**MODELOS Y PERSISTENCIA DE DATOS**

**DOCUMENTO PROYECTO**

**Mote e’ Queso**

**Alfonso Caro**

**Elver Flores**

**Giovanni Lopez**

**Yesid Gutierrez**

|  |  |  |
| --- | --- | --- |
| **Versión** | **Fecha** | **Descripción de la modificación** |
| 1.0 | 28/08/2019 | Documento inicial |

1. DESCUBRIMIENTO
2. Customers

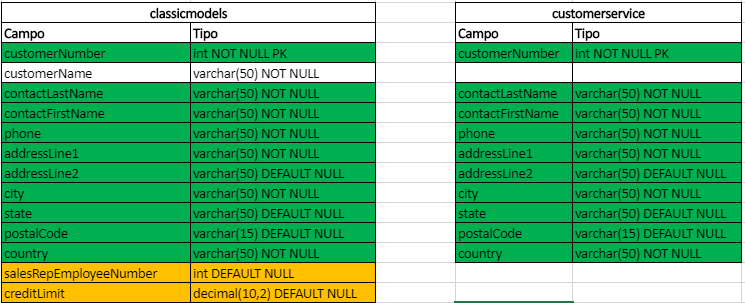


Ilustración 1 Descubrimiento de clientes

1. Employees

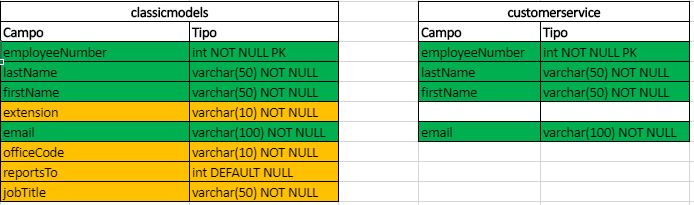


Ilustración 2 Descubrimiento de empleados

1. Products

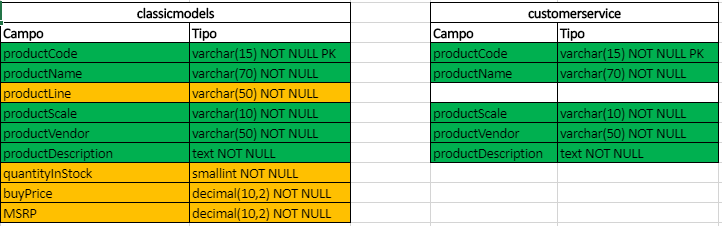


Ilustración 3 Descubrimiento de productos

1. PERFILAMIENTO
2. Rangos de los datos (máximos y mínimos)

Los tipos de datos de los campos de las tablas customers, employees y products que se encuentran en las dos bases de datos tienen los mismos tipos de datos.

1. Porcentaje de datos nulos

Base de datos classicmodels

|  |  |  |
| --- | --- | --- |
| **Tabla** | **Campos** | **Porcentaje** |
| Employees | reportsTo | 4% |
| Offices | addressLine2 | 28% |
|  | state | 43% |
| Productlines | htmlDescription | 100% |
|  | Image | 100% |
| Customers | addressLine2 | 82% |
|  | State | 60% |
|  | postalCode | 6% |
|  | salesRepEmployeeNumber | 18% |
|  | creditLimit | 0% |
| Orders | shippedDate | 4% |
|  | comments | 75% |

Ilustración 4 Porcentaje de datos nulos en classicmodels

Base de datos customerservice

|  |  |  |
| --- | --- | --- |
| **Tabla** | **Campos** | **Porcentaje** |
| Customers | addressLine2 | 81% |
|  | State | 60% |
|  | postalCode | 6% |

Ilustración 5 Porcentaje de datos nulos en customerservice

1. Estructuras en común

* Toda la información de la tabla customers está igual.
* Toda la información de la tabla employees es diferente.
* Toda la información de la tabla products está igual.

1. BODEGA DE DATOS
   1. ODS

Despues del descubrimientos de datos realizado se decide que las tablas para el ODS tomará la información de la siguiente forma:

Tabla **employees**: la fuente de datos será *classicmodels* y *customerservice*, la tabla de *classicmodels* tienes mas información, por lo tanto la tabla resultante en el ODS tendra los mismos campos de *classicmodels* con la unica diferencia que se establecen como nulos los campos *extension* y *officeCode*, *jobTitle* se le define un valor por defecto para customer services en cual será *sales*.

Tabla **customers**: ambas fuentes tienen la misma información de los clientes, pero al ser la tabla *classicmodels* mas completa, esta será la fuente de información para el ODS.

Tabla **products**: ambas fuentes tienen la misma información de los clientes, pero al ser la tabla *classicmodels* mas completa, esta será la fuente de información para el ODS, estableciendo como null por defecto los campos *productLine*, *quantityInStock*, *buyPrice*, *MSRP* ya que en la fuente customerservice no se cuenta con esta información.

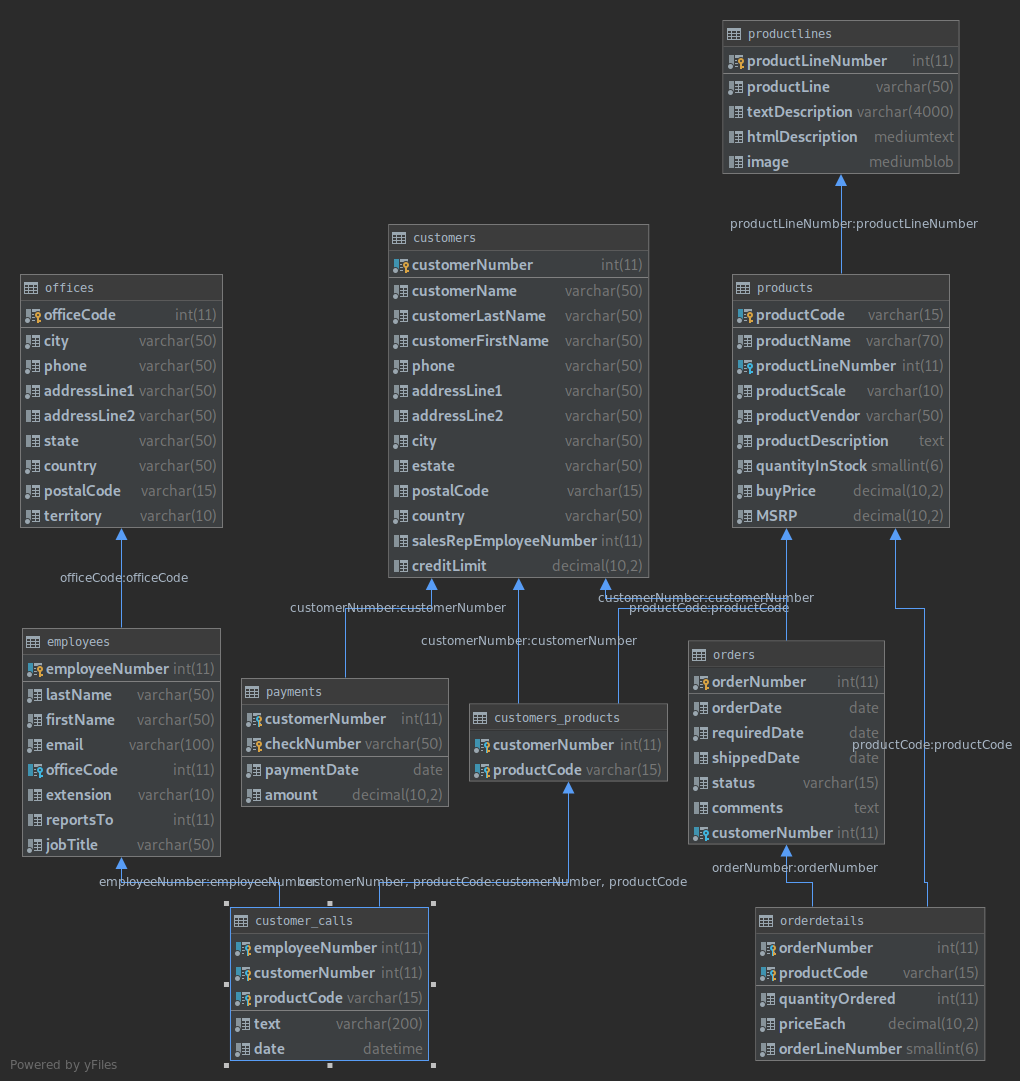
****

Ilustración 6 ODS SQL Schema

1. DWH

En la siguiente ilustración se evidencian las FACTs y los DIMs que se definieron como resultado del analisis sobre las fuentes de datos y lo que podriamos presentar al DWH

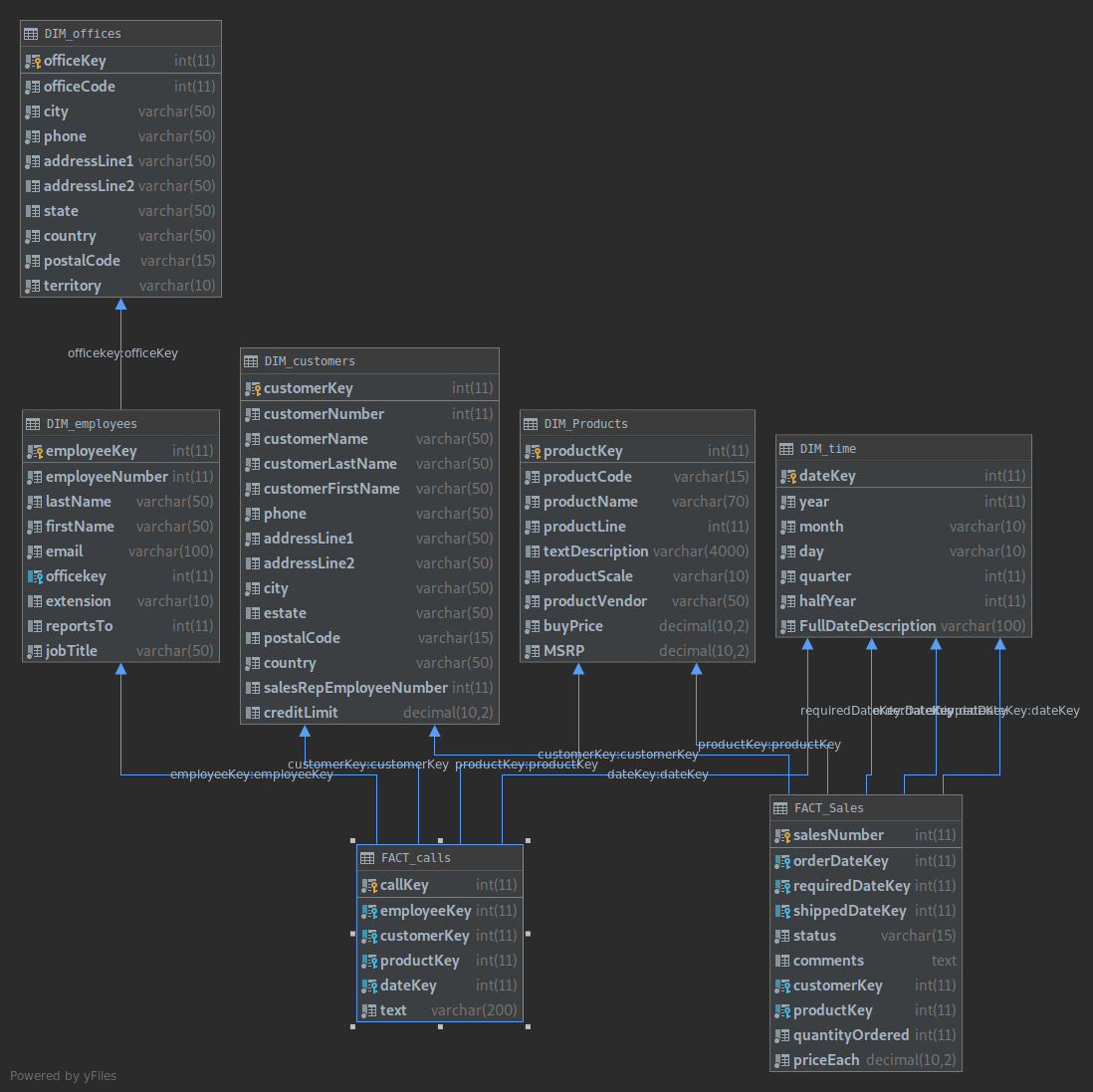
****

Ilustración 7 DWH Schema