# **Bases de Dades Analítiques**

## PAC 3: Complements de SQL

El supermercat a l'engròs *Lingo* s'ha posat de nou en contacte amb nosaltres perquè implementeu els requisits que ens han proposat.

Per a la implementació d'aquesta PAC, heu de crear una base de dades nova denominada **Lingo\_pec3** i executar el script adjunt **BBDD\_Sales (DDL).sql**. El segon script proporcionat (**BBDD\_Sales (dades).sql**) us donarà un conjunt de dades que us permetrà implementar els components requerits en els diferents apartats d'aquesta PAC. **NOTA**: les dades proporcionades no es poden modificar i els exercicis han de realitzar-se sobre la base d'aquestes dades.

Consideracions per al lliurament i realització de la PAC:

* Tot el que es demana en aquesta PAC està explicat en els blocs didàctics 1, 2 i 3 (tret que es tracti d'un exercici de recerca, l'enunciat de la qual l'especificarà). No fa falta que avanceu estudi de material d'altres blocs didàctics per a la realització d'aquesta PAC.
* En aquesta PAC treballarem el mòdul de complements de SQL. Heu d'assegurar una correcta execució de tot el que es demana: consultes CTE, funcions analítiques, etc. a més d'assegurar-se que els valors proporcionats són els que realment es demanen. Es recomana que proveu de forma exhaustiva les consultes.
* Sigueu fidels a l'enunciat. Si alguna cosa no és clar, si us plau comenteu-ho en el fòrum o directament al correu del professor col·laborador.
* Es recomana utilitzar **pgAdmin** per a la implementació de tota la PAC. Existeix una altra alternativa que és ***psql*** (línia de comandos), però és preferible que utilitzeu pgAdmin ja que és una interfície gràfica que us permetrà editar i crear SQL (així com mostrar els resultats) de forma més senzilla que psql.
* Tal com s'indica en l'enunciat, segons aplicació, la resposta als exercicis ha de lliurar-se en un fitxer *.sql* diferent, amb el nom corresponent. S'avaluarà el codi lliurat en aquests fitxers *.sql* i NO el codi que aparegui en els pantallazos del document adjunt.
* Les captures de pantalla dels exercicis (i explicacions pertinents) han de proporcionar-se en un document a part (podeu utilitzar l'enunciat de la PAC com a plantilla, **indiqueu el vostre nom en el document**, si us plau).
* S'ha de realitzar el lliurament de tots els fitxers de la PAC (tant els fitxers .*sql* com el document amb explicacions i captures de pantalla) en un fitxer comprimit .*zip*.

Consideracions per a l'avaluació de l'exercici:

* Es tindrà molt en compte l'ús de la convenció de noms, així com l'aplicació de les bones pràctiques de codificació en SQL, de consultes i de programació de procediments i disparadores. És a dir: codi amb sagnat, ús de clàusules SQL de forma correcta, comentaris, capçaleres en el procediment, etc.
* Els *scripts* proporcionats per l'estudiant amb les solucions dels exercicis han d'executar-se correctament sense cap error. Reviseu que les sentències estiguin ben formades.
* Les sentències SQL proporcionades en els scripts han de ser creades de forma manual i no mitjançant assistents que PostgreSQL/pgAdmin puguin proporcionar. Sentències SQL generades mitjançant eines penalitzaran l'exercici.
* Les sentències SQL proporcionades en els exercicis han de ser solament aquelles que demana l'enunciat i cap altra més. Qualsevol sentència afegida a majors, si està mal o provoca que el script no s'executi correctament a l'hora de corregir-ho, penalitzarà l'exercici.

**EXERCICI 1 (20%)**

Es demana modificar l'estructura de la taula Sale\_dw.tb\_product\_dim per a gestionar en nostre data warehouse l'històric de canvis que succeeixen en els productes i implementar així, segons el model dimensional de Ralph Kimball , una dimensió de tipus 2 (Slowly Changing Dimension, Type 2). Per a això, es demanen afegir els següents camps:

* una clau subrogada (que serà la **clau primària**) denominada product\_key, seguint la convenció de noms especificada en el bloc didàctic 1. Aquesta clau subrogada serà de tipus sencer i auto generada, començant amb el valor 1.
* product\_code, per tant, deixarà de ser clau primària.
* dos camps data inici i data fi (start\_date i end\_date en anglès) de tipus data/hora amb 6 dígits fraccionals. Aquests dos camps no poden permetre valors nuls.
* una clau alternativa estarà formada per les columnes product\_code i start\_date.

A més dels canvis en aquesta taula, es requereixen canvis en el procediment emmagatzemat Sale\_dw.sp\_load\_product\_dim() per a gestionar els següents requisits:

* La càrrega de dades de la dimensió de productes ha d'emmagatzemar dades històriques. Això és, cada vegada que es detecti un canvi (segons les especificacions donades en la PAC2 i la implementació proporcionada en el fitxer de DDL), ha de comprovar-se si ja existeix el producte en la dimensió:
  + Si el codi de producte no existeix en la dimensió, llavors s'inserirà una nova fila amb les noves dades del producte, categories i subcategories, amb data d'inici = 1900-01-01 00.00:00 i data de fi = 4000-01-01 00.00:00. Nous productes han de ser inserits amb una nova clau subrogada.
  + Si el codi de producte ja existeix, llavors es realitzaran dues accions:
    - S'actualitzarà la data de fi de la fila activa (aquella fila que existeix en la dimensió la data de la qual de fi és 4000-01-01).
    - S'inserirà una nova fila amb data d'inici = data/hora de sistema i data de fi = 4000-01-01, i una nova clau subrogada.
* Les consultes sobre l'existència de productes han de realitzar-se a partir del codi de producte, que és la clau de negoci en la base de dades operacional.
* Es requereix un tractament de nuls utilitzant el següent criteri:
  + Per a columnes de tipus caràcter, s'ha d'utilitzar UNKWN per als codis i UNKNOWN per a les descripcions.
  + Per a columnes de tipus numèric, s'ha d'utilitzar un valor 0.

Es demana proporcionar les sentències SQL per a crear els components especificats segons els requisits anteriors. El resultat d'aquest EXERCICI ha de lliurar-se en un fitxer anomenat **pec3\_ej1.sql**. **NOTA**: per a la taula, es demana proporcionar en el script el DROP TABLE i CREATE TABLE amb la nova estructura de les taules.

Kimball, R.; Ross, M. (2013). The Data Warehouse Toolkit (Third Edition). John Wiley & Sons, Inc.

**EXERCICI 2 (20%)**

Es demana proporcionar les sentències SQL per a crear els components especificats. El resultat d'aquest exercici ha de lliurar-se en un fitxer anomenat **pec3\_ej2.sql**.

**tb\_order\_line\_fact**

Taula de fets per a guardar informació de les línies de comanda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Taula:** tb\_order\_line\_fact **Esquema:** Sale\_dw | |  |  |  |
| **Nom columna** | **Tipus de dades** | **Accepta Nuls** | **Clau primària** | **Clau forana** |
| order\_number | Cadena de 10 caràcters fixa | No | Si |  |
| order\_line\_number | Sencer | No | Si |  |
| order\_date\_key | Sencer | No | No | Referencia a date\_key a la Taula Sale\_dw.tb\_date\_dim |
| product\_key | Sencer | No | No | Referencia a product\_key a la Taula Sale\_dw.tb\_product\_dim |
| quantity | Sencer | Si | No |  |
| unit\_price | Nombre de 12 dígits i 2 decimals. | Si | No |  |
| value\_amt | Nombre de 18 dígits i 2 decimals | Si | No |  |
| created\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | Si | No |  |
| updated\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sens zona | Si | No |  |

**sp\_load\_order\_line\_fact**

Procediment emmagatzemat que s'encarrega de realitzar la càrrega de dades de productes en la Taula tb\_order\_line\_fact. El funcionament d'aquest procediment és el següent:

* El procediment ha de realitzar un tipus de **càrrega total amb esborrat**. Per a això, el procediment ha de realitzar el següent:
  + Primer, s'han d'eliminar les dades de la Taula de fets existents.
  + A continuació, s'ha de realitzar la càrrega de les línies de comanda capturades en la Taula \_cdc. Per a això, s'han de seleccionar, utilitzant les taules tb\_order\_cdc i tb\_order\_line\_cdc de l'esquema Sale\_cdc, **aquelles línies de comanda que no estiguin eliminades.**
* En la Taula de fets de línies de comanda, product\_key referència a product\_key en la dimensió de productes del EXERCICI 3, i value\_amt és el càlcul obtingut a partir de unit\_price \* quantity.
* En la Taula de fets de línies de comanda, order\_date\_key ha de ser un sencer, que el seu serà YYYYMMDD on YYYY és l'any, MM és el mes i DD és el dia, i ha de referenciar date\_key a la dimensió de dates. Per exemple, 20170201 és l'1 Febrer 2017.
* En realitzar la càrrega de línies de comanda, pel fet que la dimensió de productes gestiona històric de dades (tal com s'ha implementat en el EXERCICI 1), cada línia de producte ha d'estar associada al producte que està actiu (és a dir, aquell producte la data del qual de fi = 4000-01-01 00.00:00).
* Heu d'assegurar-vos que les dimensions de producte i dates estan totes dues carregades abans de realitzar la càrrega dels fets.
* **El procediment ha de parar-se en cas de trobar-se amb una excepció**. En aquests casos, **el procediment ha d'assegurar que les dades anteriors a l'esborrat continuïn existint en la taula**. Amb això garantim que en la Taula de fets existeixin sempre dades perquè els usuaris puguin consumir-los.
* Una vegada que les dades anteriors han estat recuperats, el procediment ha de llançar una excepció indicant l'error que s'ha produït i notificant que les dades anteriors estan ja disponibles.
* El procediment emmagatzemat s'ha de crear en l'esquema Sale\_dw.
* Si ho desitgeu, podeu fer ús de components SQL vists en els blocs didàctics fins a aquest punt (per exemple, taules temporals, vistes, etc.). **Qualsevol component SQL que faciliti el procés de càrrega ha de crear-se en l'esquema** Sale\_stg**, i la definició d'aquests components ha de proporcionar-se com a part del lliurament**.

**EXERCICI 3 (25%)**

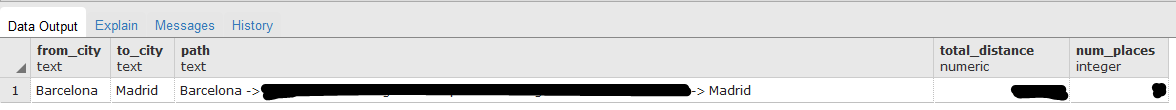
En la nostra base de dades de Surt, s'ha afegit una Taula de ciutats (Sale.tb\_city), que emmagatzema les ciutats on volem distribuir productes, i les coordenades latitud/longitud de cadascuna d'aquestes ciutats.

Com som experts en SQL, ens demanen que realitzem una consulta que ens proporcioni la ruta més curta possible des de **Barcelona** (com a ciutat origen) a **Madrid** (ciutat destino) assegurant-se que es viatja per totes les ciutats que existeixen en la base de dades una única vegada, sense repetir cap ciutat, proporcionant en el resultat de la consulta final:

* la ciutat origen (Barcelona)
* la ciutat destino (Madrid)
* el camí complet (de ciutat en ciutat, separades per “ -> “)
* la distància total recorreguda
* el Nom de ciutats recorregudes

Aquesta consulta ha de reutilitzar-se, de manera que puguem canviar fàcilment les ciutats origen i destinació. **NOTA**: no és necessari realitzar un procediment/funció, però sí que heu de construir la consulta de manera que es permeti canviar les ciutats d'origen i destinació.

Com a exemple, es mostren les dades demanades utilitzant les ciutats origen i destinació indicades respectivament.



Es demana que es lliuri aquesta consulta i una captura de pantalla dels resultats proporcionats. La definició d'aquesta consulta ha de lliurar-se en un fitxer anomenat **pec3\_ej3.sql.**

**NOTES:**

* Aquest exercici pot requerir, a més de conèixer i saber implementar consultes CTE recursivas que s'han vist en el bloc didàctic 3, investigar l'ús d'expressions i/o funcions PostgreSQL no vistes en els materials, i que podreu consultar en la documentació en línia de PostgreSQL.
* Per a calcular la distància entre dues ciutats, es proporciona la funció Sale.distance\_km, que accepta els paràmetres de latitud i longitud de les ciutats origen i destinació.

**EXERCICI 4 (10%)**

Es demana proporcionar les següents consultes en SQL mitjançant la utilització de funcions analítiques o *window functions* vistes en el bloc didàctic 3. S'ha d'utilitzar la base de dades de Sali (esquema Sale) per a solucionar el exercici. El resultat d'aquest exercici ha de lliurar-se en un fitxer anomenat **pec3\_ej4.sql.**

1. Obtenir, per a les comandes existents en la base de dades, la data del primer dia de la setmana a partir de la data de comanda (order\_date), la suma total de vendes (calculada com la quantitat multiplicada pel preu unitari) per a cada setmana, la mitjana acumulativa de les vendes (arrodonida a 2 dígits decimals) i el percentatge de creixement de les vendes de setmana en setmana, ordenat per data de comanda ascendentment. Assumiu que el primer dia de la setmana és Dilluns (i no Diumenge, com sol succeir en alguns països)..

**EXERCICI 5 (25%)**

El Servei de Salut de Catalunya disposa d'una base de dades per a registrar tota la informació sobre les visites mèdiques de pacients, els metges que treballen i les receptes que aquests dispensen.

Les relacions amb les quals es treballen són les següents:

**MEDICO** **(num\_med, nom\_med, especialidad, direccion, ciudad, telefono, sueldo, DNI)**

Informació sobre els metges. D'aquests, es guarda el Nom de metge (*num\_med*, que és clau primària), el nom (*nom\_med*), la seva especialitat (*especialidad*), l'adreça (*direccion*) i ciutat on viu (*ciudad*), telèfon (*telefono*), sou (*sueldo*) i Nom d'identificador nacional (*dni*).

El nom de metge, la seva especialitat, la ciutat on viu, el seu sou i el seu Nom d'identificador nacional mai poden ser nuls. A més, el seu sou sempre ha de ser superior a zero. El número d'identificador nacional és únic per a cada metge.

**PACIENTE (num\_pac, nom\_pac, direccion, ciudad, DNI)**

Informació sobre els pacients. D'aquests, es guarda el Nom de pacient (*num\_med*, que és clau primària), el nom (*nom\_pac*), la direcció (*direccion*), ciutat (*ciudad*) i Nom d'identificador nacional (*dni*).

El nom de pacient, la ciutat on viu i el seu Nom d'identificador nacional mai poden ser nuls. El Nom d'identificador nacional és únic per a cada pacient.

**VISITA (num\_med, num\_pac, fecha\_visita, importe)**

Informació sobre les visites que els pacients realitzen als metges. De cada visita, es guarda la següent informació: el número de metge (*num\_med*), el número de pacient (*num\_pac*), data de visita (*fecha\_visita*) i import (*importe*).

El conjunt d'atributs Nom de metge, Nom de pacient i data de visita formen la clau primària de la visita. A més, l'import de pagament de la visita no pot ser nul. Pot haver-hi pacients que no hagin realitzat cap visita, i metges que no hagin atès a pacients.

{num\_med} és clau forana de MEDICO i {num\_pac} és clau forana de PACIENTE.

**RECETA (num\_rec, medicamento, num\_med, num\_pac, fecha\_visita)**

Informació sobre les receptes que s'han dispensat els metges en les visites que han tingut amb els pacients. D'aquestes, es guarda el Nom de recepta (*num\_rec*, que és clau primària), el nom del medicament que s'ha dispensat (*medicamento*), el número de metge (*num\_med*), el número de pacient (*num\_pac*) i la data de visita (*fecha\_visita*).

El medicament mai pot ser nul. Pot haver-hi visites en les quals no s'hagin dispensat receptes (per exemple, quan el pacient solament fa revisions).

{num\_med, num\_pac, data\_visita} és clau forana de VISITA.

Donada aquesta base de dades i les següents transaccions (T1 i T2) que volen executar les següents peticions:

|  |
| --- |
| **T1** |
| UPDATE medico SET especialidad = 'Pediatria' WHERE num\_med = 236 |
| SELECT \* FROM medico WHERE num\_med = 236 |
| UPDATE medico SET sueldo = sueldo\*1.10 WHERE num\_med = 236 |
| COMMIT |

|  |
| --- |
| **T2** |
| SELECT \* FROM medico WHERE num\_med = 236 |
| COMMIT |

|  |
| --- |
| **T3** |
| INSERT INTO paciente  VALUES (123, 'Albert Pla', 'Rogent 10', 'Barcelona', '46789999') |
| SELECT \* FROM paciente |
| COMMIT |

Suposant que el contingut de les taules MEDICO i PACIENT és el que es mostra a continuació (les claus primàries de cada Taula es mostren subratllades):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **MEDICO** | | | | | | | |
| num\_med | nom\_med | especialidad | direccion | ciudad | telefono | sueldo | DNI |
| 236 | 'Pere Ba' | 'M. Interna' | 'Pi 23' | 'Lleida' | '678999000' | 3000 | '23456777' |
| 558 | 'Josema' | 'Pediatria' | 'Born 1' | 'Barcelona' | '938897898' | 2500 | '44089768' |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PACIENTE** | | | | |
| num\_pac | nom\_pac | direccion | ciudad | DNI |
| 989 | 'Manel Pi' | 'Or 23' | 'Barcelona' | '44044041' |

Suposant que el SGBD **no aplica cap mecanisme de control de concurrència**, responeu breument a les següents preguntes:

1. Proposeu, si és possible, un horari que incorpori totes les sentències SQL de T1 i de T2, i que contingui una (i només una) interferència. Heu de dir el nom de la interferència que s'està produint i el nivell mínim d'aïllament necessari de SQL per a evitar-la.
2. Proposeu si és possible un horari sense interferències que incorpori totes les sentències SQL de T2 i T3.
3. Suposeu ara que tenim una transacció T4 que s'executa de forma concurrent amb la transacció T3. El contingut de les taules és el que mostra a l'enunciat (com si les transaccions T1, T2 i T3 no s'haguessin executat). Proposeu, si és possible, quines sentències SQL hauria d'executar T4 perquè es veiés implicada en una interferència de tipus **fantasma**.

|  |  |
| --- | --- |
| **T3** | **T4** |
|  | **SENTENCIA/S SQL** |
| INSERT INTO paciente  VALUES (123, 'Albert Pla', 'Rogent 10', 'Barcelona', '46789999') |  |
|  | **SENTENCIA/S SQL** |
|  | COMMIT |
| SELECT \* FROM paciente |  |
| COMMIT |  |

Criteris de valoració

En l'enunciat s'indica el pes/valoració de cada exercici.

Per a aconseguir la puntuació màxima en els exercicis, és necessari explicar amb claredat la solució que es proposa.

Format i Data d’entrega

Heu d'enviar la PAC a la bústia de Lliurament i registre de EC disponible a l'aula (apartat Avaluació). El format de l'arxiu que conté la vostra solució pot ser .pdf, .doc i .docx. Per a altres opcions, per favor, contactar prèviament amb el vostre consultor. El Nom del fitxer ha de contenir el codi de l'assignatura, el vostre cognom i vostre Nom, així com el número d'activitat (PAC3).

La Data límit d’entrega de la PAC3 es el **12/05/2019**.

|  |
| --- |
| Nota: **Propietat intel·lectual**  En presentar una pràctica o PAC que faci ús de recursos aliens, s'ha de presentar juntament amb ella un document en què es detallin tots ells, especificant el Nom de cada recurs, el seu autor, el lloc on es va obtenir i el seu estatus legal: si l'obra està protegida per el copyright o s'acull a alguna altra llicència d'ús (Creative Commons, llicència GNU, GPL etc.). L'estudiant haurà d'assegurar-se que la llicència que sigui no impedeix específicament el seu ús en el marc de la pràctica o PAC. En cas de no trobar la informació corresponent haurà d'assumir que l'obra està protegida per el copyright.  Serà necessari, a més, adjuntar els fitxers originals quan les obres utilitzades siguin digitals, i el seu codi font, si així correspon. |