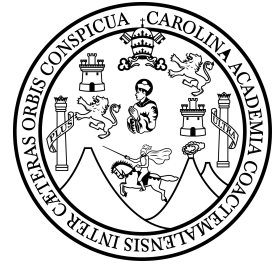


Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Seminario de Sistemas 2
Sección "N"
GRUPO C



Proyecto Fase 1
Laboratorio de Seminario de Sistemas 2

Integrantes:

Nombre	Carné
Erick Villatoro	201900907
Leonardo Roney Martínez Maldonado	201780044
Juan Diego Alvarado	201807335
Jose Mateo	201603189

Ciudad de Guatemala, 27 de diciembre de 2022

Dataset de Elecciones Electorales

Para el dataset que se escogió para este proyecto, cuenta con la información acerca de las elecciones gubernamentales de cada país de centroamérica, dividiendo la información en regiones o departamentos, los partidos políticos involucrados en las elecciones e información acerca de la población que realizó la votación. Este dataset contaba con 20,971 registros y sus campos eran los siguientes:

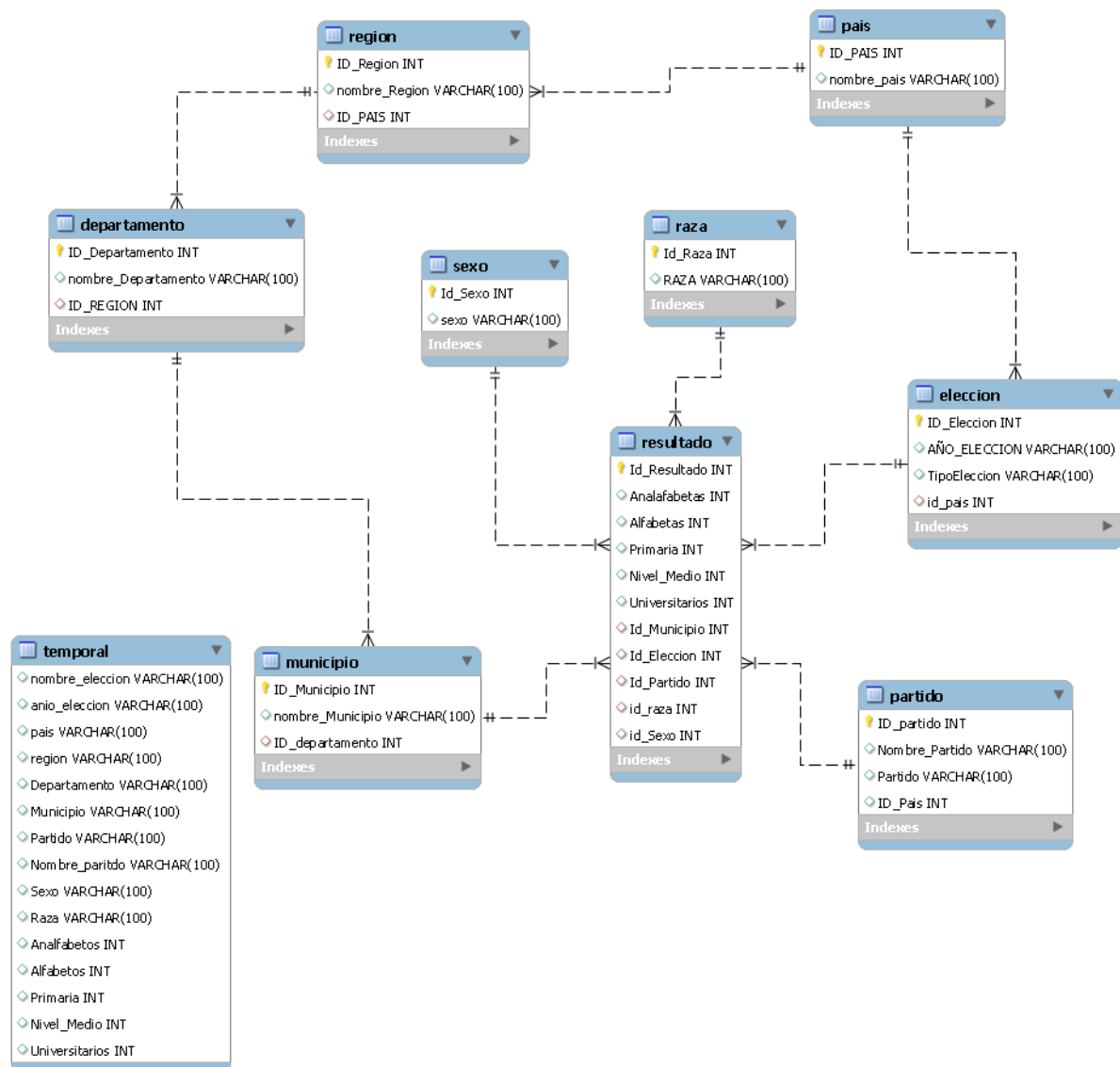
- Nombre elección
- Año elección
- País
- Región
- Departamento
- Municipio
- Partido
- Nombre Partido
- Sexo
- Raza
- Analfabetos
- Alfabetos
- Primaria
- Nivel Medio
- Universitarios

El dataset seleccionado estaba listo para utilizar, por lo que no se realizó ningún tipo de limpieza o manipulación al dataset.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	NOMBRE	AÑO_ELECCION	PAIS	REGION	DEPTO	MUNICIPIO	PARTIDO	NOMBRE_SEXO	RAZA	ANALFABETOS	ALFABETOS	SEXO	RAZA	PRIMARIA	NIVEL_MEDIO	UNIVERSITARIOS	
2	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Sensuntepeque	ARENA	Alianza Nacional	hombres	INDIGENA	2298	4800	hombres	INDIGENA	1471	450	2879
3	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Ilobasco	ARENA	Alianza Nacional	hombres	INDIGENA	2448	5536	hombres	INDIGENA	1807	2966	763
4	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Victoria	ARENA	Alianza Nacional	hombres	INDIGENA	1724	3638	hombres	INDIGENA	1421	1183	1034
5	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	San Isidro	ARENA	Alianza Nacional	hombres	INDIGENA	583	2772	hombres	INDIGENA	1747	895	130
6	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Jutiapa	ARENA	Alianza Nacional	hombres	INDIGENA	31	5010	hombres	INDIGENA	2576	602	1832
7	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Tejatepec	ARENA	Alianza Nacional	hombres	INDIGENA	883	5972	hombres	INDIGENA	2129	2947	896
8	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Dolores	ARENA	Alianza Nacional	hombres	INDIGENA	1160	4563	hombres	INDIGENA	139	1879	2545
9	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Cinquera	ARENA	Alianza Nacional	hombres	INDIGENA	689	3472	hombres	INDIGENA	275	263	2934
10	Elecciones:	2005	EL SALVADOR	REGION 1	Cabañas	Guacotect	ARENA	Alianza Nacional	hombres	INDIGENA	49	3094	hombres	INDIGENA	678	1521	895
11	Elecciones:	2005	EL SALVADOR	REGION 2	Chalatenango	Chalatenango	ARENA	Alianza Nacional	hombres	INDIGENA	904	2996	hombres	INDIGENA	2060	693	243
12	Elecciones:	2005	EL SALVADOR	REGION 2	Chalatenango	Nueva Concepción	ARENA	Alianza Nacional	hombres	INDIGENA	2807	6042	hombres	INDIGENA	1986	1752	2304
13	Elecciones:	2005	EL SALVADOR	REGION 2	Chalatenango	La Palma	ARENA	Alianza Nacional	hombres	INDIGENA	819	6088	hombres	INDIGENA	2541	2503	1044
14	Elecciones:	2005	EL SALVADOR	REGION 2	Chalatenango	Tejutla	ARENA	Alianza Nacional	hombres	INDIGENA	1732	7004	hombres	INDIGENA	2080	2113	2811
15	Elecciones:	2005	EL SALVADOR	REGION 2	Chalatenango	La Reina	ARENA	Alianza Nacional	hombres	INDIGENA	970	7497	hombres	INDIGENA	1661	2958	2878

Modelo ER de Datawarehouse

Para el modelo del data warehouse, se elaboraron 8 tablas y una tabla extra, una tabla 'temporal' en la que se realiza la carga del dataset proporcionado y de donde se extraen los datos para llenar las tablas del modelo. Por lo que el modelo Entidad-Relación planteado es el siguiente:

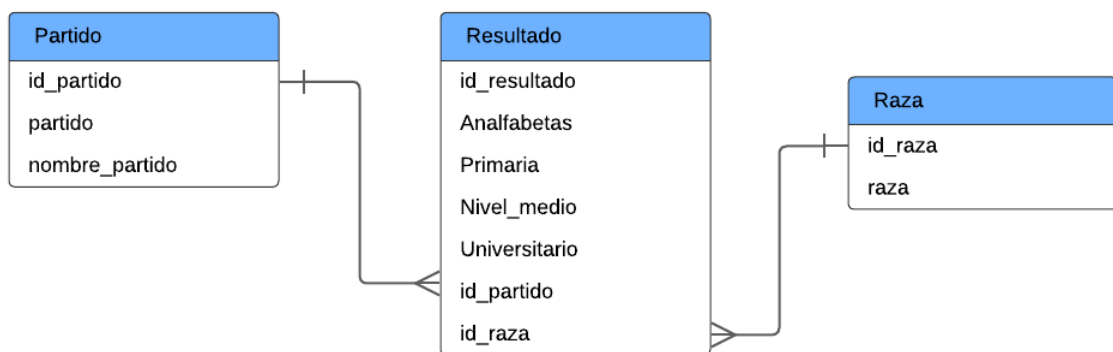


Datamart 1: Tipo de Población por Partido

Para el primer datamart se extrajeron los datos acerca de la población y los votos que realizaron por cada partido, independientemente del año de elección que poseían. Por lo que se seleccionaron las siguientes atributos de las tablas del modelo entidad-relación planteado para el data warehouse:

- Tabla Partido
 - Partido
 - Nombre del Partido
- Tabla Resultado
 - Analfabetos
 - Primaria
 - Secundaria
 - Nivel Medio
- Tabla Raza
 - Raza

El data mart posee tres tablas, con el mismo nombre que el modelo original para mantener la coherencia y las relaciones entre estas quedaron de la siguiente manera:

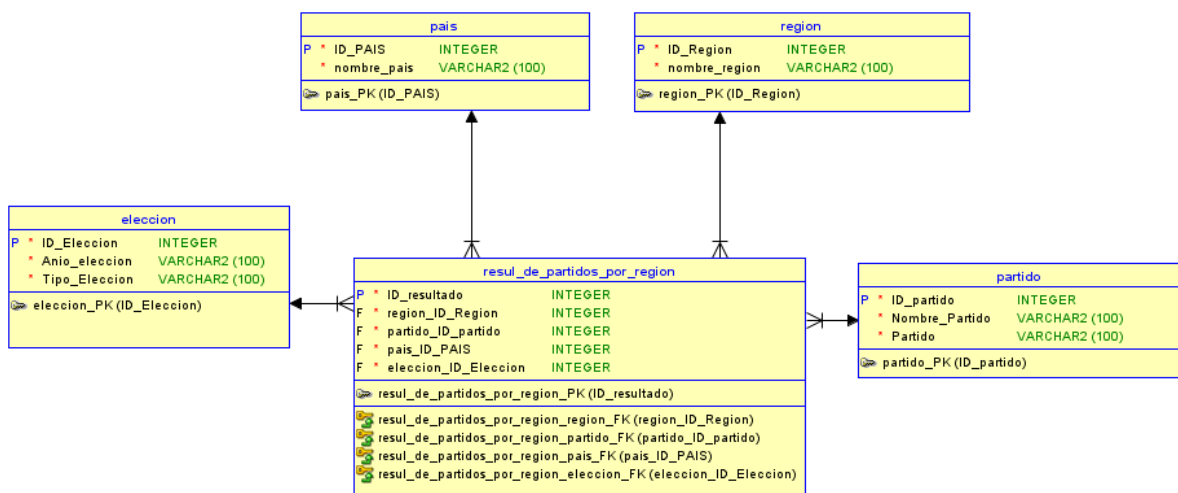


Datamart 2: Partidos por Región

Para el segundo datamart se extrajeron los datos acerca de las votaciones por región y el país en donde se realizó la votación, el año en el que se realizó la elección y el partido político al que se le acreditan los votos. Por lo que se seleccionaron las siguientes atributos de las tablas del modelo entidad-relación planteado para el data warehouse:

- Tabla Pais
 - País
- Tabla Región
 - Región
- Tabla Elección
 - Año de elección
 - Tipo de elección
- Partido
 - Nombre del partido
 - Partido
- Tabla Resultado
 - Id País
 - Id Región
 - Id Partido
 - Id Elección

El data mart posee cinco tablas, con el mismo nombre que el modelo original para mantener la coherencia y las relaciones entre estas quedaron de la siguiente manera:



Consultas

Para este apartado, se realizaron diez consultas al datawarehouse, de donde se puede destacar información para el análisis de datos posterior. Las consultas realizadas fueron las siguientes:

Consulta 1: Suma de votos por municipio de cada país en donde los votantes eran mujeres.

	País	Departamento	Sexo	SUMA
1	Costa Rica	Alajuela	mujeres	218517
2	Costa Rica	Cartago	mujeres	110331
3	Costa Rica	Guanacaste	mujeres	158382
4	Costa Rica	Heredia	mujeres	118350
5	Costa Rica	Limon	mujeres	73896
6	Costa Rica	Puntarenas	mujeres	145755
7	EL SALVADOR	Cabañas	mujeres	120818
8	EL SALVADOR	Chalatenango	mujeres	461156
9	EL SALVADOR	Cuscatlán	mujeres	214837
10	EL SALVADOR	La Libertad	mujeres	290670
11	EL SALVADOR	La Paz	mujeres	295131
12	EL SALVADOR	La Unión	mujeres	240491
13	EL SALVADOR	Morazán	mujeres	346971
14	EL SALVADOR	San Miguel	mujeres	283501
15	EL SALVADOR	San Salvador	mujeres	252681

Consulta 2: Suma de votos por municipio de cada país en donde los votantes eran hombres.

	País	Departamento	Sexo	SUMA
1	Costa Rica	Alajuela	hombres	197862
2	Costa Rica	Cartago	hombres	104783
3	Costa Rica	Guanacaste	hombres	151390
4	Costa Rica	Heredia	hombres	144354
5	Costa Rica	Limon	hombres	80228
6	Costa Rica	Puntarenas	hombres	153989
7	EL SALVADOR	Cabañas	hombres	107776
8	EL SALVADOR	Chalatenango	hombres	437621
9	EL SALVADOR	Cuscatlán	hombres	229718
10	EL SALVADOR	La Libertad	hombres	293993
11	EL SALVADOR	La Paz	hombres	286150
12	EL SALVADOR	La Unión	hombres	234099
13	EL SALVADOR	Morazán	hombres	379598
14	EL SALVADOR	San Miguel	hombres	297945
15	EL SALVADOR	San Salvador	hombres	261476

Consulta 3: Promedio de votos por cada región del país en donde los votantes eran analfabetas.

	País	Region	PromedioVotos
1	Costa Rica	REGION 1	923546.5000
2	Costa Rica	REGION 2	868196.5000
3	Costa Rica	REGION 3	694457.0000
4	EL SALVADOR	REGION 1	718645.0000
5	EL SALVADOR	REGION 2	1598851.2000
6	EL SALVADOR	REGION 3	1592646.2500
7	EL SALVADOR	REGION 4	1767436.0000
8	GUATEMALA	Región 1	1378892.0000
9	GUATEMALA	Región 2	676662.5000
10	GUATEMALA	Región 5	1231731.3333
11	GUATEMALA	Región 4	1031350.0000
12	GUATEMALA	Región 6	1455999.3333
13	GUATEMALA	Región 7	1450287.6000
14	HONDURAS	REGION 1	1273970.6000
15	HONDURAS	REGION 2	1382556.8000

Consulta 4: Promedio de votos por cada región del país en donde los votantes poseían un grado académico, es decir, alfabetas.

	País	Region	PromedioVotos
1	Costa Rica	REGION 1	319040.5000
2	Costa Rica	REGION 2	279517.5000
3	Costa Rica	REGION 3	230416.5000
4	EL SALVADOR	REGION 1	249351.0000
5	EL SALVADOR	REGION 2	534088.0000
6	EL SALVADOR	REGION 3	521894.7500
7	EL SALVADOR	REGION 4	569793.6667
8	GUATEMALA	Región 1	417451.0000
9	GUATEMALA	Región 2	234386.0000
10	GUATEMALA	Región 5	412094.0000
11	GUATEMALA	Región 4	344923.3333
12	GUATEMALA	Región 6	495393.3333
13	GUATEMALA	Región 7	480946.0000
14	HONDURAS	REGION 1	423697.4000
15	HONDURAS	REGION 2	470884.4000

Consulta 5: Suma de votos por país en donde los votantes eran mujeres analfabetas.

	País	MUJERES_ANALFABETAS
1	Costa Rica	2487802
2	EL SALVADOR	10157133
3	GUATEMALA	13388770
4	HONDURAS	11641942
5	Nicaragua	6128995
6	Panama	3400334

Consulta 6: Suma de votos por país en donde los votantes eran hombres analfabetos.

	País	HOMBRES_ANALFABETAS
1	Costa Rica	2484598
2	EL SALVADOR	10228661
3	GUATEMALA	13473450
4	HONDURAS	11528508
5	Nicaragua	6140088
6	Panama	3445602

Consulta 7: Porcentaje de votos por país de votantes analfabetas.

	País	PORCENTAJE
1	Costa Rica	37.5214
2	EL SALVADOR	37.4767
3	GUATEMALA	37.3306
4	HONDURAS	37.6752
5	Nicaragua	37.5215
6	Panama	37.5131

Consulta 8: Porcentaje de votos por país de votantes alfabetas.

	País	PORCENTAJE
1	Costa Rica	37.4731
2	EL SALVADOR	37.7406
3	GUATEMALA	37.5668
4	HONDURAS	37.3081
5	Nicaragua	37.5894
6	Panama	38.0125

Consulta 9: Suma total de votos por departamento del país de Guatemala.

	País	nombre_Departamento	suma
1	GUATEMALA	Alta Verapaz	1723126
2	GUATEMALA	Baja Verapaz	872550
3	GUATEMALA	Chimaltenango	1764670
4	GUATEMALA	Chiquimula	1213967
5	GUATEMALA	El Progreso	850522
6	GUATEMALA	Escuintla	1404489
7	GUATEMALA	Guatemala	1796343
8	GUATEMALA	Huehuetenango	3359427
9	GUATEMALA	Izabal	547163
10	GUATEMALA	Jalapa	746683
11	GUATEMALA	Jutiapa	1833716
12	GUATEMALA	Peten	1436857
13	GUATEMALA	Quetzaltenango	2593013
14	GUATEMALA	Quiche	2264208
15	GUATEMALA	Retalhuleu	957369

Consulta 10: Suma total de votos por departamento del país de Costa Rica.

	País	nombre_Departamento	suma
1	Costa Rica	Alajuela	1654358
2	Costa Rica	Cartago	830816
3	Costa Rica	Guanacaste	1221301
4	Costa Rica	Heredia	1074127
5	Costa Rica	Limon	648902
6	Costa Rica	Puntarenas	1200845

Conclusiones

- La limpieza correcta de la información es de suma importancia ya que los datos son ahora hoy en día como un recurso esencial para cualquier institución de cualquier ámbito. Si una empresa o institución tiene una mala calidad de datos es posible que no logre obtener la mayor cantidad de ganancias y que todas sus campañas de marketing estén mal orientadas llegando hasta el extremo de fracasar.
- Teniendo calidad en su información las instituciones del estado que realicen la limpieza de datos de manera correcta tendrán la capacidad de poder corregir todos los registros que se consideran inexactos, es decir todos los datos que estén incompletos o que simplemente sean corruptos, sean irrelevantes que sean pertenecientes a una base de datos.
- Tener un DW bien distribuido , es de gran ayuda ya que al momento de querer manejar grandes cantidades de información puede resultar muchas veces tedioso , por lo que tener un DW puede resultar que a futuro la información sea almacenada de la manera más óptima.
- Utilizar modelo estrella en la base de datos que se esté trabajando puede dar como resultado que al momento de realizar consultas a la base de datos resulte para el usuario final o el programador resulte sencillo ayudando así a que realice su trabajo de manera rápida, sencilla y sobre todo eficaz a diferencia de los otros modelos donde puede que se complique un poco más realizar consultas.