

# **PAC 1:** Arquitectures de bases de dades no tradicionals

## **Conceptes bàsics i models NoSQL**

**Nom: Marc Bracons Cucó**

## Exercici 1

### 1. Explica les implicacions que l'esquema sigui schemaless a les bases de dades NoSQL respecte a la independència de les dades.

El fet que les bases de dades NoSQL siguin schemaless té diverses implicacions en quan a la independència de les dades.

Es poden afegir dades sense definir prèviament el seu esquema (tal i com es fa en el model relacional). Això permet molta més flexibilitat per tractar amb dades amb diferents formats tot i que en puja la dificultat de programació [1]. A més, tot i que no hi ha un esquema explícit, sempre n'hi ha un d'implícit que condiciona com es distribuiran les dades en las bases de dates NoSQL, cosa que pot comprometre la independència de les dades [2].

### 2. Explica quines limitacions presenta l'estratègia one size fits all respecte: 1) a les consultes, 2) les propietats ACID i 3) el manteniment de l'esquema de la base de dades.

L'estratègia one size fits all presenta diverses limitacions respecte a les consultes, les propietats ACID i el manteniment de l'esquema de la base de dades.

- 1) Consultes: Al dependre principalment de sistemes de gestors de bases de dades relacionals té limitacions quan tracta de realitzar consultes complexes o sobre dades que no tenen una estructura definida [3].
- 2) Propietats ACID: Tot i que aquestes propietats són fonamentals però no sempre necessàries i poden fer més lent el sistema [4], per tant, el compliment estricte d'aquestes propietats en l'estratègia one size fits all pot no ser òptima en tots els escenaris.
- 3) Manteniment del esquema de la base de dades: La seva gestió és molt important en l'estratègia, però això implica dificultats en entorns dinàmics on els requisits poden canviar amb freqüència [5].

### 3. Explica 3 situacions genèriques on sigui més útil el model de dades en graf que el model relacional.

- 1) Base de dades sobre vins [6]: Degut a la complexitat i relacions que existeixen sobre la informació dels vins, és més útil el model de dades en graf que el model relacional, ja que la seva complexitat es pot representar d'una manera més natural mitjançant nodes i connexions, a més, el model en graf ofereix consultes flexibles i independència de l'emmagatzematge.
- 2) Xarxa social [7]: Al igual que en l'exemple anterior, la flexibilitat d'aquest model per representar relacions complexes és un punt clau. A més es poden utilitzar colors diferents per distingir entre diferents entitats.

- 3) Model sobre la relació entre estudiants, cursos i professors: Si volem veure les relacions entre els estudiants que s'inscriuen a cursos que són impartits per professors, i alguns cursos poden ser requisits d'altres cursos, el model en graf podria ser més útil per veure de forma natural i ordenada la jerarquia dels cursos així com les seves relacions amb les persones.

#### **4. Explica 3 situacions on l'ús del model d'agregació és avantatjós respecte d'altres alternatives.**

- 1) Gestió d'Enviaments de comandes [8]: Si es vol recuperar la informació d'una comanda, un model d'agregació permet dissenyar un document que tingui tots els detalls necessaris, evitant haver de realitzar múltiples operacions de combinació entre taules.
- 2) Entorns altament distribuïts [9]: Aquest model és especialment útil en aquesta situació ja que la gestió de dades és més eficaç al fer que l'agregat sigui la unitat de distribució. A més, la seva replicació a través de diferents nodes implica un augment de la disponibilitat, tot i que s'ha de mantindre la coherència de les rèpliques.
- 3) Base de dades BerkeleyDB [10]: el model agregat clau-valor és avantatjós degut a la seva simplicitat i velocitat, cada dada es guarda com un parell de clau i valor. A més, permet flexibilitat en el tipus de dades.

## Exercici 2

**1. Les bases de dades NoSQL tenen un esquema ben definit i estructurat, cosa que facilita la gestió de la base de dades i la implementació de mesures de seguretat.**

Fals. Les bases de dades NoSQL tenen un esquema de dades flexible o no tenen un esquema predefinit, cosa que en dificulta la gestió ja que aquesta es fa explícitament en el codi dels programes que accedeixen a la base de dades. Aquest tipus d'esquema compromet la independència de les dades [2].

**2. Les propietats ACID encaixen millor amb l'estratègia PUSH.**

Fals. Les propietats ACID serveixen per garantir que les transaccions de bases de dades siguin processades de manera fiable. En canvi, l'arquitectura push consisteix en l'intercanvi de flux de dades en temps real, sense necessitat de ser emmagatzemats prèviament [11].

**3. En el disseny de bases de dades en graf, no cal identificar les relacions entre les dades ni assignar-hi noms, ja que el sistema pot inferir automàticament l'estructura de la base de dades a partir de les dades ingressades inicialment.**

Fals. En el disseny de bases de dades en graf la importància de les dades recau majoritàriament en les seves relacions [12].

**4. Els models d'agregació garanteixen la integritat de les dades. És a dir, si tenim les mateixes dades (per exemple les dades d'un client) a diferents agregats de la base de dades, el sistema gestor de bases de dades s'encarregarà de garantir que les diferents ocurrències de les mateixes dades (del mateix client) sempre tinguin els valors actualitzats i consistents.**

Fals. Tot i que els models d'agregació poden introduir redundàncies en les dades, implicant que hi hagi referits de un mateix objecte. Només en cas que el domini de l'aplicació requereixi que es mantingui la coherència es duran a terme processos per fer-ho, i en cas de ser així, serà responsabilitat del programa i no del gestor de la base de dades [13].

**5. Algunes bases de dades basades en el model clau-valor ofereixen flexibilitat en permetre definir una estructura mínima per als agregats i en alguns casos permeten establir relacions entre ells, cosa que el fa adequat per a una varietat de casos d'ús en diferents dominis.**

Cert. Existeixen bases de dades clau-valor que permeten definir una estructuració mínima del agregat i relacions entre ells [14].

## Exercici 3

**Consulta 1:** Per a cada mes, any i tipus de vi (negre, blanc, rosat), volem obtenir el nombre total d'enviaments realitzats d'aquest tipus de vi, el volum mitjà de cada enviament d'aquest tipus de vi i una llista detallada dels enviaments agrupats per regió d'origen del vi. Per a cada regió d'origen, es vol conèixer el nom de la regió, el nombre d'enviaments que contenen vins d'aquesta regió, el volum total de vi enviat des d'aquesta regió i una llista detallada de cada enviament des d'aquesta regió. Per a cada enviament, cal el número de seguiment, la data d'enviament, la destinació de l'enviament, el volum del paquet de vi i el preu unitari del vi.

En el nivell superior tenim Any, Mes i Tipus de Vi ja que son les variables per les que volem fer la consulta. A un nivell inferior hi tenim els dos objectes que es demanen per a aquestes variables; per una un resum del total d'enviaments i per l'altre la llista de regions. Aquesta última ha de tindre tantes caixes com regions i dintre de cada regió tantes caixes com enviaments.

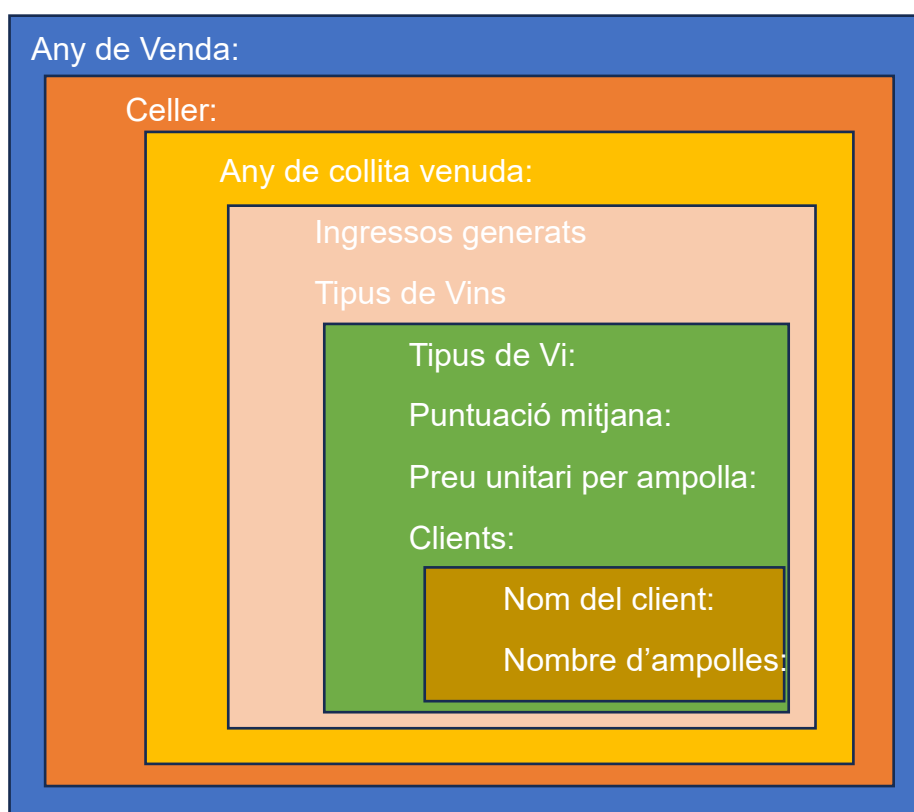


Exemple de JSON:

```
{
  "any": 2024,
  "mes": "Març",
  "tipus_de_vi": "negre",
  "resum_enviaments": {
    "total_enviaments": 150,
    "volum_mitjà": 750
  },
  "regions": [
    {
      "nom_de_la_regió": "Priorat",
      "total_enviaments_de_la_regió": 50,
      "volum_total_enviat": 37500,
      "enviaments": [
        {
          "número_de_seguiment": "123456789",
          "data_enviament": "2024-03-10",
          "destinació": "Barcelona, Espanya",
          "volum_del_paquet_de_vi": 750,
          "preu_unitari_del_vi": 15
        }
      ]
    }
  ],
}
```

**Consulta 2: A nivell anual (any de venda), per a cada celler de vins i cada anyada (any de la collita) venuda, es vol obtenir el total d'ingressos generats per la venda de vins d'aquest celler en aquesta anyada l'any de venda considerat, i per a cada tipus de vi venut mostrar el nom del tipus de vi, una llista amb els noms dels clients que van comprar aquest vi i el nombre d'ampolles comprades per client, la puntuació mitjana atorgada pels crítics i el preu unitari per ampolla.**

Al nivell superior i tenim l'any de venda, després filtrem pel celler i a continuació per l'any de collita venuda. En quan haguem fixat aquestes variables obtindrem els ingressos generats i una llista dels tipus de vins. Per a cada vi obtindrem la puntuació mitjana, el preu unitari i la llista de clients. Per a cada client el seu nom i el nombre d'ampolles comprades.





Exemple de json:

```
{
  "any_de_venda": 2024,
  "cellers": [
    {
      "nom_del_celler": "Celler de l'Empordà",
      "anyades_vendes": [
        {
          "any_de_collita_venduda": 2018,
          "ingressos_generats": 85000,
          "tipus_de_vins": [
            {
              "tipus_de_vi": "Garnatxa Negra",
              "puntuació_mitjana": 89,
              "preu_unitari_per_ampolla": 15,
              "clients": [
                {
                  "nom_del_client": "Bruce Wayne",
                  "nombre_d'ampolles": 30
                },
                {
                  "nom_del_client": "Anthony Soprano",
                  "nombre_d'ampolles": 24
                }
              ]
            }
          ]
        }
      ]
    }
  ]
}
```

## Exercici 4

### **1. Una base de dades relacional és la millor solució per a les necessitats d'emmagatzematge i processament descrites en el supòsit.**

Fals. Un sistema de monitoratge com el que es descriu a l'enunciat requereix una gran quantitat de dades generades pels sensors, a més, aquestes són a temps real. A més, com el sistema aplica estratègies personalitzades per regió es necessita que sigui flexible i escalable.

Els sistemes NoSQL i NewSQL poden realitzar millor aquestes tasques, proporcionant escalabilitat, esquemes flexibles i processament en temps real.

### **2. La utilització d'un esquema explícit per a les dades que cal emmagatzemar seria millor que fer servir un esquema implícit. A més, en cas de no fer servir un model relacional, la millor solució seria utilitzar una base de dades NewSQL.**

Fals. En un esquema explícit, les estructures de dades es defineixen de forma clara i detalla, això implica que es requereix un processament previ de les dades que s'hi emmagatzemaran. En un esquema implícit, en canvi, al no tindre les estructures definides de forma estricta, permet una flexibilitat més gran i capacitat d'adaptació, que és el que necessitem per a aplicar estratègies personalitzades a les regions. Tot i que NewSQL aporta l'escalabilitat de NoSQL amb les garanties de consistència de les bases de dades, a dia d'avui seria millor utilitzar NoSQL, ja que és l'estàndard utilitzat i encara no s'ha avançat prou en l'estandardització del NewSQL.

### **3. Segons els requisits plantejats, la solució ha de ser fàcilment escalable en un ambient distribuït.**

Cert. En l'escenari plantejat es necessita processar dades en temps real recollides de múltiples sensors distribuïts pels camps. La capacitat de manejar els grans volums de dades generats i el seu anàlisi posterior requereixen una solució fàcilment escalable.

### **4. Pel que fa a la replicació, les dades d'interès no permeten una fragmentació ni una replicació eficient. Per tant, seria millor que no hi hagi rèpliques i hi hagi una única còpia en un ambient centralitzat.**

Fals. Al tractar-se d'un ambient distribuït, la replicació pot ser bona pel sistema. Això és degut a la necessitat de que el sistema sigui escalable i, amb el gran

volum de dades generat, la replicació permet que el sistema escali per realitzar càrregues de dades des de diferents nodes. A més, en cas que fallés algun d'aquests nodes, al haver-hi replicació les dades seguirien sent accessibles.

**5. Si es considera que la millor solució és una base de dades NoSQL, aleshores el model òptim seria l'orientat a grafs.**

Fals. El model graf és útil quan les relacions entre les dades són molt importants, com podria ser el cas de les xarxes socials. Aplicar aquest model implicaria afegir una complexitat innecessària.

## Citacions

Apunts de l'assignatura Architectures de bases de dades no tradicionals de La universitat Oberta de Catalunya, curs 2023-2024.

[1] Introducción a NoSQL, Historia y motivación. Jordi Conesa i Caralt Pere Urbón Bayes, M. Elena Rodríguez González. B1\_T1\_IntroduccionNoSQL-1: Pàgina 9, primer punt.

[2] Introducción a NoSQL, Historia y motivación. Jordi Conesa i Caralt Pere Urbón Bayes, M. Elena Rodríguez González. B1\_T1\_IntroduccionNoSQL-1: Pagina 9 segon punt.

[3] La persistencia poliglota, La mejor herramienta para cada problema. Àlex Bartrolí Muñoz, M. Elena Rodríguez González, Jordi Conesa i Caralt. B1\_T2\_PersistenciaPoliglota: Pàgina 11.

[4] B2\_T2: La persistencia poliglota, La mejor herramienta para cada problema. Àlex Bartrolí Muñoz, M. Elena Rodríguez González, Jordi Conesa i Caralt. B1\_T2\_PersistenciaPoliglota: Pàgina 11.

[5] B1\_T2: La persistencia poliglota, La mejor herramienta para cada problema. Àlex Bartrolí Muñoz, M. Elena Rodríguez González, Jordi Conesa i Caralt. B1\_T2\_PersistenciaPoliglota: Pàgina 11.

[6] Modelo en grafo. Jordi Conesa i Caralt, M. Elena Rodríguez González. B2\_T4\_1\_ModelosEnGrafo: Pàgina 10.

[7] Modelo en grafo. Jordi Conesa i Caralt, M. Elena Rodríguez González. B2\_T4\_1\_ModelosEnGrafo: Pàgina 19.

[8] Modelos de agregación. Motivación. M. Elena Rodríguez González, Jordi Conesa i Caralt, Pere Urbón Bayes. B2\_T3\_1\_ModelosAgregacionMotivacion. Pàgina 7

[9] Modelos de agregación Características generales. M. Elena Rodríguez González, Jordi Conesa i Caralt, Pere Urbón. BayesB2\_T3\_2\_ModelosAgregacionCaracteristicas. Pàgina 6

[10]: Modelos de agregación Tipos. Jordi Conesa i Caralt, M. Elena Rodríguez González, Pere Urbón Bayes. B2\_T3\_3\_ModelosAgregacionTipos. Pàgina 4

[11]: B1\_T2: La persistencia poliglota, La mejor herramienta para cada problema. Àlex Bartrolí Muñoz, M. Elena Rodríguez González, Jordi Conesa i Caralt. B1\_T2\_PersistenciaPoliglota: Figura pàgina 15.

[12]: Modelo en grafo. Jordi Conesa i Caralt, M. Elena Rodríguez González.  
B2\_T4\_1\_ModelosEnGrafo. Figura pàgina 8

[13] Modelos de agregación. Motivación. M. Elena Rodríguez González, Jordi Conesa i Caralt, Pere Urbón Bayes. B2\_T3\_1\_ModelosAgregacionMotivacion.  
Pàgina 5

[14]: Modelos de agregación Tipos. Jordi Conesa i Caralt, M. Elena Rodríguez González, Pere Urbón Bayes. B2\_T3\_3\_ModelosAgregacionTipos. Pàgina 4.