

Индекси в MS SQL

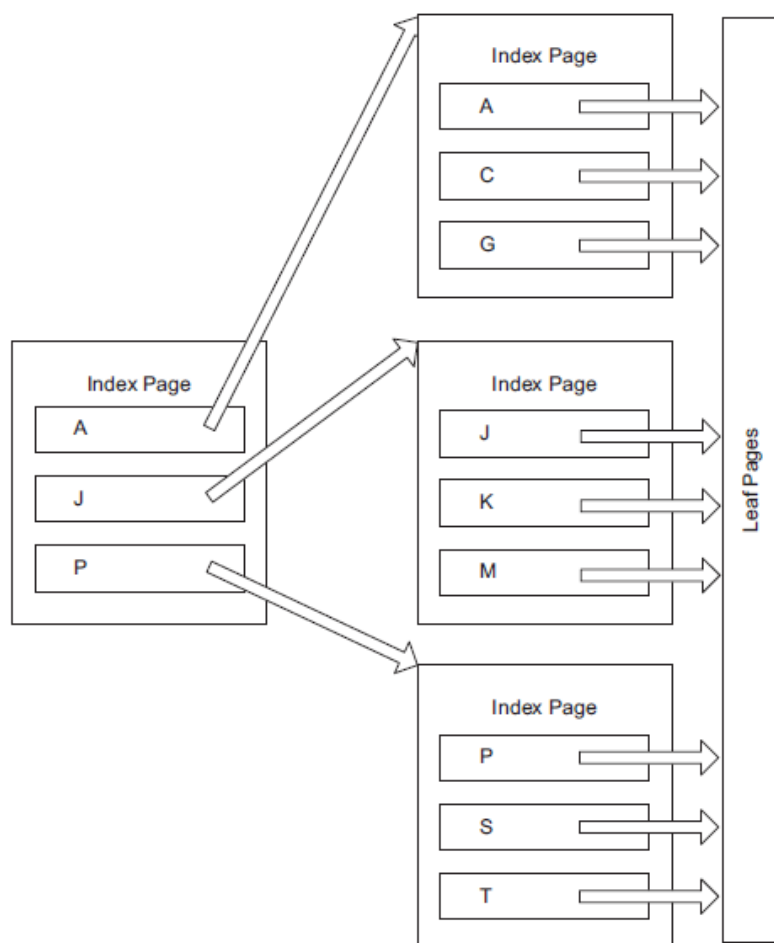
Индексът е обект в базата данни, който помага за организирането/поддръждането на данните и ускоряването на изпълнението на заявките.

Има 2 вида индекси:

- *Clustered*
- *Nonclustered*

Структура на индексите

Индексите се реализират посредством балансираните дървета (**B-trees**). Всеки възел на дървото съдържа множество от сортирани ключови стойности. Тези ключове дефинират интервали от стойности, в които се намират ключовете на наследниците на възлите. Вътрешната структура на индекса зависи от типа му. В случая когато индексът е **nonclustered** вътрешната структура на индекса се определя от наличието/отсъствието на **clustered index** за таблицата. За вътрешните възли на дървото (които не са листа), организацията е еднаква за двата типа индекси.



Clustered Indexes

Характеристики:

- Данните се подреждат по начин, дефиниран от индекса (например речник или телефонен указател). В листата на балансираното дърво се поместват именно страниците с данните, подредени по ключовите атрибути, върху които е формиран индексът. Ключови атрибути (**clustering key**) за индекса са атрибутите, които се използват в дефиницията на индекса и данните се подреждат според тях. **Не е задължително ключовите атрибути на индекса да са първичен ключ за релацията.**
- Една таблица може да има само един **clustered index**. Това е така, защото данните физически се подреждат на диска според ключовите колони от индекса.
- По подразбиране първичният ключ се създава като **clustered index**. Затова много често при извикване на заявка, която не изисква специфично подреждане на данните, резултатното множество се връща сортирано по първичен ключ (както е разположението на данните върху диска).

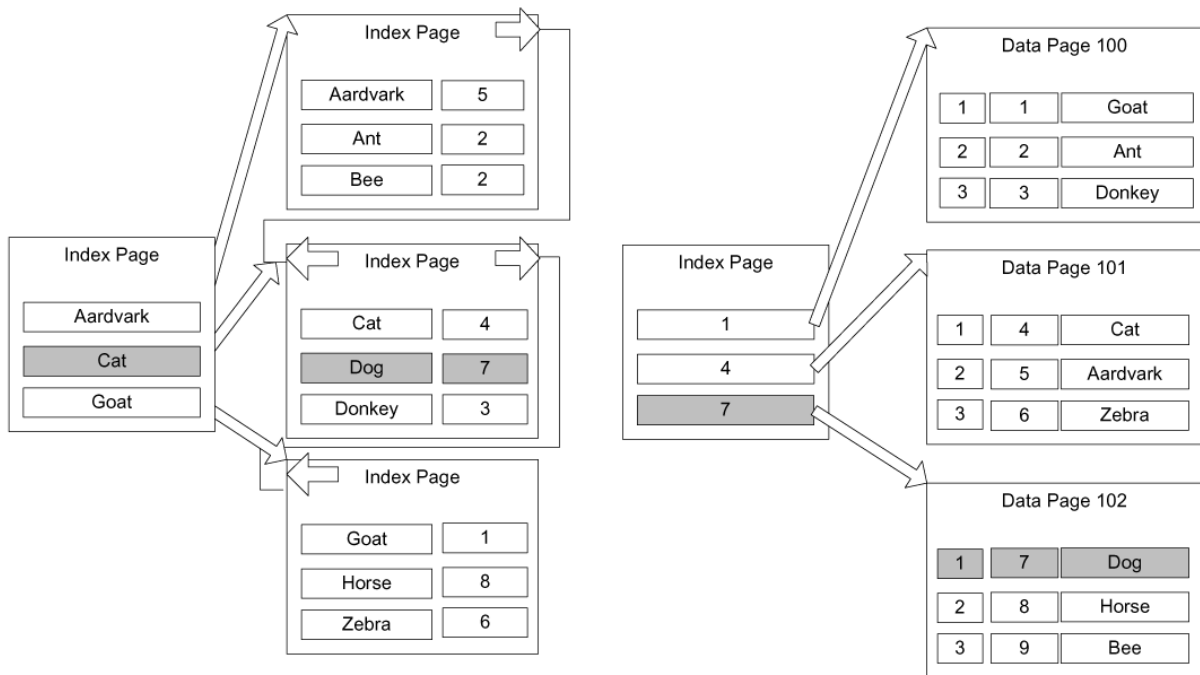
Nonclustered index

Nonclustered индексът е като индекс в книга. Таблицата не се подрежда по структурата на индекса, а се пази информация (указател, референция) за това къде точно на диска се намират стойностите на колоните. (В индекса на една книга се намират термините, които се срещат в книгата. Те са сортирани по азбучен ред и към всеки термин е посочен номер на страница, в която се използва терминът). За разлика от **clustered index nonclustered** индексът е отделен от физическото съхранение на данните. В листата на балансираното дърво, представлящо индекса, се намират указатели (**row locators**) към мястото на диска, където са разположени реалните данни.

Указателят (**row locator**) при този вид индекси се различава в зависимост от това дали таблицата, върху която се дефинира **nonclustered index** има вече дефиниран **clustered index**.

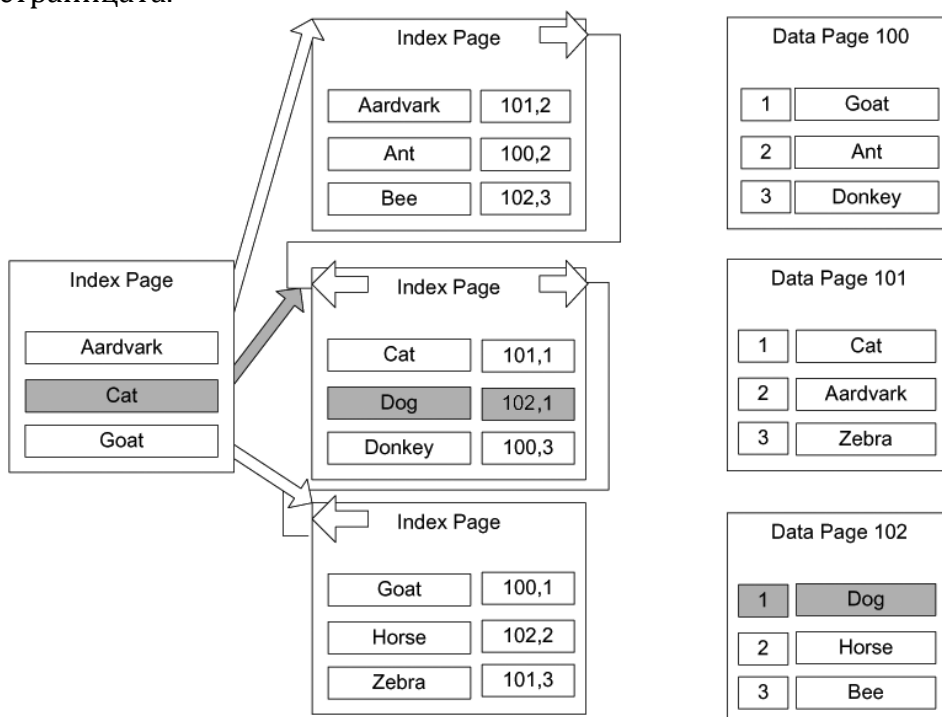
- **nonclustered index** върху таблица с дефиниран **clustered index**

Указателят в случая представлява ключовите колони за **clustered** индекса (**clustering key**).



- **nonclustered index** върху таблица БЕЗ **clustered index**

Указателят представлява номер на страница във файла с данните + номер на ред в страницата.



Кога да създаваме индекси?

- Добре е да се дефинират индекси върху колони като първични и външни ключове, които се използват често в заявки и по тях се извършват съединения.
- Индекс може да се използва за ускоряването на заявки, които търсят интервали от съседни стойности (например дати между начална и крайна дата)
- Когато често се налага да се сортират резултатните множества по дадена съвкупност от атрибути е от полза да се създаде индекс по атрибутите, участващи в сортирането
- Множество от атрибути, които участват заедно при агрегиране, са подходящ кандидат за дефиниране на индекс върху тях

Кога да НЕ създаваме индекс?

- Колоната не е достатъчно селективна, т.е. има малък брой уникални стойности
- Колоната рядко се използва в заявки

Големият брой уникални стойности и малък брой повторения за дадена колона е предпоставка за създаване на индекс, но само по себе си не е достатъчно. Индексите трябва да се дефинират и избират внимателно, защото поддръжката им при вмъкване, изтриване и промяна на данни е тежка за реализация (алгоритмите за балансиране, разпределяне и освобождаване на памет са нетривиални).

Команди за създаване и изтриване на индекс

```
CREATE [UNIQUE] [CLUSTERED | NONCLUSTERED] INDEX <indexName>  
ON <tableName> (<columnList>)
```

```
DROP INDEX <indexName> ON <tableName>
```

Информация за индекси от системните таблици може да се извлече посредством следната заявка

```
SELECT object_name(i.object_id) as object_name  
, case when i.is_unique = 1 then 'UNIQUE' else '' end + i.type_desc as index_type  
, i.name as index_name  
, user_seeks, user_scans, user_lookups, user_updates  
FROM sys.indexes i  
LEFT OUTER JOIN sys.dm_db_index_usage_stats s ON i.object_id = s.object_id  
AND i.index_id = s.index_id  
AND database_id = db_id()  
WHERE objectproperty(i.object_id, 'IsUserTable') = 1  
ORDER BY 1,3
```

където

user_seeks – брой пъти, в които индексът е използван за бърз достъп до данни чрез търсене в балансираното дърво

user_scans – брой пъти, в които данните в листата на индекса се преглеждат последователно без да се ползва ускорен достъп по колоните от индекса

user_lookups – за ***clustered*** индекси е брой пъти, в които индексът е ползван за търсене на данни при рефериране от ***nonclustered index***. Т.е. търси се по данни от колоните на ***nonclustered*** индекс и при достигане на листата указателят се търси в ***clustered*** индекса.

user_updates – показва колко пъти е променян индексът в потребителски заявки

Източник

Davidson L., Pro SQL Server 2008 Relational Database Design and Implementation, Apress