



## Enunciado de la PEC1

Diseño y uso de bases de datos analíticas (Universitat Oberta de Catalunya)

## PEC 1

### Presentación.

La primera práctica de evaluación continua consta de dos partes diferenciadas, una **teórica (40%)** y otra **práctica (60%)**.

La primera parte consiste en una serie de preguntas que tienen el objetivo de consolidar los conocimientos teóricos de los módulos **1. Introducción a las bases de datos analíticas** y **2. La construcción de la factoría de información corporativa** de la asignatura.

La segunda parte consiste en la toma de contacto con la máquina virtual proporcionada por la UOC (VDI) y el software que será necesario para realizar las prácticas de la asignatura, capacitando al estudiante para iniciar el resto de las actividades prácticas del curso.

### Objetivos y competencias

- Diferenciar claramente entre el almacén de datos y la base de datos operacional, identificando las principales diferencias.
- Conocer el contexto del almacén de datos, es decir, la Factoría de Información Corporativa, sus componentes y la función de cada uno de ellos.
- Validar el entorno de prácticas.

## Contenido.

### - Parte teórica.

Esta parte de la PEC está compuesta por 5 preguntas teóricas que tienen por objetivo comprobar la correcta comprensión del módulo 1 y 2 de la asignatura a partir de las respuestas del estudiante.

**Tiempo de lectura (estimado): 2h (módulo 1) + 5h (módulo 2)**

**Tiempo de resolución de preguntas (estimado): 5h**

### - Parte práctica.

Se realizará la configuración necesaria para acceder a las máquinas del entorno de prácticas (VDI). Cada estudiante desde cada escritorio personal validará tanto la conectividad con la base de datos como la ejecución de procesos ETL (Extracción, Transformación y Carga). Finalmente, también se realizará un ejercicio con Visual Studio y Microsoft SQL *Analysis Services* .

Esta parte consta de 4 ejercicios:

- Ejercicio 1. Configuración del entorno VDI
- Ejercicio 2. Validación de la BBDD.
- Ejercicio 3. Validación de la aplicación de ETL.
- Ejercicio 4. Microsoft SQL *Analysis Services*

Información relevante de credenciales:

a) Las credenciales para acceder a las máquinas VDI son las mismas que utilizáis para entrar en el campus de la UOC.

b) Las credenciales de acceso al SQL server para cada estudiante son:

Usuario: STUDENT\_usuariouoc

Contraseña: \$d3f4ult\$

En el primer login os solicitará cambiar el password.

**Tiempo de resolución de ejercicios prácticos (estimado): 6h**

**Tiempo de redacción de la guía paso a paso (estimado): 14h**

## **Criterios de evaluación.**

### **- Parte teórica (40%)**

Se valorarán todas las respuestas correctas por igual, 8% del total de la PEC

Todas las respuestas se deben justificar. No demostrar en la justificación la correcta comprensión teórica implica un 0% en dicha pregunta. Se valorarán positivamente todas las justificaciones/argumentaciones que no hagan referencia directa (o copia) a los recursos proporcionados por la Universidad. Las respuestas deben demostrar la asimilación de los contenidos por parte del estudiante.

### **- Parte práctica (60%)**

Se valorará de acuerdo con las evidencias de cumplimiento de los pasos indicados y aportados por el estudiante en el documento de solución, mediante capturas de pantalla y las explicaciones oportunas de las actividades realizadas, de acuerdo con los objetivos establecidos. Las capturas sin explicación no se valorarán como totalmente correctas.

#### **La nota final estará formada por:**

Pregunta 1 (8%) + Pregunta 2 (8%) + Pregunta 3 (8%) + Pregunta 4 (8%) + Pregunta 5 (8%) + Ejercicio 1 (5%) + Ejercicio 2 (20%) + Ejercicio 3 (25%) + Ejercicio 4 (10%)

## Formato y fecha de entrega

La entrega se realizará enviando un único mensaje al buzón de entrega de actividades del aula. Dicho mensaje llevará adjunto un único documento en formato word o pdf con la solución de la PEC. En el documento se debe indicar obligatoriamente el nombre completo del estudiante y los estudios que está cursando. El nombre del archivo debe ser la composición del nombre de usuario y “DW\_PEC1” (por ejemplo: si el nombre de usuario es “acabero”, el nombre del archivo debe ser “acabero\_DW\_PEC1.pdf” o “acabero\_DW\_PEC1.doc”).

Es responsabilidad única del estudiante asegurarse que entrega el documento que pretende en el lugar que la Universidad habilita con este objetivo, que es el registro EC (Evaluación Continua).

*La fecha máxima de entrega es el 28/10/2020 a las 23:59 h.*

## Formato y fecha de feedback o retorno pedagógico

El feedback o retorno pedagógico se realizará de manera individual a través del registro EC (Evaluación Continua). Además, se publicará en el foro asociado a la PEC una valoración general de los resultados donde se indicarán aquellos errores comunes u observaciones que se quieran dar con carácter general a los estudiantes.

*La solución se publicará el 05/11/2020*

*La fecha de calificación y retorno pedagógico es el 09/11/2020*

## Parte teórica (40%).

### Pregunta 1 (8%):

¿Cuáles son las diferencias existentes entre una base de datos operacional y un *Data Warehouse* desde un punto de vista de la estructura de datos? Justifica tu respuesta.

### Pregunta 2 (8%):

Enumera y describe brevemente las diferencias entre un *Data Mart* y un *Enterprise Data Warehouse*.

### Pregunta 3 (8%):

¿Qué enfoque en la construcción de la FIC consideras más adecuado? Argumenta la respuesta.

### Pregunta 4 (8%):

¿Cuáles son las principales ventajas y desventajas de disponer de un cubo OLAP? Enumera y describe brevemente.

### Pregunta 5 (8%):

Indica si las siguientes afirmaciones son verdaderas (V) o falsas (F). Justifica tu respuesta en aquellas afirmaciones que consideres falsas.

- El proceso de integración será el mismo independientemente de que el proceso de carga sea inicial o incremental. ( V o F )
- El responsable de datos o *Chief Data Officer* (CDO) debe ser capaz de implementar algoritmos de procesamiento de datos y definir modelos predictivos. ( V o F )
- Usando herramientas self-service BI varios analistas pueden obtener resultados diferentes a un mismo problema. ( V o F )
- Los datos semiestructurados son datos con una estructura de datos conocida, y se almacenan principalmente en bases de datos relacionales. ( V o F )

Se valorarán todas las justificaciones/argumentaciones que no hagan referencia directa (o copia) de los recursos proporcionados por la Universidad. Las respuestas deben demostrar la asimilación de los contenidos por parte del estudiante.

## Parte práctica (60%).

### Ejercicio 1 (5%): Configuración del entorno VDI

#### Objetivo:

Configurar la conexión al escritorio VDI que se utilizará durante todo el curso, tanto para realizar esta PEC (PEC1) como las dos prácticas siguientes (PRA1 y PRA2).

#### Desarrollo del ejercicio:

Utilizando los documentos: **DW\_Entorn\_Virtual\_VDI.pdf** (como guía de configuración y conexión) y **Importació de bases de dades, creació de vistes i cubs.pdf** (con indicaciones sobre cómo proceder per a crear cubos), el estudiante debe de llevar a cabo todo el proceso hasta obtener su escritorio VDI. Se recomienda leer atentamente dicho documento y seguir sus indicaciones.

El estudiante debe entregar, como solución a este ejercicio, una **guía paso a paso** de todo el proceso de configuración de la máquina accesible desde el escritorio de la VDI. La guía debe de incluir todas las explicaciones y capturas de pantalla que sean necesarias.

También se adjuntará una captura de pantalla del escritorio VDI y la información técnica disponible bajo el logo de la UOC (*HostName*, *UserName*, etc.), que encontraréis como fondo de escritorio.



## Ejercicio 2 (20%): Validación de la BBDD.

### Objetivo:

Toma de contacto con el Sistema Gestor de Bases de Datos facilitado con la máquina VDI (SQL Server 2017), configuración de una conexión con el servidor y creación de una base de datos simple con una tabla.

### Desarrollo del ejercicio:

#### 1) Conectar al servidor de bases de datos SQL Server

Utilizad la consola “Microsoft SQL Server Management Studio 2017 (SSMS)” instalada en la VDI para abrir una conexión con el servidor de base de datos y adjuntad las siguientes capturas de pantalla:

- a) Versión de SSMS
- b) Pantalla de conexión inicial de la consola SQL Server Management Studio (SSMS) donde se pueda observar la configuración utilizada.
  - *Server Type*
  - *Server name*
  - *Authentication*
  - *Login*
- c) Explorador de objetos del SSMS donde se puede ver el usuario de conexión.

#### 2) Configurar la base de datos DW\_DB\_XXX

XXX corresponde al nombre asignado de BBDD que habéis recibido por email.

Se debe configurar el parámetro *Recovery Model* de la BBDD como “simple” y desactivar la actualización automática de estadísticas. Adjuntad las capturas de pantalla que demuestren la correcta configuración.

#### 3) Crear una tabla con el nombre T\_PEC1

La tabla tendrá dos campos: campo1 y campo2 de la siguiente manera:

- **Campo1** será de tipo varchar(50) y permitirá valores nulos
- **Campo2** será de tipo int y no permitirá valores nulos.

Adjuntad las capturas de pantalla que demuestren la correcta creación de la tabla con sus campos.

## Ejercicio 3 (25%): Validación de la aplicación de ETL.

### Objetivo:

Toma de contacto con la aplicación de ETL suministrada en la máquina VDI (*Pentaho Data Integration*) y capacitar al estudiante para iniciar las actividades del curso

### Desarrollo del ejercicio:

Para familiarizarse con la herramienta Spoon que permite la transformación de datos, se pide cargar los datos de un archivo CSV (origen de datos) en una nueva tabla de la base de datos (destino de datos).

Describid paso a paso todas las acciones realizadas aportando las capturas de pantalla y scripts que sean necesarias. Los pasos para realizar son los siguientes:

- 1) Abrir el componente Spoon.
- 2) Capturar versión de PDI.
- 3) Crear un repositorio de tipo *file* con el nombre PEC1.
- 4) Crear una transformación y un *Job* básicos:

Anexo al enunciado de la PEC1, encontraréis un archivo **datos.csv**.

Se debe copiar el archivo de origen de datos a una de las carpetas persistentes de la máquina VDI.

#### A. Análisis preliminar del origen de datos

Es conveniente analizar los orígenes de datos antes de cargarlos, dado que nuestro origen es un archivo csv de muy pocos registros, podemos hacer un análisis preliminar directamente:

- Identificar los campos a cargar
- Determinar el tipo de dato de cada campo
- Identificar campos que deben admitir valores nulos
- Otras observaciones

#### B. Crear una tabla en la BBDD con el nombre T\_DATOS

Esta tabla permitirá cargar toda la información del archivo datos.csv en los campos correspondientes.

*C. Crear una transformación con Spoon que realice las siguientes tareas:*

- I. Extraer la información del archivo CSV*
- II. Transformar todos los datos de los campos de tipo texto a mayúsculas*
- III. Corregir los errores detectados en el análisis*
- IV. Cargar la información transformada en la tabla anteriormente creada.*
- V. Realizar la carga efectiva de la tabla y validar el resultado*

*D. Crear un job que ejecute la transformación anterior*

## Ejercicio 4 (10%): Microsoft SQL Analysis Services

### Objetivo:

Toma de contacto con la aplicación MOLAP suministrada en la máquina VDI (Microsoft SQL *Analysis Services*) y capacitar al estudiante para iniciar las actividades del curso.

### Desarrollo del ejercicio:

En este último ejercicio veremos cómo crear un cubo de SSAS y conectarlo a nuestra base de datos.

Describid paso a paso todas las acciones realizadas, aportando las capturas de pantalla que sean necesarias. Los pasos para seguir son los siguientes:

- 1) Abrir Visual Studio (SSAS)
- 2) Crear un nuevo proyecto de *Analysis Services* multidimensional
- 3) Crear un origen de datos (*DataSource*) conectado a *DW\_DB\_XXX*
- 4) Crear una vista de origen de datos (*Data Source View*)
- 5) Crear un cubo con una única tabla de hechos T\_Datos.

Lógicamente un cubo tan simple carece de funcionalidad, el objetivo de esta PEC únicamente es validar las herramientas disponibles.