

ApuntsTema2.pdf



onafolch



Desenvolupament d'Aplicacions de Dades Massives



3º Grado en Ingeniería de Datos



Escuela de Ingeniería
Universidad Autónoma de Barcelona

antes



**Descarga sin publi
con 1 coin**



Después

WUOLAH





BIG DATA

REDIS

1. REDIS

Les dades estan emmagatzemades en un únic espai d'adreces, que és la memòria, i per tant es redueix la complexitat de la solució i no cal paginar informació fora de la memòria, ni dissenyar mètodes de coherència. Tenim un accés ràpid i les estructures de dades ja no s'han d'optimitzar per al disc.

Principals problemes a resoldre:

- **Durabilitat:** què fer quan el sistema s'apaga? Hauríem de crear i distribuir còpies de dades i incrementar la redundància.
- **Capacitat:** què fer quan les necessitats de dades són més grans que l'espai de memòria? Hauríem d'afegir més servidors al sistema, de tal manera que tindríem els blocs de dades distribuïts entre diferents servidors, no són còpies.

Al afegir nous servidors, tenim nous problemes. Quan fem consultes complexes, haurem de llegir totes les dades distribuïdes entre els servidors, i necessitem una distributed global join (si per exemple tenir dos servidors i volem trobar el màxim d'algun valor, haurem de buscar en els dos i fer un join per saber quin és el màxim entre els dos elements).

L'ús de dades locals a les bases de dades NoSQL eviten la necessitat de fer consultes join entre particions.

Els avenços tecnològics van millorant, però la freqüència i la capacitat d'un microprocessador CPU no ha avançat, però ens permeten tenir sistemes de bases de dades amb processadors de molts nuclis.

Aquests avenços ens permeten guardar les dades a memòria si el nostre conjunt de dades hi cap, ja que és més ràpid que al disc. La memòria + SSD s'ha convertit en un nou estàndard (ja està predeterminat els serveis de BD al núvol), i la majoria de BD no són tant grans.

Si el conjunt de dades no cap a memòria, no podem utilitzar redis, ja que s'utilitza la memòria com a emmagatzematge principal.

El disc només s'utilitza per fer persistència (per fer còpies, ja que si el sistema cau, es perdran poques dades) i està orientat a l'estructura de dades (és més complex que el sistema de clau-valor). Redis és open source, i ens permet donar informació als usuaris en temps real (amb poca latència).

No s'utilitzen taules, es guarden objectes individuals a memòria.

Ona Folch

WUOLAH

Todos los planes de suscripción incluyen descargas sin publicidad con coins