

PEC 1 – Diseño y uso de bases de datos analíticas

Presentación.

La PEC1 busca consolidar los conocimientos teóricos de los dos primeros módulos didácticos de la asignatura (módulo 1- “Introducción a las bases de datos analíticas” y módulo 2 - “La construcción de la factoría de información corporativa”) y familiarizar al estudiantado con el entorno de prácticas.

Objetivos.

Los objetivos específicos de la actividad incluyen:

- Diferenciar entre el almacén de datos analítico y la base de datos operacional, identificando sus principales características.
- Conocer el contexto del almacén de datos, los componentes de la Factoría de Información Corporativa y la función de cada uno de ellos.
- Validar el entorno de prácticas para garantizar la disponibilidad del escritorio virtual (VDI) de cada estudiante y la correcta configuración del software necesario para llevar a cabo las actividades prácticas del curso.

En lo que respecta al entorno de prácticas el objetivo es llevar a cabo la configuración del escritorio virtual para asegurar que está accesible y validar la conectividad con la base de datos y la ejecución de procesos ETL (Extracción, Transformación y Carga). Además, se realizará una primera toma de contacto con las herramientas Microsoft Visual Studio y SQL *Analysis Services*.

Contenido.

La PEC1 consta de una parte teórica y una parte práctica.

La **parte teórica (40%)**, se evaluará mediante un **cuestionario Moodle** de 5 preguntas tipo test que buscan comprobar la correcta comprensión de los módulos 1 y 2 de la asignatura. Cada pregunta tiene una única opción válida. Antes de realizar el cuestionario se deben haber estudiado los módulos 1 y 2 con un tiempo estimado de dedicación de 2h y 5h respectivamente. Después del estudio, responder el cuestionario no debería llevar más de 30 minutos.

La **parte práctica (60%)** se llevará a cabo en la máquina virtual (VDI) y consiste en la grabación de un video con la realización de 4 ejercicios que se detallan al final de este documento. Para la realización de esta segunda parte, tanto la base de datos SQL Server como la instancia de *Analysis Services* se encuentran en el mismo servidor, por lo que el nombre de dicho servidor será el mismo y se facilitará en el espacio "Anuncios" del aula de Laboratorio de Soporte al entorno VDI.

Se recomienda al estudiantado consultar y seguir las indicaciones de los documentos y vídeos asociados a la PEC1, disponibles en el espacio Contenidos del aula. Estos recursos ayudarán a completar los ejercicios con éxito y resolver posibles dudas frecuentes.

El resultado del ejercicio práctico se entregará en formato de video, la nomenclatura del archivo se especifica en el apartado "Formato y fecha de entrega" del documento. Para preparar el vídeo, se recomienda utilizar la herramienta *ScreencastOMatic*, pero se puede utilizar cualquier otra herramienta, siempre que el video cumpla con los requerimientos indicados. Se estiman 6 horas de dedicación para resolver el ejercicio práctico y una hora para grabar/editar el video de 10 minutos.

Recursos

- Módulo 1. Introducción a las bases de datos analíticas
- Módulo 2. La construcción de la factoría de información corporativa
- Guía Estudio del Módulo 1: Introducción a las bases de datos analíticas
- Guía Estudio del Módulo 2: La construcción de la FIC
- Entorno Virtual VDI. Guía de configuración y conexión.
- FAQs entorno VDI. Resolución de dudas frecuentes.
- Importación de bases de datos, creación de vistas y cubos.
- Guía de instalación y uso de *ScreenOMatic*.
- Vídeos:
 - Primera Transformación Lectura Excel
 - Pasos Comunes en Transformaciones
 - Creación de Jobs
 - Crear Conexión a Bases de Datos
 - Procesar una dimensión con Visual Studio

Criterios de evaluación.

Parte teórica (40%)

- Se valorarán todas las respuestas correctas por igual, 8% del total de la PEC.
- Al responder cada una de las preguntas, se puede comprobar si la respuesta introducida es correcta. Si no lo es, se puede responder de nuevo, pero con una penalización de un 20% respecto a la pregunta errónea.

Parte práctica (60%)

- Se valorará mediante un vídeo resumen (máx. 10 min) que muestre evidencias de cumplimiento de los pasos indicados y aportados por el estudiante al resolver los ejercicios solicitados.
- La grabación del video visualizará los pasos realizados directamente sobre la VDI, no se evaluarán documentos PowerPoint con los pasos, ni ningún otro formato que no sea el especificado.

Penalizaciones

- Penalización de un 10% si el video no tiene comentarios.
- Penalización de un 10% si el video no se graba con la imagen del estudiante (webcam)
- Penalización de un 10% si el video tiene mala calidad (audio y/o vídeo).
- Penalización de un 10% si el video excede los 10 minutos.

Nota final:

La nota final se calculará de acuerdo con esta fórmula: 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 de la solución de la PEC1 se realizará mediante el espacio Contenidos del aula en 2 partes:

1a) Envío del **Cuestionario Moodle**.

2a) Envío de un **vídeo**, de **no más de 10 minutos de duración**.

El fichero debe tener un nombre que siga el siguiente patrón: BDA_PEC1_Apellido1_Apellido2_Nombre.extensión y se debe enviar desde el apartado “Entrega vídeo PEC1” del área de contenidos correspondiente a la PEC1.

IMPORTANTE: Tal y como indica el plan docente:

- Es responsabilidad única del estudiante asegurarse que envía el cuestionario y entrega el vídeo que pretende en el lugar del aula habilitado para ese cometido.
- Entregas realizadas fuera de los canales indicados se considerarán como No presentadas.
- Entregas pasadas las 23:59 h de la fecha límite no serán aceptadas y por tanto, no podrán ser evaluadas.

La fecha máxima de entrega es el 29/10/2024 a las 23:59h

Las PECs son actividades opcionales y, por tanto, no existe la posibilidad de recuperarlas, se deben entregar en tiempo y forma indicados. Se recomienda leer atentamente el Plan Docente de la asignatura para entender los criterios de evaluación.

Enunciado PEC1

Parte teórica (40%)

Cuestionario Moodle PEC1 (disponible en el aula)

Parte práctica (60%)

El estudiantado debe entregar un vídeo que muestre **paso a paso** la realización de los cuatro ejercicios en la propia VDI. La entrega debe incluir todas las explicaciones y pantallas que sean necesarias, y no debe superar los 10 minutos de duración.

El estudiantado puede distribuir los 10 minutos del video de la manera que considere más oportuna, no hay limitaciones en la dedicación de tiempo a cada uno de los ejercicios, siempre y cuando el tiempo total del video no supere los 10 minutos indicados y se dedique un tiempo a cada apartado.

No obstante, para facilitar la distribución de tiempo, se ofrecen unas indicaciones orientativas para los checkpoint en la creación del video.

Ejercicio	Enunciado	Tiempo recomendado
Ejercicio 1	Configuración del entorno VDI	1 minuto
Ejercicio 2	Validación de la BBDD	2 minutos
Ejercicio 3	Validación de la aplicación de ETL	4 minutos
Ejercicio 4	Microsoft SQL <i>Analysis Services</i>	3 minutos

IMPORTANTE: Para crear el proyecto de Visual Studio, hay que importarlo desde el servidor de *Analysis Services*. No hay que crear uno nuevo y no se deben borrar los componentes importados.

Como solución a cada uno de los ejercicios, se incluirá en el vídeo la realización y explicación de los pasos más significativos, que demuestren su correcta resolución.

Ejercicio 1 (5%): Configuración del entorno VDI

Objetivo:

Configurar la conexión al escritorio VDI que se utilizará durante todo el curso en las actividades que lo requieran. Las credenciales para acceder a la máquina VDI son las que utilizáis para entrar en el campus de la UOC.

Desarrollo del ejercicio:

Utilizando el documento **Entorno_Virtual_VDI** (guía de configuración y conexión) el estudiantado debe llevar a cabo todo el proceso hasta obtener una correcta configuración de su escritorio VDI. Se recomienda leer atentamente dicho documento y seguir sus indicaciones.

Como solución a este ejercicio, además de los pasos más significativos, se incluirá la pantalla del escritorio de la VDI con 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 la base de datos Microsoft SQL Server, configuración de la conexión con el servidor y creación de una tabla simple en la base de datos asignada.

Desarrollo del ejercicio:

Para acceder a la base de datos del servidor SQL Server facilitado, cada estudiante dispondrá de sus propias credenciales:

Usuario: `STUDENT_loginuoc`
Contraseña: `$d3f4ult$`

Donde `STUDENT_loginuoc` corresponde al usuario del campus de la UOC. Por ejemplo, para un estudiante con usuario del campus *pmartin*, su usuario en SQL Server sería **`STUDENT_pmartin`**. En el primer inicio de sesión se solicitará cambiar la contraseña.

2.1) Conectar al servidor de bases de datos SQL Server

Utilizando la consola “*Microsoft SQL Server Management Studio (SSMS)*” instalada en la VDI, se abrirá una conexión con el servidor de base de datos y se mostrará en el vídeo la siguiente información:

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

2.2) Configurar la base de datos *SOURCE_loginuoc*

Se pide configurar el parámetro *Recovery Model* de la BBDD como “simple” y desactivar la actualización automática de estadísticas.

2.3) Crear una tabla con el nombre T_PEC1

Se creará una tabla muy sencilla con 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.

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*).

Desarrollo del ejercicio:

Para familiarizarse con la herramienta *Spoon*, que permite la extracción, transformación y carga 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). El archivo para cargar (datos.csv) se facilita con el enunciado de la PEC1.

3.1) Abrir el componente *Spoon*.

3.2) Capturar la versión de PDI.

3.3) Crear un repositorio de tipo *file* con el nombre PEC1.

Un repositorio de tipo file permite organizar todas las transformaciones y jobs creados en Spoon en una misma ubicación. Para crearlo se debe utilizar el botón Connect que aparece en la parte superior derecha.

3.4) Crear una transformación y un *job* básicos:

Se debe copiar el archivo de origen datos.csv, 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:

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

Dado que nuestro origen de datos es un archivo csv de muy pocos registros, se puede hacer un análisis preliminar directamente.

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 preliminar.
- 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

Como parte de la solución de este apartado, se incluirá también en el vídeo la pantalla de la herramienta PDI dónde se visualiza la pestaña con los logs de su ejecución.

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

Objetivo:

Toma de contacto con la aplicación para diseñar cubos multidimensionales (OLAP) suministrada en la máquina VDI (Microsoft SQL *Analysis Services*).

Desarrollo del ejercicio:

Para la realización de este último ejercicio, se recomienda utilizar como guía el documento disponible en los recursos de la PEC1, **Importación de bases de datos, creación de vistas y cubos**, con indicaciones sobre cómo proceder para realizar la importación de la base de datos y crear cubos.

- 4.1) Abrir Visual Studio (SSAS).
- 4.2) Importar el proyecto de *Analysis Services* multidimensional desde el servidor.
- 4.3) Configurar un origen de datos (*DataSource*) conectado a SOURCE_loginuoc.
- 4.4) Crear una vista de origen de datos (*DataSourceView*).
- 4.5) Configurar el destino de los datos.
- 4.6) Crear un cubo con una única tabla de hechos T_Datos.
- 4.7) Procesar e implementar el cubo en el servidor.
- 4.8) Explorar el cubo desde el navegador integrado en SSAS (*Browser*).

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