

Indeksy

Wprowadzenie

- Indeks zdefiniowany na pliku jest dodatkową strukturą fizyczną, której celem jest przyspieszenie wykonywania operacji, które niesą dostatecznie wspierane przez podstawowe struktury logiczne danych
- Indeksy są zakładane na pojedynczych atrybutach lub zbiorach atrybutów relacji. Atrybuty te noszą nazwę **atrybutów indeksowych**.
- Indeks jest uporządkowanym plikiem rekordów indeksu (ang. data entry) o stałej długości. Rekordy indeksu zawierają dwa pola: **klucz** reprezentujący jedną z wartości występujących w atrybutach indeksowych relacji oraz **wskaźnik** do bloku danych zawierający krotkę, której atrybut indeksowy równy jest kluczowi

Wprowadzenie

- Każdy rekord indeksu, oznaczony k^* , zawiera dostateczną informację umożliwiającą wyszukanie (jednego lub więcej) rekordów danych o wartości klucza k ,
- Pytanie:
 - W jaki sposób rekordy indeksu powinny być zorganizowane, aby efektywnie wspierać wyszukiwanie rekordów o danej wartości klucza?
 - Co powinien zawierać rekord indeksu?

Rekordy indeksu

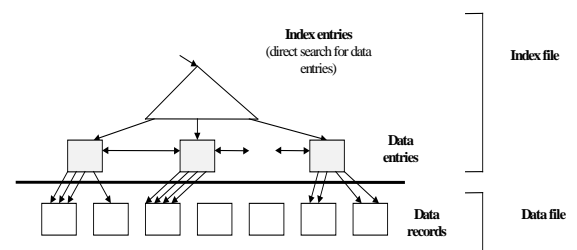
- Rekord indeksu k^* umożliwia wyszukanie rekordów danych o wartości klucza k
- Cztery typy rekordów indeksu:
 - Rekord indeksu k^* jest rekordem danych (o wartości klucza k)
 - Rekord indeksu jest parą $\langle k, rid \rangle$, gdzie rid jest identyfikatorem id rekordu danych o wartości klucza k
 - Rekord indeksu jest parą $\langle k, rid-list \rangle$, gdzie $rid-list$ jest listą identyfikatorów rekordów danych ids o wartości klucza k
 - Rekord indeksu jest parą $\langle k, bitmapa \rangle$, gdzie $bitmapa$ jest wektorem 0 i 1 reprezentującym zbiór rekordów danych

Indeksy zgrupowane i nie zgrupowane

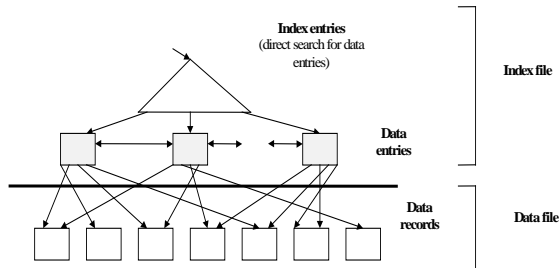
Jeżeli plik danych jest zorganizowany w ten sposób, że porządek rekordów danych jest taki sam lub zbliżony do uporządkowania rekordów indeksu, to taki indeks nazywamy indeksem zgrupowanym (ang. **clustered** index)

- Wariant (1) jest indeksem zgrupowanym z definicji
- Warianty (2), (3), i (4) mogą być indeksami zgrupowanymi wtedy i tylko wtedy, gdy rekordy danych są uporządkowane wg. wartości atrybutu indeksowego

Indeks zgrupowany



Indeks nie zgrupowany



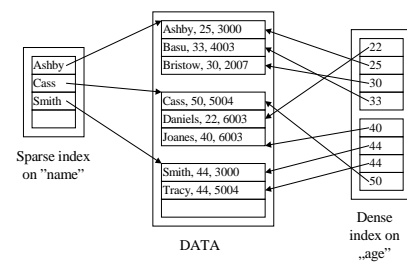
Indeksy zgrupowane i nie zgrupowane

- Komentarz:
 - Rekordy danych są rzadko uporządkowane (wysoki koszt pielęgnacji plików uporządkowanych) – wymaga stron przepełnienia
 - Indeksy zgrupowane są kosztowne w pielęgnacji
 - Plik danych może być zgrupowany tylko wg. jednego atrybutu indeksowego
 - Zapytania z przedziałem wartości: koszt zapytania zależy od typu indeksu

Indeksy rzadkie i indeksy gęste

- Indeks nazywamy **gęstym** jeżeli zawiera rekord indeksu per rekord danych
- Indeks nazywamy **rzadkim** jeżeli zawiera rekord indeksu per strona rekordów danych w pliku danych
- Plik danych nazywamy **odwróconym** względem atrybutu jeżeli istnieje indeks wtórny na tym atrybucie
- Plik danych nazywamy **w pełni odwróconym** jeżeli istnieje indeks wtórny na każdym atrybucie wtórnym

Indeksy rzadkie i indeksy gęste



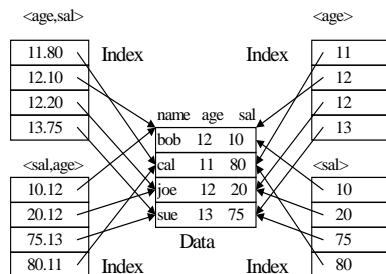
Indeksy podstawowe i indeksy wtórne

- Indeks zdefiniowany na zbiorze atrybutów zawierającym klucz podstawowy relacji nazywamy **indeksem podstawowym**
- Indeks, który nie jest indeksem podstawowym, nazywamy **indeksem wtórnym**
- **Duplikatami** nazywamy rekordy indeksu o tej samej wartości klucza
- Indeks podstawowy nie zawiera duplikatów
- Indeks wtórny może zawierać duplikaty

Indeksy o kluczach złożonych

- Klucz indeksu może zawierać kilka atrybutów – taki klucz nazywamy **kluczem złożonym**
- Problemy z realizacją zapytań:
 - zapytania punktowe: age=20; sal=10
 - zapytania z przedziałem wartości: age > 20

Indeksy o kluczach złożonych



Specyfikacja indeksów w SQL-92

```
CREATE INDEX IndAgeRating ON Students
WITH STRUCTURE = BTREE,
KEY = (age,gpa)
```