

## PAC 2

### Presentació

La PAC2 consisteix en una sèrie de preguntes amb l'objectiu de consolidar els coneixements teòrics del mòdul 3 - "Disseny multidimensional y explotació de dades" de l'assignatura i preparar l'estudiant per a la creació dels primers processos ETL com preparació per la PR2.

### Objectius

Tenint en compte el contingut del mòdul, com a objectius més específics cal assenyalar el següent:

- Comprendre els components del model multidimensional (estructures de dades, operacions i restriccions d'integritat).
- Diferenciar clarament entre el disseny conceptual, lògic i físic.
- Entendre quin és el disseny multidimensional i els problemes de disseny que presenta (en l'àmbit conceptual, així com en el lògic i físic).
- Comprendre la importància d'un adequat disseny d'un projecte de magatzem de dades, abans del seu desenvolupament i posada en marxa.
- Comprendre els mecanismes d'emmagatzematge i indexació associats a les eines multidimensionals (MOLAP, ROLAP, etc.).
- Dissenyar processos d'extracció, transformació i càrrega (ETL).
- Practicar amb l'eina PDI que s'utilitzarà per a la realització de la PR2 per a implementar els processos ETL que permetin carregar les dades en el magatzem del cas d'ús del semestre.

### Contingut.

La PAC2 consta de dues parts diferenciades: una part teòrica i una part pràctica.

- La primera part, la teòrica, es durà a terme mitjançant un qüestionari Moodle compost per 4 preguntes que tenen per objectiu comprovar la correcta comprensió del mòdul 3 per part de l'estudiant.

S'ha d'haver estudiat el mòdul 3 prèviament a la realització del qüestionari. S'estima un temps de lectura/estudi del mòdul de 12h. Després de l'estudi, respondre el qüestionari, com a molt, hauria de portar 15 minuts.

- La segona part la constitueix un únic exercici amb diversos apartats. El seu objectiu és introduir l'estudiantat en la creació de processos ETL – que formen

part de la PR2 – utilitzant l'eina Spoon (PDI), instal·lada a l'entorn VDI. Es recomana estudiar el document **“Guia i consells per al desenvolupament d'ETL”**.

El resultat d'aquest exercici pràctic es lliurarà en format vídeo, a l'apartat **Lliurament vídeo PAC2** de l'aula (veure durada màxima i la nomenclatura del nom d'arxiu a entregar a l'apartat “Format i data de lliurament” d'aquest document).

S'estimen 6h per a resoldre l'exercici pràctic i 1h per a enregistrar un vídeo de 10 minuts.

## Recursos

- Mòdul 3. Disseny multidimensional i explotació de dades
- Guia Estudi del Mòdul 3: Disseny multidimensional i explotació de dades
- Reptes 3 i 4 del material teòric-pràctic
- Guia i consells per al desenvolupament d'ETL
- Enunciat “Cas pràctic: anàlisi de l'impacte conductual de la COVID-19 a la població” i en particular d'aquest cas:
  - o Enunciat PR2
  - o Solució PR2
  - o COICOP-xml (disponible junt amb l'enunciat de la PAC2)
- Manual d'ús de ScreencastOMatic

## Criteris d'avaluació.

### Part teòrica (60%)

- Es valoraran totes les respostes correctes per igual, 15% del total de la PAC.
- L'estudiantat disposarà d'**UN intent**, que pot anar responnent, revisant i canviant fins el dia del lliurament. Quan ho consideri vàlid, enviarà el seu intent i es determinarà automàticament la seva qualificació.
- En respondre cadascuna de les preguntes, l'estudiantat té la possibilitat de comprovar si la seva resposta és correcta. Si no ho és, pot respondre-la de nou, penalitzant un 20% la puntuació d'aquella pregunta.

## **Part pràctica (40%)**

- Es valorarà mitjançant un **video resum** que mostri evidències del compliment dels passos indicats i aportats per l'estudiant en resoldre en el seu propi escriptori els exercicis sol·licitats.
- Els vídeos sense explicació no es valoraran com totalment correctes.
- La durada màxima del vídeo serà de **10 minuts**.
- Es valoraran tots els punts per igual, 6,6 % del total de la PAC.

Penalitzacions:

- Penalització d'un 10% si el vídeo no té comentaris.
- Penalització d'un 10% si el vídeo no s'enregistra amb la imatge de l'estudiantat.
- Penalització d'un 10% si el vídeo i/o l'àudio té mala qualitat.
- Penalització d'un 10% si el vídeo excedeix els 10 minuts.

**La nota final estarà formada per:**

Pregunta 1 (15%) + Pregunta 2 (15%) + Pregunta 3 (15%) + Pregunta 4 (15%)  
+ Part pràctica (40%)

## **Format i data de lliurament**

El lliurament de la solució de la PAC2 es realitzarà en dues parts:

- a) Enviament del **Qüestionari Moodle** mitjançant l'espai corresponent de l'aula.
- b) Enviament d'un **vídeo**, de **no més de 10 minuts de durada**, mitjançant l'apartat **Lliurament vídeo PAC2** de l'aula. El fitxer enviat ha de tenir un nom que segueixi el següent patró: BDA\_PAC2\_Cognom1\_Cognom2\_Nom.extensió.

**IMPORTANT: És responsabilitat única de l'estudiant assegurar-se que envia el qüestionari i entrega el vídeo que pretén en el lloc de l'aula habilitat per a aquesta comesa.**

*La data màxima de lliurament és el 05/12/2023 a les 23:59 h.*

## **Format i data de feedback o retorn pedagògic**

El feedback o retorn pedagògic es realitzarà de forma individual mitjançant el Registre AC (Avaluació Continua). A més, es publicarà al fòrum associat a la PAC una valoració general dels resultats on s'indicaran aquells errors comuns o observacions que es vulguin donar amb caràcter general als estudiants.

*La data de qualificació i retorn pedagògic és el 15/12/2023*

## **Part teòrica (60%)**

**Qüestionari Moodle PAC2 (disponible a l'aula)**

## **Part pràctica (40%)**

*L'estudiantat ha d'entregar, com a solució a aquesta part, un vídeo que il·lustri pas a pas la realització de l'exercici a la pròpia VDI. El lliurament ha d'incloure totes les explicacions i pantalles que siguin necessàries, i no ha d'excedir els 10 minuts de durada.*

A partir del fitxer "COICOP.xml", s'han de dissenyar, implementar i executar els processos d'extracció, transformació i càrrega per a la Transformació IN\_COICOP, seguint els següents passos:

- Executar la següent comanda SQL per a la creació de la taula intermèdia on s'emmagatzemaran les dades d'origen "COICOP.xml"

### **STG\_COICOP**

```
CREATE TABLE dbo.STG_COICOP(  
    [code] nvarchar(25),  
    [Description] nvarchar(300)  
)
```

b) Lectura del fitxer XML. Implementar el pas «*Get data from XML*»:

**Nom:** Lectura fitxer xml

**Component:** Get data from XML

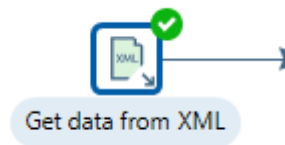
**Descripció:** Permet carregar dades d'entrada procedents d'un fitxer XML.

El primer pas de la transformació correspon a la lectura del fitxer origen.  
Com es tracta d'un fitxer .xml s'utilitzarà com a entrada el tipus «*Get data from XML*», concretament "COICOP.xml".

**Paràmetres:**

**Files / File or directory:** \${DIR\_IN}\COICOP.xml

Per a facilitar la lectura del fitxer s'utilitzarà la variable d'entorn «DIR\_IN»:



Get data from XML

Step name: Get data from XML

File Content Fields Additional output fields

XML source from field

XML source is defined in a field? ☐

XML source is a filename? ☐

Read source as Url ☐

get XML source from a field

File or directory

Regular Expression

Exclude Regular Expression

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfo
1	\${DIR_IN}\COICOP.xml			N	N
2					

En aquest pas, indicar per una banda el «*Loop XPath*» i ignorar els possibles comentaris que hi hagi dins del fitxer, A la imatge es pot veure la configuració que s'acaba de comentar:

Step name: **Get data from XML**

File | Content | **Fields** | Additional output fields

Settings

Loop XPath:  Get XPath nodes

Encoding:

Namespace aware? ☐

Ignore comments? ☒

Validate XML? ☐

Use token ☐

Ignore empty file ☐

Do not raise an error if no files ☒

Limit:

Prune path to handle large files:

Additional fields

Include filename in output? ☐ Filename fieldname:

Rownum in output? ☐ Rownum fieldname:

**Fields (camps)** A la pestanya «*Fields*», mitjançant el botó «*Select fields*» s'obtenen tots els camps del fitxer.

Step name: **Get data from XML**

File | Content | **Fields** | Additional output fields

#	Name	XPath	Element	Result type	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type	Rep
1	Code	Code	Node	Value of	String							none	N
2	Description	Description	Node	Value of	String							none	N

Get fields

**Preview:** botó «*Preview rows*» (Previsualitzar files).

Examine preview data

Rows of step: Get data from XML (100 rows)

#	Code	Description
1	TOT_X_CP041_042	Total except actual rents
2	CP01113	Bread
3	CP01116	Pasta products and couscous
4	CP01118	Other cereal products
5	CP01126_01127	Dried, salted or smoked meat and edible meat offal
6	CP01128A	Other preserved or processed meat and meat preparations
7	CP01128B	Other fresh, chilled or frozen edible meat
8	CP01131_01132	Fresh, chilled or frozen fish
9	CP01133_01134	Fresh, chilled or frozen seafood
10	CP01135	Dried, smoked or salted fish and seafood
11	CP01136	Other preserved or processed fish and seafood and fish and seafood preparations
12	CP0116A	Citrus fruits (fresh, chilled or frozen)
13	CP0116B	Bananas (fresh, chilled or frozen)

- c) Filtratge de nuls. Eliminar valors nuls al camp 'code'. Es realitzarà mitjançant el pas «*Filter rows*»:

**Nom:** Eliminar valors nuls

**Component:** Filter rows

**Descripció:** S'eliminen valors nuls

**Paràmetres:**

Filter rows

Step name: Filter rows

Send 'true' data to step: [dropdown]

Send 'false' data to step: [dropdown]

The condition:

[ ]

Code IS NOT NULL - [ ]

[ ]

+

- d) Carregar la informació transformada a la taula de base de dades mitjançant el pas «*Table Output*»:

**Nom:** Sortida Taula

**Component:** Table Output

**Descripció:** Carrega dades en una base de dades. És equivalent a l'operador d'SQL INSERT.

Carrega les dades resultants de les transformacions precedents a la taula intermèdia de l'àrea de stage: [dbo].[STG\_COICOP]

### Paràmetres:

**Connection:** Mitjançant la connexió prèviament establerta «CN\_STAGE»

**Target table:** [dbo].[STG\_COICOP]

**Truncate table :** Sí, marcat per a possibles reprocessos.

Table output

Step name: Insert into SQL Server

Connection: CN\_STG

Target schema: dbo

Target table: STG\_COICOP

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	Code	Code
2	Description	Description

- e) Mostrar la pantalla de la transformació complerta, incloent la pestanya informativa de l'execució «step metrics».

### Pantalla de la transformació completa:



### Registres incorporats: 951

- f) Realitzar la consulta a la Base de Dades i mostrar el resultat del Top 10 de registres sense ordenar. Coincideixen els 10 primers registres ordenats de forma ascendent de tots els camps segons la seva col·locació a la taula *Staging*?

```
SELECT TOP 10 * FROM STG_COICOP
```

```
SELECT TOP 10 description, code FROM STG_COICOP ORDER BY description ASC, code ASC
```