

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

## PR1 – Anàlisi i disseny del magatzem de dades

### Índex

<b>Presentació</b> .....	2
<b>Descripció</b> .....	3
<b>Criteris d'avaluació</b> .....	4
<b>Guia de mostra</b> .....	5
<b>1. Anàlisi dels requisits</b> .....	6
<b>2. Anàlisi de les fonts de dades</b> .....	7
2.1. Estimació de volumetria.....	7
<b>3. Anàlisi funcional</b> .....	8
<b>4. Disseny del model conceptual, lògic i físic del magatzem de dades</b> .....	9
4.1. Disseny conceptual.....	9
4.2. Disseny lògic .....	10
4.3. Disseny físic .....	11
<b>Format i data de lliurament</b> .....	14

## Presentació

La PR1 consisteix en l'anàlisi i el disseny multidimensional d'un magatzem de dades (*data warehouse*, DW) que faciliti informació rellevant; concretament referent a la conciliació de la vida laboral i familiar (*work-life balance*), per a qui vulgui estudiar el mercat laboral.

Per dur a terme el seu desenvolupament, s'ha publicat la informació següent:

- **Cas pràctic** (22.410\_Enunciat\_PR): document que descriu el context, las persones usuàries potencials i les fonts de dades origen d'un cas d'ús orientat a la creació d'un DW per a l'anàlisi de dades sobre conciliació de la vida laboral i familiar. A més de la descripció general del cas, detalla les fases o etapes que cal seguir per a la construcció d'aquest DW, així com els programes que s'han d'utilitzar i la bibliografia d'interès.
- **Fonts de dades** (fonts.zip): fitxer comprimit amb les fonts de dades disponibles per desenvolupar el cas pràctic.

## Descripció

A partir de l'anàlisi del context del cas i de les fonts de dades disponibles, l'estudiantat individualment ha de dissenyar i proposar un magatzem de dades que ofereixi suport a l'anàlisi de la conciliació de la vida laboral i familiar o *work-life balance*.

Per a això, **es demana** que es duguin a terme els punts següents:

**1) Anàlisi dels requisits** que inclogui les preguntes a les quals el sistema ha de donar resposta. S'han de definir un mínim de cinc preguntes i un màxim de deu.

**2) Anàlisi de totes les fonts de dades** proporcionades. Per cada font, s'elaborarà una taula amb la identificació de camps, descripcions, tipus de camp i un exemple. Addicionalment, es completarà l'anàlisi mitjançant observacions que permetin identificar les fonts i els camps que incloguin llistes de valors.

**3) Anàlisi funcional** que, d'una banda, inclogui els requisits funcionals per al disseny d'una factoria d'informació, establint la prioritat entre exigible (E) i desitjable (D), i que, de l'altra, proposi el tipus d'arquitectura per a la factoria d'informació que s'adeqüi millor al projecte.

**4) Disseny del model conceptual, lògic i físic** del DW. S'han d'identificar i dissenyar les taules de fets (*facts*), les dimensions de l'anàlisi (*dimensions*) i els atributs que permetin tenir el nivell de granularitat suficient per implementar els requisits.

El document de la solució d'aquesta activitat (PR1) ha d'incloure una descripció i una justificació de totes les accions portades a terme en els punts enumerats anteriorment.

## Criteris d'avaluació

La nota final es calcularà a partir de la suma dels apartats següents:

- 1) Anàlisi dels requisits (10 %)
- 2) Anàlisi de fonts de dades (15 %)
- 3) Anàlisi funcional (15 %)
- 4) Matriu de dimensions o mètriques més disseny conceptual (20 %)
- 5) Disseny lògic (20 %)
- 6) Disseny físic (20 %)

## Guia de mostra

A continuació, trobareu una guia creada amb la finalitat d'ajudar a aconseguir els objectius plantejats de la PR1. Aquesta guia conté cadascun dels apartats sol·licitats en la PR1 sobre un dels fets del context del nostre cas d'ús.

### Important

La guia no està completa i l'estudiantat haurà de completar-la per cobrir els objectius marcats per la PR1.

# 1. Anàlisi dels requisits

L'anàlisi dels requisits es basa a identificar les necessitats específiques que té una organització particular respecte a l'anàlisi de la informació. Normalment, en aquesta fase s'ha de ser previsor i pensar més enllà de les necessitats actuals per poder cobrir les futures.

La necessitat principal de l'organització encarregada de l'anàlisi de la conciliació de la vida laboral i familiar és disposar de la informació integrada per a la seva anàlisi i la seva difusió posterior mitjançant les eines d'intel·ligència de negoci. Aquestes eines ajudaran a facilitar la presa de decisions a totes les persones usuàries potencials per garantir el compliment, entre altres, de l'objectiu següent: analitzar l'evolució de la conciliació de la vida laboral i familiar.

A continuació s'indica la informació necessària identificada per analitzar-ho des de diferents perspectives:

- per any,
- per país,
- per sexe,
- per edat,
- per sector productiu,
- per tipus de contracte,
- per situació laboral,
- per nombre de fills,
- per edat dels fills.

Si es té en compte tota aquesta informació, el sistema podrà respondre moltes preguntes i, d'aquesta manera, aconseguirà cobrir les necessitats de les persones usuàries potencials.

Com a exemple, s'indiquen de manera específica algunes preguntes que el sistema ha de ser capaç de respondre com a mínim:

- Percentatges de persones que treballen des de casa, ordenats per la dada corresponent de menor a major.
- Mitjana d'hores treballades per país de la UE, comparada amb la mitjana anual dels EUA.

L'estudiantat, en aquest punt, haurà de completar la definició dels requisits. Per a això, haurà de completar els objectius descrits en la guia amb uns altres que permetin cobrir la necessitat principal de l'organització encarregada de l'anàlisi de la conciliació de la vida laboral i familiar. Així mateix, haurà de plantejar altres perspectives per analitzar aquests objectius i plantejar altres preguntes a les quals el sistema haurà de respondre.

## 2. Anàlisi de les fonts de dades

En aquest apartat, s'han de revisar les fonts de dades proporcionades, el tipus d'informació que contenen, el seu format i les dades que han de ser carregades. A continuació podeu veure una anàlisi detallada per cada tipus de format.

A continuació es mostra un exemple:

**1) CountryList.json.** Conté els noms dels països, els elements de codi ISO 3166-1 alpha-2 i la regió a la qual pertanyen en format JSON. L'estructura del fitxer és la següent:

Nom de camp	Descripció	Tipus	Exemple
name	Nom de país	Text	'Spain'
code	Codi	Text	'ES'
region	Regió	Text	'Southern Europe'

Total de registres: 79

### 2.1. Estimació de volumetria

En els projectes de disseny de factoria d'informació corporativa hi ha una primera fase en la qual es fa una càrrega inicial i, *a posteriori*, una segona fase per fer les càrregues incrementals de les dades noves que van arribant. Una possible estimació del volum de dades del magatzem per a la càrrega inicial de les dades seria la següent:

Fitxer	Registres	Valors	Dades
CountryList.json	79	3	237
...	...	...	...
Total			

En aquest punt, l'estudiantat ha de completar la definició de la resta de les fonts proporcionades i l'estimació de volumetria per a la càrrega inicial de totes les fonts de dades proporcionades.

### 3. Anàlisi funcional

A continuació es proposa el tipus d'arquitectura per a la factoria d'informació que s'adequa millor al projecte. Per a això, es consideren els requisits funcionals i s'estableix la prioritat entre exigible (E) i desitjable (D). En el context d'aquesta activitat, els requisits exigibles són aquells que es demanen en l'enunciat, mentre que els desitjables són els que complementen l'activitat.

A més, en termes de l'escala de prioritats, s'assigna una prioritat de l'1 al 3, en què 1 és completament prioritari per a l'activitat i 3 és no prioritari.

A continuació es descriuen alguns dels requisits funcionals per al disseny d'una factoria d'informació per a l'organització tenint en compte les consideracions de l'enunciat:

#	Requisit	Prioritat	Exigible / desitjable
1	S'extraurà de forma adequada la informació de les fonts de dades.	1	E
2	Es crearà un DW.	1	E
...	...		

En aquest punt, l'estudiantat ha de:

- 1) Completar la taula dels requisits funcionals i assignar les prioritats.
- 2) Triar l'arquitectura funcional que consideri més adequada per al cas d'estudi (triar una arquitectura funcional de totes les estudiades en els mòduls teòrics de l'assignatura i descriure'n els elements).
- 3) Identificar altres requisits funcionals.



## 4. Disseny del model conceptual, lògic i físic del magatzem de dades

### 4.1. Disseny conceptual

Per al correcte desenvolupament del DW, cal definir els fets (*facts*), les dimensions d'anàlisi (*dimensions*), les mètriques i els atributs que permetin tenir el nivell de granularitat suficient per a la presentació dels objectius. Aquests objectius s'han definit en l'anàlisi dels requisits i de les fonts de dades.

Per tant, per fer el disseny conceptual, ens serà de gran utilitat representar la matriu de dimensions i mètriques. La matriu de dimensions i mètriques (Kimball, *Data Warehouse Toolkit*, 2013) és una eina clau en el disseny del DW que representa els processos centrals de l'organització i la dimensionalitat associada. Aquest model busca proporcionar la perspectiva necessària per garantir que tota l'empresa pugui integrar les seves dades a l'entorn del DW.

Partint de l'anàlisi de requeriments i disseny del DW, aquesta eina ha de servir d'ajuda per a la creació de processos d'extracció, transformació i càrrega (ETL) que permetran la càrrega del DW. La seva finalitat és ajudar a organitzar les idees de manera que es pugui tenir una visió completa sobre com es relacionen aquests processos i quines transformacions són necessàries per arribar al model dimensional buscat.

Procés	FACT_PCT_EMPLOYEES_HOME
	Anàlisi del percentatge de persones que treballen a casa (STG_PCT_EMPLOYEES_HOME)
Dimensió/Mètrica	Percentatge de persones que treballen a casa (OBS_VALUE)
DIM_WORKTIME	N/A
...	...

Una vegada plasmada la matriu de dimensions i mètriques, de l'anàlisi de les fonts de dades i dels requisits inicials, es pot determinar que un dels fets que s'han de considerar és el següent:

- **L'evolució de la conciliació de la vida laboral i familiar.** Fa referència a la informació rellevant sobre indicadors relatius al *work-life balance*.

L'anàlisi de l'evolució d'indicadors relatius a conciliació de la vida laboral i familiar determina el disseny de la primera taula de fets, com es pot observar a continuació:

Taula de fets	Descripció
FACT_PCT_EMPLOYEES_HOME	Anàlisi del percentatge de persones que treballen a casa

En la següent taula, s'indica la mètrica de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME.

Mètriques	Descripció
OBS_VALUE	Percentatge de persones que treballen a casa

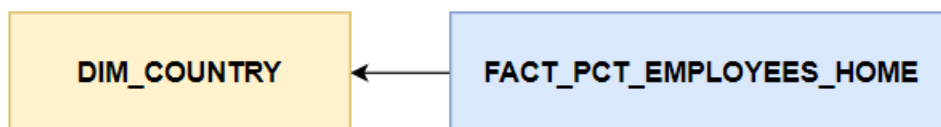
**Nota:** les taules de fets poden tenir una o diverses mètriques.

La mètrica d'aquesta taula de fets pot ser analitzada des de les diferents perspectives a partir de les dimensions. Una dimensió és DIM\_COUNTRY, que permet analitzar el percentatge de persones que treballen a casa des de la perspectiva geogràfica.

L'estudiantat ha de completar la taula de dimensions

Dimensions	Descripció
País	País de l'empresa on treballa cada persona
...	

Amb les dimensions definides en la guia, en la imatge es mostra una part del diagrama corresponent al disseny conceptual de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME.



L'estudiantat ha de completar el disseny conceptual de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME i el de la resta de les taules de fets identificades per a aquest cas d'ús.

## 4.2. Disseny lògic

Una vegada obtingut el model conceptual del DW per a l'anàlisi d'indicadors de la conciliació de la vida laboral, **passem a elaborar-ne el disseny lògic**.

A continuació es mostra, a tall d'exemple, una taula amb alguna de les mètriques identificades en el disseny conceptual de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME:

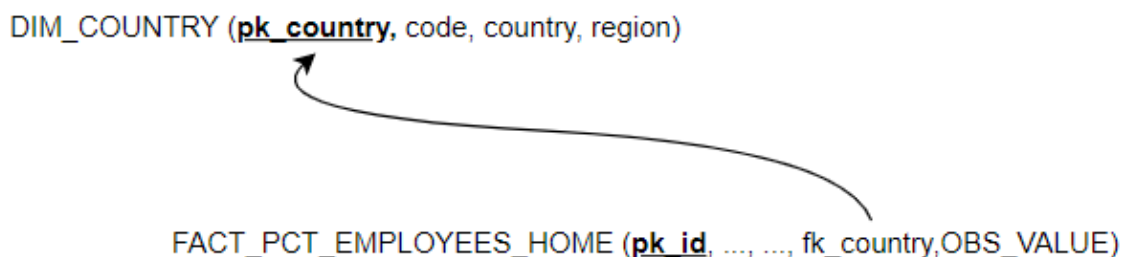
Taula de fets	Mètriques
FACT_PCT_EMPLOYEES_HOME	OBS_VALUE

Després es detallen alguns dels atributs descriptors de les dimensions de cada fet:

Dimensions	Atributs
DIM_COUNTRY	Codi del país, nom del país, regió
...	

L'estudiantat ha de completar la taula amb els atributs descriptors de les dimensions.

Part de la representació visual del **model lògic** definit en la guia per l'anàlisi de la conciliació laboral i familiar és:



L'estudiantat ha de completar el disseny lògic de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME i de la resta de les taules de fets identificades per a aquest cas d'ús.

### 4.3. Disseny físic

Una vegada que s'han determinat les taules de fets, les dimensions, les mètriques i els atributs que hi ha en el model lògic, podem passar a elaborar el disseny físic, la qual cosa significa obtenir una implementació del model lògic en termes del sistema gestor de bases de dades triat.

En aquest punt, l'estudiantat hauria de reflexionar sobre les consideracions prèvies que cal tenir en compte per al disseny físic.

En aquesta guia indiquem, com a exemple, alguns dels aspectes que cal tenir en compte:

- El **sistema gestor de bases de dades** amb el qual treballarem implementarà d'una manera concreta els diferents elements del model lògic.

L'estudiantat ha de completar la resta dels aspectes que cal tenir en compte per al correcte disseny físic de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME i de la resta de les taules de fets identificades per a aquest cas d'ús.

Per a això, detallarem els tipus de dades de cada camp que formen part de les taules de fets i dimensions.

## Dimensions

Les dimensions del model podran estar referenciades en les taules de fets utilitzant les seves claus primàries, o, en anglès, *primary keys* (PK). El model físic de les dimensions identificades és el següent:

- **DIM\_COUNTRY**: conté les dades dels països.

Nom de camp	Tipus	Mida	Exemple
<b>pk_country (PK)</b>	Numèric	8	1
code	Text	8	ES
country	Text	50	Spain
region	Text	75	Southern Europe

- ...

L'estudiantat ha de completar la definició de les dimensions del disseny físic de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME i de la resta de les taules de fets identificades per a aquest cas d'ús.

## Taules de fets

La composició del model físic de les taules de fets consistirà en la creació de taules els camps de les quals seran les mètriques, els atributs i els atributs referencials definits en el model conceptual i en el model lògic. Per crear els atributs referencials en les taules de fets, es defineixen com a claus foranes les primàries de les dimensions amb les quals estan relacionades, seguint el diagrama en estrella definit.

El model físic de les taules de fets del magatzem de dades per a l'anàlisi de la conciliació de la vida laboral i familiar està compost, entre altres, de les taules següents:

- **FACT\_PCT\_EMPLOYEES\_HOME**. És la taula física que contindrà la informació que permetrà fer l'anàlisi de la conciliació de la vida laboral i familiar. Entre altres, tindrà els camps següents:

Nom de camp	Tipus	Mida	Exemple
<b>pk_id (PK)</b>	<b>Sencer</b>	10	1
fk_country (FK)	Sencer	8	2
OBS_VALUE	Numèric	12,2	9.70
.....			

Atès que això és una guia per ajudar l'estudiantat a portar a terme el lliurament complet de l'activitat PR1, haurà de completar tant la definició del model físic de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME com la de la resta de les taules de fets identificades i les seves dimensions.

## Format i data de lliurament

El lliurament final d'aquesta activitat es farà a través de l'enllaç «Lliurament PR1» de l'espai *Continguts de l'aula*, adjuntant un únic fitxer amb la solució de la PR1, en format DOCX o PDF. El nom del fitxer a enviar estarà format per la composició del nom d'usuari i «\_BDA\_PR1». Per exemple, si el nom d'usuari és «bantich», el nom de l'arxiu ha de ser «bantich\_BDA\_PR1.pdf».

La data màxima de lliurament és el 20/11/2023 a les 23.59 hores.