

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

Ignasi Grau Muñoz

1. Anàlisi dels requisits

Uns possibles objectius pel Data Warehouse de la conciliació de la vida laboral i familiar serien els següents:

- Analitzar l'impacte de l'equilibri entre vida laboral i vida familiar a la productivitat de l'empresa o entitat. Poder analitzar si l'increment d'hores dels treballadors implica un increment significatiu de la feina produïda o en el cas contrari, que baixi la productivitat dels empleats i no hi hagi increment de producció, sinó al revés.
- Analitzar l'impacte generacional de la vida laboral i familiar. Poder veure com totes les generacions equilibren la vida laboral i familiar.
- Analitzar l'eficàcia de programes de benestar de les empreses que disposin d'un i poder estudiar l'impacte real dels programes.
- Visió global de conciliació de la vida laboral i familiar entre els diferents països perquè les persones que vulguin emigrar a un altre país puguin consultar.

Preguntes que el sistema ha de ser capaç de respondre com a mínim:

- Comparativa de la conciliació en empreses de diferents mides: Examinar com les polítiques de conciliació varien en funció de la mida de l'empresa, identificant si les petites, mitjanes i grans empreses tenen enfocaments diferents en aquesta àrea.
- Comparació de la conciliació en diferents sectors: Investigar com les pràctiques de conciliació varien entre sectors específics de l'economia, identificant els sectors que lideren en l'adopció de polítiques de conciliació i els que encara estan en una fase incipient.
- Comparativa de les polítiques de conciliació entre sectors públic i privat: Analitzar les diferències en l'adopció i implementació de polítiques de conciliació entre organitzacions del sector públic i privat, incloent-hi les pràctiques més comunes en cada àmbit.
- Anàlisi dels efectes de la conciliació en la igualtat d'oportunitats: Investigar com les pràctiques de conciliació afecten la igualtat d'oportunitats dins de l'organització, especialment en termes d'ascensos, accessos a formació i participació en projectes importants.
- Avaluació dels resultats de la conciliació en la diversitat i inclusió: Examinar com les pràctiques de conciliació contribueixen a la millora de la diversitat i inclusió dins de l'organització, incloent-hi la representació de grups minoritaris en diferents nivells jeràrquics.

2. Anàlisi de les fonts de dades

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

Total de registres: 79

2) BLS_US_weeklyhours.xlsx. Mitjana d'hores treballades setmanals als Estats Units en format excel. Dades obtingudes de US Department of Labor.

Nom del camp	Descripció	Tipus	Exemple
Year	Any	Int	2015
Period	Període	Text	M01
Label	Mes i any	Text	'2015 Jan'
Observation Value	Mitja d'hores semanals de tots els empleats	Int	34,5

Total de registres: 96

3) ESTATE_AGE_en.tsv. Llista de trams d'edats segons classificació de l'Eurostat en format tsv.

Nom del camp	Descripció	Tipus	Exemple
TOTAL	Codi pel rang d'edat	Text	MN1-9
Total	Descripció del rang d'edat	Text	From 1 to 9 minutes

Total de registres: 654

4) ESTATE_WORKTIME_en.tsv. Llista de tipus d'horaris laborals de les persones que treballen segons classificació de l'Eurostat en format tsv.

Nom del camp	Descripció	Tipus	Exemple
TOTAL	Codi pel rang de temps	Text	H1-9
Total	Descripció del rang de temps	Text	From 1 to 9 hours

Total de registres: 31

5) Ifsa_ewhun2_linear.csv. Mitjana d'hores treballades setmanals a la UE per país, sexe, edat, categoria professional i activitat econòmica en format csv.

Nom del camp	Descripció	Tipus	Exemple
DATAFLOW	versió de les dades	text	ESTAT:LFSA_EWHUN2(1.0)
LAST UPDATE	última actualització de la dada	text	15/02/23 23:00:00
freq	frequència de temps	text	A (anual)
nace_r2	activitats laborals	text	A (Agricultura)
wstatus	estats laborals	text	EMP (persona amb feina)
worktime	tipus d'horaris laborals	text	PT (Part-time)
age	edat	text	Y18-24 (From 18 to 24 years)
sex	sexe	text	M (Male)
unit	unitat de mesura	text	HR (hora)
geo	país del treballador	text	BE (Belgium)
TIME_PERIOD	any	int	2022
OBS_VALUE	mitja d'hores treballades setmanalment	text	4.5
OBS_FLAG	?	text	u

Total de registres: 480551

6) lfst_hhwahchi_linear.csv. Percentatge de persones que treballen a casa per país, grup d'edat, nombre i edat de fills en format csv.

Nom del camp	Descripció	Tipus	Exemple
DATAFLOW	versió de les dades	text	ESTAT:LFST_HHW AHCHI(1.0)
LAST UPDATE	última actualització de la dada	text	15/02/23 23:00:00
freq	frecuencia de temps	text	HR (hora)
sex	sexe	text	M (Male)
age	edad	text	Y18-24 (From 18 to 24 years)
n_child	número de fills	text	1 (1 child)
agechild	edad del fill	text	Y_LT6 (less than 6 years)
unit	unitat	text	PC (%)
geo	país del treballador	text	BE (Belgium)
TIME_PERIOD	període de temps	text	2022
OBS_VALUE	percentatge d'hores que es teletreballa	text	9.7
OBS_FLAG	?	text	u

Total de registres: 12883

7) ESTAT_N_CHILD_en.tsv. Llista del nombre de fills segons classificació de l'Eurostat en format tsv.

Nom del camp	Descripció	Tipus	Exemple
TOTAL	Codi pel número de fills	Text	2
Total	Número o rang de número de fills	Text	2 children

Total de registres: 20

8) ESTAT_NACE_R2_en.tsv. Llista d'activitats econòmiques segons classificació de l'Eurostat en format tsv.

Nom del camp	Descripció	Tipus	Exemple
TOTAL	Codi pel sector	Text	B091
Total	Descripció de les activitats econòmiques	Text	Support activities for petroleum and natural gas extraction

Total de registres: 1330

9) ESTAT_SEX_en.tsv. Llista de valors relatius al sexe de les persones que treballen segons classificació de l'Eurostat en format tsv.

Nom del camp	Descripció	Tipus	Exemple
T	Codi pel sexe	Text	M
Total	Descripció del sexe	Text	Male

Total de registres: 7

10) ESTAT_WSTATUS_en.tsv. Llista d'estats laborals segons classificació de l'Eurostat en format tsv.

Nom del camp	Descripció	Tipus	Exemple
TOTAL	Codi per la situació laboral	Text	EMP_FT
Total	Situació laboral	Text	Employed persons working full-time

Total de registres: 7

2.1. Estimació de volumetria

Fitxer	Registres	Valors	Dades
CountryList.json	79	3	237
BLS_US_weeklyhours.xlsx	96	4	384
ESTATE_AGE_en.tsv	653	2	1306
ESTATE_WORKTIME_en.tsv	31	2	64
lfsa_ewhun2_linear.csv	480551	13	6247163
lfst_hhwahchi_linear.csv	12883	12	154596
ESTAT_N_CHILD_en.tsv	19	2	38
ESTAT_NACE_R2_en.tsv	1229	2	2658
ESTAT_SEX_en.tsv	6	2	12
ESTAT_WSTATUS_en.tsv	71	2	142
Total:			6.406.600

3. Anàlisi funcional

#	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
3	El sistema ha de permetre la inserció, actualització i eliminació de dades a la base de dades.	2	E
4	Hi ha d'haver un mecanisme per garantir la integritat referencial entre taules.	3	D
5	La base de dades ha de ser escalable per gestionar un augment en el volum de dades i la càrrega de treball.	2	D
6	La base de dades ha d'admetre la integració amb altres aplicacions o sistemes mitjançant interfícies i API.	3	D
7	S'han de definir mecanismes per a la importació i l'exportació de dades des de i cap a la base de dades.	2	E

Proposaria una arquitectura de Business Intelligence amb els següents components:

Data Warehouse:

Extracció de Dades:

Utilitzar eines ETL (Extract, Transform, Load) per extreure i transformar la informació de les fonts de dades.

Creació d'un DW:

Establir un Data Warehouse centralitzat que serveixi com a magatzem de dades consolidat per a les necessitats d'anàlisi.

Manipulació de Dades:

Implementar funcionalitats que permetin la inserció, actualització i eliminació de dades dins del Data Warehouse.

Integritat Referencial:

Establir mecanismes per garantir la integritat referencial entre les taules del Data Warehouse.

Escalabilitat:

Disposar d'una arquitectura escalable per gestionar un creixement en el volum de dades i càrrega de treball. Això pot incloure particionament de taules, clustering i altres tècniques d'optimització de rendiment.

Integració amb Altres Aplicacions:

Implementar interfícies i API que permetin la integració amb altres aplicacions i sistemes. Això facilitarà la interconnexió amb altres components del sistema empresarial.

Importació i Exportació de Dades:

Definir mecanismes per importar i exportar dades des de i cap a la base de dades, facilitant la comunicació amb altres sistemes i la transferència de dades.

Aquesta arquitectura proporcionaria un entorn robust per a la gestió i anàlisi de dades, amb un focus especial en les necessitats del Business Intelligence i la facilitat d'integració amb altres components del sistema.

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

4.1. Disseny conceptual

Taula de fets 1:

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

Taula de dimensions 2:

Mètriques	Descripció
OBS_VALUE	Percentatge de persones que treballen a casa

Dimensions de la taula de fets 1:

Dimensions	Descripció
País	País de l'empresa on treballa cada persona
Edat	Edat del treballador
Edat fills	Edat dels fills del treballador
Nombre de fills	Nombre de fills del treballador
Sexe	Sexe del treballador
Any	Any on es va prendre la mesura

Taula de fets 2:

Taula de fets	Descripció
WEEKLY_HOURS_WORKED	Anàlisi del nombre mitjà d'hores treballades a l'ofici principal

Taula de dimensions 2:

Mètriques	Descripció
OBS_VALUE	Nombre mitjà d'hores setmanals treballades a l'ofici principal

Dimensions de la taula de fets 2:

Dimensions	Descripció
País	País de l'empresa on treballa cada persona
Edat	Edat del treballador
Sector	Sector en el qual treballa l'empleat
Jornada	Tipus de jornada laboral, com per exemple jornada completa o mitja jornada.
Sexe	Sexe del treballador
Status	Situació laboral del treballador
Any	Any on es va prendre la mesura

4.2. Disseny lògic

Fet 1:

Taula de fets	Mètriques
WEEKLY_HOURS_WORKED	OBS_VALUE

Dimensió	Atributs
DIM_COUNTRY	Nom del país, codi del país, regió del país
DIM_AGE	Codi del rang d'edats, rang d'edats
DIM_CHILD_AGE	Codi del rang d'edats, rang d'edats
DIM_N_CHILDS	Codi de nombre de fills, nombre de fills
DIM_SEX	Codi del sexe, sexe
DIM_YEAR	Any

Fet2:

Taula de fets	Mètriques
FACT_PCT_EMPLOYEES_HOME	OBS_VALUE

Dimensió	Atributs
DIM_COUNTRY	Nom del país, codi del país, regió del país
DIM_AGE	Codi del rang d'edats, rang d'edats
DIM_SECT	Codi del sector, nom del sector
DIM_WORK_TIME	Codi del tipus de jornada, tipus de jornada
DIM_SEX	Codi del sexe, sexe
DIM_WORK_STATUS	Codi de la situació laboral de l'empleat, situació laboral de l'empleat

4.3. Disseny físic

Aspectes que cal tenir en compte per al correcte disseny físic de les taules de fets FACT_PCT_EMPLOYEES_HOME i WEEKLY_HOURS_WORKED:

Compatibilitat de l'SGBD:

És fonamental assegurar-se que l'SGBD seleccionat sigui compatible amb els requisits i els objectius del projecte. Alguns sistemes són més adequats per a entorns específics o tipus de dades.

Rendiment i Escalabilitat:

Avaluar la capacitat de l'SGBD per gestionar el volum de dades previst i la capacitat per escalar a mesura que les necessitats de l'organització creixen.

Seguretat i Control d'Accés:

Cal considerar les funcions de seguretat del SGBD per garantir el control d'accés adequat a la informació delicada i la protecció contra possibles amenaces.

Respatller i Recuperació:

Avaluar les capacitats de suport i recuperació del SGBD per garantir la integritat de les dades i la capacitat de restaurar la base de dades en cas de fallades o pèrdues.

Integració amb Altres Tecnologies:

Assegurar-se que l'SGBD pugui integrar-se sense problemes amb altres tecnologies i eines utilitzades en l'organització, com ara eines de visualització de dades, aplicacions empresarials, etc.

Dimensions per a la taula FACT_PCT_EMPLOYEES_HOME:

Nom del camp	Tipus	Mida	Exemple
pk_country	Numèric	8	1
code	Text	5	ES
country	Text	50	Spain
region	Text	80	Southern Europe

Nom del camp	Tipus	Mida	Exemple
pk_age	Numèric	8	1
code	Text	5	Y22
age	Text	50	21 years

Nom del camp	Tipus	Mida	Exemple
pk_age_childs	Numèric	8	1
code	Text	5	Y22
age_childs	Text	50	21 years

Nom del camp	Tipus	Mida	Exemple
pk_n_child	Numèric	8	1
code	Text	5	GE4
n_child	Text	50	children or more

Nom del camp	Tipus	Mida	Exemple
--------------	-------	------	---------

pk_sex	Numèric	8	1
code	Text	5	F
sex	Text	50	Females

Nom del camp	Tipus	Mida	Exemple
pk_year	Numèric	8	1
year	Numèric	4	2022

Dimensions per a la taula WEEKLY_HOURS_WORKED:

Nom del camp	Tipus	Mida	Exemple
pk_country	Numèric	8	1
code	Text	5	ES
country	Text	50	Spain
region	Text	80	Southern Europe

Nom del camp	Tipus	Mida	Exemple
pk_age	Numèric	8	1
code	Text	5	Y22
age	Text	50	21 years

Nom del camp	Tipus	Mida	Exemple
pk_sector	Numèric	8	1
code	Text	5	G451
sector	Text	100	Sale of motor vehicles

Nom del camp	Tipus	Mida	Exemple
pk_work_time	Numèric	8	1

code	Text	5	PC50-74
work_time	Text	100	From 50 to 74 percent of a full-time

Nom del camp	Tipus	Mida	Exemple
pk_sex	Numèric	8	1
code	Text	5	F
sex	Text	50	Females

Nom del camp	Tipus	Mida	Exemple
pk_status	Numèric	8	1
code	Text	5	EMP_FT
status	Text	50	Employed persons working full-time

Taules de fets FACT_PCT_EMPLOYEES_HOME:

Nom del camp	Tipus	Mida	Exemple
pk_id	Sencer	10	1
OBS_VALUE	Numèric	12,2	9.7
fk_country	Sencer	8	2
fk_age	Sencer	8	3
fk_age_childs	Sencer	8	24
fk_n_childs	Sencer	8	6
fk_sex	Sencer	8	22
year	Sencer	4	2022

Taules de fets WEEKLY_HOURS_WORKED:

Nom del camp	Tipus	Mida	Exemple
pk_id	Sencer	10	1

OBS_VALUE	Numèric	12,2	22.5
fk_country	Sencer	8	2
fk_age	Sencer	8	3
fk_sector	Sencer	8	24
fk_work_time	Sencer	8	6
fk_sex	Sencer	8	22
fk_status	Sencer	8	33