

# PAC 1 – Disseny i ús de bases de dades analítiques

## Presentació.

Amb aquesta PAC1 es pretén ajudar a consolidar els coneixements teòrics adquirits en l'estudi dels dos primers mòduls didàctics de l'assignatura (mòdul 1- “Introducció a les bases de dades analítiques” i mòdul 2 - “La construcció de la factoria d'informació corporativa”) i familiaritzar a l'estudiantat amb l'entorn de pràctiques.

## Objectius.

Els objectius específics de l'activitat inclouen:

- Diferenciar entre el magatzem de dades analític i la base de dades operacional, identificant les seves característiques principals.
- Conèixer el context del magatzem de dades, els components de la Factoria d'Informació Corporativa i la funció de cadascun d'ells.
- Validar l'entorn de pràctiques per a garantir la disponibilitat de l'escriptori virtual (VDI) de cada estudiant i la correcta configuració del programari necessari per a dur a terme les activitats pràctiques del curs.

I pel que a la l'entorn de pràctiques l'objectiu és dur a terme la configuració de l'escriptori virtual per assegurar que sigui accessible i validar la connectivitat amb la base de dades i l'execució de processos ETL (Extracció, Transformació i Càrrega). A més a més, es realitzarà una primera presa de contacte amb les eines Microsoft Visual Studio i SQL Analysis Services.

## Contingut.

La PAC1 consta d'una part teòrica i una part pràctica.

La **part teòrica (40%)**, s'avaluarà mitjançant un **qüestionari Moodle** de 5 preguntes tipus test que busquen comprovar la correcta comprensió dels mòduls 1 i 2 de l'assignatura. Cada pregunta té una única opció vàlida. Abans de realitzar el qüestionari

s'han d'haver estudiat els mòduls 1 i 2 amb un temps estimat de dedicació de 2h i 5h respectivament. Després de l'estudi, respondre el qüestionari no hauria de portar més de 30 minuts.

La **part pràctica (60%)** es durà a terme en la màquina virtual (VDI) i consisteix en l'enregistrament d'un vídeo amb la realització de 4 exercicis que es detallen al final d'aquest document. Per a la realització d'aquesta segona part, tant la base de dades SQL Server com la instància d'*Analysis Services* corren sobre el mateix servidor i, per tant, tindran el mateix nom, que es facilitarà via l'espai Anunci a l'aula de Laboratori de Suport a l'entorn VDI.

Es recomana a l'estudiantat consultar i seguir les indicacions dels materials i vídeos associats a la PAC1, disponibles a l'espai de Continguts de l'aula. Aquests recursos ajudaran a completar els exercicis amb èxit i resoldre possibles dubtes freqüents.

El resultat de l'exercici pràctic es lliurarà en format de vídeo, la nomenclatura de l'arxiu s'especifica en l'apartat "Format i data de lliurament" del document. Per a preparar el vídeo, es recomana utilitzar l'aplicació *ScreencastOMatic*, però es pot utilitzar qualsevol altra eina, sempre que el vídeo compleixi amb els requeriments indicats. S'estimen 6 hores de dedicació per a resoldre l'exercici pràctic i una hora per a gravar/editar el vídeo de 10 minuts.

## Recursos

- Mòdul 1. Introducció a les bases de dades analítiques
- Mòdul 2. La construcció de la factoria d'informació corporativa
- Guia Estudi del Mòdul 1: Introducció a les bases de dades analítiques
- Guia Estudi del Mòdul 2: La construcció de la FIC
- Entorn Virtual VDI. Guia de configuració i connexió.
- FAQs entorn VDI. Resolució de dubtes freqüents.
- Importació de bases de dades, creació de vistes i cubs.
- Guia d'instal·lació i ús de *ScreenOMatic*.
- Vídeos:
  - Primera Transformació Lectura Excel

- Passos Comuns en Transformacions
- Creació de Jobs
- Crear Connexió a Bases de dades
- Prova processar dimensió

## **Criteris d'avaluació.**

### **Part teòrica (40%)**

- Es valoraran totes les respostes correctes per igual, 8% del total de la PAC.
- En respondre cadascuna de les preguntes, es pot comprovar si la resposta donada és correcta. Si no ho és, pot respondre de nou però amb una penalització d'un 20% sobre la puntuació de la pregunta errònia.

### **Part pràctica (60%)**

- Es valorarà mitjançant un vídeo resum (màx. 10 min) que mostri evidències de compliment dels passos realitzats per l'estudiant en resoldre els exercicis sol·licitats.
- L'enregistrament del vídeo visualitzarà els passos realitzats directament sobre la VDI, no s'avaluaran documents PowerPoint amb els passos, ni cap altre format que no sigui l'especificat.

#### Penalitzacions

- Penalització d'un 10% si el vídeo no té comentaris.
- Penalització d'un 10% si el vídeo no es grava amb la imatge de l'estudiant.
- Penalització d'un 10% si el vídeo té mala qualitat (àudio o vídeo).
- Penalització d'un 10% si el vídeo excedeix los 10 minuts.

## **Nota final:**

La nota final es calcularà d'acord a aquesta fórmula: Pregunta 1 (8%) + Pregunta 2 (8%) + Pregunta 3 (8%) + Pregunta 4 (8%) + Pregunta 5 (8%) + Exercici 1 (5%) + Exercici 2 (20%) + Exercici 3 (25%) + Exercici 4 (10%)

## Format i data de lliurament

El lliurament de la solució de la PAC1 es realitzarà mitjançant l'espai Continguts de l'aula en 2 parts:

1a) Enviament del **Qüestionari Moodle**.

2a) Enviament d'un **vídeo**, de **no més de 10 minuts de duració**.

El fitxer ha de tenir un nom que segueixi el següent patró: BDA\_PAC1\_Cognom1\_Cognom2\_Nom.extensió i cal enviar-ho des de l'apartat "Lliurament vídeo PAC1" de l'àrea de Continguts corresponent a la PAC1.

### **IMPORTANT:** Tal i com s'indica en el pla docent

- És responsabilitat única de l'estudiant assegurar-se que envia el qüestionari i lliura el vídeo que pretén en el lloc de l'aula habilitat per a aquesta comesa.
- Lliuraments realitzats fora dels canals indicats es consideraran com No presentats.
- Lliuraments passades les 23.59 h de la data límit no seran acceptats i per tant, no podran ser avaluats.

***La data màxima de lliurament és el 31/10/2023 a les 23:59h***

Les PACs són activitats opcionals i, per tant, no existeix la possibilitat de recuperar-les, s'han de lliurar dintre del termini i en la forma indicats. Es recomana llegir atentament el Pla Docent de l'assignatura per a entendre els criteris d'avaluació.

## Enunciat PAC1

### Part teòrica (40%)

Qüestionari Moodle PAC1 (disponible a l'aula)

### Part pràctica (60%)

L'estudiantat ha de lliurar un vídeo que mostri pas a pas la realització dels quatre exercicis en la pròpia VDI. El lliurament ha d'incloure totes les explicacions i pantalles que siguin necessàries, i no ha de superar els 10 minuts de durada.

**IMPORTANT:** Per a crear el projecte de Visual Studio, cal importar-lo des del servidor d'*Analysis Services*. No s'ha de crear un nou i no s'han d'esborrar els components importats.

Com a solució a cadascun dels exercicis, s'inclourà en el vídeo la realització i explicació dels passos més significatius, que demostrin la seva correcta resolució.

### Exercici 1 (5%): Configuració de l'entorn VDI

#### Objectiu:

Configurar la connexió a l'escriptori VDI que s'utilitzarà durant tot el curs en les activitats que el requereixin. Les credencials per a accedir a la màquina VDI són les que utilitzeu per a entrar al campus de la UOC.

#### Desenvolupament de l'exercici:

Utilitzant el document **Entorn\_Virtual\_VDI** (guia de configuració i connexió) l'estudiantat ha de dur a terme tot el procés fins a obtenir una correcta configuració del seu escriptori VDI. Es recomana llegir atentament aquest document i seguir les seves indicacions.

Com a solució a aquest exercici, a més a més dels passos més significatius, s'inclourà la pantalla de l'escriptori de la VDI amb la informació tècnica disponible sota el logotip de la UOC (*HostName*, *UserName*, etc.), que trobareu com a fons d'escriptori.

## Exercici 2 (20%): Validació de la BBDD.

### Objectiu:

Presa de contacte amb la base de dades Microsoft SQL Server, configuració de la connexió amb el servidor i creació d'una taula simple en la base de dades assignada.

### Desenvolupament de l'exercici:

Per a accedir a la base de dades del servidor SQL Server facilitat, cada estudiant disposarà de les seves pròpies credencials:

Usuari: `STUDENT_loginuoc`  
Paraula de pas: `$d3f4ult$`

On *STUDENT\_loginuoc* correspon a l'usuari del campus de la UOC. Per exemple, per a un estudiant amb usuari del campus *pmartin*, el seu usuari en SQL Server seria ***STUDENT\_pmartin***. En el primer inici de sessió se sol·licitarà canviar la contrasenya.

### 2.1) Connectar al servidor de bases de dades SQL Server

Fent ús de la consola “*Microsoft SQL Server Management Studio (SSMS)*” instal·lada en la VDI, s'obrirà una connexió amb el servidor de base de dades i es mostrarà en el vídeo la següent informació:

- a) Versió de SSMS
- b) Pantalla de connexió inicial de la consola SQL Server Management Studio (SSMS) on es pugui observar la configuració utilitzada.
  - *Server Type*
  - *Server name*
  - *Authentication*
  - *Login*
- c) Explorador d'objectes de SSMS on es pugui veure l'usuari de connexió.

## 2.2) Configurar la base de dades *SOURCE\_loginuoc*

Es demana configurar el paràmetre *Recovery Model* de la BBDD com a “simple” i desactivar l'actualització automàtica d'estadístiques.

## 2.3) Crear una taula amb el nom *T\_PAC1*

Es crearà una taula molt senzilla amb dos camps: *camp1* i *camp2* de la següent manera:

- **Camp1** serà de tipus *varchar(50)* i permetrà valors nuls.
- **Camp2** serà de tipus *int* i no permetrà valors nuls.

## Exercici 3 (25%): Validació de l'aplicació de ETL.

### Objectiu:

Presa de contacte amb el programari d'ETL subministrat en la màquina VDI (*Pentaho Data Integration*).

### Desenvolupament de l'exercici:

Per a familiaritzar-se amb l'eina *Spoon*, que permet l'extracció, transformació i càrrega de dades, es demana carregar les dades d'un arxiu CSV (origen de dades) en una nova taula de la base de dades (destinació de dades). L'arxiu per a carregar (*dades.csv*) es facilita amb l'enunciat de la PAC1.

### 3.1) Obrir el component *Spoon*.

### 3.2) Capturar la versió de PDI.

### 3.3) Crear un repositori de tipus *file* amb el nom PAC1.

Un repositori de tipus *file* permet organitzar totes les transformacions i jobs creats en *Spoon* en una mateixa ubicació. Per a crear-ho s'ha d'utilitzar el botó *Connect* que apareix a la part superior dreta.

### 3.4) Crear una transformació i un *job* bàsics:

S'ha de copiar l'arxiu d'origen `dades.csv`, a una de les carpetes persistents de la màquina VDI.

#### a. Anàlisis preliminar de l'origen de dades

És convenient analitzar els orígens de dades abans de carregar-los:

- Identificar els camps a carregar.
- Determinar el tipus de dada de cada camp.
- Identificar camps que admeten valors nuls.
- Altres observacions.

Atès que el nostre origen de dades és un arxiu csv de molt pocs registres, podem fer una anàlisi preliminar directament.

#### b. Crear una taula a la BBDD amb el nom `T_DADES`

Aquesta taula permetrà carregar tota la informació de l'arxiu `dades.csv` en els camps corresponents.

#### c. Crear una transformació amb Spoon que realitzi les següents tasques:

- I. Extreure la informació de l'arxiu CSV.
- II. Transformar totes les dades dels camps de tipus text a majúscules.
- III. Corregir els errors detectats en l'anàlisi preliminar.
- IV. Carregar la informació transformada en la taula anteriorment creada.
- V. Realitzar la càrrega efectiva de la taula i validar el resultat.

#### d. Crear un job que executi la transformació anterior

Com a part de la solució d'aquest apartat, s'inclourà també en el vídeo la pantalla de l'eina PDI on es visualitza la pestanya amb els logs de la seva execució.



## Exercici 4 (10%): Microsoft SQL *Analysis Services*

### Objectiu:

Presa de contacte amb l'aplicació per a dissenyar cubs multidimensionals (OLAP) subministrada en la màquina VDI (Microsoft SQL *Analysis Services*).

### Desenvolupament de l'exercici:

Per a la realització d'aquest últim exercici, es recomana utilitzar com a guia el document disponible en els recursos de la PAC1, **Importació de bases de dades, creació de vistes i cubs**, amb indicacions sobre com procedir per a realitzar la importació de la base de dades i crear cubs.

- 4.1) Obrir Visual Studio (SSAS).
- 4.2) Importar el projecte d'*Analysis Services* multidimensional des del servidor.
- 4.3) Configurar un origen de dades (*DataSource*) connectat a SOURCE\_loginuoc.
- 4.4) Crear una vista d'origen de dades (*DataSourceView*).
- 4.5) Configurar el destí de les dades.
- 4.6) Crear un cub amb una única taula de fets T\_Dades.
- 4.7) Processar e implementar el cub en el servidor.
- 4.8) Explorar el cub des del navegador integrat en SSAS (*Browser*).

Lògicament una cub tan simple manca de funcionalitat, l'objectiu d'aquesta PAC únicament és validar les eines disponibles.