МАТФ
Универзитет у Београду
Математички факултет

11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Индекси

Пројектовање база података

Василије Тодоровић, Анђела Дамњановић, Александар Стефановић

7. мај 2025.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Садржај

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Садржај

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Неефикасност упита

- Претпоставимо да имамо табелу Test(id, contents). За рад са датом табелом развијена је апликација која извршава много упита типа:
  ```
  SELECT contents
  FROM Test
  WHERE id = constant;
  ```
- Као што је познато, систем у овом случају пролази целу табелу Test ред по ред и издваја садржај одговарајућих редова. Међутим, може се десити да табела има доста редова, а да свега пар њих (рецимо 0 или 1) задовољавају услов.
- Из датог примера се види да претраживање табеле без икакве припреме унапред може бити доста неефикасно. Овај проблем могу решити *индекси*.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Индекси

- Индекси су структуре података које убрзавају претрагу табела (по цену додатног простора).
- Они се праве над једном или више колона неке табеле.
- Синтакса за прављење индекса је следећа:
  CREATE INDEX име_индекса ON име_табеле (име_поља)

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Поставља се питање како одредити да ли и које поље индексирати?

- Приметимо да нема смисла индексирати колоне које ће ретко бити претраживане јер уштеда неће бити значајна.
- То нас даље доводи до закључка да треба индексирати оне колоне које ће се претраживати често.
- Учесталост претраживања колоне зависи од базе података и њен програмер је дужан да о томе унапред размисли.
- У случају из претходног примера, често се претражује поље id (јер се налази у WHERE клаузи), те стога има смисла њега индексирати: CREATE INDEX test_id_index ON Test (id);

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- Након креирања индекса, све остало се оставља систему: одлука да ли је боље претраживати по индексу или секвенцијално, као и ажурирање индекса приликом додавања и брисања података из базе.
- За брисање индекса, користи се команда DROP INDEX име_индекса. Индекси се могу правити и брисати у било ком тренутку.
- Индекси се могу користити и за убрзање UPDATE и DELETE команди, као и код JOIN клауза.
- У PostgreSQL бази података индексима се убрзавају претраге упита који имају једну или више WHERE или JOIN клауза.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Садржај

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Типови индекса

PostgreSQL нуди разне типове индекса:

- Б-дрво
- хеш индекс
- GiST
- SP-GiST
- GIN и
- BRIN.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- Сваки од ових индекса користи различите алгоритме за индексирање (сваки алгоритам најбоље индексира тачно одређену клаузу)
- Уколико тип индекса није експлицитно задат (случај из првог примера), подразумевано се користи Б-дрво, јер његово индексирање одговара већини уобичајених ситуација и он се креира на следећи начин:
  CREATE INDEX име ON табела(поље)
- Иначе, уколико је потребно направити неки други индекс, то се ради командом:
  CREATE INDEX име ON табела USING тип_индекса (поље);

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Б-дрво

- Најефикаснији када се у упитима користе поређења или интервали.
- Да би поређење било могуће, неопходно је да се подаци из те колоне могу некако сортирати.
- PostgreSQL ће размотрити коришћење овог индекса кад год се колона која је индексирана нађе у неком од поређења која укључују следеће операторе: <, <=, =, >=, >.
- Такође, овај индкес је погодан и у случајевима када се у упиту нађу BETWEEN и IN (као комбинација претходних оператора).
- Најпогоднији у случајевима када су табеле мале и средње величине и интервалне услове.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Хеш индекс

- Хеш индекси функционишу тако што вредност која се налази у индексираној колони хешира и чува се њен 32-битни запис
- Стога се ови индекси могу користити само за једноставна поређења на једнакост.
- Најбржи су и заузимају најмање додатног простора, па су често најбоља опција у условима једнакости, као и у случајевима када треба хеширати податке који су јединствени или скоро јединствени (вредности се ретко понављају).
- Креирање хеш индекса:
  CREATE INDEX име_индекса ON име_табеле USING HASH (име_поља);

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

BRIN – Block Range Index

- Најбољи за огромне табеле где постоји природно уређење и које су сортиране (или су њихови блокови сортирани).
- Уместо индексирања сваког реда, врши се индексирање низа блокова (нпр. чува се минимална/максимална вредност блока), па се у претраживању могу прескочити читави блокови са неодговарајућим вредностима.
- Оператори који су погодни за прављење овог типа индекса: <, <=, =, >=, >.
- CREATE INDEX име_индекса ON име_табеле USING BRIN (име_поља);

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Садржај

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

До сада је било речи само о прављењу индекса над једном колоном табеле.
Постоји могућност да се индекси дефинишу и за више редова у табели.
Размотримо следећу табелу:
Test2(minimum, maximum, name)
и следећи упит:

```
SELECT name
FROM Test2
WHERE maximum = constant AND minimum = constant;
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- Пошто се у WHERE клаузи претражују подаци о две колоне из табеле, можемо их обе индексирати ради лакше претраге. CREATE INDEX test2_idx ON Test2 (minimum, maximum);
- Од горенаведених типова индекса, само хеш индекси *не подржавају* индексирање више колона. Индекси могу имати до 32 колоне.
- Често је довољан индекс над једном колоном. Индекс над више од 3 колоне је ретко користан.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Б-дрво

- Б-дрво индекс се може користити за било који подскуп индексираних колона, али је најефикаснији када се у упиту налази ограничење на *најлевљој* колони.
- Прецизно правило гласи: за ограничавање простора претраге користе се ограничења једнакости на првим позицијама у упиту, као и прво следеће поређење по неједнакости.
- На пример, уколико имамо индексе на колонама a, b, c и услов WHERE a = 5 AND b >= 42 AND c < 77, за ограничавање простора претраге користи се услов једнакости a = 5 и први услов неједнакости b >= 42.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

BRIN

- BRIN индекс се може користити за било који подскуп индексираних колона. За разлику од претходног, његова ефикасност је увек иста.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Садржај

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- PostgreSQL нуди могућност да се комбинује више индекса (или више пута употребу истог индекса) да би се обрадили случајеви које је немогуће покрити једним јединим индексом. Ово је посебно корисно код AND и OR услова.
- На пример, имамо услов: WHERE x = 42 OR x = 47, и индексирану колону x, упит који садржи претходни услов извршава се на следећи начин:
- Најпре се пронађу сви резултати који задовољавају услов x = 42 користећи индекс, а затим и они који задовољавају услов x = 47 коришћењем тог истог индекса. На крају се за добијена 2 скупа резултата ради OR операција. То је финални резултат.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- Са друге стране, уколико је услов WHERE x = 5 AND y = 6, а индексиране колоне су x и y, користе се индекси појединачно, најпре за x, па за y и на крају се изврши конјункција добијених резултата.
- Да би се комбиновало више различитих индекса, систем пролази кроз сваки потребан индекс и припрема битмапу у меморији која садржи локације редова који одговарају условима. Над добијеним резултатима у битмапама даље се врше конјукције и дисјункције које су потребне за упит.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Комбиновање индекса vs индекси са више колона

- Претпоставимо да имамо табелу која садржи поља x и y. Да ли би било боље користити комбиноване појединачне индексе или један индекс са више колона?
- Уколико се често јављају упити који се врше само над колоном x, као и упити који се врше само над колоном y и понекад је потребно претражити табелу по обе колоне, боља је опција имати одвојене индексе над x и y колонама јер ће се тада сви упити моћи извршити.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- Такође, опција је и да се направи један индекс који садржи обе колоне, који би био ефикаснији у поређењу са одвојеним индексима, али би био бескористан за упите који садрже само колону у. Дакле, шта је боља опција зависи од упита који се јављају.
- Најзад, могуће је направити и сва три наведена индекса, али они су ефикасни само у ситуацијама кад је *претрага* далеко најчешћа операција и када су *сви* упити чести. Иначе је боље размотрити једну од две алтернативе.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Садржај

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

За следеће упите потребно је размислити употребом којих индекса би они могли бити оптимизовани и написати наредбу за креирање индекса.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Услови по једнакости

```
SELECT *
FROM users
WHERE email = 'example@example.com';
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Решење: Најпре треба утврдити које колоне су кандидати за индексирање. Пошто се претрага врши по колони email, она је једини кандидат, те њу индексирамо. Остаје још питање који индекс искористити. С обзиром да је у питању поређење на једнакост, сва три горенаведена индекса долазе у обзир, али су хеш индекси најбржи и заузимају најмање додатног простора у случајевима када индексирана колона има много различитих вредности, те ће они најбоље оптимизовати упит.

```
CREATE INDEX users_idx ON users USING HASH (email);
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

```
SELECT E.ename, D.mgr
FROM Employees E, Departments D
WHERE D.dname='Toy' AND E.dno=D.dno
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Поново, требало би индексирати колоне које учествују у WHERE клаузи, тј. dname и/или dno у овом случају. Услед постојања угнеждене петље, најбоље је направити индекс над оба поља. Пошто је у питању поређење по једнакости, хеш индекс је опет повољнија опција. Такође, пошто је у питању спајање табела, ефикаснији је хеш индекс, те стога бирамо њега.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Још једно битно питање се поставља: да ли индексирати поље dno у табели Employees или Departments?
Пошто се у упиту пролази кроз угнеждену петљу, боље је индексирати ону која је већа (унутрашња петља), тј. Employees (јер ће запослених природно бити више него сектора).

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Коначно решење:

```
CREATE INDEX Departments_idx ON Departments USING HASH (dname);
CREATE INDEX Employees_idx ON Employees USING HASH (dno);
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Интервални услови

Који индекс искористити у наредном упиту уколико је табела:
а) велика и сортирана
б) мања и несортирана?

```
SELECT E.ename
FROM Employees E
WHERE E.sal BETWEEN 10000 AND 20000
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Пошто се претрага врши по колони sal, она је једини кандидат за индексирање, те њу индексирамо. Остаје још питање који индекс искористити. С обзиром да је у питању интервални услов, хеш индекс се не може искористити, док се Б-дрво и BRIN индекс могу користити.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

а) Уколико је табела Employees велика, а њени подаци су сортирани, BRIN индекс је ефикаснији.

```
CREATE INDEX Emp_idx ON Employees USING BRIN(sal);
```

б) Са друге стране, уколико подаци у табели нису сортирани, претходни индекс је готово неупотребљив, те је боље користити Б-дрво индекс.

```
CREATE INDEX Emp_idx_btree ON Employees(sal);
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Који индекс искористити у наредном упиту уколико је табела:
а) велика и сортирана
б) мања и несортирана?

```
SELECT E.dno
FROM Employees E
WHERE E.age > 40
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

У услову претраге је интервал, па је овај пример аналоган прошлом.
а) CREATE INDEX Emp_idx ON Employees USING BRIN(age);
б) CREATE INDEX Emp_idx_btree ON Employees(age);

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Селективност услова

```
SELECT E.dno, COUNT(*)
FROM Employees E
WHERE E.age > 10
GROUP BY E.dno
```

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

У овом примеру, видимо да се у услову налази само једна колона, али се у ORDER BY клаузи налази још једна која може бити кандидат за индексирање. Да ли индексирати по некој од њих две или искористити обе?
Уколико погледамо услов за године у WHERE клаузи можемо закључити да је он бескористан, тј. сигурно не постоји запослени који је млађи од 10 година, па самим тим индекс над овом колоном би био бескористан. Са друге стране, поље dno може бити корисно. Дакле, потребно је њега индексирати. Пошто само Б-дрво може да врати сортиран резултат, управо овај индекс бирамо.
Решење: CREATE INDEX Emp_idx_btree ON Employees(dno);

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

```
SELECT E.eid
FROM Employees E
WHERE E.age BETWEEN 20 AND 40
AND E.sal BETWEEN 3000 AND 5000
```

Напомена: Табела није сортирана.

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Кандидати за индекс? age, sal, или оба. Уколико је један од услова довољно селективан, можемо одабрати само ту колону за индекс. Међутим, ако претпоставимо да већина запослених има баш између 20 и 40 година и да већина запослених има мању плату (реална ситуација), можемо закључити да ниједан од ових услова појединачно није довољно селективан. Стога је боље покушати индекс над обе колоне у нади да ће комбинација ових индекса имати нешто мање резултата. Сада се још поставља питање да ли направити индекс (age, sal) или (sal, age). Уколико су оба подједнако селективна, свеједно је. Иначе, као први бирамо онај селективнији. Индекс је Б-дрво.
Решење: CREATE INDEX Emp_idx_btree ON Employees(age, sal).

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

- Који би индекс био коришћен у претходном случају да је табела била сортирана?
- Који би био редослед атрибута да је услов био e.age = 25 AND E.sal BETWEEN 3000 AND 5000? Зашто?

МАТФ
Универзитет у Београду
Математички факултет
11001 БЕОГРАД
Студентски трг 16
П.П. 550
Телефон: 011 20 27 801
Факс: 011 26 30 151
matf@matf.bg.ac.rs
www.matf.bg.ac.rs

Литература

- О индексима са званичног postgreSQL сајта