

DBVS Daugiadimensiniai indeksai

Paruošė:

Kazimieras Mickus

2002

Daugiadimensiniai indeksai

- Daugelio dimensijų reikalaujantys taikymai
- Daugiadimensinių indeksų struktūros
 - Grid failai
 - Padalintos hash funkcijos
 - Daugelio raktų indeksai
 - kd-Medžiai
 - Ketvirtainiai medžiai
 - R-Medžiai
 - Bitų žemėlapių (bitmap) indeksai

Daugelio dimensijų reikalaujantys taikymai

- Geografinės informacinės sistemos
 - Dalinio atitikimo užklauskos
 - Sferinės užklauskos
 - Artimiausio kaimyno užklauskos
 - Kur-aš-esu užklauskos
- Duomenų kubai
 - Sprendimų palaikymo sistemos

Daugiadimens. taikymų realizacija įprastom priemonėm (1)

- SQL daugiadimensinės užklausos
 - Formulių panaudojimas artimiausio kaimyno užklausiai
 - Patogus stačiakampių atvaizdavimas GIS sistemose
 - Faktų ir dimensijų lentelės duomenų kubams
- Sferinių užklausų vykdymas naudojant paprastus indeksus
 - Kiekvienai dimensijai antrinis indeksas (B-Medis)

Daugiadimens. taikymų realizacija įprastom priemonėm (2)

- Artimiausio kaimyno užklausų vykdymas naudojant paprastus indeksus
 - Naudotinos sferinės užklausos
 - Gali nebūti taško pažymėtam regione
 - Artimiausias taškas regione nebūtinai bus artimiausias apskritai

Hash-tipo struktūros daugiadimens. duomenims (1)

- Grid failai
 - Struktūra
 - Erdvė, padalinta į režius pagal kiekvieną dimensiją
 - Kiekvienas gautas regionas – bucket'as
 - Masyvas iš tiek dimensijų, kiek yra ir duomenų faile
 - Paieška
 - Įterpimas
 - Surandam ir įterpiam
 - Jei netelpa kuriam naują perteklinį bloką bucket'ui
 - Jei bloką bucket'e per daug reorganizuojam režių ribas

Hash-tipo struktūros daugiadimens. duomenims (2)

- Padalintos hash funkcijos
 - Panašiai kaip vienadimensiniu atveju
- Grid failų ir padalintų hash funkcijų hash funkcijos
 - Grid failuose bucket'o numerį taškui gražina hash funkcija nuo visų atributų
 - Padalintose hash funkcijose hash funkcija yra sąrašas nuo hash funkcijų kiekvienam atributui ir gražina bitinę išraišką
- Grid failų ir padalintų hash funkcijų palyginimas
 - Padalintos hash funkcijos netinka sferinėms ir artimiausio kaimyno užklausoms
 - Grid failai linkę palikti daug laisvų bucket'ų

Medžio tipo struktūros daugiadimens. duomenims

- Daugelio raktų indeksai
 - Skirta taškų aibėms
- kd-Medžiai
 - Skirta taškų aibėms
- Ketvirtainiai medžiai
 - Skirta taškų aibėms
- R-Medžiai
 - Skirta regionų aibėms

Daugelio raktų indeksai

- Struktūra
 - Indeksas iš indeksų kiekvienam jo lygyje kiekvienam atributui atskirai
 - Atskirų atributų indeksai nebūtinai tokie patys
- Paieška

kd-Medžiai (k-dimensijų)

- Struktūra

- Pagrindinės atminties duomenų struktūra, apibendrinanti binarinius paieškos medžius daugiadimensiniams duomenims
- Binarinis medis
- Vidinio lygio mazgai turi atributą, tam atributui dalinančią reikšmę į mažesnius ir didesnius, ir nuorodas į kairįjį ir dešinį vaikus
- Kiekvienam lygyje vis pagal skirtingą atributą pasikartojant ratu
- Lapai – blokai su tiek vietos įrašams kiek blokas gali talpinti

- Paieška

- Įterpimas

Ketvirtainiai medžiai

- Struktūra

- Kiekvienas vidinio lygio mazgas atitinka k -dimensijų kubą k dimensijose
- Jei taškų skaičius kube ne didesnis nei tilptų bloke, tai šis kubas – lapas ir taškai laikomi bloke
- Jei taškų daugiau, tai kubas – vidinio lygio mazgas su vaikais, atitinkančiais jo lygius vidinius kubus

- Paieška

R-Medžiai (1)

- **Struktūra**

- Panašūs į B-medžius, tik yra ne taškai, o n-mačiai regionai, apimantys kažkiek taškų
- Vidinio lygio mazgas atitinka vieną regioną ir iš jo eina nuorodos į vaikus – vidinius regionus
- Vidiniai regionai nebūtinai dengia visą tėvo regioną
- Vidiniai regionai gali persidengti
- Duomenys saugomi lapuose

R-Medžiai (2)

- Paieška
 - Tipinė užklausa – “kur-aš-esu”
 - Ieškom regiono pradedant nuo šaknies
 - Galimi atsakymai: nėra, vienas, keli
- Įterpimas
 - Ieškomas tinkamiausias regionas
 - Jei nėra, kažkuris regionas praplečiamas
 - Jei lape duomeniui nėra vietos, rekursiškai aukštyje atliekamas regionų skaldymas

Bitų žemėlapių indeksai (1)

- Struktūra

- Įrašai faile turi pastovius numerius $1, 2, \dots, n$
- Yra duomenų struktūra, pagal i -ąjį numerį randanti atitinkamą įrašą faile
- Bitų žemėlapių indeksas kažkuriam laukui F yra rinkinys n ilgio bitų-vektorių, po vieną kiekvienai reikšmei, kurią gali įgyti laukas F
- Reikšmės v vektorius i -ojoje pozicijoje turi 1 , jei i -asis įrašas turi v lauke F , arba 0 priešingu atveju

Bitų žemėlapių indeksai (2)

- Bitų žemėlapių indeksų motyvavimas
 - Efektyviai atsakoma į dalinio atitikimo užklausas net nesikreipiant į pačius duomenis
 - Loginių operacijų naudojimas su bitų-vektoriais leidžia efektyviai atsakyti į sferines užklausas net nesikreipiant į pačius duomenis

Bitų žemėlapių indeksai (3)

- Suspausti bitų žemėlapiai
 - Vienam laukui bendras bitų skaičius indekse yra įrašų skaičiaus ir galimų reikšmių sandauga – gali pasirodyti daug
 - Naudojama kompresija, besiremianti atskirų bitų pozicijų skaičiavimu
 - Naudojamas kodavimas leidžia operuoti bitų-vektoriais net jų pilnai neiškodavus

Bitų žemėlapių indeksai (4)

- Bitų žemėlapių indeksų valdymas
 - Bitų-vektorių duotai reikšmei paieška – naudotini kiti indeksai, pvz. B-Medžiai
 - Įrašo radimas duomenų faile – naudotini antriniai indeksai pagal įrašo numerį duomenų faile
 - Duomenų failo keitimosi valdymas
 - Kartą priskirti įrašų numeriai turi išlikti nepakitę
 - Keičiant duomenų failą turi keistis ir indeksas