

EXAMEN DE CBDE

28 de Maig del 2014

Instruccions: Respon cada pregunta al full corresponent.

L'examen dura 1 hora.

Nom i Cognoms:

PREGUNTA 1. [3p]

1 Fes un sketch de com HBase gestiona el catàleg global de la base de dades.

2 Justifica una avantatja i un inconvenient d'aquesta arquitectura respecte al Consistent Hashing de MongoDB.

.....

.....

.....

.....

3 Aplica HBase fragmentació horitzontal i/o vertical? Justifica la teva resposta.

.....

.....

.....

.....

PREGUNTA 2. [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 totes les dades i n'hi ha 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 queries
- 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
- 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 300.000 valors  
de domini diferents (és a dir, ndist(b) = 300.000)
```

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

Pel sistema 2, en el millor cas, estàs assumint *data shipping* i/o *query shipping*? Perquè?

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

PREGUNTA 3. [3p]

Considera un gestor NoSQL que implementa un entorn de recuperació NoSteal / Force i un mecanisme de control de concurrència basat en timestamping estàtic (és a dir, no en bloquejos). A més, comprova recuperabilitat en temps de commit (tal com hem assumit a les transparències i exercicis).

Si dues transaccions fan un update de la mateixa tupla, es pot generar cap problema? Explica breument el perquè i, si cal, com s'arreglaria.