

EXAMEN DE CBDE

Instruccions: Respon cada pregunta i subapartat a l'espai habilitat. L'examen dura 1h 30'.

PREGUNTA 1. [3p] (Temps de referència: 15')

Considera la següent porció de la taula T emmagatzemada a HBase i que fa referència a totes les dades de la instància que té per clau '220101'.

F_1

"220101" + TS ₁ , NOM: "OSCAR"
"220101" + TS ₂ , COGNOM: "AMETLLER"
"220101" + TS ₃ , EDAT: "30"
"220101" + TS ₄ , NOM: "ANDREU"
"220101" + TS ₅ , COGNOM: "BUENAFUENTE"
"220101" + TS ₆ , USERID: "33"

F_2

"220101" + TS ₇ , SOU: "55000"
"220101" + TS ₈ , DPT: "DATA SCIENCE"
"220101" + TS ₉ , DPT: "DATA ENG."
"220101" + TS ₁₀ , SOU: "75000"

Les claus estan compostes pel valor de la clau especificat per l'usuari i el TS d'inserció (representats com TS_x). Podeu assumir que TS_x < TS_y sii x < y.

Aquesta taula té dues famílies, F1 i F2, i està configurada per emmagatzemar dues versions per atribut com a màxim.

1) Especifica quins atributs es guarden a cada família:

F1:

F2:

2) Suposa que es fa la següent operació: put 'T', '220101', 'F2: SOU', '90000'

Modifica els fitxers de la taula T per mostrar com quedarien després d'aquesta inserció, que es fa a TS₁₁. Fes la modificació directament a sobre de les taules.

3) La clau '220101' correspon a la data d'alta per primer cop al sistema d'aquest usuari. Sabent això, emprar aquesta clau garanteix una bona distribució de les dades al llarg dels RegionServer d'HBase (suposa'n 100)? Justifica la teva resposta.

Data: 11/01/2022. Nom i Cognoms:

PREGUNTA 2. [3p] (Temps de referència: 20')

Considera la següent taula que guarda un *log* d'accions a una web de compres online (p.e., *Amazon*). La taula emmagatzema l'usuari, el moment, l'acció feta i sobre quin element l'ha feta.

UserID	Timestamp	Action	Element
Jmercade	010120210930	Click	Menu_Products
Jmercade	010120210931	Free_Text	Search_Bar
Jmercade	010120210932	Click	Product_Item
Jmercade	010120210933	Click	Product_Description
Jmercade	010120210934	Click	Buy_Product
Jmercade	010120210935	Exit	NULL
Acanyelles	010120210936	Click	Contact_Page
Ogarcia	010120210936	Click	Menu_Products
Ogarcia	010120210938	Back_Button	NULL
Acanyelles	010120210944	Free_Text	Contact_Text
Acanyelles	010120210950	Click	Send_Contact_Info
Acanyelles	010120210951	Exit	NULL
Ogarcia	010120210952	Click	Menu_About
Jmercade	010120210955	Exit	NULL
Ogarcia	010120210956	Exit	NULL

1) Suposa que s'han creat totes les estructures físiques amb RLE, diccionari i ERI per tots els atributs de la taula. La següent afirmació, és certa?

“No hi ha cap atribut del qual les estructures físiques creades a la base de dades columnar ocupin MÉS espai que el que ocuparia aquest atribut a nivell físic en una taula relacional tradicional”.

Nota: Per la base de dades relacional tradicional pots assumir el *heap file* com a estructura física de referència.

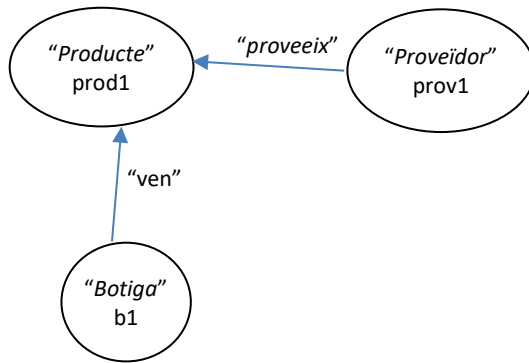
Justifica la teva resposta:

Data: 11/01/2022. Nom i Cognoms:

2) Dibuixa les estructures físiques que es crearien per emmagatzemar aquesta taula en una base de dades columnar emprant *Run-Length Encoding (RLE)*, diccionaris i *Ending Row Indexing (ERI)*. Només cal dibuixar la solució per l'atribut *userID*, però heu de dibuixar i identificar clarament les estructures finals (si dibuixes cap estructura temporal, distingeix-la de la resta).

PREGUNTA 3. [4p] (Temps de referència: 45')

Una agrupació de botigues gestiona els seus productes i proveïdors mitjançant una base de dades Neo4j com la de la figura (entre cometes l'etiqueta del node / aresta. Als nodes, a més, es mostra el valor de l'atribut *codi*):



- Els nodes de tipus **producte** tenen els següents atributs: **codi, nom, descripció**
- Els nodes de tipus **botiga** tenen els següents atributs: **codi, nom, telèfon, ciutat**
- Els nodes de tipus **proveïdor** tenen els següents atributs: **codi, nom, email**
- Les arestes de tipus **ven** tenen els següents atributs: **preu_venda, stock**
- Les arestes de tipus **proveeix** tenen els següents atributs: **preu_cost**

L'empresa vol modificar el graf actual per afegir informació sobre els clients i els productes que aquests han comprat. En concret, per cada client es vol saber el seu nom, el seu email, a quina ciutat viu, quins productes ha comprat, i a quina botiga. **De forma rellevant, el model proposat ha de permetre respondre la pregunta de l'apartat 2.**

1. Dibuixeu la vostra proposta de graf per a poder obtenir tota aquesta informació. Dibuixeu el vostre graf (seguint la mateixa notació que a la figura de dalt) a l'espai habilitat en aquesta pàgina. Si heu fet cap assumpció extra, l'heu d'explicitar a sota del vostre graf.

Data: 11/01/2022. Nom i Cognoms:

2. Feu en Cypher la següent consulta: *donat un client, retornar el preu total que ha pagat per tots els productes que ha comprat.*