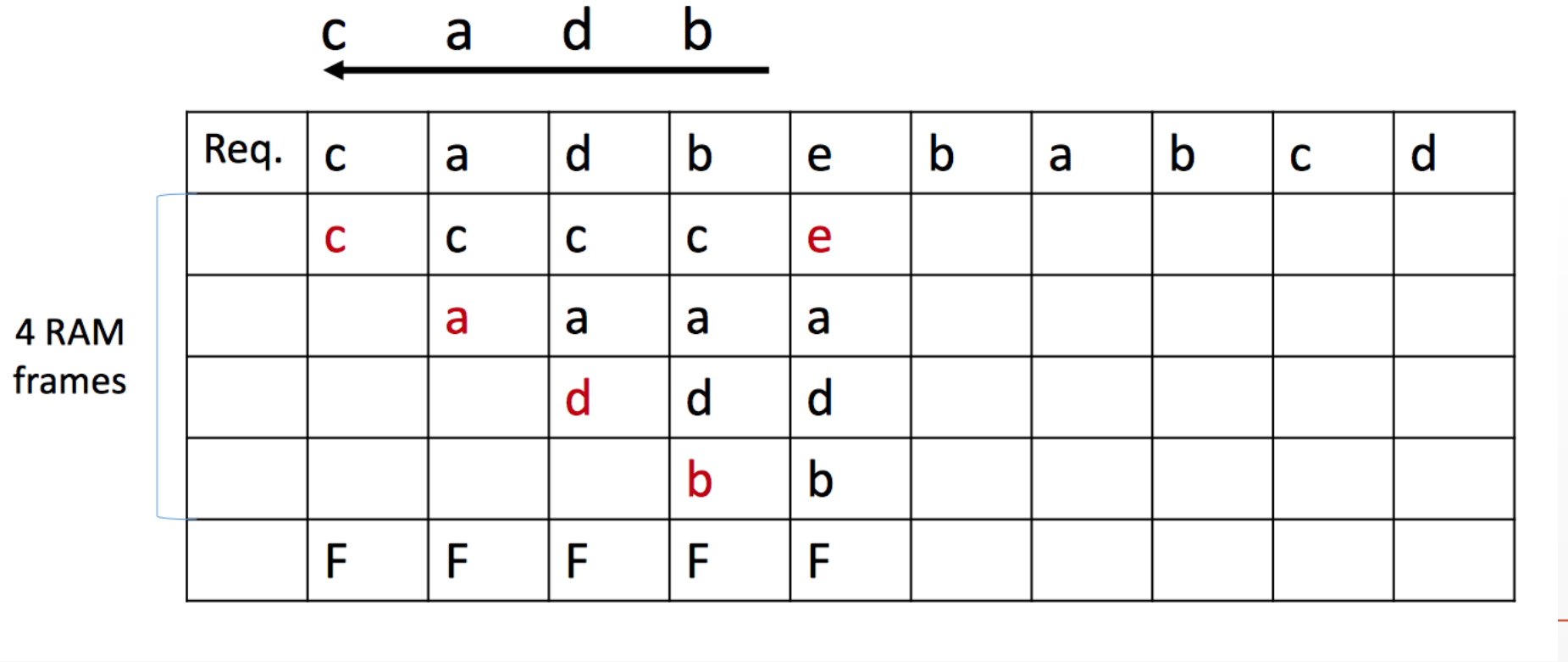
- Pamti se vreme pristupa stranici u pulu. Kada je pul pun, a tražena stranica nije u pulu uklanja se stranica koja je najranije referencirana.



Req.	c	a	d	b	e	b	a	b	c	d
	c	c	c	c						
		a	a	a						
			d	d						
				b						
	F	F	F	F						

LRU

- Pamti se vreme pristupa stranici u pulu. Kada je pul pun, a tražena stranica nije u pulu uklanja se stranica koja je najranije referencirana.



Req.	c	a	d	b	e	b	a	b	c	d
	c	c	c	c	e					
		a	a	a	a					
			d	d	d					
				b	b					
	F	F	F	F	F					

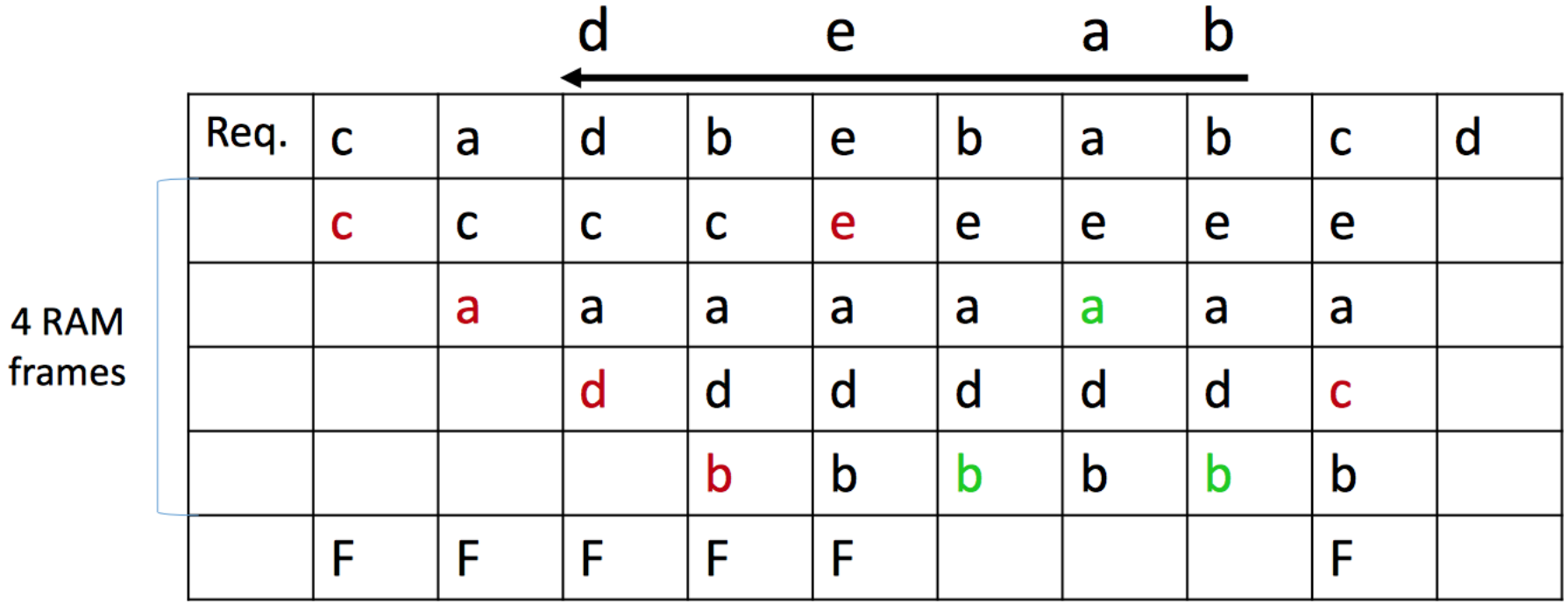
LRU

- Pamti se vreme pristupa stranici u pulu. Kada je pul pun, a tražena stranica nije u pulu uklanja se stranica koja je najranije referencirana.

	Req.	c	a	d	b	e	b	a	b	c	d
4 RAM frames		c	c	c	c	e	e	e	e		
			a	a	a	a	a	a	a		
				d	d	d	d	d	d		
					b	b	b	b	b		
		F	F	F	F	F					

LRU

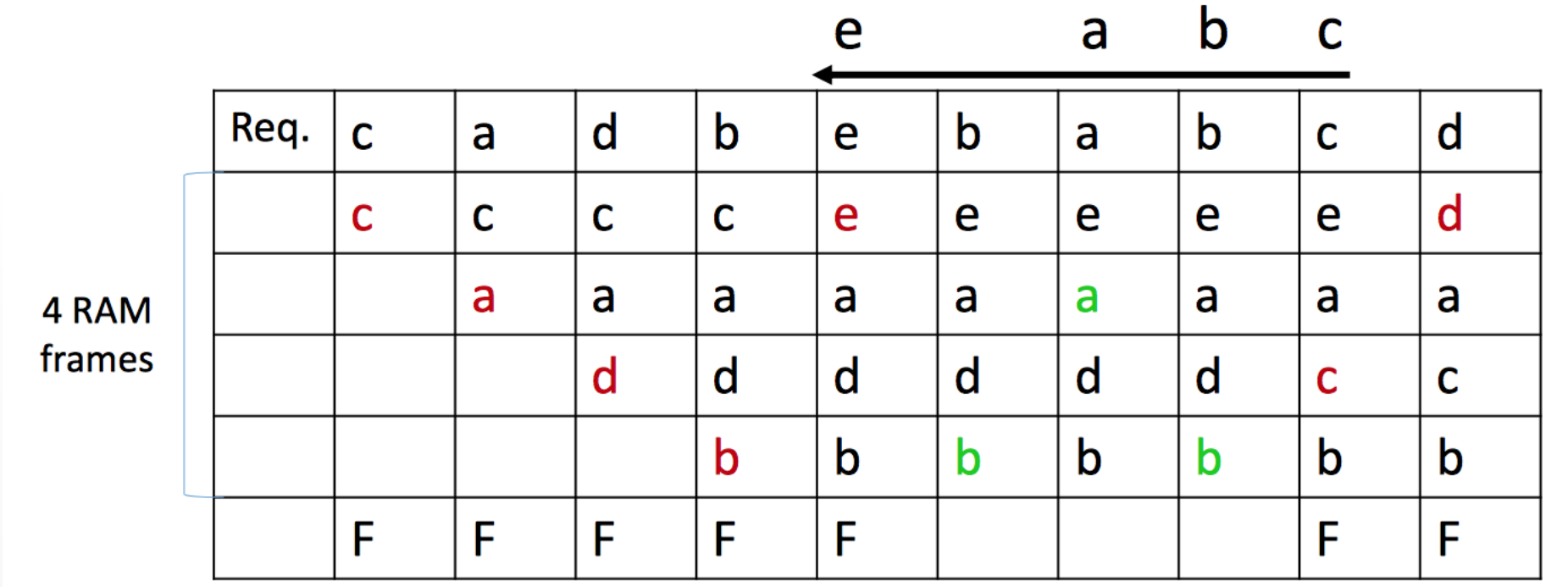
- Pamti se vreme pristupa stranici u pulu. Kada je pul pun, a tražena stranica nije u pulu uklanja se stranica koja je najranije referencirana.



Req.	c	a	d	b	e	b	a	b	c	d
	c	c	c	c	e	e	e	e	e	
		a	a	a	a	a	a	a	a	
			d	d	d	d	d	d	c	
				b	b	b	b	b	b	
	F	F	F	F	F				F	

LRU

- Pamti se vreme pristupa stranici u pulu. Kada je pul pun, a tražena stranica nije u pulu uklanja se stranica koja je najranije referencirana.

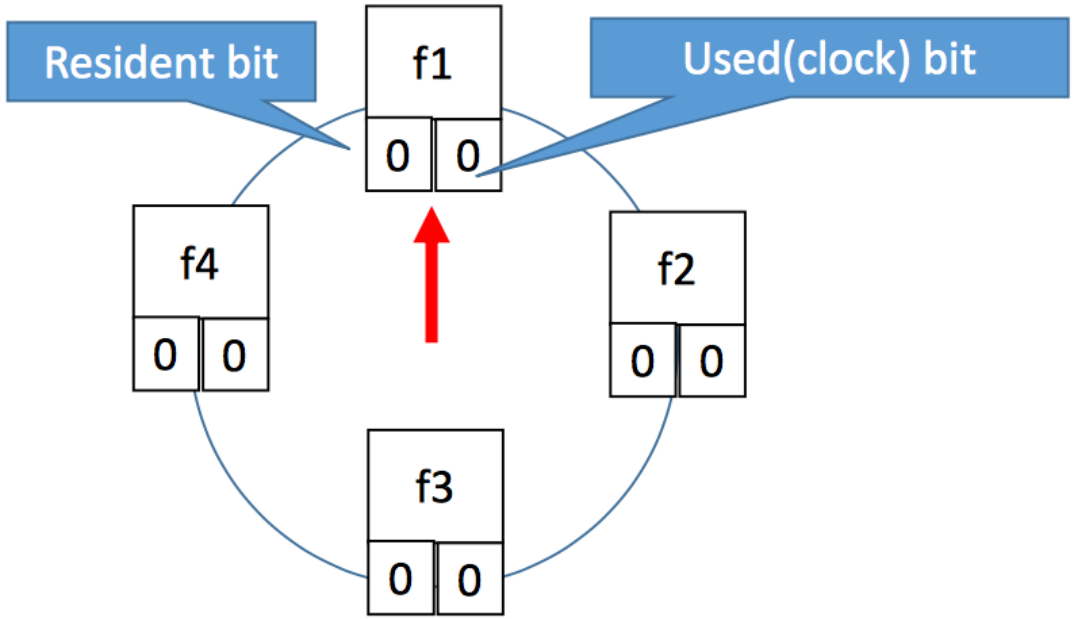


Req.	c	a	d	b	e	b	a	b	c	d
	c	c	c	c	e	e	e	e	e	d
		a	a	a	a	a	a	a	a	a
			d	d	d	d	d	d	c	c
				b	b	b	b	b	b	b
	F	F	F	F	F				F	F

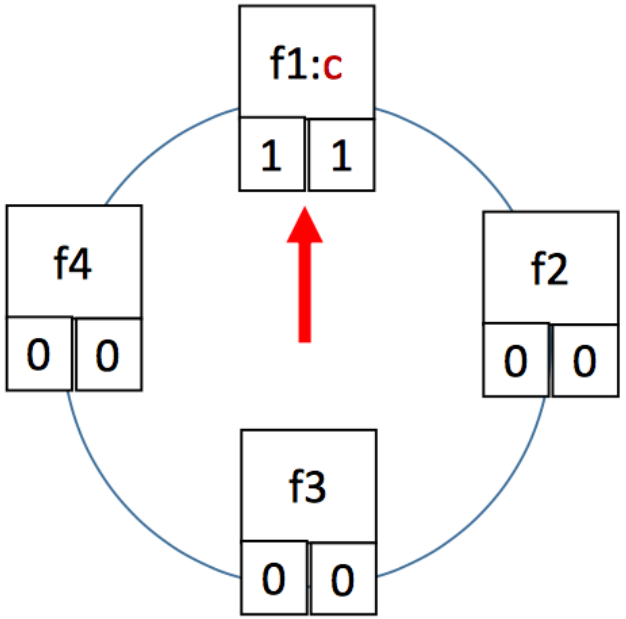
Clock (second chance)

- Ne pamte se vremena pristupa, već za svaku stranu po jedan reference bit.
- Kad god se strani koja se nalazi u pulu pristupa, reference bit se postavlja na 1.
- Informacije o stranicama se smeštaju u kružni bafer. Pamti se pokazivač na prvog kandidata za zamenu.
- Kada u bafer putu zahtevana stranica ne postoji, a bafer je pun, onda se razmatra stranica koja je kandidat za zamenu.
  - Ako je njen refernce bit 1, njegova vrednost se postavlja na 0, a razmatranje za zamenu se vrši nad narednom stranicom (pokazivač se pomera na narednu stranicu).
  - Ako je refenrce bit 0, onda se ta stranica menja novom, tj. u njen frejm se smešta nova.

	Req.	c	a	d	b	e	c	a	b	c	d
4 RAM frames	f1										
	f2										
	f3										
	f4										

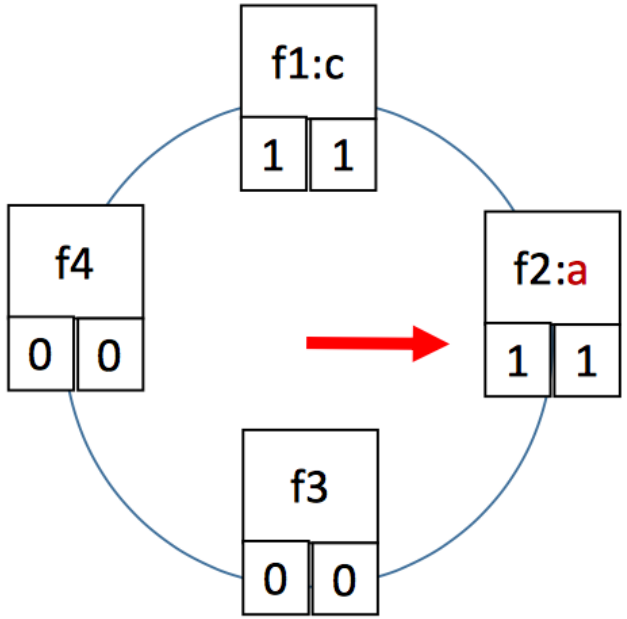


	Req.	c	a	d	b	e	c	a	b	c	d
4 RAM frames	f1	c									
	f2										
	f3										
	f4										
		F									



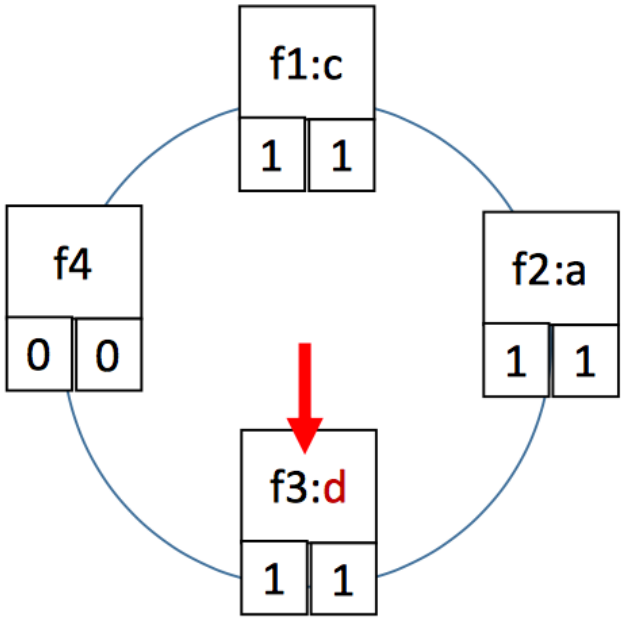
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c								
f2		a								
f3										
f4										
	F	F								



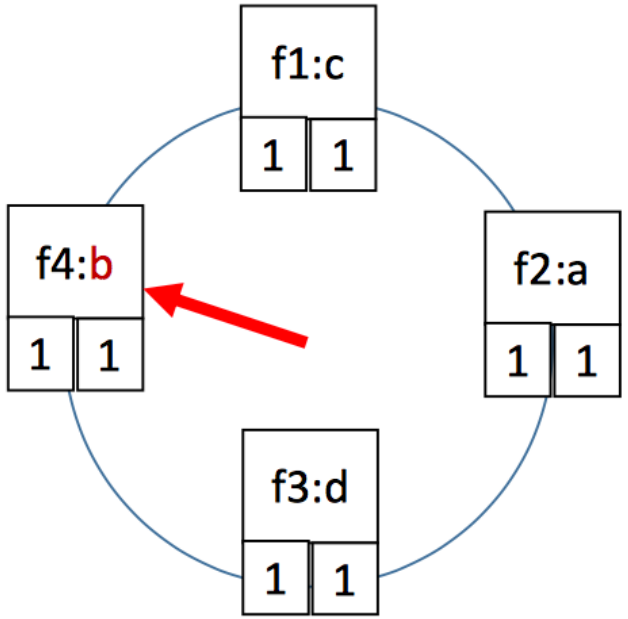
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c							
f2		a	a							
f3			d							
f4										
	F	F	F							



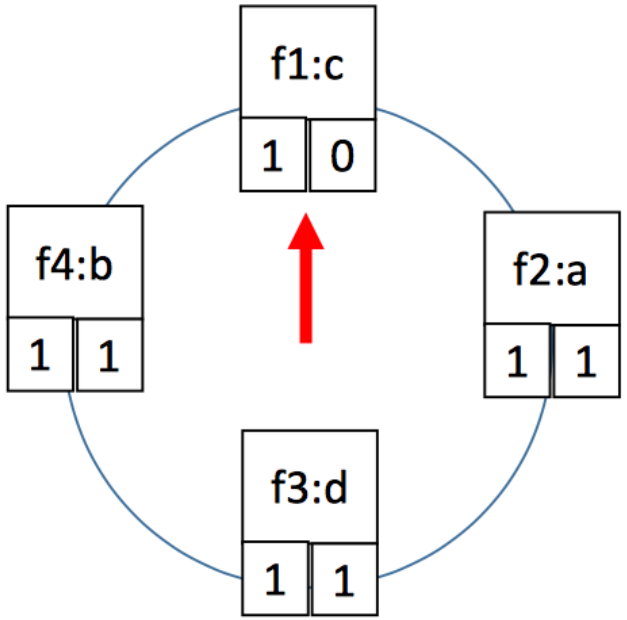
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c	c						
f2		a	a	a						
f3			d	d						
f4				b						
	F	F	F	F						



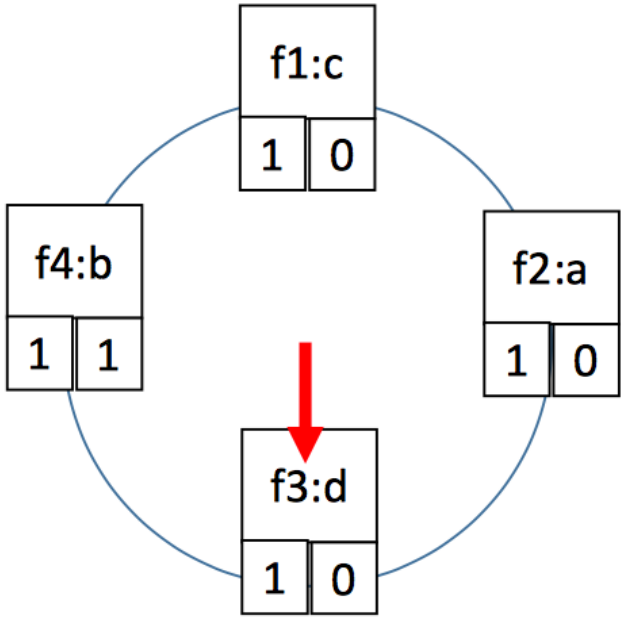
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c	c						
f2		a	a	a						
f3			d	d						
f4				b						
	F	F	F	F						



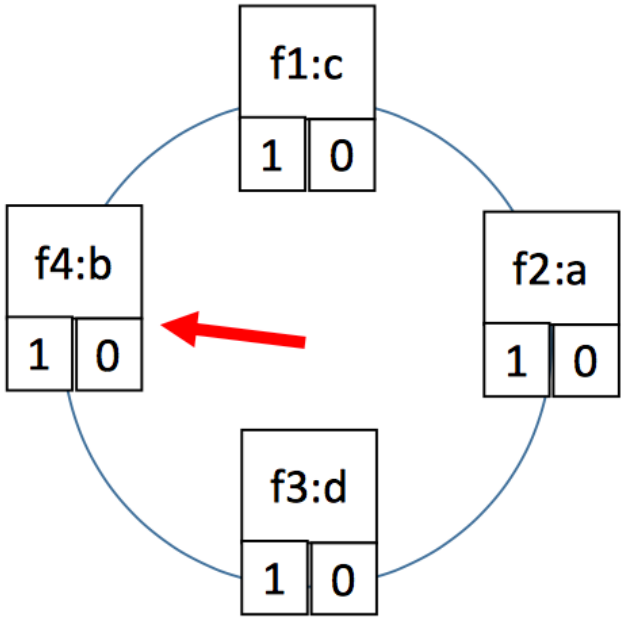
Searching for page to evict

	Req.	c	a	d	b	e	c	a	b	c	d
4 RAM frames	f1	c	c	c	c						
	f2		a	a	a						
	f3			d	d						
	f4				b						
		F	F	F	F						



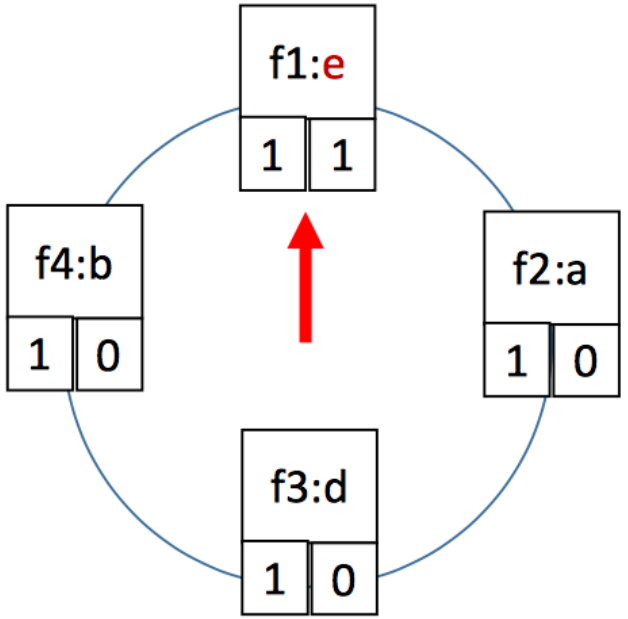
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c	c						
f2		a	a	a						
f3			d	d						
f4				b						
	F	F	F	F						



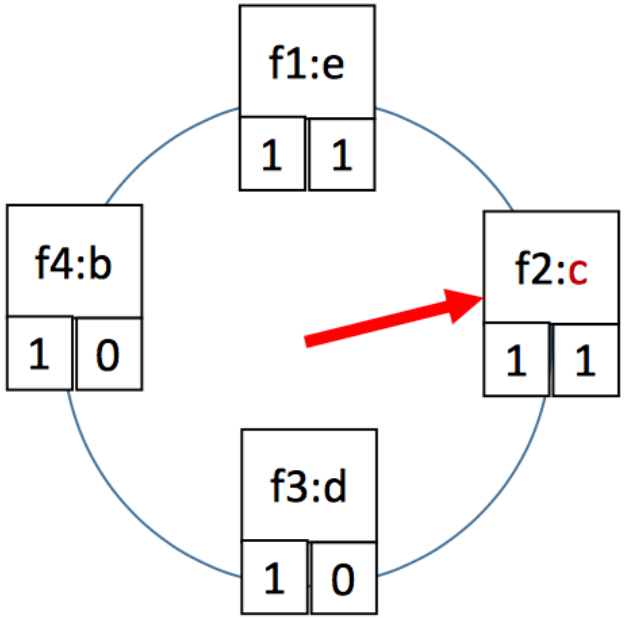
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c	c	e					
f2		a	a	a	a					
f3			d	d	d					
f4				b	b					
	F	F	F	F	F					



Evicted page c

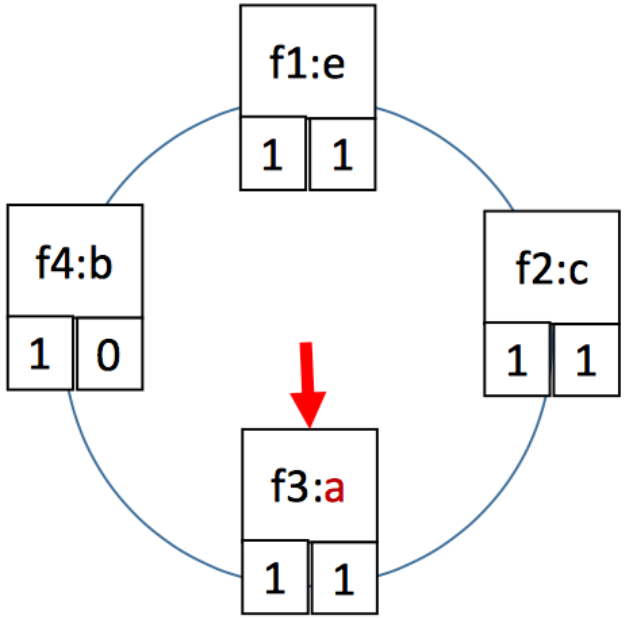
	Req.	c	a	d	b	e	c	a	b	c	d
4 RAM frames	f1	c	c	c	c	e	e				
	f2		a	a	a	a	c				
	f3			d	d	d	d				
	f4				b	b	b				
		F	F	F	F	F	F				



Evicted page a – its clock bit was 1

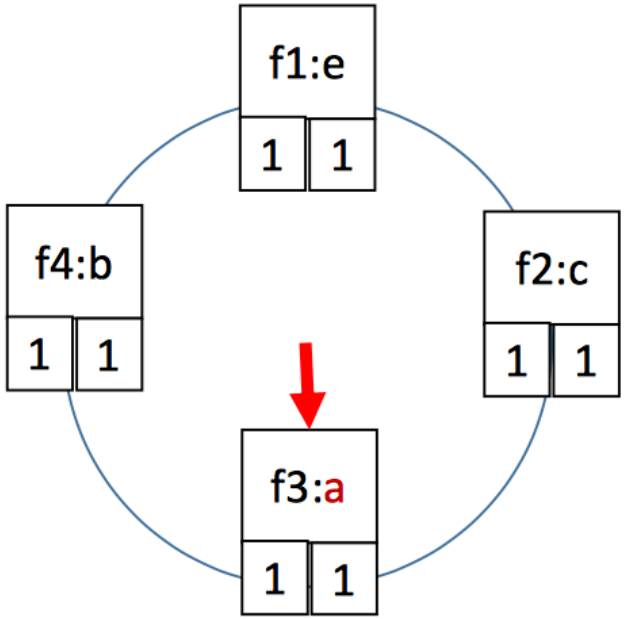
4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c	c	e	e	e			
f2		a	a	a	a	c	c			
f3			d	d	d	d	a			
f4				b	b	b	b			
	F	F	F	F	F	F	F			



Evicted page d

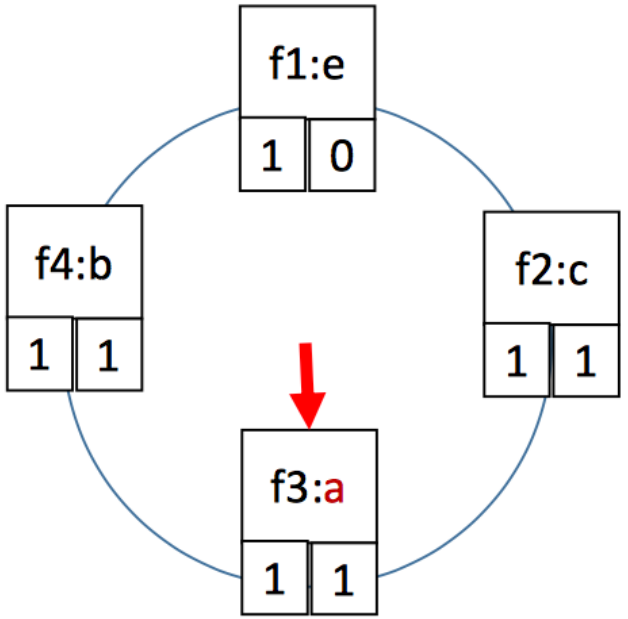
	Req.	c	a	d	b	e	c	a	b	c	d
4 RAM frames	f1	c	c	c	c	e	e	e	e		
	f2		a	a	a	a	c	c	c		
	f3			d	d	d	d	a	a		
	f4				b	b	b	b	b		
		F	F	F	F	F	F	F			



We know that b is in buffer

4 RAM frames

Req.	c	a	d	b	e	c	a	b	c	d
f1	c	c	c	c	e	e	e	e	e	
f2		a	a	a	a	c	c	c	c	
f3			d	d	d	d	a	a	a	
f4				b	b	b	b	b	b	
	F	F	F	F	F	F	F			



Where does *d* go:

A: frame 1

B: frame 2

C: frame 3

D: frame 4

Sequential Flooding

A: a, b
B: c, d, e, f

```
for each record i in A
    for each record j in B
        do something with i and j
```

	req.	a	a	a	a	b	b	b	b
3 frames for A - enough		a	a	a	a	a	a	a	a
						b	b	b	b
		F				F			
	req.	c	d	e	f	c	d	e	f
3 frames for B < |B|		c	c	c	f	f	f	e	e
			d	d	d	c	c	c	f
				e	e	e	d	d	d
		F	F	F	F	F	F	F	F

Sequential flooding
– each request –
page fault

LRU happens to evict exactly the page which we will need next!

LRU - K

- Na osnovu istorijskih podataka DMBS procenjuje kada će sledeći put određena stranica biti tražena.
- Pamti se K vremenskih trenutaka pristupa i izračunava interval između dva uzastopna pristupa stranici.

Još neke tehnike

- **Prioritetizacija (Priority hints)** – DBMS na osnovu konteksta korišćenja stranica može da definiše njihov prioritet.
- **Lokalizacija (Localization)** – pamti se pristup stranicama koje se koriste za izvršavanje jednog upita (kružni bafer definisan za jedan upit).
- **Pozadinsko pisanje na disk (Background writing)** – DBMS periodično, mimo potrebe za trenutnim oslobađanjem frejma, zapisuje prljave strane na disk i postavlja dirty flag na 0 ili samu stranu/frejm obeležava kao sledeći za upis.

Politika alokacije i optimizacija

Buffer Management

Optimizacija rada bafer pul-a

Politika alokacije

- Globalna politika: Donošenje odluka o alokaciji frejmova sa ciljem optimizacije rada celog sistema.
- Lokalna politika: Alokacija frejmova za pojedinačne tredove/upite.
- Većina sistema koristi kombinaciju.

Optimizacija rada bafer pula

- Višestruki bafer pulovi
- Pre-Fetching
- Scan Sharing
- Zaobilaženje bafer pula

Višestruki bafer pulovi

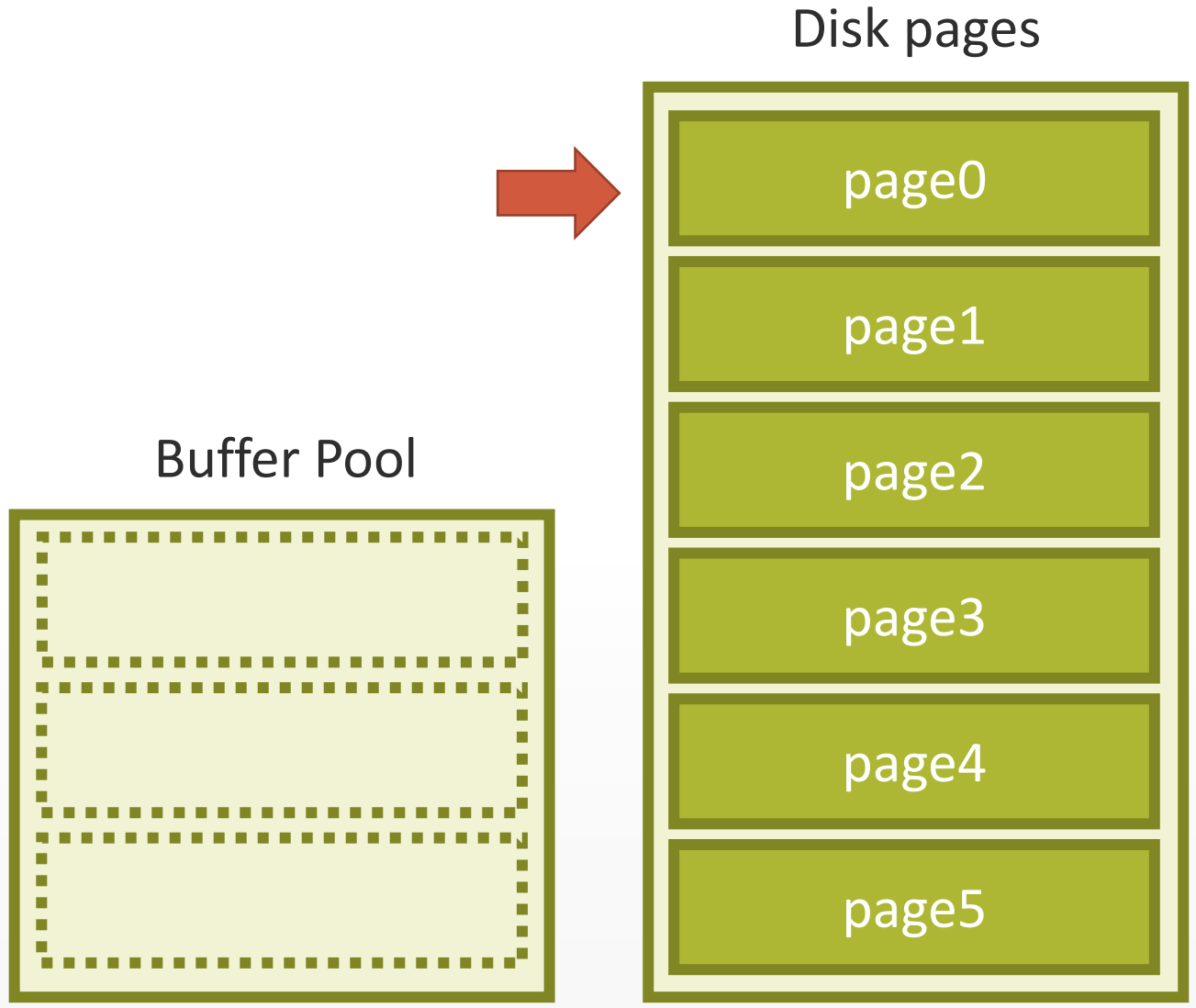
- Većina DBMS-ova ne koristi samo jedan bafer pul za podršku radu celog sistema/svih tredova.
  - Per-database buffer pool
  - Per-page type buffer pool
  - Per-table buffer pool

PRE-FETCHING

- Ukoliko se zahtevi mogu predvideti (sekventcijalna pretraga) tada se strane mogu učitavati i unapred – pre-fetched
- Na osnovu plana izvršavanja upita DMBS-ovi (mogu da) vrše dobavljanje stranica pre nego što je za njima iskazana potreba.
- Primeri:
  - Sekvencijalno čitanje
  - Sekvencijalna pretraga indeksa

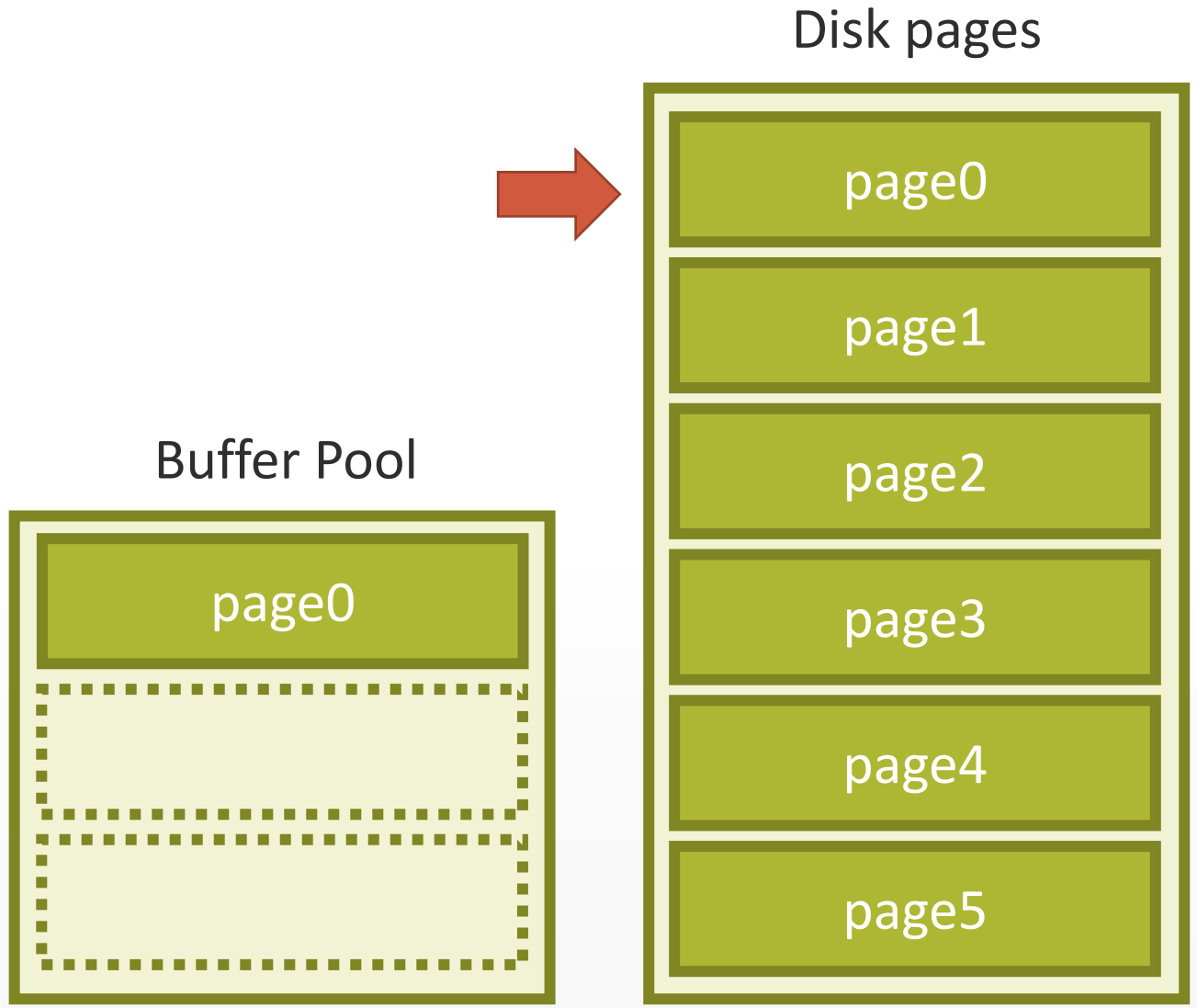
PRE-FETCHING

- Sekvencijalno čitanje



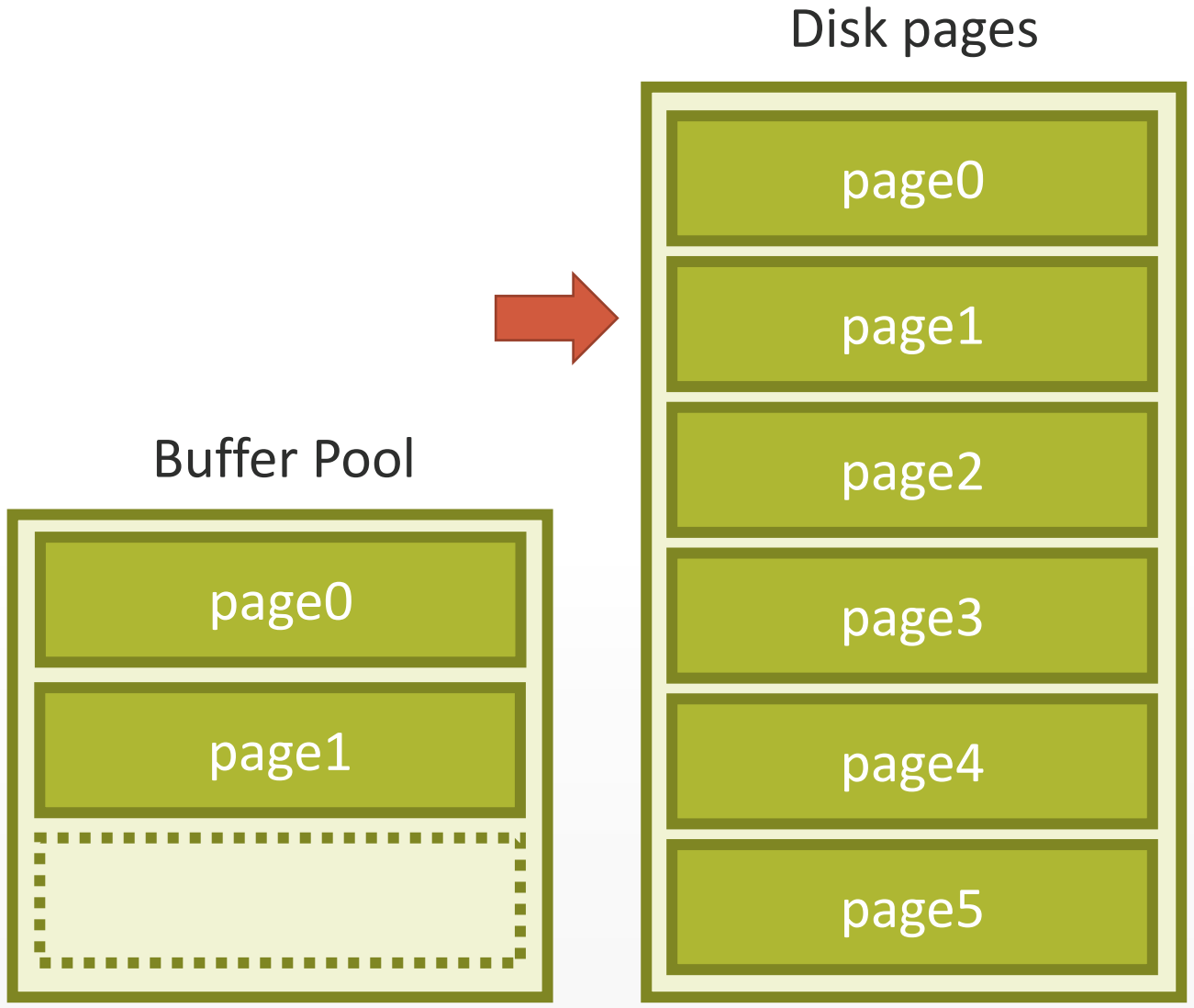
PRE-FETCHING

- Sekvencijalno čitanje



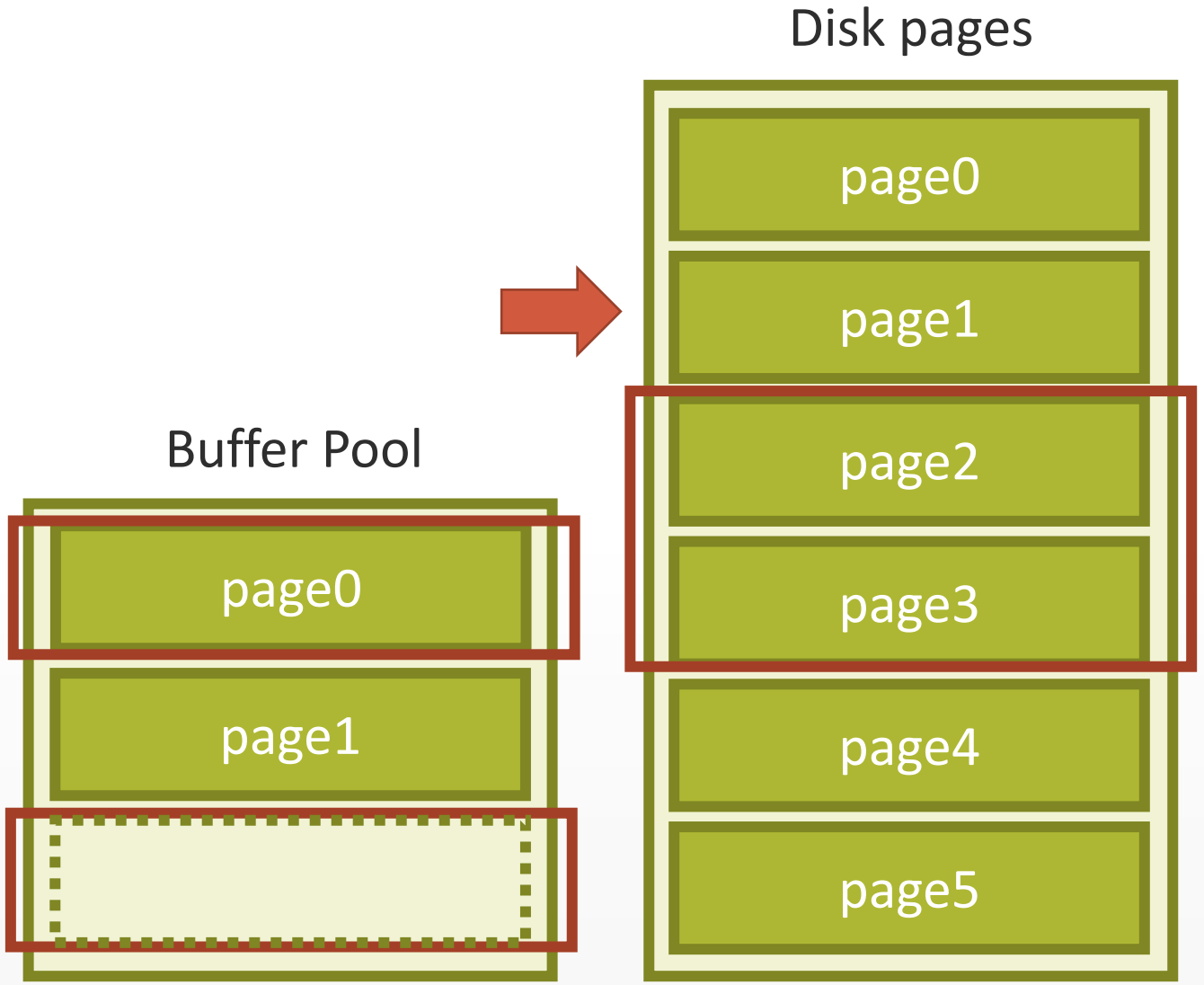
PRE-FETCHING

- Sekvencijalno čitanje



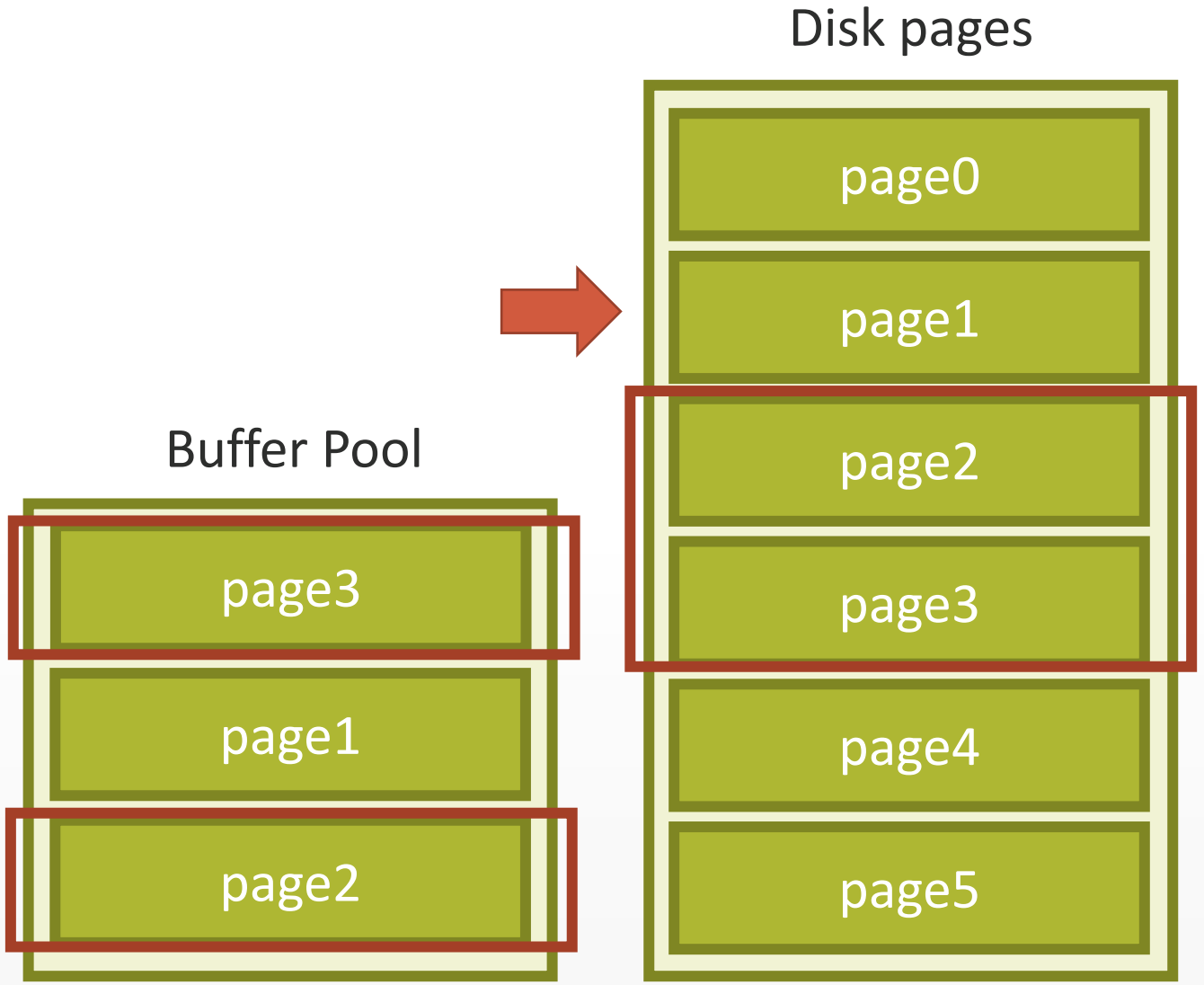
PRE-FETCHING

- Sekvencijalno čitanje



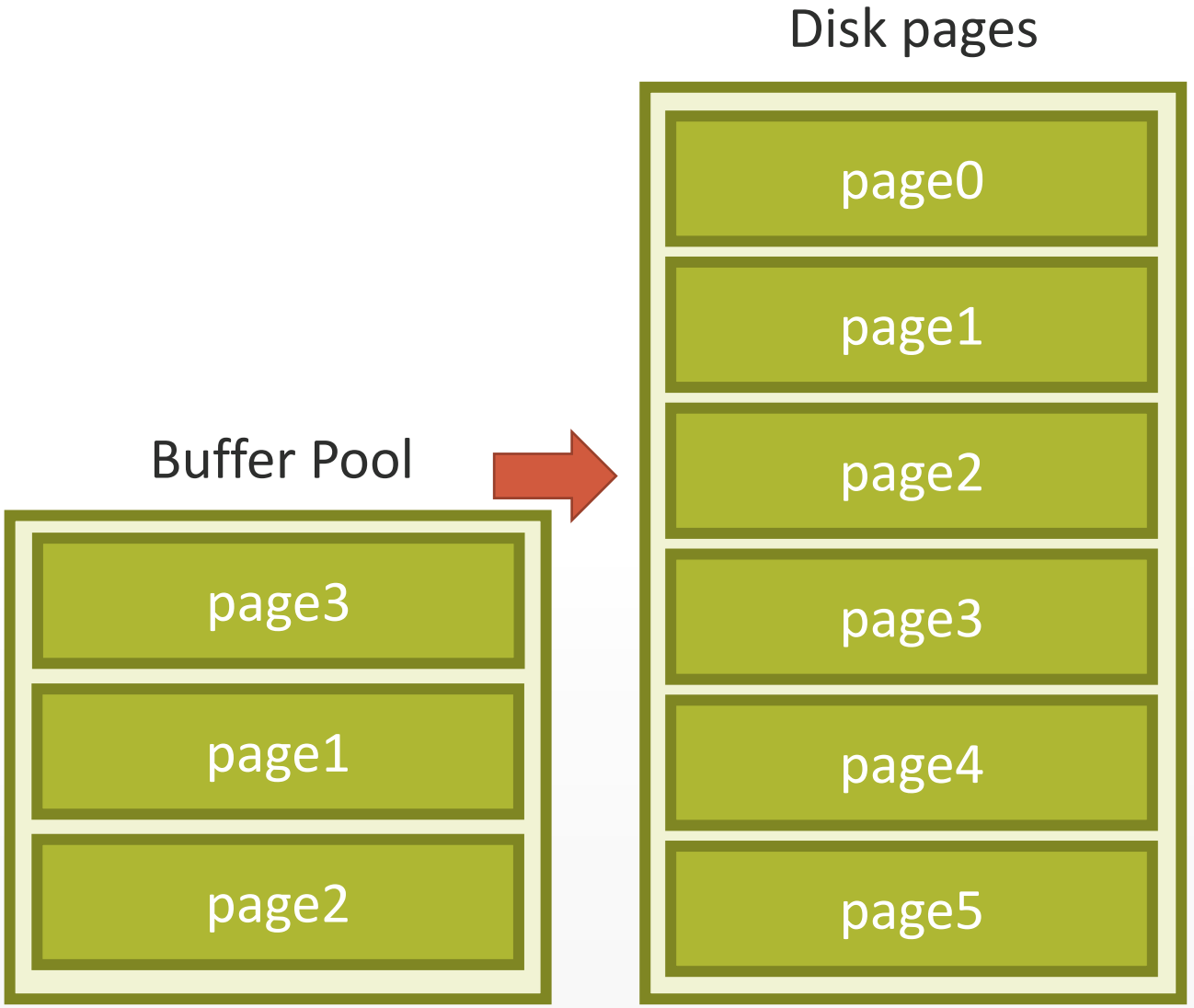
PRE-FETCHING

- Sekvencijalno čitanje

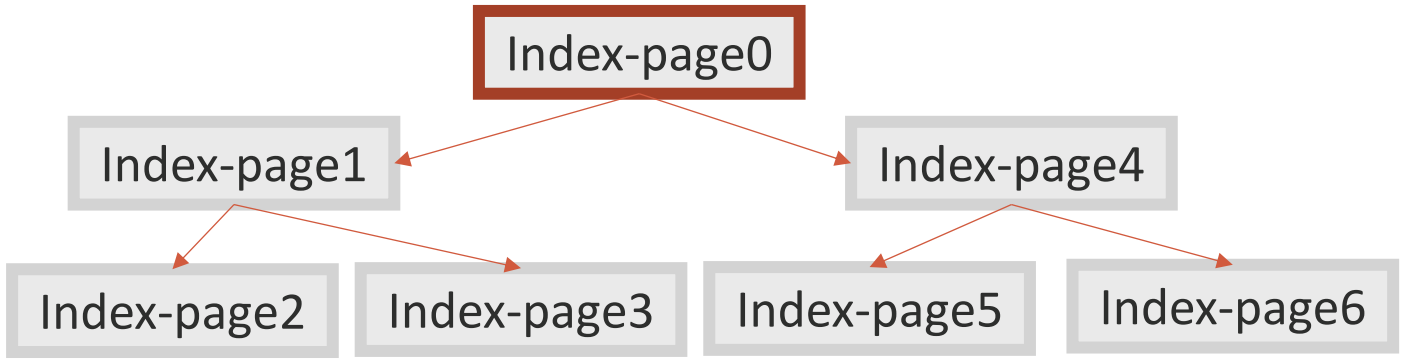


PRE-FETCHING

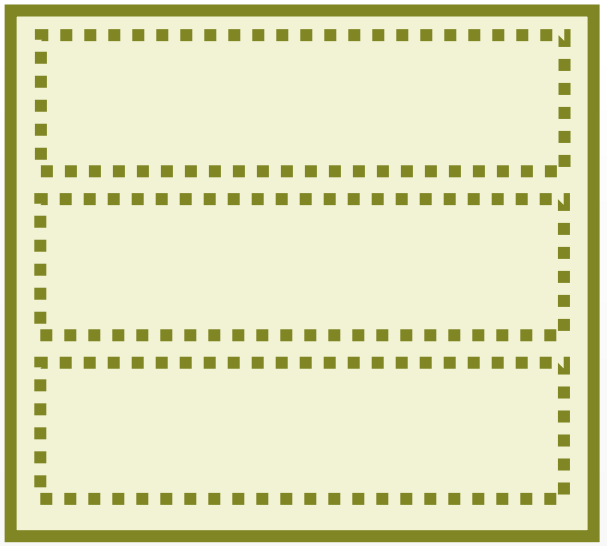
- Sekvencijalno čitanje



PRE-FETCHING



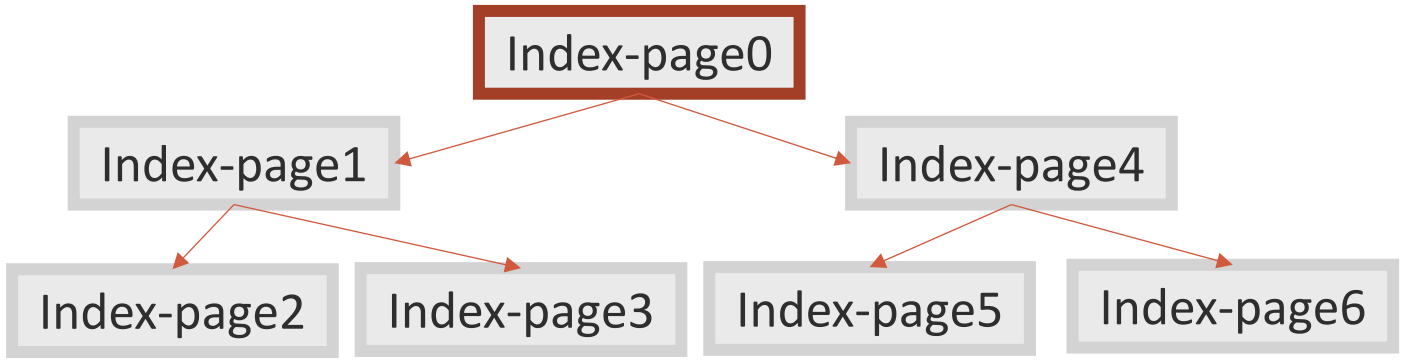
Buffer Pool



Disk pages

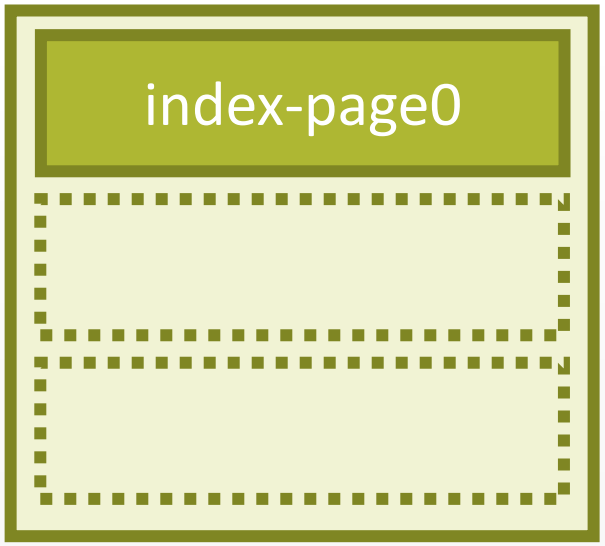
- Sekvencijalno čitanje indeksa

PRE-FETCHING



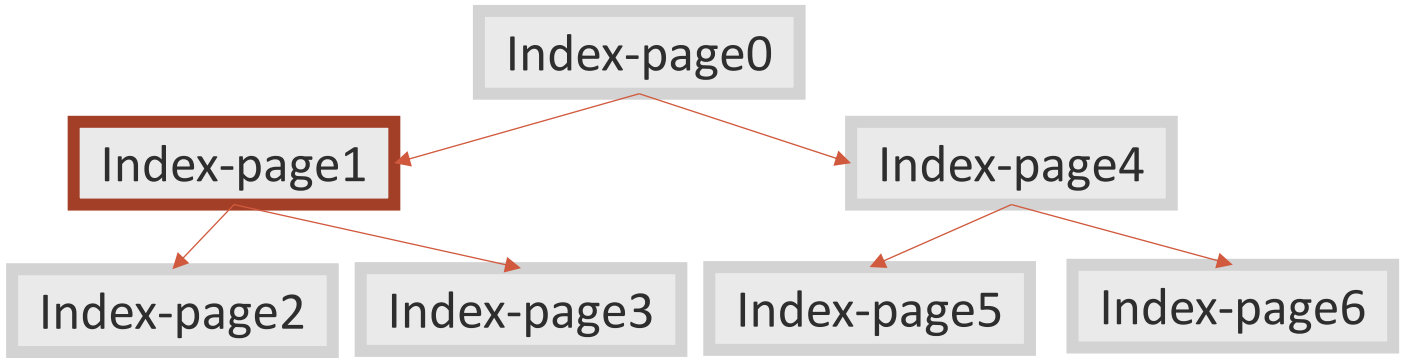
Disk pages

Buffer Pool



- Sekvencijalno čitanje indeksa

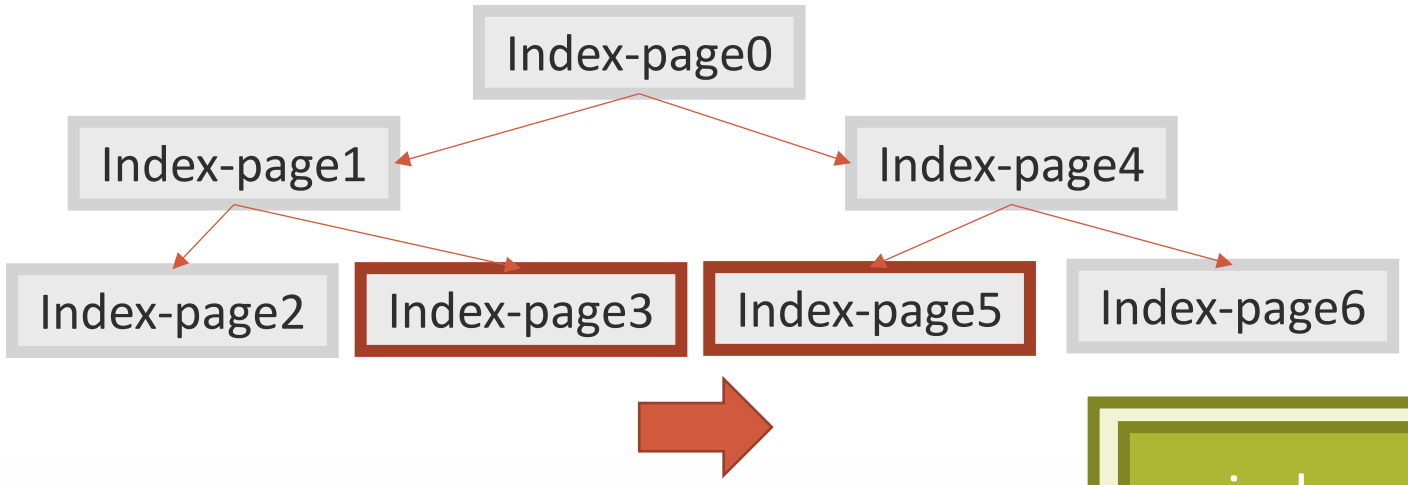
PRE-FETCHING



Buffer Pool

- Sekvencijalno čitanje indeksa

PRE-FETCHING



Disk pages

Buffer Pool

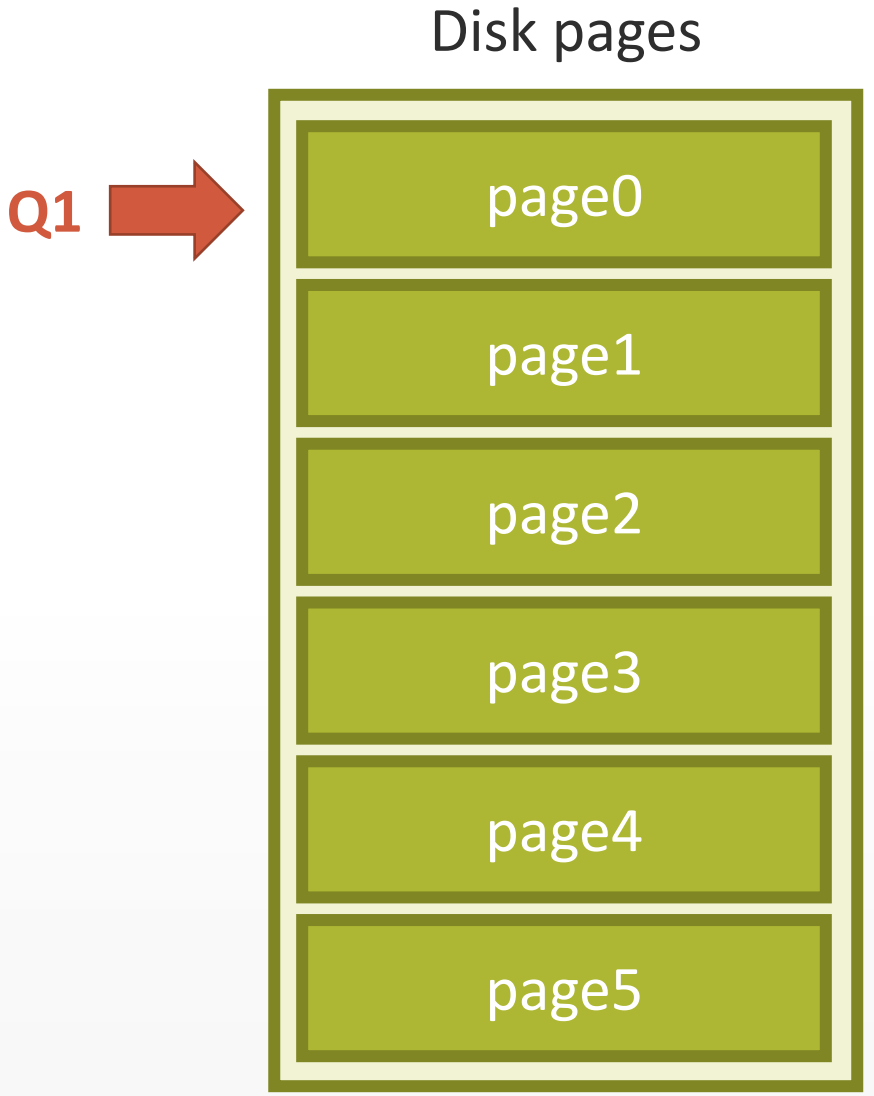
- Sekvencijalno čitanje indeksa

DELJENJE KURSORA

- DBMS može da dozvoli da više upita koristi jedan kursor za čitanje tabele ili međurezultata.
- Ako se pokrene izvršavanje upita koji čita podatke koje već neko drugi čita, DBMS će preusmeriti kursor na postojeći.
  - DMS vodi evidenciju o tome gde se novi upis pridružio u čitanju.
- MS SQL, IBM DB2
- Oracle podržava deljenje kursora samo za identične upite.

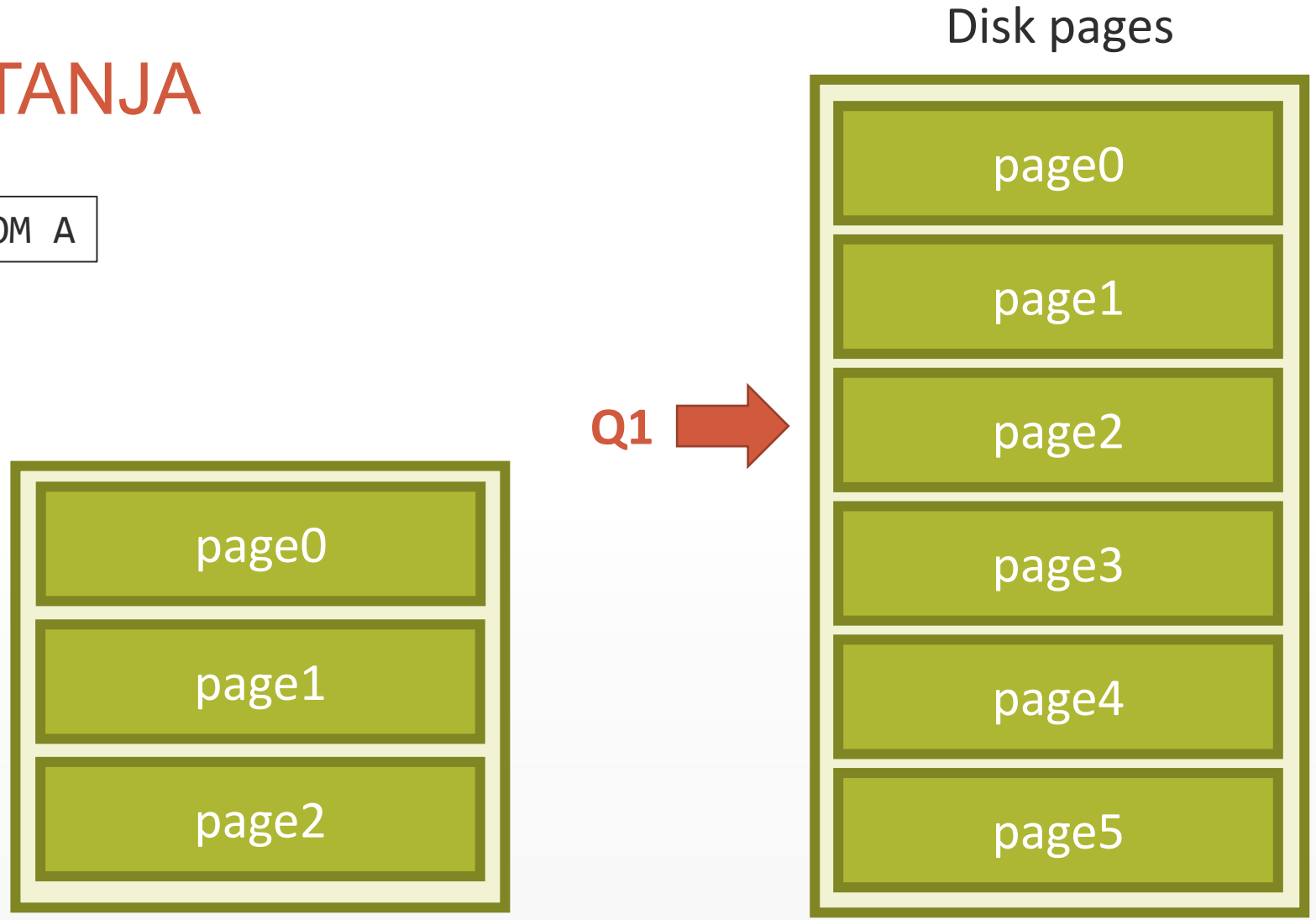
DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`



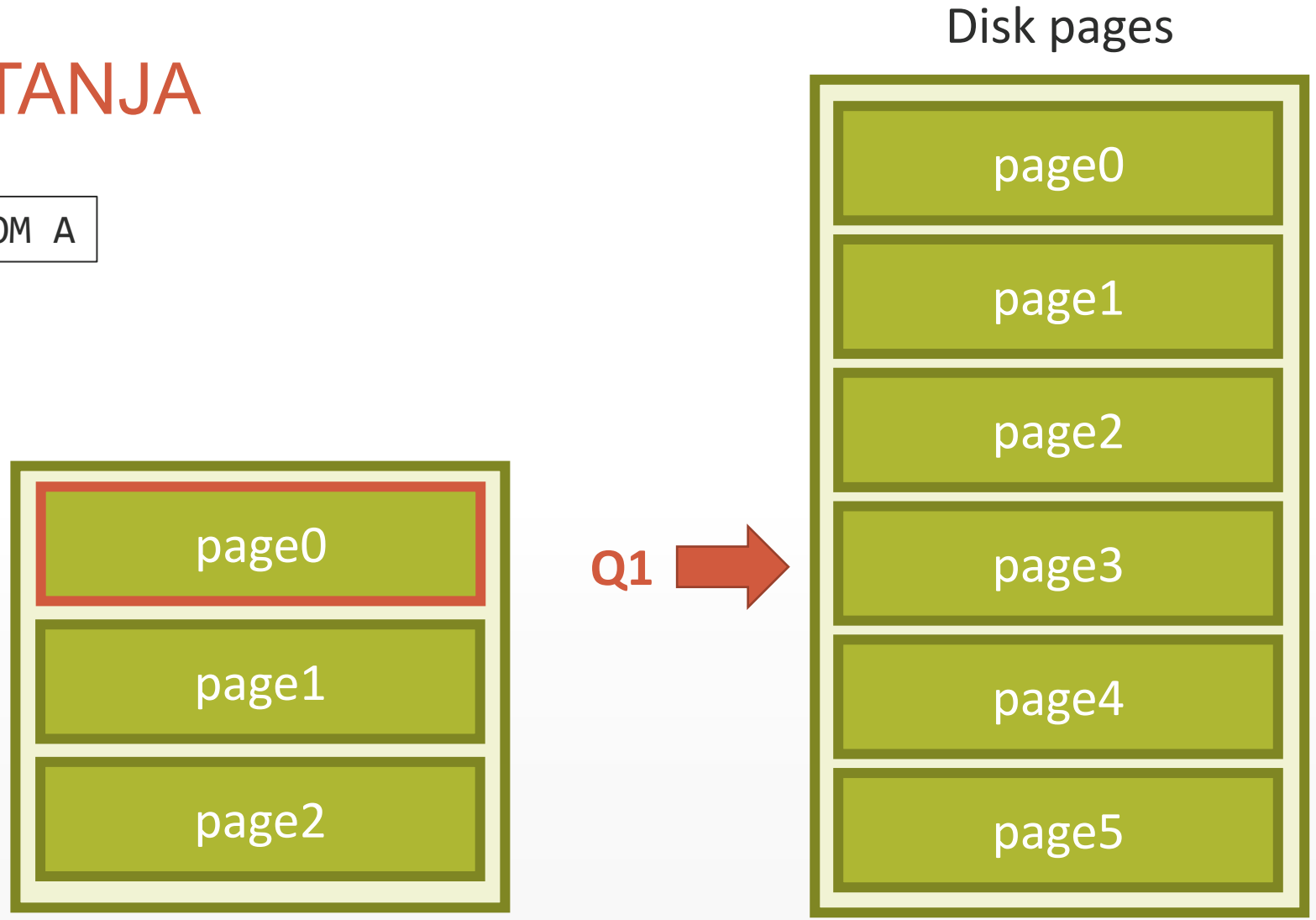
DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`



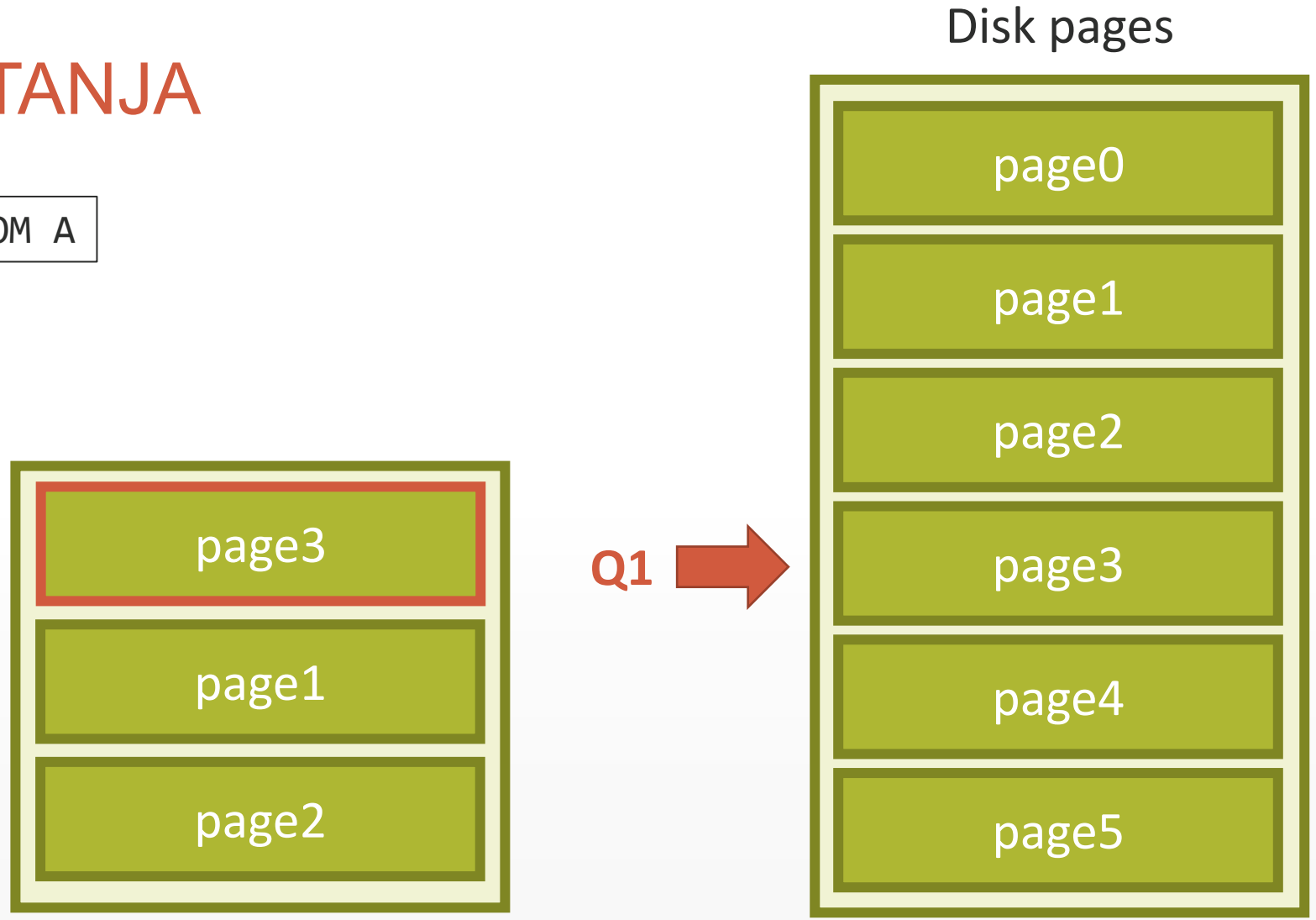
DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`



DELJENJE ČITANJA

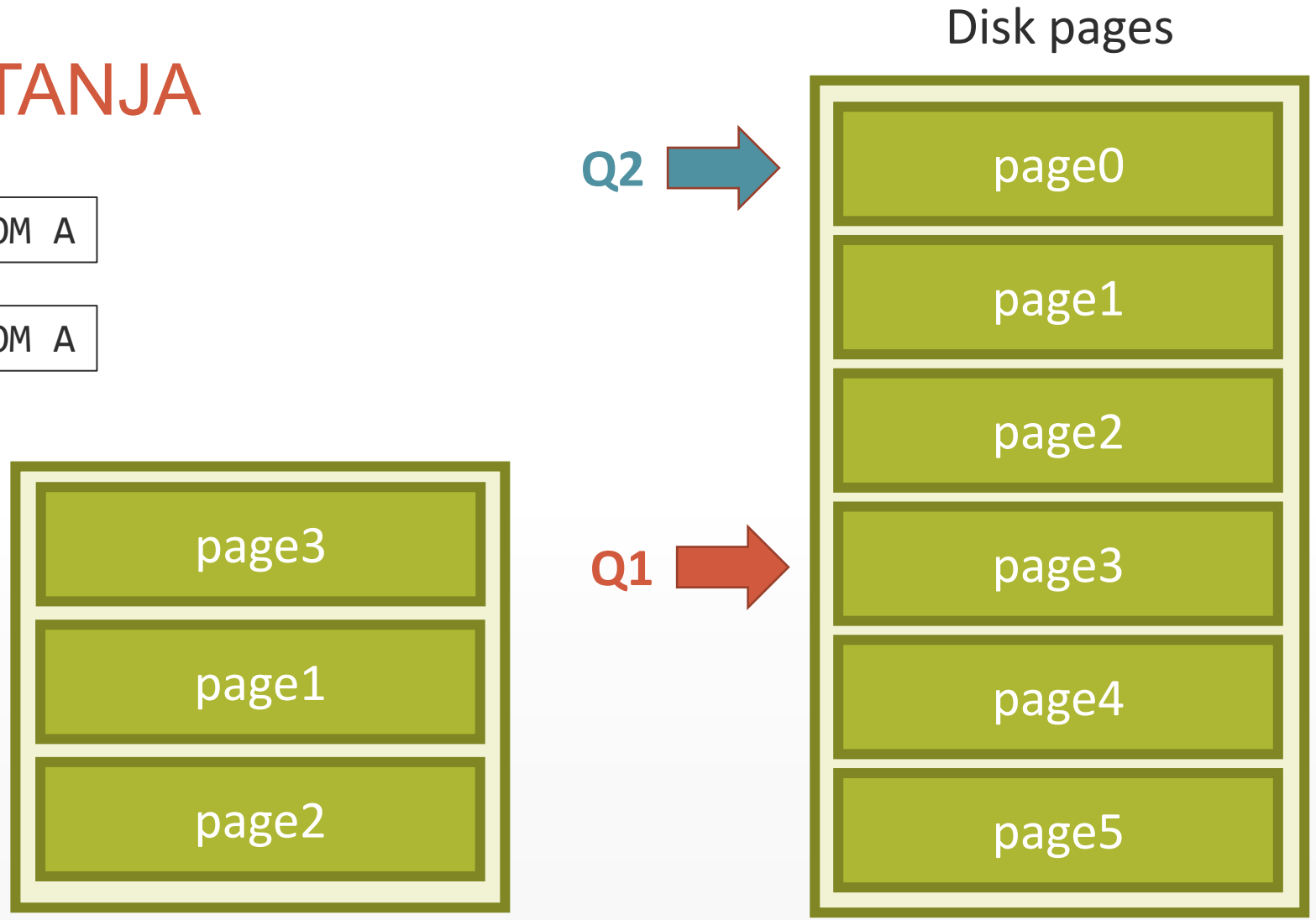
Q1 `SELECT SUM(VAL) FROM A`



DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`

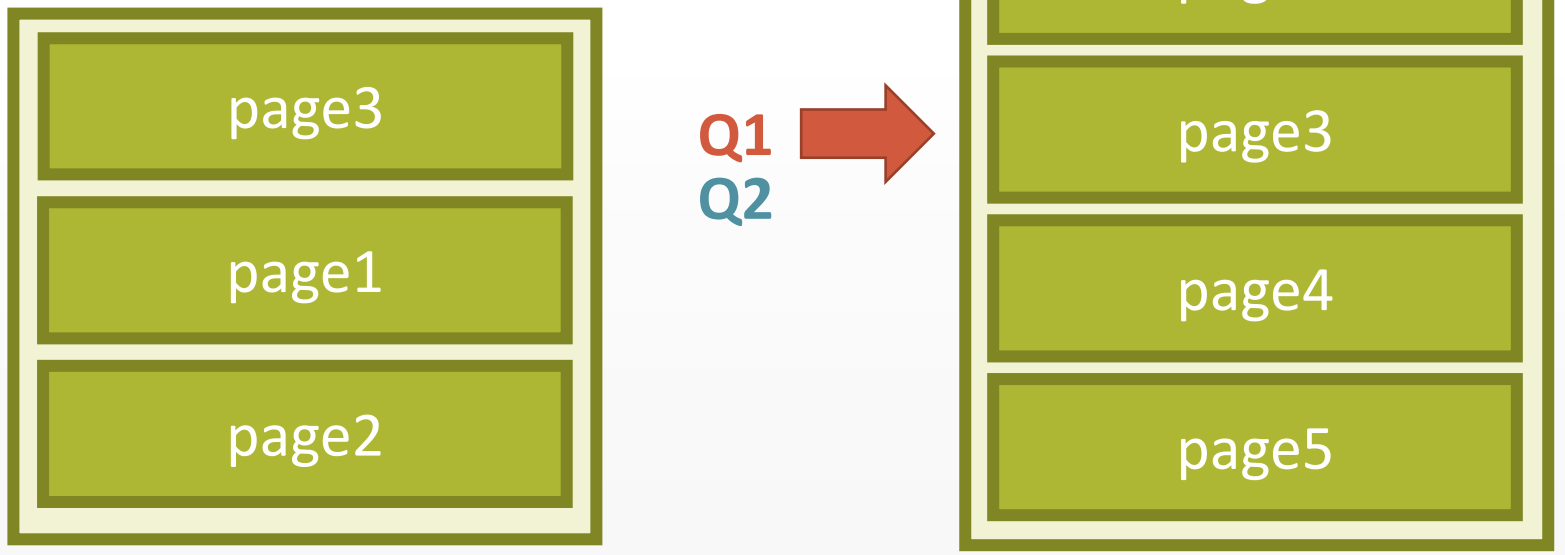
Q2 `SELECT AVG(VAL) FROM A`



DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`

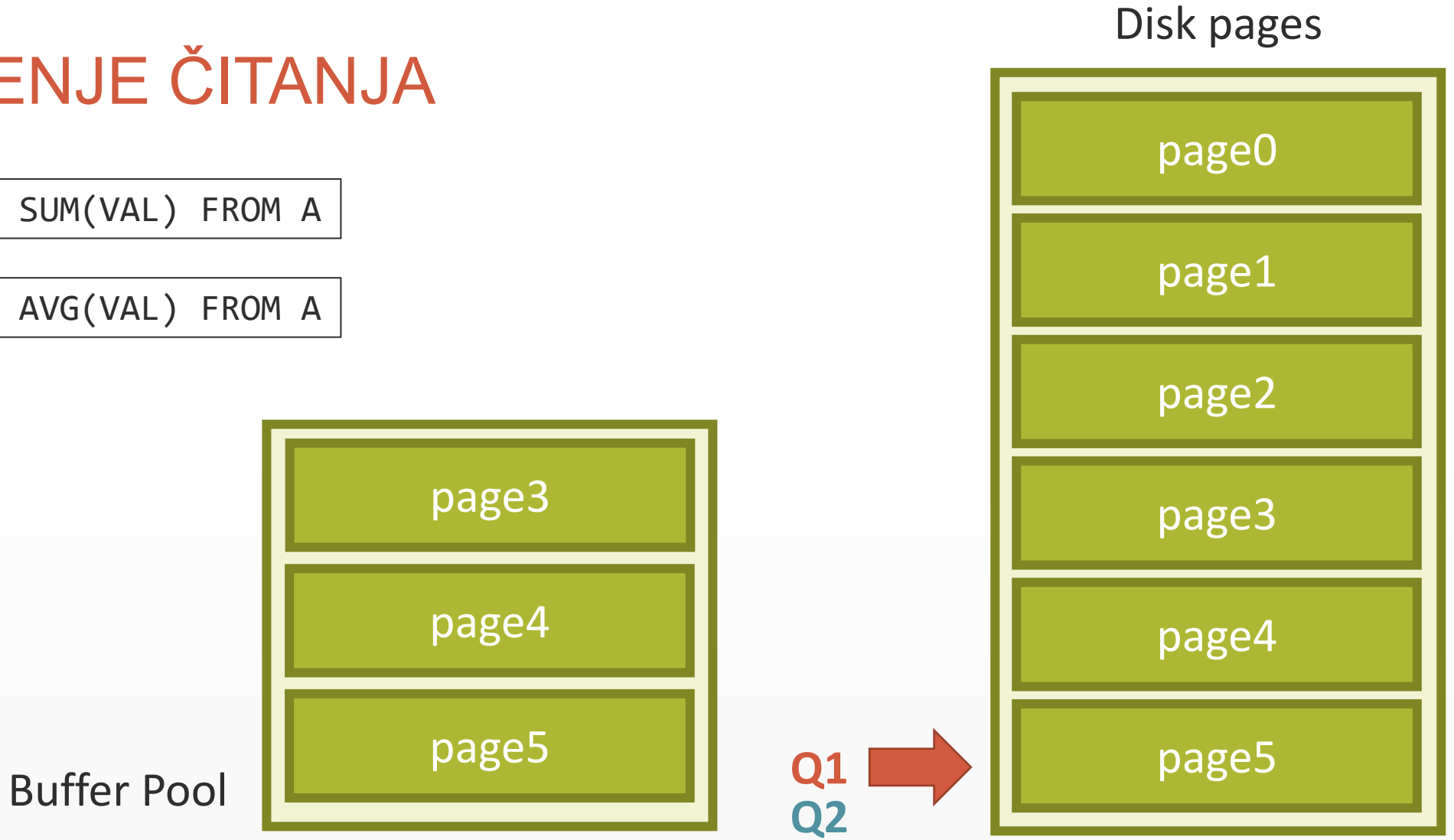
Q2 `SELECT AVG(VAL) FROM A`



DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`

Q2 `SELECT AVG(VAL) FROM A`



DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`

Q2 `SELECT AVG(VAL) FROM A`

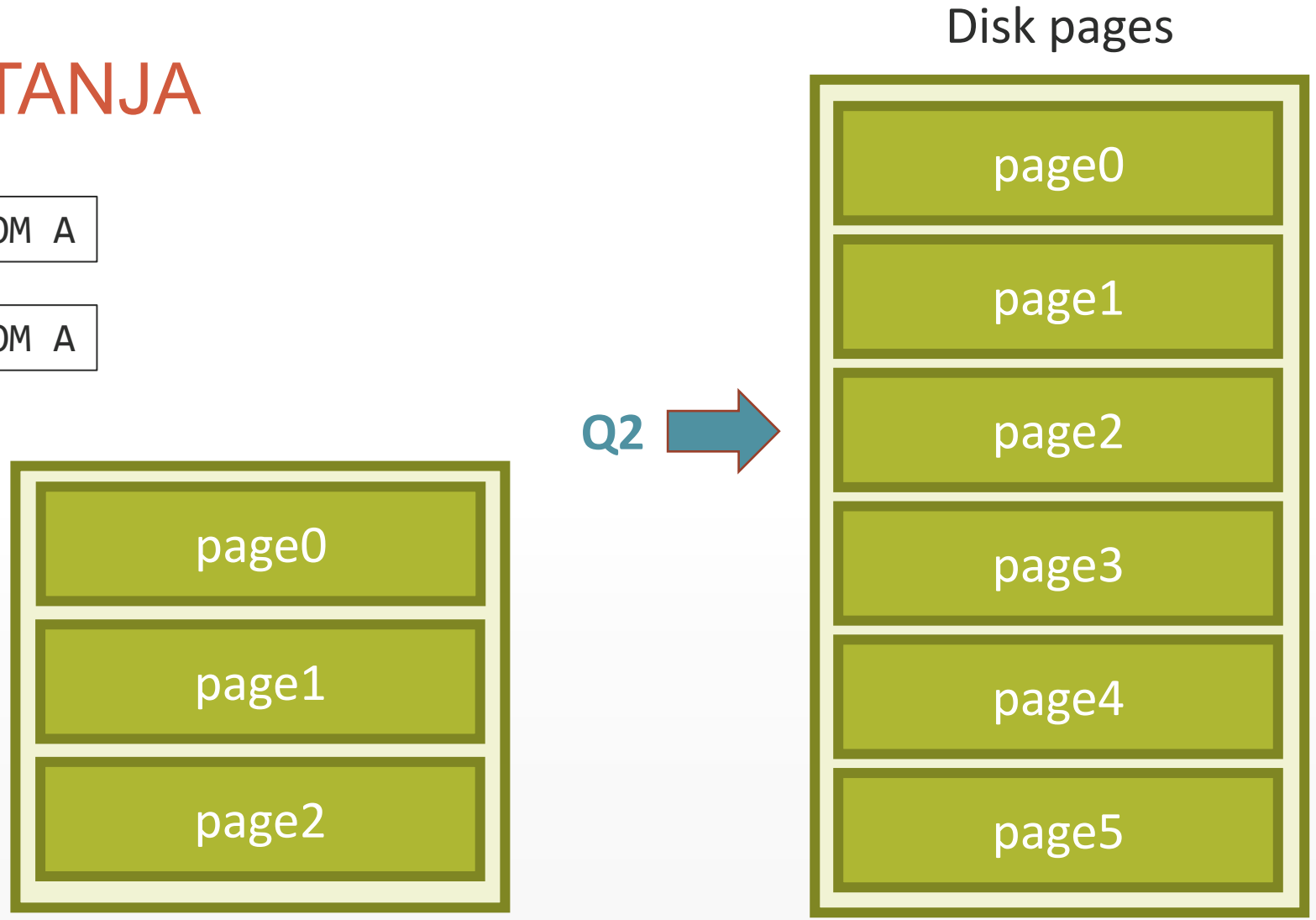
Buffer Pool

Q2

DELJENJE ČITANJA

Q1 `SELECT SUM(VAL) FROM A`

Q2 `SELECT AVG(VAL) FROM A`



Zaobilaženje bafer pula

- Operatori (koji obavljaju operacije nad podacima) mogu biti definisani tako da ne koriste bafer pul.
- Operator sekvencijalnog skeniranja ne skladištiti preuzete stranice u bafer pulu. Umesto toga, koristi se lokalna memorija operatora.
  - Ako operator treba da pročita veliku sekvencu stranica koje su susedne na disku i koje se neće ponovo koristiti.
- Zaobilaženje bafer pula se, takođe, može koristiti za privremene podatke potrebne operatorima sortiranja ili spajanja.