

DISTINCT in de SELECT-opdracht



dubbele uitkomsten

In hoofdstuk 7 van het boek 'Databases en SQL' ontbreekt een stukje over het weglaten van dubbele uitkomsten bij SELECT-opdrachten. Deze download bevat een uitleg over hoe we dubbele uitkomsten bij SELECT-opdrachten kunnen vermijden.

voorbeeld met dubbele uitkomsten

Stel, we willen de nummers en namen weten van werknemers die als vervanger van andere werknemers geregistreerd staan. We kunnen nagaan wie dat zijn met de volgende SELECT-opdracht:

```
SELECT WERKNEMER.W#, WNAAM
FROM VERVANGING, WERKNEMER
WHERE VERVANGING.VERVANGER=WERKNEMER.W#
```

Dit levert als uitkomst op:

```
-----
WERKNEMER.W#  WNAAM
-----
      5      DEKKER
     13      MOL
      2      BROUWER
      5      DEKKER
      9      JANSEN
     14      PIETERS
      7      EVERS
     11      LI
     12      LI
     13      MOL
      4      CHIN
     11      LI
     14      PIETERS
     12      LI
     11      LI
-----
```

Het valt op dat de werknemers 5, 11, 12, 13 en 14 meer dan eens in het resultaat voorkomen. Zij komen net zo vaak voor als dat zij vervanger zijn voor iemand. Het dubbel voorkomen in de uitkomst vertroebelt alleen maar de uitkomst en levert voor het oplossen van de vraag geen bijdrage.

dubbel komt door joinen

Het dubbel in de uitkomst opgenomen zijn van werknemers is een ongewenst neveneffect van het joinen. Iemand (bijvoorbeeld werknemer 5) komt voor elke werknemer die hij of zij vervangt (werknemer 1 en 4) in de join opnieuw in een rij voor. De rijen in de join als geheel verschillen van elkaar. Beperken we de rijen tot wat we in de uitkomst zien, dan komen sommige rijen vaker voor.

hoe onderdrukken we dubbele uitkomsten

We kunnen het vaker tonen van dezelfde uitkomst onderdrukken door bij SELECT de frase DISTINCT op te nemen. Hierdoor worden uitkomsten hooguit één keer opgenomen.

De SELECT-opdracht wordt dan:

```
SELECT DISTINCT WERKNEMER.W#, WNAAM
FROM VERVANGING, WERKNEMER
WHERE VERVANGING.VERVANGER=WERKNEMER.W#
```

De uitkomst wordt dan:

```
-----
WERKNEMER.W#  WNAAM
-----
      2      BROUWER
      4      CHIN
      5      DEKKER
      7      EVERS
      9      JANSEN
     11      LI
     12      LI
     13      MOL
     14      PIETERS
-----
```

dubbele rijen weg / uitkomst is gesorteerd

Twee dingen vallen op. Ten eerste valt op dat alle uitkomsten die eerder vaker voorkwamen, nog maar één keer zijn opgenomen. Ten tweede valt op dat het resultaat gesorteerd is. Het gesorteerd zijn heeft te maken met de wijze waarop dubbele uitkomsten opgespoord worden. Dit gebeurt door de uitkomsten te sorteren en vervolgens iedere uitkomst met de vorige te vergelijken. Is de uitkomst gelijk aan de vorige dan wordt deze weggelaten. Vervolgens wordt een volgende genomen en wordt gekeken of deze gelijk is aan de net opgenomen uitkomst. Zo ja, dan wordt ook die weggelaten en wordt verdergegaan. Zo nee, dan wordt de uitkomst opgenomen en verdergegaan. Dit wordt herhaald we totdat alle resultaten gevonden zijn.

download over DISTINCT in de COUNT-functie

De frase DISTINCT gebruiken we niet alleen in SELECT-opdrachten. DISTINCT gebruiken we mogelijk ook in de COUNT-functie. In zo'n geval gebruiken we de COUNT-functie voor een kolom. DISTINCT zorgt er dan voor dat waarden die in de kolom vaker voorkomen slechts één keer meegeteld worden voor het bepalen van het resultaat van de COUNT-functie.

Hoe dit precies gaat leg ik uit aan de hand van voorbeelden in het stuk 'DISTINCT in de COUNT-functie' dat je ook van deze website kunt downloaden.