

## LABORATOR 7: GRUPAREA DATELOR

### OPERATORII ROLLUP ȘI CUBE

În clauza *GROUP BY* a unei cereri se pot utiliza operatorii *ROLLUP* și *CUBE*. Acești operatori sunt disponibili începând cu versiunea *Oracle8i*.

#### Operatorul *ROLLUP*

Acest operator furnizează valori agregat și superagregat corespunzătoare expresiilor din clauza *GROUP BY*. Operatorul *ROLLUP* poate fi folosit pentru extragerea de statistici și informații totalizatoare din mulțimile rezultate. Acest operator poate fi util la generarea de rapoarte, diagrame și grafice.

Operatorul *ROLLUP* creează grupări prin deplasarea într-o singură direcție, de la dreapta la stânga, de-a lungul listei de coloane specificate în clauza *GROUP BY*. Apoi, se aplică funcția agregat acestor grupări. Dacă sunt specificate  $n$  expresii în operatorul *ROLLUP*, numărul de grupări generate va fi  $n + 1$ . Liniile care se bazează pe valoarea primelor  $n$  expresii se numesc linii obișnuite, iar celelalte se numesc linii superagregat.

Dacă în clauza *GROUP BY* sunt specificate  $n$  coloane, pentru a produce subtotaluri fără operatorul *ROLLUP* ar fi necesare  $n + 1$  instrucțiuni *SELECT* conectate prin *UNION ALL*. Aceasta ar face execuția cererii inefficientă pentru că fiecare instrucțiune *SELECT* determină accesarea tabelului. Operatorul *ROLLUP* determină rezultatele efectuând un singur acces la tabel și este util atunci când sunt implicate multe coloane în producerea subtotalurilor.

Așadar ***GROUP BY ROLLUP*** (*expr\_1*, *expr\_2*, ..., *expr\_n*) generează ***n+1 tipuri de linii***, corespunzătoare următoarelor grupări:

- *GROUP BY* (*expr\_1*, *expr\_2*, ..., *expr\_n-1*, *expr\_n*)
- *GROUP BY* (*expr\_1*, *expr\_2*, ..., *expr\_n-1*)
- ...
- *GROUP BY* (*expr\_1*, *expr\_2*)
- *GROUP BY* (*expr\_1*)
- *GROUP BY* () – corespunzător absenței clauzei *GROUP BY* și deci, calculului funcțiilor grup din cerere pentru întreg tabelul.

**Exemplul 1:** Pentru fiecare departament având codul  $< 50$ , să se afișeze media salariilor angajaților cu același *JOB\_ID*, media salariilor pe departament, media salariilor pentru toți angajații.

```
SELECT department_id, job_id, AVG(salary)
FROM employees
WHERE department_id <= 50
GROUP BY ROLLUP (department_id, job_id);
```

DEPARTMENT_ID	JOB_ID	AVG(SALARY)
10	AD_ASST	4400
10		4400
20	MK_MAN	13000
20	MK_REP	6000
20		9500
30	PU_CLERK	2780
30	PU_MAN	11000
30		4150
40	HR_REP	6500
40		6500
50	SH_CLERK	3215
50	ST_CLERK	2785
50	ST_MAN	7280
50		3475.55556
		3840

În rezultatul prezentat anterior se pot distinge 3 tipuri de linii.

- Prima linie afișează media salariilor din departamentul 10 pentru angajații care sunt "AD\_ASST".
- A doua linie afișează media salariilor din departamentul 10. Coloana *JOB\_ID* este *NULL*.
- Ultima linie afișează media tuturor salariaților. Coloanele *DEPARTMENT\_ID*, *JOB\_ID* sunt completate cu valoarea *NULL*.

## Operatorul *CUBE*

Operatorul *CUBE* grupează liniile selectate pe baza valorilor tuturor combinațiilor posibile ale expresiilor specificate și returnează câte o linie totalizatoare pentru fiecare grup. Acest operator este folosit pentru a produce mulțimi de rezultate care sunt utilizate în rapoarte. În vreme ce *ROLLUP* produce subtotalurile doar pentru o parte dintre combinațiile posibile, operatorul *CUBE* produce subtotaluri pentru toate combinațiile posibile de grupări specificate în clauza *GROUP BY*, precum și un total general.

Dacă există  $n$  coloane sau expresii în clauza *GROUP BY*, vor exista  $2^n$  combinații posibile superagregat. Din punct de vedere matematic, aceste combinații formează un cub  $n$ -dimensional, de aici provenind numele operatorului. Pentru producerea de subtotaluri fără ajutorul operatorului *CUBE* ar fi necesare  $2^n$  instrucțiuni *SELECT* conectate prin *UNION ALL*.

**Exemplul 2:** Să se afișeze mediia salariilor angajaților pe departamente, pe job-uri, în cadrul departamentului pe job-uri, media salariilor tuturor angajaților.

```
SELECT department_id, job_id, AVG(salary)
FROM employees
WHERE department_id <= 50
GROUP BY CUBE (department_id, job_id)
```

Apar în plus linii în care *department\_id* este *NULL* care afișează media salariilor pentru angajații cu un anumit *job\_id*

Pentru determinarea modului în care a fost obținută o valoare totalizatoare cu *ROLLUP* sau *CUBE*, se utilizează funcția: **GROUPING**(expresie)

Aceasta întoarce:

- valoarea 0, dacă expresia a fost utilizată pentru calculul valorii agregat
- valoarea 1, dacă expresia nu a fost utilizată.

### Exemplul 3:

```
SELECT department_id, job_id, AVG(salary) medie,
GROUPING(department_id) dn, GROUPING(job_id) jid
FROM employees
GROUP BY ROLLUP(department_id, job_id);
```

DEPARTMENT_ID	JOB_ID	MEDIE	DN	JID
90	AD_VP	17000	0	0
90		19333.3333	0	1
100	FI_ACCOUNT	7920	0	0
100	FI_MGR	12000	0	0
100		8600	0	1
110	AC_ACCOUNT	8300	0	0
110	AC_MGR	12000	0	0
110		10150	0	1
	SA_REP	7000	0	0
		7000	0	1
		6461.68224	1	1

Ultimele 3 linii au coloanele DN, JID completate cu 0 0, 0 1 respectiv 1 1 deoarece se ia în calcul grupul angajaților care au completat department\_id cu *NULL*

Dacă se dorește obținerea numai a anumitor grupări superagregat, acestea pot fi precizate prin intermediul clauzei :

**GROUPING SETS** ((expr\_11, expr\_12, ..., expr\_1n), (expr\_21, expr\_22, ...expr\_2m), ...)

### Exemplul 4: Să se afișeze suma salariilor pentru

- fiecare departament și, în cadrul său, pentru fiecare job\_id
- pentru tot tabelul.

```
SELECT department_id, job_id, AVG(salary) medie,
GROUPING(department_id) dn, GROUPING(job_id) jid
FROM employees
GROUP BY GROUPING SETS((department_id, job_id),())
```

**Exercițiul 5:** a) Să se afișeze numele departamentelor, titlurile job-urilor si valoarea medie a salariilor, pentru:

- fiecare departament și, în cadrul său pentru fiecare job;
- fiecare departament (indiferent de job);
- întreg tabelul.

b) Analog cu a), afișând și o coloană care arată intervenția coloanelor *department\_name*, *job\_title*, în obținerea rezultatului. În această coloană se vor afișa textele: “media pe departament”, “media pe departament si job”, “”media pentru toti angajatii””.

**Exercițiul 6:** Să se afișeze numele departamentelor, titlurile job-urilor și valoarea medie a salariilor, pentru:

- fiecare departament și, în cadrul său pentru fiecare job;
- fiecare departament (indiferent de job);
- fiecare job(indiferent de departament)
- întreg tabelul.

**Exercițiul 7:** Să se afișeze numele departamentelor, numele job-urilor, codurile managerilor, maximul și suma salariilor pentru:

- fiecare departament și, în cadrul său, fiecare job;
- fiecare job și, în cadrul său, pentru fiecare manager;
- întreg tabelul.

**Exercițiul 8:** Afișați primele trei departamente în funcție de numărul de angajați pe care îi are în subordine directă șeful de departament.

Pentru fiecare departament din lista se cer: numele, numărul de angajați subordonați managerului de departament, media salariilor din acel departament.

**Exercițiul 9:** Afișați pentru fiecare angajat câte zile a lucrat în fiecare oras.

**Exercițiul 10:** Afișați media salariilor pentru toți salariații care

- au același număr de superiori ierarhici;
- pentru toți salariații care au același număr de superiori ierarhici și care lucrează în același departament.

## INSTRUCȚIUNEA CASE

**Exercițiul 11:** Afișați denumirile joburilor și o coloană denumită statistică ce va conține:

- media generală a salariilor, pentru job-ul având salariul maxim;
- media generală a salariilor, pentru job-ul având salariul minim;
- salariul minim, pentru fiecare din celelalte job-uri;
- salariul minim posibil pentru joburile care nu sunt încă ocupate de un angajat.

Includeți în coloană un string care să precizeze semnificația (de exemplu, 'media\_min: 5000')

**Exercițiul 12:** Afișați pentru fiecare manager numele sau, numărul de angajați din subordine (directă sau indirectă) și diferența dintre cel mai mare comision al unui subordonat și comisionul sau.