

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

Autors: Marc Álvarez Brotons
Carlos Calonge Cebrián
Sergio Chica Lucena
David Díaz Arias

Índex

Introducció	2
Context	3
Persones usuàries potencials	4
Fonts de dades	5
Enunciat	7
1. PR1: Anàlisi i disseny del magatzem de dades	7
2. PR2: Càrrega de les dades	7
3. PR3: Explotació de les dades	8
Programes	9
Bibliografia	10

Introducció

El cas «**Magatzem de dades per a l'anàlisi de la conciliació de la vida laboral i familiar**» ha estat creat per practicar el disseny i la implementació del magatzem de dades (*data warehouse*, DW) com a sistema d'emmagatzematge per a l'anàlisi de dades.

El disseny, el desenvolupament i la implantació d'un sistema de DW en qualsevol organització comporta dur a terme un projecte que pot durar mesos o fins i tot anys, en funció de l'abast del projecte, de la naturalesa i el grau de maduresa de l'organització, així com de la participació d'equips multidisciplinaris que van implementant diferents projectes en un procés de millora contínua del magatzem.

L'objectiu d'aquest cas no és desenvolupar un DW que doni resposta a totes les necessitats, sinó entendre i utilitzar les metodologies per desenvolupar aquest tipus de projectes en un context real, passant per totes les fases que comprenen els projectes d'aquesta tipologia:

- 1) Anàlisi, disseny i implementació.** Consisteix a desenvolupar i implementar un DW que permeti la gestió de la informació disponible.
- 2) Càrrega.** Implica dissenyar i implementar els processos de càrrega de dades necessàries per disposar d'informació en el DW implementat en l'etapa anterior.
- 3) Explotació.** Té objectiu de dur a terme una anàlisi multidimensional d'aquesta informació que en permeti l'explotació.

Amb la finalitat de poder desenvolupar un projecte al més específic possible, l'estudiantat haurà d'afrontar el repte de desenvolupar un DW que descrigui una part dels serveis que es poden oferir, basant-se en dades tractades en el cas i que formarien part d'un sistema real.

A partir del context que es descriu a continuació, l'estudiantat adquirirà un coneixement bàsic de l'entorn tecnològic, comprendrà les necessitats existents i definirà la proposta més adequada que les satisfaci.

Mitjançant el desenvolupament del cas, l'estudiantat es trobarà amb els problemes, els dubtes i les dificultats que es plantegen en un projecte de característiques similars.

Context

A la fi del novembre de 2022 el Ministeri de Treball i Economia d'Espanya va fer una inspecció de treball simultània en els edificis on se situen les grans multinacionals de consultoria i auditoria, conegudes popularment pel nom de **Big Four** (Deloitte, PwC, EY i KPMG). Aquesta inspecció es va fer amb l'objectiu de revisar el control d'horaris al qual estan subjectes totes les companyies espanyoles des del 2019.

La consultoria i l'auditoria és un dels sectors en què es realitzen jornades que poden superar àmpliament vuit hores diàries de mitjana. Aquest sector no és l'únic en el qual moltes companyies creen una cultura laboral, que no imposen explícitament, en la qual es premia la dedicació excessiva excusant-se en diversos factors com la competitivitat o la possibilitat d'ascendir.

D'altra banda, com a conseqüència dels efectes de la pandèmia de la COVID-19 i la millora en les telecomunicacions, s'ha observat que una part important de l'activitat laboral que antigament es feia en oficina es pot fer perfectament a casa amb resultats iguals o millors quant a productivitat i permet, a més, una millor conciliació de la vida laboral i familiar (*work-life balance*) per a les persones treballadores.

No obstant això, moltes empreses continuen mostrant un gran recel davant el model del teletreball i han començat a sol·licitar a les persones que hi treballen que tornin a l'oficina presencial a temps complet o parcial. Paral·lelament, s'ha demostrat que una part important de les persones que teletreballen continuen treballant les mateixes hores i fins i tot més que en una oficina convencional, amb jornades plenes de videoconferències ininterrompudes que també posen llastre a la productivitat.

Aquesta situació ha portat a molts analistes a plantejar el dubte següent: és la conciliació de la vida laboral i familiar un mite o una realitat?

Persones usuàries potencials

Com a fase inicial del disseny del sistema d'anàlisi de la conciliació de la vida laboral i familiar, identifiquem els requisits de les persones usuàries potencials perquè el sistema els pugui tenir en compte en donar resposta a les seves necessitats i generar informació que els pugui resultar útil.

Les persones usuàries finals que faran ús del sistema són:

- Les **administracions**. Amb la informació disponible en el sistema integrat, els governs i les administracions (especialment el Ministeri de Treball i Economia de cada país) disposaran d'informació per establir les seves polítiques d'actuació enfront de possibles abusos de certs sectors i organitzacions sobre les persones treballadores.
- Els **sindicats** i altres organitzacions en defensa de les persones treballadores. Amb dades objectives sobre la productivitat i la conciliació de la vida laboral dels empleats actius, poden pressionar tant les administracions com les organitzacions empresarials perquè s'introdueixin opcions més flexibles en benefici de l'empleat.
- Les **organitzacions empresarials**. Amb les dades que ofereix el magatzem, les empreses poden fer ajustos que permetin millorar les condicions laborals de l'empleat i augmentar la productivitat. Les organitzacions multinacionals també podran observar possibles desequilibris entre els seus de diferents països.
- **Unió Europea**. Amb la informació que proporciona el sistema, es podrà donar una millor resposta i un paquet de directrius en concordança amb els ministeris corresponents de cada país membre.
- Els **mitjans de comunicació**. Molt relacionats amb el perfil de periodista de la dada, mereixen un punt específic, ja que duen a terme un ús diferent de la informació més orientat a la difusió, la denúncia, la reivindicació...
- La **població activa**. Consultant les dades integrades, poden disposar d'informació molt valuosa per valorar millor la seva situació i permetre constituir grups de pressió per millorar les seves condicions laborals.

Fonts de dades

Un dels objectius d'aquest cas d'estudi és integrar les diverses fonts de dades (i formats) proporcionades per poder fer diferents tipus d'anàlisis. En concret, disposem d'informació detallada sobre la mitjana d'hores treballades per setmana, el percentatge d'hores treballades a casa o l'índex de productivitat per país.

La relació de fitxers de *dades obertes* que utilitzarem per a la càrrega inicial és la següent:

#	Nom de fitxer	Descripció	Font
1	lfsa_ewhun2_linear.csv.gz	Mitjana d'hores treballades setmanals a la UE per país, sexe, edat, categoria professional i activitat econòmica.	https://ec.europa.eu/eurostat/databrowser/view/lfsa_ewhun2/default/table?lang=en
2	lfst_hhwahchi_linear.csv.gz	Percentatge de persones que treballen a casa per país, grup d'edat, nombre i edat de fills.	https://ec.europa.eu/eurostat/databrowser/view/lfst_hhwahchi/default/table?lang=en
3	BLS_US_weeklyhours.xlsx	Mitjana d'hores treballades setmanals als Estats Units. Dades obtingudes de US Department of Labor.	https://beta.bls.gov/dataviewer/view/timeseries/ces0500000002
4	CountryList.json	Conté els noms dels països en ordre alfabètic, agrupats per zona i amb el seu codi ISO 3166-1 alpha-2 en format JSON.	Github Gist: https://gist.github.com/ccalonge/3c5657b5aeade00a02d2b4acb38f48e3
5	ESTAT_AGE_en.tsv	Llista de trams d'edats segons classificació de l'Eurostat.	<i>Codelists extrets de:</i> https://ec.europa.eu/eurostat/databrowser/view/lfsa_ewhun2/default/table?lang=en
6	ESTAT_N_CHILD_en.tsv	Llista del nombre de fills segons classificació de l'Eurostat.	<i>Codelists extrets de:</i> https://ec.europa.eu/eurostat/databrowser/view/lfst_hhwahchi/default/table?lang=en
7	ESTAT_NACE_R2_en.tsv	Llista d'activitats econòmiques segons classificació de l'Eurostat.	<i>Codelists extrets de:</i> https://ec.europa.eu/eurostat/databrowser/view/lfsa_ewhun2/default/table?lang=en
8	ESTAT_SEX_en.tsv	Llista de valors relatius al sexe de les persones que treballen segons classificació de l'Eurostat.	<i>Codelists extrets de:</i> https://ec.europa.eu/eurostat/databrowser/view/lfsa_ewhun2/default/table?lang=en

#	Nom de fitxer	Descripció	Font
9	ESTAT_WORKTIME_en.tsv	Llista de tipus d'horaris laborals de les persones que treballen segons classificació de l'Eurostat.	<i>Codelists</i> extrets de: https://ec.europa.eu/eurostat/data-browser/view/lfsa_ewhun2/default/table?lang=en
10	ESTAT_WSTAT_US_en.tsv	Llista d'estats laborals segons classificació de l'Eurostat.	<i>Codelists</i> extrets de: https://ec.europa.eu/eurostat/data-browser/view/lfsa_ewhun2/default/table?lang=en

Aquestes dades es rebran actualitzades periòdicament i, per tant, seran càrregues incrementals necessàries per a la seva integració en el *data warehouse*. El desenvolupament d'aquests processos futurs queda fora de l'abast d'aquesta activitat.

Enunciat

1. 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 haurà de dissenyar i proposar un DW que permeti i faciliti l'anàlisi de la conciliació de la vida laboral i familiar o *work-life balance*. Mitjançant la metodologia de disseny d'un DW proposat en l'assignatura, l'estudiantat ha de dur a terme:

- **L'anàlisi de requisits.** Com a resultat, es generarà un document que descrigui les preguntes per les quals les persones usuàries potencials esperen resposta del sistema.
- **L'anàlisi de fonts de dades.** S'han de revisar les fonts de dades proporcionades, el tipus d'informació que contenen, el seu format i la quantitat que representen per a la càrrega inicial.
- **L'anàlisi funcional.** S'ha de proposar el tipus d'arquitectura per a la factoria d'informació que s'adeqüi millor al projecte (per exemple, si és necessari un *datamart* operacional o una estructura de càrrega intermèdia).
- **El disseny del model conceptual, lògic i físic del DW.** S'han d'identificar i dissenyar les taules de fets, les dimensions i els atributs que descriuen la informació.

Per a aquest apartat, l'estudiantat ha de preparar un document (solució PR1) en el qual es detalli cadascuna de les seccions anteriors.

S'haurà de tenir en compte que per al desenvolupament del DW cal definir correctament els fets (*facts*), les dimensions d'anàlisi (*dimensions*) i els atributs que ens permetin tenir el nivell de granularitat suficient per a la mesura i presentació dels objectius que es defineixin en l'anàlisi de requisits.

2. PR2: Càrrega de les dades

A partir de la solució oficial de la primera pràctica (PR1), l'estudiantat haurà de dissenyar, implementar i executar els processos d'extracció, transformació i càrrega de les dades (ETL) de les fonts de dades proporcionades.

Les tasques que s'hauran de dur a terme són les següents:

- Identificació dels processos d'extracció, transformació i càrrega de dades (ETL) cap al DW.
- Disseny i desenvolupament dels processos d'ETL mitjançant les eines de disseny proporcionades.
- Implementació amb els treballs dels processos d'ETL per a la seva càrrega efectiva planificada.

3. PR3: Explotació de les dades

Després de la càrrega efectiva de les dades en el magatzem (PR2), s'ha d'implementar un cub multidimensional per explotar la informació com a suport a la presa de decisions de les persones usuàries potencials.

La finalitat del model *multidimensional online analytical processing* (MOLAP) serà respondre les preguntes definides en l'anàlisi de requisits.

Programes

Per al present cas, la UOC proporciona un entorn d'infraestructura d'escriptoris virtuals (VDI) amb tot el programari preconfigurat amb les característiques següents:

- Sistema operatiu: Windows 10.
- Base de dades: base de dades remota Microsoft SQL Server 2016, accessible des de client mitjançant SQL Server Management Studio 17.
- Eina per a la creació de cubs OLAP: Visual Studio 2017.
- Eina de disseny d'ETL: Spoon - Pentaho Data Integration 9.3.
- Eina de creació d'informes: Power BI Desktop.

Bibliografia

Material de l'assignatura *Data warehouse* de la UOC:

Inmon, W. H.; Imhoff, C.; Sousa, R. (2001). *Corporate Information Factory*. Nova York: John Wiley & Sons Inc.

Inmon, W. H. (2005). *Building the Data Warehouse*. Nova York: John Wiley & Sons Inc.

Inmon, W. H.; Strauss, D.; Neushloss, G. (2010). *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. The Morgan Kaufman Sèries in Data Management Systems. Burlington (DT): Morgan Kaufmann Publishers.

Kimball, R. (2013). *The Data Warehouse Toolkit* (3a ed.). Nova York: John Wiley & Sons Inc.

Krishnan, K. (2013). *Data Warehousing in the Age of Big Data*. The Morgan Kaufmann Series on Business Intelligence. Burlington (DT): Morgan Kaufmann Publishers.

Enllaços web:

MSDN Analysis Services Tutorial

<https://docs.microsoft.com/es-es/analysis-services/analysis-services-tutorials-ssas?view=asallproducts-allversions>

Tutorial Pentaho Data Integration

<http://wiki.pentaho.com/display/eai/latest+Pentaho+Data+Integration+%28aka+Kettle%29+Documentation>