

Data Analytics
Informe Taller Práctico Data Analytics

PRESENTADO A: Ing. Luis Ernesto Rodriguez Olaya

PRESENTADO POR: Germán Homero Morán.

Septiembre.20 del 2021

Contenido

1.	REQUERIMIENTOS FUNCIONALES	3
1.1	RESTAURACION DEL BACKUP	3
1.2	PROCESO DEL TALLER	3
2.	ELABORACION ETLs VISUAL STUDIO 2019	4
3.	BASE DE DATOS SQL SERVER	6
4.	DISEÑO DEL TABLERO EN POWER BI	8
5.	INSTALACIONES Y HERRAMIENTAS	13
6.	CONCLUSIONES	14

TABLA DE FIGURAS

Figura 1:	Diagrama de flujo creación de ETLs	4
Figura 2:	Bloque de código para obtener la edad del Alumno.....	5
Figura 3:	Configuración del bloque columna derivada para el cálculo de la edad	5
Figura 4:	Configuración de las relaciones tabla de destino fctMatricula	6
Figura 5:	Restauración BD DWHMEN	7
Figura 6:	Registros almacenados en la tabla MaestroEstudiantes.....	7
Figura 7:	Carga de datos en la tabla de hecho fctMatricula	8
Figura 8:	Conexión con el servidor de la BD SQL Server en Power BI	9
Figura 9:	Selección de la tabla de hechos fctMatricula con todas la relaciones	9
Figura 10:	Relaciones entre la tabla hechos y dimensiones.....	10
Figura 11:	Dimensiones en Power BI	11
Figura 12:	Interfaz gráfica para el diseño de tablero en Power BI	11
Figura 13:	Tablero principal Power BI	12

1. REQUERIMIENTOS FUNCIONALES

En este ejercicio, aunque es muy sencillo, les servirá para aprender a observar y analizar los datos que se tienen, las ETL's que usted construirá las cuales son muy sencillas.

1.1 RESTAURACION DEL BACKUP

- Restaurar el Backup adjunto (DWHMEN_V2.bak). Se adjunta quía para restaurar el backup "Proceso restauración Backup.docx". Esta restauración se hará en la Base de datos creada por ustedes en su equipo. La base de datos que se restaurara se llama "DWHMEN".

1.2 PROCESO DEL TALLER

- Después de la restauración de la Base de datos, se debe llenar la tabla "fctMatricula" (tabla vacía) que reposa en la base de datos "DWHMEN" (Base restaurada). La tabla origen es "MaestroEstudiantes", la cual debe ser leída y en la secuencia del flujo debe buscar las llaves o SK, para luego ser insertado cada registro en la tabla "fctMatricula" (Tabla de hechos).
- Construir las respectivas ETLs para llenar la tabla de hechos "fctMatricula".
- Una vez llenada la tabla de hechos "fctMatricula", deberán construir un reporte en power BI que responda las siguientes preguntas de negocio:
 - Cantidad de Instituciones
 - Cantidad de sedes
 - Cantidad de Estudiantes o alumnos
 - Cantidad de instituciones por municipio
 - Cantidad de instituciones por tipo
 - Cantidad y porcentaje de participación alumnos por género
 - Cantidad de Alumnos por género y municipio
 - Cantidad y porcentaje de participación de alumnos matriculados por tipo de institución
 - Cantidad de alumnos matriculados por tipo de institución y género
- El reporte debe permitir filtrar la información por la dimensión de tiempo.

- Finalizada el ejercicio por favor enviar en un archivo .zip con sus nombres y apellidos (Eje: BI_nombre1_nombre2_apellido1_apellido_2.zip) la siguiente información:
 - Backup de la base “DWHMEN”
 - Proyecto de las ETLs
 - Reporte Power BI
- En caso de no poderse adjuntar el archivo del Backup, por el Sitio de Wow, favor enviarlo adjunto al correo luisr@bextsa.com, junto con los otros documentos.

2. ELABORACION ETLs VISUAL STUDIO 2019

Inicialmente se crea el diagrama de flujo para cargar los datos en la tabla de hechos “**factMatricula**”, En la Figura 1 se muestra la estructura del diagrama de flujo diseñado, se realizaron algunas conversiones de datos, especialmente con el tamaño de las variables.

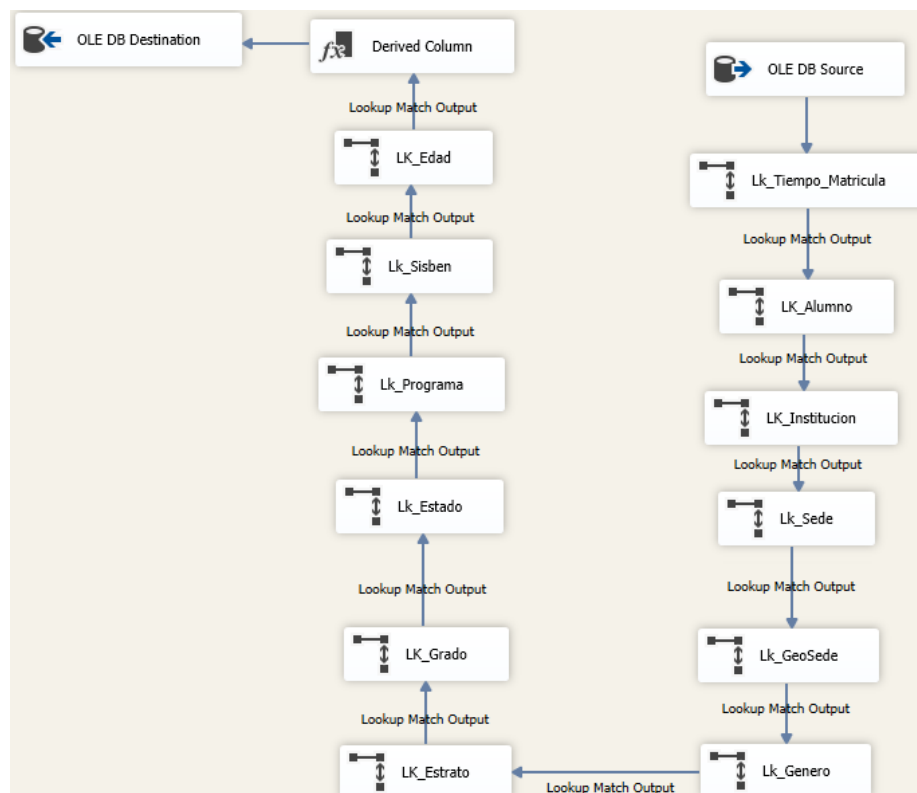


Figura 1: Diagrama de flujo creación de ETLs

También se adiciono una **columna derivada** para calcular la edad de los empleados. El bloque **Lk-Edad** permite extraer la fecha de nacimiento de cada uno de los alumnos registrados en la tabla MaestroEstudiantes, en este caso para hacer la consulta se relaciona el **idAlumno** de la tabla “MaestroEstudiantes” con el **idAlumno** de la tabla **DWH. DimAlumnos** tal como se muestra en el bloque de código de la Figura 2. De esta manera es posible relacionar ambas tablas y extraer la fecha de nacimiento de cada Alumno Matriculado.

```
SELECT
    skAlumno, fechaNacimiento

FROM [DWH].[MaestroEstudiantes] me
INNER JOIN [DWH].[DimAlumnos] da ON da.idAlumno= me.idAlumno
```

Figura 2: Bloque de código para obtener la edad del Alumno

La configuración de la **columna derivada** para el cálculo de la edad de cada registro se muestra en la Figura 3

Especifique las expresiones utilizadas para crear nuevos valores de columna e indique si los valores actualizan las columnas existentes o bien llenan columnas nuevas.

Variables y parámetros

Columnas

fechaNacimiento

idAlumno

numeroDocumento

codDepartamento

codMunicipio

codPais

idInstitucion

idSede

fecha

Idgenero

Funciones matemáticas

Funciones de cadena

Funciones de fecha y hora

Funciones NULL

Conversiones de tipo

Operadores

Descripción:

Nombre de columna d...	Columna derivada	Expresión	Tipo de datos	
DC_Edad	<agregar como colum...	YEAR(GETDATE()) - YEAR(fechaNacimiento)	four-byte signed integ...	

Figura 3: Configuración del bloque columna derivada para el cálculo de la edad

Finalmente una vez creado todos los bloques LK_Matricula, Lk_Intitucion, Lk_Sede, LK_Programa, etc, se realiza la conexión con la tabla “fctMatricula”, mediante el bloque “OLE DB DESTINATION” la configuración de este bloque se muestra en la Figura 4, donde se asignan cada una de las variables a su correspondiente valor.

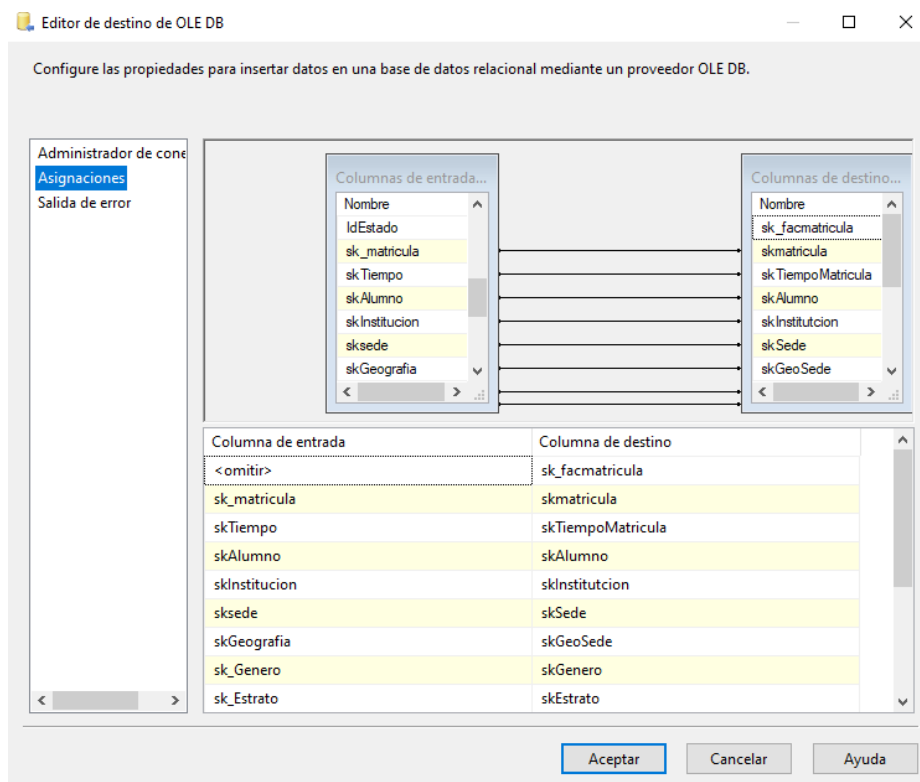


Figura 4: Configuración de las relaciones tabla de destino fctMatricula

3. BASE DE DATOS SQL SERVER

Para este proceso se debió previamente haber instalado SQL Server con todas las herramientas necesarias, además teniendo en cuenta la guía práctica de **restauración del Backup** se hace la carga de la base de datos. En la Figura 5 se muestra que la restauración del **backup** se realizó de manera correcta, ya que se dispone de la base de datos **DWMEN**.

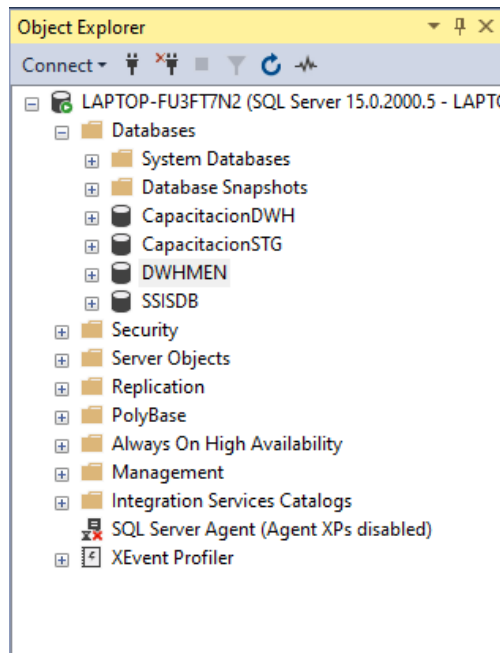


Figura 5: Restauración BD DWHMEN

Además, en la Figura 6 se puede observar que todas las dimensiones se restauraron correctamente, en este caso se muestra la dimensión **DWH.MaestroEstudiantes** donde se observa todos los registros que tiene almacenado esa tabla .

The screenshot shows a SQL query executed in the 'SQLQuery8.sql' window. The query selects all columns from the 'MaestroEstudiantes' table in the 'DWH' database. The results are displayed in a table with 11 columns and 11 rows.

sk_matricula	idAlumno	numeroDocumento	codDepartamento	codMunicipio	codPais	idInstitucion	idSede	fecha	Idgenero	IdEstrato	IdGrado	codPrograma	IdEstado
1	7010389	1128432113	76	76001	57	1828	1828	2014-01-01	56	999	999	52059	999
2	142488	1095802889	11	11001	57	9110	9110	2013-01-01	57	999	999	91157	999
3	5607787	80902196	11	11001	57	1101	1101	2012-07-01	56	999	999	31	999
4	142488	1095802889	68	68001	57	3201	3201	2012-07-01	57	999	999	9350	999
5	999240	92013126722	23	23001	57	1113	1113	2011-07-01	56	999	999	4863	999
6	1865966	1015413292	11	11001	57	1301	1301	2014-01-01	56	999	999	102133	999
7	1868918	1032399957	11	11001	57	1806	1806	2011-07-01	56	999	999	10550	999
8	1868918	1032399957	11	11001	57	1806	1806	2016-01-01	56	999	999	54091	999
9	1875714	80828563	11	11001	57	1105	1105	2009-07-01	56	999	999	166	999
10	1875714	80828563	11	11001	57	1101	1101	2016-01-01	56	999	999	20	999
11	1875714	80828563	11	11001	57	1101	1101	2012-01-01	56	999	999	20	999

Figura 6: Registros almacenados en la tabla MaestroEstudiantes

Después de haber restaurado el **backup** se realiza la carga de datos en la tabla **fctMatricula** teniendo en cuenta la tabla **MaestroEstudiantes**. Para este proceso de carga de datos se ejecuta el diagrama de flujo previamente diseñado y el resultado obtenido se muestra en la Figura 7.

SQLQuery53.sql -...FT7N2\German (65))

```

SELECT [sk_facmatricula]
,[skmatricula]
,[skTiempoMatricula]
,[skAlumno]
,[skInstitucion]
,[skSede]
,[skGeoSede]
,[skGenero]
,[skEstrato]
,[skSiaben]
,[skGrado]
,[Edad]
,[skPrograma]
,[skEstado]
FROM [DimMien].[DimH].[fctMatricula]

```

Results

	sk_facmatricula	skmatricula	skTiempoMatricula	skAlumno	skInstitucion	skSede	skGeoSede	skGenero	skEstrato	skSiaben	skGrado	Edad	skPrograma	skEstado
1	1	1	20140101	23559	32587	1892	1007	75	17	21	69	31	11411	91
2	2	2	20130101	23471	32497	5492	149	76	17	21	69	32	1203	91
3	3	3	20120701	23537	1115	64773	149	75	17	21	69	36	356	91
4	4	4	20120701	23471	2036	2115	847	76	17	21	69	32	14951	91
5	5	5	20110701	23474	1127	64906	429	75	17	21	69	29	1419	91
6	6	6	20140101	23484	32628	64935	149	75	17	21	69	32	5006	91
7	7	7	20110701	23485	32546	1088	149	75	17	21	69	34	9795	91
8	8	8	20160101	23485	32546	1088	149	75	17	21	69	34	12522	91
9	9	9	20090701	23488	1119	64878	149	75	17	21	69	37	1026	91
10	10	10	20160101	23488	1115	64773	149	75	17	21	69	37	469	91
11	11	11	20120101	23488	1115	64773	149	75	17	21	69	37	469	91

Query executed successfully. LAPTOP-FU3FT7N2 (15.0 RTM) LAPTOP-FU3FT7N2\German... DWHMEN 00:00:00 7,156 rows

Figura 7: Carga de datos en la tabla de hecho fctMatricula

En este caso se observa que la tabla “**facMatricula**” ya tiene datos y corresponde a cada una de las claves que se relacionan con las dimensiones.

4. DISEÑO DEL TABLERO EN POWER BI

Una vez añadida la información a la tabla de hechos “**fctMatricula**” y configuradas todas las relaciones existentes entre cada una de las dimensiones se procede a realizar el diseño del tablero en la herramienta POWER BI.

Inicialmente se realiza a conexión con el servidor para obtener la base de datos donde reposan las tabla de hecho y dimensiones. En la Figura 8 se muestra la conexión con el servidor.

Base de datos SQL Server

Servidor ①

LAPTOP-FU3FT7N2

Base de datos (opcional)

Modo Conectividad de datos ①

☒ Importar

☐ DirectQuery

► Opciones avanzadas

Aceptar

Cancelar

Figura 8: Conexión con el servidor de la BD SQL Server en Power BI

Posteriormente se debe seleccionar la base de datos **DWMAN** y la tabla de hechos con todas las tablas relacionadas para realizar la carga de información como se muestra en la Figura9.

Navegador

Opciones de presentación		DWH.fctMatricula					
▲	LAPTOP-FU3FT7N2 [4]	skTiempoMatricula	skAlumno	skInstitucion	skSede	skGeoSede	skC
▶	CapacitacionDWH	20140101	23559	32587	1892	1007	
▶	CapacitacionSTG	20130101	23471	32497	5492	149	
▲	DWHMEN [21]	20120701	23537	1115	64773	149	
<input type="checkbox"/>	DEPARTAMENTOS	20120701	23471	2036	2115	847	
<input checked="" type="checkbox"/>	DIM.Estado	20110701	23474	1127	64906	429	
<input checked="" type="checkbox"/>	DIM.Estrato	20140101	23484	32628	64935	149	
<input checked="" type="checkbox"/>	DIM.Genero	20110701	23485	32546	1088	149	
<input checked="" type="checkbox"/>	DIM.Grado	20160101	23485	32546	1088	149	
<input type="checkbox"/>	DIM.MaestroPruebaEstudiantes	20090701	23488	1119	64878	149	
<input type="checkbox"/>	DIM.prueba	20160101	23488	1115	64773	149	
<input type="checkbox"/>	DIM.Sisben	20120101	23488	1115	64773	149	
<input type="checkbox"/>	DOCENTES	20140701	23513	1124	64790	362	
<input checked="" type="checkbox"/>	DWH.DimAlumnos	20090701	23527	672	64699	1007	
<input checked="" type="checkbox"/>	DWH.DimGeografia	20070101	23527	672	64699	1007	
<input type="checkbox"/>	DWH.DimHomologa	20100101	23532	18259	1528	149	
<input checked="" type="checkbox"/>	DWH.DimInstitucion	20130701	23470	672	64699	1007	
<input checked="" type="checkbox"/>	DWH.DimProgramas	20100101	23470	672	64699	1007	
<input checked="" type="checkbox"/>	DWH.DimSede	20130701	23476	32497	5492	149	
<input checked="" type="checkbox"/>	DWH.DimTiempo	20110101	23480	1621	35	149	
		20130701	23480	1621	35	149	
		20100701	23482	1132	1266	149	
		20130101	23482	32497	5492	149	
		20110701	23495	32587	1892	1007	

Figura 9: Selección de la tabla de hechos fctMatricula con todas la relaciones

Una vez se realiza la carga de la información se puede observar las distintas relaciones de cada una de las dimensiones con la tabla de hechos como se muestra en la figura 10.

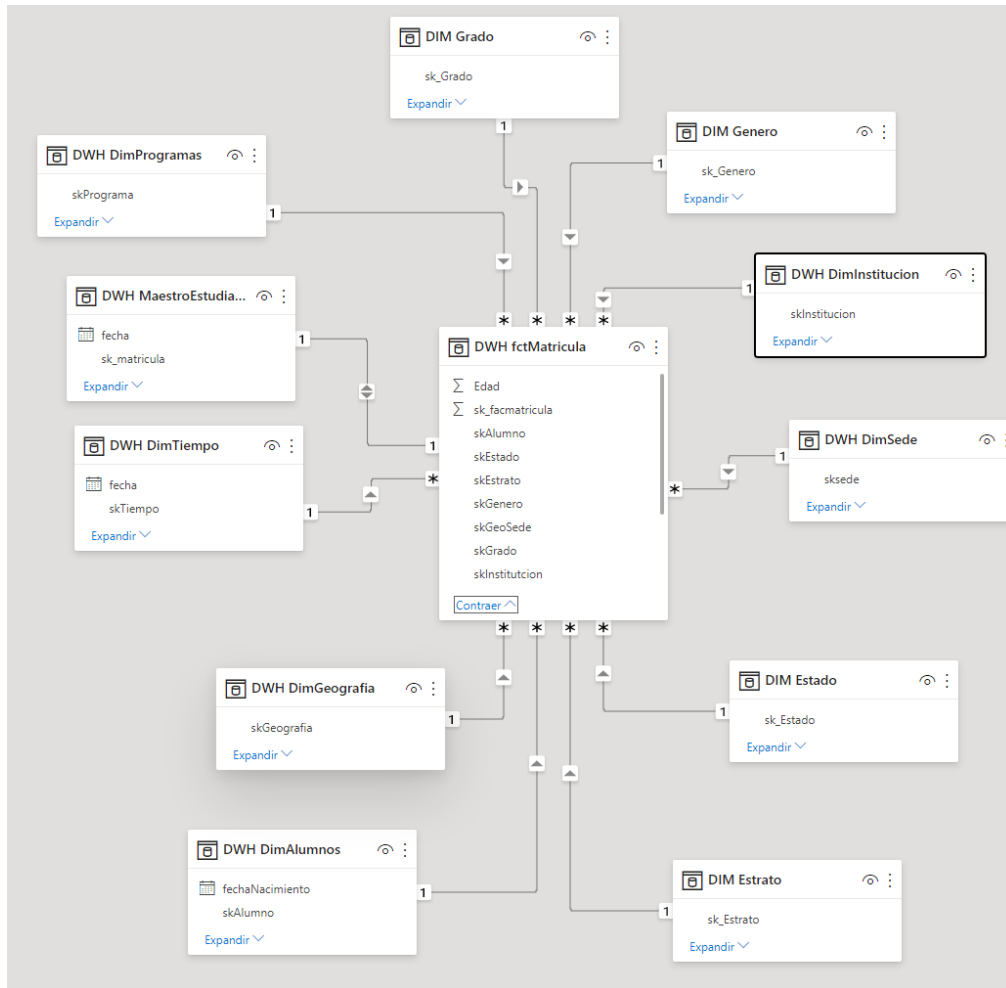


Figura 10: Relaciones entre la tabla hechos y dimensiones

En power BI también se puede observar cada una de las dimensiones y realizar operaciones con las diferentes columnas. Algunas de las operaciones que se pueden realizar son: sumas, restas, promedios, porcentajes, etc. En la Figura 11 se muestra cada una de las dimensiones en Power BI, en este caso se muestra la tabla de hechos **fctMatricula** donde se observa las columnas de las llaves foráneas que relacionan con otras dimensiones.

Archivo Inicio Ayuda Herramientas de tablas

Nombre: DWH fctMatricula

Marcar como tabla de fechas

Administrador relaciones

Nueva Medida Nueva medida rápida Nueva columna Nueva tabla

Estructura

Matricula	skmatricula	skTiempoMatricula	skAlumno	skInstitucion	skSede	skGeoSede	skGenero	skEstrato	skSisen	skGrado	Edad	skPrograma	skEstado
29	29	20080101	23531	32628	64935	149	75	17	21	69	38	8363	91
61	61	20100101	23557	32628	64935	149	76	17	21	69	35	8363	91
213	213	20130701	23531	32628	64935	149	75	17	21	69	38	8363	91
288	288	20110701	23475	32628	64935	149	76	17	21	69	37	8363	91
290	290	20100101	23475	32628	64935	149	76	17	21	69	37	8363	91
293	293	20070701	23477	32628	64935	149	75	17	21	69	33	8363	91
294	294	20090101	23477	32628	64935	149	75	17	21	69	33	8363	91
508	508	20090701	23557	32628	64935	149	76	17	21	69	35	8363	91
510	510	20130101	23557	32628	64935	149	76	17	21	69	35	8363	91
578	578	20120101	23477	32628	64935	149	75	17	21	69	33	8363	91
579	579	20120701	23477	32628	64935	149	75	17	21	69	33	8363	91
581	581	20130701	23477	32628	64935	149	75	17	21	69	33	8363	91
583	583	20110701	23477	32628	64935	149	75	17	21	69	33	8363	91
863	863	20090701	23531	32628	64935	149	75	17	21	69	38	8363	91
865	865	20080701	23531	32628	64935	149	75	17	21	69	38	8363	91
866	866	20100101	23531	32628	64935	149	75	17	21	69	38	8363	91
997	997	20070701	23557	32628	64935	149	76	17	21	69	35	8363	91
998	998	20110701	23557	32628	64935	149	76	17	21	69	35	8363	91
999	999	20070701	23557	32628	64935	149	76	17	21	69	35	8363	91
1000	1000	20080701	23557	32628	64935	149	76	17	21	69	35	8363	91
1001	1001	20090101	23557	32628	64935	149	76	17	21	69	35	8363	91
1078	1078	20100101	23437	32628	64935	149	76	17	21	69	33	8363	91

Campos

Buscar

- DIM Estado
- DIM Estrato
- DIM Genero
- DIM Grado
- DWH DimAlumnos
- DWH DimGeografia
- DWH DimInstitucion
- DWH DimProgramas
- DWH DimSede
- DWH DimTiempo
- DWH fctMatricula
- DWH MaestroEstudiant...

Tabla: DWH fctMatricula (7,156 filas)

Figura 11: Dimensiones en Power BI

Finalmente se realiza el diseño del tablero, teniendo en cuenta cada una de las visualizaciones y campos específicos. La interfaz gráfica en Power BI para diseñar el tablero se muestra en la Figura 12

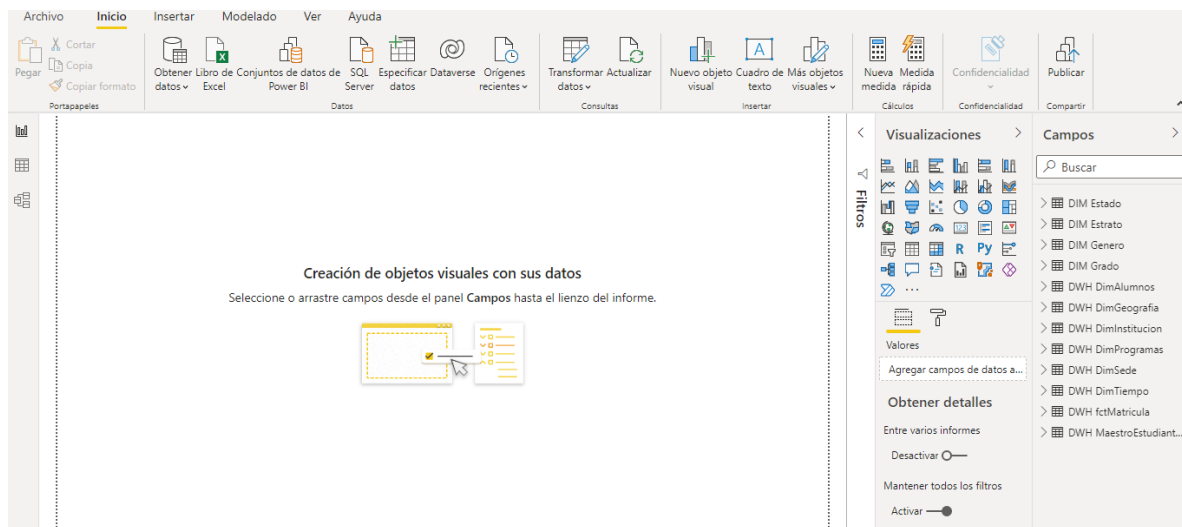


Figura 12: Interfaz gráfica para el diseño de tablero en Power BI

El diseño Final del tablero se muestra en la Figura 13 donde inicialmente se observa la ventana principal **Main_page**, en este caso se tienen diferentes gráficos y visualizaciones que permiten dar respuesta a cada una de las preguntas planteadas al inicio del documento.

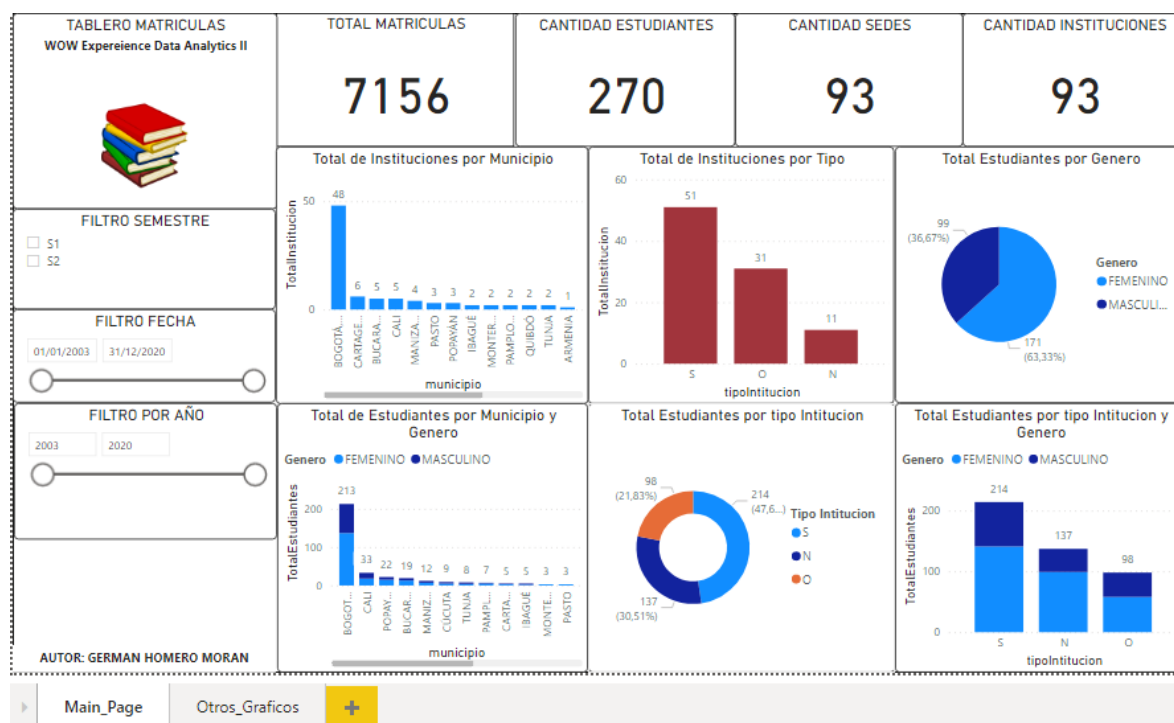


Figura 13: Tablero principal Power BI

Este tablero muestra información relevante de las matrículas, teniendo en cuenta diferentes aspectos como: número de estudiantes, instituciones, sedes, municipios, entre otros. Por medio de las visualizaciones y los respectivos filtros se puede acceder a la información de manera granular.

Los principales filtros de este tablero son:

- Filtro por semestre: S1 y S2
- Filtro por fecha de matrícula:
- Filtro por años de matrícula
- Adicionalmente es posible filtrar por género, tipo de institución y municipios.

En base a este tablero es posible se responde las preguntas definidas anteriormente:

- Cantidad de Instituciones: **93**
- Cantidad de sedes: **93**
- Cantidad de Estudiantes o alumnos: **270**
- Cantidad de instituciones por municipio: **Se analiza según el tablero**

- **Bogota: 48**
- **Cartagena: 6**
- **Bucaramanga:5**
- Cantidad de instituciones por tipo: **Se analiza según el tablero**
 - **Tipo O: 51**
 - **Tipo S: 31**
 - **Tipo N: 11**
- Cantidad y porcentaje de participación alumnos por género: **Se analiza según el tablero**
 - **Maculino: 99 equivale a 36.97 % del total**
 - **Femenino: 171 equivale 63.335 del total**
- Cantidad de Alumnos por género y municipio: **El análisis se realiza según el tablero debido a que son múltiples municipios**
- Cantidad y porcentaje de participación de alumnos matriculados por tipo de institución: **Según la gráfica se tiene**
 - **Tipo s: 214 equivale a 47.6**
 - **Tipo O: 98 equivale a 21.83**
 - **Tipo N: 137 equivale a 30.5**
- Cantidad de alumnos matriculados por tipo de institución y género: **El análisis se realiza según el tablero debido a que son múltiples municipios.**

5. INSTALACIONES Y HERRAMIENTAS

Para el correcto desarrollo de esta práctica se instaló cada uno de las siguientes herramientas:

- Instalar sql server 2019
URL:<https://www.microsoft.com/es-es/sql-server/sql-server-downloads> (Modo desarrollador)
- Instalar SQL Server Management Studio (SSMS)

URL:<https://docs.microsoft.com/es-mx/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver15>

- Instalar Visual Studio 2019 y SQL Server Data Tools (SSDT)

URL:<https://docs.microsoft.com/es-mx/sql/ssdt/download-sql-server-data-tools-ssdt?view=sql-server-ver15>

En esta instalación se debe tener en cuenta que adicionalmente se debe descargar el paquete *integration services y Analysis services* , posteriormente se deben agregar esas extensiones Visual estudio para poder realizar la practica .

- Instalar Power BI desktop

URL: <https://powerbi.microsoft.com/es-es/downloads/>

Se descargar la última versión de Power Bi desktop.

Adicionalmente se debe tener en cuenta los recursos suministrados por el profesor

- Grabaciones de clase
- Documento restauración de backup
- Documento proceso backup

6. CONCLUSIONES

- Power BI es una herramienta muy intuitiva y facilita en gran medida el análisis de grandes volúmenes de información, además existe gran documentación lo que facilita el desarrollo de tableros.
- Las herramientas SQL server y visual estudio 2019 mediante la integración de **integration services** facilitan en gran medida los procesos de ETL (Extracción, Transformación y Carga) de datos. Esta interfaz gráfica basada en componentes hace que la limpieza y transformación de los datos sea más sencilla .
- Esta práctica permitió afianzar los conceptos de clase y enfrentarse de manera real al manejo y tratamiento de **Data WareHouse** y a partir de esta base de datos crear informes y reportes que pueden ser utiles para las entidades.