

Tentamen Databases

21 april 2016

17:00 - 19:00, Educatorium-Gamma

- **Lees eerst onderstaande aanwijzingen.**
- Vermeld op elk vel je naam en studentnummer. Indien één van deze zaken ontbreekt, wordt het vel niet nagekeken.
- Vul alle antwoorden in op deze tentamenbundel. Het bijgeleverde tentamenpapier is uitsluitend bedoeld als kladpapier.
- Bij elke meerkeuze vraag is één antwoord correct (of optimaal), tenzij anders vermeld. Omcirkel de letter voor dit antwoord. Als je je keuze wilt corrigeren, kras de letter dan door en omcirkel alsnog de juiste letter. Als je veelvuldig je antwoord hebt gecorrigeerd, vermeld dan de letter van het juiste antwoord in de kantlijn links van de vraag en kras de andere letters door.
- Achter het nummer van elke vraag staat hoeveel punten deze vraag waard is. Het maximale aantal punten dat je kunt verdienen is 120.
- Toon bij het inleveren je collegekaart.
- Schrijf en formuleer duidelijk.
- Je mag een A4 met aantekeningen raadplegen.
- Het tentamen duurt 2 uur.
- Gebruikte afko's:
 - RDB = relationele database
 - 2PL = two-phase locking
 - 2PC = two-phase commit
 - 3NF = derde normaalvorm
 - BCNF = Boyce-Codd normaalvorm
 - COORD = coördinator
 - CTP = coöperatief terminatieprotocol van 2PC
 - DP = dependency preserving
 - FD = functional dependency
 - RA = relationele algebra
 - SQL = SQL (Structured Query Language)
- **Succes!**

Studentnr:	Achternaam:
------------	-------------

Algemeen (16 punten)

De vragen onder de noemer Algemeen worden bij een fout antwoord met een negatieve score (het aantal vermelde punten) beoordeeld. Bij ontbrekend antwoord is de score 0. Reden hiervoor is de verwachte scoringskans van 50% bij louter gokken.

De gezamenlijke score over de vragen onder de noemer Algemeen (dus vraag 1-8) kan niet negatief worden.

Vraag 1 [2]

Als op basis van een casus door verschillende competente ontwerpers een ER-diagram wordt opgesteld, dan zal dit, behoudens variaties in naamgeving, altijd hetzelfde ER-diagram zijn.

A: juist

B: onjuist

Vraag 2 [2]

De volgende voorwaarde is voldoende en noodzakelijk voor de eigenschap DP: elke FD uit de gegeven FD-set overleeft de projectie op de decompositie.

A: juist

B: onjuist

Vraag 3 [2]

De volgende voorwaarde is voldoende en noodzakelijk voor de eigenschap DP: elke FD uit de closure van de gegeven FD-set overleeft de projectie op de decompositie.

A: juist

B: onjuist

Vraag 4 [2]

De stelling $X \twoheadrightarrow Y \Rightarrow X \rightarrow Y$ is

A: juist

B: onjuist

Vraag 5 [2]

Als niet alle deelnemers aan het CTP gestemd hebben na het verzoek van de oorspronkelijke COORD, leidt het CTP tot Abort.

A: juist

B: onjuist

Studentnr:	Achternaam:
-------------------	--------------------

Vraag 6 [2]

Het terminatieprotocol van 2PC kan ook tot blokkade leiden als er geen sprake is van een netwerkpartitie.

A: juist

B: onjuist

Vraag 7 [2]

Een B-tree is geschikt voor range queries.

A: juist

B: onjuist

Vraag 8 [2]

Twee dezelfde SQL-queries kunnen leiden tot dezelfde methodes voor de berekening van het resultaat van de query op verschillende Database Management Systemen.

A: juist

B: onjuist

Dependencies (16 punten)

De vragen over Dependencies worden bij een fout antwoord met een negatieve score beoordeeld. Bij ontbrekend antwoord is de score 0. De gezamenlijke score over de vragen onder de noemer Dependencies kan niet negatief worden.

We hebben het volgende relatieschema met gegevens betreffende films.

IMDB (filmid, genre, company, director, actorid, year, title, character)

Voor elke film bestaat een unieke, identificerende code. Hetzelfde geldt voor acteurs. Een film kan mogelijk passen bij verschillende genres. Elke film wordt uitgebracht door een filmmaatschappij (company) en heeft een titel en een regisseur (director). Binnen een film speelt een acteur een karakter. Een karakter kan in meerdere films terugkomen. Denk aan *Q* en *miss Moneypenny* en de held zelf in James Bond-films. In *Dr. Strangelove* speelt Peter Sellers zowel *Dr. Strangelove* als *Lionel Mandrake* als *Merkin Muffley*. Gedurende de opnamen van *The Imaginarium of Doctor Parnassus* overleed Heath Ledger. De rol van *Tony* werd overgenomen door Johnny Depp, Jude Law en Colin Farrell.

Geef van de volgende FD's aan of deze wel of niet geldig zijn. Geef desgewenst een korte toelichting, maar alleen als je vermoedt dat de geldigheid van de FD afhankelijk is van een bepaalde interpretatie.

Studentnr:	Achternaam:
-------------------	--------------------

Vraag 9 [1]

De FD *filmid* → *genre*

A: geldt

B: geldt niet

Vraag 10 [1]

De FD *filmid* → *company*

A: geldt

B: geldt niet

Vraag 11 [1]

De FD *title* → *company*

A: geldt

B: geldt niet

Vraag 12 [1]

De FD *director* → *company*

A: geldt

B: geldt niet

Vraag 13 [1]

De FD *company* → *director*

A: geldt

B: geldt niet

Vraag 14 [1]

De FD *filmid, actorid* → *company*

A: geldt

B: geldt niet

Vraag 15 [1]

De FD *character* → *actorid*

A: geldt

B: geldt niet

Vraag 16 [1]

De FD *actorid* → *character*

A: geldt

B: geldt niet

Studentnr:	Achternaam:
-------------------	--------------------

Vraag 17 [1]

De FD *character, filmid* \rightarrow *actorid*

A: geldt

B: geldt niet

Vraag 18 [1]

De FD *actorid, filmid* \rightarrow *character*

A: geldt

B: geldt niet

Vraag 19 [1]

De FD *character, actorid* \rightarrow *filmid*

A: geldt

B: geldt niet

Vraag 20 [1]

De FD *company, actorid* \rightarrow *filmid*

A: geldt

B: geldt niet

Vraag 21 [4]

We beperken ons nu even tot de projectie op *(company, director, actorid)*.

A: De MVD *company* \rightarrow *director* geldt

B: De MVD *company* \rightarrow *actorid* geldt

C: De MVD's *company* \rightarrow *director* en *company* \rightarrow *actorid* gelden beide

D: De MVD's *company* \rightarrow *director* en *company* \rightarrow *actorid* gelden geen van beide

Studentnr:	Achternaam:
------------	-------------

Minimal covers en 3NF (16 punten)

Stel we hebben een relatieschema $R(ABCDEFGG)$ en een set FDs $F = \{C \rightarrow AG, A \rightarrow DF, CD \rightarrow FB, B \rightarrow G, F \rightarrow G\}$.

Vraag 22 [6]

Vul in onderstaande ruimte in welke FDs veranderd worden door de linkerkantreductie en hoe deze er daarna uitzien.

Vraag 23 [6]

Vul in onderstaande ruimte in welke FDs verwijderd worden in verband met overtolligheid.

Vraag 24 [4]

Vul in onderstaande ruimte in welke verliesvrije 3NF, DP decompositie uit de minimal cover gegenereerd kan worden. Geef daartoe de relatieschema's en voor elk schema de candidate keys.

Studentnr:	Achternaam:
------------	-------------

BCNF (16 punten)

De vragen onder de noemer BCNF worden bij een fout antwoord met een negatieve score beoordeeld. Bij ontbrekend antwoord is de score 0. De gezamenlijke score over de vragen onder de noemer BCNF kan niet negatief worden.

We hebben een schema $R(ABCDE)$ en een FDset

$$F = \{ABC \rightarrow D, C \rightarrow E, D \rightarrow B\}$$

Vraag 25 [2]

Is de decompositie (ABCD), (CE) verliesvrij?

A: ja

B: nee

Vraag 26 [2]

Is de decompositie (ABCD), (CE) in 3NF?

A: ja

B: nee

Vraag 27 [2]

Is de decompositie (ABCD), (CE) in BCNF?

A: ja

B: nee

Vraag 28 [2]

Is de decompositie (ABCD), (CE) DP?

A: ja

B: nee

Vraag 29 [2]

Is de decompositie (ACD), (BD), (CE) verliesvrij?

A: ja

B: nee

Vraag 30 [2]

Is de decompositie (ACD), (BD), (CE) in 3NF?

A: ja

B: nee

Studentnr:	Achternaam:
-------------------	--------------------

Vraag 31 [2]

Is de decompositie (ACD), (BD), (CE) in BCNF?

A: ja

B: nee

Vraag 32 [2]

Is de decompositie (ACD), (BD), (CE) DP?

A: ja

B: nee

Queries (16 punten)

Een bedrijf dat gegevens bijhoudt over 1-daagse wielrenklassiekers heeft een database met daarin een tabel waarin gegevens van renners worden bijgehouden, een tabel waarin gegevens van wedstrijden worden bijgehouden en een tabel waarin de deelname van renners aan wedstrijden wordt bijgehouden, maar alleen als de wedstrijd is uitgereden.

Elke renner heeft een uniek rennummer. Daarnaast wordt van elke renner de naam, het geslacht, de nationaliteit en de ploeg bijgehouden. Van elke wedstrijd wordt de naam, het kalenderjaar en de lengte van het parcours bijgehouden. Eveneens heeft elke wedstrijd een uniek wedstrijdnummer. De registratie van de deelname van een renner aan een wedstrijd gaat vergezeld van de tijd waarin de renner de wedstrijd verreden heeft, de finishpositie (1 voor de winnaar, etc.) en het aantal punten dat de renner in de wedstrijd behaald heeft. Voor inschrijvingen aan een wedstrijd die nog niet verreden is wordt een aparte tabel gebruikt.

We gebruiken het symbool % voor de division.

R (*nr*, *naam*, *geslacht*, *ploeg*, *land*)

WR (*nr*, *nr*, *tijd*, *positie*, *punten*)

W (*nr*, *naam*, *jaar*, *lengte*)

We hebben de volgende queries en expressies:

Q1: Welke renners hebben elke editie van de Amstel-Goldrace (AG) sinds 2010 uitgereden?

Q2: Welke renners hebben in hetzelfde jaar zowel de Ronde van Vlaanderen (RV) als Parijs-Roubaix (PR) gewonnen.

Q3: In welke wedstrijd(en) zijn sinds 2013 uitsluitend Belgen op de derde plaats gefinisht?

Q4: Welke renner heeft in zijn carrière het hoogste aantal punten gescoord?

E1: $\pi_{rnr,rnaam}(\sigma_{jaar \geq 2010} \wedge wnaam = 'AG')(W) \bowtie WR \bowtie R$

E2: $\pi_{rnr,rnaam}(\sigma_{wnaam='RV' \wedge wnaam='PR'}(W \bowtie \sigma_{positie=1}(WR)) \bowtie R)$

E3: $\pi_{rnr,rnaam}(\sigma_{wnaam='RV' \wedge wnaam='PR'}(W \bowtie \sigma_{positie=1}(WR) \bowtie R))$

E4: $\pi_{rnr,rnaam}(R \bowtie \pi_{rnr,wnr}(WR) \% \pi_{wnr}(\sigma_{jaar \geq 2010} \wedge wnaam = 'AG')(W))$

E5: $\pi_{wnaam}(\sigma_{land='Belgie'}(R) \bowtie \sigma_{positie=3}(WR) \bowtie \sigma_{jaar \geq 2013}(W))$

E6: $\pi_{wnaam}(W) - \pi_{wnaam}(\sigma_{land='Belgie'}(R) \bowtie \pi_{rnr,wnr}(\sigma_{positie=3}(WR)) \bowtie \sigma_{jaar \geq 2013}(W))$

E7: $\pi_{wnaam}(W) - \pi_{wnaam}(\sigma_{land \neq 'Belgie'}(R) \bowtie \pi_{rnr,wnr}(\sigma_{positie=3}(WR)) \bowtie \sigma_{jaar \geq 2013}(W))$

E8:

```
SELECT rnr, rnaam FROM R
WHERE NOT EXISTS (
    SELECT * FROM WR, W
    WHERE wnaam <> 'AG' AND jaar >= 2010 AND WR.rnr = W.wnr
)
```

E9:

```
SELECT rnr, rnaam FROM R
WHERE rnr NOT IN (
    SELECT rnr FROM WR, W
    WHERE wnaam <> 'AG' AND jaar >= 2010 AND WR.rnr = W.wnr
)
```

E10:

```
SELECT rnr, rnaam FROM R
WHERE rnr NOT IN (
    SELECT rnr FROM WR, W
    WHERE wnaam = 'AG' AND jaar < 2010 AND WR.rnr = W.wnr
)
```

E11:

```
SELECT wnr, wnaam FROM W
WHERE wnr NOT IN (
    SELECT wnr FROM W
    WHERE jaar >= 2013
    AND wnr IN (
        SELECT wnr FROM WR, R
        WHERE WR.rnr = R.rnr AND positie = 3 AND land <> 'Belgie'
    )
))
```

E12:

```
SELECT wnr, wnaam FROM W
WHERE wnr NOT IN (
    SELECT W.wnr FROM WR, R, W
    WHERE WR.rnr = R.rnr AND WR.wnr = W.wnr
    AND positie = 3 AND jaar >= 2013 AND land <> 'Belgie'
)
```

E13:

```
SELECT rnr, rnaam FROM R, WR
WHERE R.rnr = WR.rnr
GROUP BY rnr, rnaam
HAVING MAX(COUNT(punten))
```

E14:

```
SELECT rnr, rnaam FROM R, WR
WHERE R.rnr = WR.rnr
GROUP BY rnr, rnaam
HAVING MAX(SUM(punten))
```

Studentnr:	Achternaam:
-------------------	--------------------

E15:

```
SELECT rnr, rnaam FROM R, WR
WHERE R.rnr = WR.rnr
GROUP BY rnr, rnaam
HAVING SUM(punten) >= ALL (
    SELECT SUM(punten) FROM R, WR
    WHERE R.rnr = WR.rnr
    GROUP BY rnr, rnaam)
```

E16:

```
SELECT rnr, rnaam FROM R, WR
WHERE R.rnr = WR.rnr
GROUP BY rnr, rnaam
HAVING SUM(punten) >= ALL (
    SELECT SUM(punten) FROM WR
    GROUP BY rnr)
```

De relatie tussen queries en expressies is many-to-many en optional.

Vraag 33 [4]

Geef hieronder aan welke expressies corresponderen met Q1.

--

Vraag 34 [4]

Geef hieronder aan welke expressies corresponderen met Q2.

--

Vraag 35 [4]

Geef hieronder aan welke expressies corresponderen met Q3.

--

Vraag 36 [4]

Geef hieronder aan welke expressies corresponderen met Q4.

--

Studentnr:	Achternaam:
------------	-------------

Concurrency (14 punten)

We beschouwen de volgende twee schedules.

<i>S1</i>					<i>S2</i>				
T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
	w(x)				w(y)				
r(z)				w(x)				w(y)	
	w(y)				r(y)				
r(y)				w(y)	w(z)				w(x)
		w(z)						r(z)	
			w(y)		w(x)				
				r(z)	r(x)				

Vraag 37 [2]

Geef een opsomming van *alle* pijlen van de precedentiegraaf $G(S_1)$. Gebruik voor een pijl dit format: $T_i \rightarrow T_j$. Sorteert de lijst oplopend voor T_i .

Vraag 38 [2]

Is S_1 serializeerbaar? Beargumenteer. Zo ja, geef een equivalente seriële schedule van S_1 .

Studentnr:	Achternaam:
-------------------	--------------------

Vraag 39 [2]

Geef een opsomming van *alle* pijlen van de precedentiegraaf $G(S_2)$, wederom gesorteerd.

Vraag 40 [2]

Is S_2 serializeerbaar? Beargumenteer. Zo ja, geef een equivalente seriële schedule van S_2 .

Vraag 41 [6]

Welke van de transacties hierboven vertonen geen 2PL-gedrag? (Nul of meer antwoorden mogelijk.)

A: T_1 van S_1

B: T_2 van S_1

C: T_3 van S_1

D: T_4 van S_1

E: T_5 van S_1

F: T_1 van S_2

G: T_2 van S_2

H: T_3 van S_2

I: T_4 van S_2

J: T_5 van S_2

Studentnr:	Achternaam:
------------	-------------

Recovery (16 punten)

We beschouwen nonquiescent recovery met **REDO** logging. Hieronder vind je een log met after images.

```
<START T1>
<T1, B, 10>
<COMMIT T1>
<START T3>
<START T2>
<T2, A, 5>
<T3, C, 10>
<START CKPT (T2, T3)>
<T2, D, 15>
<START T4>
<START T5>
<T4, F, 25>
<T5, H, 12>
<COMMIT T2>

<COMMIT T3>
<END CKPT>
<COMMIT T4>
<T5, G, 30>
```

Stel dat een crash optreedt direct na `<COMMIT T2>` . (Het resterende gedeelte van de log wordt dan niet geschreven.)

Vraag 42 [3]

Welk gedeelte van de log file wordt gescand?

Studentnr:	Achternaam:
-------------------	--------------------

Vraag 43 [3]

Welke transacties worden redone?

Stel nu dat een crash optreedt direct na `<T5, G, 30>` .

Vraag 44 [3]

Welk gedeelte van de log file wordt gescand?

Vraag 45 [3]

Welke transacties worden redone?

Vraag 46 [4]

We kijken nu naar een andere log, maar dan in het kader van **UNDO** logging.

`<START T3>`

`<START T2>`

`<T2, A, 5>`

`<T3, C, 10>`

`<T2, C, 15>`

`<COMMIT T2>`

`<COMMIT T3>`

Welk bezwaar kun je inbrengen tegen deze executievolgorde?

Studentnr:	Achternaam:
------------	-------------

Query processing (10 punten)

De vragen onder de noemer *Query processing* worden bij een fout antwoord met een negatieve score beoordeeld. Bij ontbrekend antwoord is de score 0. De gezamenlijke score over de vragen onder de noemer *Query processing* kan niet negatief worden.

Geef van de hieronder genoemde algebraïsche equivalenties aan of deze gelden. De context is verzamelingen, geen bags. Bij vraag 47 en 48 geldt dat de schema's van R en S identiek zijn en dat de schema's van T en U identiek zijn

$\Gamma_{X,F}$ staat voor *group by* op attribootset X met aggregate functie F .

Vraag 47 [2]

$$(R \bowtie T) \cap (S \bowtie U) = (R \cap S) \bowtie (T \cap U)$$

A: ja

B: nee

Vraag 48 [2]

$$(R \cup S) \bowtie (T \cup U) = (R \bowtie T) \cup (S \bowtie U) \cup (R \bowtie U) \cup (S \bowtie T)$$

A: ja

B: nee

Vraag 49 [2]

$$\pi_L(\Gamma_{X,F}(R)) = \Gamma_{X,F}(\pi_L(R)),$$

waarbij de attributen van L bevat zijn in X

A: ja

B: nee

Vraag 50 [2]

$$\sigma_p(\Gamma_{X,F}(R)) = \Gamma_{X,F}(\sigma_p(R)),$$

waarbij de attributen waaraan p refereert bevat zijn in X

A: ja

B: nee

Vraag 51 [2]

$$\Gamma_{X,F}(R) \cap \Gamma_{X,F}(S) = \Gamma_{X,F}(R \cap S),$$

waarbij de schema's van R en S identiek zijn

A: ja

B: nee

Einde