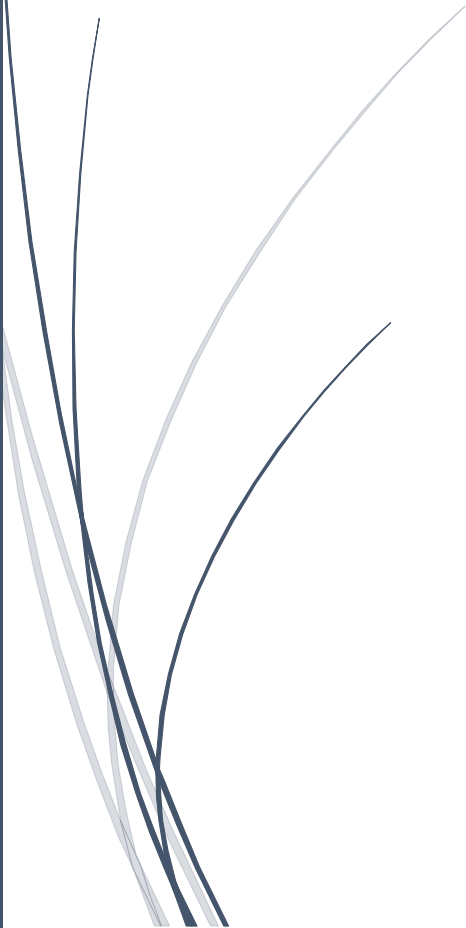




10-3-2017

# LAB BOOK

MARIA DOLORES SESMERO POZO  
JUAN ANGEL PIQUERAS LOPEZ;JUAN  
ANGEL PIQUERAS LOPEZ



<b>Martes, 21 Febrero 2017.....</b>	<b>1</b>
<b>Miércoles, 22 Febrero 2017.....</b>	<b>3</b>
<b>Jueves, 23 Febrero 2017.....</b>	<b>4</b>
<b>Viernes, 24 Febrero 2017.....</b>	<b>5</b>
<b>Sábado, 25 Febrero 2017 .....</b>	<b>6</b>
<b>Lunes, 27 Febrero 2017 .....</b>	<b>7</b>
<b>Martes, 28 Febrero 2017.....</b>	<b>8</b>
<b>Miercoles, 8 Marzo 2017.....</b>	<b>9</b>
<b>Jueves, 9 Marzo 2017.....</b>	<b>10</b>
<b>Viernes, 10 Marzo 2017 .....</b>	<b>11</b>
<b>ANEXO.....</b>	<b>12</b>
<b>Miércoles, 22 Febrero 2017 .....</b>	<b>12</b>
<b>Jueves, 23 Febrero 2017 .....</b>	<b>12</b>
<b>Viernes, 24 Febrero 2017.....</b>	<b>12</b>
<b>Sábado, 25 Febrero 2017.....</b>	<b>12</b>
<b>Lunes, 27 Febrero 2017.....</b>	<b>13</b>
<b>Martes, 28 Febrero 2017 .....</b>	<b>13</b>
<b>Miércoles, 8 Marzo 2017 .....</b>	<b>13</b>
<b>Jueves, 9 Marzo 2017 .....</b>	<b>13</b>
<b>Viernes, 10 Marzo 2017.....</b>	<b>13</b>

**Martes, 21 Febrero 2017**

## **1. Selección del tema para realizar el Data Warehouse**

En primer lugar hemos buscado los posibles temas que podríamos usar para realizar el data warehouse.

El tema seleccionado ha sido "Registro de identificación de animales de compañía de la CAE (REGIA)" tomado de "<http://opendata.euskadi.eus>".

El archivo que usaremos se trata de un csv, llamado "fueraEuskadi.csv" que contiene datos como la fecha de nacimiento de los animales, fecha de alta e implantación de chip en los mismos, así como el identificador del chip, el tipo de animal, sexo, raza, y localización del animal.

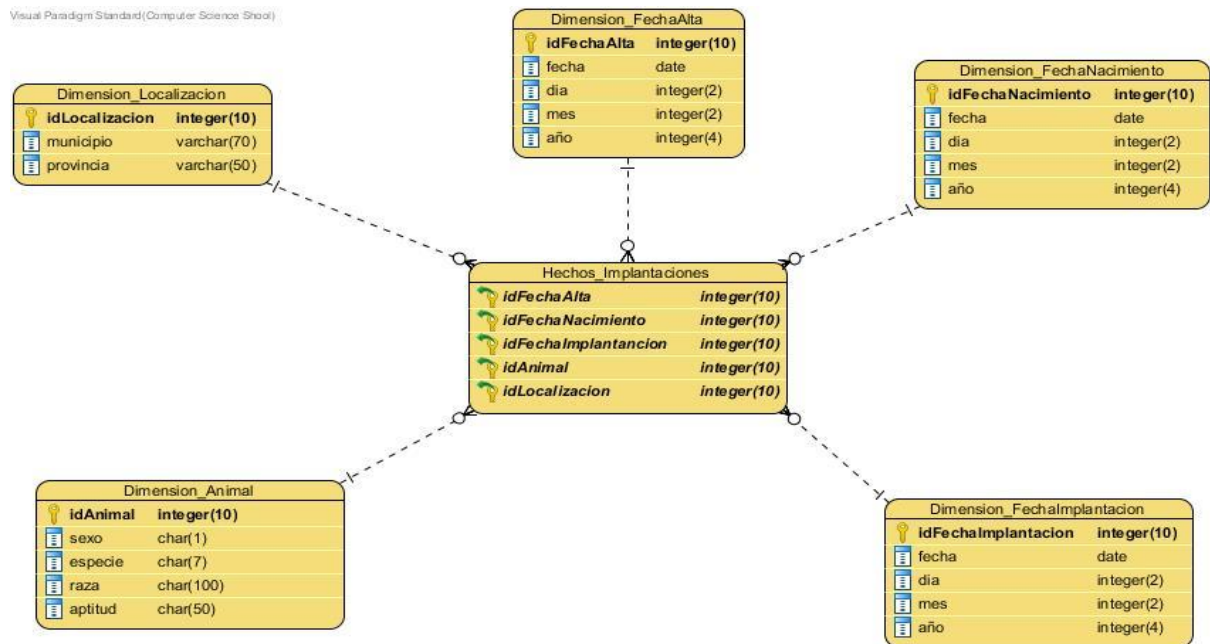
Miércoles, 22 Febrero 2017

## 1. Diseño preliminar del DW en estrella

Una vez seleccionado el tema y los datos con los que trabajaremos hemos procedido a realizar un diseño en papel de cómo quedaría finalmente el data warehouse y cuáles eran las tablas de hechos y dimensiones necesarias, así como los campos de cada una de las tablas.

Este diseño puede ser que cambie a lo largo del proyecto, ya que puede que encontremos errores o mejores formas de hacer el data warehouse.

El diseño en estrella que en primer lugar se desea conseguir es el siguiente:



**Jueves, 23 Febrero 2017**

## **1. Limpieza de los datos del archivo CSV**

Después de realizar el diseño del data warehouse hemos llevado a cabo la limpieza de los datos del csv eliminando los campos innecesarios para nuestro data warehouse y las filas que contenían datos erróneos o intangibles.

La limpieza del archivo .csv se realiza en un archivo Python llamado "cleanData.py" donde se leen todos los datos del archivo original, se realiza la limpieza de campos y registros necesaria y se guardan los datos en un nuevo archivo .csv llamado "datos.csv". Solo nos quedamos con los campos que son necesarios para rellenar las tablas del data warehouse.

**Viernes, 24 Febrero 2017**

## **1. Limpieza de datos CSV**

Hemos continuado con la limpieza del archivo .csv ya que nos hemos dado cuenta de que ciertos caracteres como la "ñ" o las tildes no las reconoce bien así que había que corregirlo.

## **2. Creación de tablas del Data Warehouse**

A continuación, hemos creado otro archivo .py, llamado "createTables.py", para la creación de la base de datos así como la creación de las tablas de dimensiones y hechos necesarias para el data warehouse.

En este archivo aparte de la creación de las tablas de hechos y dimensiones se crea una tabla llamada "ANIMALES" donde se guardan todos los datos almacenados en el archivo "datos.csv", y es la usada para el filtrado y selección de los datos que serán insertados en las tablas de hechos y dimensiones.

## **3. Inserción de datos en las tablas de dimensiones**

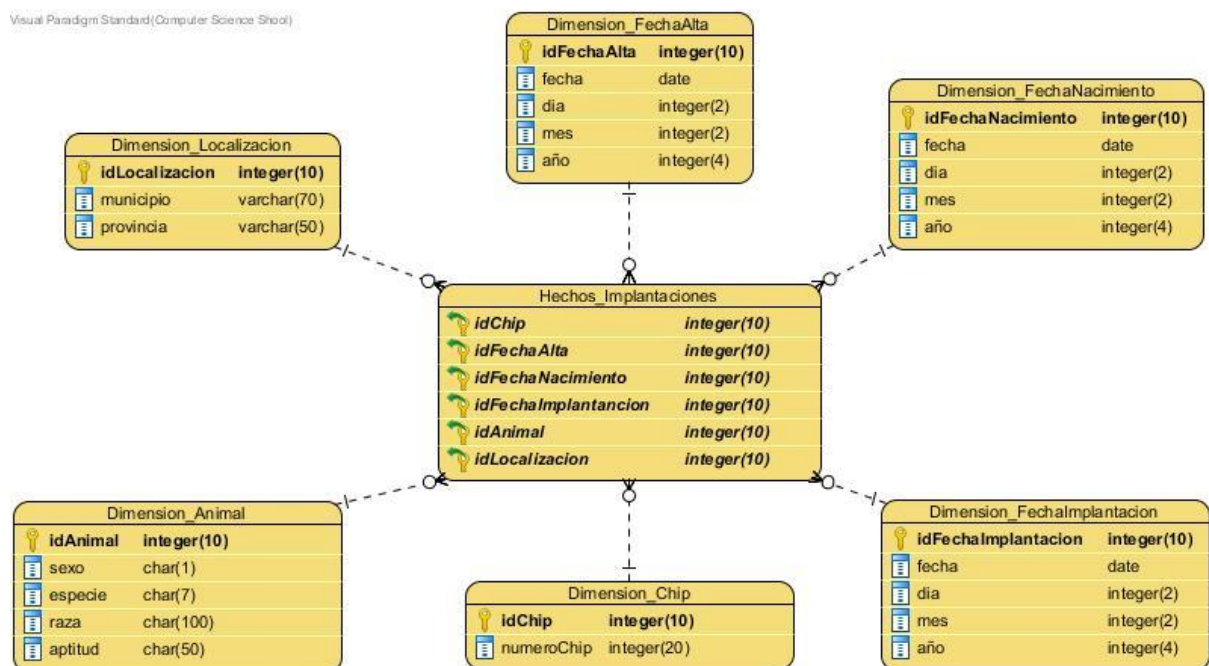
También se ha creado otro archivo .py, llamado "insertData.py", para la extracción de datos del archivo "datos.csv" e introducirlos en la tabla de "ANIMALES", así como la inserción de los datos correspondientes en las tablas de dimensiones.

Sábado, 25 Febrero 2017

## 1. Modificación del diseño en estrella

Nos hemos dado cuenta que podría ser útil añadir una tabla de dimensión para el chip para poder diferenciar a cada animal, ya que el resto de campos podrían coincidir pero el número de chip es un valor único.

El nuevo diseño en estrella es el siguiente:



## 2. Modificación de los archivos existentes e inserción de datos en las tablas de dimensiones

Hemos modificado los archivos ya que teníamos que crear una nueva tabla de dimensiones y hemos continuado con la inserción de datos en las tablas de dimensiones, y corregido algunos errores que se producían.

**Lunes, 27 Febrero 2017**

### **1. Inserción de los datos en la tabla de hechos**

Hemos introducido los datos correspondientes en la tabla de hechos haciendo uso de de la tabla de ANIMALES que es donde están todos los datos almacenados, y las tablas de dimensiones para introducir en la tabla de hechos cada identificador correspondiente en función de los datos que haya en la tabla de ANIMALES.



**Martes, 28 Febrero 2017**

**1. Dudas acerca de aplicar el término medida a nuestro Data WareHouse.**

Se ha preguntado sobre las medidas y que medidas interesantes se podrían añadir a la tabla de hechos. Se nos ha recomendado el siguiente enlace: [Almacenes de Datos Matilde Celma](#) para poder comprender mejor las medidas a través de ejemplos, etc.

**2. Revisión del LabBook para la entrega parcial.**

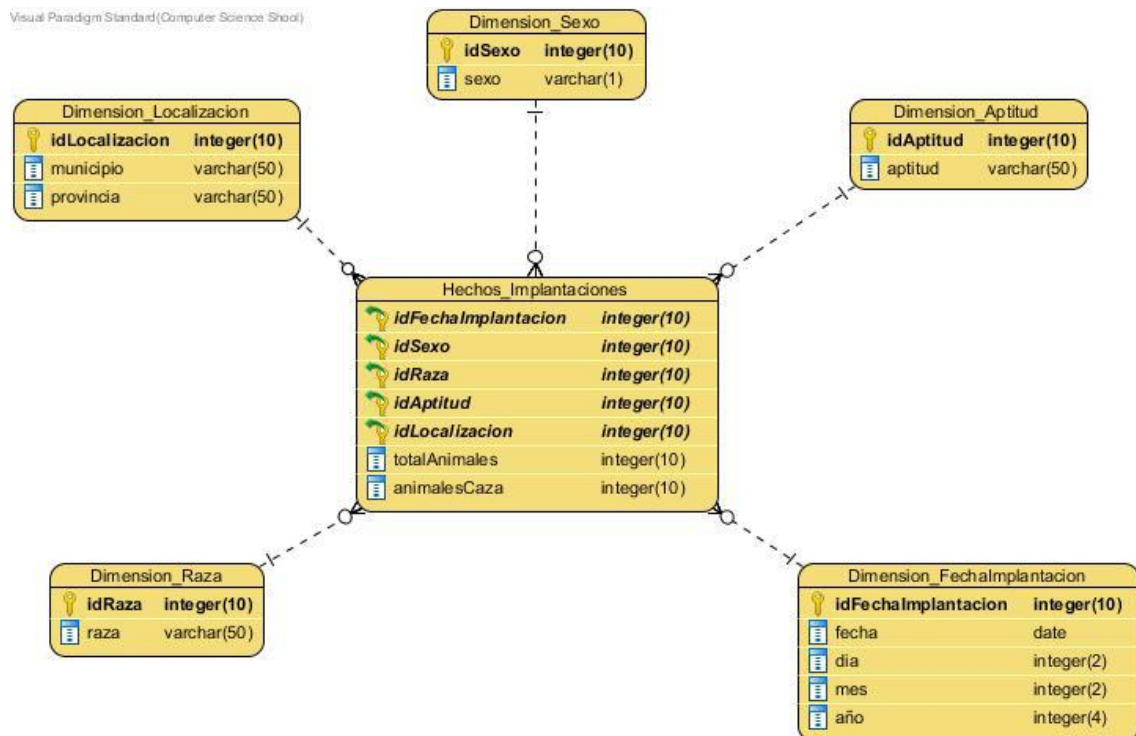
**3. Creación y revisión del documento de referencias bibliográficas y resumen propuesta.**

Miercoles, 8 Marzo 2017

## 1. Modificación del diseño en estrella

En el proceso de la creación e inserción de las medidas nos hemos dado cuenta de que teníamos que modificar el diseño del Data Warehouse ya que teníamos tablas de dimensiones que no eran necesarias para las medidas que habíamos creado.

El diseño en estrella que finalmente hemos usado es el siguiente:



## 2. Modificación tablas de dimensiones y hechos

Una vez modificado el diseño en estrella hemos tenido que modificar los archivos utilizados para la creación de las tablas e inserción de los datos en el Data Warehouse.

**Jueves, 9 Marzo 2017**

### **1. Realización de algunas consultas**

Hemos realizado una serie de consultas que se pueden hacer al Data Warehouse para mostrar su funcionalidad y el uso para el que se ha diseñado.

## **Viernes, 10 Marzo 2017**

Hemos comprobado que todo funcionaba correctamente y modificado algún archivo si era necesario.

También, hemos terminado la documentación y realizado la presentación del trabajo.

## **ANEXO**

### **Miércoles, 22 Febrero 2017**

#### **Tiempo dedicado**

2 horas

#### **Problemas encontrados**

Decisión de separar los campos "sexo", "especie", "raza" y "aptitud" en cuatro tablas de dimensiones o juntarlos en una tabla de dimensión llamada "Dimension\_Animal".

### **Jueves, 23 Febrero 2017**

#### **Tiempo dedicado**

4 horas

#### **Problemas encontrados**

Comprobar los distintos datos introducidos en el csv para encontrar posibles datos incoherentes.

### **Viernes, 24 Febrero 2017**

#### **Tiempo dedicado**

3 horas

#### **Problemas encontrados**

Problemas con la codificación que tenía el csv ya que no reconocía bien ciertos caracteres como la “ñ” o las vocales con tilde.

### **Sábado, 25 Febrero 2017**

#### **Tiempo dedicado**

3 horas

**Lunes, 27 Febrero 2017**

**Tiempo dedicado**

2.5 horas

**Martes, 28 Febrero 2017**

**Tiempo dedicado**

4 horas

**Miércoles, 8 Marzo 2017**

**Tiempo dedicado**

4 horas

**Jueves, 9 Marzo 2017**

**Tiempo dedicado**

3 horas

**Viernes, 10 Marzo 2017**

**Tiempo dedicado**

3 horas