

Organizacija podataka u slogu

Tuple layout

Slogovi i polja (vrednosti atributa)

- Slog predstavlja kolekciju polja (koja su slogom povezana u logičku celinu)
 - Slog ima definisanu strukturu / format
 - Polje nosi podatak koji je određenog tipa
 - Slogovi se beleže kao niz bitova/bajtova. **Zadatak DBMSa je da ih protumači i izdvoji pojedine delove kao vrednosti atributa jedne torke.**
 - Pitanja:
 - Kako se beleže formati?
 - Kako se beleže pojedini tipovi podataka?
- Dodatno: kako oba utiču na organizaciju slogova u fajlu?
-

Tipovi podataka / polja – celobrojne vrednosti

- Većina DBMSova beleži celobrojne vrednosti koristeći sopstvene C/C++ tipove prema IEEE-754 standardu
- Vrednosti su fiksne dužine
- INTEGER, BIGINT, SMALLINT, ...

00000000

00100011

Tipovi podataka / polja – realne vrednosti

- Brojevi varijabiljne preciznosti
 - C/C++ tipovi prema IEEE-754 standardu
 - FLOAT, REAL

```
#include <stdio.h>

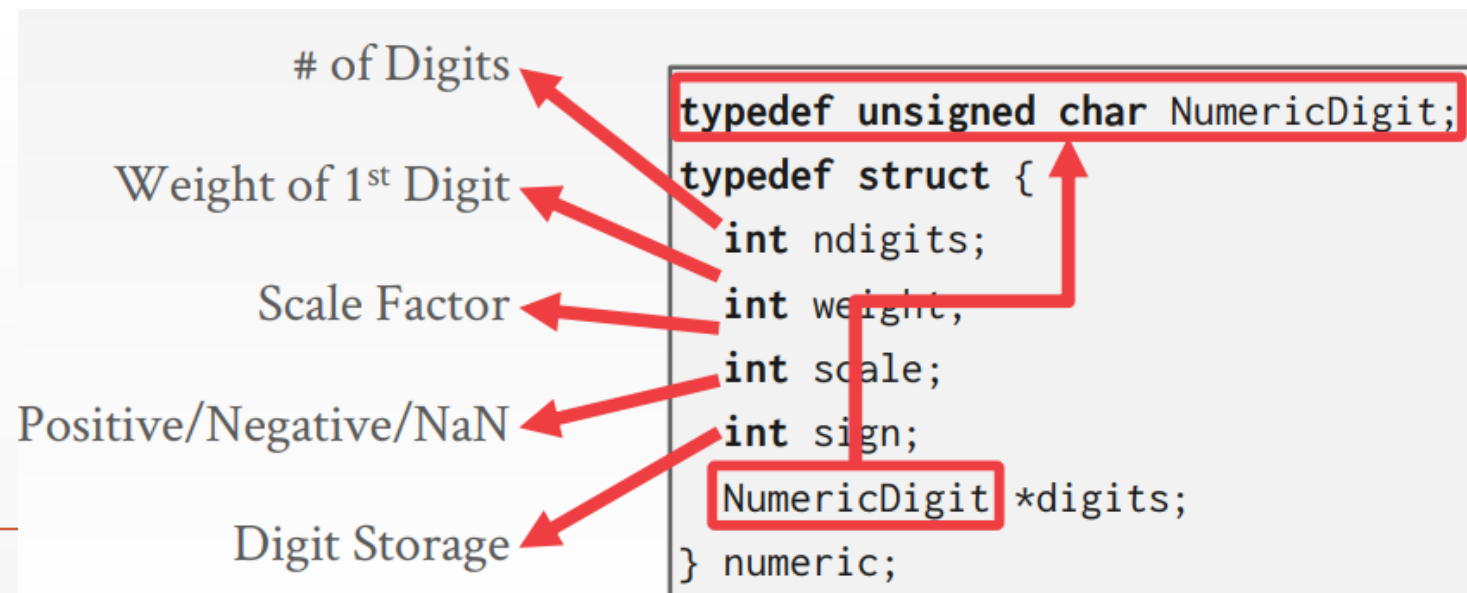
int main(int argc, char* argv[]) {
    float x = 0.1;
    float y = 0.2;
    printf("x+y = %.20f\n", x+y);
    printf("0.3 = %.20f\n", 0.3);
}
```

```
x+y = 0.30000001192092895508
0.3 = 0.299999999999999998890
```

Tipovi podataka / polja – realne vrednosti

- Brojevi fiksne preciznosti
- Proizvoljna preciznost i opseg
 - Obično sačuvani striktnom reprezentacijom, varijabilne dužine sa dodatnim metapodacima.
 - Operacije nad njima su sporije.
 - NUMERIC, DECIMAL

Postgres



Tipovi podataka / polja – karakteri, bool, stringovi

- Karakteri
 - Koriste se različite kodne šeme
 - Primer ASCII

Example:

A: 1000001

a: 1100001

5: 0110101

LF: 0001010

- Boolean

e.g., TRUE
FALSE

1111 1111
0000 0000

- String of characters

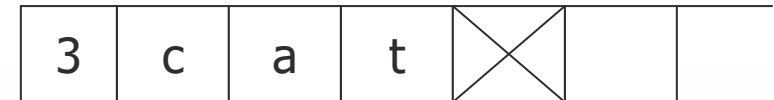
Null terminated

e.g.,



Data dužina

e.g.,



[Char, Varchar / Nchar, Nvarchar](https://docs.microsoft.com/en-us/sql/relational-databases/collations/collation-and-unicode-support?view=sql-server-ver15#Locale_Defn)

https://docs.microsoft.com/en-us/sql/relational-databases/collations/collation-and-unicode-support?view=sql-server-ver15#Locale_Defn

Tipovi podataka / polja – veliki podaci

- Bag of bits

Length	Bits
--------	------

BLOB (Binary Large Object)

- Većina DBMSova ne dozvoljava slogu da prevaziđe veličinu jedne strane.
 - Oni koji dozvoljavaju koriste overflow stranice (stranice prekoračenja), pri čemu svaki deo torke na jednoj strani sadrži referencu na stranicu u koju je smešten nastavak.
 - Neki sistemi dozvoljavaju čuvanje u eksternim fajlovima (na primer slike). Mana ovakvog skladištenja je to što su izmene nad ovako sačuvanim podacima van kontrole DBMSa (oporavak od otkaza, na primer, ih neće obuhvatiti).
-

Tipovi podataka / polja – datum, vreme

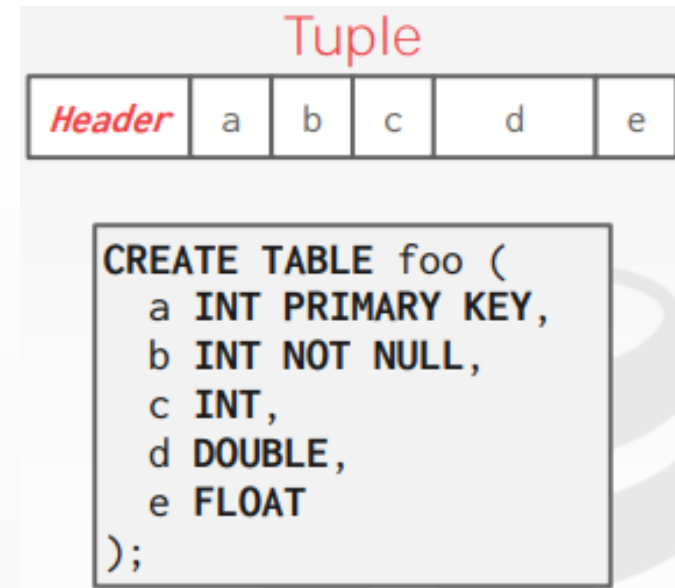
- TIME, DATE; TIMESTAMP
 - Datum
 - e.g.: - Integer, # days since Jan 1, 1900
 - 8 characters, YYYYMMDD
 - 7 characters, YYYYDDD
 - Vreme
 - e.g. - Integer, seconds since midnight
 - characters, HHMMSSFF
-

Sistemiški katalog

- Metapodaci o bazama podataka.
 - Čuva podatke opise:
 - Tabela, kolona, indeksa, pogleda
 - Kosrisnika, dozvola,
 - Kao i statistike
 - DMBS ima sopstvene kodove za beleženje i čitanje.
-

Tuple layout (Postavka sloga)

- Slogovi mogu sadržati zaglavlje u kom se mogu naći
 - Tip sloga ili celu šemu sloga
 - Dužinu sloga
 - Vremenski pečat
 - ...



Uključivanje šeme u slog

Example: Include record type in record

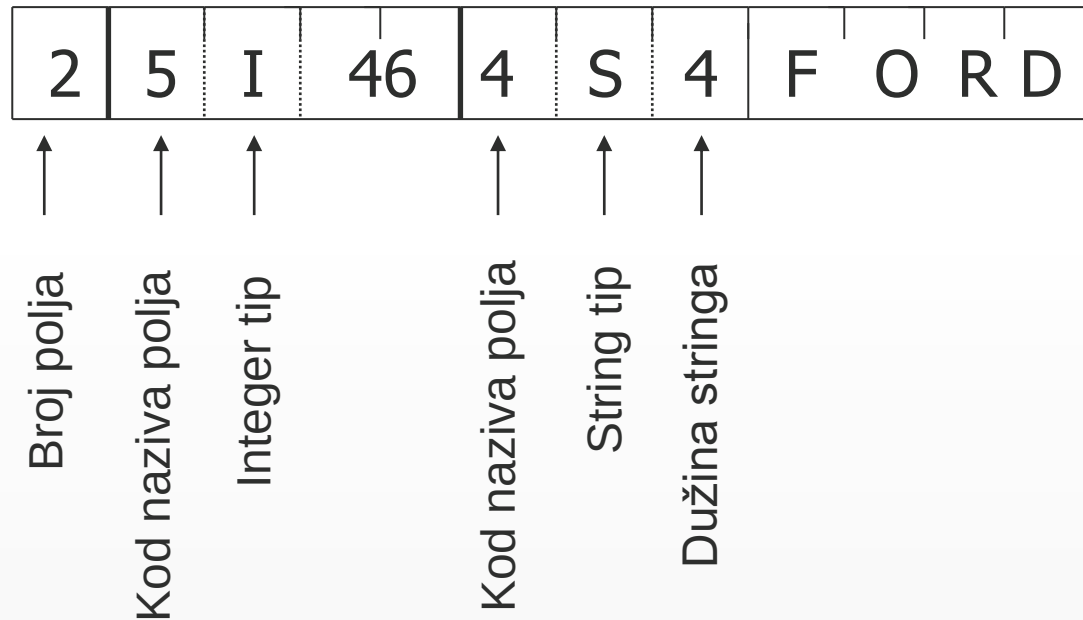


↑
Record type
tells me what
to expect
(i.e. points to
schema)

←
record length

Uključivanje šeme u slog

- Samoopisujući format



Kodovi naziva polja mogu biti bilo kakvi tagovi.

Tipovi slogova

- FORMAT: FIKSNI ili PROMENLJIVI
- DUŽINA: FIKSNA ili PROMENLJIVA

Struktura, tj. format sloga je opisan **ŠEMOM**.

Tipovi slogova – Fiksni format

- Šema sloga sadrži
 - Broj polja
 - Tip svakog polja
 - Redosled polja
 - Značenje svakog polja

Employee record

(1) E#, 2 byte integer

(2) E.name, 10 char.

(3) Dept, 2 byte code

} Schema

55	s	m	i	t	h					02
----	---	---	---	---	---	--	--	--	--	----

83	j	o	n	e	s					01
----	---	---	---	---	---	--	--	--	--	----

} Records

Promenljivi format

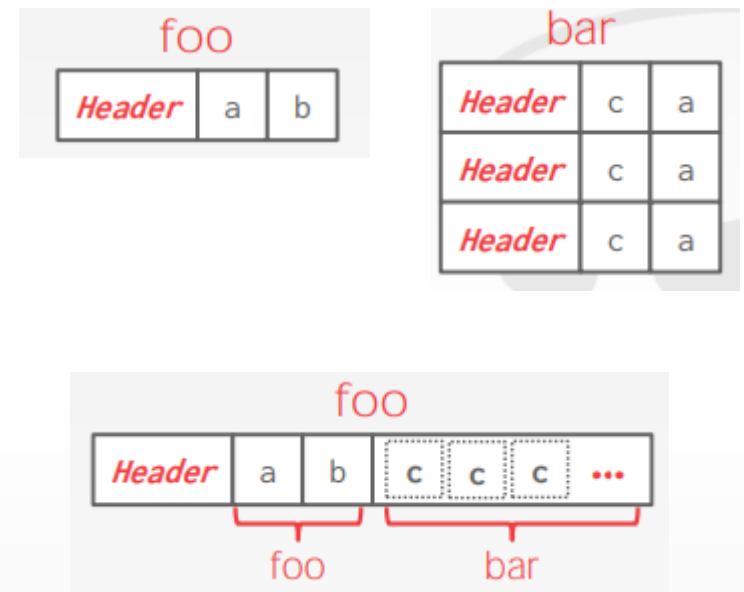
- Pogodan za
 - Retke slogove
 - Ponavljajuća polja
 - Promenljive definice formata
- Primer: promenljivi format sa ponavljajućim poljima

Employee → one or more → children

3	E_name: Fred	Child: Sally	Child: Tom
---	--------------	--------------	------------

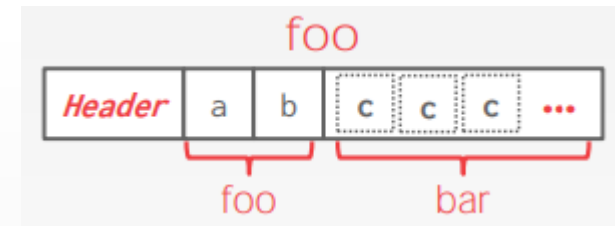
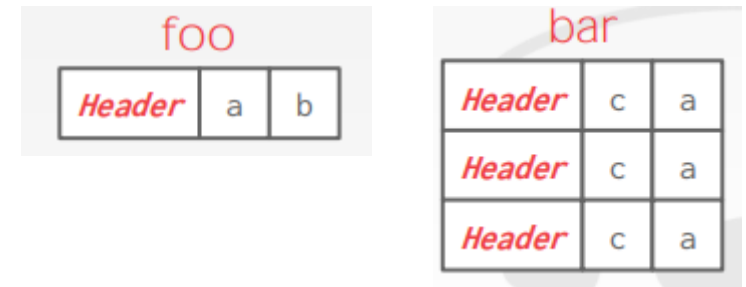
Denormalizovane torke

- Na jednoj strani se mogu naći i torke više tabela, vezanih stranim ključem.
- Prednosti i mane?



Denormalizovane torke

- Na jednoj strani se mogu naći i torke više tabela, vezanih stranim ključem.
- Zahtevi za odgovarajućim join-om će imati manje čitanja.
- Ažuriranje je skuplje.
- Nisu sve pretrage 'jeftinije'.



Četiri osnovna načina beleženja više slogova unutar strane

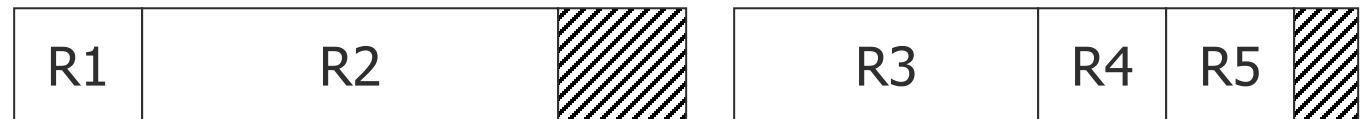
- Razdvajanje slogova markerima



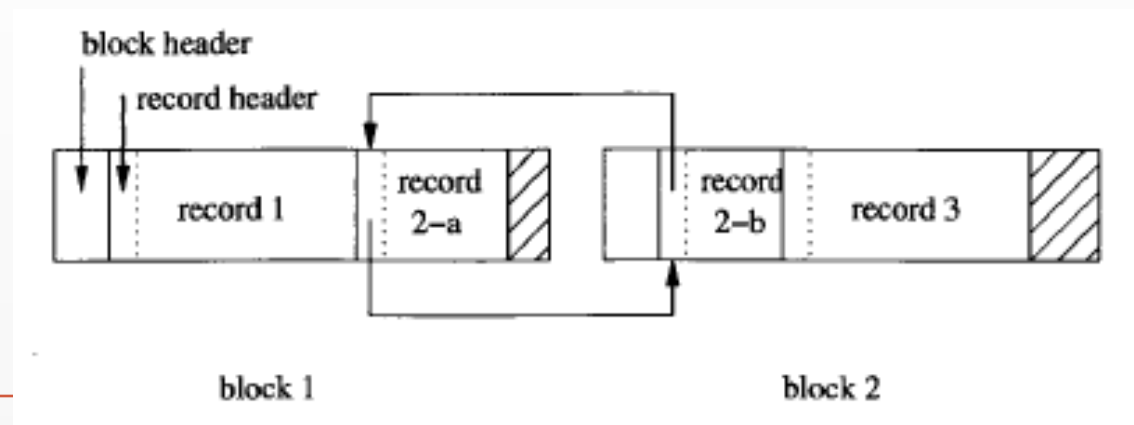
- Slogovi fiksne dužine – ne moraju imati separator
 - Uvođenje specijalnog markera kao separatora
 - Zadavanje dužine sloga (ili ofseta)
 - Unutar svakog sloga
 - U zaglavlju bloka
-

Četiri osnovna načina beleženja više slogova unutar strane

- Razdvajanje slogova markerima
- Razapeti vs. celoviti
- Nerazapeti slogovi – jednostavniji, postojanje neiskorišćenog prostora

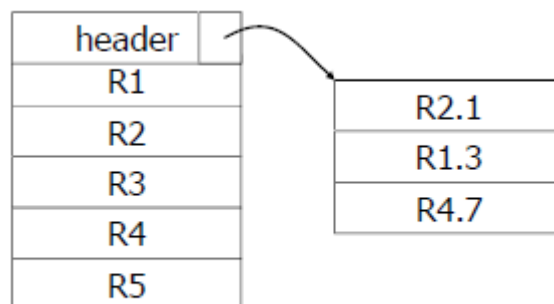
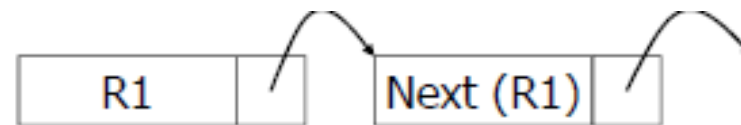


- Razapeti – neophodni kada je veličina sloga veća od veličine sloga



Četiri osnovna načina beleženja više slogova unutar strane

- Razdvajanje slogova markerima
- Razapeti vs. celoviti
- Uređivanje
- Fajlovi čiji su slogovi uređeni prema vrednosti ključa – fajlovi sa sekvencijalnim uređenjem. (uređenje ubrzava pretragu, doduše po ključu)



- Opcije za uređivanje
 - Slogovi su fizički uzastopno smešteni
 - Slogovi čine povezanu listu
 - Prostor prekoračenja

Četiri osnovna načina beleženja više slogova unutar strane

- Razdvajanje slogova markerima
- Razapeti vs. celoviti
- Uređivanje
- Preusmeravanje

