

**E**scola **S**uperior de **E**nxeñaría **I**nformática

|  |
| --- |
| PRÁCTICA INTELIGENCIA DE NEGOCIO Y ALMACENES DE DATOS  (curso 2020-2021)  **Código Grupo: G1-6**  **Título: Datawarehouse para empresa de software** |

**Alumna/o: David Rodríguez Martínez**

**Alumna/o: Julio Patricio DaSilva**

**Alumna/o: Diego Enrique Fontán Lorenzo**

**Alumna/o: Diego de Miguel Fernández**

**ÍNDICE**

1. **Descripción del proceso**
2. **Selección de la granularidad**
3. **Identificación de las dimensiones**
4. **Selección de las medidas asociadas al hecho**
5. **Almacenamiento de valores precalculados en la tabla de hechos**
6. **Terminación de las tablas de dimensión**
7. **Selección de la duración de la base de datos**
8. **Control de las dimensiones lentamente cambiantes**
9. **Modos de consulta y prioridades de ejecución**

# Descripción del proceso

* 1. Descripción textual de la organización

El objetivo de esta organización consiste en el desarrollo de software por parte del equipo de desarrolladores y su posterior puesta en venta para clientes particulares.

* 1. Modelo E/R original correspondiente al sistema OLTP preexistente

Enlace al modelo E/R:

[**https://drive.google.com/file/d/1EFTJeFou1GsIRsjMD9IWVNgw\_5nCxBZ1/view?usp=sharing**](https://drive.google.com/file/d/1EFTJeFou1GsIRsjMD9IWVNgw_5nCxBZ1/view?usp=sharing)

* 1. Modelo E/R con las entidades que almacenan información de interés para el mercado de datos

Enlace al modelo E/R:

[**https://drive.google.com/file/d/1dnohbNLMwvOPhx1nLhqWeC8ES8edNqho/view?usp=sharing**](https://drive.google.com/file/d/1dnohbNLMwvOPhx1nLhqWeC8ES8edNqho/view?usp=sharing)

* + 1. Explicar la incorporación de nuevos atributos a los preexistentes en la organización, en su caso.

Las tablas siguientes con sus correspondientes atributos serán necesarios para poder realizar clasificaciones a la hora de realizar consultas OLAP.

Tabla **TICKET**:

* **ID:** Identificador del ticket
* **Descripción:** Texto que describirá el asunto del ticket
* **Estado:** Indica la situación del ticket: Presentado, en revisión, resuelto…
* **Título:** Titulo que describe el ticket
* **Fecha:** Fecha de creación del ticket
* **Ultima\_actualización:** Última actualización del ticket

Tabla **FEEDBACK**:

* **Comentario:** Comentario que da el cliente en el feedback
* **Valoración:** Valoración numérica que da el cliente en el feedback
  1. **Descripción de las fuentes externas**

|  |  |  |
| --- | --- | --- |
| **FUENTE** | **Directorio de empresas** | |
| **Descripción** | Dataset que contiene información de 900 empresas que residen fiscalmente en España. Contiene datos como el nombre, dirección y código postal, datos de contacto así como página web y sector al que pertenecen. También se indica la actividad que desarrollan. | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** |
| NOMBRE | STRING | Nombre comercial de la empresa |
| DIRECCION | STRING | Dirección fiscal donde se encuentra la empresa |
| C.POSTAL | INT | Código postal donde reside la empresa |
| ZONA | STRING | Indica el tipo de zona urbana donde se encuentra la empresa |
| TELEFONO | STRING | Teléfono de contacto de la empresa |
| FAX | STRING | Numero FAX de la empresa |
| EMAIL | STRING | Email de contacto de la empresa |
| WEB | STRING | Página web de la empresa |
| CONTACTO | STRING | Nombre de la persona de contacto |
| SECTOR | STRING | Sector empresarial en el que se ubica la empresa |
| ACTIVIDAD1 | STRING | Actividad principal que desarrolla la empresa |
| ACTIVIDAD2 | STRING | Actividad secundaria que desarrolla la empresa |
| LAT | FLOAT | Coordenada de latitud de la ubicación de la empresa |
| LON | FLOAT | Coordenada de longitud de la ubicación de la empresa |

* 1. Descripción textual de la actividad a modelar

La actividad a modelar consistirá en el hecho de vender las licencias a los clientes. Esta actividad será medida por características como el precio, las unidades vendidas, el número de clientes o las valoraciones del producto.

# Selección de la granularidad

* 1. Descripción del gránulo de la actividad a modelar y de cada una de las dimensiones

Para la actividad **ventas de licencias:**

Se almacenará la información de las ventas **diarias** de las licencias de **cada producto** que se vendan a **cada cliente**.

Dimensiones:

* **Qué:** Venta de licencias
* **A quien:** Venta de licencias dirigidas a particulares o empresas.
* **Quien:** El equipo de desarrollo crean los programas para su venta mediante las licencias.
* **Cuando:** Nos interesa saber el número de ventas que se realizan diariamente.

Para la actividad **solución de incidencias:**

Se desea almacenar el número de tickets abiertos **diariamente** de **cada programa** emitidos por **cada cliente**.

Dimensiones:

* **Qué:** Tickets almacenados de cada programa
* **Cuando:** Diariamente
* **Quien:** Cada cliente que genere un ticket
  1. Razonar por qué no se ha optado por una granularidad mayor o menor

No tendría sentido ajustar las ventas a una granularidad mayor ya que no sería relevante la información de ventas de cada hora o una frecuencia menor.

# Identificación de las dimensiones

* 1. Descripción textual de las dimensiones (a nivel general, sin detalle de los atributos)
  2. Diagrama de las dimensiones

# Selección de las medidas asociadas al hecho

* 1. Descripción textual de las medidas y de dónde se obtienen (a qué atributo(s) se corresponde en la BD OLTP).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MEDIDA | TIPO | DESCRIPCIÓN | TABLA/FUENTE ORIGEN | ATRIBUTO/CAMPO ORIGEN |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. Diagrama con la incorporación de las medidas

# Almacenamiento de valores precalculados en la tabla de hechos

* 1. Definición de valores precalculados y descripción acerca de cómo se obtienen
  2. Diagrama con incorporación de los valores precalculados

# Terminación de las tablas de dimensión

* 1. Descripción detallada de las dimensiones. Para cada uno de sus atributos indicar:
     1. Cómo se obtiene (atributo(s) específico(s) en OLTP, campo calculado, fuente externa X, etc.)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DIMENSIÓN | (nombre dimensión) | | | |
| Descripción de la dimensión |  | | | |
| ATRIBUTO | TIPO | DESCRIPCIÓN | TABLA/FUENTE ORIGEN | ATRIBUTO/CAMPO ORIGEN |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* + 1. Establecimiento de la jerarquía entre los atributos de cada dimensión (p.ej: día, mes, año)

|  |  |
| --- | --- |
| DIMENSIÓN | (nombre dimensión) |
| Jerarquía |  |

* + 1. Conversión de datos: numerización/etiquetado, discretización (hacer al menos 2)

|  |  |  |
| --- | --- | --- |
| DIMENSIÓN |  | |
| ATRIBUTO | TIPO DE  CONVERSIÓN | DESCRIPCIÓN |
|  |  |  |
|  |  |  |

* + 1. Cómo se debe tratar en caso de datos erróneos

|  |  |
| --- | --- |
| DIMENSIÓN | (nombre dimensión) |
| ATRIBUTO | DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO |
|  |  |
|  |  |

* + 1. Cómo se debe tratar en caso de datos faltantes

|  |  |
| --- | --- |
| DIMENSIÓN | (nombre dimensión) |
| ATRIBUTO | DESCRIPCIÓN DEL TIPO DE PROBLEMA Y CÓMO RESOLVERLO |
|  |  |
|  |  |

* 1. Diagrama final de diseño de la BD Datawarehouse versión ROLAP (incorporando claves)

# Selección de la duración de la base de datos

* 1. Indicación de la duración elegida y justificación

# Control de las dimensiones lentamente cambiantes

* 1. Determinación de las dimensiones lentamente cambiantes y cuáles son los atributos afectados.
  2. Para cada uno de los atributos cambiantes indicar cómo será tratado (Tipo 1, 2 o 3). *Es necesario incluir al menos un atributo cambiante Tipo 2.*

# Modos de consulta y prioridades de ejecución

* 1. Descripción textual de las consultas OLAP