

# T-106.5241 Distributed Databases

Exam, May 10th, 2010

Write the following clearly on top of every paper you submit: "T-106.5241, May 10th, 2010", your full name, student ID and study programme, and the **total number of papers** you submit.

1. (6p) Explain the following concepts briefly:
  - (a) horizontal partitioning
  - (b) one-phase commit protocol
  - (c) weak mutual consistency of replicas
  - (d) semijoin
  - (e) hash partitioning
  - (f) page server (also known as: data server with page shipping)
2. (a) (4p) Describe the two-phase commit protocol that was used on the course (i.e., the version that is based on the presumed abort property). What messages are transferred between the different sites? What log records are produced? When is the log forced to disk? You can assume that the transaction commits successfully and that none of the sites crash.  
(b) (2p) In the two-phase commit protocol, there exists a situation where a cohort (=participant) may need to wait for a long time before it can either commit or abort its subtransaction. What is this situation?
- 3.(a-c) (3p) Are the following combinations possible under the quorum consensus protocol?  $n$  = number of replicas,  $p$  = read quorum,  $q$  = write quorum.
  - (a)  $n = 15, p = 1, q = 1$
  - (b)  $n = 15, p = 4, q = 12$
  - (c)  $n = 15, p = 8, q = 7$(d) (1p) In the parts (a) to (c) that were possible, how many of the sites containing replicas can crash, so that the quorum consensus protocol can still continue without needing to wait for the crashed sites?  
(e) (2p) Give an example of one-item deadlock in a replicated database.
4. How should the following queries be parallelized in a parallel database system? The values of attribute  $A$  are evenly distributed in the range  $[0..1000]$ . The relations  $r(\underline{A}, B)$  and  $s(\underline{A}, C)$  are range-partitioned (with the same ranges) to all processors. The final result of the query should be given to a single processor.
  - (a) (1p) **select count(distinct  $A$ ) from  $r$ .**
  - (b) (2p) **select distinct  $B$  from  $r$  where  $A > 500$ .**
  - (c) (1p) **select count(\*) from  $r, s$  where  $r.A = s.A$  and  $B < 20$  and  $C < 50$ .**
  - (d) (2p) **select distinct  $C$  from  $r, s$  where  $r.A = s.A$  and  $B < 20$ .**
5. (a) (4p) Site  $s_1$  of a distributed database crashes. At the time of the crash, the contents of the log on disk are the following:  
101:  $\langle \text{begin-checkpoint} \rangle$   
102:  $\langle \text{transaction-table}, \{\} \rangle$   
103:  $\langle \text{page-table}, \{\} \rangle$   
104:  $\langle \text{end-checkpoint} \rangle$   
105:  $\langle T_1, s_0, T, B \rangle$   
106:  $\langle T_1, W, p, i, x, \dots \rangle$   
107:  $\langle T'_1, s_0, T', B \rangle$   
108:  $\langle T'_1, W, q, j, y, \dots \rangle$   
109:  $\langle T_1, \{s_0, s_1, s_2\}, P \rangle$   
110:  $\langle S, B \rangle$   
111:  $\langle S, W, r, k, z, \dots \rangle$   
112:  $\langle S, \{s_1, s_2\}, P \rangle$   
Restart recovery is performed using the ARIES algorithm. What are the contents of the reconstructed transaction table at the end of the analysis phase? What happens to the transactions  $T_1, T'_1$  ja  $S$  in the undo phase? Does the recovery algorithm need to acquire any locks in this situation?  
(b) (2p) In a page server (also known as data server with page shipping), what happens when the server  $s$  notices that a client  $c$  has crashed? What does  $c$  need to do when it later restarts?

# T-106.5241 Hajautetut tietokannat

Tentti, 10. 5. 2010

Kirjoita jokaisen palauttamasi paperin ylälaitaan selvästi: "T-106.5241, 10. 5. 2010", koko nimesi, opiskelijanumerosi ja koulutusohjelmasi sekä palauttamiesi **paperien kokonaismäärä**.

1. (6p) Selitä lyhyesti seuraavat käsitteet:
  - a) vaakasuora osittaminen (horizontal partitioning)
  - b) yksivaiheinen sitoutumiskäytäntö (one-phase commit protocol)
  - c) toisinteiden heikko keskinäinen johdonmukaisuus (weak mutual consistency of replicas)
  - d) puoliliitos (semijoin)
  - e) hajautusositus (hash partitioning)
  - f) sivupalvelin (page server / data server with page shipping)
2. a) (4p) Kuvaa kaksivaiheinen sitoutumiskäytäntö (two-phase commit protocol) kurssimateriaalissa esitetyllä tavalla (eli keskeytysoletukseen, presumed abort, perustuen). Mitä viestejä eri pisteiden välillä liikkuu? Mitä lokikirjauksia tehdään? Milloin loki pakotetaan levyllä? Oleta, että transaktio sitoutuu onnistuneesti eikä mikään piste romahda.  
b) (2p) Missä tilanteessa kaksivaiheisen sitoutumiskäytännön osallinen (cohort, participant) voi joutua odottamaan pitkäänkin ennen kuin pystyy joko sitouttamaan tai keskeyttämään oman alitransaktionsa?
3. a-c) (3p) Ovatko seuraavat yhdistelmät mahdollisia päätösvaltakäytännössä (quorum consensus protocol)?  $n$  = toisinteiden määrä,  $p$  = lukuvallan koko (read quorum),  $q$  = kirjoitusvallan koko (write quorum).
  - a)  $n = 15, p = 1, q = 1$
  - b)  $n = 15, p = 4, q = 12$
  - c)  $n = 15, p = 8, q = 7$d) (1p) Niissä kohdista a-c), jotka edellä olivat mahdollisia, kuinka moni toisinnuspiste voi romahtaa ilman että päätösvaltakäytännön soveltaminen estyy?  
e) (2p) Anna esimerkki yhdestä tietoalkiosta johtuvasta lukkiumasta (one-item deadlock) toisinetussa tietokannassa.
4. Miten seuraavat kyselyt kannattaa rinnakkaistaa rinnakkaistietokannassa? Attribuutin  $A$  arvot ovat jakautuneet tasaisesti välille  $[0..1000]$ . Relaatiot  $r(A, B)$  ja  $s(A, C)$  on osaväliositettu (range-partitioned) samoilla osaväleillä kaikille prosessoreille. Kyselyn tulos halutaan yhdelle prosessorille.
  - a) (1p) **select count(distinct A) from r.**
  - b) (2p) **select distinct B from r where A > 500.**
  - c) (1p) **select count(\*) from r, s where r.A = s.A and B < 20 and C < 50.**
  - d) (2p) **select distinct C from r, s where r.A = s.A and B < 20.**
5. a) (4p) Hajautetun tietokannan pisteen  $s_1$  lokin sisältö levyllä häiriön sattuessa on seuraava:
  - 101:  $\langle \text{begin-checkpoint} \rangle$
  - 102:  $\langle \text{transaction-table}, \{\} \rangle$
  - 103:  $\langle \text{page-table}, \{\} \rangle$
  - 104:  $\langle \text{end-checkpoint} \rangle$
  - 105:  $\langle T_1, s_0, T, B \rangle$
  - 106:  $\langle T_1, W, p, i, x, \dots \rangle$
  - 107:  $\langle T'_1, s_0, T', B \rangle$
  - 108:  $\langle T'_1, W, q, j, y, \dots \rangle$
  - 109:  $\langle T_1, \{s_0, s_1, s_2\}, P \rangle$
  - 110:  $\langle S, B \rangle$
  - 111:  $\langle S, W, r, k, z, \dots \rangle$
  - 112:  $\langle S, \{s_1, s_2\}, P \rangle$Häiriöstä elvytään ARIES-algoritilla. Mikä on analyysivaiheen tuloksena rekonstruoidun transaktiotaulun sisältö? Mitä transaktioille  $T_1, T'_1$  ja  $S$  tehdään elvytyksen peruutusvaiheessa? Täytyykö elvytysalgoritmien varata mitään lukkoja tässä tilanteessa?  
b) (2p) Mitä sivupalvelinjärjestelmässä (page server, data server with page shipping) tapahtuu, kun palvelin  $s$  huomaa, että asiakas  $c$  on romahtanut? Entä mitä  $c:n$  pitää tehdä, kun se myöhemmin käynnistyy uudestaan?