

EXAMEN DE CBDE

9 de Gener del 2020

Instruccions: Respon cada pregunta al full corresponent.

L'examen dura 1 hora.

Nom i Cognoms:

PREGUNTA 1. [3p]

Considera la següent història de transaccions en una base de dades que implementa time-stamping estàtic sense cap mecanisme per garantir la recuperabilitat:

#TS	T1	T2	T3
10	BoT		
20		BoT	
30			BoT
40	R(A)		
50	W(A)		
60		R(A)	
70			R(A)
80			R(B)
90		R(B)	
100		COMMIT	
110			R(D)
120			COMMIT
130	ROLLBACK		

Es cancel·larà alguna transacció? Justifica la teva resposta

.....
.....

Un cop aplicat l'algoritme de time-stamping, la història resultant és lliure d'interferències? Justifica la teva resposta

.....
.....

Finalment, digues si la següent afirmació és correcta (justifica la teva resposta):

El time-stamping dinàmic es diferencia del time-stamping estàtic en que aquest evita certs tipus d'interferències que el time-stamping estàtic no pot evitar

.....
.....

PREGUNTA 2. [4p]

Suposa el següent esquema denormalitzat a implementar en HBase:

Compres (llibreId, data, preu, unitats, autor, editorial, ISBN, anyPublicacio)

I les següents consultes:

- | | |
|--|--|
| a) SELECT llibreId, SUM(unitats)
WHERE editorial = 'RBA'
GROUP BY llibreId | b) SELECT AVG(preu)
WHERE data BETWEEN '01/01/xxxx' AND
'31/12/XXXX'
GROUP BY editorial |
| c) SELECT autor, editorial,
COUNT(*)
GROUP BY autor, editorial | d) SELECT editorial, AVG(unitats)
WHERE data BETWEEN '01/01/xxxx' AND
'31/12/XXXX'
GROUP BY editorial |

Les consultes tenen la següent freqüència, a 50%, b 20%, c 10% i d 20%

Ja has decidit que la clau d'HBase serà la indicada a la signatura de la taula. Però et cal decidir quines famílies crear. Empra la tècnica *attribute affinity matrix* amb una afinitat superior o igual al 70% **per decidir quines famílies d'HBase implementar**. Identifica clarament les famílies que crearies i els atributs que contindria cada familia.

Nom i Cognoms:

PREGUNTA 3. [3p]

PROCESSAMENT DE DADES DISTRIBUÏT

Entre les combinacions següents (base de dades, framework de processament): HBase / MapReduce, MongoDB / Aggregation Framework i Neo4J / Cypher ordena-les (de major a menor) segons els següents criteris:

Capacitat per distribuir dades:

Capacitat de processament en paral·lel:

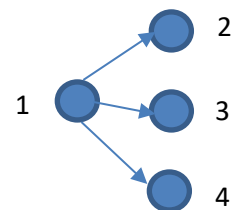
Finalment, digues si la següent afirmació és correcta (justifica la teva resposta):

La distribució de dades i el paral·lelisme que obté un sistema són conceptes independents sense relació entre ells

.....
.....
.....

GRAPH DATABASES

Dibuixa sota aquesta línia la matriu d'adjacència del següent graf:



Neo4J, internament, representaria aquest graf com una llista d'adjacència. Dibuixa l'estructura que Neo4J crearia per implementar aquest graf. Pots obviar l'existència d'etiquetes (labels) i atributs i centrar-te només en els nodes i les arestes. Considera els identificadors de la figura com els identificadors que Neo4J ha assignat a cada node.