# **Bases de Dades Analítiques**

## PAC 5: Altres BD per a entorns analítics: Magatzems de columnes i BD en memoria

En aquesta última pràctica de l'assignatura, aprofundirem en les principals característiques i utilitat de les bases de dades orientades a columnes. D'igual forma, veurem més detalladament les bases de dades in-memory així com la seva implementació en alguna eina comercial existent.

Per a completar aquesta pràctica teniu a la vostra disposició tot el material explicat en el bloc didàctic 5 així com el material disponible en el material d'aprenentatge associat a aquest bloc didàctic. Algun dels exercicis presentats són exercicis de recerca on és possible que necessiteu consultar informació addicional. Recordeu identificar la bibliografia addicional utilitzada en cadascun dels exercicis.

El primer exercici serà d'anàlisi i estarà orientat al fet que s'analitzi l'eina SAP HANA. A partir de la presentació sol·licitada s'intentarà valorar el vostre grau d'anàlisi, síntesi i capacitat de cerca d'informació.

El segon exercici estarà centrat en la lectura d'un article que analitza la situació de les bases de dades in-memory mostrant un ampli ventall de possibles bases de dades, casos d'ús i aplicacions possibles. L'objectiu d'aquest exercici és que us familiaritzeu amb les diferents tipologies existents de bases de dades in-memory.

Finalment, el tercer exercici us proposarà un exercici pràctic a partir d'una arquitectura de base de dades modelat en un magatzem de columnes. En aquest exercici heu de realitzar suposats a partir dels requeriments expressats en l'enunciat i realitzar propostes de solució per als plantejaments realitzats.

**EXERCICI 1** **(35%)**

Prepareu una presentació en PowerPoint sobre l'eina SAP HANA. La presentació ha d'exposar de forma clara i concisa la informació relacionada amb la matèria. Els següents punts han de ser exposats en la presentació:

* Breu introducció a SAP HANA
* Possibles tipus de configuració de la Base de dades
* Arquitectura de Base de dades
* Principals característiques
* Conclusions
* Bibliografia

La presentació ha de tenir una longitud màxima de 10 slides sense comptar la portada (en el cas que inclogueu una) i les slides de la bibliografia en la qual us heu basat.

Recordeu que en el material docent adjunt de la PEC5 podeu consultar diferents papers (articles en revistes) i llibres relacionats amb SAP HANA. Addicionalment, podeu utilitzar qualsevol bibliografia que estigui a la vostra disposició, incloent-la en la documentació de la presentació.

Per a accedir als recursos de biblioteca associats al material docent d'aquest bloc, heu de tenir en compte els següents aspectes:

1. Assegurar-vos que accediu amb perfil d'estudiant. La selecció de perfil es realitza en la pàgina principal de campus, una vegada hàgiu entrat. És un menú desplegable que està en el marge superior dret. Una vegada seleccionat aquest perfil, els vostres missatges aniran precedits per la lletra e (estudiant).
2. Biblioteca demanarà identificació (cognom i número de carnet UOC). En el cas que no tingueu clar quin és el vostre número de carnet o encara no el tingueu, podeu usar com a número vostre DNI (sense lletra) precedit pel número 20 (20NúmeroDNI sense lletra).

**EXERCICI 2** **(25%)**

Leer el artículo “In-memory big data management and processing. A survey” disponible en el material de aprendizaje asociado a este bloque didáctico.

El objetivo de este ejercicio es razonar sobre los siguientes conceptos:

* Motivos por los que surge esta nueva tecnología de base de datos
* Ventajas y puntos débiles que veis del uso de esta nueva tecnología

Adicionalmente, podéis utilizar cualquier bibliografía que este a vuestra disposición, incluyéndola en el documento de vuestra repuesta.

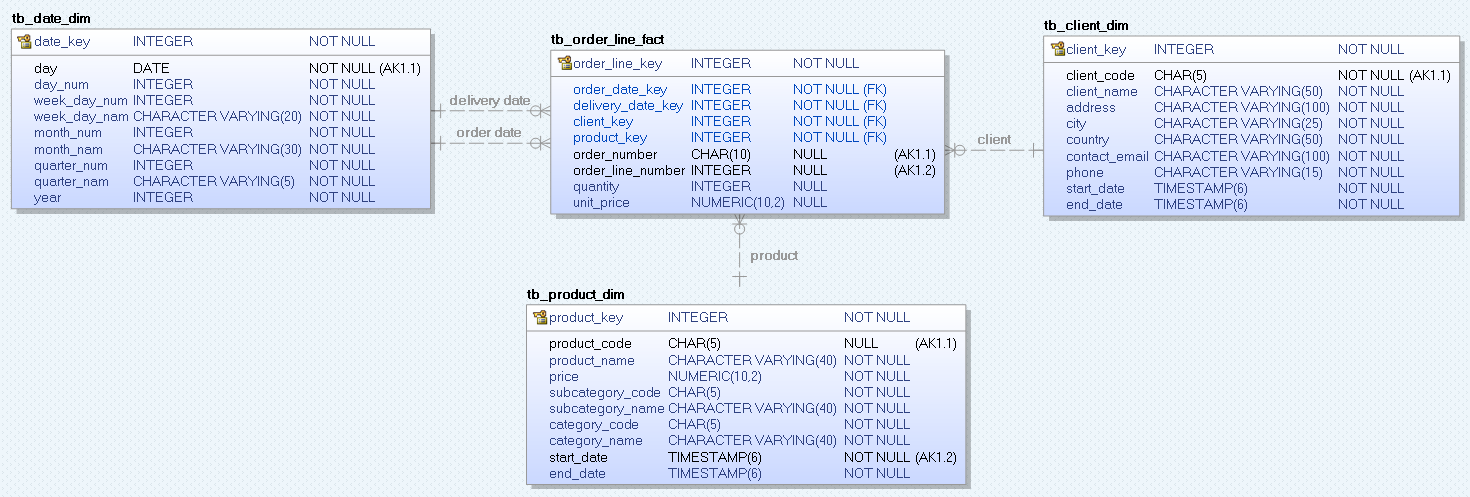
La extensión máxima del documento con vuestra respuesta no debe superar 2 páginas

Per a accedir als recursos de biblioteca associats al material docent d'aquest bloc, heu de tenir en compte els següents aspectes:

1. Assegurar-vos que accediu amb perfil d'estudiant. La selecció de perfil es realitza en la pàgina principal de campus, una vegada hàgiu entrat. És un menú desplegable que està en el marge superior dret. Una vegada seleccionat aquest perfil, els vostres missatges aniran precedits per la lletra e (estudiant).
2. Biblioteca demanarà identificació (cognom i número de carnet UOC). En el cas que no tingueu clar quin és el vostre número de carnet o encara no el tingueu, podeu usar com a número vostre DNI (sense lletra) precedit pel número 20 (20NúmeroDNI sense lletra).

**EXERCICI 3 (40%)**

Suposeu que tenim el següent model dimensional en nostre data warehouse. Aquest model es carrega a partir de la base de dades Surt que heu vist al llarg d'aquest semestre.



* tb\_date\_dim: taula de dimensió de dates. Cada fila representa un dia de l'any, amb els seus respectius detalls (per exemple, 1 Gener 2016). La clau primària és la clau subrogada date\_key i la clau alternativa és day.
* tb\_client\_dim: taula de dimensió de clients, que es carrega a partir de la taula sale.tb\_client. Vegeu que no s'ha modelat la jerarquia de clients per a fer el model més senzill. La clau primària és la clau subrogada client\_key i la clau alternativa és client\_code.
* tb\_product\_dim: taula de dimensió de productes que heu creat i que heu carregat a partir de la taula sale.tb\_product mitjançant el procediment emmagatzemat sp\_load\_product\_dim() en la PEC3.
* tb\_order\_line\_dim: taula de fets de línies de comanda. Es carrega a partir de les taules sale.tb\_order i sale.tb\_order\_line.

Contesteu als següents apartats i expliqueu cadascun d'ells breument, suposant que aquest model està implementat en un magatzem de columnes.

1) Sabem que la nostra dimensió de productes té un nombre de files elevat, concretament 10 milions de files (recordeu que aquesta dimensió manté l'històric de productes segons han canviat categories, subcategories, etc.).

Volem crear una projecció que contingui les columnes category\_name i subcategory\_name, i volem aplicar algorismes de compressió sobre totes dues columnes.

Es demana proporcionar quins algorismes de compressió poden donar-nos un bon resultat per a reduir l'espai d'emmagatzematge de totes dues columnes en la nova projecció i expliqueu breument el per què. Proporcionar diferents suposats si ho considereu oportú (no és necessari realitzar i proporcionar els càlculs d'espai de les noves columnes sobre la base dels algorismes que presenteu).

2) Suposant que els nostres usuaris realitzen les següents consultes:

* El 65% de les consultes mostra informació de surt per país del client i client, de forma anual sobre la base de la data de comanda (order\_dóna't).
* Un 15% de les consultes es mostra informació de tots els productes (nom, categoria, subcategoria, preu, data d'inici i data de fi), que s'utilitzarà per a importar a una base de dades de gestió de dades mestres.
* La resta de consultes que llancen els usuaris són consultes ad hoc (consultes no predefinides), és a dir, no sabem el tipus de consulta que s'espera sobre aquest model dimensional.

Definir les projeccions necessàries per a donar una solució de rendiment òptima al tipus de consultes que s'han especificat anteriorment. No heu de definir més de 3 projeccions en total (sense comptar les superproyecciones de cadascuna de les taules). Explicar breument la clau d'ordenació seleccionada i per a quines consultes es crea la projecció.

Per a resoldre aquest exercici heu d'usar el material en vídeo (i les anotacions associades) disponibles en el recurs d'aprenentatge associat a aquest bloc didàctic. Assegurar-vos, si us plau, de visualitzar els vídeos (i les seves anotacions) en l'ordre adequat. Aquest ordre està especificat a l'aula en la descripció del bloc didàctic 5, concretament en l'apartat de Presentació de continguts.

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).

És necessari lliurar un arxiu zip o rar que contingui:

* La presentació (.pdf, .ppt i .pptx) corresponent a l'exercici 1
* Un document (.pdf, .doc i .docx) amb les respostes als exercicis 2 i 3

El nom dels diferents arxius (incloent l'arxiu comprimit) ha de contenir el codi de l'assignatura, el vostre cognom i el vostre nom, així com el número d'activitat (PEC5).

La data límit per a lliurar la PAC és el **23/06/2018**.

|  |
| --- |
| 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. |