

# Solución PRA1- Diseño y construcción de un almacén de datos

## Caso Práctico: Sistema Integrado de Egresados Universitarios

A partir del análisis del contexto del caso y de las fuentes de datos disponibles, el estudiante deberá diseñar y proponer un almacén de datos que ofrezca soporte al funcionamiento del sistema integrado de estudiantes egresados universitarios.

### 1. Análisis de requerimientos

El análisis de requerimientos se basa en identificar las necesidades específicas que tiene una organización particular respecto al análisis de la información. Normalmente en esta fase, se debe ser previsor y pensar más allá de las necesidades actuales para poder cubrir las futuras.

La necesidad principal de la Comisión Europea es disponer de información integrada sobre los egresados para su análisis y difusión mediante herramientas de inteligencia de negocio que le permita, por un lado, la mejora en la toma de decisiones tanto a instituciones universitarias como a la sociedad y por otro, el cumplimiento de principios de transparencia y eficiencia.

En este caso, el objetivo es diseñar un almacén de datos que lo permita, a través de un proyecto, que incluye la creación e implementación de un modelo relacional, el diseño e implementación de procesos ETL, el diseño e implementación del modelo OLAP y, por último, el diseño de las consultas solicitadas en el enunciado.

A continuación, identificamos las siguientes necesidades de información:

1. Conocer la evolución temporal del número de egresados en el sistema educativo universitario.
2. Esta evolución debe poderse analizar desde diferentes perspectivas:
  - a. Tipo de Universidad (P.Ej: Universidades Privadas).
  - b. Modalidad. (P.Ej: No Presencial)
  - c. Universidad (P.Ej: Oberta de Catalunya).
  - d. Rama de enseñanza. (P.Ej: Ciencias Sociales y Jurídicas)
  - e. Ámbito de Estudio. (P.Ej: Ciencias de la educación)
3. Conocer el perfil de los estudiantes egresados en el curso académico 2016-2017, en términos de características personales como sexo y edad.
4. Analizar la incorporación de los graduados universitarios del curso 2009-2010 al mercado laboral en 2014.

## 5. Realizar la comparativa entre egresados universitarios en España y otros países.

Si se tiene en cuenta toda esta información, el sistema podrá responder a múltiples preguntas y de esta manera, cubrir las necesidades de los usuarios potenciales.

De forma específica, se pide que el sistema debe como mínimo ser capaz de dar respuesta a las siguientes preguntas:

- Top 10 de universidades con mayor número de egresados.
- Ranking de ramas de conocimiento con mayor número de estudiantes egresados insertados.
- Evolución en los últimos 5 años del número de egresados universitarios por tipo de universidad.
- Evolución del número de egresados universitarios por modalidad de impartición y rama de conocimiento.
- Ranking de ámbitos de estudios con mayor número de egresados menores de 25 años en el curso académico 2016-2017.
- Ranking de edad con mayor número de personas egresadas en el curso académico 2016-2017.
- Ámbito de estudios con mayor número de mujeres egresadas.
- Ámbito de estudios con mayor número de hombres egresados.
- Tipo de universidad y rama de conocimiento con menor número de estudiantes egresados insertados.
- Ranking de países con mayor porcentaje de estudiantes jóvenes con estudios superiores completos.

## 2. Análisis de fuentes de datos

En este apartado se deben revisar las fuentes de datos proporcionadas, qué tipo de información contienen, cuál es su formato, y qué datos deben ser cargados.

Veamos a continuación un análisis detallado por cada tipo de formato.

### Ficheros Planos.

Los ficheros planos desde el origen vienen con las siguientes características:

- Formato: CSV
- Primera línea con etiquetas de los campos.
- Separador de campos: Punto y coma (;)
- Campos de Texto: Entre comillas ("")

**SEGR1.csv:** Series de número de egresados Universidades Privadas, modalidad, universidad y rama de enseñanza. Cursos 2009-2017.

Nombre Campo	Tipo	Ejemplo
TIPO_UNIVERSIDAD	Texto	"Universidades Privadas"
MODALIDAD	Texto	"No Presencial"
UNIVERSIDAD	Texto	"Oberta de Catalunya"
RAMA_ENSEÑANZA	Texto	"Ciencias Sociales y Jurídicas"
EGR_C16_17	Numérico	3364
EGR_C15_16	Numérico	3129
EGR_C14_15	Numérico	3355
EGR_C13_14	Numérico	2718
EGR_C12_13	Numérico	2975
EGR_C11_12	Numérico	3903
EGR_C10_11	Numérico	3073
EGR_C09_10	Numérico	2513

- Total de registros: 160

**SEGR2.csv:** Series de número de egresados de universidades públicas, modalidad, universidad y rama de enseñanza. Cursos 2009-2017.

Nombre Campo	Tipo	Ejemplo
TIPO_UNIVERSIDAD	Texto	"Universidades Públicas"
MODALIDAD	Texto	"No Presencial"
UNIVERSIDAD	Texto	"Nacional de Educación a Distancia"
RAMA_ENSEÑANZA	Texto	"Ciencias Sociales y Jurídicas"
EGR_C16_17	Numérico	4482
EGR_C15_16	Numérico	5640
EGR_C14_15	Numérico	5245
EGR_C13_14	Numérico	5442
EGR_C12_13	Numérico	5491
EGR_C11_12	Numérico	4663
EGR_C10_11	Numérico	4967
EGR_C09_10	Numérico	3695

Total de registros: 250

**ISCED\_2013.csv:** Clasificación normalizada internacional de educación (ISCED-F 2013). Clasificación del campo de estudio a 5 Niveles.

Nombre Campo	Tipo	Ejemplo
COD_RAMA	Texto	"2"
NOM_RAMA	Texto	"Ciencias Sociales y Jurídicas"
COD_RAMA_N2	Texto	"03"
NOM_RAMA_N2	Texto	"Ciencias sociales, periodismo y documentación"
COD_RAMA_N3	Texto	"031"
NOM_RAMA_N3	Texto	"Ciencias sociales y del comportamiento"
COD_RAMA_N4	Texto	"0311"

NOM_RAMA_N4	Texto	"Economía"
COD_RAMA_N5	Texto	"031101"
NOM_RAMA_N5	Texto	"Economía"

Total de registros: 162

**grad\_5sc.csv:** Perfil de Egresados de grado y master por Ámbito de Enseñanza (4º Nivel de la Clasificación ISCED), Sexo y Grupos de Edad. Curso 2016-2017.

Nombre Campo	Tipo	Ejemplo
COD_AMBITO	Texto	"0311 - Economía"
SEXO	Texto	"Mujeres"
EDAD	Texto	"De 25 a 30 años"
NUM_EGR_NV1	Numérico	249
NUM_EGR_NV2	Numérico	214

Total de registros: 703

## Ficheros Excel.

**03003.xls:** Titulados universitarios según su situación laboral en 2014 por sexo, tipo de universidad y rama de conocimiento del curso académico 2009-2010.

Nombre Campo	Tipo	Ejemplo
TIPO_UNIVERSIDAD	Texto	"Universidades Públicas"
SEXO	Texto	"Mujeres"
RAMA_ENSEÑANZA	Texto	"Ciencias de la salud"
TRABAJANDO	Numérico	14648
EN DESEMPLEO	Numérico	2490
INACTIVO	Numérico	1128

Total de registros: 20

De este fichero Excel no se cargarán datos sobre totales por tipo universidad, por sexo y rama de conocimiento porque son campos redundantes, y se pueden calcular.

**educ\_uoe\_grad01.xls:** Comparación del número de egresados universitarios entre España y otros Países de los cursos 2013-2017.

Nombre Campo	Tipo	Ejemplo
GEO/TIME	Texto	"Finland"
2013	Numérico	52730

2014	Numérico	53878
2015	Numérico	56829
2016	Numérico	56066
2017	Numérico	56136

Total de registros: 12

Este fichero Excel, contiene 3 hojas (Data, Data2 y Data3), el total de personas egresadas, mujeres egresadas y hombres egresados respectivamente. Sólo se cargará la información de la primera hoja con el total de egresados por curso académico ya que es suficiente para el análisis comparativo entre países que se solicita en los requerimientos.

**edat\_ifse\_03.xls:** Porcentaje de egresados universitarios jóvenes (Entre 20 y 29 años).con estudios superiores completos

Nombre Campo	Tipo	Ejemplo
GEO/TIME	Texto	"Finland"
2013	Numérico	50,6
2014	Numérico	50,8
2015	Numérico	54,4
2016	Numérico	53,9
2017	Numérico	53,6

Total de registros: 12

La unidad es número de personas egresadas por cada mil habitantes.

Este fichero Excel, contiene 2 hojas (Data y Data2), con el total de personas egresadas por cada mil habitantes y jóvenes egresados de entre 20 y 29 años por cada mil habitantes, dado que en los requerimientos sólo se hace referencia al análisis sobre jóvenes egresados, y para no complicar la práctica, sólo se tendrá en cuenta la información de la segunda hoja (Data2).

En los proyectos de diseño de factoría de información corporativa existe una primera fase en la que se realiza una carga inicial, y a posteriori, una segunda fase para realizar las cargas incrementales de los datos nuevos que nos van llegando.

Una estimación del volumen de datos de nuestro almacén para la carga de los datos que disponemos sería:

Fuente de datos	Valores a almacenar	Total registros
SEGR1 <input type="checkbox"/> fichero anual <input type="checkbox"/> Tipo Universidad <input type="checkbox"/> Modalidades <input type="checkbox"/> 32 Universidades <input type="checkbox"/> 5 Ramas	Cursos académicos:8 datos	1 fichero x 1 Tipo Univ x 3 Modalidades x 32 Universidades x 5 Ramas x 8 datos = 3840

SEGR2 <input type="checkbox"/> fichero anual <input type="checkbox"/> Tipo Universidad <input type="checkbox"/> Modalidades <input type="checkbox"/> 50 Universidades <input type="checkbox"/> 5 Ramas	Cursos académicos: 8 datos	1 fichero x 1 Tipo Univ x 3 Modalidades x 50 Universidades x 5 Ramas x 8 datos = 6000
grad_5sc <input type="checkbox"/> fichero anual <input type="checkbox"/> 88 Ámbitos Estudio <input type="checkbox"/> 2 Sexo <input type="checkbox"/> 4 Rangos Edad	Niveles Académicos : 2 datos	1 fichero x 88 Ámbitos x 2 Sexo x 4 Edades x 2 datos = 1408
03003 <input type="checkbox"/> 1 fichero anual <input type="checkbox"/> 2 Tipo Universidad <input type="checkbox"/> 2 Sexo <input type="checkbox"/> 5 Ramas	Situaciones Laborales : 3 datos	1 fichero x 2 Tipos x 2 Sexo x 5 Ramas x 3 datos = 60
educ_uoe_grad01 <input type="checkbox"/> ficheros anuales <input type="checkbox"/> 12 Países	Cursos académicos: 5 datos	1 ficheros x 12 Países x 5 datos = 60
edat_lfse_03 <input type="checkbox"/> fichero anual <input type="checkbox"/> 12 Países	Cursos académicos: 5 datos	1 fichero x 12 Países x 5 datos = 60
ISCED_2013 <input type="checkbox"/> fichero anual	Titulaciones <input type="checkbox"/> 162 datos	1 fichero x 162 datos= 162
	TOTAL	=11.590

### 3. Análisis Funcional

A continuación, se propone el tipo de arquitectura para la factoría de información que mejor se adecua al proyecto.

Para ello vamos a considerar los requisitos funcionales y establecer para cada uno de ellos una prioridad, que podrá ser exigible (E) o deseable (D). En el contexto de esta actividad, los requerimientos exigibles son aquellos que demanda el enunciado y los deseables son aquellos que complementan la actividad.

Por otro lado, en términos de la escala de prioridades asignamos una prioridad de 1 a 3 siendo 1 completamente prioritario para la actividad y 3 no prioritario para la actividad.

A continuación, se describen los requerimientos funcionales para el diseño de una factoría de información para nuestra organización, bajo las consideraciones del enunciado:

#	Requerimiento	Prioridad	Exigible / Deseable
1	Se extraerá de forma adecuada la información de las fuentes de datos (considerando sólo la información relevante).	1	E
2	Se creará un almacén de datos.	1	E
3	Se cargará la información de las personas egresadas en el almacén de datos.	1	E
4	Se creará un modelo OLAP para consultas multidimensionales de los usuarios.	2	E
5	Se crearán los informes estáticos solicitados.	2	E
6	Se redactará un manual de carga de datos incremental	3	D

Cabe comentar que en un caso genérico real podemos encontrar también otros requerimientos funcionales:

- Creación de procesos de calidad de datos.
- Creación de *data marts* (si se analizan otras áreas).
- Automatizar cada proceso de carga de *data marts* (según sus necesidades).
- Creación de procesos de cargas totales e incrementales.
- Creación de un repositorio de metadatos de gestión del almacén de datos, así como de los procesos ETL.

Así mismo, dado que estos sistemas frecuentemente forman parte de la implementación de un sistema de inteligencia de negocio, la lista de requerimientos funcionales sería mucho mayor.

En términos de la arquitectura funcional tenemos los siguientes elementos:

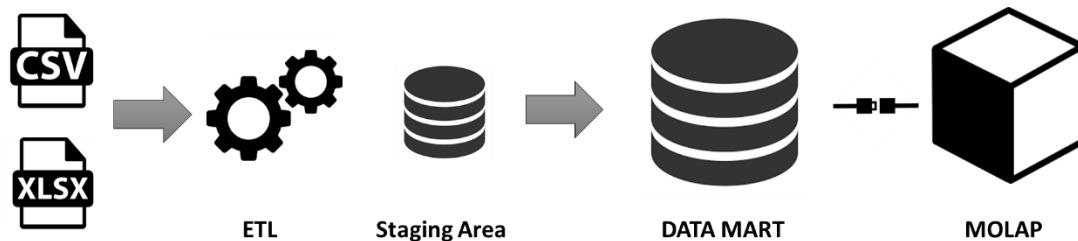
1. Las fuentes de datos están compuestas por los ficheros planos csv y hojas de cálculo xls obtenidos de las webs del Ministerio de Educación (MECD), Eurostat y *OpenData*. Estos ficheros contienen información sobre la evolución de las personas egresadas desde 2012 a 2017, características personales como el sexo y edad de egresados universitarios en el curso 2016-2017, inserción al mercado laboral en 2014 de los egresados del curso 2009-2010 y datos comparativos entre España y otros países de los egresados entre 2012 y 2017.
2. La arquitectura de la factoría de información de personas egresadas puede estar formada por varios elementos alojados en la misma máquina:
  - *Staging Area* (opcional): En el caso de tener múltiples fuentes (ficheros, bases de datos, servicios RSS....) es conveniente cargarlas para consolidar la información en una estructura de carga intermedia que puede ser creada en la misma base de datos.

Esta área del DW también puede servir para entender, simplificar y consolidar el proceso ETL.

- *Data Mart* de personas egresadas: al centrarnos en una única área temática como es el análisis de egresados universitarios, es más correcto considerar que se está creando un *data mart* en lugar de un almacén de datos corporativo.
- *MOLAP*: a partir de la información del *datamart* se creará un cubo multidimensional.

Según lo comentado anteriormente, en diseño de la arquitectura funcional se puede optar por usar un área intermedia (*Staging Area*), incluida dentro de la misma base de datos, cuyos objetos se identificarán con un prefijo en los nombres.

El siguiente gráfico resume los elementos de la arquitectura para esta actividad:



O también sería correcto utilizar una arquitectura sin *Staging Area*, identificando las tablas intermedias con un prefijo en el nombre, por ejemplo: IN\_nombre\_tabla\_intermedia.



## 4. Diseño del modelo conceptual, lógico y físico del almacén de datos

### Diseño Conceptual.

Para el correcto desarrollo del DW es preciso definir los hechos (*facts*), dimensiones de análisis (*dimensions*), las métricas y los atributos que nos permitan tener el nivel de granularidad suficiente para presentación de los objetivos que se han definido en el análisis de requerimientos y de las fuentes de datos.



Del análisis de las fuentes de datos se determinan que disponemos de datos anuales, correspondientes a datos acumulados de cada curso académico y que los hechos son:

- Personas egresadas
- Personas egresadas insertadas, que la situación laboral es trabajando.

Teniendo en cuenta los requerimientos solicitados, los hechos identificados se analizarán para resolver cuatro necesidades principales de los usuarios:

1. Análisis temporal de las personas egresadas.
2. Caracterización de las personas egresadas.
3. Análisis de las personas egresadas insertadas (cuya situación laboral es trabajando).
4. Comparativa de las personas egresadas entre España y otros países.

El análisis temporal de las personas egresadas, determina el diseño de la primera tabla de hechos:

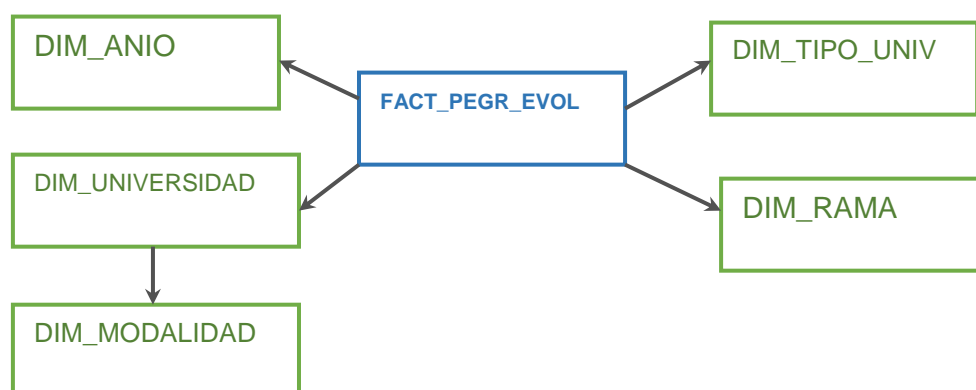
Tabla de hechos	Descripción
Fact_PersonasEgresadas_Evolutivo	Recoge el número de egresados por curso académico.

Esta tabla de hechos, además de la métrica del número personas egresadas, se analizará desde diferentes perspectivas:

Dimensiones	Descripción
Año	Número de personas egresadas por Curso Académico
TipoUniversidad	Número de egresados por Tipo de Universidad
Universidad	Número de personas egresadas por Universidad
Rama Enseñanza	Número de personas egresadas por Rama Enseñanza (Primer Nivel de Clasificación ISCED).

A partir de las dimensiones y la tabla de hechos identificados, se construye el modelo conceptual, siendo tanto las dimensiones como los hechos, entidades independientes que forman parte de nuestro modelo de estrella/copo de nieve.

El diseño conceptual para esta tabla de hechos y sus dimensiones es:



Para la caracterización de las personas egresadas, identificamos una segunda la tabla de hechos:

Tabla de hechos	Descripción
Fact_PersonasEgresadas_Perfil	Recoge el número de egresados del curso 2016-2017.

Esta tabla de hechos almacenará la métrica del número de personas egresadas que será un campo calculado de la suma de las personas egresadas de grado (NUM\_EGR\_NV1) más las personas egresadas de master (NUM\_EGR\_NV2) y será analizada desde diferentes perspectivas:

Dimensiones	Descripción
Año	Número de personas egresadas por Curso Académico
Sexo	Número de personas egresadas por Sexo
Edad	Número de personas egresadas por intervalo de Edad
Rama Enseñanza	Número de personas egresadas por Rama Enseñanza, utilizando el nivel, Ámbito de Enseñanza, que corresponde al 4º Nivel de jerarquía de la dimensión Rama Enseñanza.

En nuestro caso tendremos un modelo en estrella conceptual para cada tabla de hechos con dimensiones comunes como la dimensión tiempo y rama de enseñanza. El diseño conceptual para esta tabla de hechos y sus dimensiones es:



La tabla de hechos que permitirá obtener información sobre la situación laboral en 2014 de las personas egresadas en el curso académico 2009-2010 es:

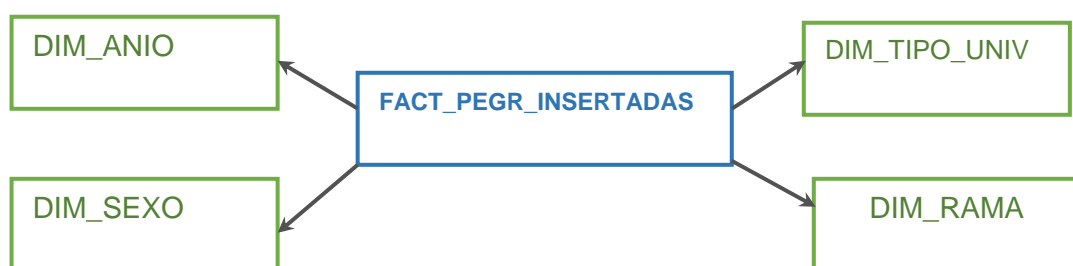
Tabla de hechos	Descripción
Fact_PersonasEgresadas_Insertadas	Recoge el número de egresados insertados en 2014.

Esta tabla de hechos contendrá 3 métricas, número de estudiantes egresados trabajando, desempleados e inactivos) y será analizada desde diferentes perspectivas:

Dimensiones	Descripción
Año	Número de personas egresadas insertadas por Curso Académico

TipoUniversidad	Número de egresados insertados por Tipo de Universidad
Sexo	Número de personas egresadas insertados por Sexo
Rama de Enseñanza	Número de personas egresadas insertados por Rama de Enseñanza (Primer Nivel de Clasificación ISCED)

El diseño conceptual para esta tabla de hechos y sus dimensiones es:



Por último, la tabla de hechos que recoge la información de Eurostat con datos de personas egresados de otros países en cuatro años académicos y que permitirá realizar un análisis comparativo es:

Tabla de hechos	Descripción
Fact_PersonasEgresadas_Comparativa	Recoge el número de egresados por países.

Contendrá dos métricas, el número de egresados y porcentaje de egresados jóvenes con estudios superiores o terciarios. Esta tabla será analizada desde diferentes perspectivas:

Dimensiones	Descripción
Año	Número de personas egresadas por Curso Académico
País	Número de personas egresadas por País

El diseño conceptual para esta tabla de hechos y sus dimensiones es:



## Diseño Lógico.

Una vez obtenido el modelo conceptual del almacén de datos de personas egresadas, compuesto de cuatro tablas de hechos y ocho dimensiones; vamos a detallar las métricas de cada una de las tablas de hechos y sus atributos con el objetivo de diseñar el modelo lógico.

A continuación, se muestra una tabla con las métricas que contiene cada tabla de hechos que compone el modelo lógico del almacén de personas egresadas.

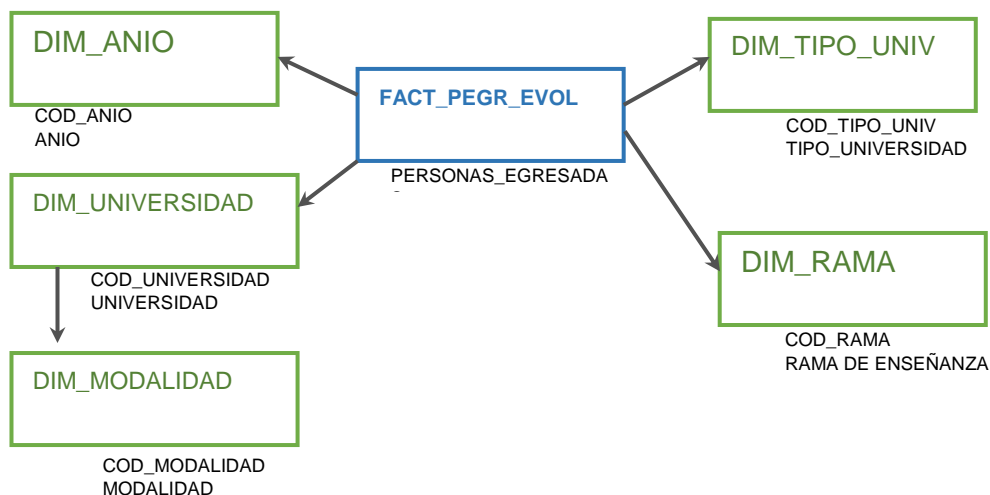
Tabla de Hechos	Métricas
FACT_PEGR_EVOLUTIVO	PERSONAS_EGRESADAS
FACT_PEGR_PERFIL	PERSONAS_EGRESADAS
FACT_PEGR_INSERTADAS	PEGRES_TRABAJANDO PEGRES_DESEMPLEADAS PEGRES_INACTIVAS
FACT_PEGR_COMPARATIVA	PERSONAS_EGRESADAS PORCENTAJE_EGR_JÓVENES

Lo siguiente que veremos son los atributos que contiene cada tabla de hechos. Los atributos junto con las métricas, nos permitirán realizar los diferentes análisis de los requerimientos.

En la siguiente tabla, se muestran los atributos descriptores con las referencias a sus dimensiones de la tabla de hechos FACT\_PEGR\_EVOLUTIVO:

Dimensiones	Atributos descriptores
DIM_ANIO	COD_ANIO
DIM_TIPO_UNIV	COD_TIPO_UNIV
DIM_UNIVERSIDAD	COD_UNIVERSIDAD
DIM_RAMA	COD_RAMA

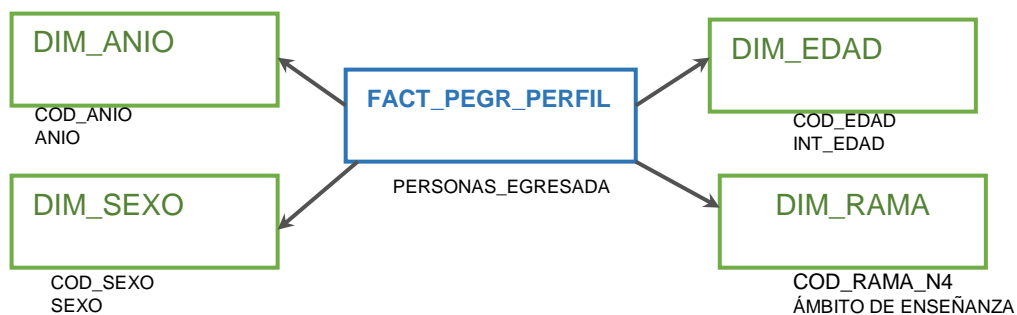
En la siguiente imagen se muestra el diseño del modelo lógico propuesto para la tabla de hechos FACT\_PEGR\_EVOLUTIVO.



Los atributos descriptores de la tabla de hechos **FACT\_PEGR\_PERFIL**:

Dimensiones	Atributos descriptores
DIM_ANIO	COD_ANIO
DIM_SEXO	COD_SEXO
DIM_EDAD	COD_EDAD
DIM_RAMAS	COD_RAMAS_N4

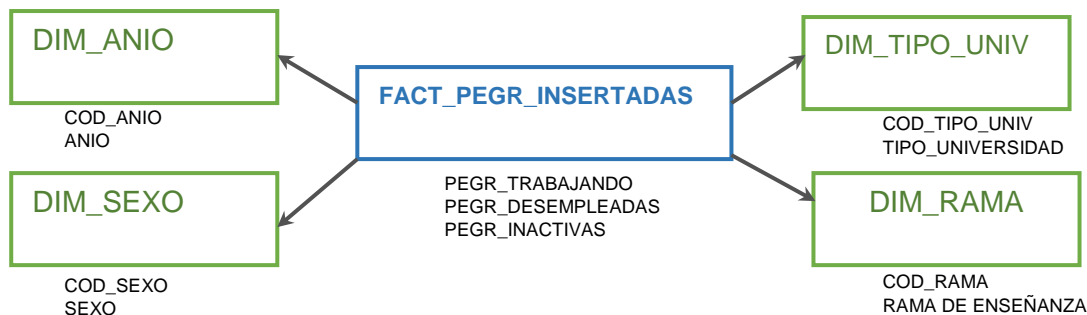
El diseño del modelo lógico propuesto para la tabla de hechos **FACT\_PEGR\_PERFIL** sería:



Los atributos descriptores a las dimensiones en la tabla de hechos **FACT\_PEGR\_INSERTADAS**:

Dimensiones	Atributos descriptores
DIM_ANIO	COD_ANIO
DIM_TIPO_UNIV	COD_TIPO_UNIV
DIM_SEXO	COD_SEXO
DIM_RAMAS	COD_RAMAS

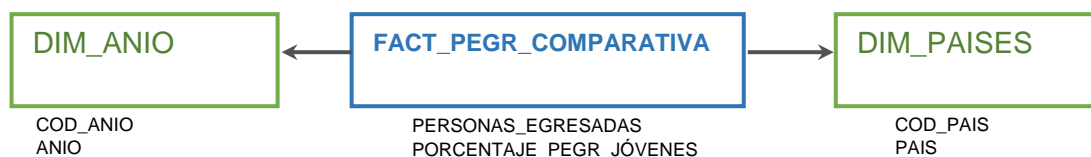
En la imagen se muestra su correspondiente modelo lógico:



Por último, en la siguiente tabla vemos los atributos de la tabla de hechos FACT\_PEGR\_COMPARATIVA.

Dimensiones	Atributos descriptores
DIM_ANIO	COD_ANIO
DIM_PAIS	COD_PAIS

Y su modelo lógico sería:



## Diseño Físico.

Para el correcto diseño físico del almacén debemos tener en cuenta diversos aspectos:

- El tipo de base de datos con el que trabajemos, puesto que cada una de ellas tiene su particularidad.
- El diseño físico debe estar orientado a generar un buen rendimiento en el procesamiento de consultas.
- La definición de los procesos de administración del DW.
- La revisión periódica del diseño físico inicial para validar que continúa dando respuesta a las necesidades del cliente.

Una vez determinados qué tablas de hechos, dimensiones, métricas y atributos existen en nuestro modelo, podemos determinar las claves foráneas que debe definirse en el modelo físico.

En este paso también es necesario tener en cuenta el tamaño adecuado de los atributos (por ejemplo, qué longitud tiene una cadena o si los numéricos contienen decimales). También es relevante acordarse de crear correctamente las claves primarias en las dimensiones.

Dado que nuestro modelo de almacén está compuesto de más de una tabla de hechos, también debemos revisar las dimensiones que hemos definido en el diseño conceptual y lógico de cada *fact* aplicando una visión conjunta del modelo. Esto nos permitirá definir dimensiones comunes, como año o tipo de universidad y así simplificar el modelo final y conseguir un rendimiento óptimo en la ejecución de los análisis.

Como es lógico, primero se crean las tablas de dimensiones y posteriormente las tablas de hechos. De esta forma creamos cada una de las tablas de nuestro almacén de datos.

## Dimensiones.

Las dimensiones del modelo podrán ser referenciadas en las tablas de hechos utilizando sus claves primarias o *primary key* (CK/PK). El modelo físico de las dimensiones es:

- **DIM\_ANIO:** Corresponde a la dimensión temporal de nuestro almacén. Disponemos de datos anuales, correspondientes a datos acumulados de cada curso académico. En nuestro caso es muy simple. Normalmente es bastante más complicada, ya que puede incluir días, días de la semana, festivos, semestres, cuatrimestres, etc. La clave primaria será el año de carga de datos y una descripción para describir los cursos académicos a los que pertenecen los datos.

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_ANIO</b> (CP/PK)	Numérico	4	2011
DESC_ANIO	Texto	50	'2009-2010'

- **DIM\_TIPO\_UNIV:** Contiene los valores de los tipos de universidades existentes (1. Universidades Públicas; 2. Universidades Privadas)

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_TIPO_UNIV</b> (CP/PK)	Numérico	1	2
DESC_TIPO_UNIV	Texto	30	'Universidades Privadas'

- **DIM\_RAMA:** La dimensión rama, contiene información sobre la clasificación normalizada internacional de educación (ISCED-F 2013) a 5 Niveles.

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_RAMA</b> (CP/PK)	Numérico	5	1
COD_RAMA	Texto	1	1
DESC_RAMA	Texto	50	'Artes y Humanidades'
COD_RAMA_N2	Texto	2	'01'
DESC_RAMA_N2	Texto	100	'Educación'

COD_RAMA_N3	Texto	3	'011'
DESC_RAMA_N3	Texto	100	'Educación'
COD_RAMA_N4	Texto	4	'0111'
DESC_RAMA_N4	Texto	150	'Ciencias de la educación'
COD_RAMA_N5	Texto	6	'011101'
DESC_RAMA_N5	Texto	200	'Pedagogía'

- **DIM\_MODALIDAD:** Contiene las modalidades de impartición de las universidades (1. Presencial; 2. No Presencial; 3. Especiales)

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_MODALIDAD</b> (CP/PK)	Numérico	1	3
DESC_MODALIDAD	Texto	50	'Especiales'

- **DIM\_UNIVERSIDAD:** Contiene la información de cada una de las universidades sobre la que se recoge información.

Según el análisis de las fuentes de datos, se ha considerado que la modalidad de impartición es una propiedad de cada universidad, por esta razón en nuestro modelo será un atributo de la dimensión de universidad en lugar de ser un atributo de la tabla de hechos.

Si las necesidades de los usuarios requieren realizar análisis sobre la información de la modalidad de impartición, podrá modificarse el modelo y utilizar la dimensión modalidad en las tablas de hechos.

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_UNIVERSIDAD</b> (CP/PK)	Numérico	3	40
DESC_UNIVERSIDAD	Texto	100	'Internacional de Andalucía'
SK_DIM_MODALIDAD (CA/FK)	Numérico	1	3

El atributo **SK\_DIM\_MODALIDAD**, hace referencia al tipo de modalidad de impartición de cada universidad y cuyos valores se encuentran en la tabla creada para la dimensión modalidad (**DIM\_MODALIDAD**). Es por esta razón, que se crea una restricción de Clave Ajena o *Foreign Key* en el atributo **SK\_DIM\_MODALIDAD** en la dimensión universidad. Y así garantizar la integridad referencial.

- **DIM\_SEXO:** Contiene los valores correspondientes al sexo de las personas egresadas (1. Mujeres y 2. Hombres)



Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_SEXO</b> (CP/PK)	Numérico	1	1
DESC_SEXO	Texto	10	'Hombres'

- **DIM\_EDAD:** Contiene los valores correspondientes al intervalo de edad de las personas egresadas (1. Menos de 25 años; 2. De 25 a 30 años; 3. De 31 a 40 años y 4. Más de 40 años)

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_EDAD</b> (CP/PK)	Numérico	1	3
DESC_INT_EDAD	Texto	50	'De 31 a 40 años'

- **DIM\_PAIS:** Contiene la relación de los nueve países con los que se realizarán el análisis comparativo de personas egresadas.

Nombre Campo	Tipo	Tamaño	Ejemplo
<b>SK_DIM_PAIS</b> (CP/PK)	Numérico	3	11
DESC_PAIS_ES	Texto	50	'Suiza
DESC_PAIS_EN	Texto	200	'Switzerland'

## Tablas de Hechos.

El modelo físico de las tablas de hechos consistirá en la creación de las tablas cuyos campos serán claves foráneas a las dimensiones del modelo estrella del su diseño lógico y de las métricas.

El modelo físico de las tablas de hechos del almacén integrado de personas egresadas está compuesto de las siguientes tablas:

- **FACT\_PEGR\_EVOLUTIVO:** Es la tabla física que contendrá la información que permitirá realizar el análisis evolutivo de las personas egresadas, concretamente tendrá los siguientes campos:

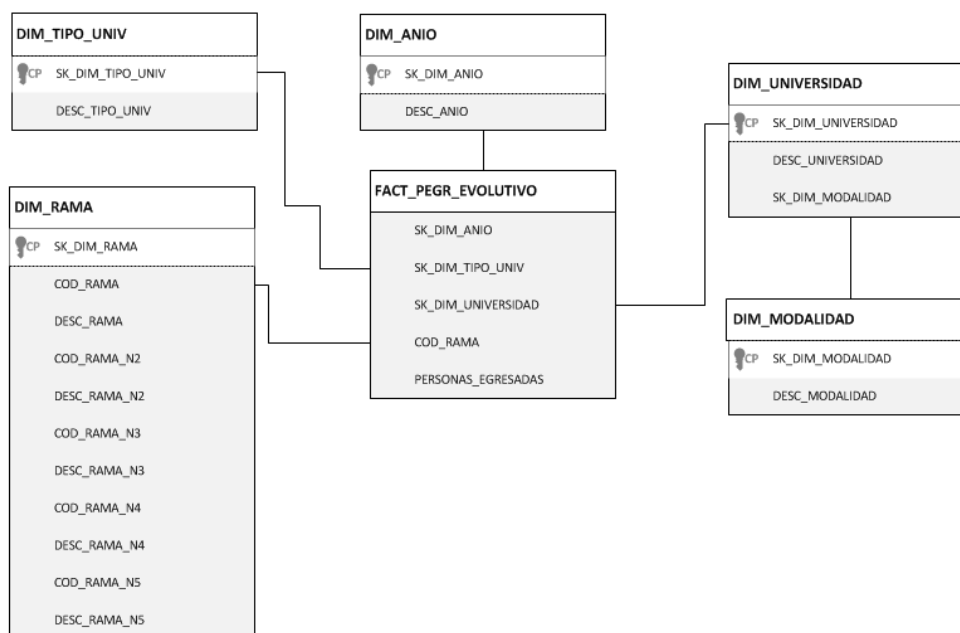
Nombre Campo	Tipo	Tamaño	Ejemplo
SK_DIM_ANIO	Numérico	4	2017
SK_DIM_TIPO_UNIV	Numérico	1	1
SK_DIM_UNIVERSIDAD	Numérico	3	9
COD_RAMA	Texto	1	2
PERSONAS_EGRESADAS	Numérico	8	1351

En la siguiente imagen se muestra el diseño del modelo físico<sup>1</sup> para la tabla de hechos FACT\_PEGR\_EVOLUTIVO.

El atributo descriptor COD\_RAMA en la tabla de hechos FACT\_PEGR\_EVOLUTIVO, no puede definirse como clave foránea de DIM\_RAMAS, ya que en dicha dimensión no es clave primaria al no contener valores únicos. En la clasificación de estudios se presenta N campos de estudio para cada rama de enseñanza. Por ejemplo, para la rama de CIENCIAS (3), hay varios campos de estudio (Biología, Bioquímica, etc.) en la clasificación que pertenecen a dicha rama.

Aunque la relación física entre la tabla de hechos y la dimensión no pueda realizarse, se hará más en la creación del cubo para la explotación de datos.

Hay otras soluciones que, si se justifican, podemos darlas por válidas. Por ejemplo, normalizando la dimensión, es decir, creando una tabla por cada nivel de la clasificación. Con el fin de simplificar la solución, se ha optado por tratar el dato de la rama de conocimiento como un atributo en la tabla de hechos.



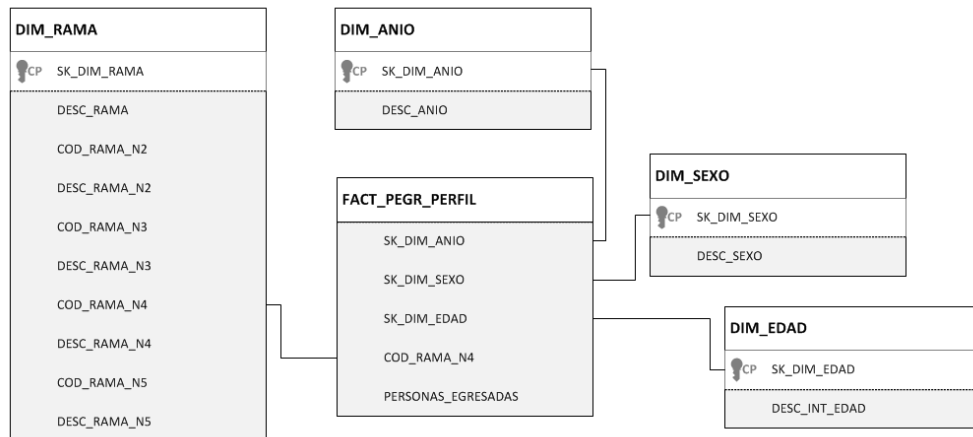
- **FACT\_PEGR\_PERFIL:** Es la tabla física que contendrá la información que permitirá realizar el análisis de la caracterización por sexo y edad de las personas egresadas, concretamente está compuesta de los siguientes campos:

Nombre Campo	Tipo	Tamaño	Ejemplo
SK_DIM_ANIO	Numérico	4	2018

<sup>1</sup> El diseño se ha realizado utilizando la herramienta Microsoft Visio.

SK_DIM_SEXO	Numérico	1	1
SK_DIM_EDAD	Numérico	1	2
COD_RAMA_N4	Texto	4	'0111'
PERSONAS_EGRESADAS	Numérico	8	320

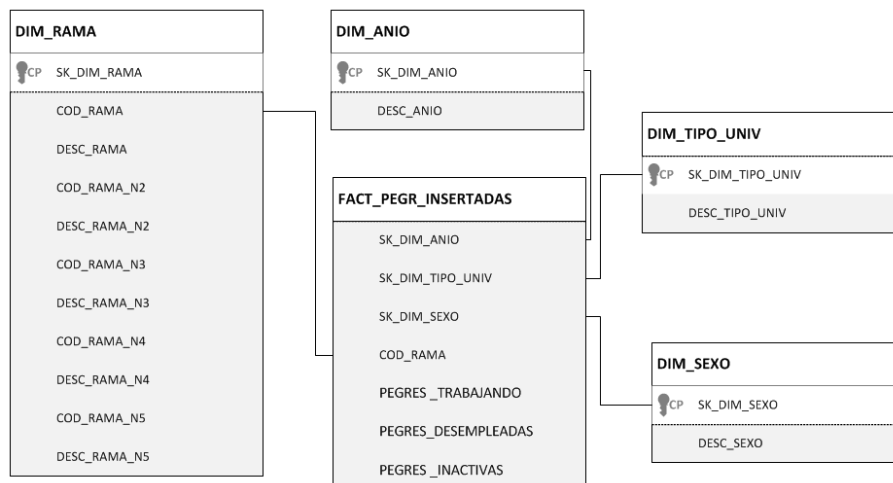
El diseño del modelo físico para la tabla de hechos FACT\_PEGR\_PERFIL:



- **FACT\_PEGR\_INSERTADAS:** Es la tabla física que contiene información sobre la inserción laboral de las personas egresadas en el curso académico 2009-2010, sus campos son:

Nombre Campo	Tipo	Tamaño	Ejemplo
SK_DIM_ANIO	Numérico	4	2011
SK_DIM_TIPO_UNIV	Numérico	1	2
SK_DIM_SEXO	Numérico	1	1
COD_RAMA	Texto	1	1
PEGR_TRABAJANDO	Numérico	8	2674
PEGR_DESEMPLEADOS	Numérico	8	1122
PEGR_INACTIVOS	Numérico	8	458

El diseño del modelo físico para la tabla de hechos FACT\_PEGR\_INSERTADAS:



- **FACT\_PEGR\_COMPARATIVA:** Es la tabla que contiene información del número personas egresadas y el porcentaje de jóvenes con estudios superiores completos por países, sus campos son:

Nombre Campo	Tipo	Tamaño	Ejemplo
SK_DIM_ANIO	Numérico	4	2013
SK_DIM_PAIS	Numérico	3	2
PERSONAS_EGRESADAS	Numérico	8	52730
PORCENTAJE_PEGR_JOVENES	Numérico	8	51

El diseño del modelo físico para la tabla de hechos **FACT\_PEGR\_COMPARATIVA** se muestra en la siguiente figura:

