

Carrera: Ingeniería en Informática y Sistemas
Curso: Análisis de Datos
Catedrático: Stanly Bolaños



MANUAL TÉCNICO

Manuel Andoni Zamora Hernández

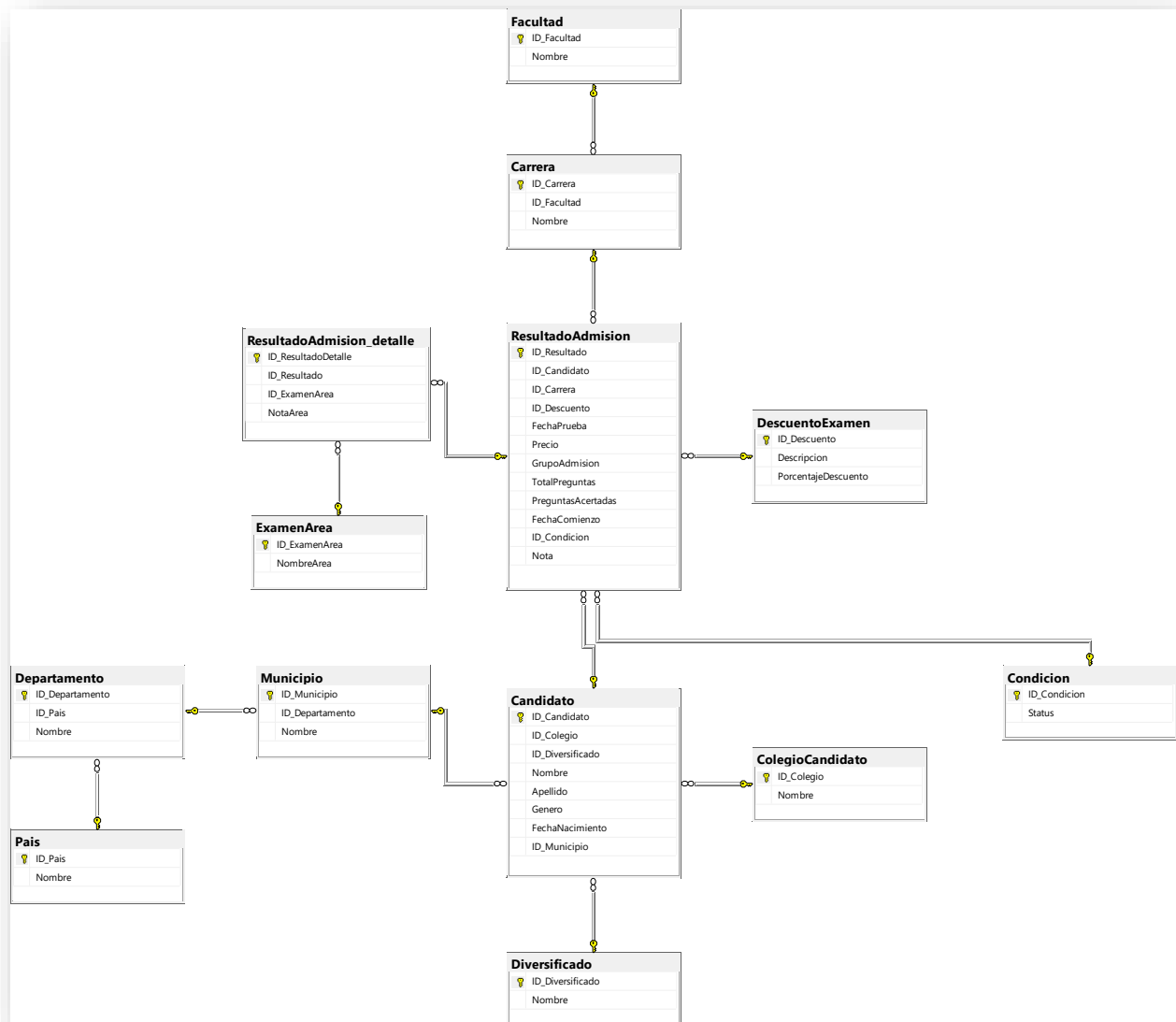
Joseph Abraham Soto Gutiérrez

Pedro Pablo Pineda Izquierdo

Tabla de contenido

Modelo de la Base de Datos	3
Tablas	3
Modelo Data Warehouse	5
Paquete de Integración SSIS	7
Carga de Datos	7
Paquete de analisis SSAS	10
Reportes	11

Modelo de la Base de Datos



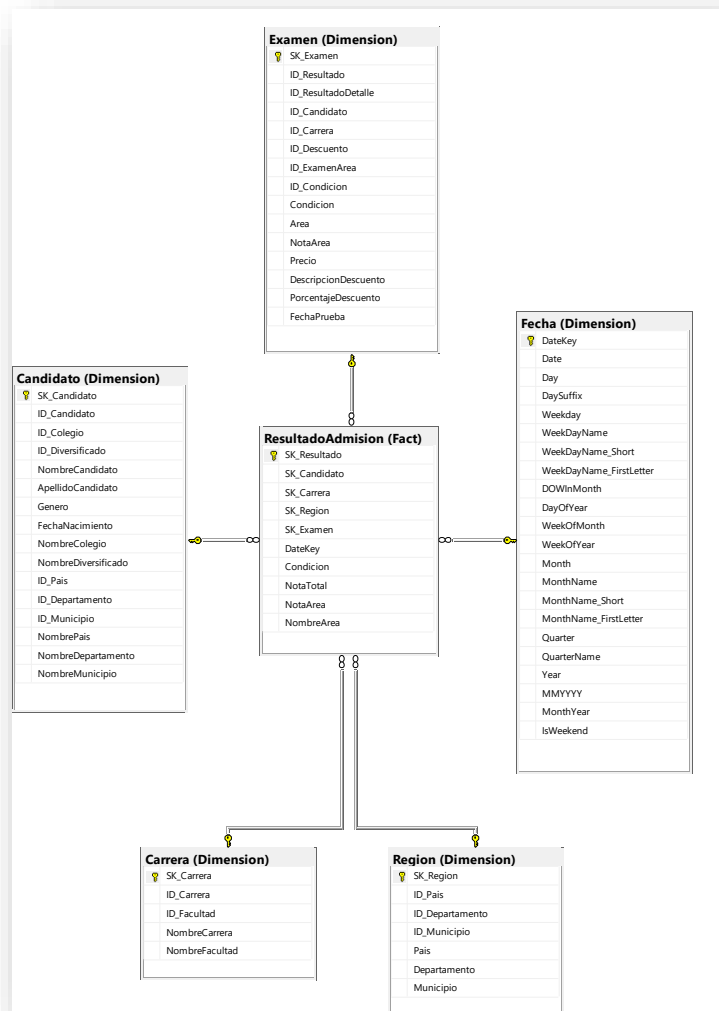
Tablas

- **Candidato** contiene la información respectiva al candidato como lo es el nombre, fecha de nacimiento (calculado mediante la edad), Genero y sus respectivos datos al momento de realizar la prueba.
- **Carrera** contiene las carreras ofrecidas por las distintas facultades de la universidad, son identificadas por un nuevo número al momento de crear un nuevo registro en la base de datos
- **ColegioCandidato** contiene el nombre de los distintos colegios donde los candidatos se encuentran estudiando o se graduaron de ellos.

- **Condicion** la tabla condición ayuda a identificar el estado en el que se encuentre el estudiante si en caso de haber perdido las pruebas se le concede de forma condicionada continuar con la solicitud. Se identifica por medio de las siglas:
 - PED - Pendiente Entrevista Director
 - PTD - Pendiente Tarea Director
 - AP - Asignado Propedéutico
 - PG - Propedéutico Ganado
- **Departamento** permite almacenar los distintos departamentos que tenga un país
- **DescuentoExamen** tabla que registra los distintos descuentos otorgados a los estudiantes para poder ser aplicados al momento de pagar las cuotas mensuales. Actualmente no es utilizada, pero puede implementarse en un futuro.
- **Diversificado** registra las diferentes carreras de las que pueden estar cursando el candidato actualmente
- **ExamenArea** en esta tabla se registran las diferentes áreas de exámenes existentes siendo actualmente, pero con la capacidad de actualizarse las siguientes:
 - Area Basica
 - Habilidades Verbales
 - Habilidades Cuantitativas
 - Español
 - Matemáticas
 - Ciencias naturales
 - Ciencias sociales
 - Area Especifica
 - Matemática para el cálculo
 - Física
 - Química
- **Facultad** esta tabla almacena las distintas facultades que la universidad puede tener para cubrir las áreas de estudio que tengan actualmente.
- **Municipio** la tabla de municipio almacena identifica a un nivel mas exacto el lugar en el que vive
- **Pais** identifica el país del que proviene el candidato
- **ResultadoAdmision** esta tabla contiene la información del estado de la admisión del usuario, así como el resultado que obtuvo en general y permite tener toda la información accesible mediante las relaciones.
- **ResultadoAdmision_detalle** contiene la información respecto al detalle de notas en cada examen que realizo el estudiante

Nota: encontrará un archivo de respaldo (backup) de todos los modelos en la carpeta *Backups*.

Modelo Data Warehouse



Para el modelo data warehouse se escogió un esquema de tipo estrella debido a que la información que se tiene actualmente puede representarse en dimensiones separadas de forma completa, este modelo cuenta con las dimensiones Carrera, Región, Fecha, Candidato y Examen. El modelo estrella presenta la ventaja de ser más simple por lo que ofrece un mayor rendimiento en comparación al modelo copo de nieve.

La desnormalización fue efectuada alrededor de la tabla de Hechos, Resultado de admisión. Esto es debido a que las notas totales nos permiten hacer varios análisis que le dan valor a la universidad. Las dimensiones nos permitirán dar contexto a las notas para poder entender varios aspectos de como se comportan estos datos.

Tablas

Dimensiones

Las siguientes dimensiones permiten hacer filtrado de la data fácilmente e identificar los datos para su fácil análisis.

- Candidato
- Carrera
- Fecha
- Examen
- Region

Hechos

La tabla de hechos contiene las medidas aditivas y semi aditivas necesarias para comparar los resultados y poder realizar análisis.

- ResultadoAdmision

Staging

Estas tablas apoyan al esquema para poder realizar la carga de datos

- Candidato
- Carrera
- Examen
- Fecha
- Region

Region (Staging)	
🔑 SK_Region	
ID_Pais	
ID_Departamento	
ID_Municipio	
Pais	
Departamento	
Municipio	

Fecha (Staging)	
🔑 DateKey	
Date	
Day	
DaySuffix	
Weekday	
WeekDayName	
WeekDayName_Short	
WeekDayName_FirstLetter	
DOWInMonth	
DayOfYear	
WeekOfMonth	
WeekOfYear	
Month	
MonthName	
MonthName_Short	
MonthName_FirstLetter	
Quarter	
QuarterName	
Year	
MMYYYY	
MonthYear	
IsWeekend	

Examen (Staging)	
🔑 SK_Examen	
ID_Resultado	
ID_ResultadoDetalle	
ID_Candidato	
ID_Carrera	
ID_Descuento	
ID_ExamenArea	
ID_Condicion	
Condicion	
Area	
NotaArea	
Precio	
DescripcionDescuento	
PorcentajeDescuento	
FechaPrueba	

Candidato (Staging)	
🔑 SK_Candidato	
ID_Candidato	
ID_Colegio	
ID_Diversificado	
NombreCandidato	
ApellidoCandidato	
Genero	
FechaNacimiento	
NombreColegio	
NombreDiversificado	
ID_Pais	
ID_Departamento	
ID_Municipio	
NombrePais	
NombreDepartamento	
NombreMunicipio	

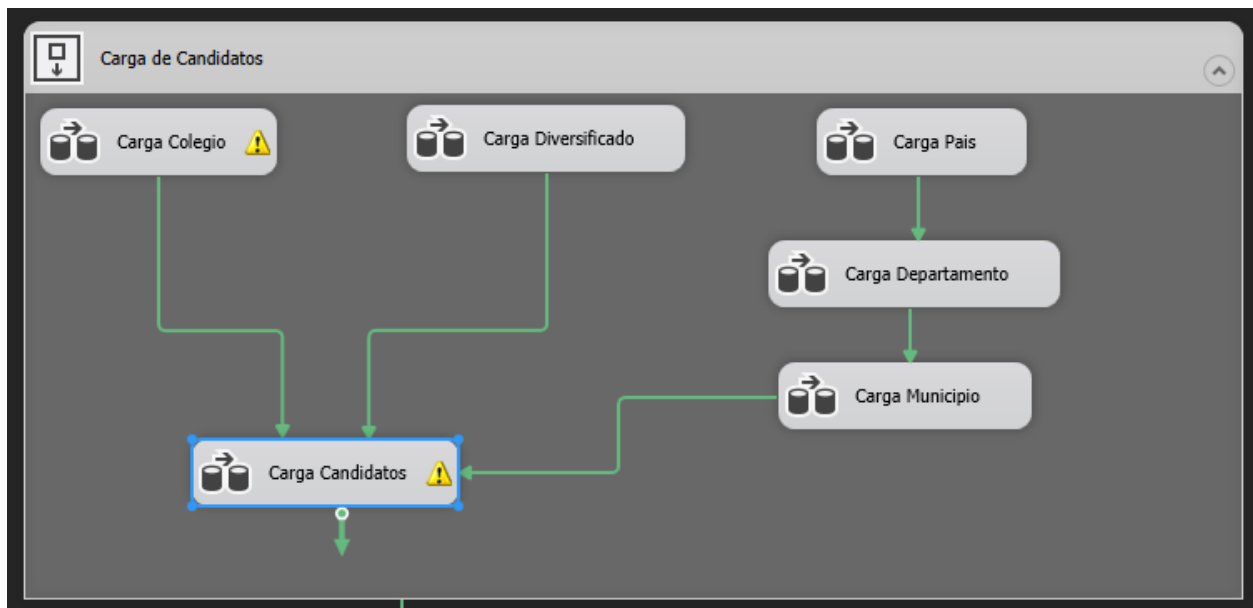
Carrera (Staging)	
🔑 SK_Carrera	
ID_Carrera	
ID_Facultad	
NombreCarrera	
NombreFacultad	

Paquete de Integración SSIS

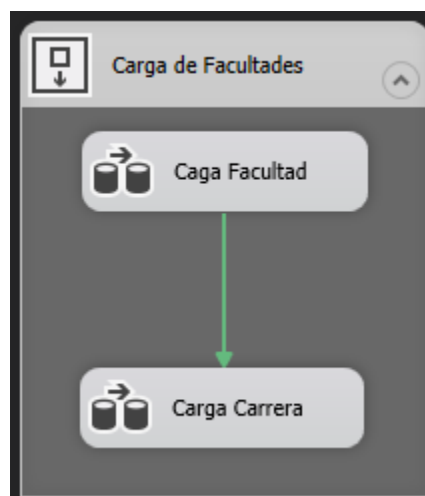
Carga de Datos

Para la carga de datos se escogió utilizar el proceso de ETL debido a que nos ofrece una gran habilidad de transformar la data y manipularla antes de que llegue al DataWarehouse. Al mismo tiempo permite separar varios aspectos de la lógica y debido a que la carga de datos se hace en fechas específicas del año puede ser manejable.

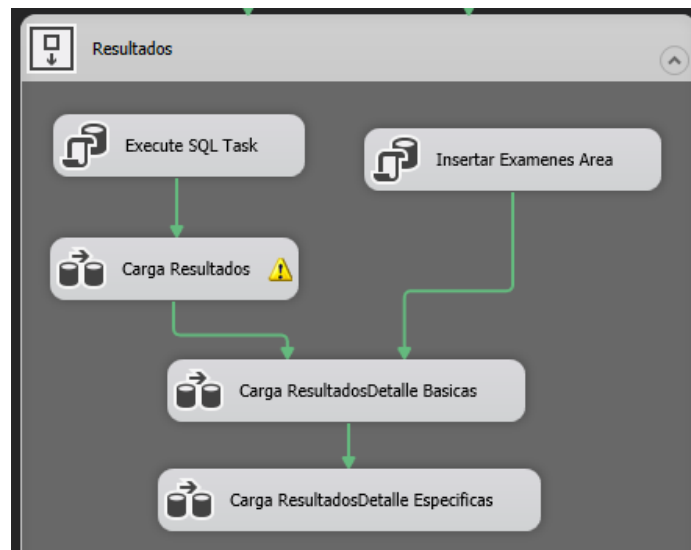
Primero se realiza la carga de datos de candidatos, siendo las tablas que no tienen dependencias hacia otra las primeras por ser cargadas.



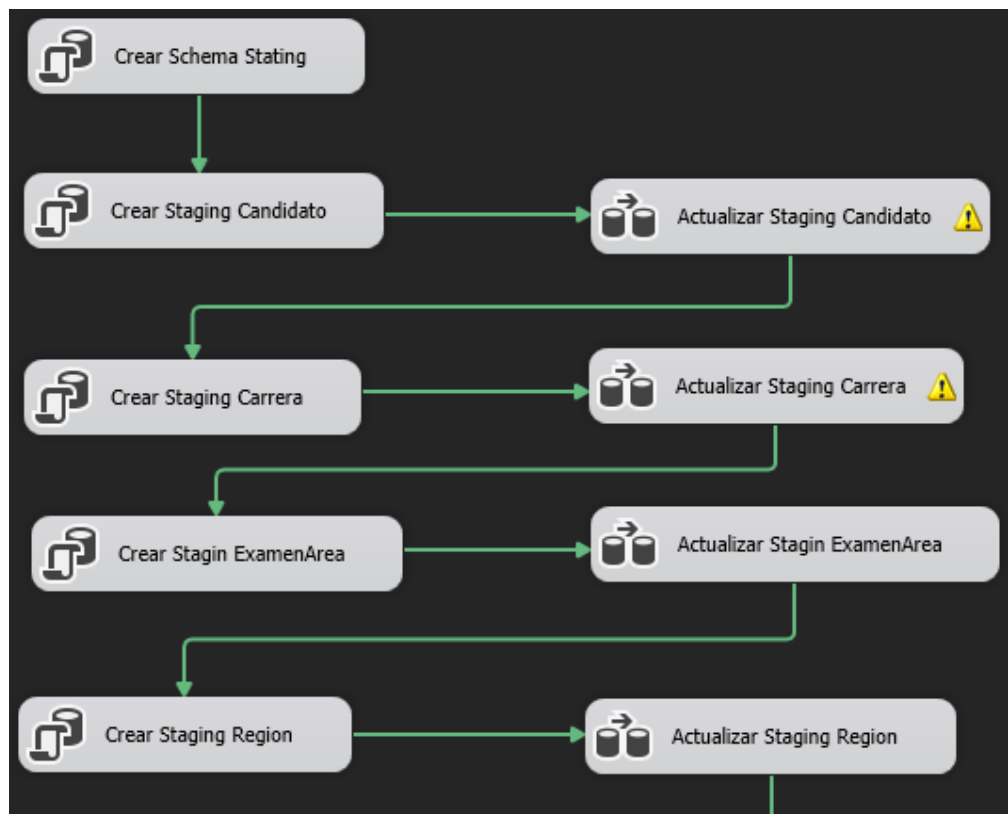
Se extraen la data de facultades y la carga de carreras que aspiran los candidatos



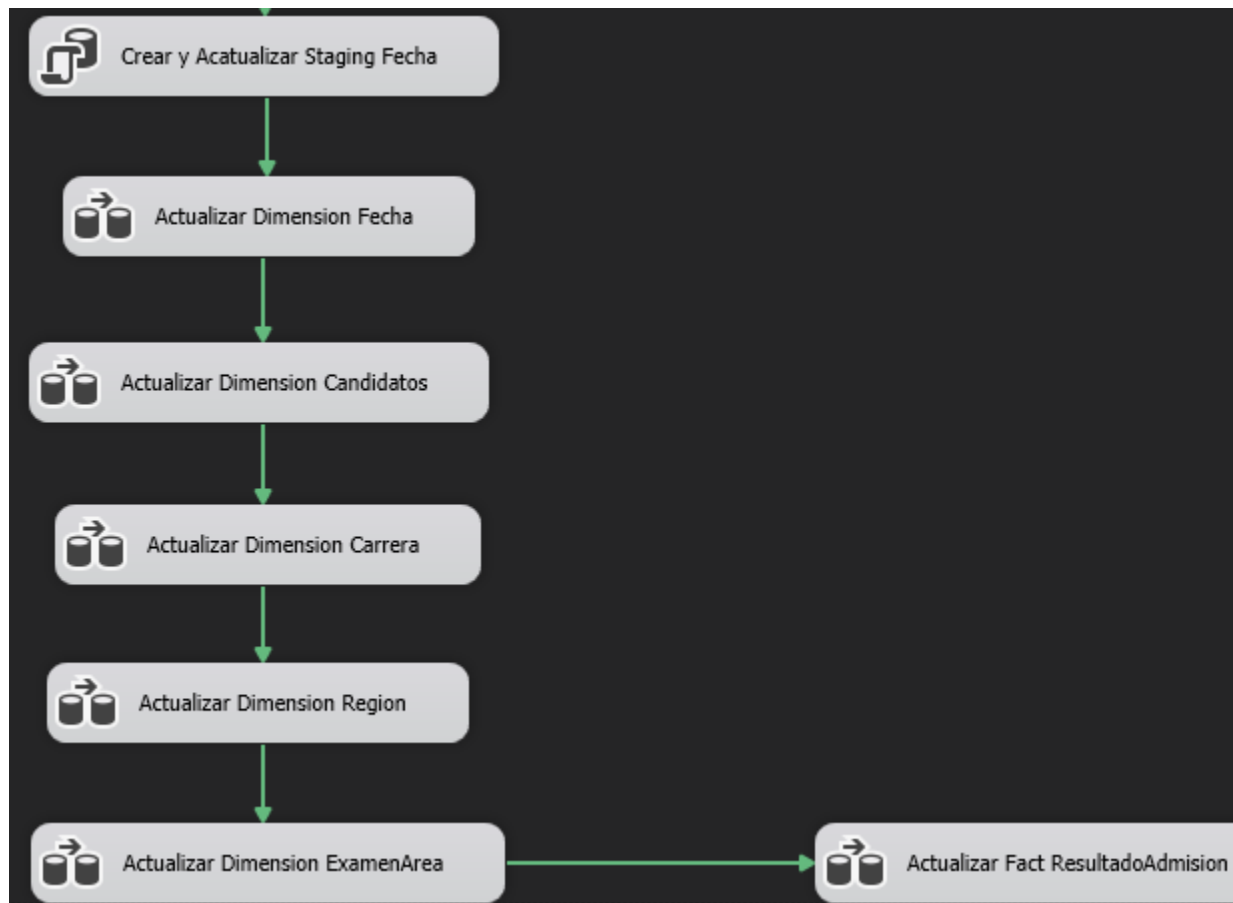
Por ultimo se cargan los resultados de los estudiantes en su respectivo ID por examen.



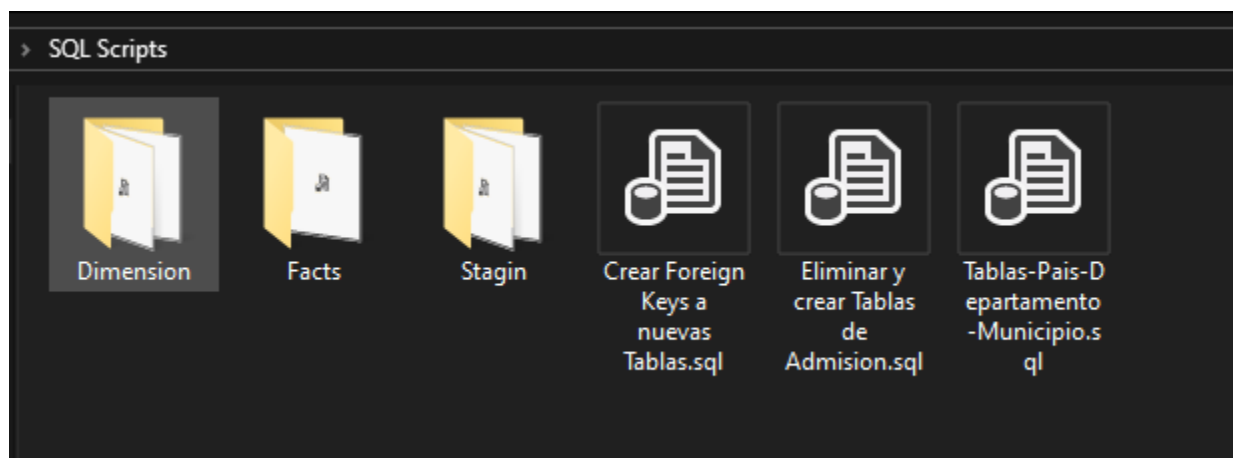
Luego se pasa a la carga de datos al modelo DataWarehouse, para esto se crean las tablas de Staging y procede a cargar los datos en cada una de ellas



Por último, se actualizan las fechas para poder obtener información adicional y se procede a la carga en de cada una de las dimensiones.

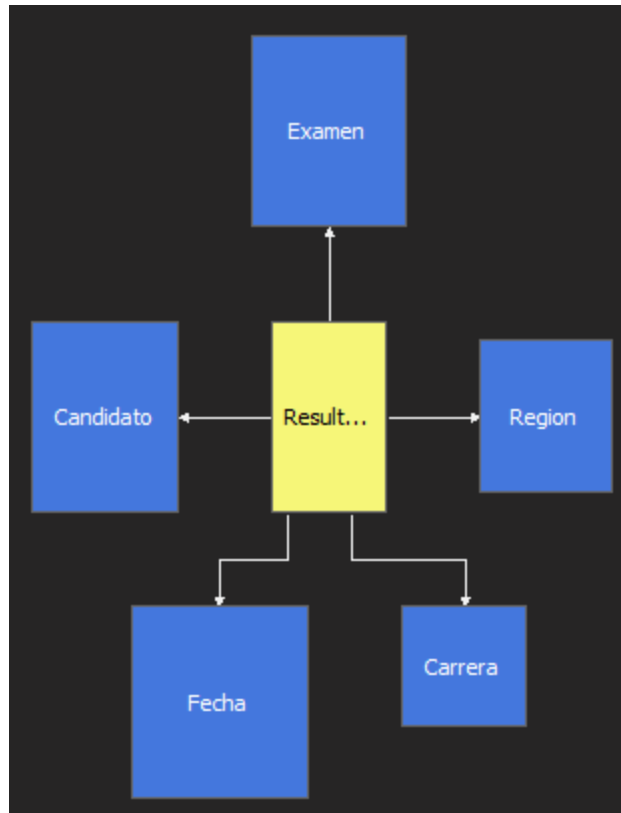


Si en caso usted desea acceder a los Scripts de todas las tablas en Staging, Dimension y Fact, puede ver los mismo en la carpeta *SQL Scripts*. Allí encontrará estos y otros archivos a su disposición para modificarlos:



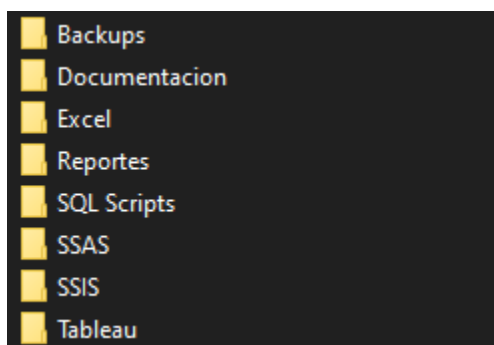
Paquete de análisis SSAS

Se utilizó un cubo MOLAP para tener análisis de múltiples dimensiones. Debido a que se crea una copia esto nos permite tomar la data que se desee analizar y trabajar con ella sin tener que recurrir a una búsqueda en las particiones de la base de datos. Al mismo tiempo nos permite tener mayor eficiencia al momento de realizar consultas grandes y que requieran de computo.



Reportes

Los archivos de reportes los encontrará disponibles en el directorio del proyecto:



La carpeta *Reportes* contiene los recursos para actualizar o agregar reportes. *Reportes* contiene los archivos que puede administrar en Tableau Desktop. Por tal, necesita el software Tableau Desktop y una licencia para su uso. Dentro de la carpeta *Reportes* encontrará los archivos:

- Reportes-Candidatos.twb
- Reportes.twb

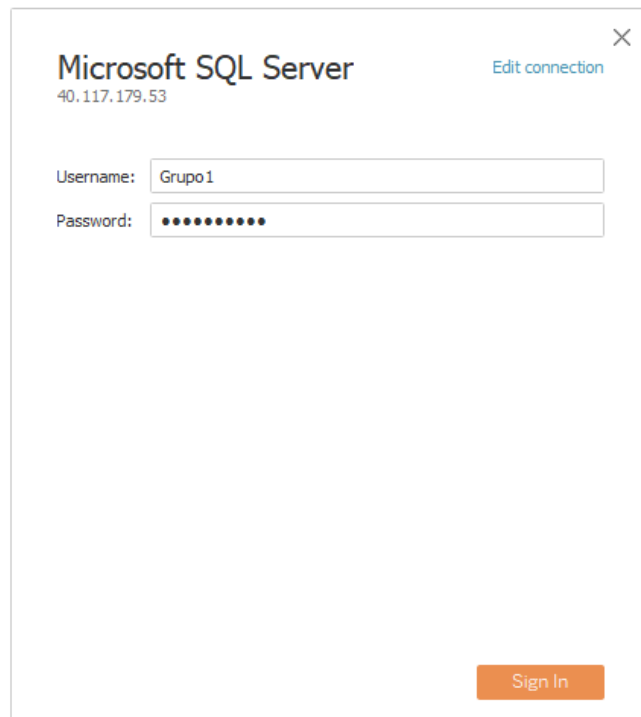
La lista de reportes incluidos es:

- Gráficas de cantidades de candidatos por
 - Año
 - Carrera
 - Diversificado
- Gráficas de mapas de candidatos por
 - País
 - Departamento
 - Municipio
- Gráficas de notas de evaluaciones de admisión por
 - Admitidos vs Condicionados vs No Admitidos
 - Departamento (Mapa con escala de colores)
 - Diversificado
 - Áreas de evaluación

Reportes-Candidatos.twb

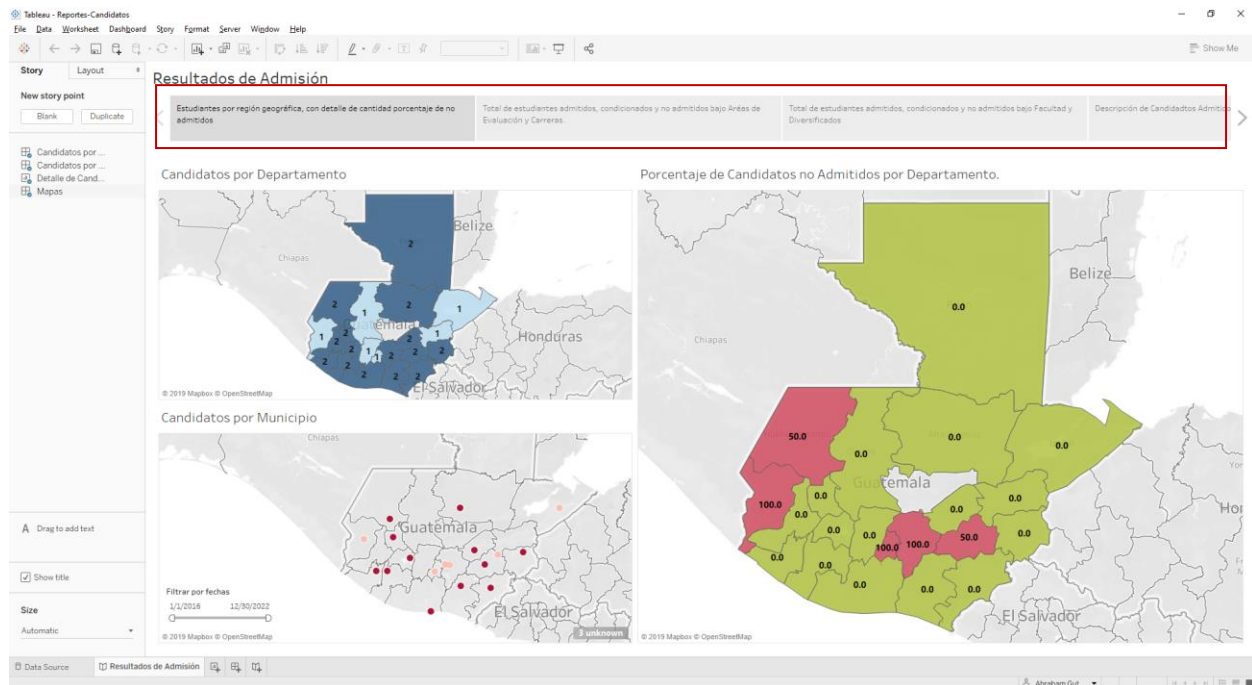
Contiene graficas de candidatos según su estado de aplicación: admitidos, no admitidos, condicionados. Los datos se presentan en diferentes formatos, incluyendo mapas geográficos, gráficos de barras; se incluye un filtro por fecha que actualiza el reporte en tiempo real.

Abra el archivo *Reporte-Candidatos.twb*, se le solicitará un usuario y contraseña para acceder a los datos que alimenten el reporte. Ingrese las credenciales provistas por su departamento de IT:



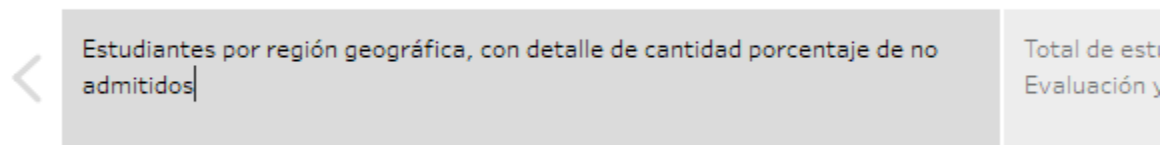
A screenshot of a Microsoft SQL Server connection dialog box. The title bar says "Microsoft SQL Server" with a close button (X) in the top right corner. Below the title, the IP address "40.117.179.53" is displayed. To the right of the IP address is a link that says "Edit connection". Below this, there are two input fields: "Username:" with the text "Grupo1" and "Password:" with a masked password represented by ten dots. At the bottom right, there is an orange button labeled "Sign In".

La primera vista será un Story de Tableau. Este agrupa varios gráficos en una misma vista y agrega una breve descripción:

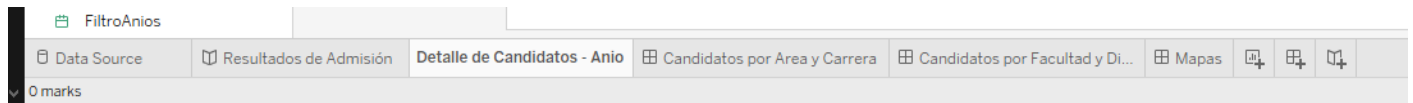
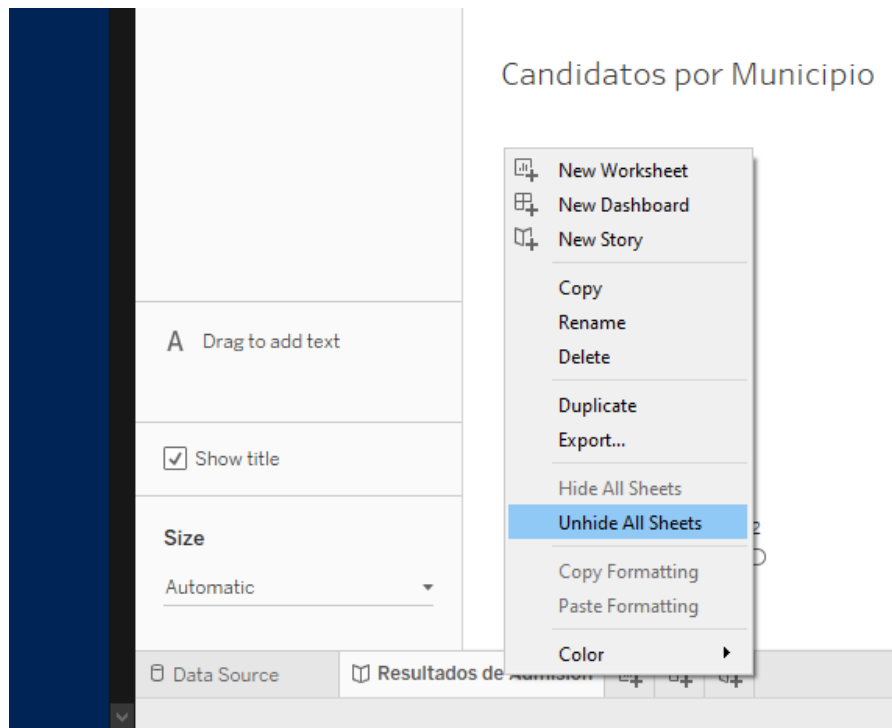


Puede editar esta descripción dando doble clic sobre la misma:

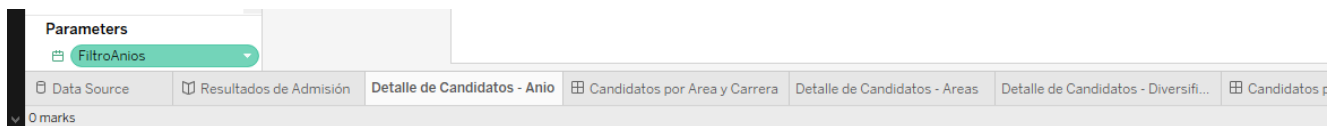
Resultados de Admisión



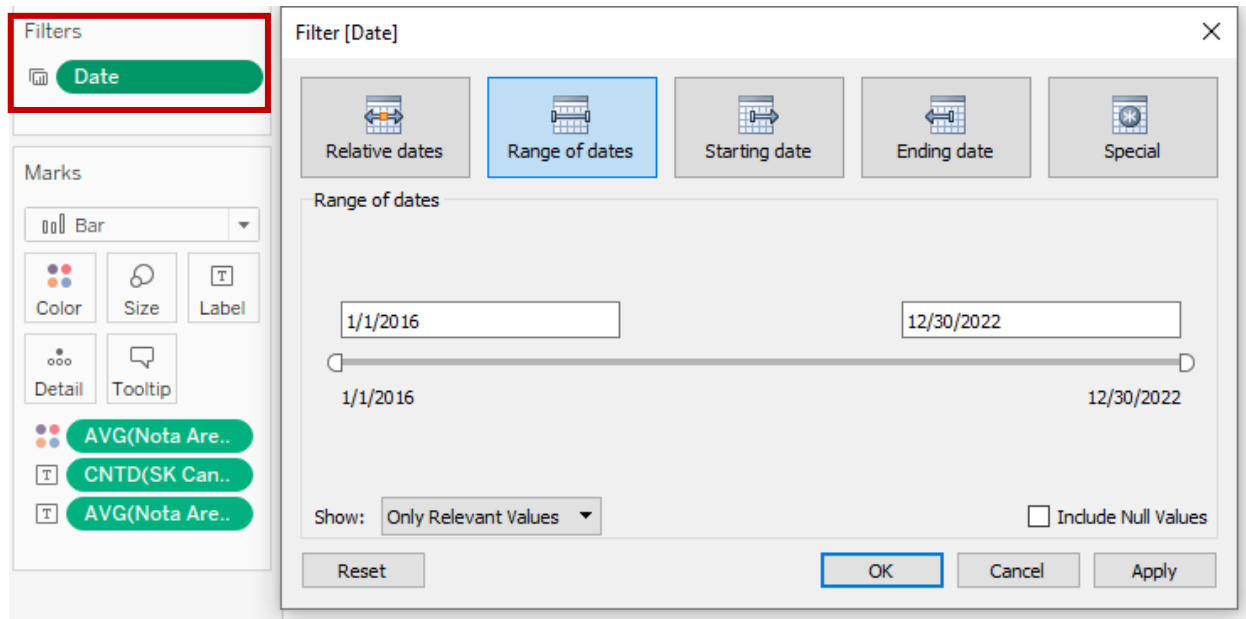
En la parte inferior vera la pestaña del Story, para mostrar las demás secciones del reporte presione clic derecho y seleccione *Unhide All Sheets*



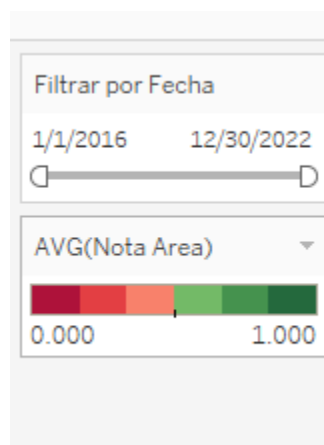
Ahora podrá ver todas los dashboards. Repita el paso anterior sobre cada dashboard para ver los reportes detalladamente y modificar sus propiedades. Deberá observar el siguiente formato:



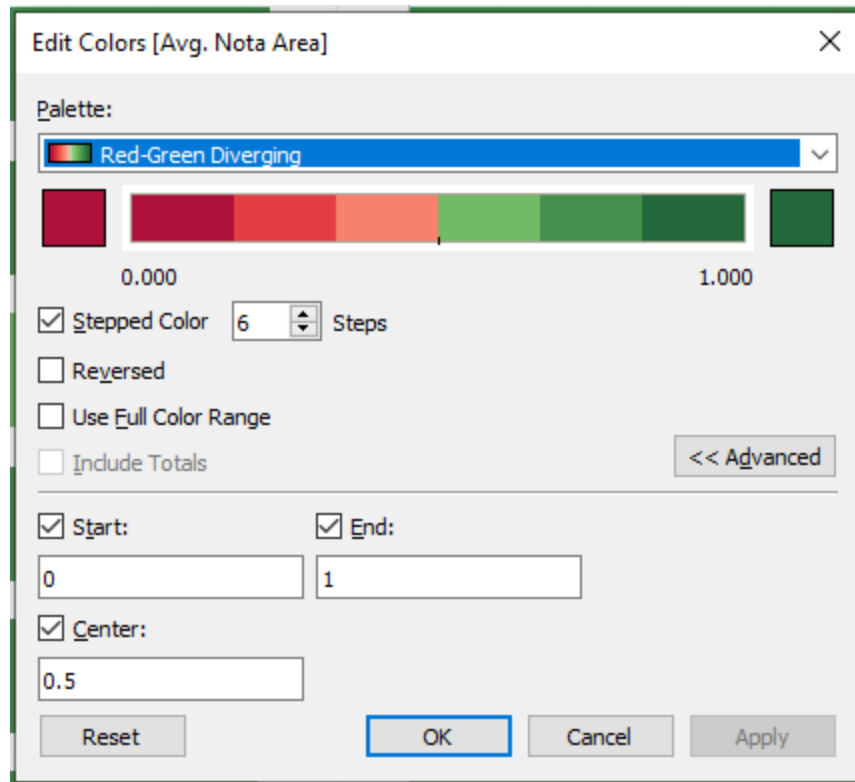
Todos los reportes cuentan con un filtro por fecha en la sección de filtros. Al dar doble clic puede modificar este filtro para ajustar los rangos y valores admitidos:



A la derecha de su pantalla vera otras tarjetas para manejar o administrar los filtros. En este reporte debe ver al menos la tarjeta de Fecha. Otras tarjetas disponibles son Colores.

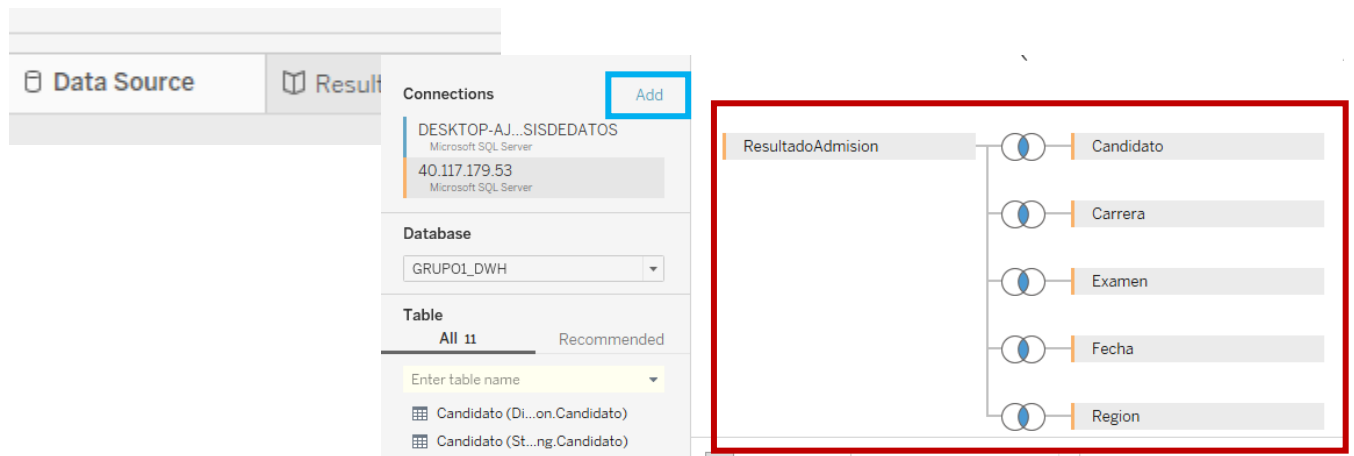


Los colores puede modificarlos a su conveniencia. La escala puede estar expresada en valor porcentuales (0 a 1) o en valor absolutos (valor mínimo a máximo). De doble clic sobre la tarjeta para modificar las propiedades:



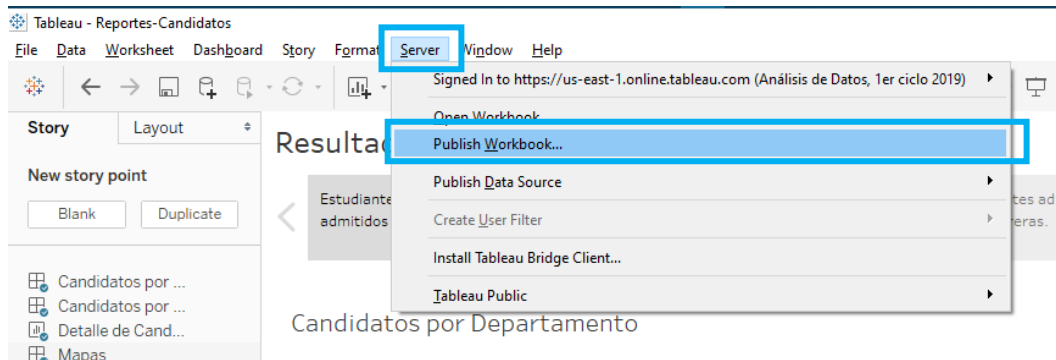
- Palette: indica la escala de color elegida. Puede elegir una predefinida de este campo o indicar una personalizada dando doble clic sobre cualquiera de los dos cuadros grandes Rojo y Verde.
- Stepped Color indica la forma en que se dividirá la escala de colores y cuantos se mostrarán en la gráfica.
- Reversed puede utilizarse para invertir la escala de colores.
- Advanced contiene modificadores a la escala, de forma que los colores se ajustan a este. Para el ejemplo, el rojo es el punto Start y cualquier valor igual a 0 (para el ejemplo) se mostrará como un color Rojo en la gráfica, y mientras se acerca a 1 cambiará a verde. Center determina como Stepped Color se distribuye entre los valores.

En la parte inferior izquierda encontrará la sección DataSource. Esta le permite indicar o modificar la fuente de datos.

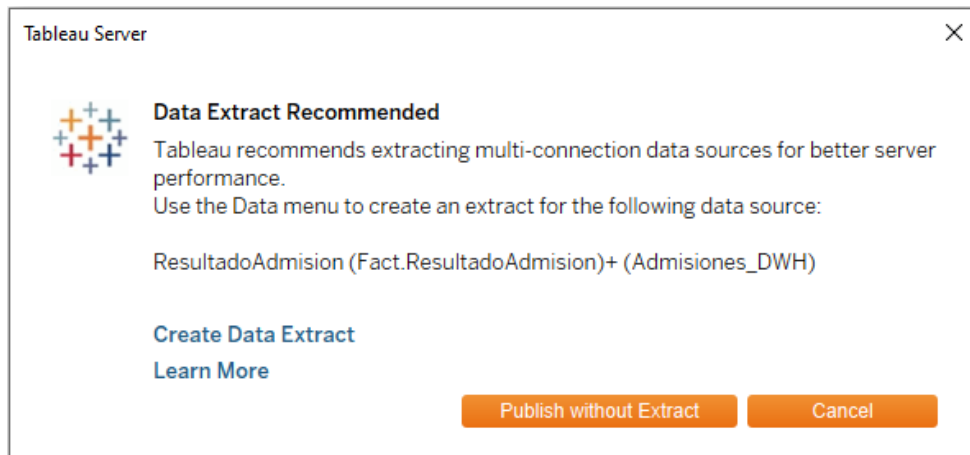


Es probable que usted no pueda ver ambas conexiones o que estas no sean iguales. Lo importante es que la parte derecha de la imagen, encerrada en un cuadro rojo, permanezca igual, pues esta es la fuente de datos que se necesita para generar los reportes.

Si tiene cambios que desea actualizar al servidor, entonces deberá buscar el menú *Server*



Si se le solicita responder la siguiente ventana, indique *Public without Extract*.



En la ventana *Publish Workbook*, indique todos los campos según su configuración en Tableau Online. Recuerde, al indicar el nombre, si otro proyecto ya fue nombrado igual este se sobrescribirá con la nueva información. Finalmente, de clic en *Publish*

A dialog box titled 'Publish Workbook to Tableau Online'. It contains several sections: 'Project' with a dropdown menu showing 'GRUPO1'; 'Name' with a dropdown menu showing 'Resultados de Admisión' and a red warning message below it: 'Workbook name is already in use. Publishing will overwrite the existing workbook.'; 'Description' with a text box containing 'Descripción de los resultados de admisión'; 'Tags' with an 'Add' link; 'Sheets' with 'All' and an 'Edit' link; 'Permissions' with 'Set to existing workbook default' and an 'Edit' link; 'Data Sources' with '1 embedded in workbook' and an 'Edit' link; and 'More Options' with two checked checkboxes: 'Show sheets as tabs' and 'Show selections'. A green 'Publish' button is at the bottom right.

Detalles técnicos

Todas las funcionalidades se desarrollaron utilizando el siguiente software:

- Bases de Datos:
 - SQL Server 2017
 - SQL Server Management Studio 18
- Esquema multidimensional
 - SQL Server Integration Services
 - SQL Server Analysis Server
 - Visual Studio 2017
- Servidor
 - Internet Information Services 10.0 Express
- Reportes
 - Tableau Software
 - Tableau Desktop
 - Tableau Online