

Duomenų Bazės

AGREGACINĖS F-JOS IR DUOMENŲ GRUPAVIMAS

Agregatinės f-jos

Agregatinės funkcijos - tai funkcijos, kurias galima pritaikyti eilučių aibėms: visoms eilutėms, tenkinančioms frazėje WHERE nurodytą sąlygą, arba eilučių grupei, apibrėžiamai fraze GROUP.

Agregatinės funkcijos bet kokiam eilučių rinkiniui apskaičiuoja vieną reikšmę. Šios rūšies funkcijos dažnai vartojamos SELECT frazėje, nurodant stulpelį, pagal kurio reikšmes apskaičiuojamas funkcijos rezultatas.

Agregatinės funkcijos vartojamos ir frazėje HAVING, kai funkcijos reikšmė skaičiuojama kiekvienai eilučių grupei.

Agregatinės f-jos

Keletas dažnai vartojamu agregatiniu funkciiu:

Agregatinė funkcija	Rezultatas
SUM([DISTINCT] <reiškinys>)	(Skirtingų) ne NULL reikšmių suma
AVG([DISTINCT] <reiškinys>)	(Skirtingų) ne NULL reikšmių vidurkis
COUNT([DISTINCT] <reiškinys>)	(Skirtingų) ne NULL reikšmių kiekis
COUNT(*)	Eilučių kiekis aibėje
MAX(<reiškinys>)	Maksimali reikšmė
MIN(<reiškinys>)	Minimali reikšmė

Jei visos funkcijoje nurodyto stulpelio reikšmės yra NULL, arba stulpelis yra tuščias, tai funkcijų SUM, AVG, MIN, MAX rezultatas yra NULL, o funkcijos COUNT – nulis.

Agregatinės f-jos

Visų vykdytojų skaičius arba, tiksliau, eilučių kiekis lentelėje Vykdytojai gali būti sužinomas sakiniu:

```
SELECT COUNT(*) FROM Vykdytojai.
```

Visų vykdytojų, dalyvaujančių bent viename projekte, skaičius - sakiniu:

- `SELECT COUNT(DISTINCT Vykdytojas) FROM Vykdymas.`

Agregatinės f-jos

Bendrą kiekį valandų, kurias vykdytojas Nr. 1 skiria visiems projektams, galima sužinoti įvykdžius užklausą:

```
SELECT SUM(Valandos) AS "Vykdytojo Nr. 1 valandos"  
FROM Vykdimas  
WHERE Vykdytojas = 1
```

Užduotis

Sukurkime papildomą lentelę įmonės statistikos puslapyje ir išveskime informaciją apie ją (lentelė pateikiama faile statistika2.php):

Įmonės statistika:	
Įmonėje dirbančių žmonių skaičius	0
Vidutinis darbo užmokestis	1500 EUR
Minimalus darbo užmokestis	1500 EUR
Maksimalus darbo užmokestis	1500 EUR

Grupavimas užklausoje

Tarkime, kiekvienam projektui reikia apskaičiuoti, kiek visi vykdytojai bendrai skiria laiko jo vykdymui. Šiam uždaviniui spręsti iki šiol išdėstytų kalbinių priemonių nepakanka.

Nesunku apskaičiuoti bendrą valandų kiekį kuriam nors vienam projektui, pvz. Nr. 1:

```
SELECT SUM(Valandos) FROM Vykdymas WHERE Projektas = 1.
```

Šiame SQL sakinyje paieškos sąlygoje yra fiksuotas vienas projektas. Visoms tenkinančioms sąlygų eilutėms yra pritaikyta funkcija valandoms sumuoti. Pradiniam mūsų uždaviniui spręsti, šią stulpelių funkciją reikia pritaikyti kiekvienai eilučių grupei, atitinkančiai visus skirtingus projektus.

Grupavimas užklausoje

Eilučių grupavimą užklausoje realizuoja speciali konstrukcija GROUP BY, kurią pavartojus uždavinio sprendinys atrodo taip:

```
SELECT Projektas, SUM(Valandos) AS Valandos FROM Vykdymas  
GROUP BY Projektas.
```

Visos lentelės *Vykdymas* eilutės grupuojamos taip, kad į kiekvieną grupę patenka eilutės su vienodomis stulpelio *Projektas* reikšmėmis. Po grupavimo, SELECT frazė yra taikoma kiekvienai grupei (o ne eilutei), suformuojant vieną užklaustos rezultato eilutę.

Grupavimas užklausoje

Kadangi, pritaikant SELECT frazę grupei eilučių, suformuojama tik viena rezultato eilutė, tai visi šios frazės reiškiniai turi būti vienareikšmiškai apskaičiuojami, t.y. turi įgyti vieną reikšmę visoje grupėje.

Užklauso su grupavimu SELECT frazėje, su nedidelėmis išimtimis, tegali būti:

- stulpelis, paminėtas frazėje GROUP BY (grupavimo stulpelis);
- konstanta;
- reiškinys pagal konstantas ir grupavimo stulpelius;
- agregatinė (stulpelių) funkcija (SUM, MIN, MAX, COUNT, AVG ir kt.), kuri eilučių grupei pateikia vieną reikšmę.

Grupavimas užklausoje

Prieš GROUP BY frazę galima vartoti paieškos sąlygą, kuri tikrinama prieš eilučių grupavimą. Jei norime paieškos sąlygą pritaikyti ne kiekvienai lentelės eilutei, bet kiekvienai grupei, tai frazė GROUP BY vartojama kartu su fraze HAVING.

Sąlyga grupėms yra tikrinama po grupavimo. Grupės “atstovas” į užklauso rezultata įtraukiamas tik tuomet, kai grupė tenkina paieškos sąlygą, nurodytą frazėje HAVING.

Grupavimas užklausoje

Tokiu būdu, jei užklausoje su grupavimu yra pavartota frazė WHERE, tai, iš pradžių, kiekvienai eilutei tikrinama ši sąlyga ir visos eilutės, tenkinančios šią sąlygą, yra sugrupuojamos. Jei užklausoje yra pavartota ir frazė HAVING, tai pastarojoje frazėje nurodyta sąlyga yra tikrinama kiekvienai grupei.

Pavyzdžiui, numeriai projektų, kuriuos vykdo daugiau negu vienas vykdytojas, gali būti sužinomi užklausa:

```
SELECT Projektas FROM Vykdymas  
GROUP BY Vykdytojas HAVING COUNT(*) > 1.
```

Grupavimas užklausoje

Galima grupuoti ir pagal kelis stulpelius. Tuomet sakoma, kad grupuojama grupės viduje.

Tarkime, kad mums reikia sužinoti, kiek yra vykdytojų pagal kiekvieną baigtą aukštąją mokyklą ir turimą ategoriją. Užklausiai sudaryti, reikia apibrėžti grupavimą pagal du lentelės *Vykdytojai* stulpelius: *Išsilavinimas* ir *Kategorija*:

```
SELECT Išsilavinimas, Kategorija, COUNT(*) AS “Vykdytojų Skaičius”  
FROM Vykdytojai  
GROUP BY Išsilavinimas, Kategorija  
ORDER BY Išsilavinimas.
```

Grupavimas užklausoje

Tarkime, reikia sužinoti, kiek kiekvienas vykdytojas skiria valandų visiems projektams vykdyti. Be to, mus domina tik tie vykdytojai, kurie visų projektų vykdymui skiria daugiau nei 300 valandų. Jeigu mums pakanktų vykdytojo numerio, tai uždavinį galėtume spręsti panašiai kaip uždavinį, kuriame valandos buvo skaičiuojamos kiekvienam projektui. Jei mus domina vykdytojų pavardės, užklausą reikia formuluoti dviem lentelėms:

```
SELECT Pavardė, SUM(Valandos) AS "Visos Valandos"  
FROM Vykdytojai, Vykdymas WHERE Nr = Vykdytojas  
GROUP BY Pavardė HAVING SUM(Valandos) > 300  
ORDER BY "Visos Valandos".
```

Grupavimas užklausoje

Sąlygoje grupei (frazėje HAVING) negalima vartoti SELECT frazėje apibrėžto stulpelio pavadinimo “*Visos Valandos*”. Tokio stulpelio pavadinimas grupėse negalioja.

Jo tiesiog nėra grupėse. Stulpelio pavadinimas prasmingas tik užklauso rezultato kontekste, todėl jį galima vartoti apibrėžiant eilučių tvarką