****

**UNIVERSIDAD TÉCNICA DE MACHALA**

Maestría en Software

**Asignatura:**

Ciencia de Datos e Inteligencia de Negocios

**Tema:**

## Tarea Práctica1. Diseño DW, proceso ETL y dashboard BI del caso Camaronera.

Docente: Ing. Bertha Mazón, Mg. Inf.

Estudiantes:

Ing. Jimmy Fernando Castillo Crespín

Ing. Paquita Alejandra Cuadros García

2021-2022

**Caso de estudio: diseño e implementación de un sistema de inteligencia de negocios aplicado al sector camaronero**

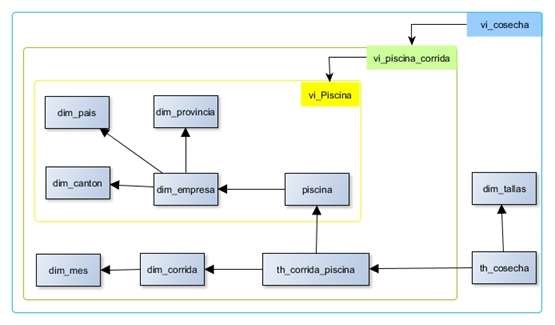
El consorcio de empresas ABC maneja N empresas que producen camarón; cada empresa tiene a cargo una camaronera que está ubicada en un país, provincia y cantón; Cada camaronera está organizada por piscinas. Cada piscina tiene un número de hectáreas de producción. El proceso de producción de camarón se denomina "corrida", cada corrida tiene un número secuencial, fecha de inicio y fin, estado (concluido o en proceso) y una o más piscinas donde se siembra el camarón. Por cada piscina que es parte de una corrida, se registra el costo de inversión, el número de larvas de camarón sembradas, la cantidad en Kilogramos de camarón cosechados clasificados por talla (Grande, Mediano y Pequeño) y el total de ingresos por la venta de la cosecha ($ por venta total y por talla).  La empresa puede tener una o más corridas activas, pero una piscina no puede estar en más de una corrida activa en el mismo periodo de tiempo.

Aplicando un proceso metodológico de inteligencia de negocios, realizar el diseño de un data warehouse que satisfaga los requerimientos planteados; luego, el proceso ETL con datos simulados desde Excel; a continuación, la creación de Cubos OLAP; y finalmente, el diseño e implementación de un dashboard BI como aplicación EIS. Aplicar el proceso metodológico de CRISP-DM. En este caso se omite la actividad de Evaluación.

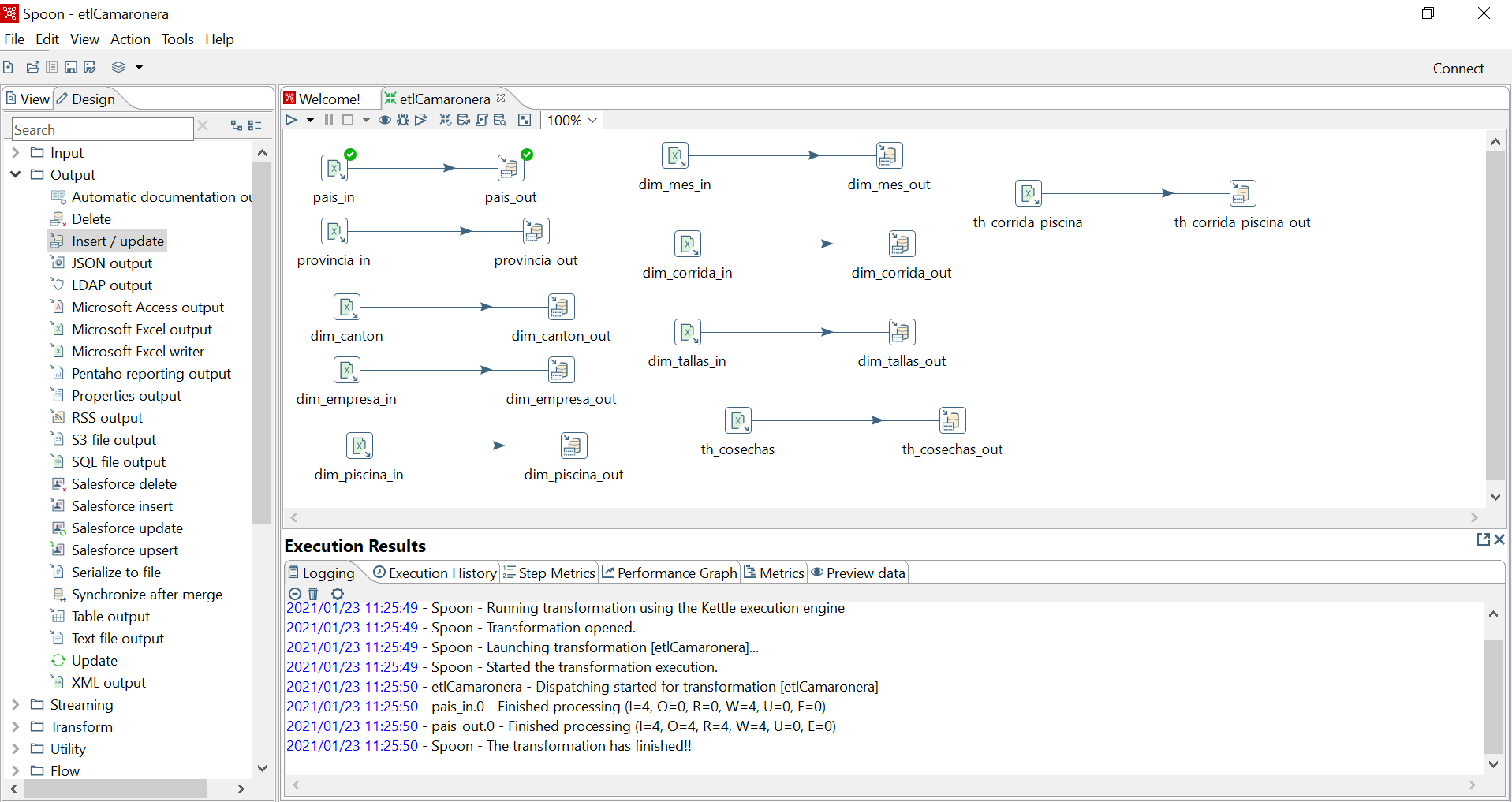
**Requerimientos del negocio (preguntas clave)**

1. ¿Cuántos **Kilogramos de camarón** se han producido por talla y empresa en el último año?
2. ¿Cuáles son **los gastos (inversión),** **los ingresos y la utilidad** por empresa en un año determinado?
3. ¿Cuántas **hectáreas de producción de camarón** tiene el consorcio por empresa, corridasactiva y piscina en el último año?
4. ¿Cuántas **hectáreas de producción de camarón** dispone el consorcio por empresa, piscina y país?
5. ¿Cuál es el **número** promedio **de larvas** **de camarón por hectárea** que se siembran por empresa y camaronera?
6. ¿Cuál es la empresa que ha generado **más** **utilidades** por año?
7. ¿Cuál es la empresa que ha **producido más camarón** por talla y año?

**Diseño del DataWarehouse**



**Proceso ETL**



Las sentencias SQL proporcionadas por este proceso:

CREATE TABLE dim\_canton

(

can\_cod int, can\_nom varchar(150), prov\_cod int

);CREATE INDEX idx\_dim\_canton\_lookup ON dim\_canton(can\_cod, can\_nom, prov\_cod)

CREATE TABLE dim\_empresa

(

emp\_cod int

, emp\_nom varchar(100)

, emp\_nom\_cam varchar(100)

, pais\_cod int

, prov\_cod int

, can\_cod int

);CREATE INDEX idx\_dim\_empresa\_lookup ON dim\_empresa(emp\_cod, emp\_nom, emp\_nom\_cam, pais\_cod, prov\_cod, can\_cod);

CREATE TABLE dim\_piscina

(

pis\_cod int

, pis\_nom varchar(20)

, pis\_num\_hect DOUBLE PRECISION

, emp\_cod int

)

;CREATE INDEX idx\_dim\_piscina\_lookup ON dim\_piscina(pis\_cod, pis\_nom, pis\_num\_hect, emp\_cod);

CREATE TABLE dim\_mes

(

mes\_codigo int

, mes varchar(20)

, trimestre varchar(10)

, semestre varchar(10)

);CREATE INDEX idx\_dim\_mes\_lookup ON dim\_mes(mes\_codigo, mes, trimestre, semestre);

CREATE TABLE dim\_corrida

(

cor\_cod int

, cor\_fec\_ini TIMESTAMP

, cor\_fec\_fin TIMESTAMP

, cor\_est varchar(20)

, anio int

, mes\_codigo int

);CREATE INDEX idx\_dim\_corrida\_lookup ON dim\_corrida(cor\_cod, cor\_fec\_ini, cor\_fec\_fin, cor\_est, año, mes\_codigo);

CREATE TABLE dim\_tallas

(

tall\_cod int

, tall\_nombre varchar(100)

, tall\_cam varchar(100)

);CREATE INDEX idx\_dim\_tallas\_lookup ON dim\_tallas(tall\_cod, tall\_nombre, tall\_cam);

CREATE TABLE th\_cosecha

(

cos\_cant\_kg DOUBLE PRECISION

, cos\_pre\_kg DOUBLE PRECISION

, cos\_ing DOUBLE PRECISION

, cor\_cod int

, pis\_cod int

, tal\_cod int

, emp\_cod int

);CREATE INDEX idx\_th\_cosecha\_lookup ON th\_cosecha(cos\_cant\_kg, cos\_pre\_kg, cos\_ing, cor\_cod, pis\_cod, tal\_cod, emp\_cod);

CREATE TABLE th\_corrida\_piscina

(

cp\_num\_larv int

, cp\_cos\_inv DOUBLE PRECISION

, cp\_total\_ing DOUBLE PRECISION

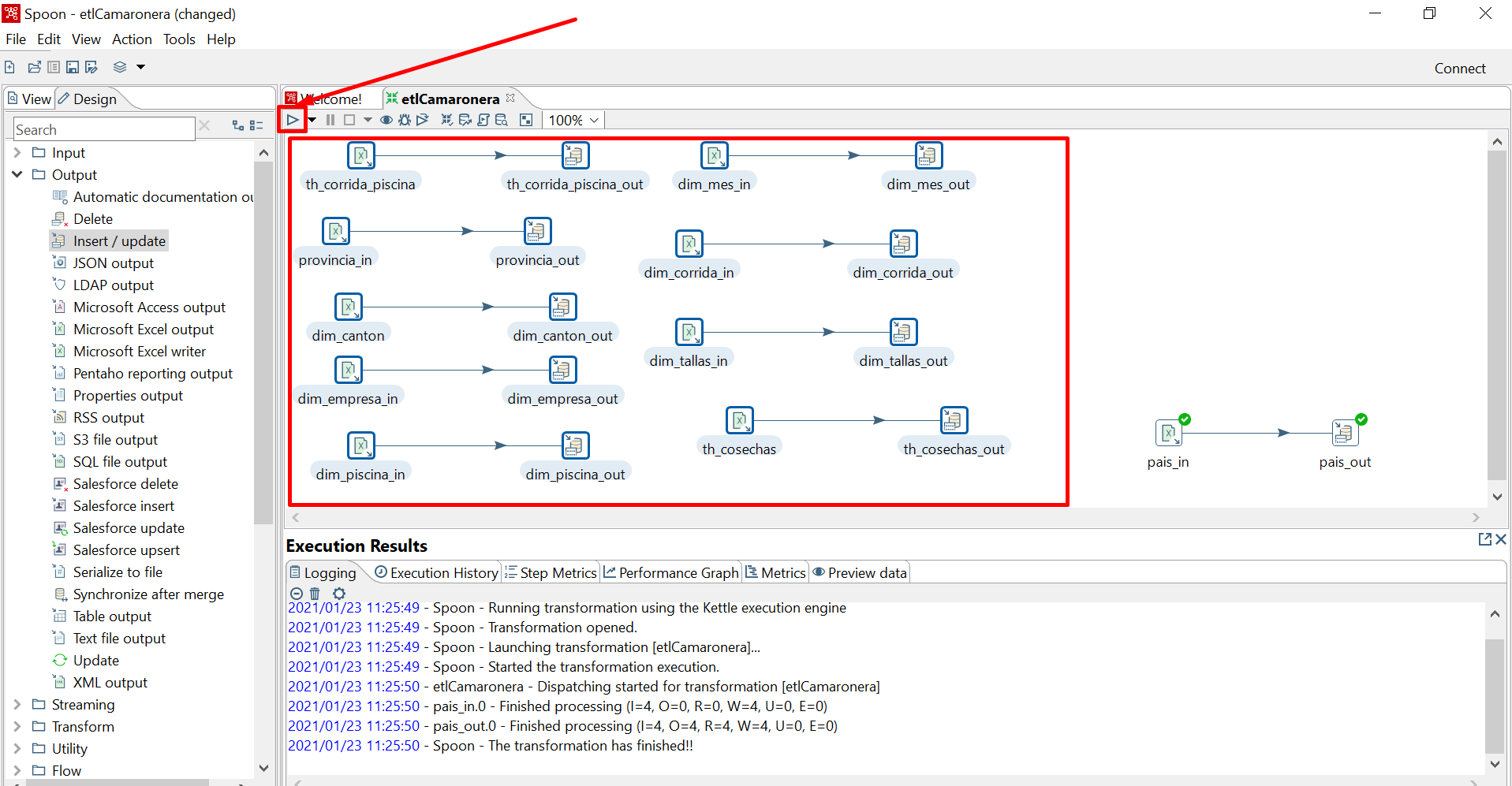
, emp\_cod int

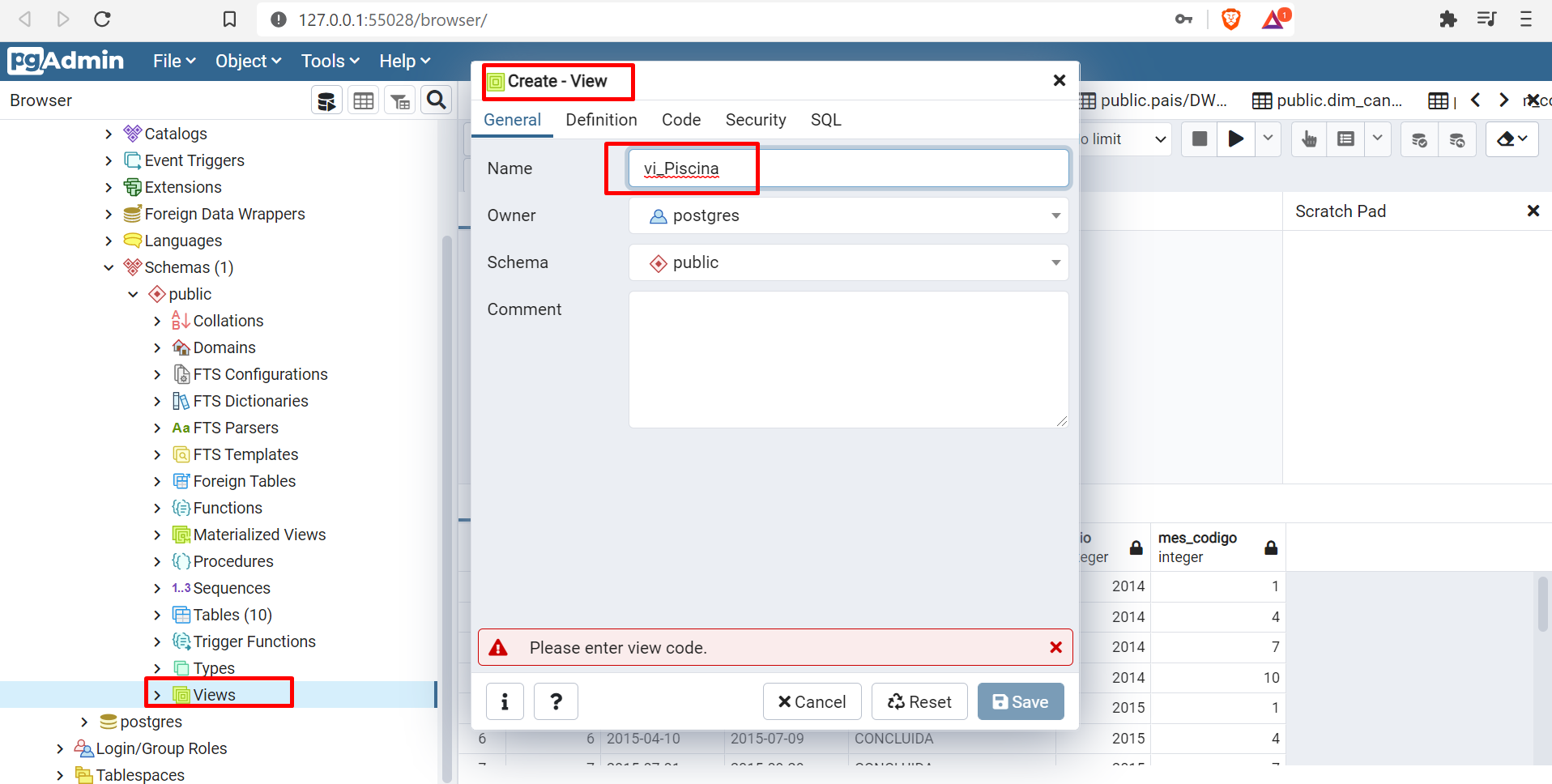
, pis\_cod int

, cor\_cod int

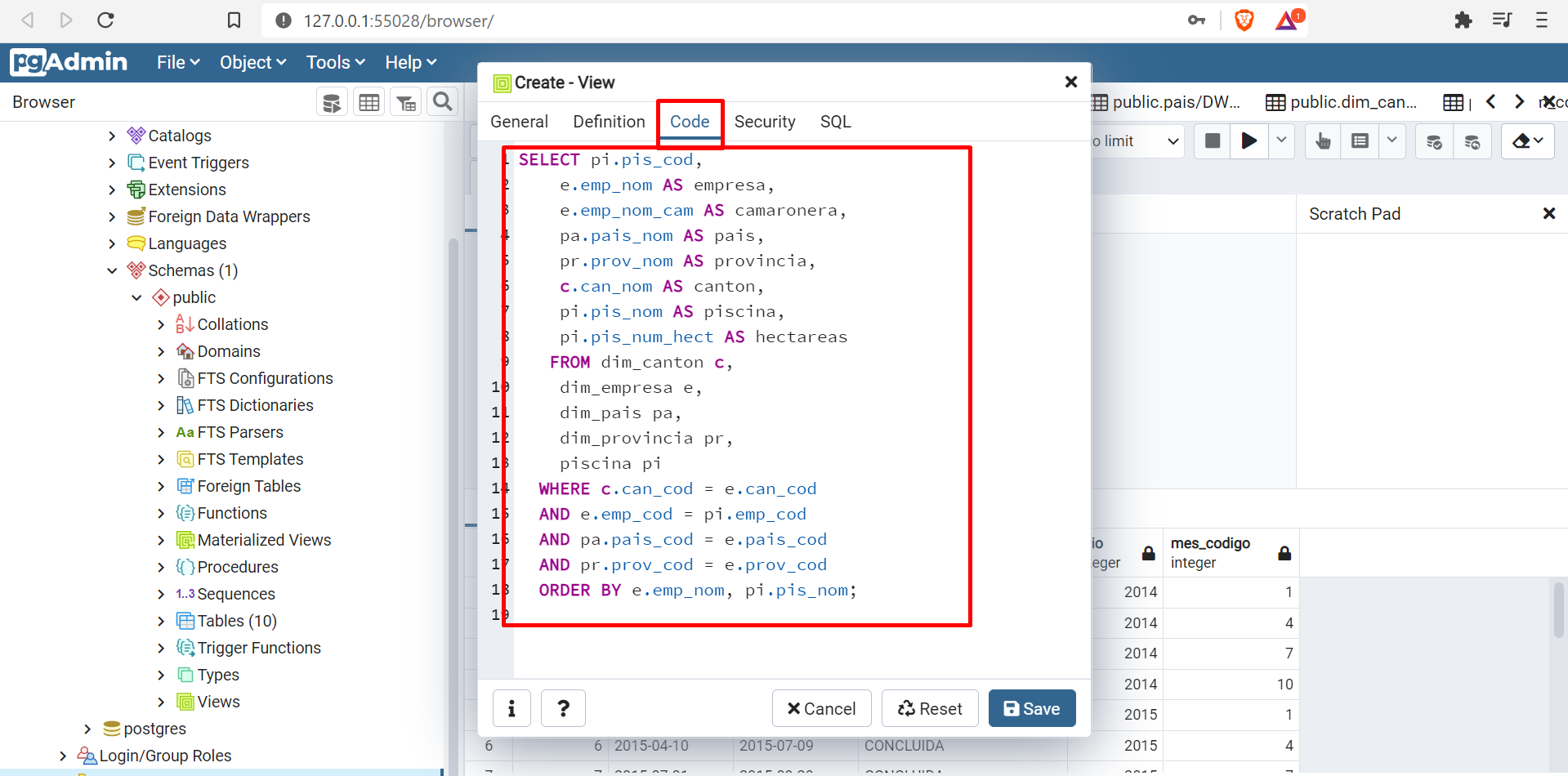
);CREATE INDEX idx\_th\_corrida\_piscina\_lookup ON th\_corrida\_piscina(cp\_num\_larv, cp\_cos\_inv, cp\_total\_ing, emp\_cod, pis\_cod, cor\_cod);

Ahora se procederá con el proceso de carga de los datos desde Pentaho Data Integration hacia la BD Postgresql, para eso se selecciona los archivos que desea ejecutar y dar click en el icono del play (run).

  
Ahora se procederá con la creación de las Vistas en PostgreSql, como ejemplo se tomará la vista para la piscina (vi\_Piscina), se debe dar click derecho en la Views, luego dar click en create y dar nombre a la vista, en este caso se llamará vi\_Piscina.



Luego en la pestaña code, codificamos la sentencia sql para crear la vista.



SCRIPT SQL VISTA **vi\_Piscina**

SELECT pi.pis\_cod,

e.emp\_nom AS empresa,

e.emp\_nom\_cam AS camaronera,

pa.pais\_nom AS pais,

pr.prov\_nom AS provincia,

c.can\_nom AS canton,

pi.pis\_nom AS piscina,

pi.pis\_num\_hect AS hectareas

FROM dim\_canton c,

dim\_empresa e,

dim\_pais pa,

dim\_provincia pr,

dim\_piscina pi

WHERE c.can\_cod = e.can\_cod

AND e.emp\_cod = pi.emp\_cod

AND pa.pais\_cod = e.pais\_cod

AND pr.prov\_cod = e.prov\_cod

ORDER BY e.emp\_nom, pi.pis\_nom;

Repetimos este proceso con el resto de vistas.

SCRIPT SQL VISTA vi\_piscina\_corrida

SELECT pi.empresa,

pi.camaronera,

pi.pis\_cod,

pi.piscina,

cr.cor\_cod AS corrida,

cr.cor\_est AS estado,

pi.pais,

pi.provincia,

pi.canton,

cr.anio AS "Año",

mes.semestre,

mes.trimestre,

mes.mes,

cp.cp\_num\_larv AS "Num larvas",

cp.cp\_num\_larv::double precision / pi.hectareas AS "Larvas por Hectarea",

cp.cp\_cos\_inv AS "Costo de inversion",

i.ingresos AS "Total ingresos",

round(( i.ingresos - cp.cp\_cos\_inv)::numeric, 3) AS "Utilidad monto",

round((( i.ingresos - cp.cp\_cos\_inv) / cp.cp\_cos\_inv \* 100::double precision)::numeric, 3) AS "Utilidad %"

FROM

dim\_mes mes,

"vi\_Piscina" pi,

dim\_corrida cr,

th\_corrida\_piscina cp,

(SELECT

c.pis\_cod,

c.cor\_cod,

sum(c.cos\_cant\_kg \* c.cos\_pre\_kg) AS ingresos

FROM th\_cosecha c

GROUP BY c.emp\_cod, c.pis\_cod, c.cor\_cod

ORDER BY c.emp\_cod, c.pis\_cod, c.cor\_cod) i

WHERE mes.mes\_codigo = cr.mes\_codigo

AND pi.pis\_cod = cp.pis\_cod

AND cr.cor\_cod = cp.cor\_cod

AND cp.pis\_cod = i.pis\_cod

AND cp.cor\_cod = i.cor\_cod

ORDER BY pi.empresa, pi.piscina, cr.cor\_cod;

SCRIPT SQL VISTA vi\_cosecha

SELECT pc.empresa,

pc.camaronera,

pc.piscina,

pc.corrida,

pc.estado,

pc.pais,

pc.provincia,

pc.canton,

pc."Año",

pc.semestre,

pc.trimestre,

pc.mes,

t.tall\_nombre AS "Talla",

c.cos\_cant\_kg AS "Cantidad kg",

c.cos\_pre\_kg AS "Precio kg",

c.cos\_cant\_kg \* c.cos\_pre\_kg AS "Total ingreso"

FROM "vi\_piscina\_corrida" pc,

dim\_tallas t,

th\_cosecha c

WHERE c.tal\_cod = t.tall\_cod AND c.pis\_cod = pc.pis\_cod AND c.cor\_cod = pc.corrida

ORDER BY pc.empresa, pc.piscina, pc.corrida;

SCRIPT SQL VISTA vi\_hectarea\_piscina

SELECT pi.empresa,

pi.camaronera,

pi.pis\_cod,

pi.hectareas,

pi.piscina,

cr.cor\_cod AS corrida,

cr.cor\_est AS estado,

pi.pais,

pi.provincia,

pi.canton,

cr.anio AS "Año",

mes.semestre,

mes.trimestre,

mes.mes,

cp.cp\_num\_larv AS "Num larvas",

((cp.cp\_num\_larv)::double precision / pi.hectareas) AS "Larvas por Hectarea",

cp.cp\_cos\_inv AS "Costo de inversion",

i.ingresos AS "Total ingresos",

round(((i.ingresos - cp.cp\_cos\_inv))::numeric, 3) AS "Utilidad monto",

round(((((i.ingresos - cp.cp\_cos\_inv) / cp.cp\_cos\_inv) \* (100)::double precision))::numeric, 3) AS "Utilidad %"

FROM dim\_mes mes,

"vi\_Piscina" pi,

dim\_corrida cr,

th\_corrida\_piscina cp,

( SELECT c.pis\_cod,

c.cor\_cod,

sum((c.cos\_cant\_kg \* c.cos\_pre\_kg)) AS ingresos

FROM th\_cosecha c

GROUP BY c.emp\_cod, c.pis\_cod, c.cor\_cod

ORDER BY c.emp\_cod, c.pis\_cod, c.cor\_cod) i

WHERE ((mes.mes\_codigo = cr.mes\_codigo) AND (pi.pis\_cod = cp.pis\_cod) AND (cr.cor\_cod = cp.cor\_cod) AND (cp.pis\_cod = i.pis\_cod) AND (cp.cor\_cod = i.cor\_cod))

ORDER BY pi.empresa, pi.piscina, cr.cor\_cod;

SCRIPT SQL VISTA vi\_pregunta5

SELECT DISTINCT pc.empresa,

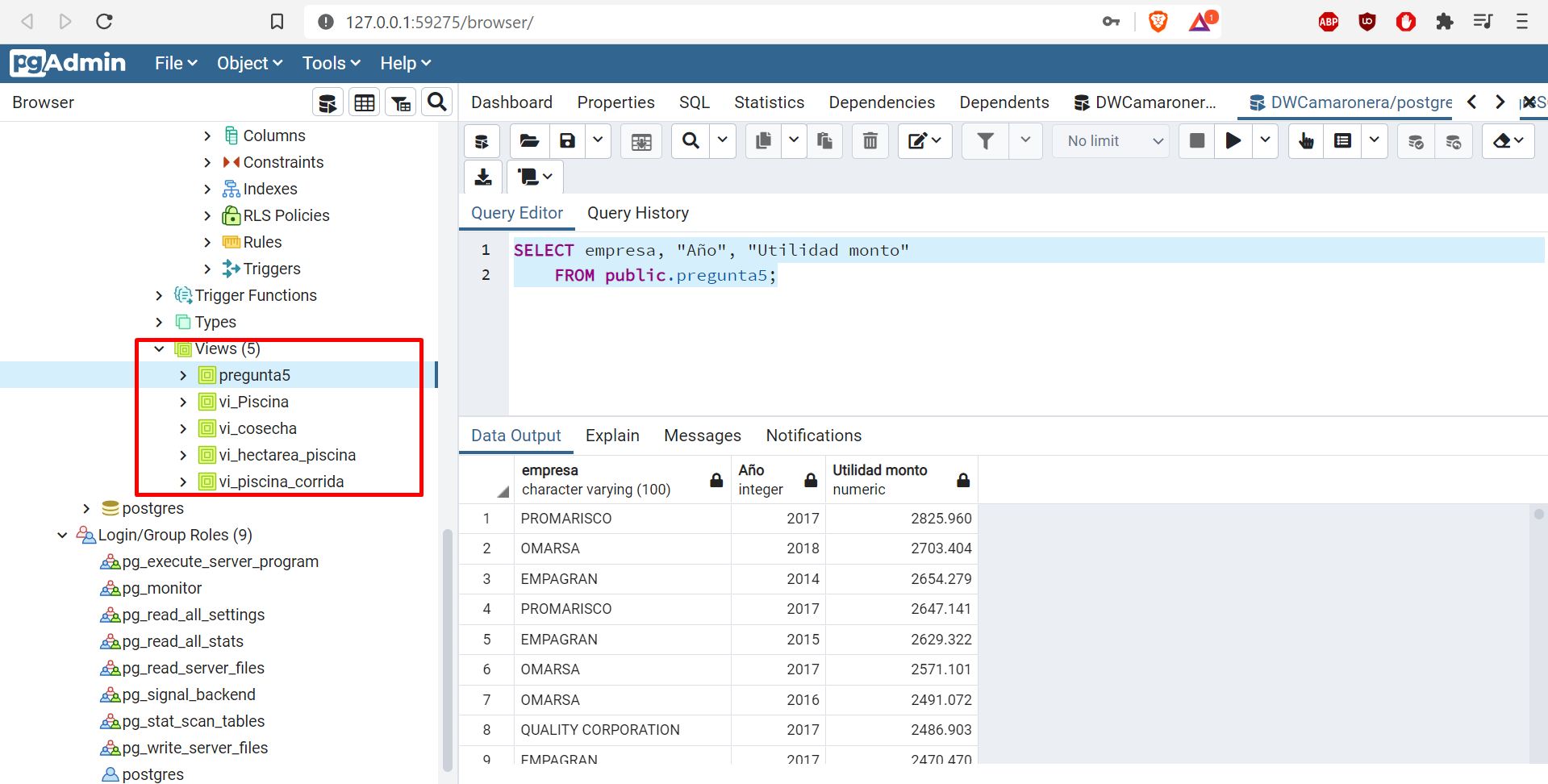
pc."Año",

pc."Utilidad monto"

FROM vi\_piscina\_corrida pc

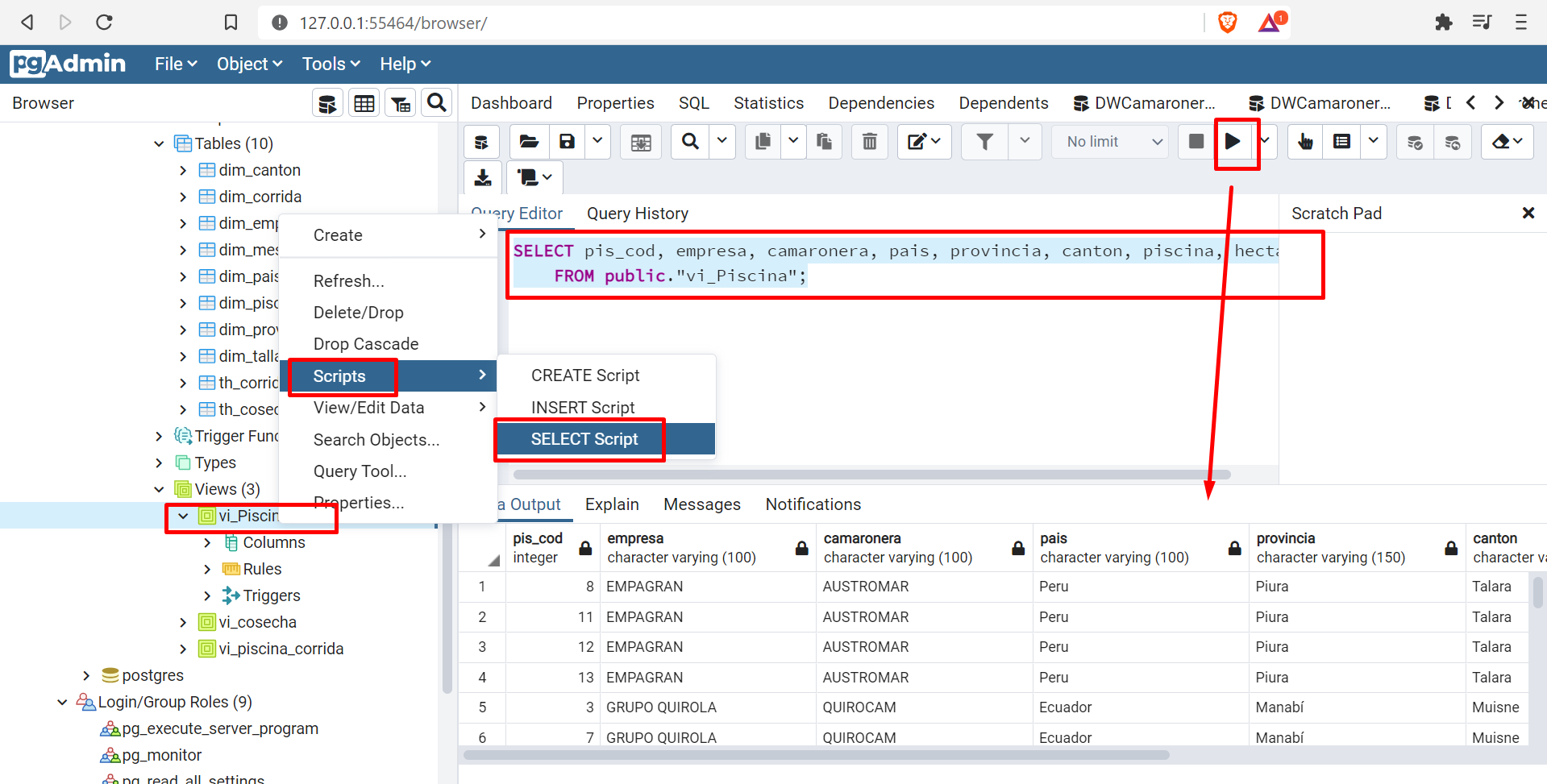
ORDER BY pc."Utilidad monto" DESC;

El resultado final tendremos 5 vistas creadas en nuestra base de datos.



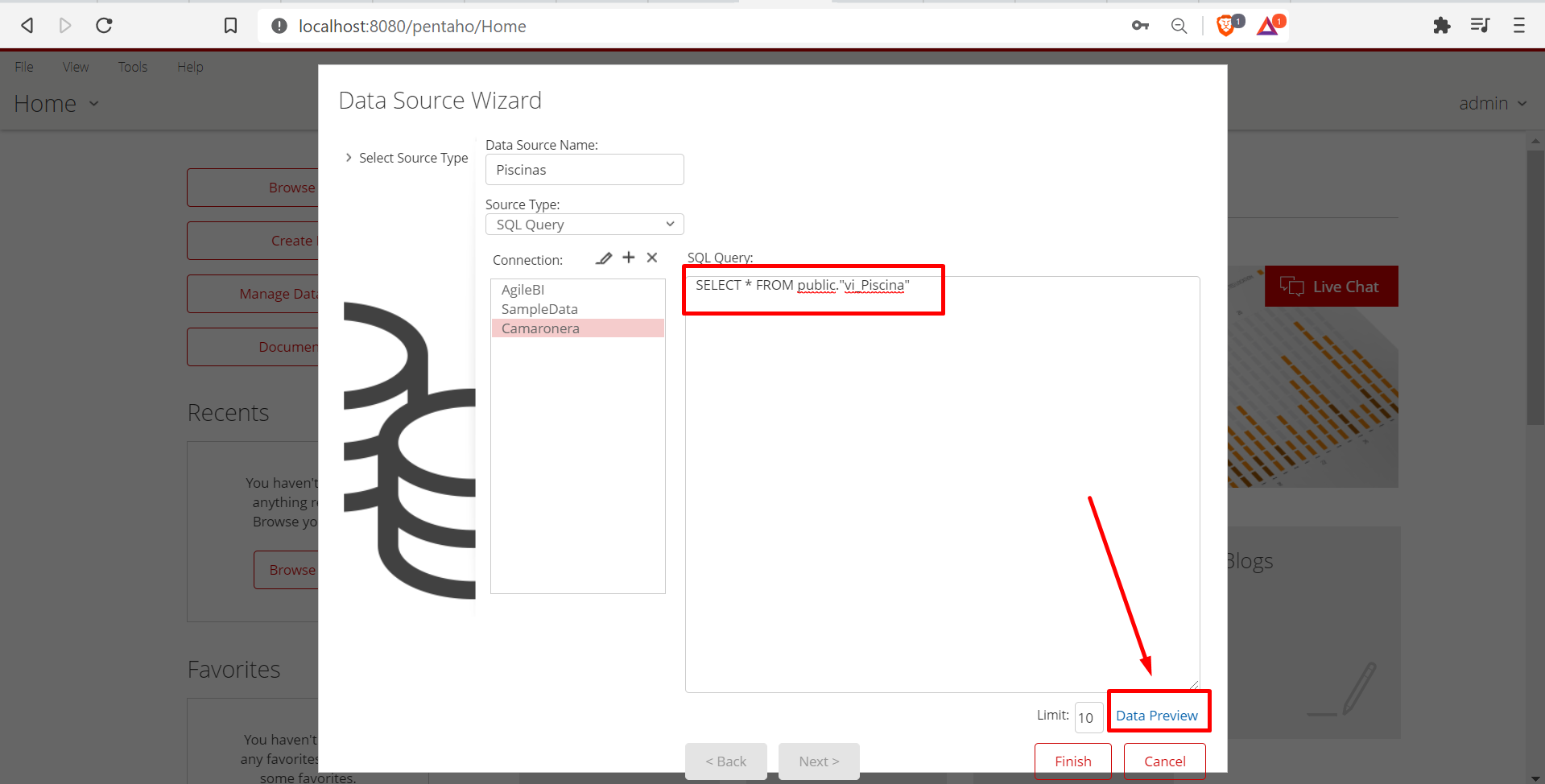
**Creación del primer cubo OLAP (Piscinas)**

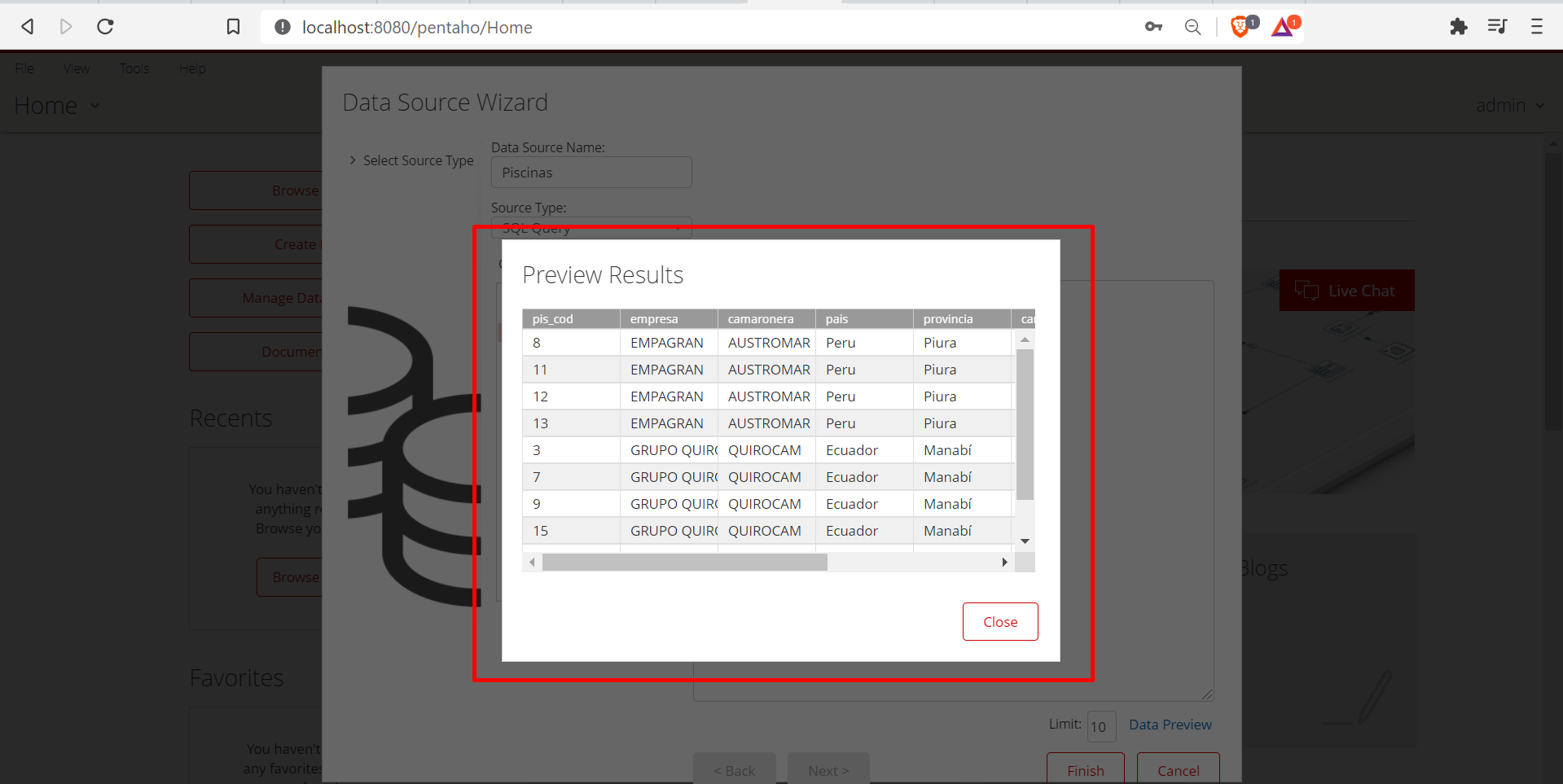
Ahora hay que crear un código sql que permitar ejecutar las vistas, para esto tienen que dirigirse a su base de datos dar clic derecho en la vista de piscinas, seleccionar la opción Scripts, luego la opción de Select Script y dar clic en el icono de play (run script) para comprobar que la vista se ejecuta correctamente, el código generado hay que copiarlo y pegarlo en Pentaho Server.



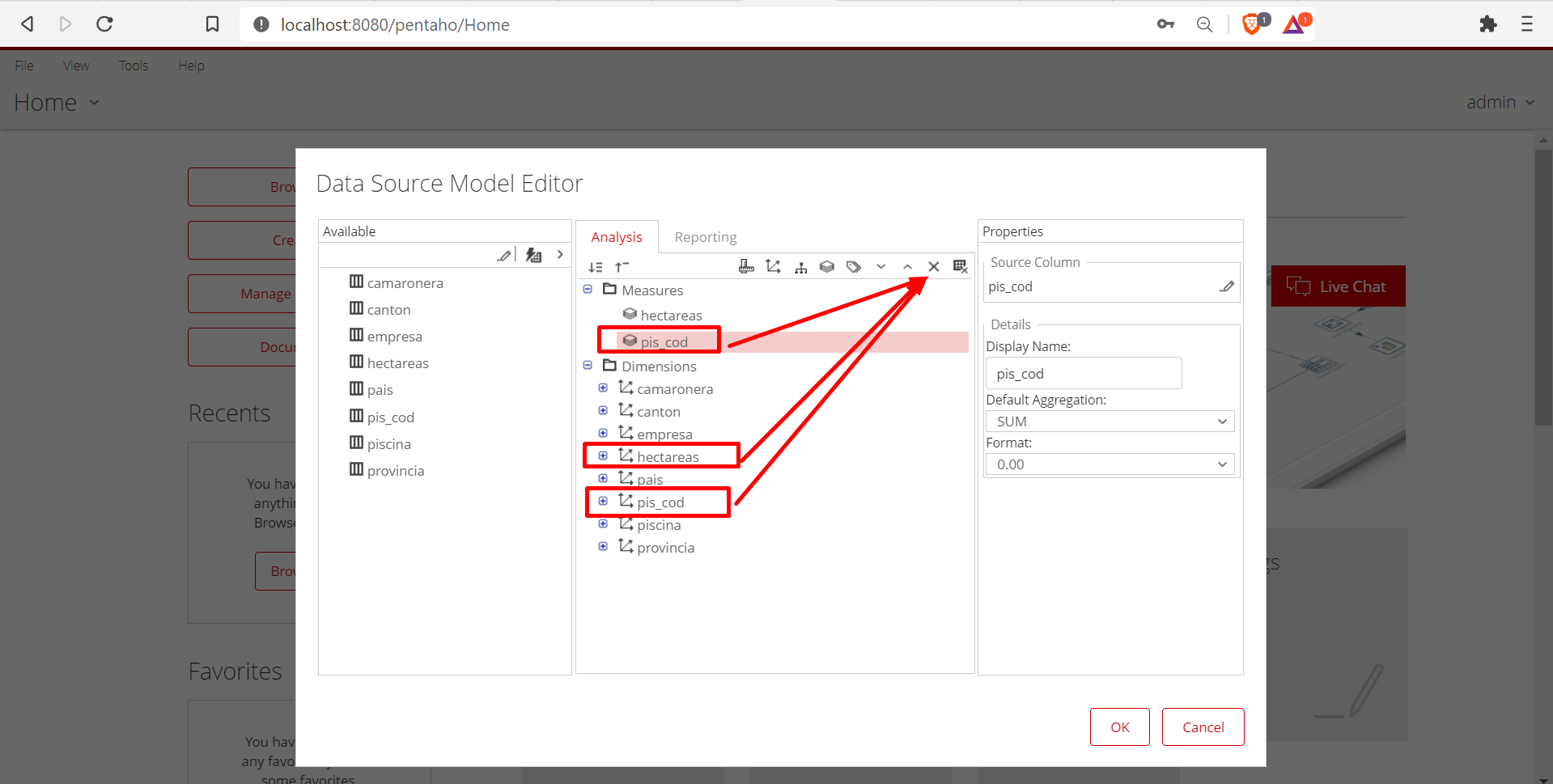
Para simplificar un poco el query, se obtendrá todos los campos quedando finalmente elquery de la siguiente manera: SELECT \* FROM public."vi\_Piscina"

Para comprobar que el query de consulta de la vista se ejecuta correctamente desde Pentaho Serve hay que dar clic en Data Preview y asi nos mostrará los datos de la vista.

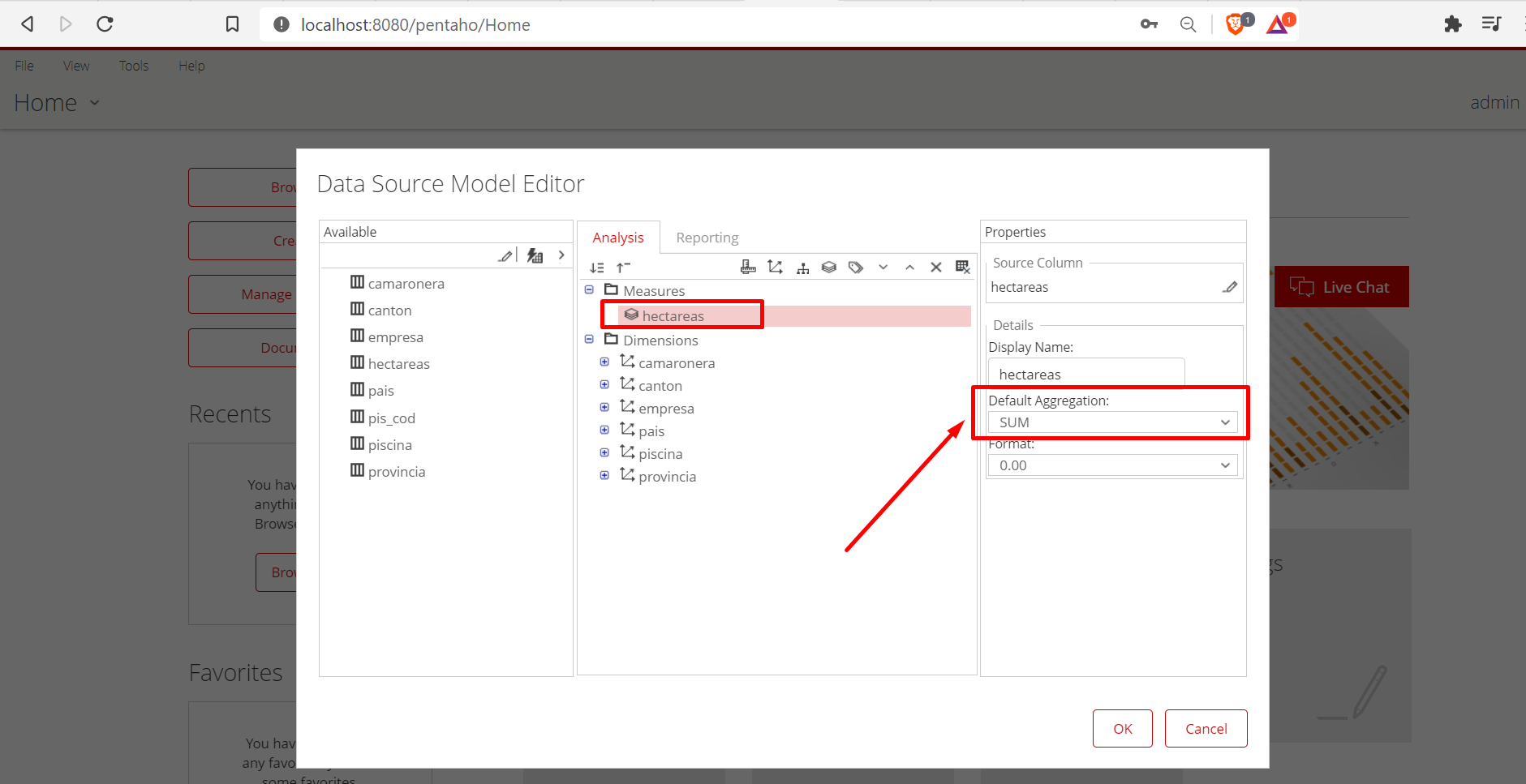


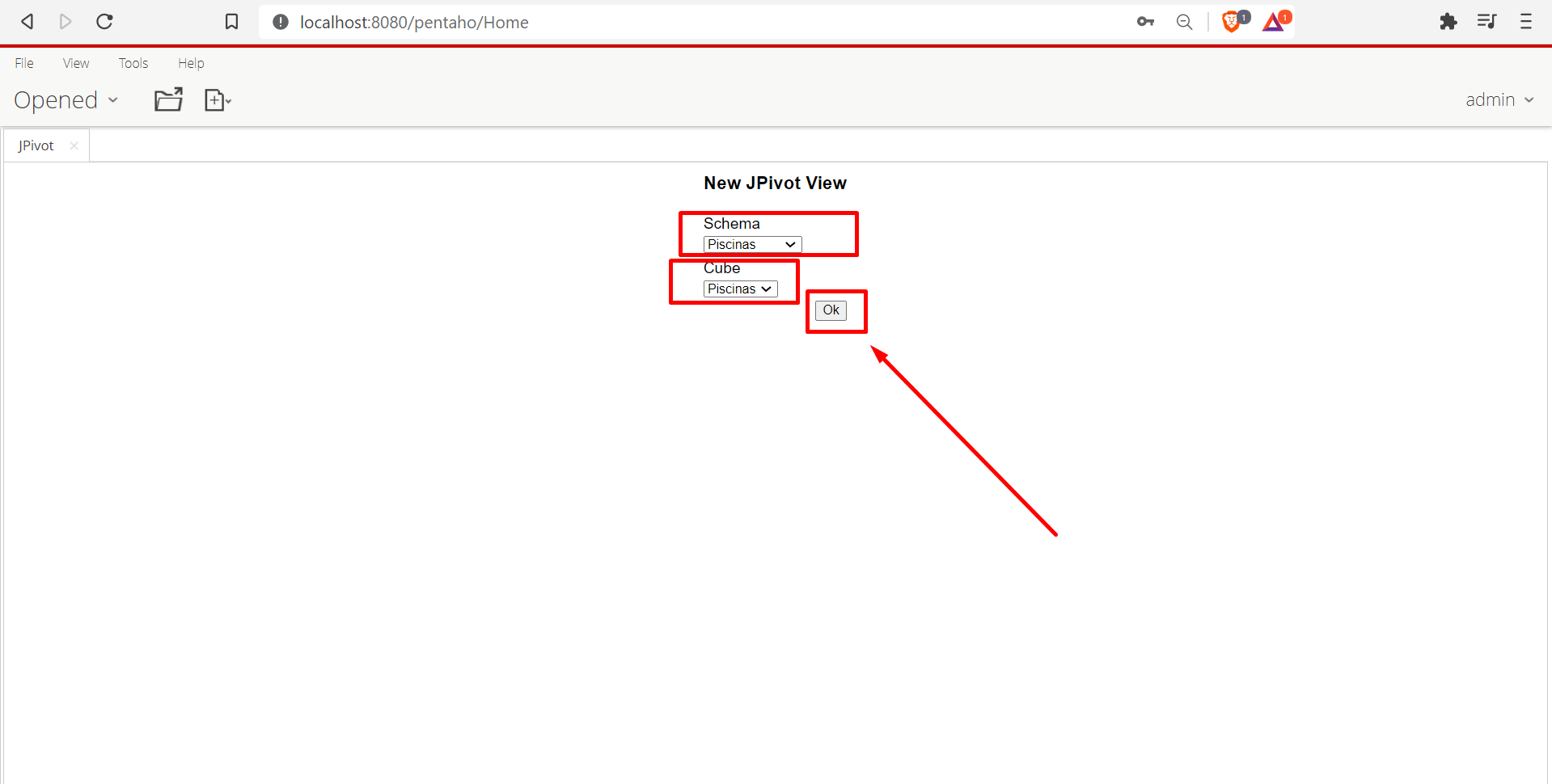


Una vez comprobada la conexión, procedemos a dar click en finish.



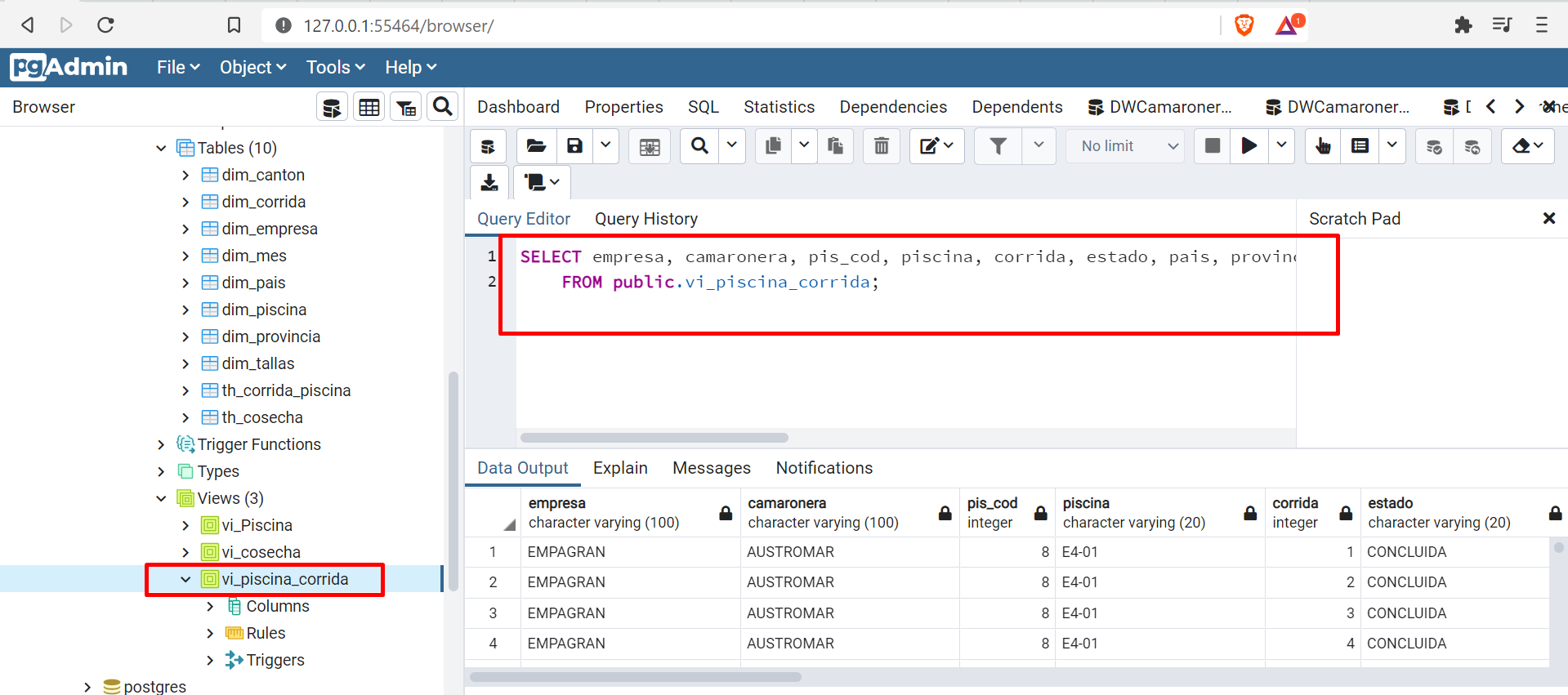
Antes de proceder a dar clic en ok hay que asegurarnos que la función de agregación de la medida de hectáreas este en SUM.





**Creación del segundo cubo OLAP (Piscina Corrida)**

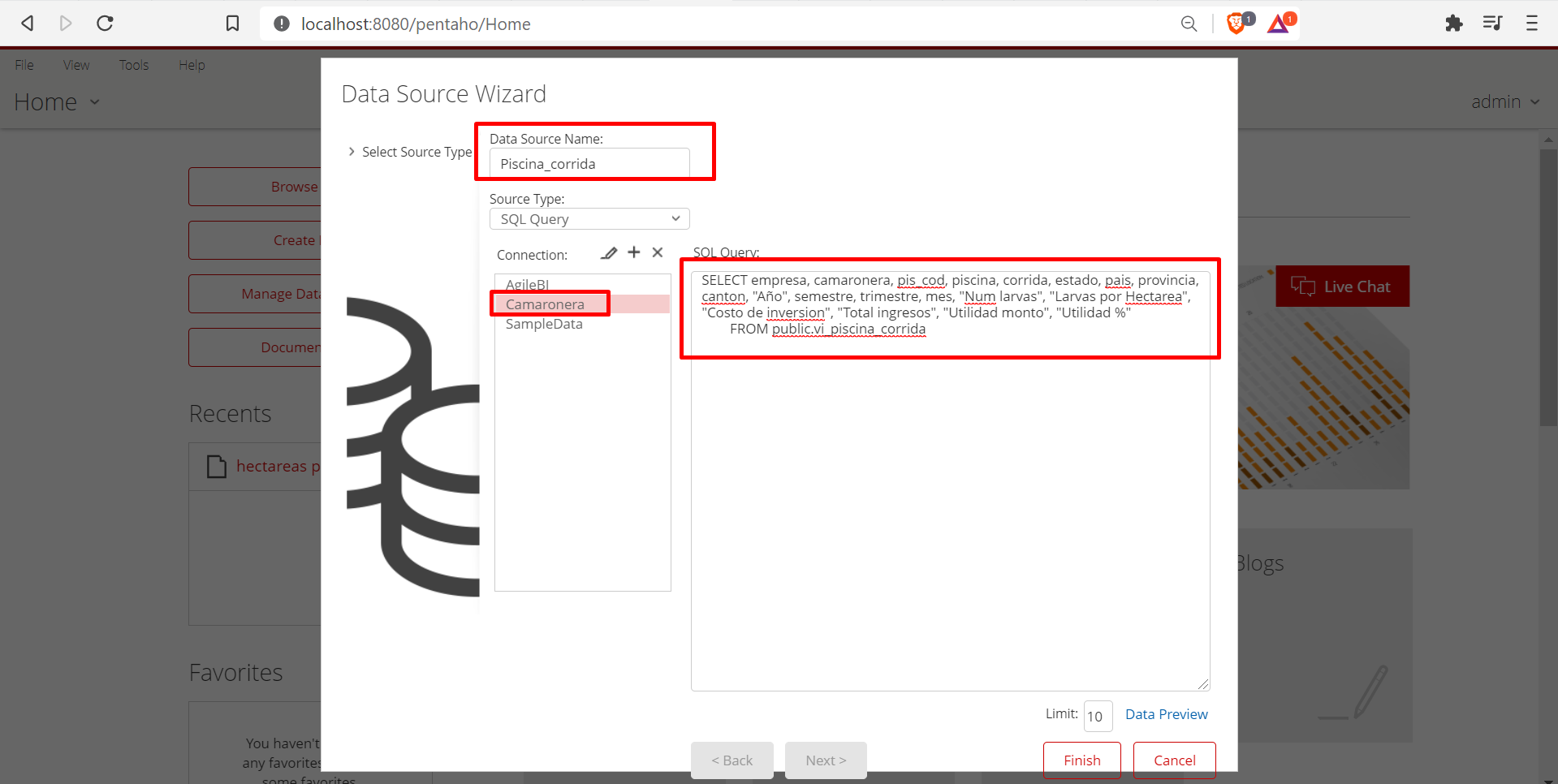
De igual forma que en el anterior cubo hay que obtener el código sql de la vista.



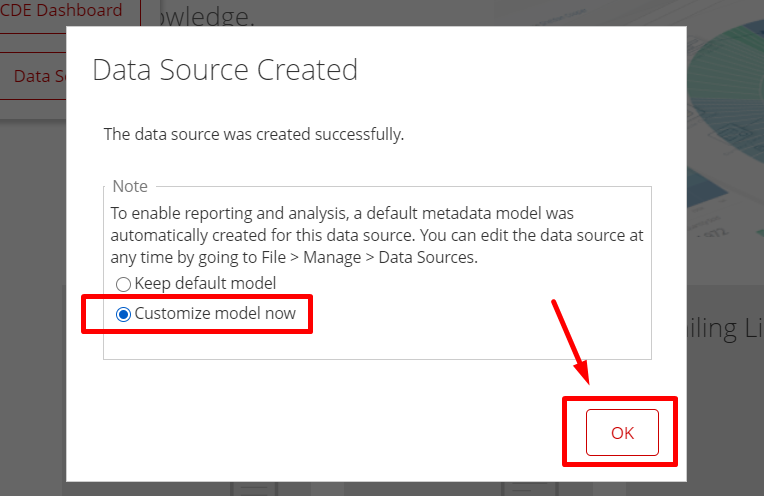
**Código sql:**

SELECT empresa, camaronera, pis\_cod, piscina, corrida, estado, pais, provincia, canton, "Año", semestre, trimestre, mes, "Num larvas", "Larvas por Hectarea", "Costo de inversion", "Total ingresos", "Utilidad monto", "Utilidad %" FROM public.vi\_piscina\_corrida

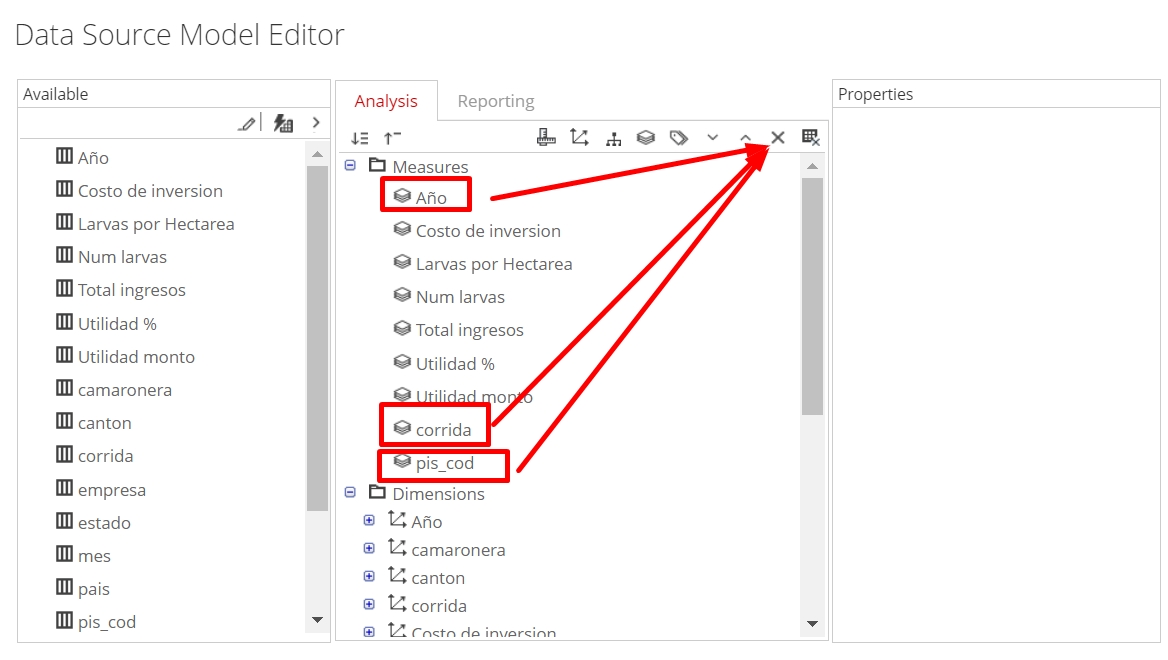
Luego creamos un nuevo data source de nombre Piscina\_corrida, seleccionamos la conexión de Camaronera y pegamos la consulta sql de la vista.

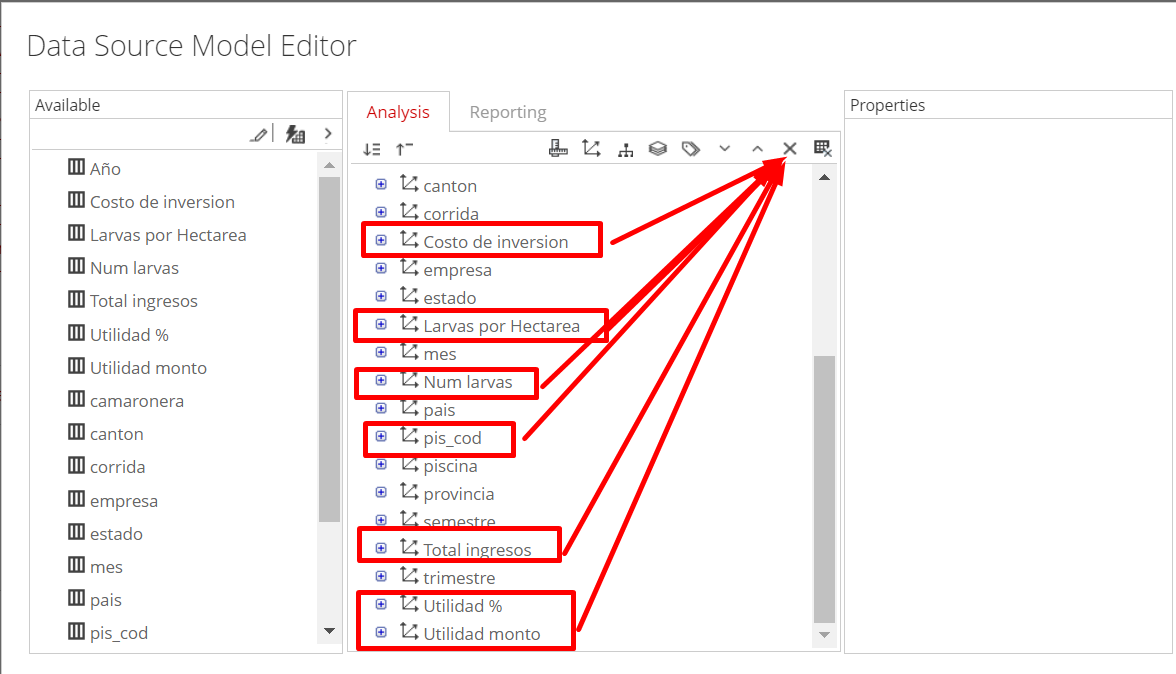


Seleccionar la opción de customize model now y dar clic en ok.

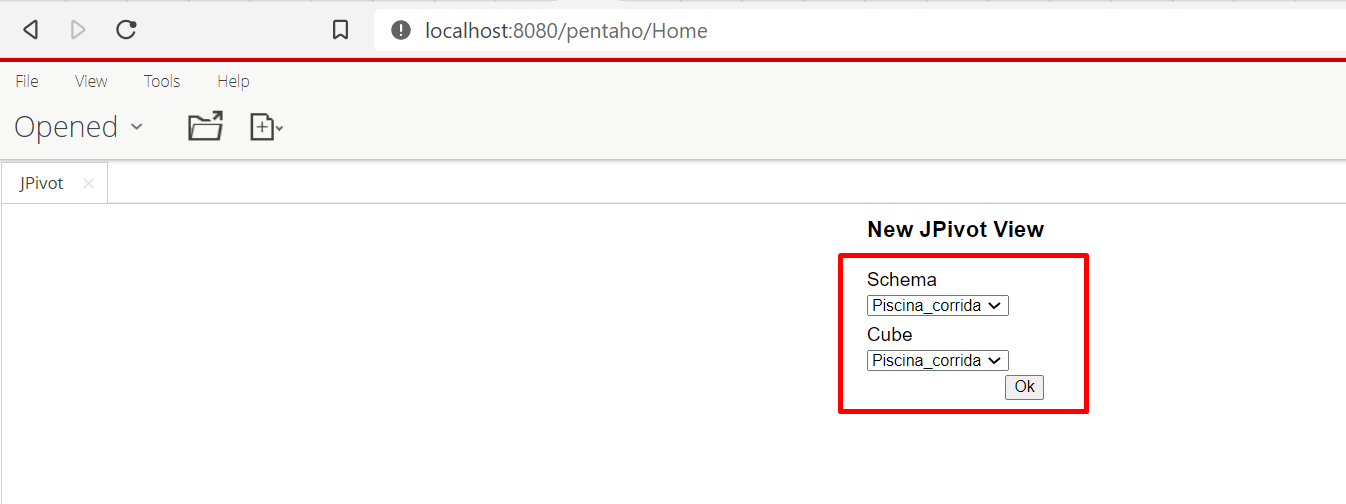


De igual forma se procede a eliminar a aquellas medidas y dimensiones que no se utilizará.



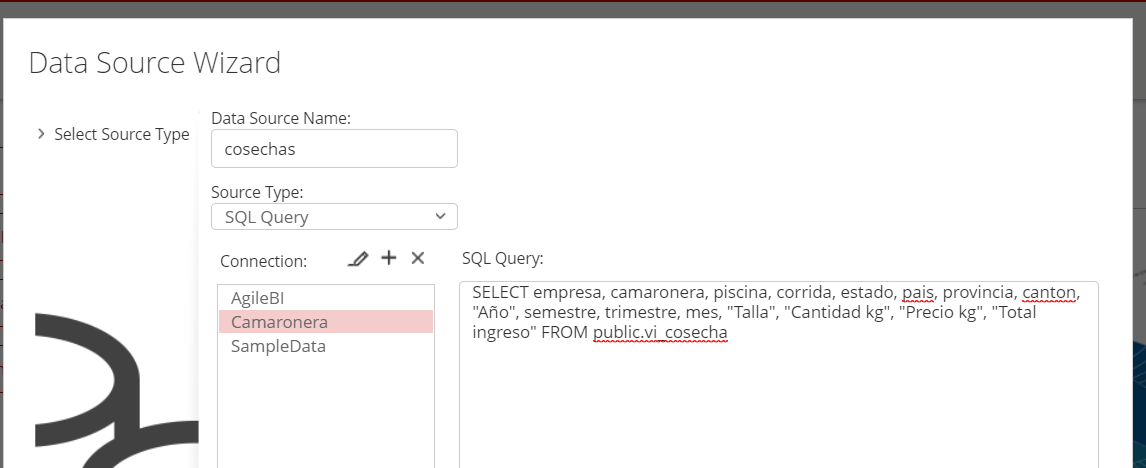


Creamos nuevamente la vista del cubo y seleccionamos Piscina\_corrida

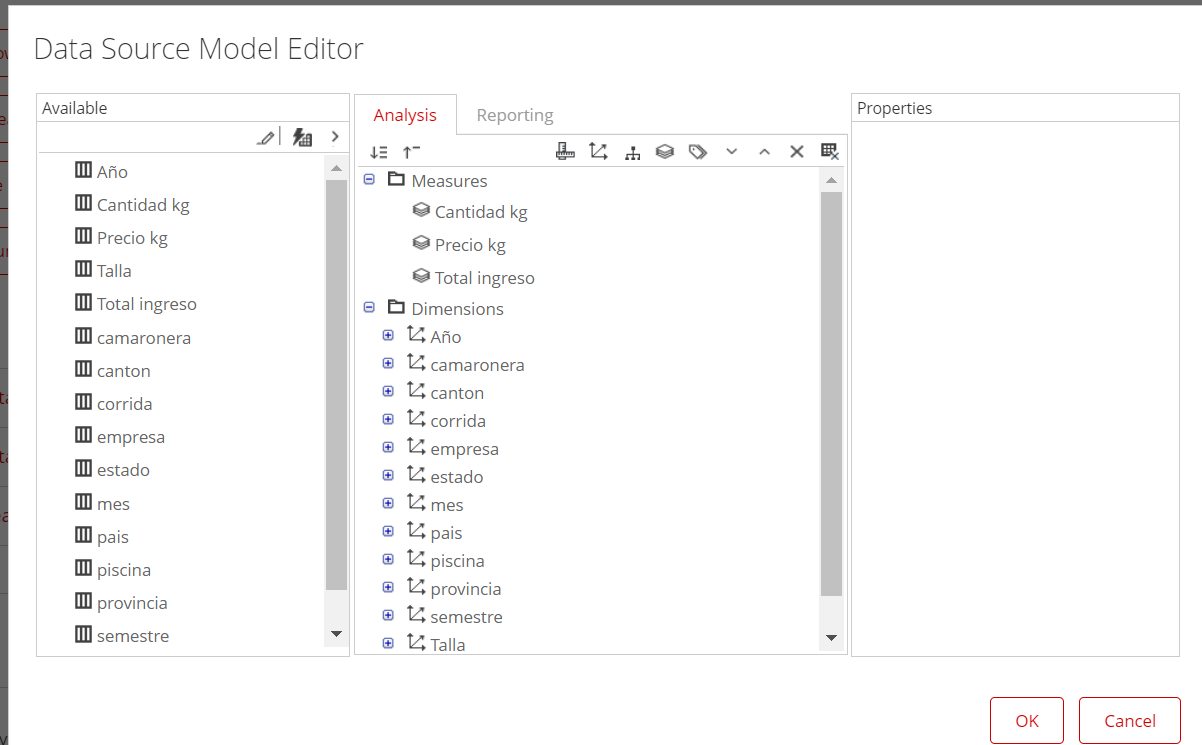


**Creación del tercer cubo OLAP (Cosechas)**

Creamos la conexión con la vista de cosechas.



Customizamos la vista sql.



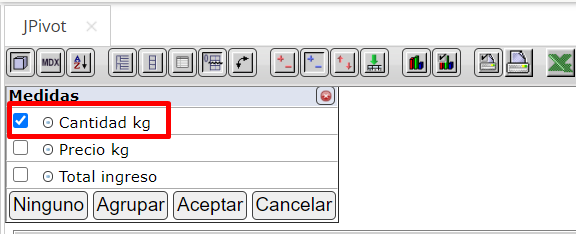
Creamos la vista con JPivot.



**Contestación de preguntas de Negocio**

**1.- ¿Cuántos Kilogramos de camarón se han producido por talla y empresa en el último año?**

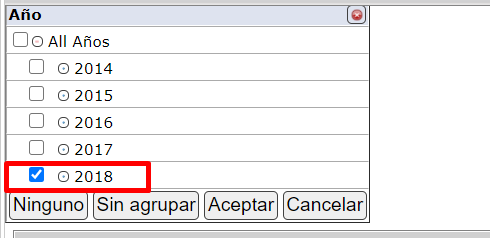
Medidas:



Dimensiones:



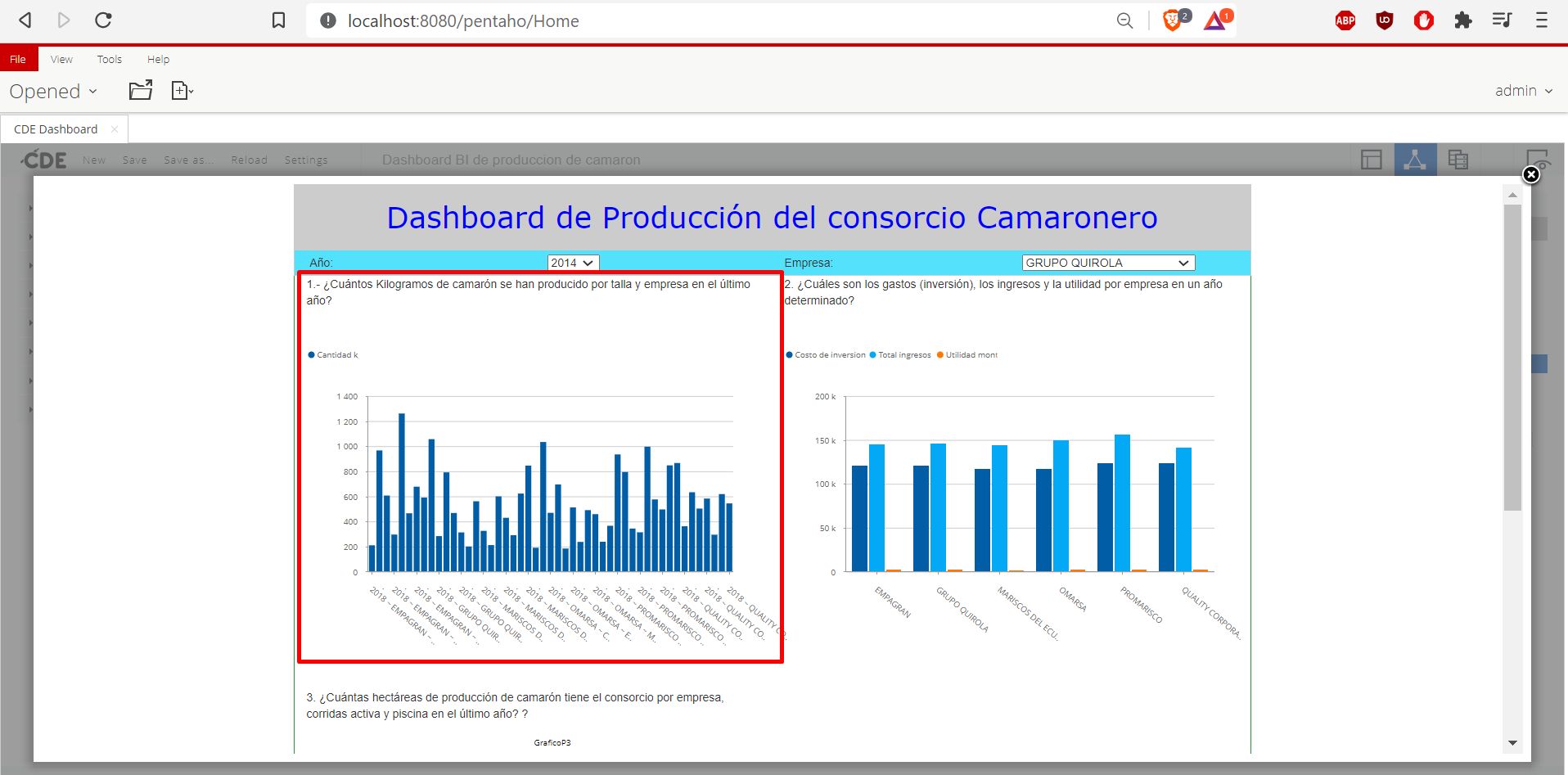
Filtro: El último año.



**Resultado:**

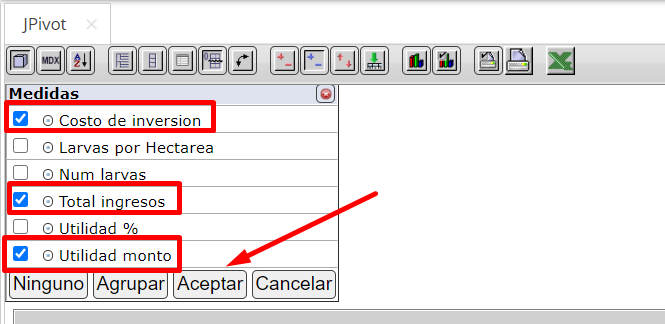


**DashBoard BI en Pentaho**



**2.- ¿Cuáles son los gastos (inversión), los ingresos y la utilidad por empresa en un año determinado?**

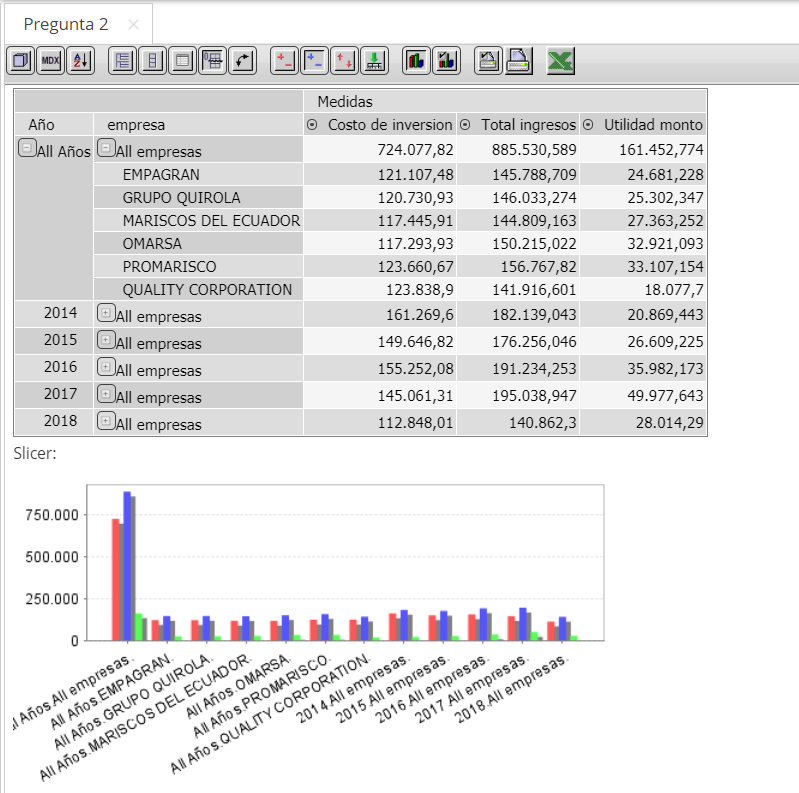
Las medidas que se van a presentar son costos de inversión, total de ingresos y utilidad monto.



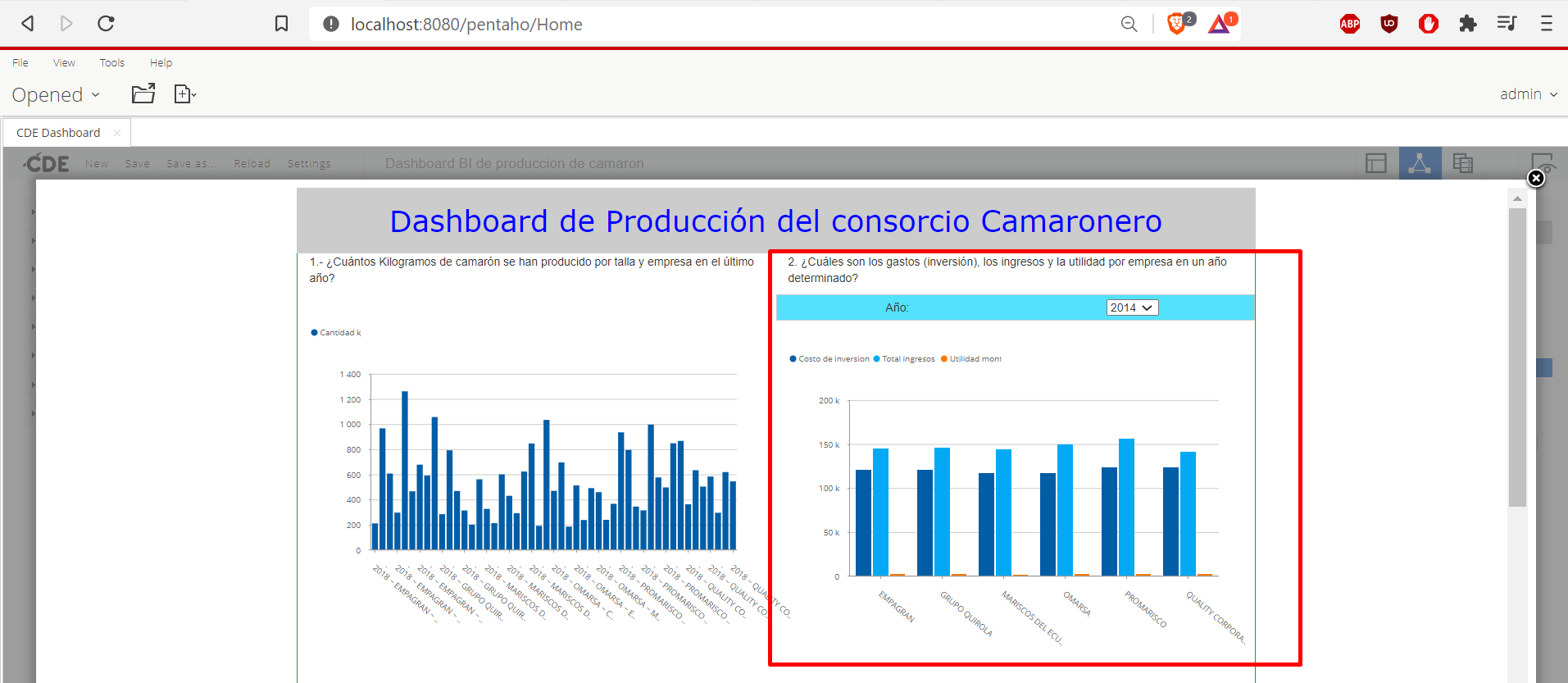
Y las dimensiones seleccionamos año y empresa.



Quedando como resultado lo siguiente:

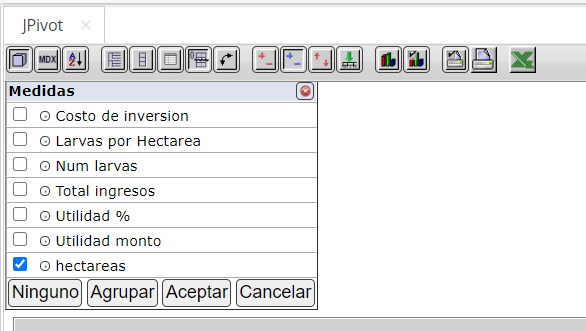


**DashBoard BI en Pentaho**



**3.- ¿Cuántas hectáreas de producción de camarón tiene el consorcio por empresa, corridas activa y piscina en el último año?**

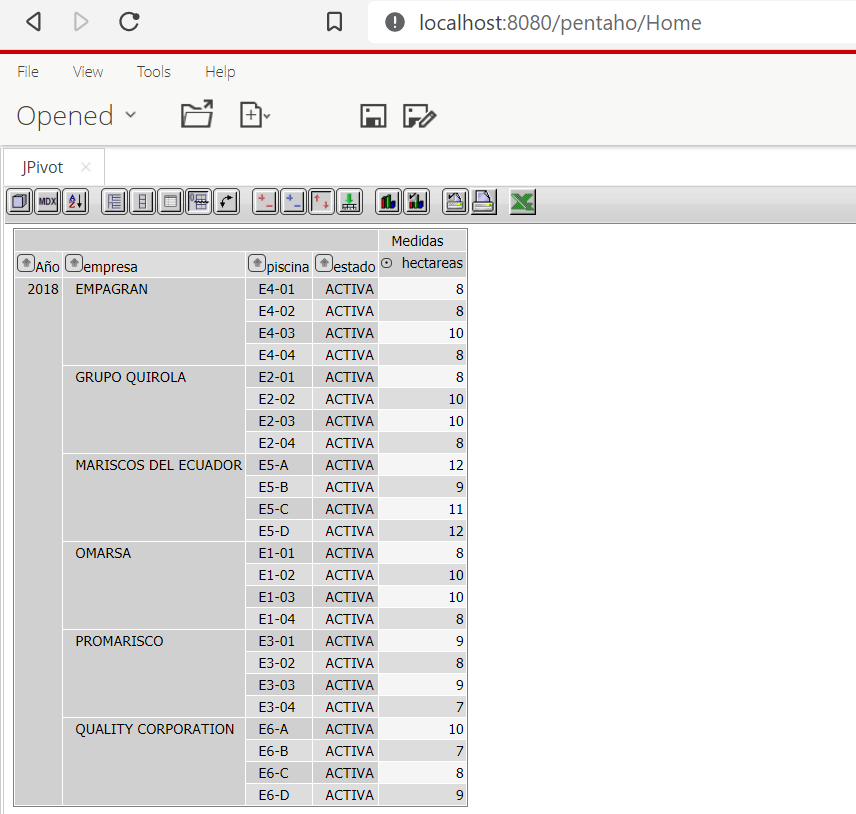
**Medidas:**



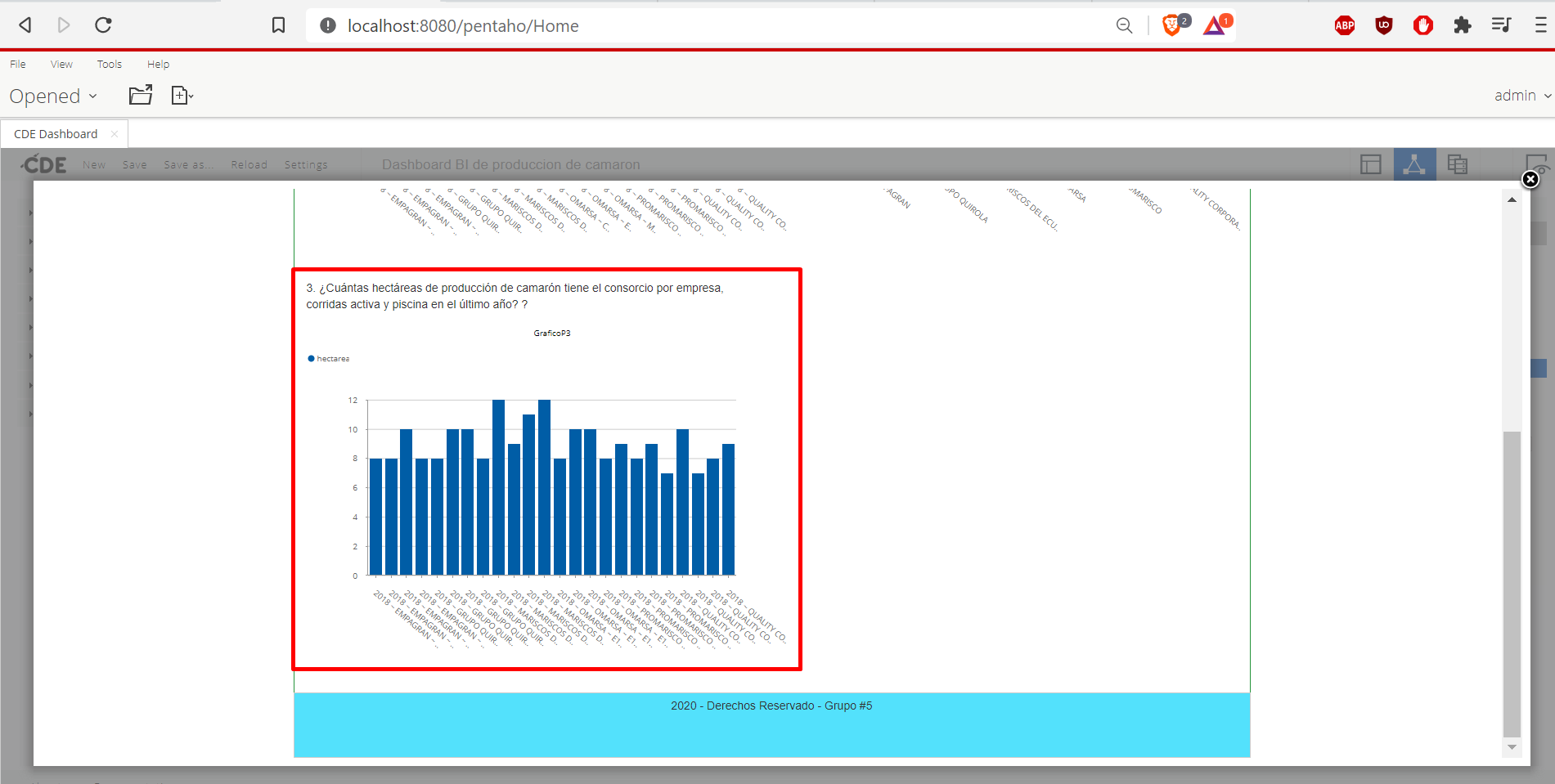
**Dimensiones:**



**Resultado:**



**Dashboard BI en Pentaho**



**4.- ¿Cuántas hectáreas de producción de camarón dispone el consorcio por empresa, piscina y país?**

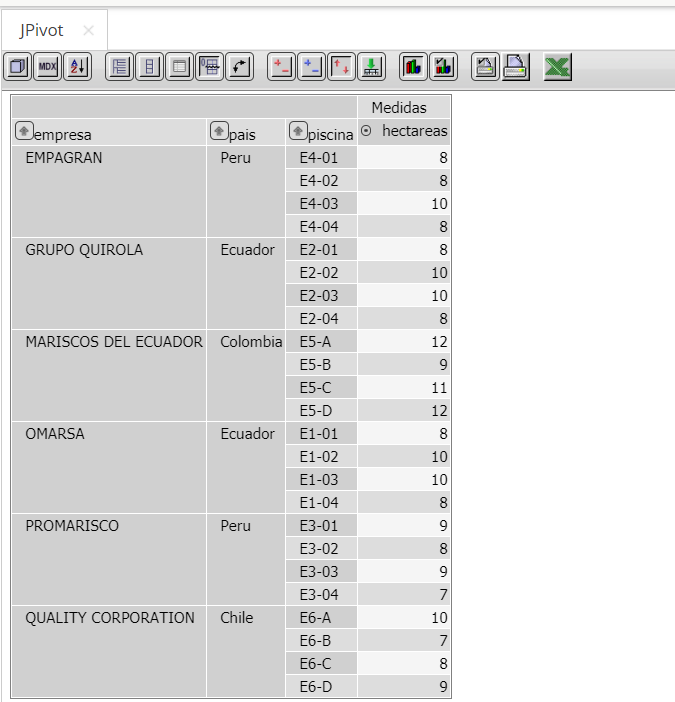
**Medidas:**



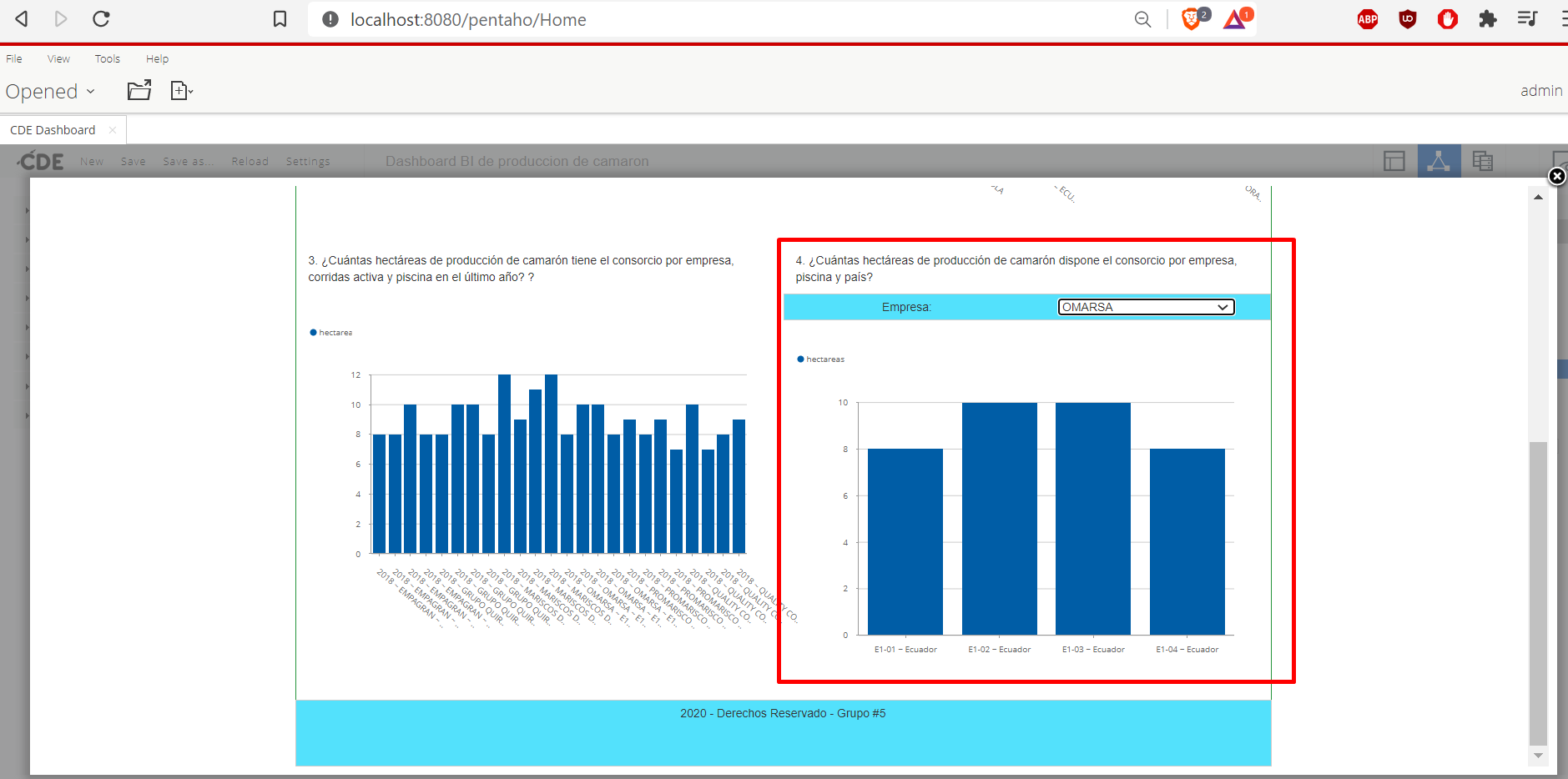
**Dimensiones:**



**Resultado:**

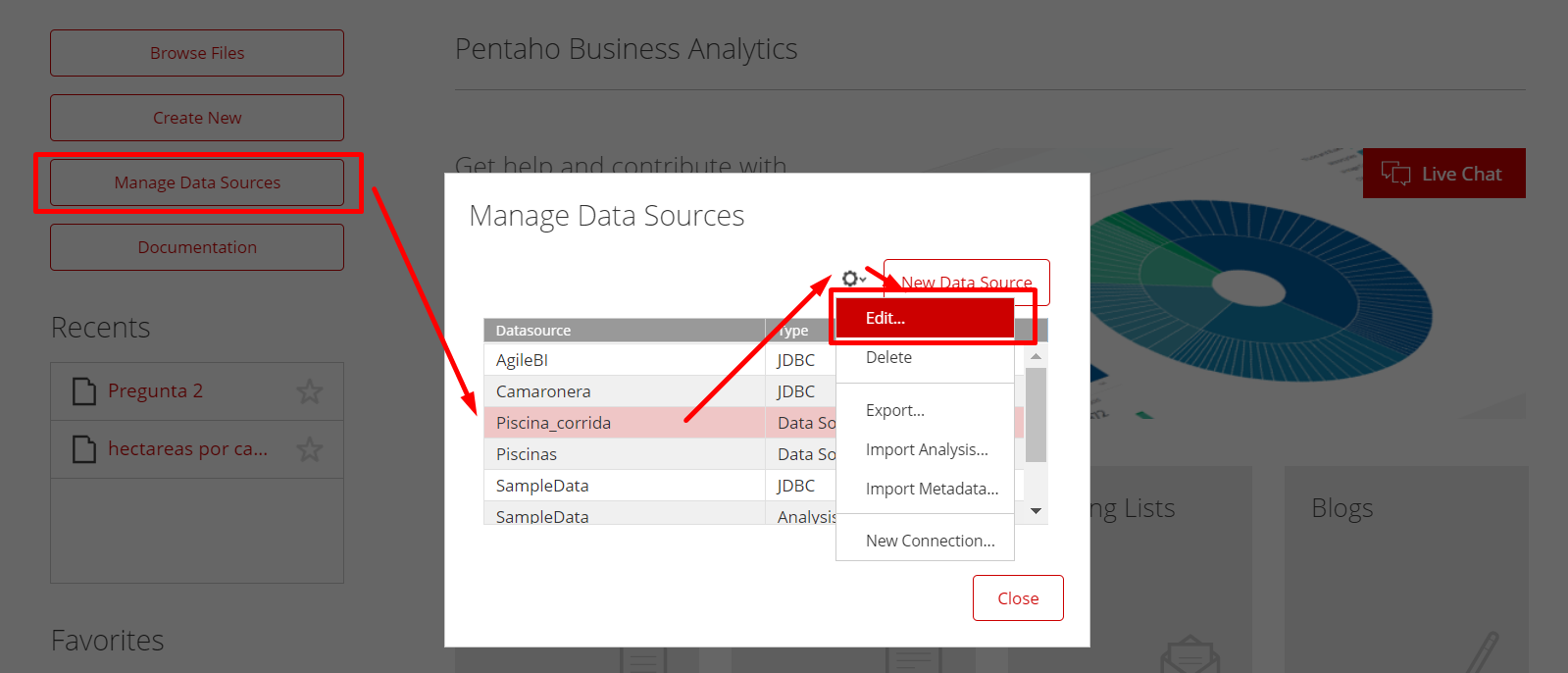


**Dashboard BI en Pentaho**

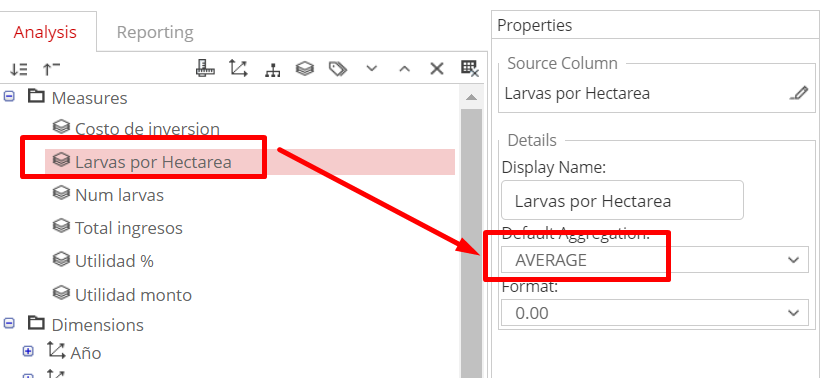


**5.- ¿Cuál es el número promedio de larvas de camarón por hectárea que se siembran por empresa y camaronera?**

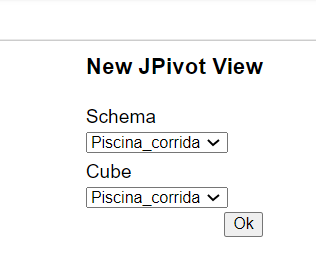
Como es promedio, hay que cambiar la función de agregación de SUM a AVG (promedio) para esto hay que editar el data source, dar clic en el menú manage data sources y seleccionar Piscina\_corrida, luego dar click en configuraciones y seleccionar la opción editar.



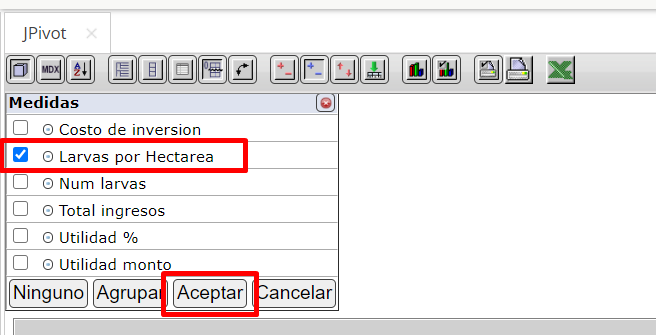
Y modificamos la función de agregación de la medida de larvas por hectarea de SUM por Average.



Volvemos a crear un cubo OLAP



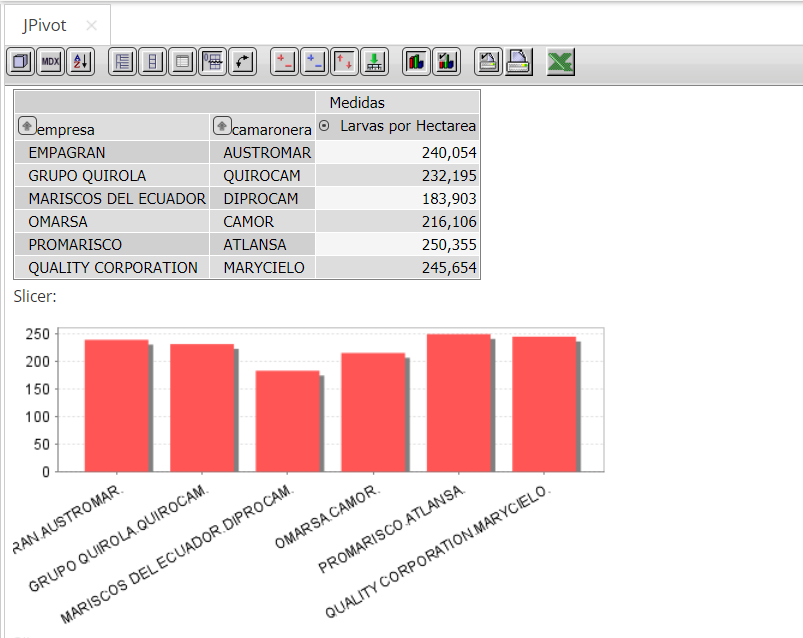
Modificamos la medida en este caso es larvas por hectáreas.



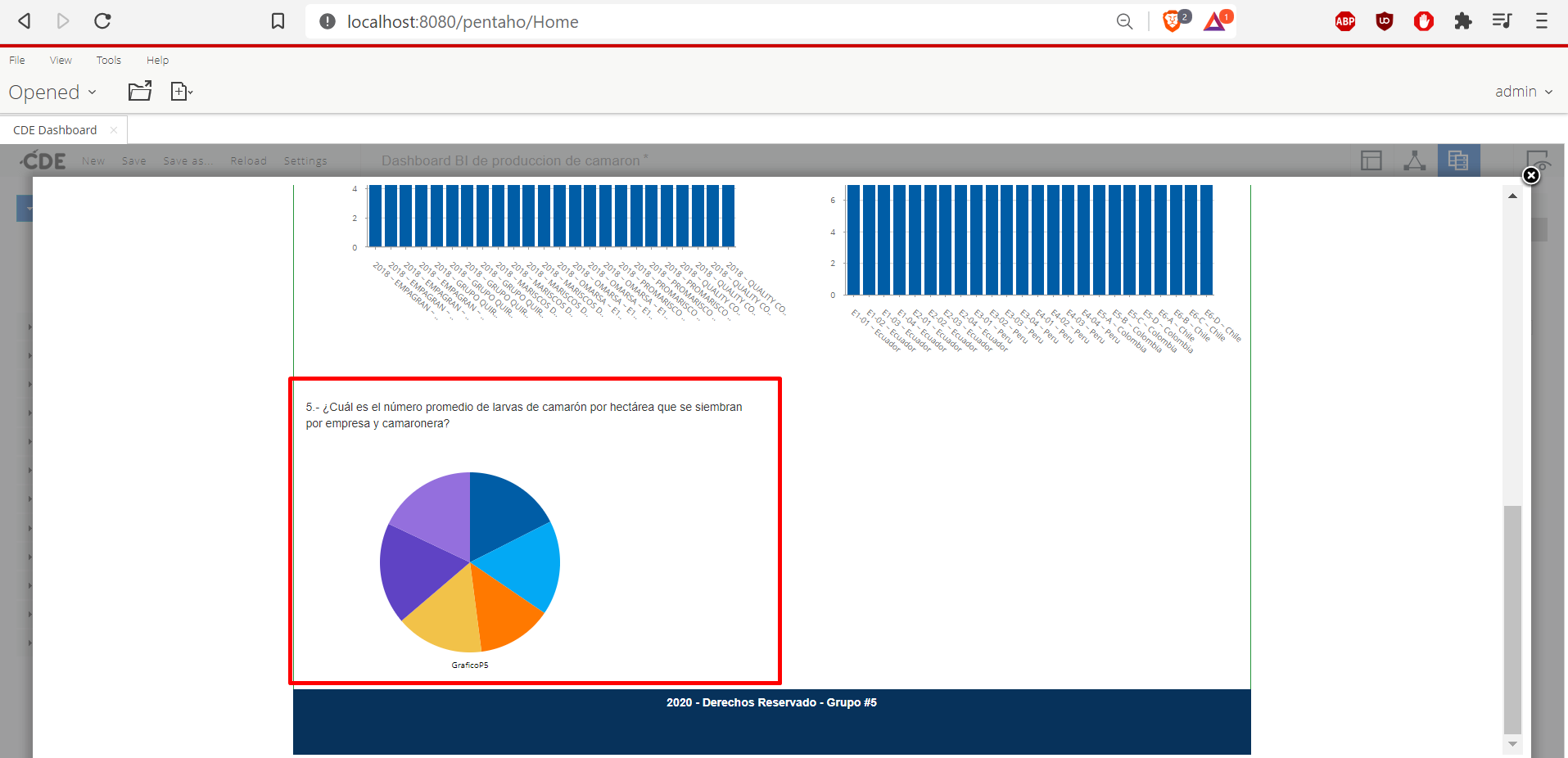
Y en las dimensiones seleccionamos empresa y camaronera.



Y el cubo nos quedará de la siguiente manera:



**Dashboard BI en Pentaho**



**6.- ¿Cuál es la empresa que ha generado más utilidades por año?**

SELECT empresa, "Año", utilidad FROM public.vi\_pregunta5;

SELECT DISTINCT pc.empresa,

pc."Año",

sum(pc."Utilidad monto") AS utilidad

FROM vi\_piscina\_corrida pc

GROUP BY pc."Año", pc.empresa

ORDER BY (sum(pc."Utilidad monto")) DESC;

SELECT pc.empresa,

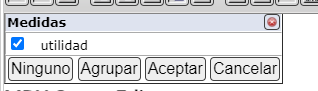
pc."Año",

pc."Utilidad monto"

FROM vi\_piscina\_corrida pc

WHERE pc."Utilidad monto" = (SELECT MAX("Utilidad monto") FROM vi\_piscina\_corrida pc where pc."Año" = 2014 )

**Medidas:**



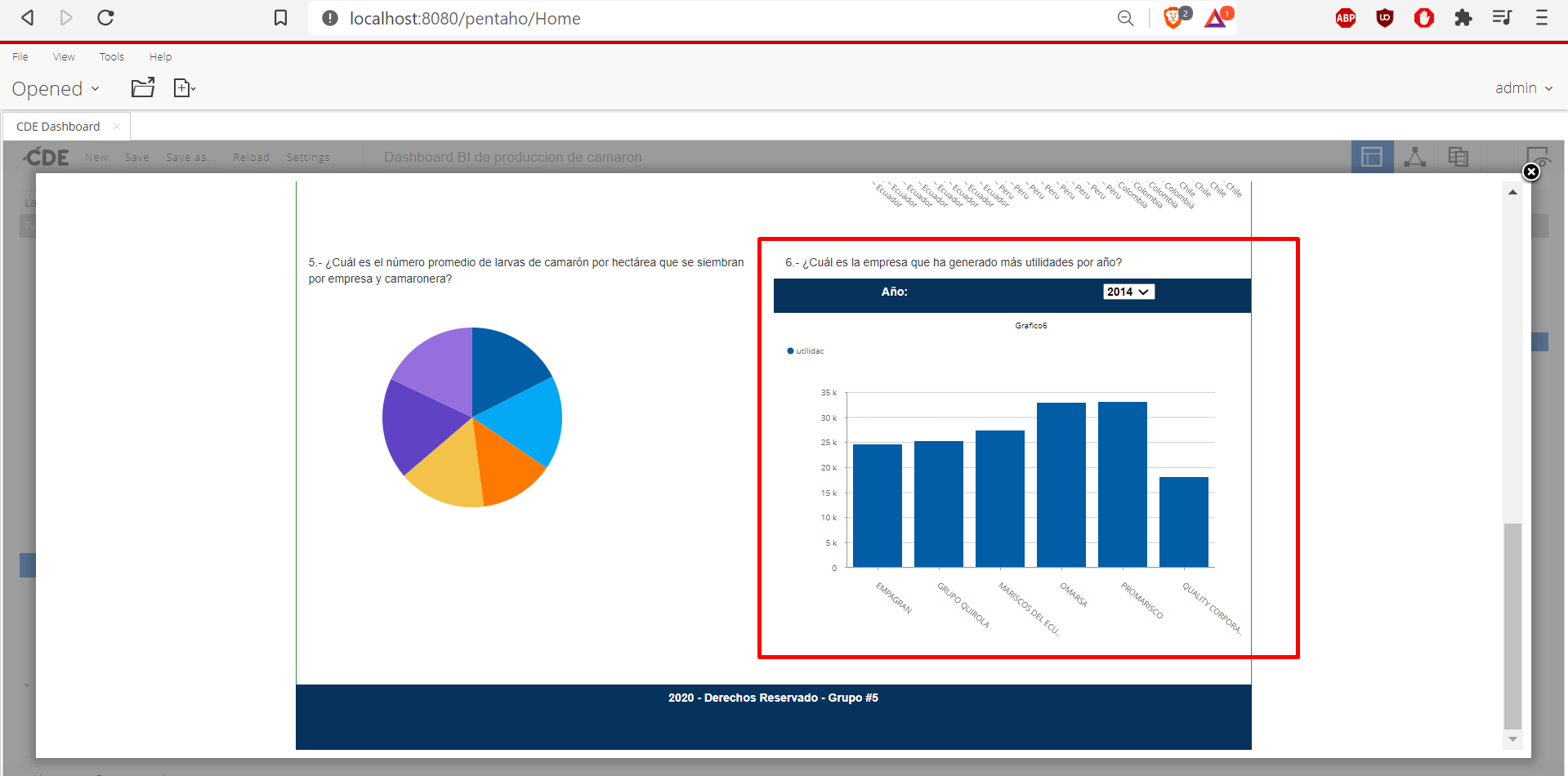
**Dimensiones:**



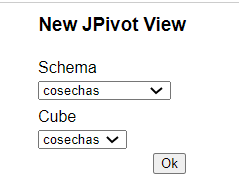
**Resultado:**



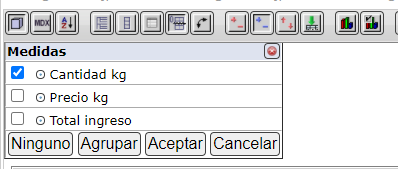
**Dashboard BI Pentaho**



**7.- ¿Cuál es la empresa que ha producido más camarón por talla y año?**



**Medidas:**



**Dimensiones:**



**Resultados:**



**Dashboard BI Pentaho**

