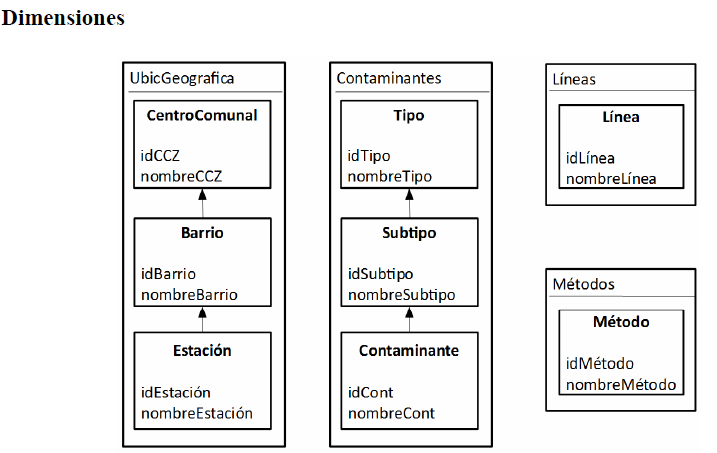
Un informe que describa la solución propuesta. Este documento deberá incluir al

menos:

**Análisis de** **requerimientos**.

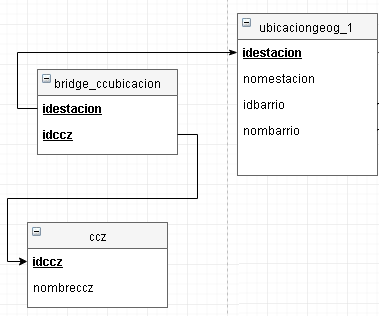
**Diseño conceptual**

El diseño conceptual del que decidimos partir fue el de la solución final suministrado por las profesoras.



**Diseño lógico**

Cuando analizamos los datos de los barrio y centros comunales y previo a realizar la carga de la ubicación geográfica, nos topamos con los casos en que había barrios que pertenecían a más de un centro comunal, para resolver esto decidimos modificar el diseno y resolverlo mediante una bridge table entre la tabla ubicación geográfica y los centros comunales como hicimos en la entrega anterior pero respecto a los problemas de los hogares:



Esto sucede por ejemplo con el barrio centro que pertenece al ccz01 y ccz02, "MERCADO MODELO Y BOLIVAR" que tiene el ccz 03 y ccz 11, "PUNTA GORDA" con el ccz07 y ccz08 entre otros.

d. Implementación de las relaciones dimensionales y dimensiones sobre

*Pentaho BI Server* (archivos .xml generados)

**Tablas y proceso de carga**

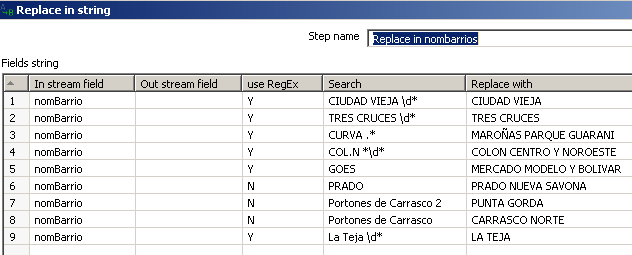
Ubicación geográfica, idestacion y nombre de los barrios.

La fuente utilizada fue aire-estaciones.csv, del campo explicación tratamos de obtener el barrio pero nos encontramos que en varios casos no nos coincidían con los nombres de los barrios del archivo barrios.shp, por lo tanto luego de analizar los datos ya que no eran muchas estaciones, decidimos en varios casos sobreescribir el barrio en forma fija ya que por ejemplo la estación 04 que se encuentra en Museo blanes devuelve el barrio aires puros al utilizar las coordenadas suministradas, pero en verdad pertenece al barrio

"PRADO NUEVA SAVONA", esto lo verificamos además buscando la dirección en el siguiente enlace para confirmalo:

<http://sig.montevideo.gub.uy/mapas/mapa-principal>

En resumen, la modificación realizada fue:



Lineas: para esta tabla se utilizo una base intermedia aux que contenia el contenido del archivo v\_uptu\_lsv.shp en una tabla de nombre líneas, para la clave principal se utilizo el campo gid que venia en dicho archivo y como nombre de la línea se realizo una combinación entre el numero de la línea y uno de los destinos, es decir que si una fila era

"16009409";"405";"PEÑAROL - PARQUE RODÓ";"B"

El valor B indica que el sentido es hacia el sur del país y si dice A, indica que el destino es hacia el norte, por lo que si dice B se decidio elegir para el nombre el segundo destino, asi que caso en nuestra tabla líneas de la base dwh, la tupla clave;nombre será

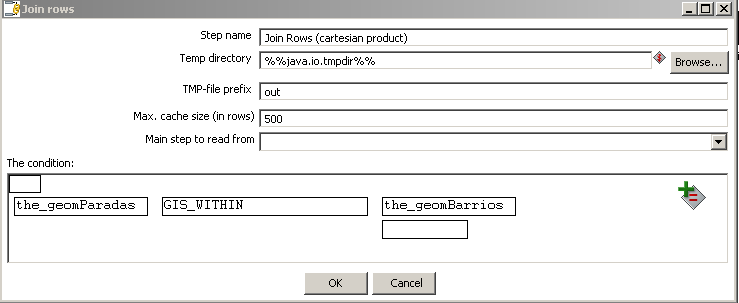
"16009409";"405 - PARQUE RODÓ"

De esta forma nos parecio similar a como sucede en la vida real.

Relacionar las líneas con las paradas y con los barrios.

Para esta parte se utilizo geokettle ya que no fue posible utilizar la extensión postgis porque nos devolvía un resultado vacio al utilizar la función whitin.

Lo primero que se hizo fue cargar los shapefiles de paradas y líneas, luego se realizo un join por el campo cod\_variant, una vez obtenido el resultado se procedio a realizar otro join con el shapefile de los barrios por el campo geom yla función whitin:



De esta forma obtuvimos los ómnibus a los cuales les corresponden las paradas indicadas y posteriormente que paradas pertenecen a que barrio.

Luego se procedio a eliminar las filas que no eran necesarias y finalmente nos quedamos con los campos gid-nrobarrio (lineas\_barrios) que se guardó en una tabla de la base auxiliar para luego poder relacionar a las líneas con la estación en la en la carga de las mediciones.

f. Documentación sobre todos los componentes que se incluyeron en la

solución para satisfacer los requerimientos.

**Problemas de calidad de datos**

Los problemas encontrados en los datos fueron varios, algunos los detectamos al analizar los datos y en otros casos por la cantidad de los mismos, aparecieron cuando realizábamos la carga a medida que teníamos errores.

Los problemas que encontramos y resolvimos fueron los siguientes:

1. Contaminantes:
   1. Fuente: contaminantes.txt
   2. Problema: conversion de string a integer, había espacios al final de los nombres de los tipos, no era lo mismo ‘primario’ que ‘primario  ’
   3. Solucion: hubo que aplicar trim.
2. Mediciones
   1. Fuente: archivos csv de datos históricos de mediciones de aire.
   2. Problema 1: en las delimitaciones por coma, en el campo que hacia referencia a idmetodo, en algunos casos había espacios al final del nombre y antes de la coma por lo cual la fk de la tabla mediciones hacia la tabla métodos no era valida.
   3. Solucion aplicar trim en el campo idmetodo de los csv al cargar el archivo en el ktr de carga de mediciones.
   4. Problema 2 Los fuentes no tienen medidas de la siguiente estacion: Colon ,Corresponde a E8 (Colón) Ubicacion Laboratorio MTOP
   5. Solucion 2 se dejo ingresada para el futuro la estación en la tabla ubicación geográfica aunque no haya medidas para ella.
   6. Problema 3: datos de la polucion en los archivos csv que o no estaban o contenían texto y no valores numéricos
   7. Solucion3 se descartaron las filas.
3. Lineas:
   1. Fuente v\_uptu\_lsv.zip
   2. Problema 1 los caracteres especiales no se visualizaban correctamente.
   3. Solucion 1 se probo con varias codificaciones y la que funciono correctamente para lo tildes y enes fue UTF-8
   4. Problema 2, había una línea 0 con geom con valor null
   5. Solucion 2 se descarto esa línea porque no aportaba a la solución
   6. Problema 3 se trato de utilizar los puntos xy, pero no obtuvimos el resultado esperado.
   7. Solucion 3 se utilizo el campo geom
4. Paradas
   1. Fuente v\_uptu\_paradas.shp
   2. Problema Con la codificación por los caracteres especiales
   3. Solucion se utilizo la codificación UTF-8
5. Barrios
   1. Fuente Barrios.shp
   2. Problema Con la codificación por los caracteres especiales
   3. Solucion se utilizo la codificación IBM00858

h. Esbozo de un plan de testeo de la solución.