

2. FEJEZET

Egytáblás csoportosító lekérdezések

Elméleti összefoglaló

Csoportképzés

Adattáblák feldolgozásánál gyakran van szükség arra, hogy az attribútumok (oszlopok) értékei (a mezőértékek) szerint a tábla sorait csoportosítsuk. Gondoljunk egy főiskola diákjainak az évfolyamuk, valamint a tankörszámuk alapján történő csoportosítására, amely például a tankörök tanulmányi átlagának meghatározásához nélkülözhetetlen.

A csoportosítást is tartalmazó SELECT utasítás általános felépítését már bemutattuk az 1. fejezetben. A csoportosítás szempontjából a következőképpen osztályozhatjuk egy csoportosításra kijelölt tábla oszlopait. *Csoportképzőnek* nevezzük azokat az oszlopkifejezéseket (illetve oszlopokat), amelyek szerint elvégezzük a csoportosítást. (Itt az oszlopkifejezések oszlopokra vonatkozó egysoros függvényekből, illetve műveletekből álló kifejezéseket jelölnek.) Ezen oszlopkifejezések vesszővel elválasztott listája áll a GROUP BY utasításrészben, és ez a lista vagy ennek egy része szerepelhet a szelekciós listában. Az e kifejezésekben nem szereplő (tehát *nem-csoportképző*) oszlopok viszont nem állhatnak, vagy csak csoportfüggvények paramétereiként állhatnak a szelekciós listában. Az alábbi utasításvázlat az elmondottakat szemlélteti:

```
SELECT ..., X+Y, ..., f(U), g(V)
FROM ...
WHERE ...
GROUP BY ..., X+Y
HAVING X+Y < 53;
```

ahol X és Y csoportképző attribútumok, U és V pedig nem-csoportképző attribútumok.

Mint a fenti példából is látható, csoportosítás esetén lehetőség van akár a csoportba kerülő sorok szűrésére (a WHERE utasításrészsel), akár a listába kerülő csoportok szűrésére (a HAVING utasításrészsel.)

Csoportfüggvények

Csoportképzés esetén alkalmazhatjuk az úgynevezett többsoros, azaz csoportfüggvényeket, melyek a sorok GROUP BY utasításrésszel kijelölt csoportjára, illetve ennek hiányában a tábla összes sorára vonatkoznak.

- AVG([DISTINCT] *oszlopkifejezés*)
az oszlopkifejezés átlagát adja meg a csoportra vonatkozóan,
- STDDEV([DISTINCT] *oszlopkifejezés*)
az oszlopkifejezés szórását (az átlagtól való eltérés értékét) adja meg a csoportra vonatkozóan,
- VARIANCE([DISTINCT] *oszlopkifejezés*)
az oszlopkifejezés szórásnégyzetét (a varianciát) adja meg a csoportra vonatkozóan,
- SUM([DISTINCT] *oszlopkifejezés*)
az oszlopkifejezés összegét adja meg a csoportra vonatkozóan,
- MIN([DISTINCT] *oszlopkifejezés*)
megadja az oszlopkifejezés legkisebb értékét a csoportban,
- MAX([DISTINCT] *oszlopkifejezés*)
megadja az oszlopkifejezés legnagyobb értékét a csoportban,
- COUNT([* | DISTINCT *oszlopkifejezés* | *oszlopkifejezés*])
* használata esetén megadja a csoport elemeinek (sorainak) számát,
DISTINCT *oszlopkifejezés* használata esetén megadja a csoport oszlopkifejezésre vonatkozóan különböző, de nem NULL értékű elemeinek (sorainak) számát,
oszlopkifejezés használata esetén megadja a csoport oszlopkifejezésre vonatkozó nem NULL értékű elemeinek (sorainak) számát.



Megjegyzés

- A DISTINCT használata esetén a csoportfüggvény értékének meghatározásában csak egyszer veszi figyelembe az azonos sorokat.
- A csoportfüggvények egymásba ágyazhatók.

2.1. példa

Listázza ki a dolgozók részlegenkénti átlagfizetését és ennek szórását egész értékre kerekítve.

Megoldás

```
SELECT deptno          AS részleg,
       ROUND(AVG(sal),0) AS átlag,
       ROUND(STDDEV(Sal),0) AS szórás
FROM emp
GROUP BY deptno;
```

Eredmény

RÉSZLEG	ÁTLAG	SZÓRÁS
10	2917	1894
20	2175	1123
30	1567	668

2.2. példa

Listázza a 2000 USD-nál nagyobb átlagjövedelmű részlegeket a jövedelem szerint rendezve.

Megoldás

```
SELECT deptno                                AS részlegazonosító,
       ROUND(AVG(sal+NVL(comm,0))) AS átlagjövedelem
FROM emp
GROUP BY deptno
HAVING ROUND(AVG(sal+NVL(comm,0))) > 2000
ORDER BY átlagjövedelem ASC;
```

Eredmény

RÉSZLEGAZONOSÍTÓ	ÁTLAGJÖVEDELEM
20	2175
10	2917

**Megjegyzés**

A feladatok megfogalmazása a gyakorlatban általában nem pontos, de a „józan ész” alapján azért ki lehet találni, hogy mire is van szüksége a felhasználónak. Tekintsük a fenti példát. Itt a „Listázza a ... részlegeket” kifejezés jelentése: „Listázza a ... részlegeknek a feladat szempontjából lényeges adatait”, vagyis az azonosítóját és az átlagjövedelmét, illetve a „rendezve” kifejezés jelentése egyaránt lehet „növekvően rendezve” vagy „csökkenően rendezve”, ám az első a gyakoribb. A feladatot természetesen már e megfontolások szellemében oldottuk meg.

2.3. példa (Csoportfüggvények egymásba ágyazása)

Listázza a részlegenkénti legnagyobb fizetéseket, a legkisebb, a legnagyobb és az átlagos részlegenkénti legnagyobb fizetéseket, a részlegenkénti létszámokat és a legkisebb, a legnagyobb és az átlagos részlegenkénti létszámokat.

Megoldás

(Az alábbiakban szereplő PROMPT SQL*Plus-utasítás segítségével lehet megadott szöveget a képernyőre kiírni. Részletesen lásd 4. fejezet.)

PROMPT Részlegenkénti legnagyobb fizetések:

```
SELECT deptno, MAX(sal)
FROM emp
GROUP BY deptno;
```

PROMPT A legkisebb, a legnagyobb és az átlagos

PROMPT részlegenkénti legnagyobb fizetések:

```
SELECT MIN(MAX(sal)), MAX(MAX(sal)), ROUND(AVG(MAX(sal)))
FROM emp
GROUP BY deptno;
```

PROMPT Részlegenkénti létszámok:

```
SELECT deptno, COUNT(*)
FROM emp
GROUP BY deptno;
```

PROMPT A legkisebb, a legnagyobb és az átlagos-
részlegenkénti létszámok:

```
SELECT MIN(COUNT(*)), MAX(COUNT(*)), ROUND(AVG(COUNT(*)),2)
FROM emp
GROUP BY deptno;
```

Eredmény

Részlegenkénti legnagyobb fizetések:

DEPTNO	MAX(SAL)
10	5000
20	3000
30	2850

A legkisebb, a legnagyobb és az átlagos
részlegenkénti legnagyobb fizetések:

MIN(MAX(SAL))	MAX(MAX(SAL))	ROUND(AVG(MAX(SAL)))
2850	5000	3617

Részlegenkénti létszámok:

DEPTNO	COUNT(*)
10	3
20	5
30	6