

**Universidad de San Carlos de Guatemala**  
**Facultad de Ingeniería**  
**Escuela de Ciencias y Sistemas**  
**Laboratorio de Seminario de Sistemas 2**

## **Documentación - Proyecto 1 - Fase 3**

### **Grupo #21**

<b>Nombre</b>	<b>Carné</b>
<b>Walter Josué Paredes Sol</b>	<b>201504326</b>
<b>Fernando Josué Flores Valdez</b>	<b>201504385</b>
<b>Gonzalo Antonio García Solares</b>	<b>201318652</b>
<b>Byron David Cermeño Juárez</b>	<b>201313734</b>

# Título del Proyecto

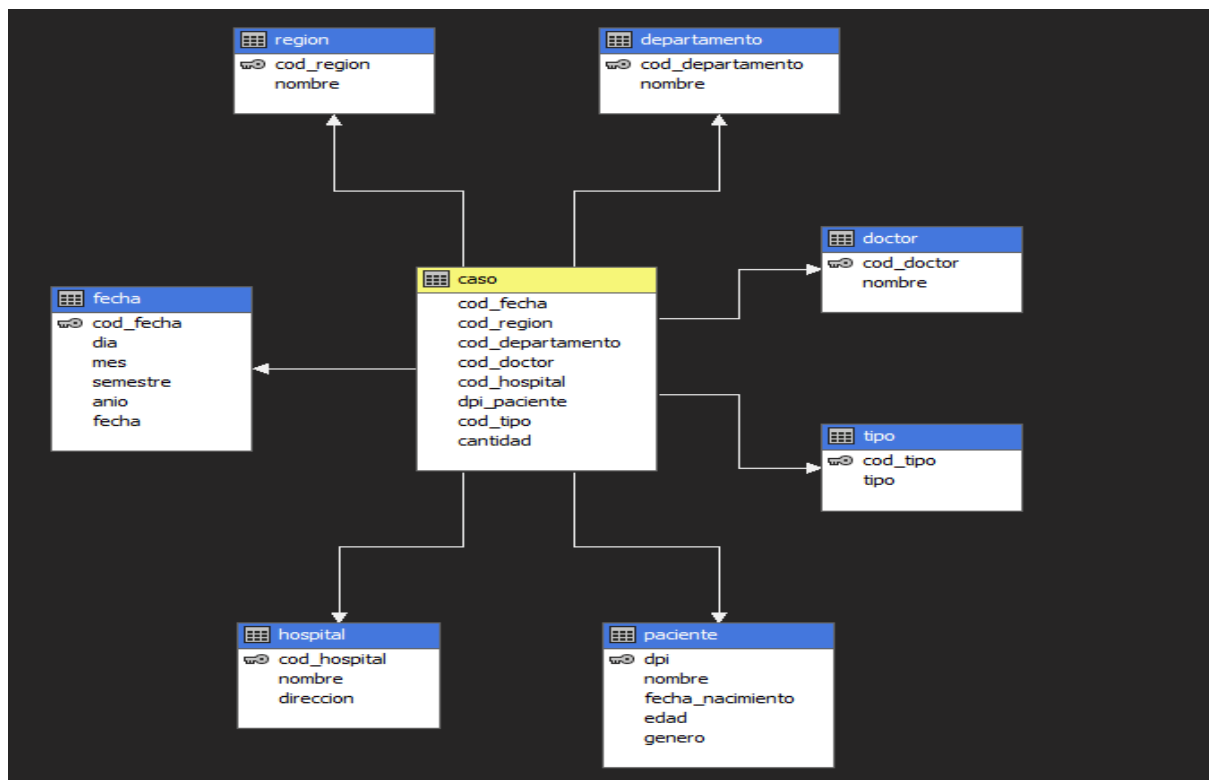
## Procesos de BI aplicados a COVID-19

Se utilizaron herramientas para el análisis de datos y convertirlos en información que genere valor, en este caso, la pandemia fue un punto y aparte en que tan valiosa es la tecnología para procesar dicha información y así mitigar la infección a nivel nacional.

## Características y funciones de la solución

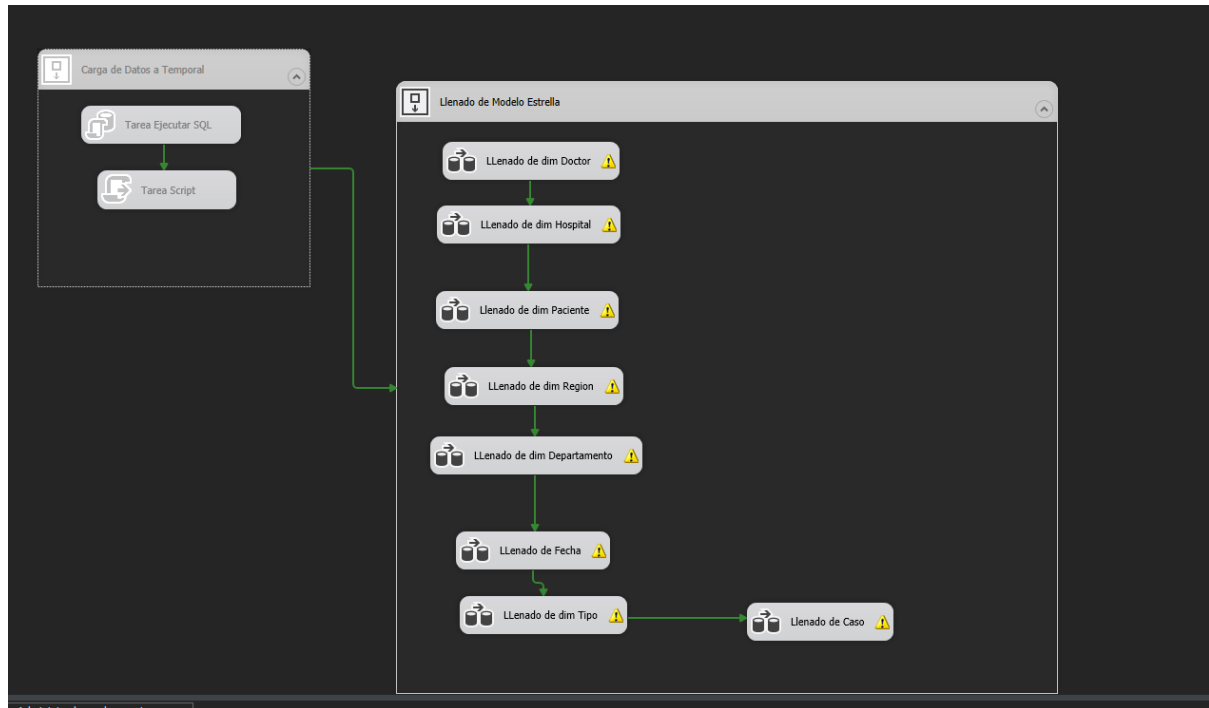
- Debe contar con la identificación de datos clave para la reportería.
- Tendrá que tener en cuenta el tipo de acceso que requieren los distintos usuarios actuales, y posibles usuarios futuros.
- Una correcta segmentación de los datos para una filtración efectiva.
- Se tiene que considerar los datos mínimos necesarios para los requerimientos actuales, y así mismo procurar no dejar cerrado a cambios el modelo.
- La segmentación debe permitir relacionar a nuevas tablas para futuros cambios.
- Como mínimo debe cumplir con tablas que permitan el filtrado básico de los hechos.

## Modelo estrella (ER)



# CARGA DE DATOS CON ETL

Se muestra el flujo general para la carga



## FLUJO DETALLADO:

Script para la carga de la tabla temporal

```
using System;
using System.IO;
using System.Data;
using System.Data.SqlClient;
using System.Windows.Forms;

//Se utiliza para validar que no existan errores que interrumpan la ejecución
//al intentar conectarnos a la ODBrigen
try
{
    //Creamos la variable de tipo SQL que nos permita una nueva conexión
    string server = "localhost";
    string database = "test";
    string uid = "root";
    string password = "admin";
    string port = "3306";
    string connectionString = "Server=" + server + ";Uid=" + uid + ";Pwd=" + password + ";Database=" + database + ";";
    MySqlConnection mySqlConnection = new MySqlConnection(connectionString);
    //MySqlConnection mySqlConnection = Dts.Connections["ss2projectog21.mysql.database.azure.com.httdtemp"].AcquireConnection(Dts.Transaction) as MySqlConnection;

    // Indicamos el servidor y la base de datos a la que nos conectaremos

    mySqlConnection.Open();

    //Variable para verificar que el linea tenga contenido
    int counter = 0;
    string linea;

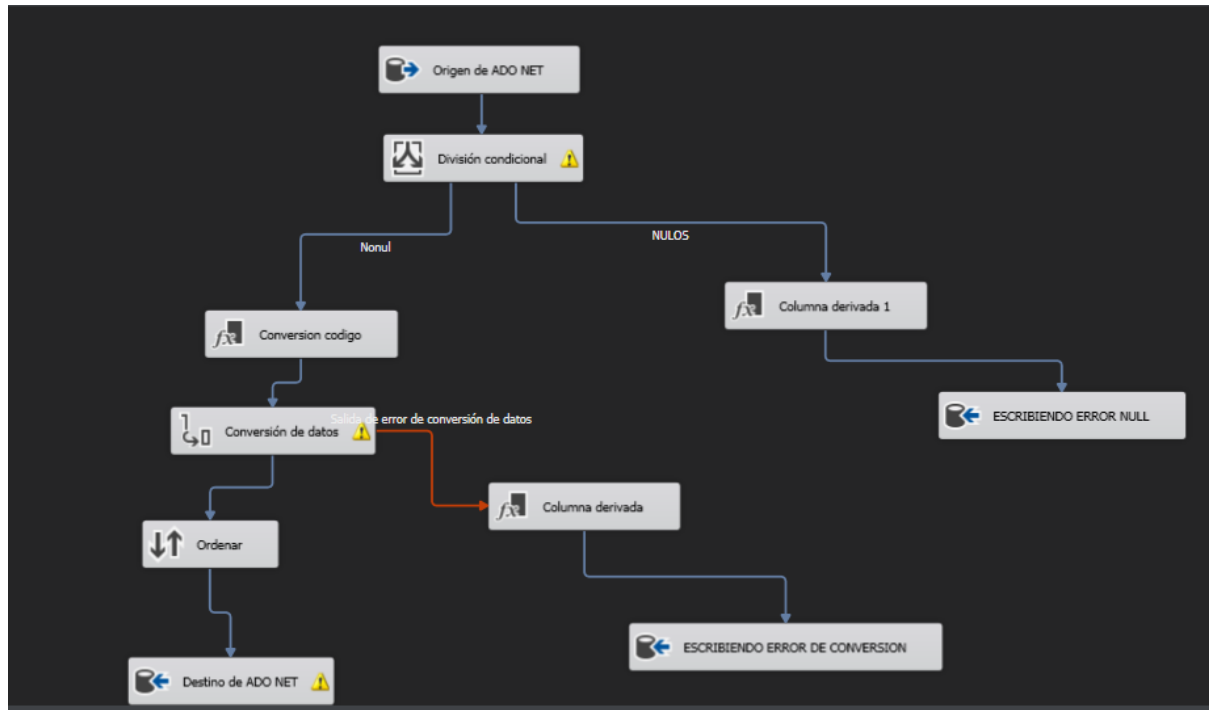
    //Creamos la variable de tipo StreamReader para realizar la lectura del contenido del archivo
    StreamReader archivo_lectura = new StreamReader(archivoactual);

    //Se aplica un split para eliminar de la ruta del archivo el simbolo \ y as? obtener
    //el nombre del archivo para compararlo
    string[] nombrearchivo = archivoactual.Split("\\").ToArray();
    while ((linea = archivo_lectura.ReadLine()) != null)
    {
        if (counter > 0)
        {
            string[] campos = linea.Split(Delimitador.ToCharArray()[0]);
            string query = "INSERT INTO tmp(file,COD_HOSPITAL,NOMBRE_HOSPITAL,DIRECCION_HOSPITAL,REGION,DEPARTAMENTO,COD_DOCTOR,NOMBRE_DOCTOR,FECHACASO,DPI,NOMBREPACIENTE,FECHANACPERSONA,GE";
            MySqlCommand comandoSql = new MySqlCommand(query, mySqlConnection);
            comandoSql.ExecuteNonQuery();
            counter++;
        }
        archivo_lectura.Close();
        MessageBox.Show("Archivo DataH1.csv fue analizado correctamente");
        //Se procede a Cerrar la conexión de la base de datos utilizada

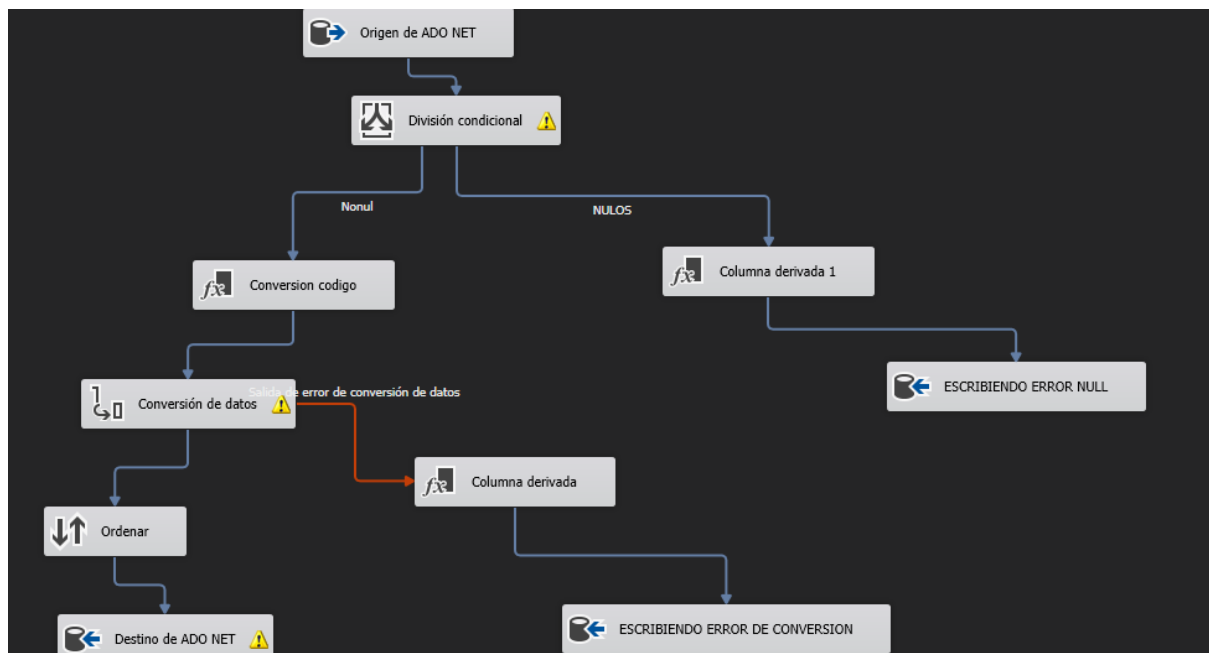
        mySqlConnection.Close();
        //Se genera un resultado de éxito para continuar con la siguiente secuencia
        Dts.TaskResult = (int)ScriptResults.Success;
    }
}
catch (Exception exception)
```

# Llenado de dimensiones

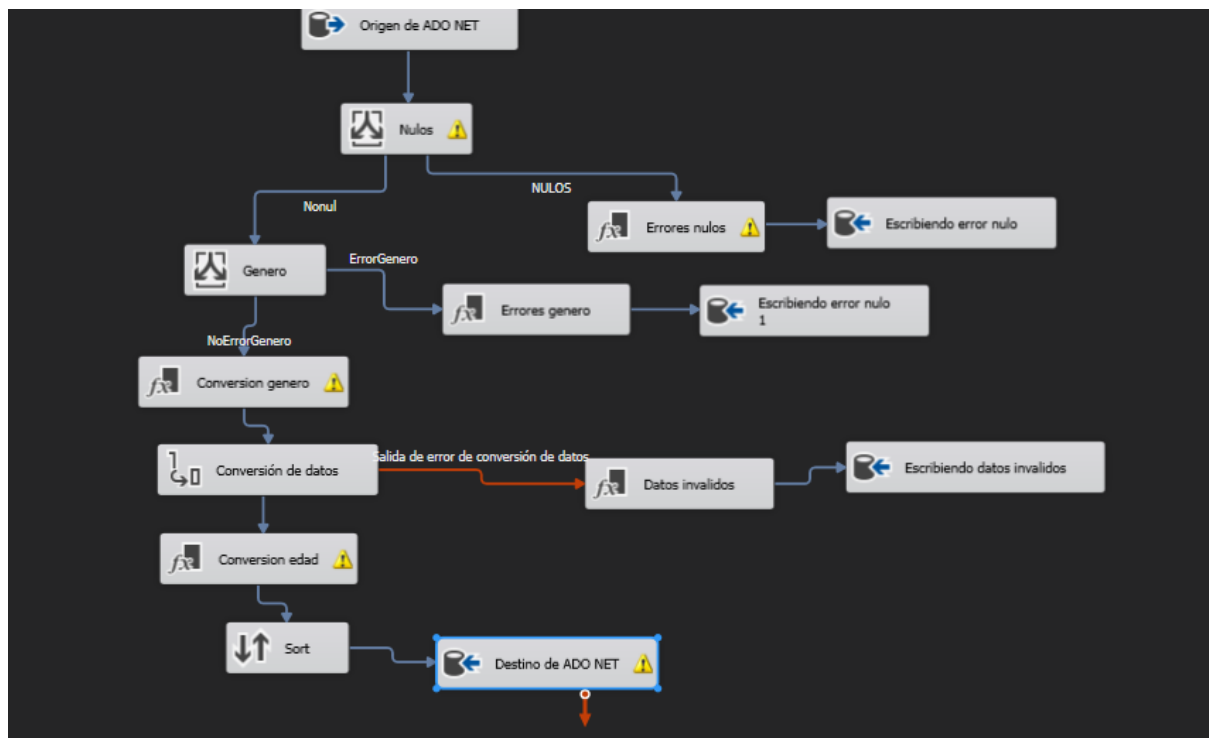
Doctor:



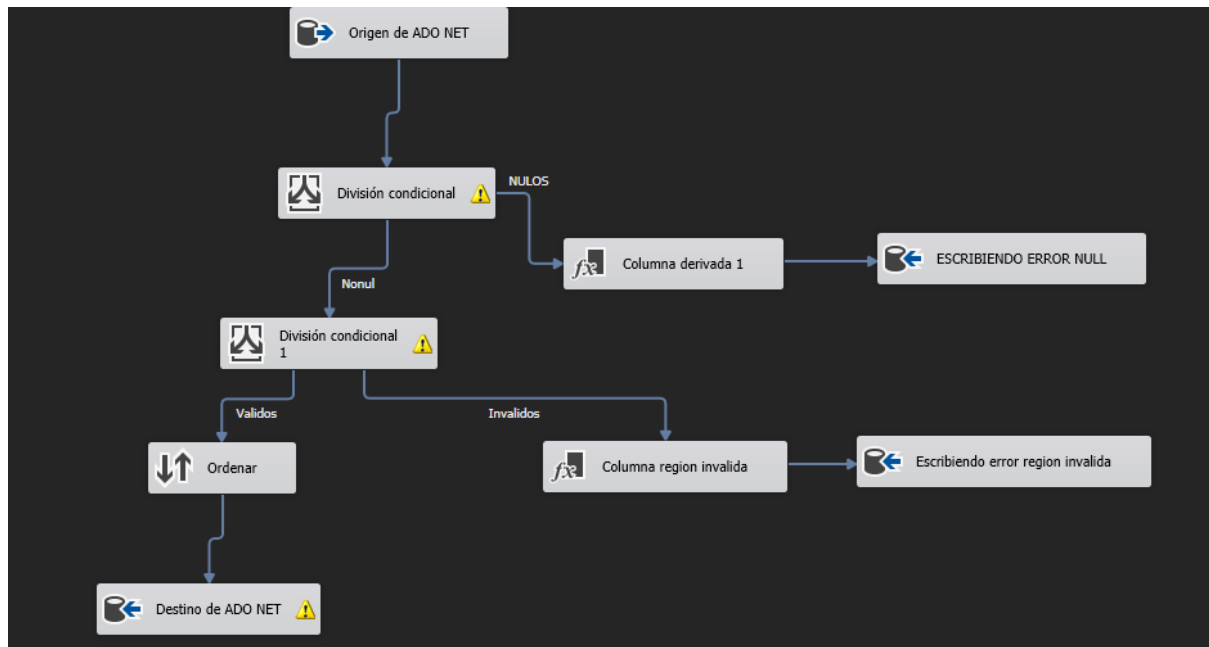
Hospital:



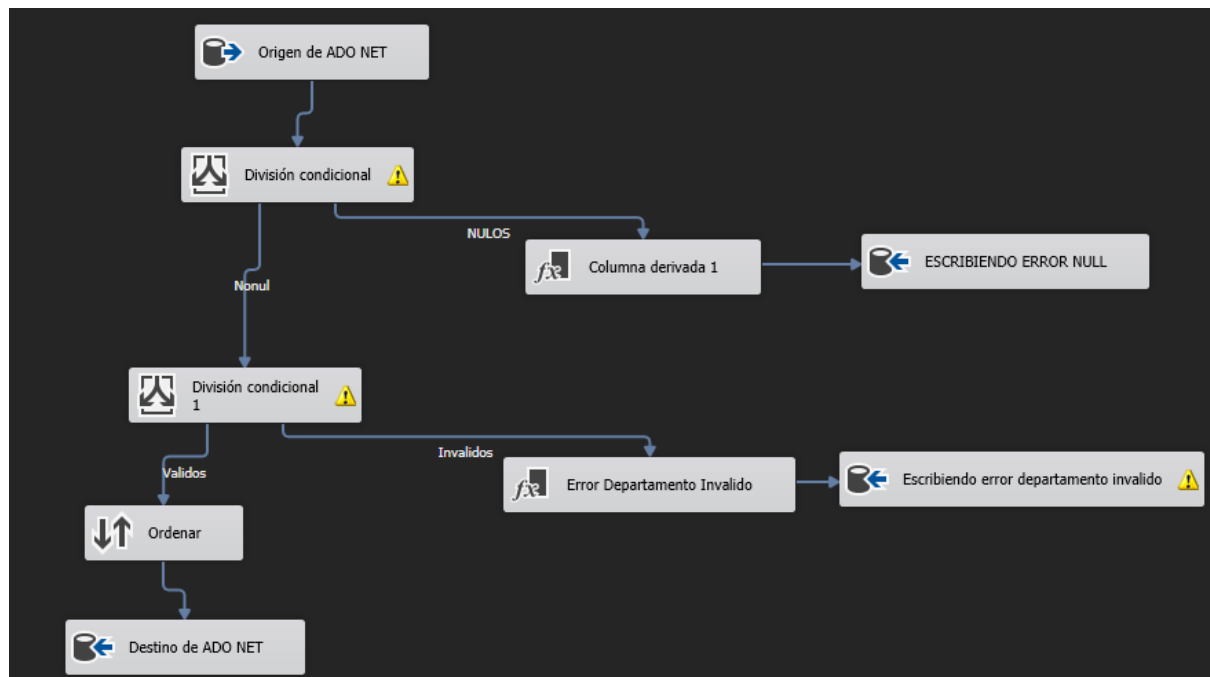
Paciente:



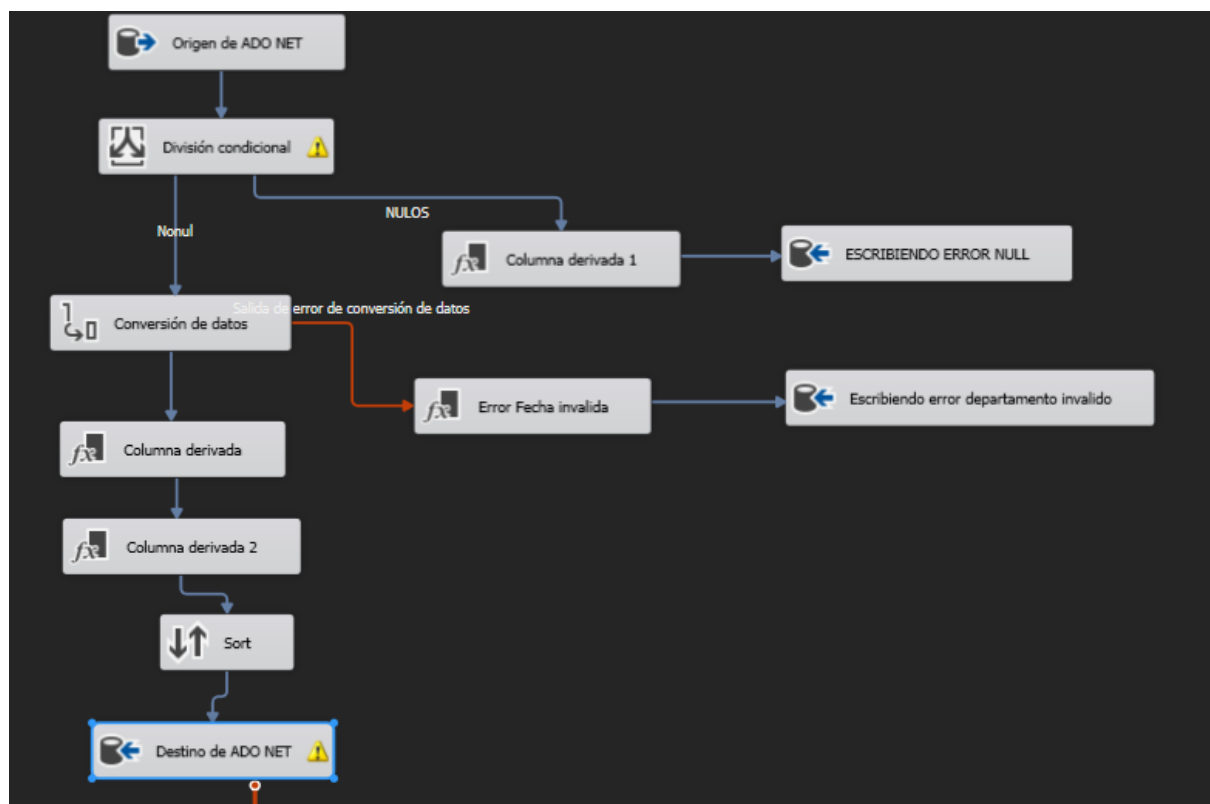
Region:



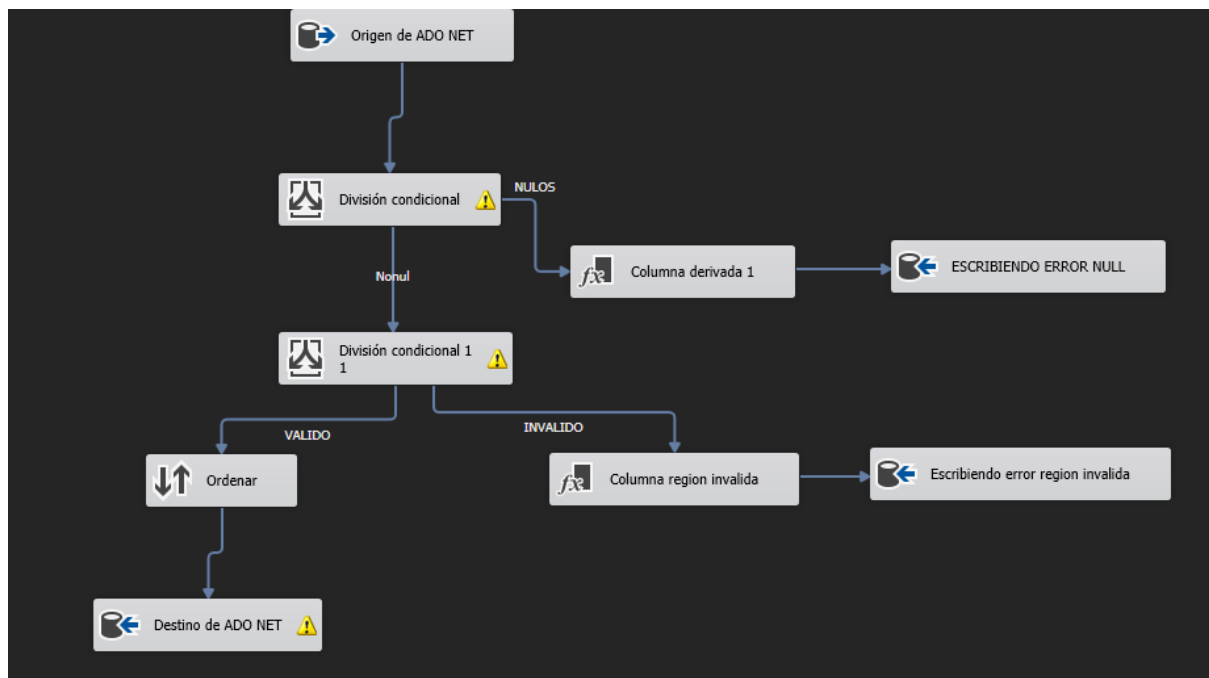
Departamento:



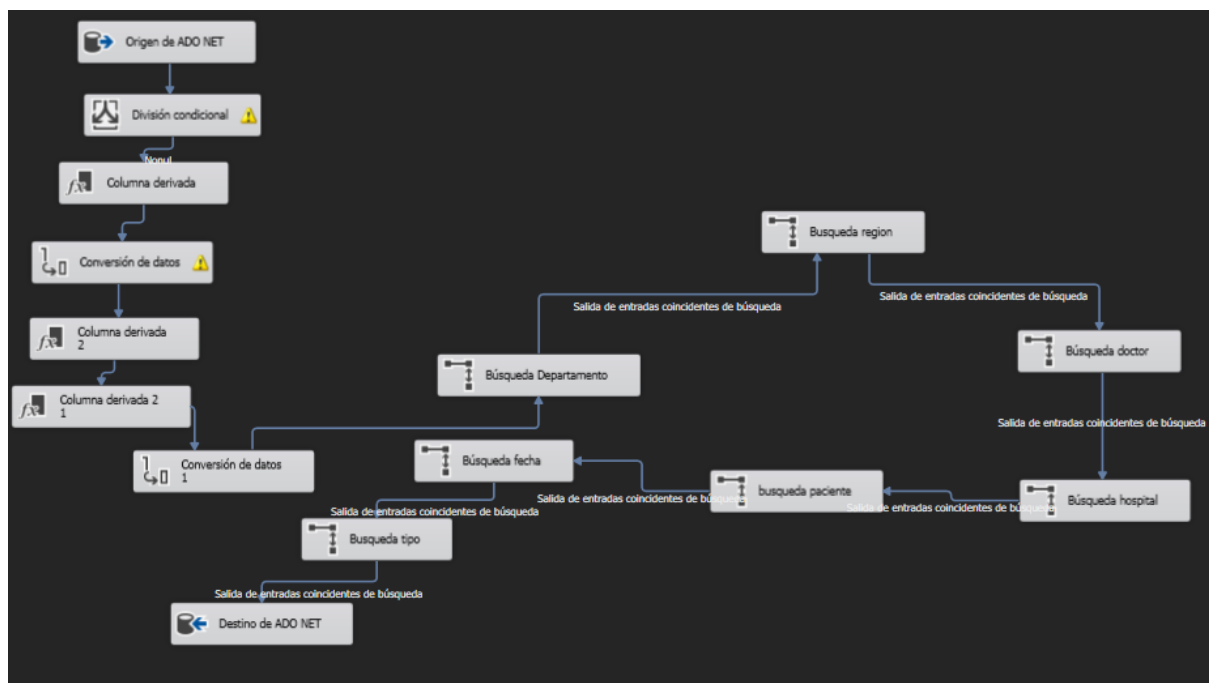
Fecha:



## Tipo:



## LLENADO DE LA TABLA DE HECHOS

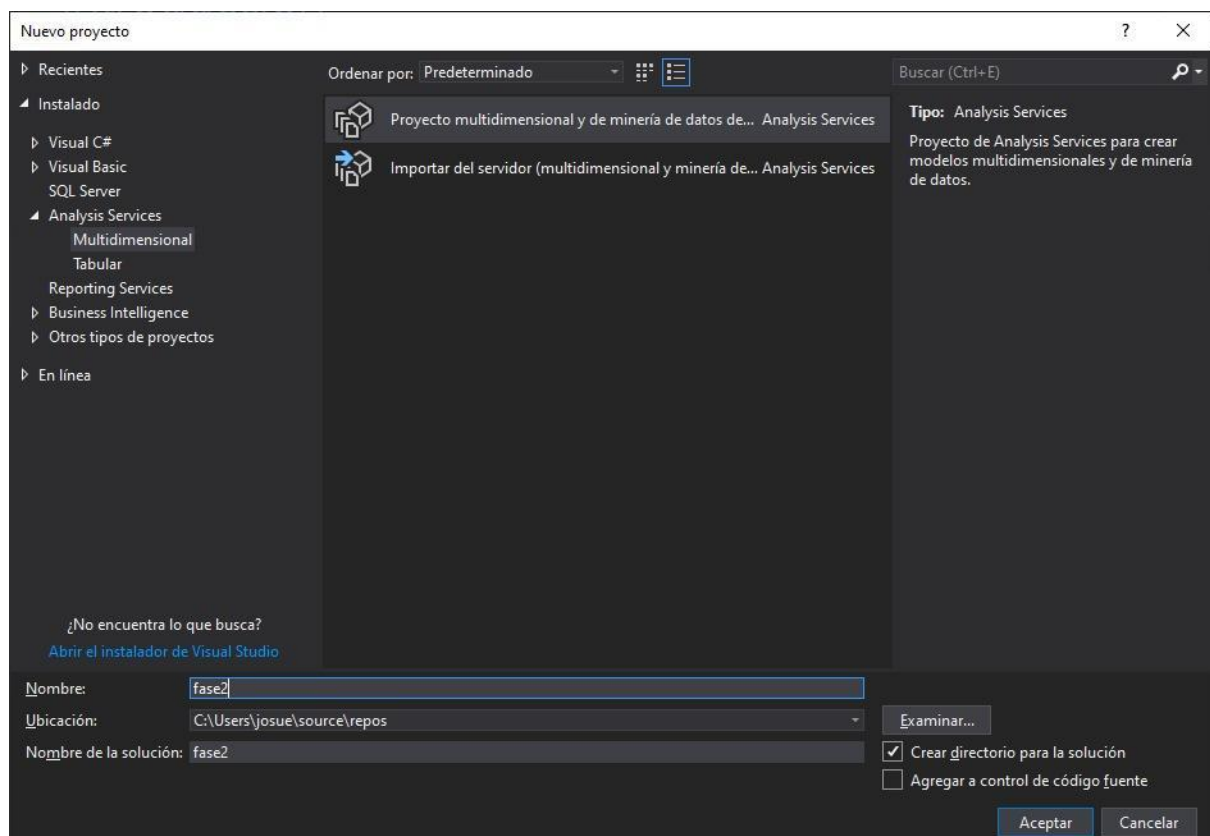


A partir de la data proporcionada por la organización se concluye que es fundamental poder segmentar los datos de una forma en la que predomine el acceso y relación de todos los involucrados en un caso detectado. Para la fácil manipulación y acceso de los datos la modularización de los datos se convierte en una necesidad.

Consideramos que el segmentar los datos y filtrarlos por cada una de las variables involucradas en un caso, es decir localidad, personal, e incluso el tiempo es la mejor forma para poder disponer de una amplia cantidad de combinaciones para generar reportes de una forma rápida, efectiva y eficiente.

Tomando en cuenta que el generar este tipo de estructuras es una forma de organizar los datos para presentar reportes e informes a distintos tipos de usuario dentro de una misma organización, es fundamental que los datos no dependan unos de otros, al menos no más allá de un par de niveles, en este caso se cuenta con una tabla hechos, donde las dimensiones giran a su alrededor, estando a un solo nivel mas de acceso las distintas relaciones; asegurando así la velocidad y efectividad de cada uno de los reportes previamente definidos.

## Creación del Cubo OLAP



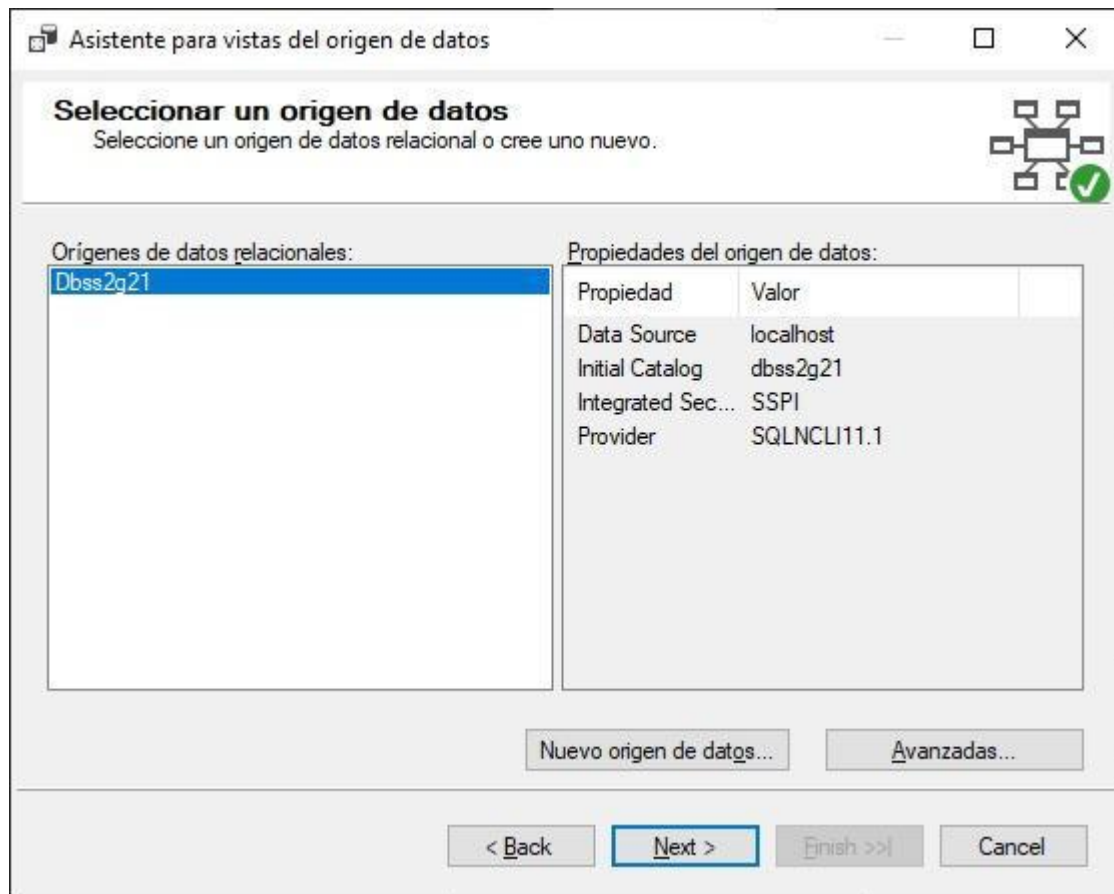
Comenzamos con la creación de un proyecto de Analysis Services, del tipo multidimensional.



Creamos la conexión de datos, con la base donde tenemos toda nuestra información. Accedemos por medio de localhost al servicio de sql server management e ingresamos a la base de datos que estamos utilizando.

En este paso debemos poner especial atención ya que a partir de acá se pueden originar muchos errores que no se verán reflejados hasta que se intente recrear la estructura, y se deberá rearmar y reconfigurar todos los pasos previos a la creación de la estructura.

Se recomienda utilizar localhost como el punto de acceso a la base de datos y la autenticación con un usuario de windows (donde luego se pedirá proporcionar el nombre y la contraseña con la que se configuró el equipo de windows en un inicio).



Nuestra base de datos se llama DBss2g21, por lo que la elegimos y procedemos.

Para asegurarnos que nuestra fuente de datos es la correcta podemos revisar la descripción de cada una de las propiedades de la conexión a base de datos, como podemos ver en este caso, estas corresponden a la conexión de datos que inicialmente habíamos preconfigurado.

Este origen de datos será la fuente que nos proveerá de la data inicial para poder generar el cubo a partir de información ya filtrada, y estandarizada en los pasos previos del proyecto.

**Asistente para vistas del origen de datos**

**Seleccionar tablas y vistas**  
 Seleccione los objetos de la base de datos relacional que deben incluirse en la vista del origen de datos.

Objetos disponibles:

Nombre	Tipo
error (dbo)	Tabla

Filtrar:

☐ Mostrar objetos del sistema

Objetos incluidos:

Nombre	Tipo
caso (dbo)	Tabla
departamento (dbo)	Tabla
doctor (dbo)	Tabla
fecha (dbo)	Tabla
hospital (dbo)	Tabla
paciente (dbo)	Tabla
region (dbo)	Tabla

Agregar tablas relacionadas

< Back   Next >   Finish >>   Cancel

Se agregan las tablas a utilizar, en este caso, todas menos la del error.

Estas tablas incluyen tanto las denominadas como dimensiones así como la tabla de hechos, es en realidad pasar todas nuestras estructuras de datos ya definidas en el modelo estrella, pero que serán de utilidad para transformarlas a la estructura aún más compleja llamada cubo.

Ya que estos datos son los que vamos a utilizar para combinarlos y generar una estructura óptima para las búsquedas y filtrados, excluimos por completo nuestros registros de error ya que estos eran útiles nada más para identificar cuáles son los datos y registros que no están bien conformados y que por tanto debemos excluir de cualquier análisis al que sometamos el cubo.

Asistente para cubos

**Seleccionar método de creación**  
Se pueden crear cubos usando tablas existentes, creando un cubo vacío o generando tablas en el origen de datos.

¿Cómo desea crear el cubo?

☒ Usar tablas existentes

☐ Crear un cubo vacío

☐ Generar tablas en el origen de datos

Plantilla:

(Ninguno)

Descripción:

Cree un cubo basado en una o varias tablas de un origen de datos.

< Back   Next >   Finish >>   Cancel

Luego debemos crear una vista de la base de datos, con las tablas que sean de nuestra ayuda para presentar los datos.

Esta vista será la que de alguna forma relacionarán todos los datos que hemos filtrado y diseñado en nuestro modelo estrella, para ello utilizaremos las tablas ya existentes y que hemos añadido a nuestras estructuras de la fuente de datos creada.

Asistente para cubos

### Seleccionar tablas de grupo de medida

Seleccione una vista o diagrama de origen de datos y, a continuación, seleccione las tablas que se usarán para los grupos de medida.

Vista del origen de datos:

Vista

Tablas de grupo de medida:

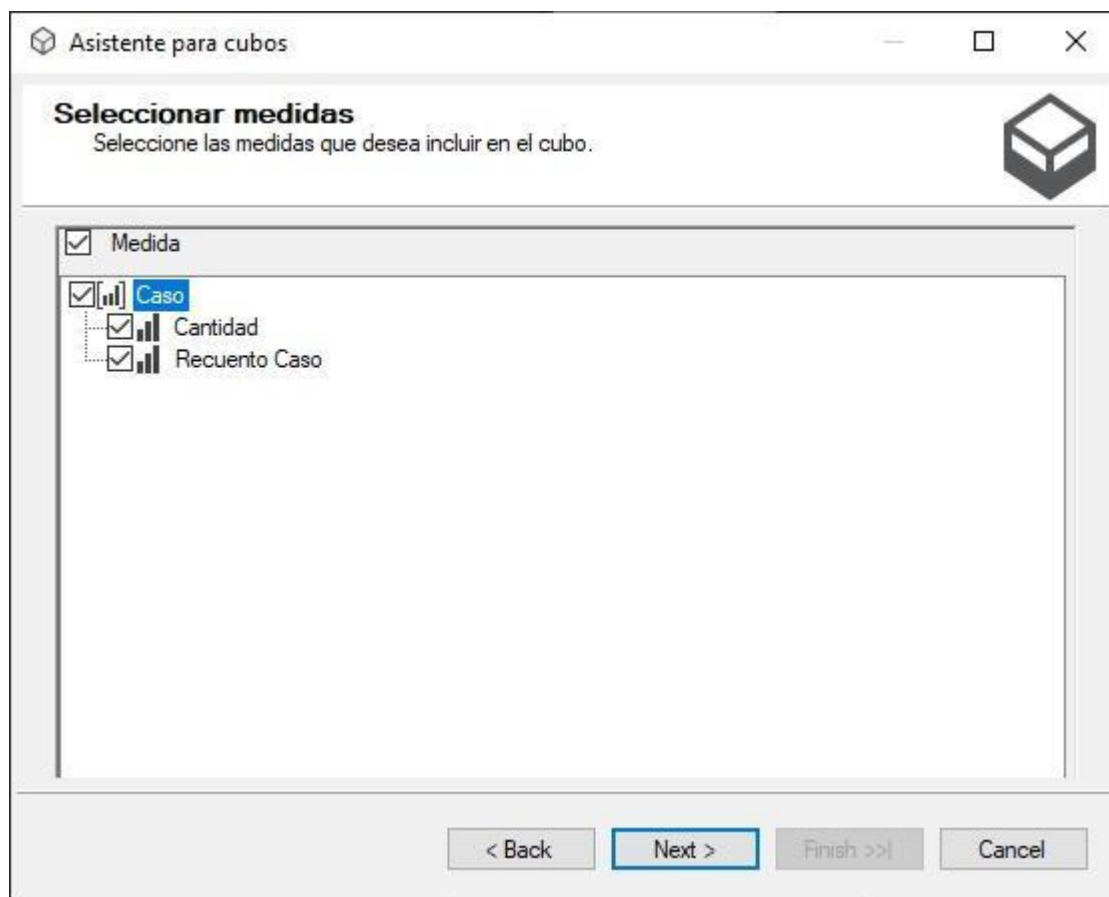
Sugerir

<input checked="" type="checkbox"/>	caso
<input type="checkbox"/>	departamento
<input type="checkbox"/>	doctor
<input type="checkbox"/>	region
<input type="checkbox"/>	paciente
<input type="checkbox"/>	hospital
<input type="checkbox"/>	fecha

< Back   Next >   Finish >>   Cancel

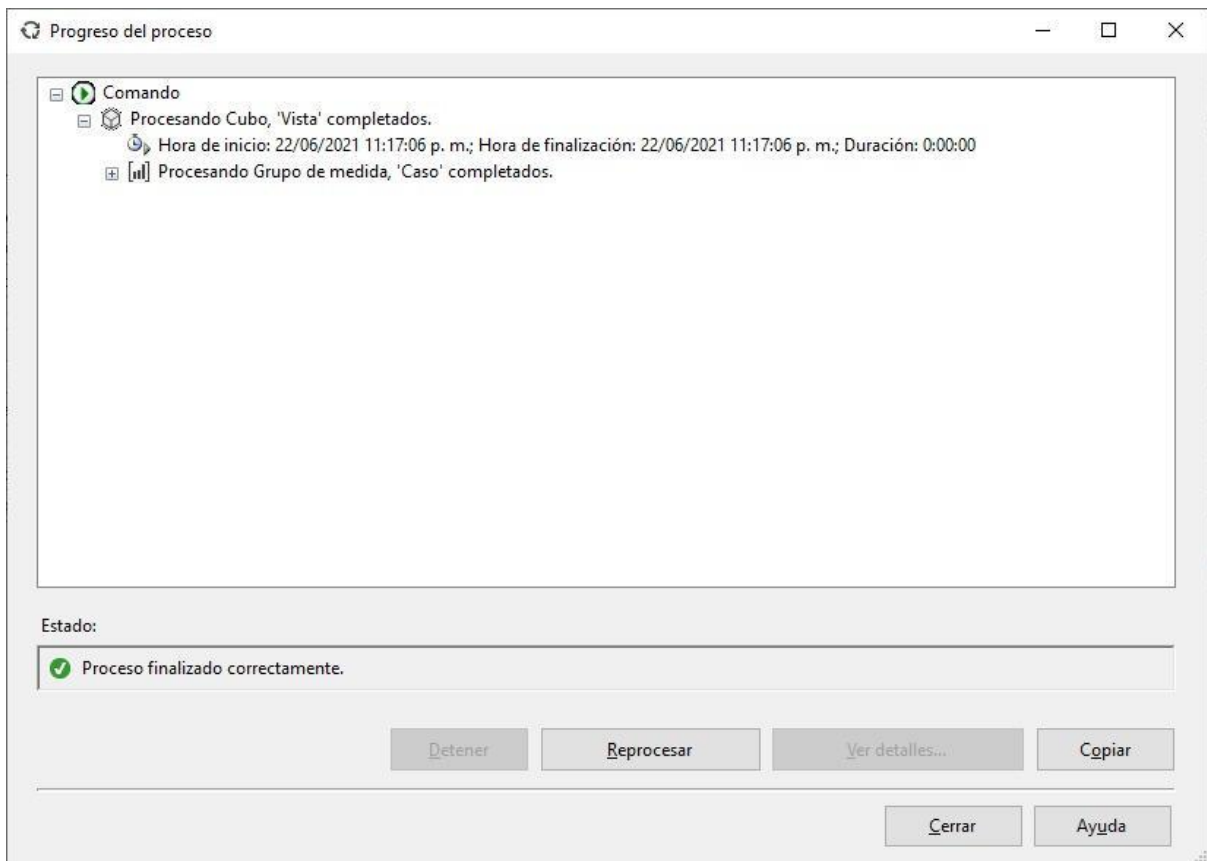
Por último, elegimos el grupo de medida, en esta ocasión es el caso sobre el que vamos a trabajar.

Debido a que estamos generando una nueva estructura de datos, es necesario seleccionar los datos de cada una de las tablas relacionadas en el cubo, para ello podemos omitir las que no vamos a utilizar si en dado caso queremos editar y reformular el cubo o al momento de generarlo nos percatamos de que hay datos que no nos son útiles.

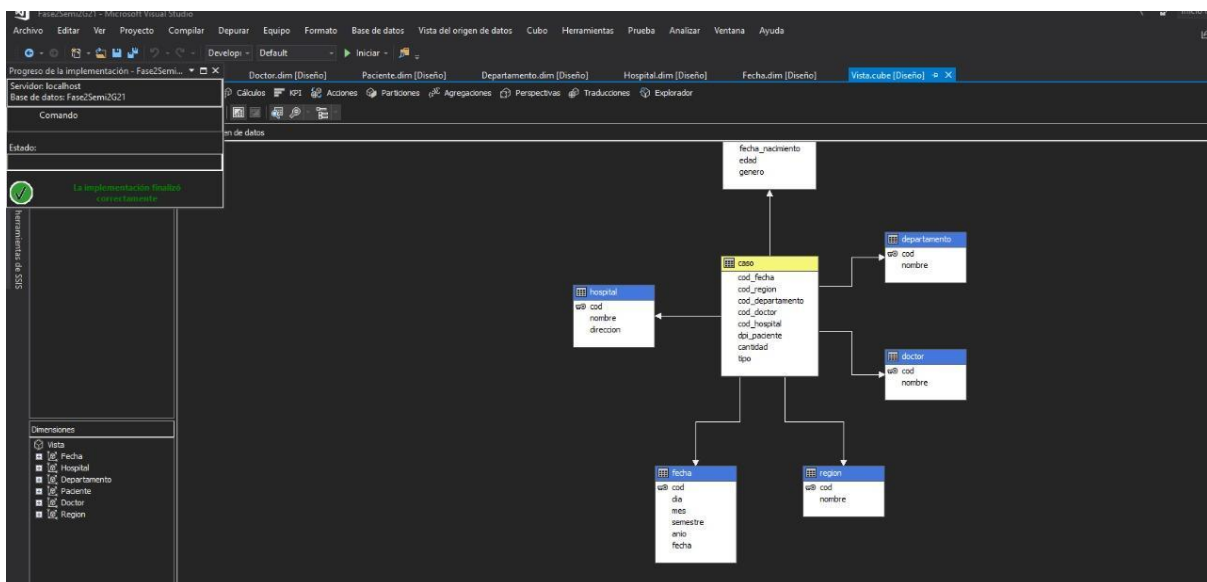


Una vez seleccionadas las medidas se concluye con la configuración de la creación de cubos.

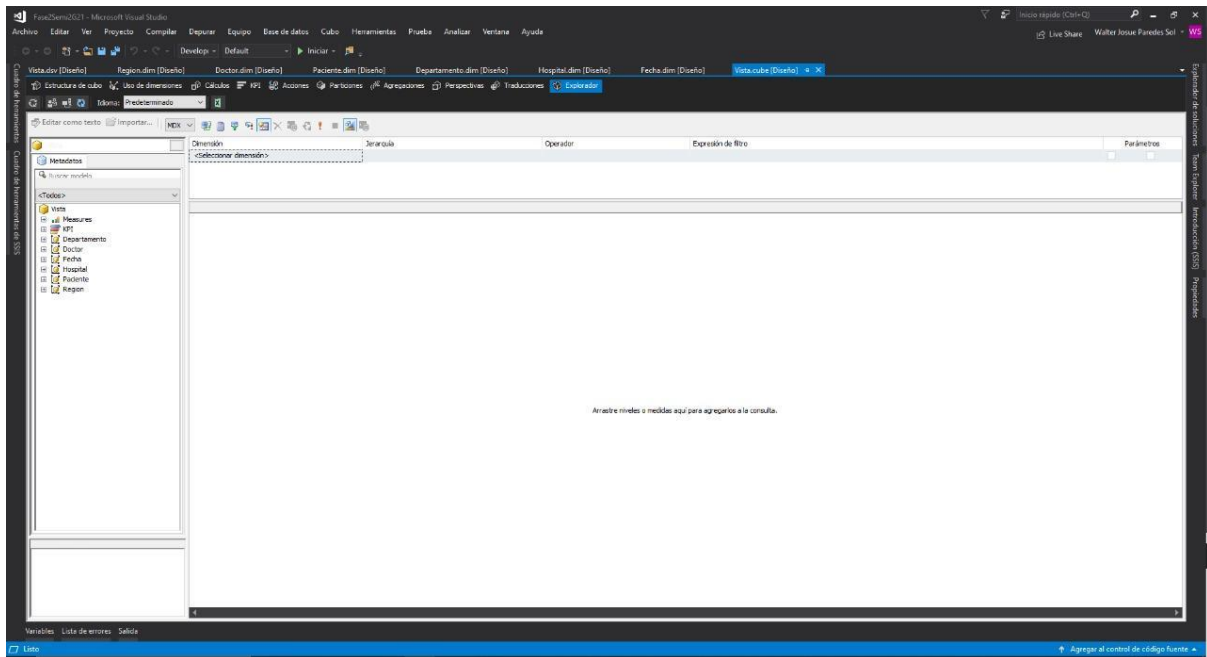
Luego de elegir una dimensión o fuente de datos del modelo, nos da opción a verificar y validar que campos de la vista queremos o deseamos involucrar dentro de la estructura, si por alguna razón a este punto decidimos que no necesitaremos uno o más campos de una dimensión acá podemos seleccionar o deseleccionar estos campos.



Para finalizar, procesamos el cubo y la información que este contiene.



La información del mismo se procesa exitosamente, notificación visible en la esquina superior izquierda de la pantalla, lo cual indica que el proceso ha sido correcto.

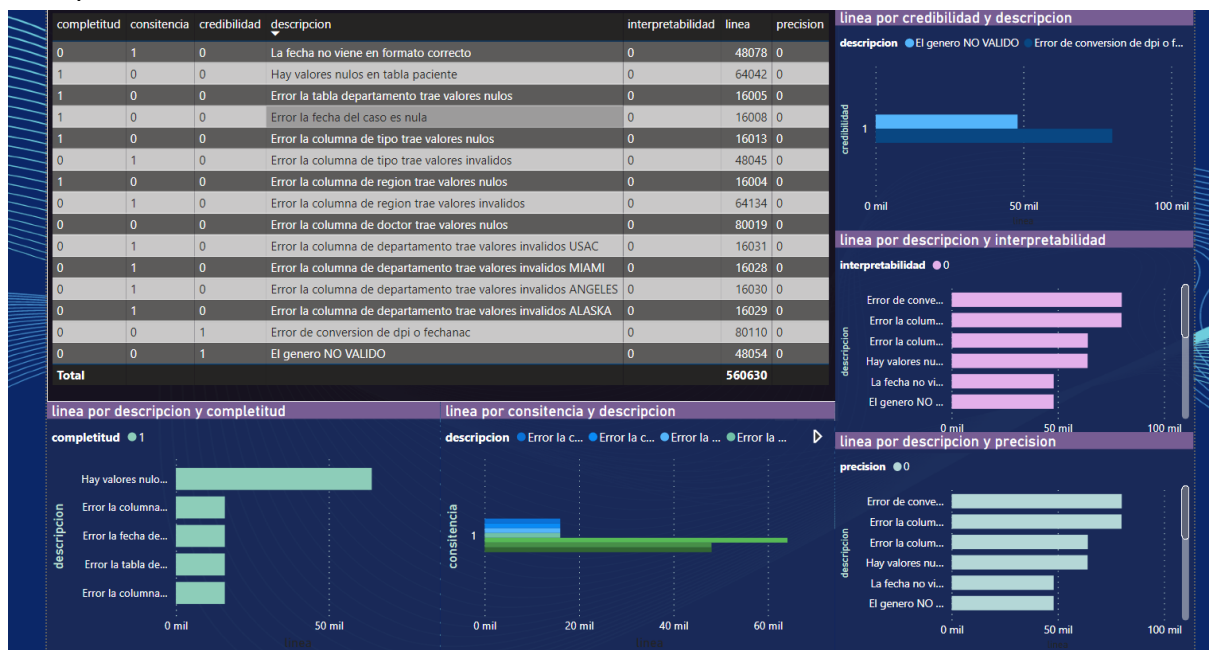


Por último, tenemos la vista final del cubo OLAP generado con la información que se recolectó desde la base de datos y las variables que se tomaron en cuenta para nuestro proyecto.

## CREACIÓN DE LOS REPORTES CON EL CUBO

### Reporte 1 - Calidad de los datos

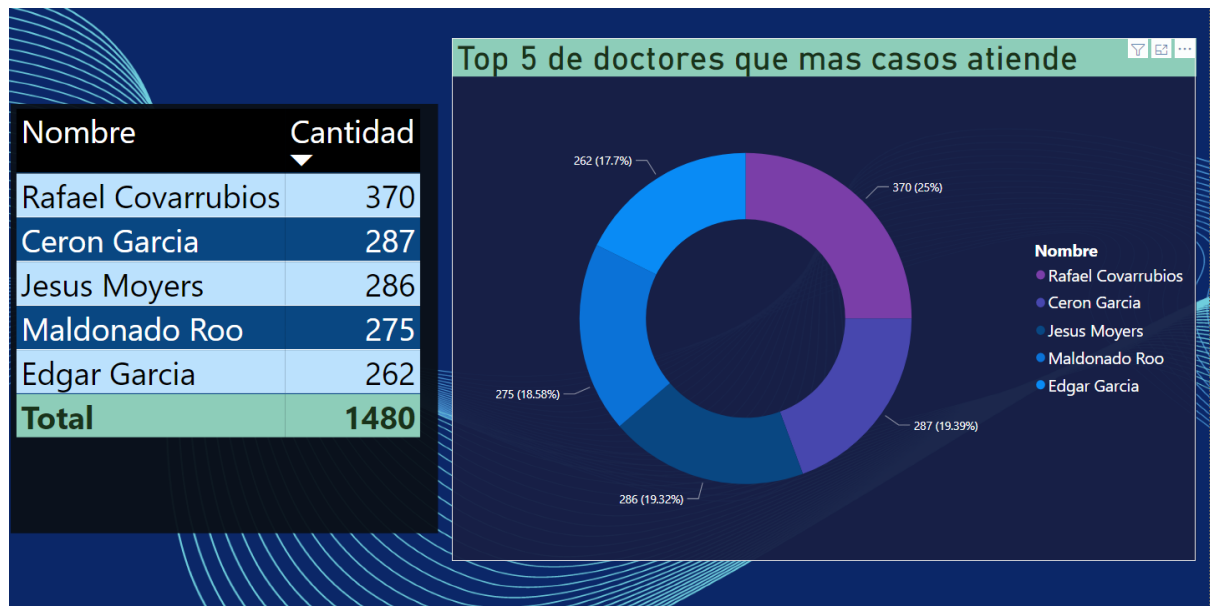
En este reporte se puede observar la calidad de la fuente de nuestros datos, luego de realizar el proceso de ETL. Podemos observar que contamos con errores de completitud, consistencia y credibilidad. Por el otro lado, no se encontraron errores de interpretabilidad, ni de precisión.





## Reporte 2 - Top 5 de doctores que más casos atiende

En este reporte se observa que un doctor principalmente atendió más que todos, pero los siguientes 4 atendieron una cantidad de casos muy cercana.



## Reporte 3: Dashboard 1 - Hospitales por casos

Se observa en la gráfica de los top5 con más casos activos,recuperados y muertos que los hospitales dentro de la capital son los que más casos han tenido siendo el San Juan de dios con más casos activos, y el Roosevelt con más recuperados y más muertos.

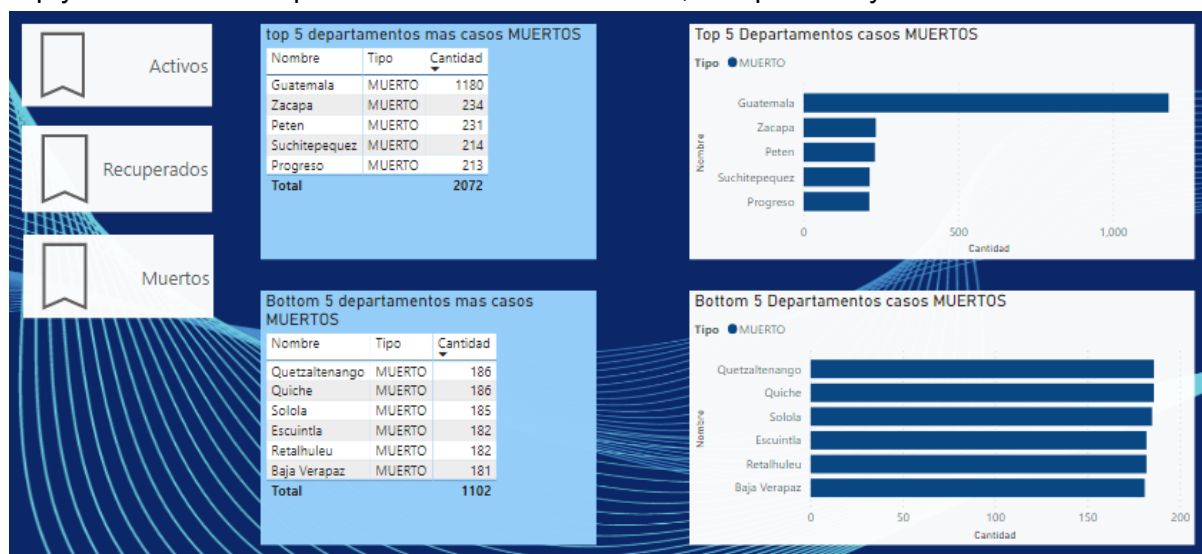


Al observar en los gráficos siguientes de los que menos activos,recuperados y muertos han tenido podemos observar que los hospitales del interior han atendido menos casos y muchos con un mayor porcentaje de muertes que recuperados.



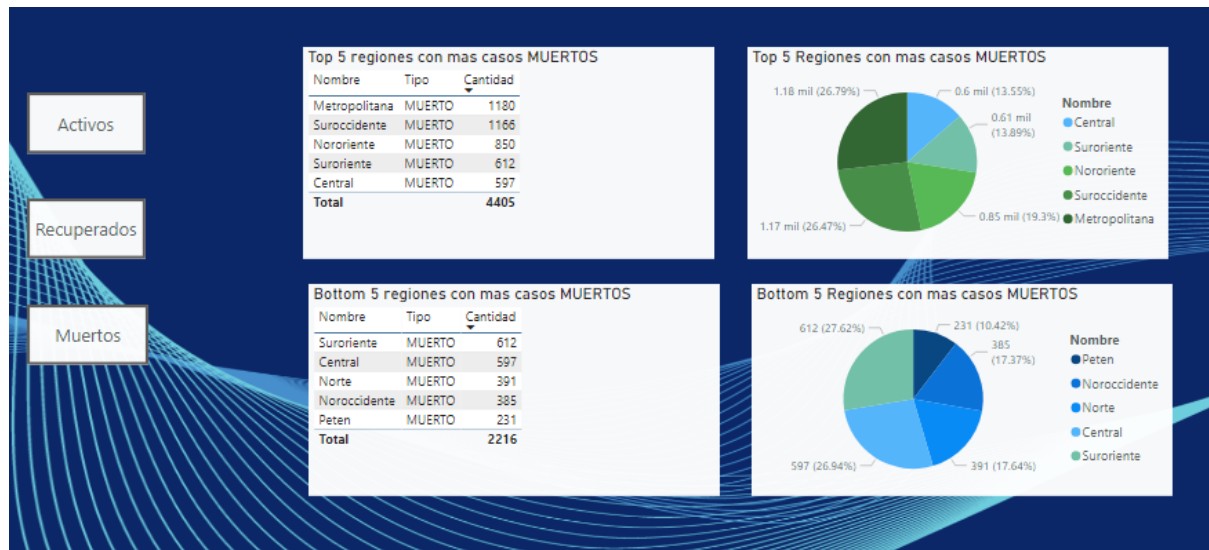
## Reporte 4: Dashboard 2 - Departamentos por caso

Top y bottom 5 de hospitales con más casos activos, recuperados y muertos.



## Reporte 5: Dashboard 3 - Regiones por caso

Top y bottom 5 de departamentos con más casos activos, recuperados y muertos.

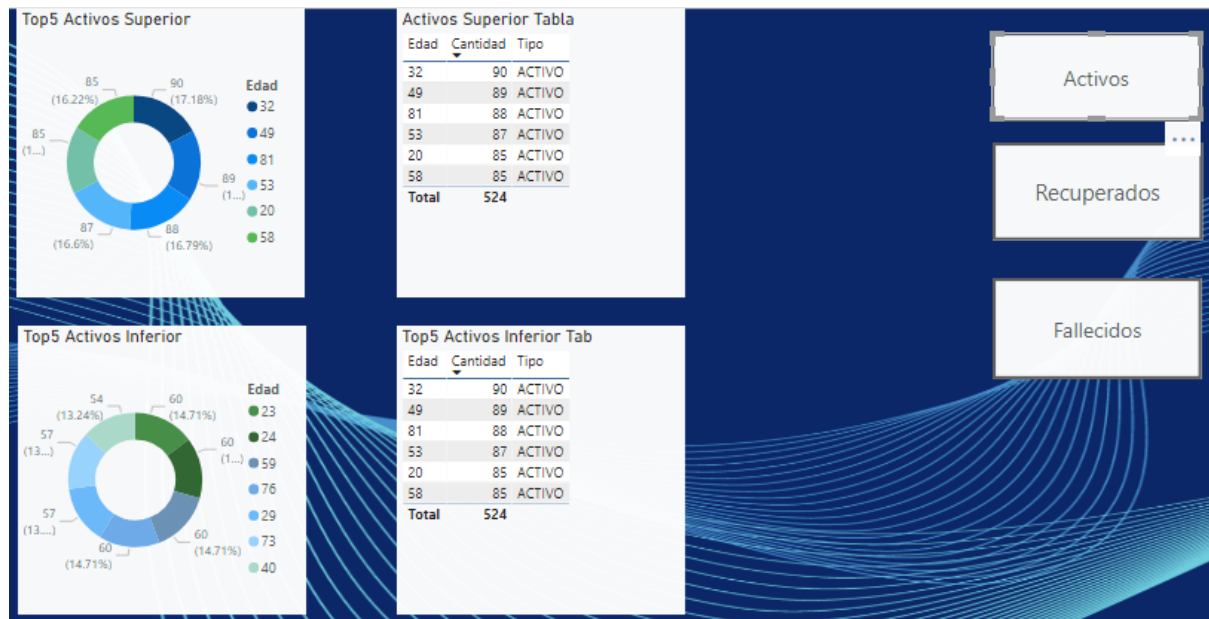


## Reporte 6: Dashboard 4 - Casos por genero



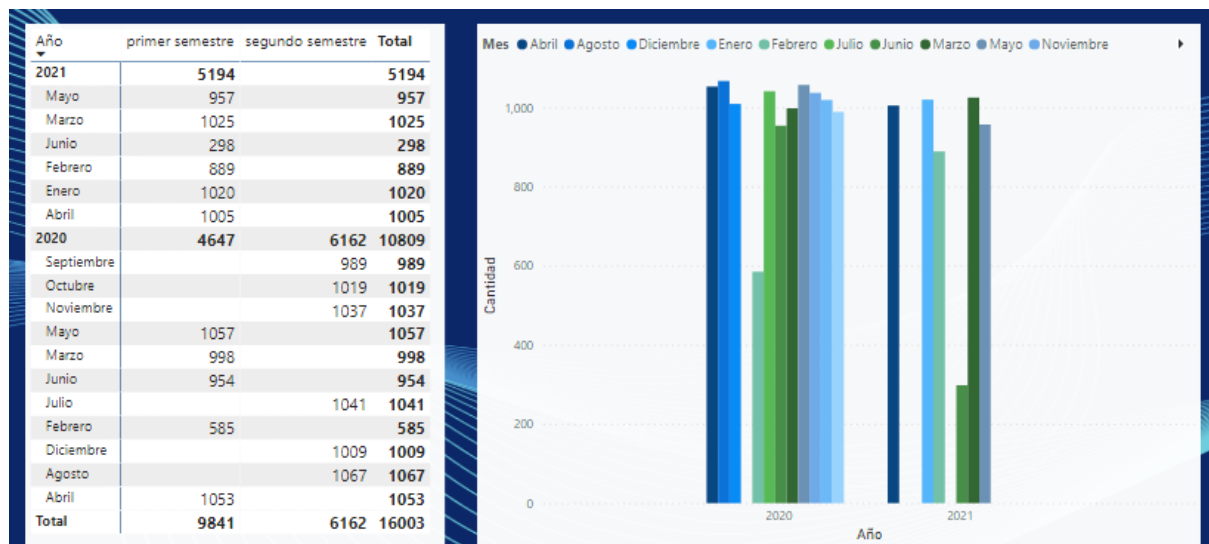
## Reporte 7: Dashboard 5 - Casos por edad

Reporte que detalla la cantidad de activos, recuperados y fallecidos Top - Bottom 5.



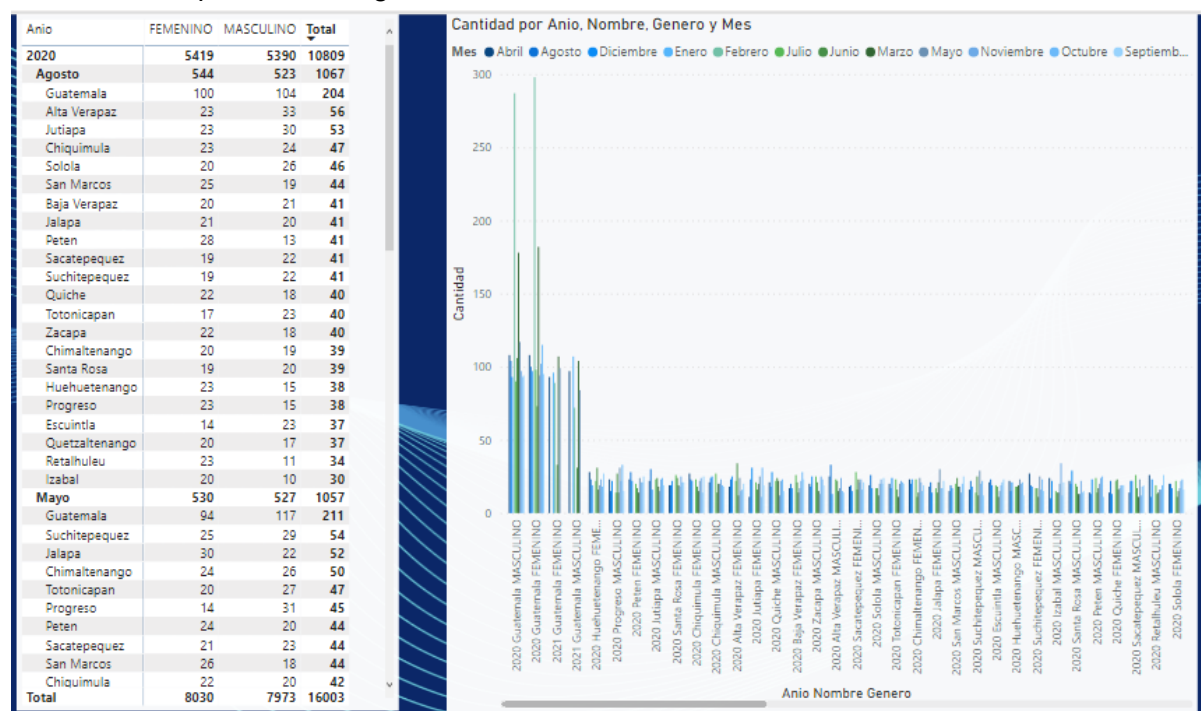
## Reporte 8 - Más casos registrados por fecha

Fecha con más casos registrados, incluir cantidades ordenadas descendente para cada dimensión.



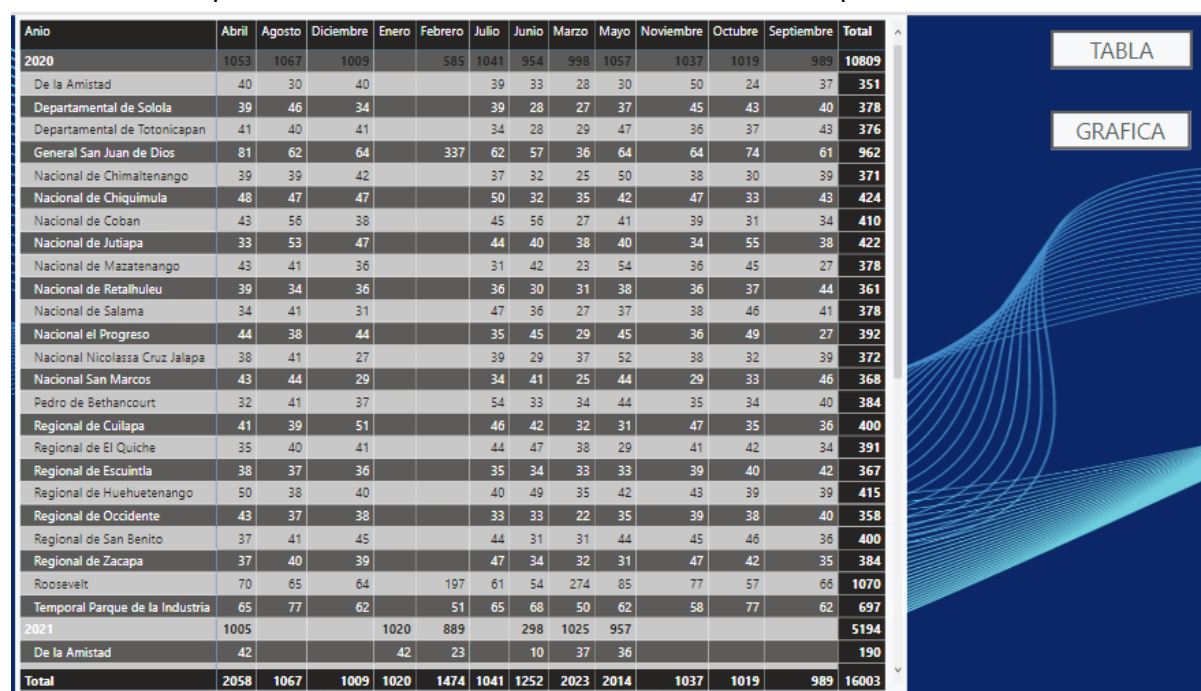
## Reporte 9 - Más casos registrados por género

Casos registrados por género, clasificados por mes del año, pudiendo navegar por año->mes->departamento->género.

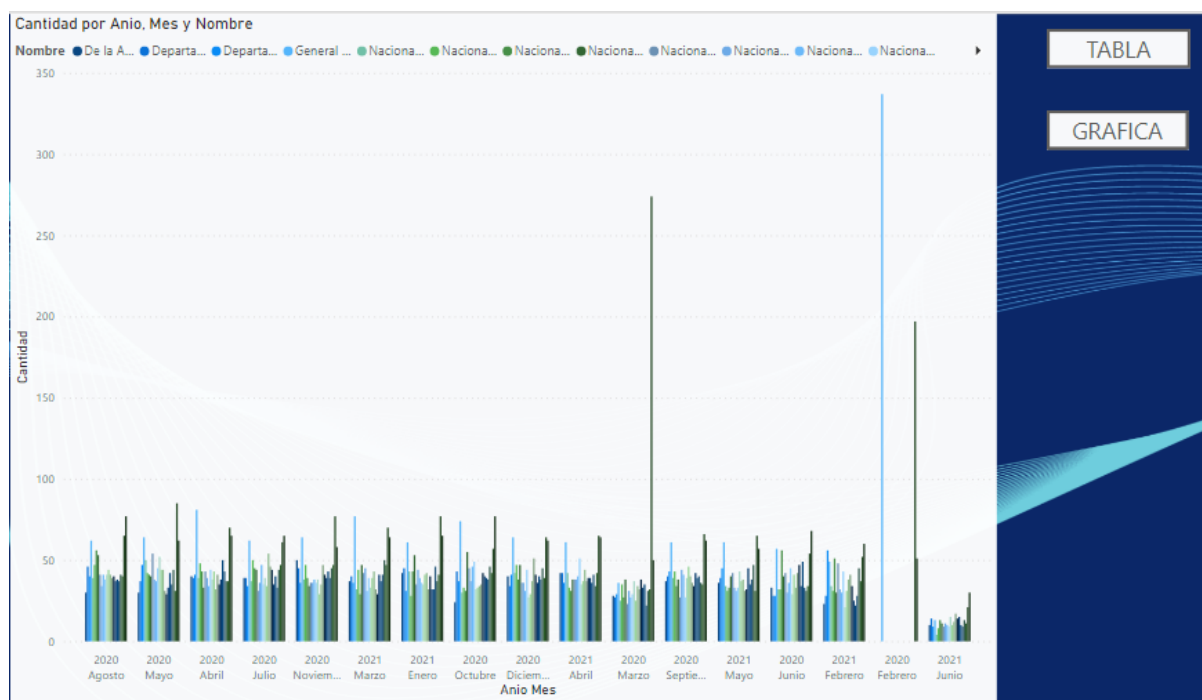


## Reporte 10 - Casos atendidos por hospital

Casos atendidos por hospital, clasificados por mes del año, pudiendo navegar por año->mes->hospital, incluir cantidades ordenadas descendente para cada dimensión.







## Reporte 11 - Casos atendidos por doctor

Casos atendidos por doctor, clasificados por mes del año, pudiendo navegar por año->mes->hospital->doctor, incluir cantidades ordenadas descendente para cada dimensión.

Año	Abril	Agosto	Diciembre	Enero	Febrero	Julio	Junio	Marzo	Mayo	Noviembre	Octubre	Septiembre	Total
2020	1053	1067	1009		585	1041	954	998	1057	1037	1019	989	10809
De la Amistad	40	30	40			39	33	28	30	50	24	37	351
Bernardo Sanchez	11	10	9			12	10	6	10	16	4	15	103
Jorge Pardo	17	12	18			10	9	9	8	16	14	16	129
Julio Estrada	12	8	13			17	14	13	12	18	6	6	119
Departamental de Solola	39	46	34			39	28	27	37	45	43	40	378
Luisa Mazariegos	16	12	13			10	13	10	12	20	11	12	129
Maria Estrada	15	16	11			16	7	10	9	9	18	14	125
Roberta Manrique	8	18	10			13	8	7	16	16	14	14	124
Departamental de Totonicapán	41	40	41			34	28	29	47	36	37	43	376
Josue Ramirez	15	12	18			13	9	9	16	14	13	18	137
Leonel Lima	9	18	9			11	12	12	18	11	11	13	124
Miguel Manriquez	17	10	14			10	7	8	13	11	13	12	115
General San Juan de Dios	81	62	64		337	62	57	36	64	64	74	61	962
Arturo Arevalo	19	16	9		37	17	13	7	14	14	18	11	175
Ceron Garcia	20	11	19		96	17	12	7	7	9	8	14	220
Edgar Garcia	12	7	10		86	11	11	8	12	12	18	13	200
Edgar Ortiz	13	13	11		76	5	11	6	14	12	17	12	190
Gustavo Ortiz	17	15	15		42	12	10	8	17	17	13	11	177
Nacional de Chimaltenango	39	39	42			37	32	25	50	38	30	39	371
Hector Vera	20	14	14			14	8	7	12	8	8	12	117
Manuel Soto	7	14	13			11	10	6	17	14	11	15	118
Moreno Cruz	12	11	15			12	14	12	21	16	11	12	136
Nacional de Chiquimula	48	47	47			50	32	35	42	47	33	43	424
Cindy Garcia	16	13	8			19	7	12	17	17	12	15	136
Karla Ruiz	15	17	20			17	8	8	13	16	8	14	136
Sofia Soto	17	17	19			14	17	15	12	14	13	14	152
Total	2058	1067	1009	1020	1474	1041	1252	2023	2014	1037	1019	989	16003

