



Tietokantojen perusteet

SQL – Koostekyselyjä

Aggregointifunktiot

- Aggregointi-, kooste- eli yhteenvetofunktioilla tarkoitetaan operaatioita, jotka tekevät koosteita (yhteenvetoja) tietokantaan tallennetuista tiedoista.
 - COUNT - lukumäärä
 - SUM - summa
 - MAX - suurin arvo
 - MIN - pienin arvo
 - AVG - aritmeettinen keskiarvo
- COUNT riveille, kaikenlaisille sarakkeille
- SUM, AVG numeerisille tietotyypeille
- MIN, MAX tietotyypeille, joille pystytään määräämään järjestys (myös merkkijonoille, päivämäärille)
- Aggregointifunktio(i)ta soveltava koostekysely (yhteenvetokysely) voi tuottaa **yhden tulosrivin** koko taulun datan perusteella tai **useita ryhmäkohtaisia tulosrivejä**, jotka on laskettu kunkin ryhmän datan perusteella.

Aggregointifunktiot

```
SELECT AVG(palkka), SUM(palkka), COUNT(palkka)
FROM tyontekija;
```

avg	sum	count
3860.1000000000000000	19300.50	5

On muodostettu yksi tulosrivi taulun kaikkien rivien perusteella.

```
SELECT AVG(palkka) AS palkka_ka,
       SUM(palkka) AS palkkasumma,
       COUNT(palkka) AS lkm
FROM tyontekija;
```

palkka_ka	palkkasumma	lkm
3860.1000000000000000	19300.50	5

Eri järjestelmät nimeävät tulossarakkeet eri tavoin.

Tulossarakkeet voi tarvittaessa nimetä uudelleen AS-operaattorin avulla.

```
SELECT ROUND(AVG(palkka),2) AS keskipalkka
FROM tyontekija;
```

keskipalkka
3860.10

Pyöristetään keskipalkka kahden desimaalin tarkkuuteen ROUND-funktiolla.

Aggregointifunktiot

COUNT(*)

- Lasketaan rivien lukumäärä.

COUNT(sarakenimi)

- SQLite ja PostgreSQL: Lasketaan niiden ei-tyhjien arvojen (ts. muiden kuin NULL-arvojen) määrä ko. sarakkeessa.

ttnro	etunimi	sukunimi	saika	kotikunta	palkka	puhelin	osastonro	esimiesnro
88	Jukka	Susi	1957-11-10	Tampere	5500.00	444 1234	1	
33	Ville	Viima	1975-12-08	Nokia	4000.50	444 4343	5	88
12	Pekka	Puro	1985-01-09	Tampere	3000.00		5	33
98	Jenni	Joki	1961-06-20	Lempäälä	4300.00	444 4488	4	88
99	Alli	Kivi	1988-07-19	Nokia	2500.00	444 5555	4	98

```
SELECT COUNT(*)  
FROM tyontekija;
```

```
count  
-----  
5
```

```
SELECT COUNT(palkka)  
FROM tyontekija;
```

```
count  
-----  
5
```

```
SELECT COUNT(esimiesnro)  
FROM tyontekija;
```

```
count  
-----  
4
```

Aggregointifunktiot

- Summaa, keskiarvoa, minimiä ja maksimia laskettaessa ei huomioida tyhjäarvoja.
- Jos sarakkeessa on vain tyhjäarvoja
 - COUNT(sarakenimi) antaa tulokseksi nollan (0)
 - muut funktiot antavat tulokseksi tyhjäarvon (null)
- Huom! Jatkossa tarkista aggregointifunktioiden toiminta käyttämäsi tietokannanhallintajärjestelmän manuaalista (esim. sarakkeiden nimeäminen ja COUNT(sarakenimi)-funktion käyttäytyminen NULL-arvojen suhteen)

Aggregointifunktiot

- COUNT-, SUM- ja AVG-operaatioihin voidaan lisätä mukaan **DISTINCT**-määre, jolloin laskennassa käsitellään vain toisistaan poikkeavia arvoja (duplikaattiarvot eliminoidaan ennen aggregointifunktion soveltamista).

```
SELECT COUNT(DISTINCT kotikunta) AS eri_kotikuntia  
FROM tyontekija;
```

```
eri_kotikuntia  
-----  
3
```

Aggregointifunktiot - Ryhmittely

- Aggregointifunktioiden yhteydessä voidaan käyttää GROUP BY ja HAVING -määreitä.
- **GROUP BY** -määreellä kerrotaan, minkä sarakkeiden suhteen tehdään ryhmittely yhteenveto-operaatioita varten.
 - Kaikki ne rivit, joilla on samat arvot annetuissa sarakkeissa, muodostavat oman ryhmänsä.
 - **Huom!** Tässä kohtaa **tyhjäarvot käsitellään samana arvona!!!!**
 - Nyt muodostetaan yksi tulosrivi yhtä ryhmää kohden.
 - Tulossarakkeissa voi olla aggregointifunktioiden lisäksi vain niitä sarakkeita, jotka esiintyvät GROUP BY -määreessä.

Aggregointifunktiot

Aggregointifunktioiden tuottamien sarakkeiden lisäksi **tulossarakkeina voi olla vain GROUP BY -määreessä annettuja sarakkeita.**

```
SELECT kotikunta, COUNT(kotikunta) AS 1km  
FROM tyontekija  
GROUP BY kotikunta;
```

ttnro	etunimi	sukunimi	saika	kotikunta	palkka	puhelin	osastonro	esimiesnro
98	Jenni	Joki	1961-06-20	Lempäälä	4300.00	444 4488	4	88

33	Ville	Viima	1975-12-08	Nokia	4000.50	444 4343	5	88
99	Alli	Kivi	1988-07-19	Nokia	2500.00	444 5555	4	98

88	Jukka	Susi	1957-11-10	Tampere	5500.00	444 1234	1	
12	Pekka	Puro	1985-01-09	Tampere	3000.00		5	33

```
kotikunta | 1km  
-----+-----  
Lempäälä | 1  
Tampere  | 2  
Nokia    | 2
```

Ryhmittely tehdään kotikunta-sarakkeen arvojen suhteen.

Tulostaulussa yksi rivi kullekin ryhmälle.

Aggregointifunktiot – ehtoja ryhmille

- HAVING-lauseella voidaan antaa ehtoja, jotka GROUP BY -operaatiolla tehdyn ryhmän on toteutettava, jotta ryhmä otetaan tulostauluun mukaan.
 - Ehdossa testataan ryhmälle lasketun aggregointifunktion arvoa.
- HAVING COUNT(*) > 1 valitsee vain sellaiset GROUP BY -operaatiolla määritellyt ryhmät, joissa on enemmän kuin yksi rivi.

```
SELECT kotikunta, COUNT(kotikunta) AS lkm
FROM tyontekija
GROUP BY kotikunta
HAVING COUNT(*) > 1;
```

kotikunta	lkm
Tampere	2
Nokia	2

Tulokseen mukaan
ryhmät, joissa vähintään
2 riviä

Kyselyn evaluointi

- SQL-kyselyn yleisrakenne
SELECT tulostietomäärittely
FROM taulut
[WHERE liitos- ja valintaehdot]
[GROUP BY ryhmittelytekijät]
[HAVING ryhmärajoitteet]
[ORDER BY järjestystekijät]
- Yksinkertainen kyselynevaluointialgoritmi: kyselyn osien suoritusjärjestys
 1. FROM-osa
 2. WHERE-osa
 3. GROUP BY -osa
 4. HAVING-osa
 5. SELECT-osa
 6. ORDER BY -osa

Huom! Tietokannanhallintajärjestelmä ei evaluoi kyselyä tällä tavoin – algoritmin tarkoituksena on auttaa ymmärtämään, miten kyselyn tulos muodostuu.

Kyselyn evaluointi

1. FROM-osa

- Jos FROM-osassa ei ole liitosoperaatioita,
 - muodostetaan annettujen **taulujen karteellinen tulo eli ristitulo**, joka on tämän vaiheen tulostaulu.
- Jos FROM-osassa on liitosoperaatio(ita),
 - muodostetaan liitoksen (karteellinen tulo ja liitosehto) antama tulostaulu ja
 - lisätään tulostauluun mahdollisen ulkoliitoksen antamat ”lisärivit”.
 - (Käydään läpi kaikki liitosoperaatiot vastaavalla tavalla.)

2. WHERE-osa

- Poistetaan edellisen vaiheen tulostaulusta ne rivit, jotka eivät täytä liitos- ja valintaehdoja.
- Huom! Jos FROM-osassa on annettu useita tauluja (ilman liitosoperaatioita), tarvitaan yleensä liitosehto(ja) taulujen tietojen yhdistämiseksi. Jos tauluja on N kappaletta, tarvitaan yleensä N-1 liitosehtoa. (Huom. Liitosehto voi koostua alkeisehdoista.)

Kyselyn evaluointi

3. GROUP BY -osa

- Jaetaan rivit ryhmiin ryhmittelytekijöiden (sarakkeiden) mukaisesti.

4. HAVING-osa

- Poistetaan ne ryhmät, jotka eivät täytä ryhmille asetettuja rajoitteita (ehtoja).

5. SELECT-osa

- Lasketaan aggregointifunktioiden arvot kullekin ryhmälle.
- Poistetaan kaikki sarakkeet, jotka eivät esiinny SELECT-listassa.
- Tuotetaan yksi tulosrivi ryhmää kohden.

6. ORDER BY -osa

- Järjestetään rivit järjestystekijöiden (sarakkeiden) mukaisesti.

Aggregointifunktiot - kyselyn evaluointi

Haetaan osaston numero, osaston nimi ja osaston valvomien projektien lukumäärä. Tulokseen otetaan mukaan ainoastaan sellaiset osastot, jotka valvovat yhtä useampaa projektia. Tulokset järjestetään osastojen nimien mukaan nousevaan järjestykseen.

```
SELECT o.onro, o.onimi, COUNT(*) AS projektien_lkm
FROM osasto o, projekti p
WHERE o.onro = p.onro
GROUP BY o.onro, o.onimi
HAVING COUNT(*) > 1
ORDER BY o.onimi;
```

onro	onimi	projektien_lkm
4	Hallinto	2
5	Tutkimus	3

Aggregointifunktiot - kyselyn evaluointi

Vaiheen 1 (FROM) tulostaulu

FROM osasto o, projekti p

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
1	Pääkonttori	88	1989-06-19	1	Tuote X	5
4	Hallinto	98	1992-01-01	1	Tuote X	5
5	Tutkimus	33	2000-05-22	1	Tuote X	5
1	Pääkonttori	88	1989-06-19	2	Tuote Y	5
4	Hallinto	98	1992-01-01	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
1	Pääkonttori	88	1989-06-19	3	Tuote Z	5
4	Hallinto	98	1992-01-01	3	Tuote Z	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
1	Pääkonttori	88	1989-06-19	10	Uudet edut	4
4	Hallinto	98	1992-01-01	10	Uudet edut	4
5	Tutkimus	33	2000-05-22	10	Uudet edut	4
1	Pääkonttori	88	1989-06-19	20	TYKY-liikunta	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4
5	Tutkimus	33	2000-05-22	20	TYKY-liikunta	4

Aggregointifunktiot - kyselyn evaluointi

Vaiheen 2 (WHERE) tulostaulu:

karteesisesta tulosta poistettu ne rivit, jotka eivät täytä liitos- ja valintaehtoja

```
WHERE o.onro = p.onro
```

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

Aggregointifunktiot - kyselyn evaluointi

Vaiheen 3 (GROUP BY) tulostaulu:

rivit jaettu ryhmiin ryhmittelytekijöiden (sarakkeiden) mukaisesti
(tyhjiä rivejä on käytetty havainnollistamaan ryhmittelyä)

GROUP BY o.onro, o.onimi

onro	onimi	johtajanro	aloituspv	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

Aggregointifunktiot - kyselyn evaluointi

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

Vaiheen 4 (HAVING) tulostaulu:

tulostaulusta poistetaan ne ryhmät, jotka eivät täytä ryhmille asetettuja rajoitteita
(Tässä tapauksessa ei poistettu yhtään ryhmää.)

HAVING COUNT(*) > 1

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

Aggregointifunktiot - kyselyn evaluointi

onro	onimi	johtajanro	aloituspm	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

Vaiheen 5 (SELECT) tulostaulu:

Laskettu aggregointifunktioiden arvot kullekin ryhmälle.

Poistettu kaikki sarakkeet, jotka eivät esiinny SELECT-listassa.

Tuotettu yksi tulosrivi ryhmää kohden.

```
SELECT o.onro, o.onimi, COUNT(*) AS projektien_lkm
```

onro	onimi	projektien_lkm
5	Tutkimus	3
4	Hallinto	2

Aggregointifunktiot - kyselyn evaluointi

onro	onimi	projektien_lkm
5	Tutkimus	3
4	Hallinto	2

Vaiheen 6 (ORDER BY) tulostaulu ja kyselyn lopullinen tulostaulu:

Lajiteltu tulosrivit osaston nimen mukaisesti nousevaan järjestykseen.

ORDER BY o.onimi

onro	onimi	projektien_lkm
4	Hallinto	2
5	Tutkimus	3

Aggregointifunktiot – kyselyn evaluointi

Haetaan kullekin osastolle osaston numero, osaston nimi ja osaston valvomien projektien lukumäärä. Tulokseen halutaan kaikki osastot – myös sellaiset, joilla ei ole yhtään projektia valvottavana. Tulokset järjestetään projektien lukumäärän mukaan nousevaan järjestykseen.

```
SELECT o.onro, o.onimi, COUNT(pnro) AS projektien_lkm
FROM osasto o LEFT OUTER JOIN projekti p
      ON o.onro = p.onro
GROUP BY o.onro, o.onimi
ORDER BY projektien_lkm;
```

onro	onimi	projektien_lkm
1	Pääkonttori	0
4	Hallinto	2
5	Tutkimus	3

FROM

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
1	Pääkonttori	88	1989-06-19	1	Tuote X	5
4	Hallinto	98	1992-01-01	1	Tuote X	5
5	Tutkimus	33	2000-05-22	1	Tuote X	5
1	Pääkonttori	88	1989-06-19	2	Tuote Y	5
4	Hallinto	98	1992-01-01	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
1	Pääkonttori	88	1989-06-19	3	Tuote Z	5
4	Hallinto	98	1992-01-01	3	Tuote Z	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
1	Pääkonttori	88	1989-06-19	10	Uudet edut	4
4	Hallinto	98	1992-01-01	10	Uudet edut	4
5	Tutkimus	33	2000-05-22	10	Uudet edut	4
1	Pääkonttori	88	1989-06-19	20	TYKY-liikunta	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4
5	Tutkimus	33	2000-05-22	20	TYKY-liikunta	4

FROM-vaiheen tulostaulu

FROM osasto o LEFT OUTER JOIN projekti p ON o.onro = p.onro

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
1	Pääkonttori	88	1989-06-19			
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

GROUP BY -vaiheen tulostaulu:

rivit jaettu ryhmiin ryhmittelytekijöiden (sarakkeiden) mukaisesti
(tyhjiä rivejä on käytetty havainnollistamaan ryhmittelyä)

GROUP BY o.onro, o.onimi

onro	onimi	johtajanro	aloituspvm	pnro	pnimi	onro
5	Tutkimus	33	2000-05-22	1	Tuote X	5
5	Tutkimus	33	2000-05-22	2	Tuote Y	5
5	Tutkimus	33	2000-05-22	3	Tuote Z	5
1	Pääkonttori	88	1989-06-19			
4	Hallinto	98	1992-01-01	10	Uudet edut	4
4	Hallinto	98	1992-01-01	20	TYKY-liikunta	4

Aggregointifunktiot – kyselyn evaluointi

SELECT-vaiheen tulostaulu:

Laskettu aggregointifunktioiden arvot kullekin ryhmälle.

Poistettu kaikki sarakkeet, jotka eivät esiinny SELECT-listassa.

Tuotettu yksi tulosrivi ryhmää kohden.

```
SELECT o.onro, o.onimi, COUNT(pnro) AS projektien_lkm
```

onro	onimi	projektien_lkm
5	Tutkimus	3
1	Pääkonttori	0
4	Hallinto	2

Aggregointifunktiot – kyselyn evaluointi

ORDER BY -vaiheen tulostaulu ja kyselyn lopullinen tulostaulu:

Lajiteltu tulosrivit projektien lukumäärän mukaan nousevaan järjestykseen.

ORDER BY projektien_lkm

onro	onimi	projektien_lkm
1	Pääkonttori	0
4	Hallinto	2
5	Tutkimus	3