

EXAMEN DE CBDE

14 de Gener del 2015

Instruccions: Respon cada pregunta al full corresponent. L'examen dura 1h 30'

Nom i Cognoms:

PREGUNTA 1. [4p]

Considera els dos sistemes que es mostren a continuació:

Sistema 1: Base de dades centralitzada

- És una base de dades centralitzada amb un únic node
- La base de dades ocupa 30TB. Per simplicitat, assumeix que conté una única taula T amb 500.000.000 de tuples, i cada tupla té 5 atributs
- La latència de llegir de disc és de 5ms
- L'ample de banda màxim que pot aconseguir el disc és de 100Mb/s

Sistema 2: Base de dades distribuïda

- Consta de 10 nodes connectats per una LAN. Qualsevol d'ells pot llençar consultes
- La base de dades ocupa 30TB i també té 500.000.000 de tuples (amb 5 atributs per tupla). Pots assumir que la taula T s'ha distribuït de forma uniforme entre els nodes mitjançant una fragmentació horitzontal. No hi ha replicació
- A cada node s'aplica, a més, una fragmentació vertical pura (un atribut per fragment). Pots assumir que tots els atributs ocupen el mateix espai a disc
- La latència de llegir de disc és de 5ms i la de la xarxa 1 ms
- L'ample de banda màxim que pot aconseguir el disc és de 100Mb/s
- Els nodes estan connectats a través d'una LAN amb ample de banda màxim de 10Mb/s

Suposa que l'única *query* del sistema és:

```
SELECT DISTINCT(b) FROM T, on b és un atribut de T amb 100 valors de  
domini diferents (és a dir, ndist(b) = 100)
```

No hi ha índexs ni cap altra estructura definida en el sistema.

En el millor cas, quant trigarà (en segons) en fer un accés seqüencial de T en el sistema 1?

Latència

Lectura seqüencial

Total

I en el sistema 2 en el millor cas?

Latència

Lectura seqüencial

Total

Que canviaria si la query fos `SELECT SUM(b) FROM T`? Justifica la resposta.

.....

.....

Si has tingut que fer cap assumptió addicional (de les explicitades a l'enunciat), explica-la aquí:

.....

.....

PREGUNTA 2. [4p]

- 1 Identifica les dues estructures bàsiques que utilitzen els gestors de bases de dades distribuïts per implementar el catàleg global de la base de dades.

1.1 Dona un exemple de gestor que implementi cada una d'aquestes estructures.

- 2 Durant el curs hem fet pràctiques amb *HBase / MapReduce*, *MongoDB / Aggregation Framework* i *Neo4J / Cypher*. Ordena de major a menor aquestes parelles *gestor / llenguatge de consultes* segons el grau de paral·lelisme que poden assolir en el millor cas. Justifica la teva resposta.

- 3 Considera un gestor amb un sistema de recuperació basat en logging i steal / no force. Justifica si cal que aquest sistema de recuperació implementi les funcionalitats UNDO i REDO.

Nom i Cognoms:

PREGUNTA 3. [2p]

Calcula (pots deixar indicada l'operació) la completesa (Q_{cm}) per cadascun dels atributs i també per a la relació següent:

Persones

Nom	Cognom	Telèfon
Alberto Abelló	<NULL>	123456789
Oscar	Romero	987654321
Toni	<NULL>	<NULL>
Xavier	Burgués	<NULL>

Q_{cm} (Nom) =

Q_{cm} (Cognom) =

Q_{cm} (Telèfon) =

Q_{cm} (Persones) =