



**SOLID
QUALITY
MENTORS**

Información Geográfica en el Data Warehouse

Javier Torrenteras

jtorrenteras@solidq.com

Salvador Ramos

sramos@solidq.com

Comencemos con preguntas 😊



- ¿Has creado una dimensión Geografía?
- ¿Qué atributos tenía?
- ¿Qué jerarquías tenía?
- Tenía una jerarquía:
 - País / Provincia / Población / Código Postal?
- ¿Cómo resolviste el caso de España?

Atributos

Código Postal
Hemisferio **Estado** **Localidad**
Continente **Departamento**
País **Zona**
Región **Población**
Provincia **Calle**
Ciudad **Dirección** **Cantón**
Municipio **Parroquia**

Atributos Geoespaciales

Coordenada X

Longitud

Coordenada Y

Altitud

Latitud

Posicion Geografica

Atributos de Negocio !!!

Nº de habitantes
Almacén **Zona de compras**
Tamaño: Pequeña/Mediana/Grande
Temperatura máxima **Zona**
Idioma **Clasificación A**
Clasificación B **Clima** **Sede**
Temperatura mínima
Municipio **Area de ventas**

Jerarquías

País	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
España	Región / Comunidad autónoma	Provincia	Población	Localidad / Pedanía / Zona Postal	C. Postal
Francia		Departamento	Población		C. Postal
Ecuador		Provincia	Cantón	Parroquia	C. Postal
Argentina		Provincia	Población		C. Postal
USA		Estado	Población		C. Postal
		Vaticano			

Jerarquías de Negocio !!!

Región de Ventas

Zona de Ventas

SubZona de Ventas

Zona de Compras

SubZona de Compras

Región

Dirección Territorial

Sede

Empresa

País

Centro de trabajo

Códigos Postales en España

Localidad != Población

Atención al cliente | Asesoría

Particulares Empresas Oficina Virtual

Códigos postales | Oficinas | Localizador

Particulares > Códigos postales > Códigos postales

Consulta códigos postales

Búsqueda por provincia

Por favor, seleccione País, provincia y localidad:

País: España Andorra

Provincia: MADRID

Localidad: MADRID

Dirección¹: Número:

(1) Escriba la dirección sin incluir el tipo de vía

Buscar

ESPAÑA - MADRID				
Localidad	Calle	Impar	Par	C.P.
MADRID				Delimitar búsqueda

Atención al cliente | Asesoría

Oficina

Localidad: goñar

Dirección¹: Número:

(1) Escriba la dirección sin incluir el tipo de vía

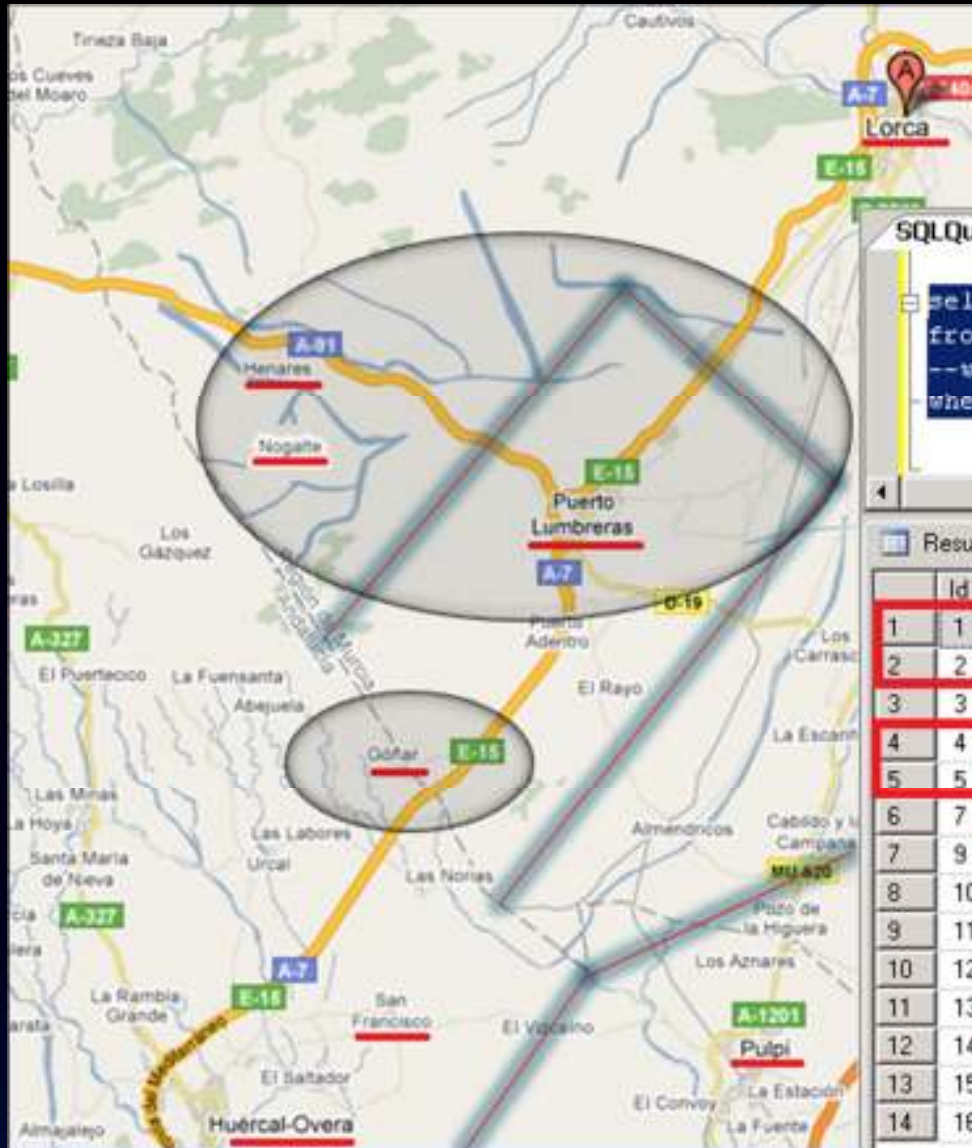
Buscar

España					
Provincia	Localidad	Calle	Impar	Par	C.P.
ALMERIA	GOÑAR				04690
MURCIA	GOÑAR				30891

España		
Provincia	Localidad	C.P.
ALMERIA	GOÑAR	04690
ALMERIA	NORIAS, LAS (HUERCAL OVERA)	04690

España		
Provincia	Localidad	C.P.
MURCIA	ESPARRAGAL DE PUERTO LUMBRERAS	30891
MURCIA	GOÑAR	30891
MURCIA	PUERTO ADENTRO	30891

Códigos Postales en España



SQLQuery1.sql - (local)\...\Admi...)*

```
select *  
from MiGeografia  
--where Pedania = 'goñar'  
where CodPostal in(30800, 30890, 30891, 04600, 04690)
```

Results Messages

	Id	CodPostal	Provincia	Poblacion	Pedania
1	1	04600	ALMERIA	PULPI	LOS AZNARES
2	2	04600	ALMERIA	HUERCAL OVERA	HUERCAL OVERA
3	3	04600	ALMERIA	HUERCAL OVERA	SAN FRANCISCO
4	4	04690	ALMERIA	HUERCAL OVERA	GOÑAR
5	5	30891	MURCIA	PUERTO LUMBRERAS	GOÑAR
6	7	30891	MURCIA	PUERTO LUMBRERAS	ESPARRAGAL
7	9	30890	MURCIA	PUERTO LUMBRERAS	PUERTO LUMBRERAS
8	10	30890	MURCIA	LORCA	NOGALTE
9	11	30890	MURCIA	LORCA	HENARES
10	12	30890	MURCIA	LORCA	ZARZALICO
11	13	30890	MURCIA	LORCA	BEJAR
12	14	30800	MURCIA	LORCA	LORCA
13	15	30800	MURCIA	LORCA	PULGARA
14	16	30800	MURCIA	LORCA	TIATA

Códigos Postales en España

- 1 Código Postal → Varias Localidades
- 1 Localidad → Varios Códigos Postales
- 1 Población → Varios Códigos Postales
- 1 Población → Varias Localidades
- 1 Localidad → Varias Poblaciones
 - *Localidad = Pedanía = Zona Postal*
- ¿Necesitamos el concepto de “Localidad” del servicio de Correos en nuestro Negocio?
 - ¿Utilizaremos la base de datos del servicio de Correos?

Códigos Postales en España

- Si creamos una dimensión geografía que incluya el atributo Código Postal, deberemos incluir una columna "Localidad" (Pedanía o Zona).
- La Business Key será: Provincia, Población, "Localidad", Código Postal
 - *En otro caso, podría devolver más de una fila*
- En otros países no existiría el nivel "Localidad"
- Podemos no incluir el atributo Código Postal en la dimensión Geografía, dejándolo en otras dimensiones como Cliente, Proveedor, Empleado

Creando la Dimensión Geografía

- Incluyamos Claves Subrogadas
 - SCD
- Seleccionemos los atributos necesarios
 - Habituales (“oficiales”)
 - Geo-espaciales
 - Y de Negocio !!!
- Seleccionemos las jerarquías necesarias
- Analizando siempre nuestras necesidades de negocio (actuales y a medio plazo)
 - No subamos el nivel de complejidad “por si acaso”

Jerarquía Típica

- Común a gran número de casos



- En España, el Código Postal no es un nivel inferior a Población
 - *Población – Código Postal es una relación N:M*

Origen de datos

- Identificar orígenes de datos
 - Habitualmente varios
- Identificar estructura de tablas
 - De cada uno de los orígenes de datos

Origen de datos

- Tablas que sólo tienen datos geográficos
 - Tablas: País, Provincia, Estado, Población, Códigos Postales
- Claves externas desde otras entidades
 - Cliente: CódigoPoblación, CódigoProvincia, CódigoPaís
 - Proveedor: CódigoPoblación, CódigoProvincia, CódigoPaís
 - Empleado: CódigoPoblación, CódigoProvincia, CódigoPaís

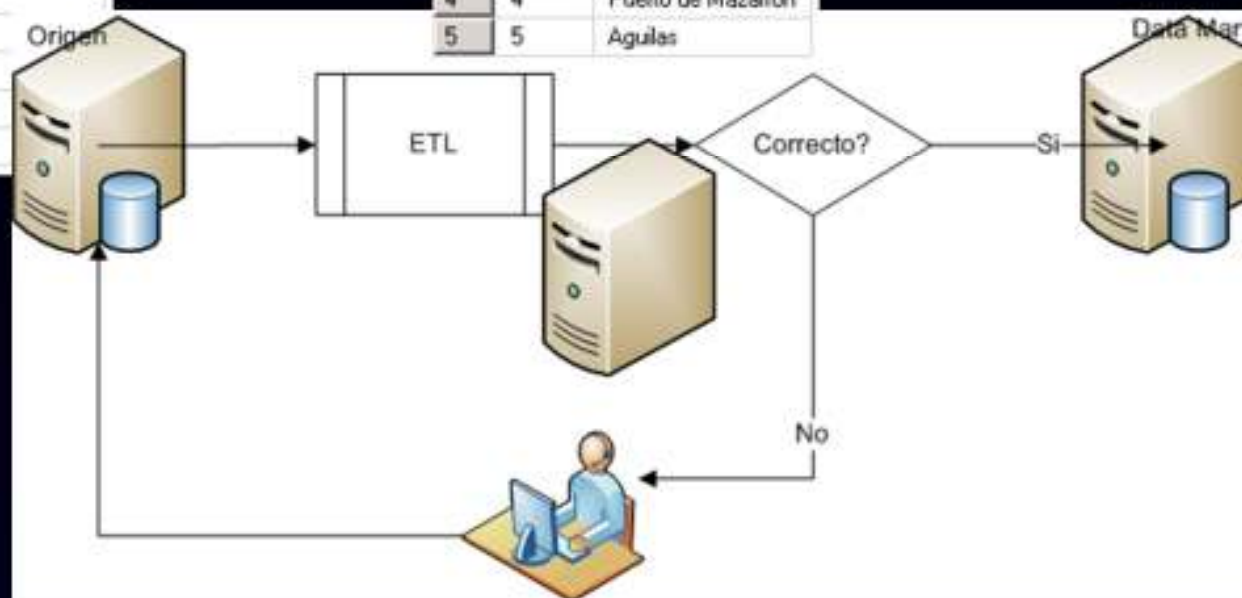
Origen de datos

- No hay tablas que sólo tienen datos geográficos
- Datos incluidos en otras entidades (como strings)
 - Cliente: NombrePoblación, NombreProvincia, NombrePaís
 - Proveedor: NombrePoblación, NombreProvincia, NombrePaís
 - Empleado: NombrePoblación, NombreProvincia, NombrePaís
- Este caso requiere mayores esfuerzos para la limpieza de datos

Limpiando datos

	Codigo	Nombre	Poblacion
1	1	Pepe	Pto. Lumberas
2	2	Juan	Puerto Lubreras
3	3	Salvador	Puerto Lumberas
4	4	Jose	Puerto Mazarron
5	5	Felipe	Puerto de Mazarron
6	6	Pedro	Puerto de Mazarrón
7	7	Andres	Muca
8	8	Alejandro	Murcia
9	9	Miguel	Aglas
10	10	Carlos	Aguila
11	11	Antonio	Aguilas

	Codigo	Nombre
1	1	Puerto Lumberas
2	2	Lorca
3	3	Murcia
4	4	Puerto de Mazarrón
5	5	Aguilas



DEMO



- Limpiando datos
 - Fuzzy Lookup

Consideraciones en Analysis Services

- No permite guardar atributos de tipo Geo-espacial
- Si tenemos distintos niveles en nuestra jerarquía, podemos utilizar “Ragged Hierarchies” (*jerarquías irregulares*)
 - Propiedad HideMemberIf

Mostrando los datos en mapas

- ¿Qué nivel de detalle quiero mostrar?
 - Continente, provincia, región, calle
- ¿Necesito ahora añadir información geoespacial?
 - No (siempre)
- ¿De donde obtengo los mapas de España?
 - Instituto nacional de estadística
 - <http://www.ine.es/prodyser/pcaxis/pcaxis.htm>
 - Se pueden cruzar con nuestros datos a nivel de nombre de población o provincia
 - Bing Maps
 - Se pueden cruzar con nuestros datos geoespaciales

DEMO



- Mostrando datos en mapas

Conclusiones

- La Dimensión Geografía, junto con la dimensión Tiempo son de las más habituales
- Es una dimensión importante para el Negocio
- Analicémosla siempre desde el punto de vista de ese negocio en particular
- No compliquemos el diseño "por si acaso"
 - Tener países con distinto número de niveles añade complejidad
- Una buena limpieza aportará Valor al Negocio
- Facilitemos la representación gráfica en mapas

¿Preguntas?

