# **Bases de Dades Analítiques**

## PAC 2: El llenguatge SQL: procediments i disparadors

El supermercat a l’engròs Lingo, ara que disposa de la base de dades que hem construït a la UOC com a part de l'assignatura **Bases de dades per a Data Warehouse**, ha obtingut el pressuposto necessari per a implementar una base de dades que els hi serveixi per a realitzar anàlisi de dades i informes. De nou, s'han posat en contacte amb nosaltres perquè implementem els requisits que ens han proposat.

Per a la implementació d'aquesta PAC, heu de crear una base de dades nova denominada **lingo\_pec2** 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 i 2 (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 procediments/funcionis i disparadors. En tractar-se d'objectes més complexos, heu d'assegurar una correcta execució d'aquests. Es recomana que proveu de forma exhaustiva els components programats i us creeu els vostres casos de prova utilitzant la base de dades proporcionada. També treballarem amb més d'un esquema. Sigueu ordenats i assegureu que els objectes es creen en l'esquema adequat.
* 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 comandes), 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 les captures de pantalla del document adjunt. Les captures de pantalla dels exercicis (i explicacions pertinents) han de proporcionar-se en un document apart (podeu fer servir l’enunciat de la PEC como a plantilla, **indiqueu el vostre nom al 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 disparadors. 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 als 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 als exercicis han de ser només aquelles que demana l'enunciat i cap altra més. Qualsevol sentència afegida, si està mal o provoca que el script no s'executi correctament a l'hora de corregir-ho, penalitzarà l'exercici.

**EXERCICI 1 (20%)**

El supermercat Lingo ens comenta que hem de crear unes estructures, en un esquema diferent on estan les taules de vendes, per a emmagatzemar informació d'auditoria dels canvis produïts en la base de dades de vendes.

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 **pec2\_ej1.sql**.

**sale\_cdc**

Nom de l'esquema de base de dades on s'han de crear els components especificats a continuació. L'esquema ha de crear-se dins de la base de dades **lingo\_pec2** especificada a l'inici d'aquest enunciat.

**tb\_client\_cdc**

Taula per emmagatzemar informació sobre els canvis produïts en la taula de clients. Les columnes marcades en vermell són les noves columnes que s'han afegit a la definició de la taula tb\_client per a crear aquesta nova taula.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taula:** tb\_client\_cdc **Esquema:** sale\_cdc | |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** |
| client\_code | Cadena de 5 caràcters fixa | No | Si |
| client\_name | Cadena de 40 caràcters variable | No | No |
| address | Cadena de 120 caràcters variable | Si | No |
| city | Cadena de 25 caràcters variable | Si | No |
| country | Cadena de 50 caràcters variable | No | No |
| contact\_email | Cadena de 100 caràcters variable | Si | No |
| phone | Cadena de 15 caràcters variable | Si | No |
| parent\_client\_code | Cadena de 5 caràcters fixa | Si | No |
| created\_by\_user | Cadena de 10 caràcters variable | No | No |
| created\_date | Data (sense temps) | Si | No |
| updated\_date | Data (sense temps) | Si | No |
| operation | Cadena de 1 caràcter fixa | No | No |
| user\_id | Cadena de 10 caràcters variable | No | No |
| operation\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | No | No |

**tb\_category\_cdc**

Taula per emmagatzemar informació sobre els canvis produïts a la taula de categories de producte. Les columnes marcades en vermell són les noves columnes que s'han afegit a la definició de la taula tb\_category per a crear aquesta nova taula.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taula:** tb\_category\_cdc **Esquema:** sale\_cdc | |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** |
| category\_code | Cadena de 5 caràcters fixa | No | Si |
| category\_name | Cadena de 60 caràcters variable | No | No |
| created\_by\_user | Cadena de 10 caràcters variable | No | No |
| created\_date | Data (sense temps) | Si | No |
| updated\_date | Data (sense temps) | Si | No |
| operation | Cadena de 1 caràcter fixa | No | No |
| user\_id | Cadena de 10 caràcters variable | No | No |
| operation\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | No | No |

**tb\_subcategory\_cdc**

Taula per a emmagatzemar informació sobre els canvis produïts a la taula de subcategories de producte. Les columnes marcades en vermell són les noves columnes que s'han afegit a la definició de la taula tb\_subcategory per a crear aquesta nova taula.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taula:** tb\_subcategory\_cdc **Esquema:** sale\_cdc | |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** |
| subcategory\_code | Cadena de 5 caràcters fixa | No | Si |
| category\_code | Cadena de 5 caràcters fixa | Si | No |
| subcategory\_name | Cadena de 60 caràcters variable | No | No |
| created\_by\_user | Cadena de 10 caràcters variable | No | No |
| created\_date | Data (sense temps) | Si | No |
| updated\_date | Data (sense temps) | Si | No |
| operation | Cadena de 1 caràcter fixa | No | No |
| user\_id | Cadena de 10 caràcters variable | No | No |
| operation\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | No | No |

**tb\_product\_cdc**

Taula per a emmagatzemar informació sobre els canvis produïts a la taula de productes. Les columnes marcades en vermell són les noves columnes que s'han afegit a la definició de la taula tb\_product per a crear aquesta nova taula.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taula:** tb\_product\_cdc **Esquema:** sale\_cdc | |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** |
| product\_code | Cadena de 5 caràcters fixa | No | Si |
| subcategory\_code | Cadena de 5 caràcters fixa | Si | No |
| product\_name | Cadena de 60 caràcters variable | No | No |
| price | Número de 12 dígits y 2 decimals. | No | No |
| created\_by\_user | Cadena de 10 caràcters variable | No | No |
| created\_date | Data (sense temps) | Si | No |
| updated\_date | Data (sense temps) | Si | No |
| operation | Cadena de 1 caràcter fixa | No | No |
| user\_id | Cadena de 10 caràcters variable | No | No |
| operation\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | No | No |

**tb\_order\_cdc**

Taula per a emmagatzemar informació sobre els canvis produïts a la taula de comandes. Les columnes marcades en vermell són les noves columnes que s'han afegit a la definició de la taula tb\_order per a crear aquesta nova taula.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taula:** tb\_order\_cdc **Esquema:** sale\_cdc | |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** |
| order\_number | Cadena de 10 caràcters fixa | No | Si |
| client\_code | Cadena de 5 caràcters fixa | No | No |
| order\_date | Data (sense temps) | No | No |
| delivery\_date | Data (sense temps) | Si | No |
| reception\_date | Data (sense temps) | Si | No |
| created\_by\_user | Cadena de 10 caràcters variable | No | No |
| created\_date | Data (sense temps) | Si | No |
| updated\_date | Data (sense temps) | Si | No |
| operation | Cadena de 1 caràcter fixa | No | No |
| user\_id | Cadena de 10 caràcters variable | No | No |
| operation\_timestamp | Data/hora con precisión de 6 dígits fraccionals, sense zona | No | No |

**tb\_order\_line\_cdc**

Taula per a emmagatzemar informació sobre els canvis produïts a la taula de línies de comandes. Les columnes marcades en vermell són les noves columnes que s'han afegit a la definició de la taula tb\_order\_line per a crear aquesta nova taula.

|  |  |  |  |
| --- | --- | --- | --- |
| **Taula:** tb\_order\_line\_cdc **Esquema:** sale\_cdc | |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** |
| order\_number | Cadena de 10 caràcters fixa | No | Si |
| order\_line\_number | Entero | No | Si |
| product\_code | Cadena de 5 caràcters fixa | No | No |
| quantity | Entero | No | No |
| unit\_price | Número de 12 dígits y 2 decimals. | No | No |
| created\_by\_user | Cadena de 10 caràcters variable | No | No |
| created\_date | Data (sense temps) | Si | No |
| updated\_date | Data (sense temps) | Si | No |
| operation | Cadena de 1 caràcter fixa | No | No |
| user\_id | Cadena de 10 caràcters variable | No | No |
| operation\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | No | No |

**EXERCICI 2 (35%)**

Per a fer ús de les taules \_cdc que hem definit en l'exercici anterior, necessitem capturar els canvis produïts en les taules de l'esquema sale i emmagatzemar el canvi en cadascuna de les taules \_cdc. Per a això, necessitem definir un disparador en cadascuna de les taules de l'esquema de sale que realitzi les següents accions:

* Capturar totes les operacions efectuades (INSERT, UPDATE i DELETE) en la taula \_cdc corresponent, en manera **CONDENSADA**: cada operació sobre una fila d'una taula de l'esquema de vendes **inserirà** una fila nova en la taula \_cdc amb les dades corresponents **si aquesta no existeix**, o **s'actualitzarà** la fila en la taula de \_cdc **en cas que aquesta sí que existeix**i. És a dir, la taula \_cdc haurà de reflectir l'últim estat de la fila respecte a la base de dades de vendes.
* Capturar els canvis a nivell de fila amb el canvi aplicat a la fila. Això és, s'emmagatzemarà cada fila canviada en el seu estat nou (no l'estat anterior al canvi, a excepció dels esborrats).
* A més de la informació guardada en vendes, s'ha d'emmagatzemar per a cada fila l'operació realitzada (I, O, D per a INSERT, UPDATE i DELETE respectivament), la data i hora de l'operació, i l'usuari que ha realitzat aquesta operació, segons la definició de les taules \_cdc definides anteriorment.
* No és necessari que captureu les excepcions.
* Tots els objectes necessaris per a la implementació d'aquesta funcionalitat han de ser creats (si és possible sobre la base de la definició de l'objecte) en l'esquema sale\_cdc.

Es demana proporcionar, **per a cadascuna de les taules de surt, la definició dels components en llenguatge SQL que ens permeti generar i emmagatzemar la informació especificada**. La definició d'aquests components ha de lliurar-se en un fitxer anomenat **pec2\_ej2.sql.**

**Suggeriment**: proveu els procediments i disparadors per a assegurar que funcionen de forma correcta per a totes les taules. Per exemple, creeu un client, actualitzeu-lo i esborreu-lo.

**EXERCICI 3 (45%)**

Es demana que es creïn els següents components.

**surt\_dw**

Nom de l'esquema de base de dades on s'han de crear les taules de *data warehouse*. L'esquema ha de crear-se dins de la base de dades **lingo\_pec2** especificada en l'inici d'aquest enunciat.

**surt\_stg**

Nom de l'esquema de base de dades on s'han de crear taules temporals, d'errors i qualsevol estructura de suport per a la càrrega de dades en sale\_dw. L'esquema ha de crear-se dins de la base de dades **lingo\_pec2** especificada en l'inici d'aquest enunciat.

**tb\_product\_dim**

Taula per a emmagatzemar informació detallada sobre els productes i la seva jerarquia de categories. Cada fila a la taula representarà un producte.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Taula:** tb\_product\_dim **Esquema:** sale\_dw | |  |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** | **Comentaris** |
| product\_code | Cadena de 5 caràcters fixa | No | Si |  |  | |
| product\_name | Cadena de 50 caràcters variable | No | No |  |  |
| price | Número de 12 dígits y 2 decimals. | No | No |  |  |
| subcategory\_code | Cadena de 5 caràcters fixa | Si | No |  |  |
| subcategory\_name | Cadena de 50 caràcters variable | Si | No |  |  |
| category\_code | Cadena de 5 caràcters fixa | Si | No |  |  |
| category\_name | Cadena de 50 caràcters variable | Si | No |  |  |

**tb\_product\_err**

Taula per a emmagatzemar informació dels errors produïts a la carrega de dades de tb\_product\_dim.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Taula:** tb\_product\_err **Esquema:** sale\_stg | |  |  |  |
| **Nom columna** | **Tipus de dada** | **Accepta Nuls** | **Clau primària** | **Comentaris** |
| product\_code | Cadena de 5 caràcters fixa | No | Si |  |  | |
| error\_timestamp | Data/hora con precisió de 6 dígits fraccionals, sense zona | No | Si |  |  |
| error\_code | Text | Si | No | Codi de error generat |  |
| error\_message | Text | Si | No | Missatge generat por el error |  |

**sp\_load\_product\_dim**

Procediment emmagatzemat que s'encarrega de realitzar la càrrega de dades de productes a la taula tb\_product\_dim. El funcionament d'aquest procediment és el següent:

* El procediment ha de realitzar un tipus de **càrrega incremental**. Per això, s'han de seleccionar, utilitzant les taules tb\_category\_cdc, tb\_subcategory\_cdc i tb\_product\_cdc de l'esquema sale\_cdc, **els productes que hagin sofert un canvi, o bé els productes la categoria dels quals o subcategoria, hagi sofert un canvi.**
* Per a capturar aquells codis de productes de les taules \_cdc, s'ha de verificar que el canvi hagi succeït en els 10 últims minuts (comparant la data i hora del sistema amb la data i hora en què ha succeït l'operació).
* El procediment ha d'obtenir els detalls del producte (nom i preu), i el codi i nom de la subcategoria i categoria associades, **detalls que han d'obtenir-se estrictament de les taules \_cdc**. Aquest punt és molt important.
* El procediment ha de realitzar la inserció o actualització dels productes en la taula tb\_product\_dim i mai eliminar-los. En el cas que la fila a processar ja existeixi en la taula, aquesta ha d'actualitzar-se.
* **El procediment no ha de parar-se en cas de trobar-se amb una excepció**, és a dir, ha d'acabar de forma correcta en totes les execucions. És necessari capturar les següents excepcions i gestionar-les segons s'indica a continuació:
  + Si en inserir la fila aquesta ja existeix en la taula (violació de clau primària), s'intentarà actualitzar mitjançant un UPDATE, però no es llançarà cap excepció.
  + Si es produeix un error en la inserció (diferent al produït en inserir una fila amb clau primària existent) o en l'actualització, s'ha d'inserir un registre en la taula tb\_product\_err amb el codi de producte que ha causat l'error, la data i hora en què s'ha produït l'error, el codi d'error i el missatge d'error. Això ens permetrà analitzar per què ha fallat la càrrega i quin producte ha causat la fallada.
* Heu de tenir en compte que si es produeix un canvi en una categoria o subcategoria, s'han d'actualitzar aquells productes en la taula tb\_product\_dim que estiguin associats a aquestes categories o subcategories (encara que les dades del producte no s'hagin modificat). El procediment emmagatzemat ha de gestionar aquest tipus d'escenaris de forma correcta.
* El procediment emmagatzemat s'ha de crear a l'esquema sale\_stg.
* 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**.

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 **pec2\_ej3.sql.**

**Responeu també a la següent pregunta**: Quines conseqüències té el fet d'implementar aquest procediment de la forma que s'ha descrit? Expliqueu breument proposant un exemple, si ho considereu necessari.

Criteris de valoració

A 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 (PEC1).

La Data límit d’entrega de la PAC1 es el **14/04/2019**.

|  |
| --- |
| 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. |