

## PEC 2

### Presentación

La PEC2 consiste en una serie de preguntas con el objetivo de consolidar los conocimientos teóricos del módulo 3 - "Diseño multidimensional y explotación de datos" de la asignatura y preparar al estudiantado para la creación de los primeros procesos ETL como preparación para la PR2.

### Objetivos

Teniendo en cuenta el contenido del módulo, como objetivos más específicos cabe señalar:

- Comprender los componentes del modelo multidimensional (estructuras de datos, operaciones y restricciones de integridad).
- Diferenciar claramente entre el diseño conceptual, lógico y físico.
- Entender cuál es el diseño multidimensional y los problemas de diseño que presenta (en el ámbito conceptual, así como en el lógico y físico).
- Comprender la importancia de un adecuado diseño de un proyecto de almacén de datos, antes de su desarrollo y puesta en funcionamiento.
- Comprender los mecanismos de almacenamiento e indexación asociados a las herramientas multidimensionales (MOLAP, ROLAP, etc.).
- Diseñar procesos de extracción, transformación y carga (ETL).
- Practicar con la herramienta PDI que se utilizará para la realización de la PR2 para implementar los procesos ETL que permitan cargar los datos en el almacén del caso de uso del semestre.

### Contenido.

La PEC2 consta de dos partes diferenciadas: una parte teórica y una parte práctica.

- La primera parte, la teórica, se llevará a cabo a través de un cuestionario Moodle compuesto por 4 preguntas que tienen por objetivo comprobar la correcta comprensión del módulo 3 por parte del estudiantado.

Debe haberse estudiado el módulo 3 con anterioridad a la realización del cuestionario. Se estima un tiempo de lectura/estudio del módulo de 12h. Después del estudio, responder el cuestionario, como mucho, os debería llevar 15 minutos.

- La segunda parte la constituye un único ejercicio con varios apartados. Su objetivo es introducir al estudiantado en la creación de procesos ETL -que forman parte de la PR2- utilizando la herramienta Spoon (PDI), instalada en el entorno VDI. Se recomienda consultar el documento **“Guía y consejos para el desarrollo de ETL”**.

El resultado de este ejercicio práctico se entregará en **formato vídeo**, en el apartado *Entrega vídeo PEC2* del aula (ver la duración máxima y la nomenclatura del nombre del archivo a entregar en el apartado “Formato y fecha de entrega” de este documento).

Se estiman 6h para resolver el ejercicio práctico y 1h para grabar el vídeo de 10 minutos

## Recursos

- Módulo 3. Diseño multidimensional y explotación de datos
- Guía Estudio del Módulo 3: Diseño Multidimensional y Explotación de Datos
- Retos 3 y 4 del material teórico-práctico
- Guía y consejos para el desarrollo de ETL
- Enunciado “Caso práctico: análisis del impacto conductual de la COVID-19 en la población” y en particular de este caso:
  - Enunciado PR2
  - Solución PR2
- Fichero “Payment\_type.xls” (disponible junto al enunciado de la PEC2)
- Manual de uso de ScreencastOMatic

## Criterios de evaluación.

### Parte teórica (60%)

- Se valorarán todas las respuestas correctas por igual, 15% del total de la PEC.
- El estudiantado individualmente dispondrá de **UN intento**, que puede ir respondiendo, revisando y cambiando hasta el día de la entrega. Cuando lo dé por bueno, enviará su intento y se determinará automáticamente su calificación.
- Al responder cada una de las preguntas, el estudiantado tiene la posibilidad de comprobar si su respuesta es correcta. Si no lo es, puede responder de nuevo, penalizando un 20% la puntuación de esa pregunta.

## **Parte práctica (40%)**

- Se valorará mediante un **vídeo resumen** que muestre evidencias de cumplimiento de los pasos indicados y aportados por el estudiantado al resolver en su propio escritorio los ejercicios solicitados.
- Los vídeos sin explicación no se valorarán como totalmente correctos.
- La duración máxima del vídeo será de **10 minutos**.
- Se valorarán todos los apartados por igual, 8% del total de la PEC.

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 estudiantado.
- Penalización de un 10% si el video y/o audio tiene mala calidad.
- Penalización de un 10% si el video excede los 10 minutos.

**La nota final estará formada por:**

Pregunta 1 (15%) + Pregunta 2 (15%) + Pregunta 3 (15%) + Pregunta 4 (15%)  
+ Parte práctica (40%)

## **Formato y fecha de entrega**

La entrega de la solución de la PEC2 se realizará en 2 partes:

- a) Envío del **Cuestionario Moodle** a través del espacio correspondiente del aula.
- b) Envío de un **vídeo**, de **no más de 10 minutos de duración**, en el apartado **Entrega vídeo PEC2** del aula. El fichero enviado debe tener un nombre que siga el siguiente patrón: BDA\_PEC2\_Apellido1\_Apellido2\_Nombre.extensión.

**IMPORTANTE:** Tal y como indica el plan docente:

- Es responsabilidad única del estudiantado 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 03/12/2024 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 al estudiantado.

***La fecha de calificación y retorno pedagógico es el 13/12/2024***

### **Parte teórica (60%)**

**Cuestionario Moodle PEC2 (disponible en el aula)**

### **Parte práctica (40%)**

El estudiantado individualmente debe entregar, como solución a esta parte, un **vídeo que ilustre paso a paso** la realización del ejercicio 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	Creación de la tabla intermedia	2 minutos
Ejercicio 2	Lectura del fichero .xls	3 minutos
Ejercicio 3	Carga del contenido en la tabla de la BBDD	3 minutos
Ejercicio 4	Mostrar la ejecución completa	1 minuto
Ejercicio 5	Ejecutar la query y mostrar resultados	1 minuto

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.

A partir del fichero “Payment\_type”, se debe diseñar, implementar y ejecutar los procesos de extracción, transformación y carga para la Transformación IN\_PAYMENT\_TYPES, siguiendo los siguientes pasos:

- a) Ejecutar el siguiente comando SQL para la creación de la tabla intermedia donde se almacenará los datos del origen “Payment\_type.xls”

```
DROP TABLE IF EXISTS dbo.STG_PAYMENT_TYPES ;  
CREATE TABLE dbo.STG_PAYMENT_TYPES(  
    [id] int NOT NULL,  
    [payment_type] NVARCHAR(100) NOT NULL  
)
```

- b) Lectura del fichero xls. Realizar el paso «Microsoft Excel *input*»:

**Nombre:** Microsoft Excel Input

**Componente:** Microsoft Excel input

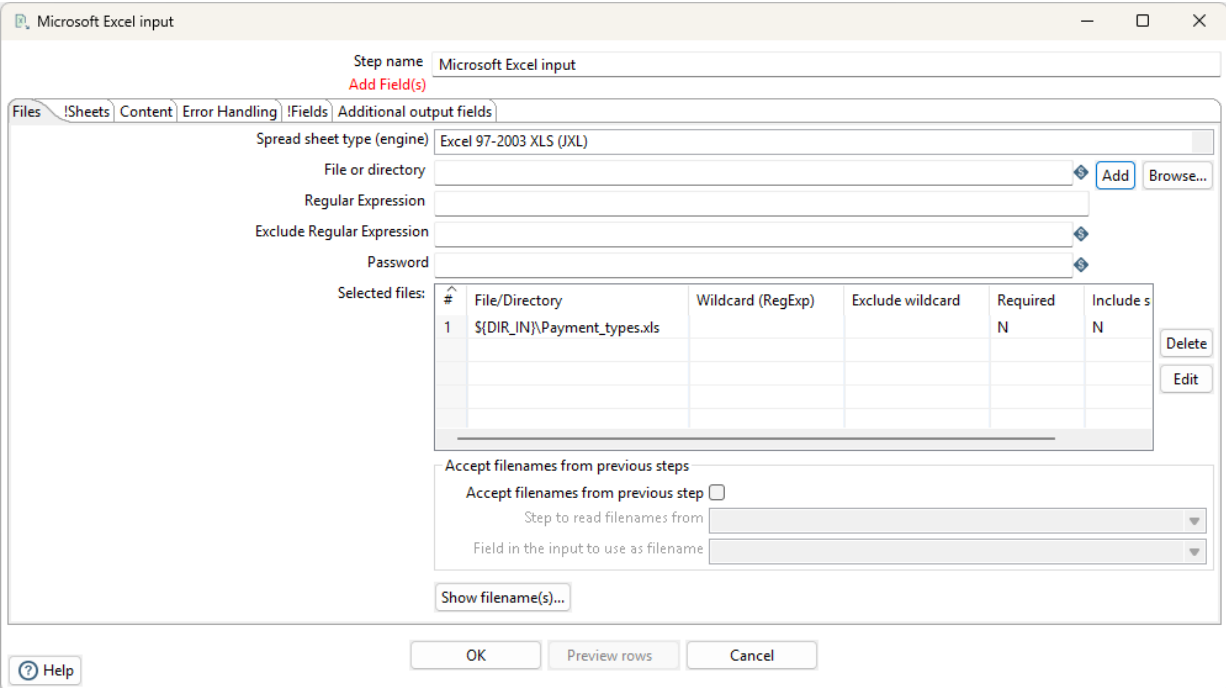
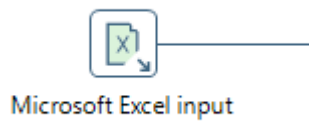
**Descripción:** Permite cargar datos de entrada provenientes de un fichero xls.

El primer paso de la transformación corresponde a la lectura del fichero origen. Como se trata de un fichero .xls se utilizará como entrada el tipo «Microsoft Excel *input*», concretamente “payment\_type.csv”.

### Parámetros:

**Files / File or directory:** \${DIR\_IN}\Payment\_type.xls

Para facilitar la lectura del fichero se utiliza la variable de entorno «DIR\_IN», aunque para esta PEC2 serviría con seleccionar el path del propio fichero.



Microsoft Excel input

Step name: Microsoft Excel input

Add Field(s)

Files | Sheets | Content | Error Handling | Fields | Additional output fields

Spread sheet type (engine): Excel 97-2003 XLS (JXL)

File or directory:  Add Browse...

Regular Expression:

Exclude Regular Expression:

Password:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include
1	\${DIR_IN}\Payment_types.xls			N	N

Delete Edit

Accept filenames from previous steps

Accept filenames from previous step ☐

Step to read filenames from:

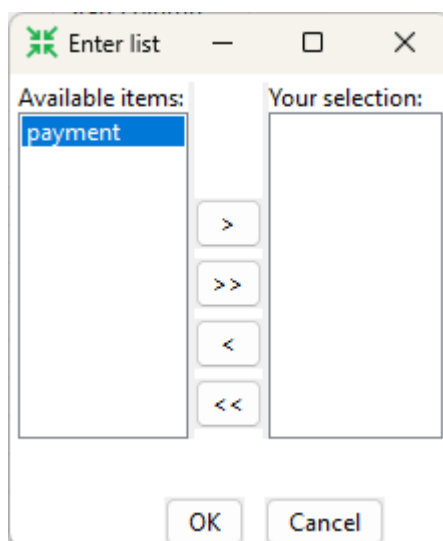
Field in the input to use as filename:

Show filename(s)...

OK Preview rows Cancel

En este paso, indicamos la ubicación del archivo en “file or directory” y con el botón “Add” lo añadimos al grid de “selected files” para que quede como en la imagen superior.

**Sheets (hojas)** En la pestaña «*Sheets*», mediante el botón «*Get sheetname(s)*» se obtienen todas las hojas del fichero. En este caso es solo una con el nombre “payment”



**Fields (campos)** En la pestaña «*Fields*», mediante el botón «*Get fields form header row...*» se obtienen todos los campos del fichero.

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	id	Number			none	N				
2	Payment_type	String			none	N				

**Preview:** botón «*Preview rows*» (Previsualizar filas).

Examine preview data			
Rows of step: Microsoft Excel input (6 rows)			
#	id	Payment_type	
1	0.0	Credit card	
2	1.0	Cash	
3	2.0	No charge	
4	3.0	Dispute	
5	4.0	Unknown	
6	5.0	Voided trip	
7			

- c) Cargar la información transformada en la tabla de base de datos mediante el paso «*Table Output*»:

**Nombre:** Insert into SQL Server

**Componente:** Table Output

**Descripción:** Carga datos en una tabla de base de datos. Es equivalente al operador de SQL INSERT.

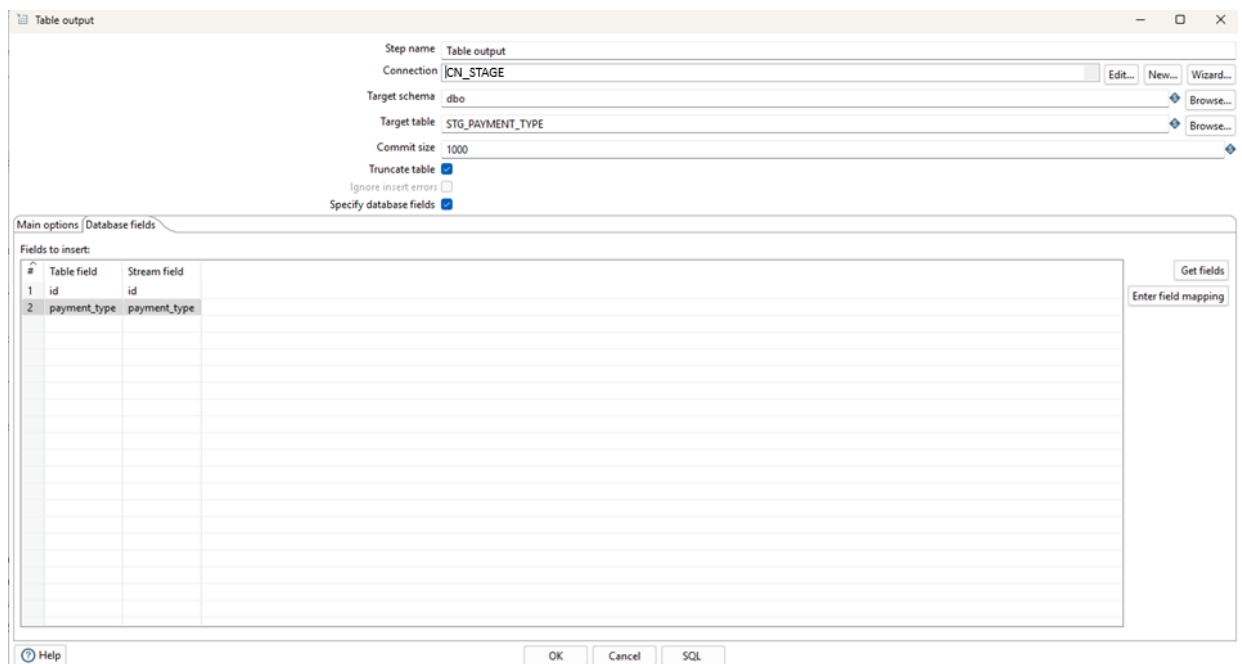
Carga los datos resultantes de las transformaciones precedentes en la tabla intermedia del stage area: [dbo].[STG\_PAYMENT\_TYPES]

**Parámetros:**

**Connection:** Mediante la conexión previamente establecida «*CN\_STAGE*» o en su defecto la creada durante la PEC1

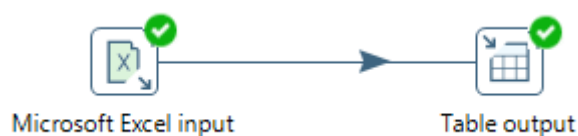
**Target table:** [dbo].[STG\_PAYMENT\_TYPES]

**Truncate table:** Sí, marcado para posibles reprocesos.



- d) Mostrar la pantalla de la transformación completa, incluyendo la pestaña informativa de ejecución «*step metrics*».

**Pantalla de la transformación completa:**

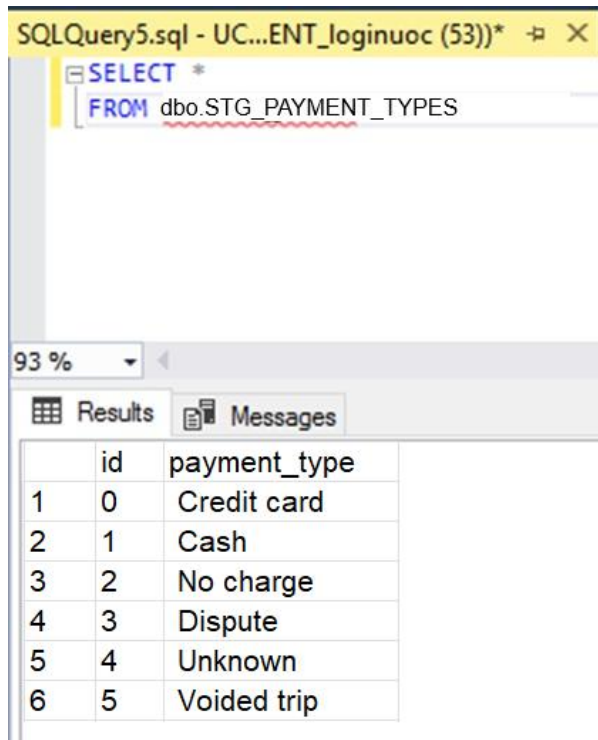




### Registros incorporados: 6

- e) Realizar la consulta en la Base de Datos y mostrar el resultado de registros sin ordenar

```
SELECT * FROM dbo.STG_PAYMENT_TYPES
```



The screenshot shows a SQL Server Enterprise Manager interface. At the top, a query window titled 'SQLQuery5.sql - UC...ENT\_loginuoc (53))\*' contains the SQL query: `SELECT * FROM dbo.STG_PAYMENT_TYPES`. Below the query window, the 'Results' tab is active, displaying a table with 6 rows and 3 columns: 'id', 'payment\_type', and an unlabeled column. The data is as follows:

	id	payment_type
1	0	Credit card
2	1	Cash
3	2	No charge
4	3	Dispute
5	4	Unknown
6	5	Voided trip