

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

## Solució PR1 – Anàlisi i disseny del magatzem de dades

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 (*data warehouse*, DW) per a l'anàlisi de la conciliació de la vida laboral i familiar o *work-life balance*.

### Índex

<b>1. Anàlisi dels requisits.....</b>	<b>2</b>
<b>2. Anàlisi de les fonts de dades.....</b>	<b>4</b>
2.1. Estimació de volumetria.....	7
<b>3. Anàlisi funcional.....</b>	<b>8</b>
<b>4. Disseny del model conceptual, lògic i físic del magatzem de dades.....</b>	<b>10</b>
4.1. Disseny conceptual.....	10
4.2. Disseny lògic.....	13
4.3. Disseny físic.....	15

# 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 o *work-life balance* és disposar de la informació integrada per a la seva anàlisi i 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 dels objectius següents, entre altres:

- Analitzar l'evolució de la conciliació de la vida laboral i familiar mitjançant el percentatge de persones que treballen a casa.
- Analitzar la mitjana d'hores treballades a la Unió Europea.

El disseny d'un DW d'un projecte inclou la creació i implementació d'un model dimensional o multidimensional, el disseny i la implementació de processos d'extracció, transformació i càrrega (ETL) i del model OLAP i, finalment, el disseny de les consultes establertes en l'enunciat.

A continuació s'indica la informació necessària identificada:

1) Analitzar l'evolució de la conciliació de la vida laboral i familiar mitjançant el percentatge de persones que treballen a casa:

- per any,
- per tram d'edat,
- per país,
- per sexe,
- per nombre de fills,
- per tram d'edat dels fills.

2) Analitzar la mitjana d'hores treballades a la Unió Europea:

- per any,
- per tram d'edat,
- per país,
- per sexe,
- per sector productiu,
- per situació laboral (*work estatus*),
- per temps treballat (*work time*).

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.

A continuació s'indiquen de manera específica les 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.
- TOP-5 de sectors productius amb més hores treballades de mitjana durant el 2021.
- 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).
- Percentatges d'hores treballades a casa segons el nombre de fills. El resultat es mostrarà arrodonit a 1 decimal.
- 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.

## 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.

**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) Ifst\_hhwahchi\_linear.csv.** Conté la informació relativa al percentatge de persones que treballen a casa.

Nom de camp	Descripció	Tipus	Exemple
DATAFLOW	Codi relatiu a l'origen de dades del fitxer	Text	'ESTAT:LFST_HHWA HCHI(1.0)'
LAST_UPDATE	Data d'actualització del registre	Data-hora	'15/02/23 23:00:00'
freq	Codi de freqüència de càlcul de l'indicador	Text	'A'
sex	Codi de gènere de la persona treballadora	Text	'F'
age	Tram d'edat	Text	'Y18-24'
n_child	Nombre de fills	Sencer	1
agechild	Tram d'edat dels fills	Text	'Y6-11'
unit	Codi de la unitat	Text	'PC'
geo	Codi geogràfic	Text	'ES'
TIME_PERIOD	Any de l'indicador	Sencer	2019
OBS_VALUE	Percentatge de persones treballadores a casa	Numèric	10.8
OBS_FLAG	Indicador de seguiment	Text	'b'

Total de registres: 12.884

3) **lfsa\_ewhun2\_linear.csv**. Conté la informació relativa a la mitjana d'hores treballades.

Nom de camp	Descripció	Tipus	Exemple
DATAFLOW	Codi relatiu a l'origen de dades del fitxer	Text	'ESTAT:LFSA_EWHUN2(1.0)'
LAST_UPDATE	Data d'actualització del registre	Data-hora	'15/02/23 23:00:00'
freq	Codi de freqüència de càlcul de l'indicador	Text	'A'
nace_r2	Sector de l'ofici (agricultura...)	Text	'A'
wstatus	Estat de la persona treballadora	Text	'CFAM'
worktime	Horari treballat	Text	'FT'
age	Tram d'edat	Text	'Y15-24'
sex	Codi de gènere de la persona treballadora	Text	'F'
unit	Codi de la unitat	Text	'HR'
geo	Codi geogràfic	Text	'ES'
TIME_PERIOD	Any de l'indicador	Sencer	2019
OBS_VALUE	Promig hores treballades	Numèric	35.1
OBS_FLAG	Indicador de seguiment	Text	'b'

Total de registres: 480.552

4) **BLS\_US\_weeklyhours.xlsx**. Conté informació relativa a la mitjana d'hores treballades en EUA.

Nom de camp	Descripció	Tipus	Exemple
Year	Any de l'indicador	Text	'2015'
Period	Mes de l'indicador	Text	'M10'
Label	Any i mes de l'indicador	Text	'2015 Oct'
Observation Value	Mitjana d'hores treballades	Numèric	34.5

Total de registres: 98

5) **ESTAT\_AGE\_en.tsv**. Conté informació relativa al tram d'edat en un fitxer de format tsv.

Nom de camp	Descripció	Tipus	Exemple
code	Codi de tram d'edat	Text	'Y105'
description	Descripció del tram d'edat	Text	'105 years'

Total de registres: 654

**6) ESTAT\_N\_CHILD\_en.tsv.** Conté informació relativa al nombre de fills de les persones que treballen en un fitxer de format tsv.

Nom de camp	Descripció	Tipus	Exemple
code	Codi de nombre de fills	Text	'GE1'
description	Descripció del nombre de fills	Text	'1 child or more'

Total de registres: 20

**7) ESTAT\_SEX\_en.tsv.** Conté informació relativa al sexe de les persones que treballen en un fitxer de format tsv.

Nom de camp	Descripció	Tipus	Exemple
code	Codi del sexe	Text	'F'
description	Descripció del sexe	Text	'Females'

Total de registres: 7

**8) ESTAT\_NACE\_R2\_en.tsv.** Conté informació relativa al sector productiu de les persones que treballen en un fitxer de format tsv.

Nom de camp	Descripció	Tipus	Exemple
code	Codi del sector productiu	Text	'ENV'
description	Descripció del sector	Text	'Environment'

Total de registres: 1.330

**9) ESTAT\_WORKTIME\_en.tsv.** Conté informació relativa a la quantitat d'hores que treballen a casa les persones en un fitxer de format tsv.

Nom de camp	Descripció	Tipus	Exemple
code	Codi de la quantitat d'hores treballades a casa per la persona	Text	'PC1-24'

description	Descripció de la quantitat d'hores treballades a casa	Text	'From 1 to 24 percent of a full-time'
-------------	---	------	---------------------------------------

Total de registres: 31

**10) ESTAT\_WSTATUS\_en.tsv.** Conté informació de la situació laboral de les persones que treballen en un fitxer de format tsv.

Nom de camp	Descripció	Tipus	Exemple
code	Codi de la situació laboral de la persona treballadora	Text	'POP'
description	Descripció de la situació laboral de la persona treballadora	Text	'Population'

Total de registres: 72

## 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
lfst_hhwahchi_linear.csv	12.884	12	154.608
lfsa_ewhun2_linear.csv	480.552	13	6.247.176
BLS_US_weeklyhours.xlsx	98	4	392
ESTAT_AGE_en.tsv	654	2	1.308
ESTAT_N_CHILD_en.tsv	20	2	40
ESTAT_SEX_en.tsv	7	2	14
ESTAT_NACE_R2_en.tsv	1.330	2	2.660
ESTAT_WORKTIME_en.tsv	31	2	62
ESTAT_WSTATUS_en.tsv	72	2	144
Total	<b>495.727</b>	44	<b>6.406.641</b>

### 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 demana 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 la qual 1 és completament prioritari per a l'activitat i 3 és no prioritari.

A continuació es descriuen els 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 manera adequada la informació de les fonts de dades.	1	E
2	Es crearà un magatzem de dades.	1	E
3	Es carregarà la informació en el magatzem de dades per fer l'anàlisi de la conciliació de la laboral i familiar.	1	E
4	Es crearà un model OLAP per a consultes multidimensionals de les persones usuàries.	2	E
5	Es crearan els informes estàtics sol·licitats.	2	E
6	Es redactarà un manual de càrrega de dades inicial i incremental.	3	D

Cal comentar que, en un cas genèric real, es poden trobar també altres requisits funcionals, com els que es mostren a continuació:

- anàlisi de viabilitat i anàlisi de riscos;
- creació de processos de qualitat de dades;
- creació d'aparadors de dades (*datamarts*) (si s'analitzen altres àrees);
- creació de processos de càrregues incrementals;
- creació d'un repositori de metadades de gestió del magatzem de dades i dels processos d'ETL, que permeti fer la traçabilitat al llarg del cicle de vida de les dades.

Així mateix, atès que aquests sistemes formen part sovint de la implementació d'un sistema d'intel·ligència de negoci, la llista de requisits funcionals seria molt més llarga, com ho pot ser l'administració de seguretat quant a dades i persones usuàries.

En termes de l'arquitectura funcional, existeixen els elements següents:

- Les fonts de dades de les quals es disposa són les següents:
  - un fitxer en format JSON,

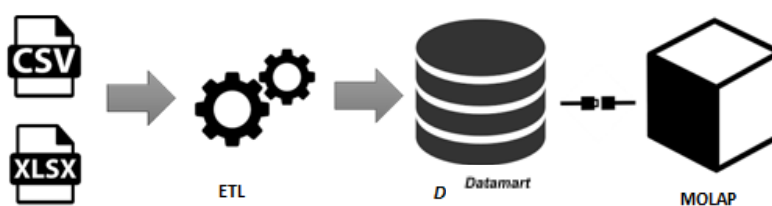


- dos fitxers en format CSV,
  - un fitxer en format XLSX,
  - sis fitxers en format TSV.
- L'arquitectura de la factoria d'informació pot estar formada per diversos elements allotjats en la mateixa màquina:
    - Àrea intermèdia (*staging area*) (opcional): en el cas de tenir moltes fonts (fitxers, bases de dades, serveis RSS, etc.), és convenient carregar-les per consolidar la informació en una estructura de càrrega intermèdia, que pot ser creada en la mateixa base de dades.
    - Aquesta àrea del DW també pot servir per entendre, simplificar i consolidar els processos d'ETL.
    - *Datamart* per a l'anàlisi dels indicadors de conciliació de la vida laboral i familiar. En centrar-nos en una única àrea temàtica, és més correcte considerar que s'està creant un *datamart* en lloc d'un DW corporatiu.
    - MOLAP: a partir de la informació de l'aparador de dades, es crearà un cub multidimensional.

Segons el que s'ha explicat anteriorment, es podria triar entre dos dissenys per a l'arquitectura funcional. D'una banda, es pot considerar una arquitectura funcional que usi una àrea intermèdia i que es creï dins de la base de dades, els objectes de la qual s'identificaran mitjançant un prefix en els noms. La figura següent resumeix els elements de l'arquitectura necessaris per a aquesta activitat:



D'altra banda, també seria correcte utilitzar una arquitectura sense àrea intermèdia (*staging area*) que identifiqui les taules intermèdies en el *datamart* amb un prefix en el nom, com, per exemple, IN\_nombre\_tabla\_intermedia.



En aquesta solució es proposa un disseny que utilitza una àrea intermèdia que, en tenir una única base de dades, simulem amb els prefixos en els noms de les taules intermèdies, els objectes de les quals s'identificaran amb un prefix en els noms (IN\_).

## 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 resultats. Aquests resultats s'han definit en l'anàlisi de requisits i de les fonts de dades.

Per tant, per elaborar 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 magatzem de dades, 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 ETL que permetran la càrrega del magatzem de dades. 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.

	FACT_PCT_EMPLOYEES_HOME	FACT_HOURS_WORKED
Procés	Anàlisi del percentatge de persones que treballen a casa (STG_PCT_EMPLOYEES_HOME)	Anàlisi de la mitjana d'hores treballades (STG_HOURS_WORKED)
Dimensió/Mètrica	Percentatge de persones que treballen a casa (OBS_VALUE)	Hores treballades (OBS_VALUE) Mitjana d'hores treballades als EUA (OBS_VALUE_USA)
DIM_WORKTIME	N/A	X
DIM_WORKSTATUS	N/A	X

<b>DIM_COUNTRY</b>	X	X
<b>DIM_SECTOR</b>	N/A	X
<b>DIM_SEX</b>	X	X
<b>DIM_AGE</b>	X	X
<b>DIM_N_CHILD</b>	X	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 els fets que hem de considerar són els següents:

- **evolució del percentatge de les persones que treballen a casa;**
- **evolució de la mitjana d'hores treballades.**

L'anàlisi del percentatge de les persones que treballen a casa determina el disseny de la primera taula de fets, com es pot observar a continuació:

<b>Taula de fets</b>	<b>Descripció</b>
FACT_PCT_EMPLOYEES_HOME	Anàlisi del percentatge de persones que treballen a casa

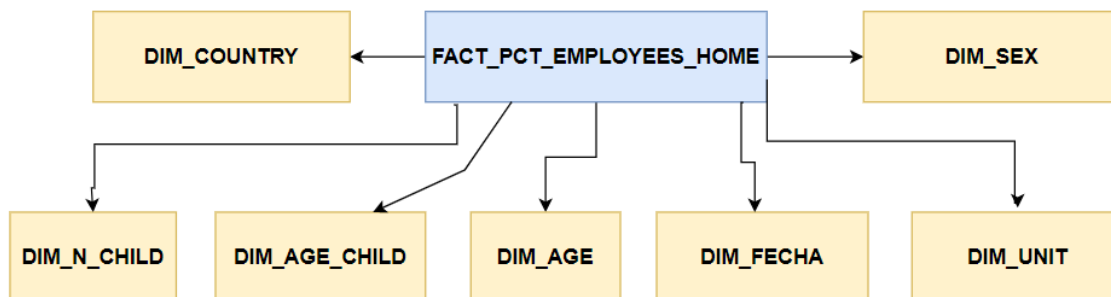
En la taula següent s'indica la mètrica de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME.

<b>Mètriques</b>	<b>Descripció</b>
OBS_VALUE	Percentatge de persones que treballen a casa

La mètrica d'aquesta taula de fets es podrà analitzar des de les diferents perspectives, a partir de les dimensions següents:

<b>Dimensions</b>	<b>Descripció</b>
Temps	Any de registre de l'indicador
País	País de l'empresa on treballa
Sexe	Gènere de la persona
Edat	Tram d'edat de la persona
Nombre de fills	Nombre de fills de la persona
Edat dels fills	Tram d'edat dels fills
Unitats	Unitats de mesura del percentatge de persones que treballen a casa

El disseny conceptual per a aquesta taula de fets (FACT\_PCT\_EMPLOYEES\_HOME) i les seves dimensions amb un **disseny en estrella** és el següent:



Aquest model considera les fonts de dades següents:

- CountryList.json
- Ifst\_hhwahchi\_linear.csv
- ESTAT\_AGE\_en.tsv
- ESTAT\_N\_CHILD\_en.tsv
- ESTAT\_SEX\_en.tsv

D'altra banda, l'**anàlisi de l'evolució de la mitjana d'hores treballades** determinarà el disseny de la segona taula de fets, com es pot observar a continuació:

Taula de fets	Descripció
FACT_HOURS_WORKED	Anàlisi de la mitjana d'hores treballades

En la taula següent s'indica la mètrica de la taula de fets FACT\_HOURS\_WORKED:

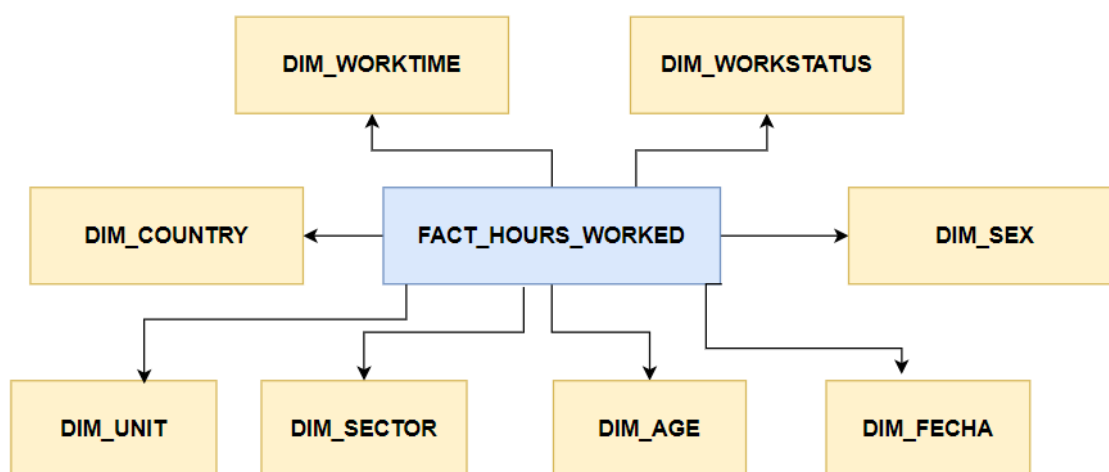
Mètriques	Descripció
OBS_VALUE	Hores treballades
OBS_VALUE_USA	Mitjana d'hores treballades mitjanes als EUA

La mètrica d'aquesta taula de fets es podrà analitzar des de les diferents perspectives, a partir de les dimensions següents (algunes comunes amb la primera taula de fets):

Dimensions	Descripció
Temps	Any de registre de l'indicador
País	País de l'empresa on treballa
Sexe	Gènere de la persona
Edat	Tram d'edat de la persona
Sector productiu	Sector productiu de l'ocupació
Estat de la persona que treballa	Estat actual (en actiu, a l'atur...)
Hores de la persona que treballa	Tipus de contracte o nombre d'hores

Unitats	Unitats de mesura d'hores treballades
---------	---------------------------------------

El disseny conceptual per a aquesta taula de fets (FACT\_HOURS\_WORKED) i les seves dimensions amb un **disseny en estrella** és el següent:



Aquest model considera les fonts de dades següents:

- CountryList.json
- Ifsa\_ewhun2\_linear.csv
- BLS\_US\_weeklyhours.xlsx
- ESTAT\_AGE\_en.tsv
- ESTAT\_SEX\_en.tsv
- ESTAT\_NACE\_R2\_en.tsv
- ESTAT\_WORKTIME\_en.tsv
- ESTAT\_WSTATUS\_en.tsv

## 4.2. Disseny lògic

Una vegada obtingut el model conceptual del magatzem de dades de l'anàlisi d'indicadors de la conciliació de la vida laboral i familiar, es procedeix a elaborar el disseny lògic.

Tenint en compte que utilitzarem tecnologia relacional i que el model de dades serà el multidimensional, es descriu el model lògic en termes de taules, atributs i claus primàries i foranes.

El primer pas correspon a identificar les mètriques de cada fet, que es representarà mitjançant una taula de fets. En el nostre cas, hem identificat aquests dos fets i les seves mètriques:

Taula de fets	Mètriques
FACT_PCT_EMPLOYEES_HOME	Percentatge de persones que treballen a casa
FACT_HOURS_WORKED	Hores treballades

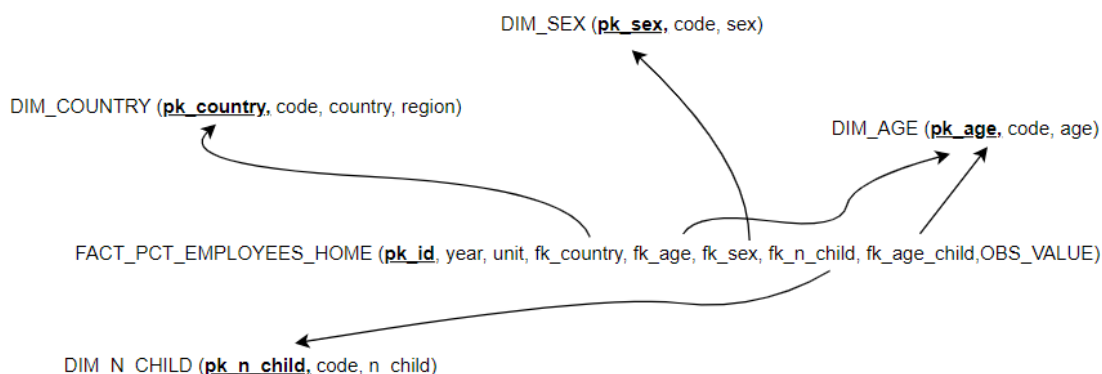
Després es detallen els atributs de les dimensions de cada fet. Específicament, els atributs de les dimensions de la taula de fets FACT\_PCT\_EMPLOYEES\_HOME es mostren en la taula següent:

Dimensions	Atributs
DIM_COUNTRY	code, country, region
DIM_SEX	code, sex
DIM_AGE	code, age
DIM_N_CHILD	code, n_child

A més, la dimensió de DIM\_AGE servirà també per al camp AGE\_CHILD del fitxer d'origen.

D'altra banda, cal comentar que les dimensions de *data* i *unitats* especificades en el model conceptual no requeriran crear cap taula de dimensió individual en no tenir atributs ni jerarquies d'agregació, però ens ajudaran a identificar les instàncies del fet. Per tant, estarem davant dues dimensions degenerades.

La representació visual del model lògic per a l'anàlisi de l'**evolució del percentatge de persones que treballen a casa** seria la següent:

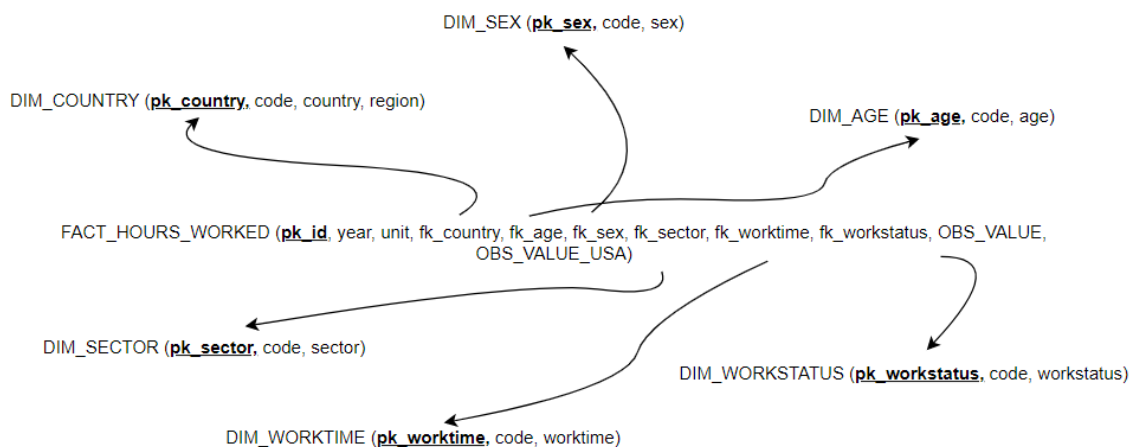


Procedim igual amb el fet FACT\_HOURS\_WORKED. En la taula següent es mostren els atributs de les dimensions d'aquesta taula de fets:

Dimensions	Atributs
DIM_COUNTRY	code, country, region
DIM_SEX	code, sex
DIM_AGE	code, age
DIM_SECTOR	code, sector
DIM_WORKTIME	code, worktime
DIM_WORKSTATUS	code, workstatus

De la mateixa forma que per a l'altra taula de fets, les dimensions de *data* i *unitats* especificades en el model conceptual no requeriran crear cap taula de dimensió individual en no tenir atributs ni jerarquies d'agregació, però ens ajudaran a identificar les instàncies del fet. Per tant, estarem davant dues dimensions degenerades.

La representació visual del **model lògic** de la taula de fets i les seves dimensions per a l'**anàlisi de l'evolució d'hores treballades** és la següent:



## 4.3. Disseny físic

Una vegada que s'han determinat les taules de fets, les dimensions, les mètriques i els atributs que existeixen 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.

A més, per al correcte disseny físic del magatzem, s'han de tenir en compte els aspectes següents:

- el **sistema gestor de bases de dades** amb el qual es treballarà, ja que implementarà d'una manera concreta els diferents elements del model lògic;
- l'ajust del **disseny físic** a les particularitats del nostre sistema gestor de bases de dades, amb la finalitat d'obtenir un bon rendiment en el processament de consultes;
- la **revisió periòdica del disseny físic inicial**, per validar que continua donant resposta a les necessitats del client.

Com que utilitzarem SQL Server, i aquest és un sistema gestor de bases de dades relacional, en aquesta etapa haurem de considerar, entre altres coses, la implementació de les claus primàries i foranes en les taules de fets i en les de dimensions.

En aquest pas, també és necessari tenir en compte la grandària adequada dels atributs (per exemple, la longitud dels camps de textos o si els valors numèrics contenen decimals).

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

Atès que el model de magatzem està compost per més d'una taula de fets, també s'han de revisar les dimensions que s'han definit en el disseny conceptual i en el lògic de cada fet i aplicar una visió conjunta del model per determinar si en el model del magatzem hi haurà dimensions comunes o conformades, com DIM\_COUNTRY o DIM\_AGE, entre altres, i, així, simplificar el model final i aconseguir un rendiment òptim en l'execució de les anàlisis.

Com és lògic, primer es creen les taules de dimensions i posteriorment les taules de fets, ja que contenen atributs referencials a aquelles. D'aquesta manera, es crea cadascuna de les taules del magatzem de dades.

## Dimensions

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



- **DIM\_COUNTRY.** Conté les dades dels països. La dimensió és comuna en tot el model dissenyat i permet analitzar els fets des d'un punt de vista geogràfic.

A més, la dimensió conformada DIM\_COUNTRY s'utilitzarà tant per analitzar l'evolució del percentatge de persones que treballen a casa com per a l'evolució de la mitjana d'hores treballades.

Nom de camp	Tipus	Mida	Exemple
<b>pk_country (PK)</b>	Numèric	8	1
code	Text	10	ES
country	Text	50	Spain
region	Text	75	Southern Europe

- **DIM\_AGE.** Conté les dades de trams d'edat. La dimensió és comuna en tot el model dissenyat i la podem usar indistintament en els atributs d'AGE com en els d'AGE\_CHILD.

A més, la dimensió conformada DIM\_AGE s'utilitzarà per analitzar tant l'evolució del percentatge de persones que treballen a casa com l'evolució de la mitjana d'hores treballades.

Nom de camp	Tipus	Mida	Exemple
<b>pk_age (PK)</b>	Numèric	8	1
code	Text	15	GE1
age	Text	60	1 child or more

- **DIM\_SEX.** Conté informació de gènere. La dimensió conformada DIM\_SEX ens permet analitzar si hi ha diferències per sexe en el percentatge de persones que treballen a casa o la mitjana d'hores.

Nom de camp	Tipus	Mida	Exemple
<b>pk_sex (PK)</b>	Numèric	8	1
code	Text	10	F
sex	Text	60	Females

- **DIM\_N\_CHILD.** Conté informació del nombre de fills. Aquesta dimensió ens permetrà analitzar si hi ha diferències en el percentatge d'hores treballades a casa segons el nombre de fills.

Nom de camp	Tipus	Mida	Exemple
<b>pk_n_child (PK)</b>	Numèric	8	1

code	Text	10	GE1
n_child	Text	50	1 child or more

- **DIM\_SECTOR:** conté informació del sector productiu de les ocupacions. Aquesta dimensió ens permetrà analitzar si hi ha diferències entre sectors respecte a la mitjana d'hores treballades.

Nom de camp	Tipus	Mida	Exemple
<b>pk_sector (PK)</b>	Numèric	8	1
code	Text	25	ENV
sector	Text	550	Environment

- **DIM\_WORKTIME.** Conté informació del tipus de contracte. Ens permet analitzar si hi ha diferències en la mitjana d'hores treballades segons el tipus de contracte.

Nom de camp	Tipus	Mida	Exemple
<b>pk_worktime (PK)</b>	Numèric	8	1
code	Text	25	PC1-24
worktime	Text	250	From 1 to 24 percent of a full-time

- **DIM\_WORKSTATUS.** Conté informació de l'estat laboral de les persones que treballen.

Nom de camp	Tipus	Mida	Exemple
<b>pk_workstatus (PK)</b>	Numèric	8	1
code	Text	25	POP
workstatus	Text	250	Population

## Taula de fets

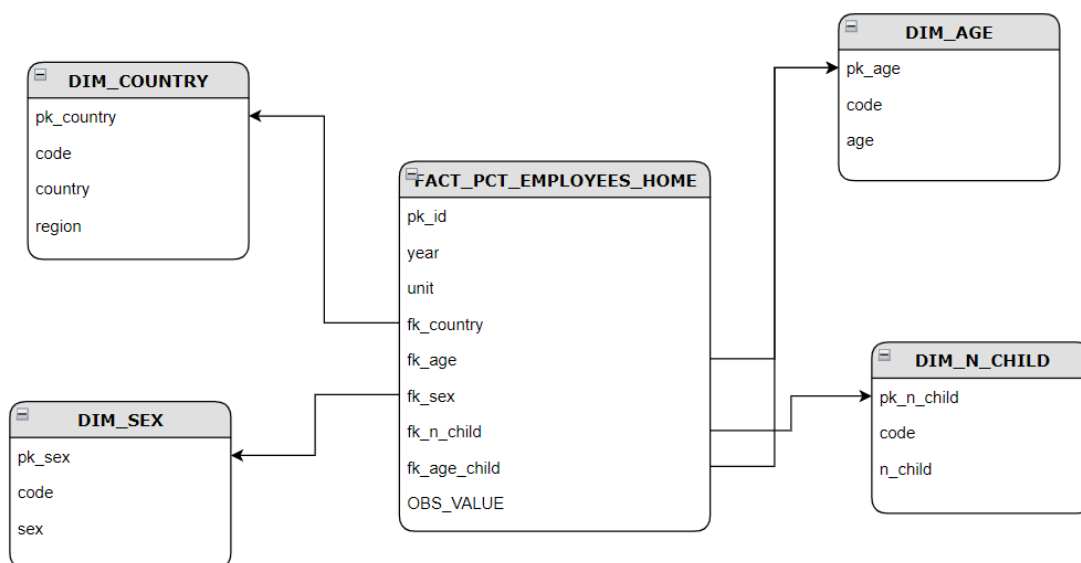
La composició del model físic de les taules de fets consistirà en la creació de taules, els camps de les quals seran les mètriques, els atributs i els atributs referencials definits en el model conceptual i en el model lògic. Per crear els atributs referencials en les taules de fets, es defineixen com 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 de les taules següents:

- **FACT\_PCT\_EMPLOYEES\_HOME.** És la taula física que contindrà la informació que permetrà fer l'anàlisi de l'evolució del percentatge de persones que treballen a casa. Tindrà els camps següents:

Nom de camp	Tipus	Mida	Exemple
pk_id (PK)	Sencer	10	1
year	Sencer	4	2019
unit	Text	25	PC
fk_country (FK)	Sencer	8	2
fk_age (FK)	Sencer	8	1
fk_sex (FK)	Sencer	8	3
fk_n_child (FK)	Sencer	8	5
fk_age_child (FK)	Sencer	8	6
OBS_VALUE	Numèric	12,2	9.70

En la imatge següent es mostra una possible implementació del **disseny del model físic** per a la taula de fets **FACT\_PCT\_EMPLOYEES\_HOME**:



- **FACT\_HOURS\_WORKED.** És la taula física que contindrà la informació que permetrà fer l'anàlisi de la mitjana d'hores treballades. Tindrà els camps següents:

Nom de camp	Tipus	Mida	Exemple
pk_id (PK)	Sencer	10	1
year	Sencer	4	2019
unit	Text	25	HR
fk_country (FK)	Sencer	8	2
fk_age (FK)	Sencer	8	1
fk_sex (FK)	Sencer	8	3
fk_sector (FK)	Sencer	8	5
fk_worktime (FK)	Sencer	8	7
fk_workstatus (FK)	Sencer	8	2
OBS_VALUE	Numèric	12,2	35.10
OBS_VALUE_USA	Numèric	12,2	34.50

En la imatge següent es mostra una possible implementació del **disseny del model físic** per a la taula de fets **FACT\_HOURS\_WORKED**:

