

# Cas pràctic: Magatzem de dades per l'anàlisi de la conciliació laboral i familiar o *work-life balance*

## PR3 – Explotació de dades

### Índex

Presentació.....	2
Descripció.....	3
Criteris d'avaluació .....	5
Guia de mostra .....	6
Format i data de lliurament.....	9

## Presentació

A partir de l'*script* facilitat en aquesta pràctica, l'estudiantat individualment ha d'implementar els cubs multidimensionals necessaris per a l'explotació de dades i anàlisi posterior amb l'objectiu de facilitar la presa de decisions a les persones usuàries potencials.

L'objectiu d'aquesta activitat és implementar un model *online analytical processing* (OLAP) per a l'anàlisi multidimensional de la informació que permeti respondre les preguntes definides en l'anàlisi de requeriments del cas pràctic.

L'*script* facilitat al costat d'aquest enunciat conté les instruccions necessàries per a la creació de taules i la càrrega de dades.

## Descripció

Més concretament, aquesta tercera part del cas pràctic consisteix a dissenyar un model OLAP que permeti donar resposta a les qüestions següents:

- 1) Percentatge de persones ocupades que van treballar des de casa durant l'any 2020. Els resultats es mostraran ordenats de menor a major.
- 2) Top 5 de sectors productius amb més hores treballades de mitjana durant 2021.
- 3) Evolució anual del percentatge de persones que treballen a casa a Espanya i Portugal des de l'any 2018 fins al 2022 (tots dos inclosos).
- 4) Percentatge d'hores treballades a casa segons el nombre de fills. El resultat es mostrarà arrodonit a un decimal.
- 5) Mitjana d'hores treballades als països de la UE durant l'any 2021, ordenats per regió i arrodonint el resultat a dos decimals. El resultat ha de mostrar la mitjana mensual.

Per comprovar el correcte desenvolupament i disseny del cub multidimensional, el document de solució ha d'incloure com a mínim una descripció detallada dels passos següents:

### Disseny del model OLAP

- 1) Creació del projecte en Visual Studio (VS) i revisió de l'origen de dades.<sup>1</sup>
- 2) Creació de la vista d'origen de dades.
- 3) Creació i implementació del cub.
- 4) Configuració de jerarquies, dimensions i atributs.
- 5) Processament i resolució d'errors. Explotació de dades.
- 6) Disseny de cinc consultes per donar resposta a les qüestions plantejades en aquest mateix enunciat.

El document de solució de la PR3 ha d'incloure:

- La descripció de totes les accions portades a terme.
- Les captures de pantalla que demostrin la correcta realització dels apartats enumerats anteriorment.

<sup>1</sup> En crear el projecte en VS i connectar a la base de dades DEST\_loginuoc que cada alumne té assignada, es crearà un cub *default* juntament amb altres objectes que no han de ser esborrats ni utilitzats en la realització de la pràctica.

- Les captures de pantalla que demostrin la correcta definició de les consultes, la visualització del resultat de les explotacions i la seva interpretació correcta.

I, explícitament, aquestes captures de pantalla:

- Test de la connexió, on es vegi el servidor, l'usuari i la base de dades.
- Captura de la implementació completa del projecte des de VS on es vegi el nom de la base de dades de destinació (DEST\_loginuoc) i les dates d'inici i fi del procés.
- En VS, captura dels objectes creats després de la implementació del projecte: *DataSources*, *DataSourceViews*, *Cubes* i *Dimensions*.
- En SQL Management Studio, captura dels objectes creats després de la implementació del projecte: *DataSources*, *DataSourceViews*, *Cubes* i *Dimensions*, on també es vegi el nom de la base de dades corresponent a l'alumne (DEST\_loginuoc).
- Per a cadascuna de les consultes sol·licitades, captura completa on es vegi l'entorn de treball (VS, SSMS, PowerBI) i la sortida de l'execució.

S'ha de tenir en compte que, per a l'avaluació, es podrà considerar, a més del document lliurat amb la solució de la PR3, la implementació en la màquina virtual.

Totes les respostes a les preguntes formulades en l'explotació de dades s'han de fer a partir del cub OLAP implementat en la PR3, independentment de l'eina utilitzada per a la seva anàlisi i explotació, ja sigui el navegador de VS, Power BI o Excel.

No es donaran per vàlides les consultes efectuades directament sobre el model relacional subjacent (*datamart*) en SQL ni en cap altre llenguatge de consulta o anàlisi de dades.

Es disposa d'una plantilla per a la resolució de la PR3.<sup>2</sup>

<sup>2</sup> Recordeu que no heu d'esborrar els objectes creats automàticament en crear i configurar el projecte (base de dades de destinació i vista de dades per defecte).

## Criteris d'avaluació

- Disseny del model OLAP (50 %)
- Explotació de dades (50 %)
  - Consulta 1 (0 %) – Guia de mostra
  - Consulta 2 (12,5 %)
  - Consulta 3 (12,5 %)
  - Consulta 4 (12,5 %)
  - Consulta 5 (12,5 %)

## Guia de mostra

Aquesta guia de mostra pretén servir d'exemple per ajudar a aconseguir els objectius plantejats de la PR3.

### Disseny del model OLAP

El disseny del cub es farà creant un projecte multidimensional i de mineria de dades en VS. S'hauran de definir els orígens de les dades, les vistes dels orígens de dades, els cubs, les relacions d'atributs, les dimensions i les jerarquies necessàries per fer la implementació de la solució i les explotacions sol·licitades en l'enunciat de la PR3.

A continuació es mostra un exemple d'explotació de dades que es pot fer després de la implementació de la solució. L'estudiantat ha d'incloure tots els apartats sol·licitats en aquest enunciat. Per accedir al visor OLAP, s'ha d'entrar en la finestra Explorador del cub creat en el projecte de VS.

Atès que els conjunts de dades (*data sets*) utilitzats contenen mètriques mitjana i percentuals, per a les següents explotacions de dades es prendran com a vàlides la suma de mitjanes, la suma de percentatges i la mitjana de mitjanes, com a aproximacions estadístiques que indiquen un ordre de magnitud, sense entrar a valorar l'exactitud dels resultats obtinguts o estrictament la seva correcció formal.

### **Percentatge de persones que treballen des de casa per país durant l'any 2020, ordenats per la dada corresponent de menor a major.**

Aquesta anàlisi es pot fer utilitzant l'explorador de cubs que proporciona VS. Aquesta eina és una interfície gràfica senzilla que permet fer de manera ràpida consultes no gaire complexes i enriquir-les utilitzant el llenguatge MDX.

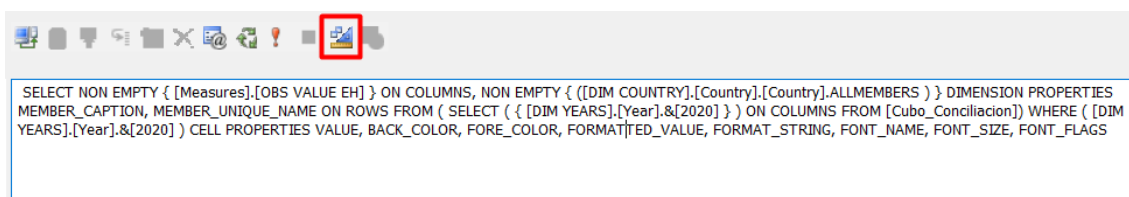
Des de l'àrea de metadades s'arrossegaran a l'àrea d'anàlisi l'atribut de jerarquia «Country» de la dimensió DIM COUNTRY i la mesura OBS VALUE EH. També serà necessari incloure «Year»="2020" en l'àrea de filtres.

Dimensió	Jerarquia	Operador	Expressió de filtre
DIM YEARS	Year	Igual	{ 2020 }
<Seleccionar dimensió>			

Country	OBS VALUE EH
Austria	1291,6
Belgium	1716,1
Bulgaria	18,2
Croatia	439,5
Cyprus	275,2
Czech Republic	671,4
Denmark	1762,8
Estonia	1073,4
Finland	1756,6
France	1342,7
Germany	983,9
Hungary	489,4
Ireland	1349,9
Italy	732,8
Latvia	125,1
Lithuania	344,8
Luxembourg	1868,6

Es pot apreciar que l'ordenació de les dades no és l'esperada. Interessa obtenir el resultat per OBS VALUE EH de menor a major. En aquest cas, s'utilitzarà el llenguatge MDX per enriquir aquesta consulta, en concret la funció ORDER. Per a això, és necessari accedir a l'editor de codi fent clic en l'eina Mode de Disseny.



Aquesta acció mostra una sentència MDX amb la transcripció del disseny actual.

```

SELECT
    NON EMPTY { [Measures].[OBS VALUE EH] } ON COLUMNS,
    NON EMPTY { ([DIM COUNTRY].[Country].[Country].ALLMEMBERS ) } ON ROWS
FROM (
    SELECT ( { [DIM YEARS].[Year].&[2020] } ) ON COLUMNS
FROM [Cubo_Conciliacion])
WHERE ( [DIM YEARS].[Year].&[2020] )

```

Es modificarà aquesta sentència per afegir la clàusula ORDER.

```

SELECT
  NON EMPTY { [Measures].[OBS VALUE EH] } ON COLUMNS,
  NON EMPTY {
    ORDER(( [DIM COUNTRY].[Country].[Country].ALLMEMBERS ),
      [Measures].[OBS VALUE EH],
      ASC ) } ON ROWS
FROM (
  SELECT ( { [DIM YEARS].[Year].&[2020] } ) ON COLUMNS
  FROM [Cubo_Conciliacion]
  WHERE ( [DIM YEARS].[Year].&[2020] )

```

S'obtindrà el resultat amb l'ordre desitjada:

Country	OBS VALUE EH		
Bulgaria	18,2	Spain	828,5
Montenegro	70,7	Poland	829,3
Romania	96,3	Slovenia	946,3
Latvia	125,1	Portugal	980,4
North Macedonia	167,1	Germany	983,9
Cyprus	275,2	Estonia	1073,4
Lithuania	344,8	Malta	1125,7
Croatia	439,5	Austria	1291,6
Slovakia	476,8	France	1342,7
Hungary	489,4	Ireland	1349,9
Serbia	499,7	Belgium	1716,1
Turkey	552,8	Finland	1756,6
Czech Republic	671,4	Denmark	1762,8
Italy	732,8	Luxembourg	1868,6
		Netherlands	2475

En els resultats es pot apreciar la relació de països ordenats de menor a major segons el percentatge de persones que treballen des de casa durant l'any 2020. Es pot observar que Bulgària és el país amb un menor percentatge de persones que treballen a casa, mentre que els Països Baixos és el país amb més valor.



## Format i data de lliurament

El lliurament final d'aquesta activitat es farà a través de l'enllaç «Lliurament PR3» de l'espai *Continguts de l'aula*, adjuntant un únic fitxer amb la solució de la PR3, en format DOCX o PDF. El nom de l'arxiu a enviar estarà format per la composició del nom d'usuari i «\_BDA\_PR3». Per exemple, si el nom d'usuari és «bantich», el nom de l'arxiu ha de ser «bantich\_BDA\_PR3.pdf».

La data màxima de lliurament és el 08/01/2024 a les 23.59 hores.