

# SQL para Ciencia de Datos

Profesor: Miguel Adrianzen

17 de febrero de 2022





# Agenda

1. Datos vs Información
2. Niveles de analítica
3. Arquitectura de Bases de datos
4. Conceptos de bases de datos
5. Entorno Sql server



# Datos vs Información

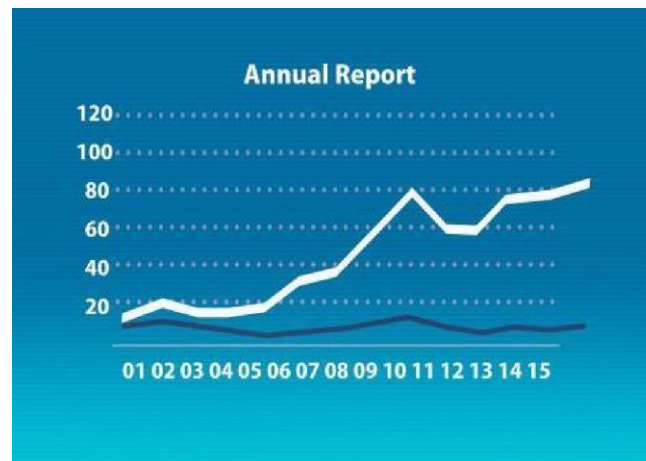
## Datos

	A	B	C	D	E	F	G	H	I
1	PERIODO	COD_SPSA	DESCRIPCION	MARCA	ESTADO_P	COD_LOCA	COD_LOCAL	DESCRIPCION	ESTADO_LOCA
2	2020-02-02 a	20171310	STARBUCKS	STARBUCKS	ACTIVO	P109			
3	2020-02-02 a	20171306	STARBUCKS	STARBUCKS	ACTIVO	P048			
4	2020-02-02 a	20171308	STARBUCKS	STARBUCKS	ACTIVO	P001			
5	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	V090			
6	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	P048			
7	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	P106			
8	2020-02-02 a	20171306	STARBUCKS	STARBUCKS	ACTIVO	P001			
9	2020-02-02 a	20171306	STARBUCKS	STARBUCKS	ACTIVO	P135			
10	2020-02-02 a	20171309	STARBUCKS	STARBUCKS	ACTIVO	P723			
11	2020-02-02 a	20171310	STARBUCKS	STARBUCKS	ACTIVO	P723			
12	2020-02-02 a	20171310	STARBUCKS	STARBUCKS	ACTIVO	P048			
13	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	P061			
14	2020-02-02 a	20171308	STARBUCKS	STARBUCKS	ACTIVO	P723			
15	2020-02-02 a	20181625	STARBUCKS	STARBUCKS	INACTIVO	P001			
16	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	P081			
17	2020-02-02 a	20171308	STARBUCKS	STARBUCKS	ACTIVO	P133			
18	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	P001			
19	2020-02-02 a	20171310	STARBUCKS	STARBUCKS	ACTIVO	P061			
20	2020-02-02 a	20171306	STARBUCKS	STARBUCKS	ACTIVO	P723			
21	2020-02-02 a	20171308	STARBUCKS	STARBUCKS	ACTIVO	P061			
22	2020-02-02 a	20171307	STARBUCKS	STARBUCKS	ACTIVO	P010			
23	2020-02-02 a	20171310	STARBUCKS	STARBUCKS	ACTIVO	P465			

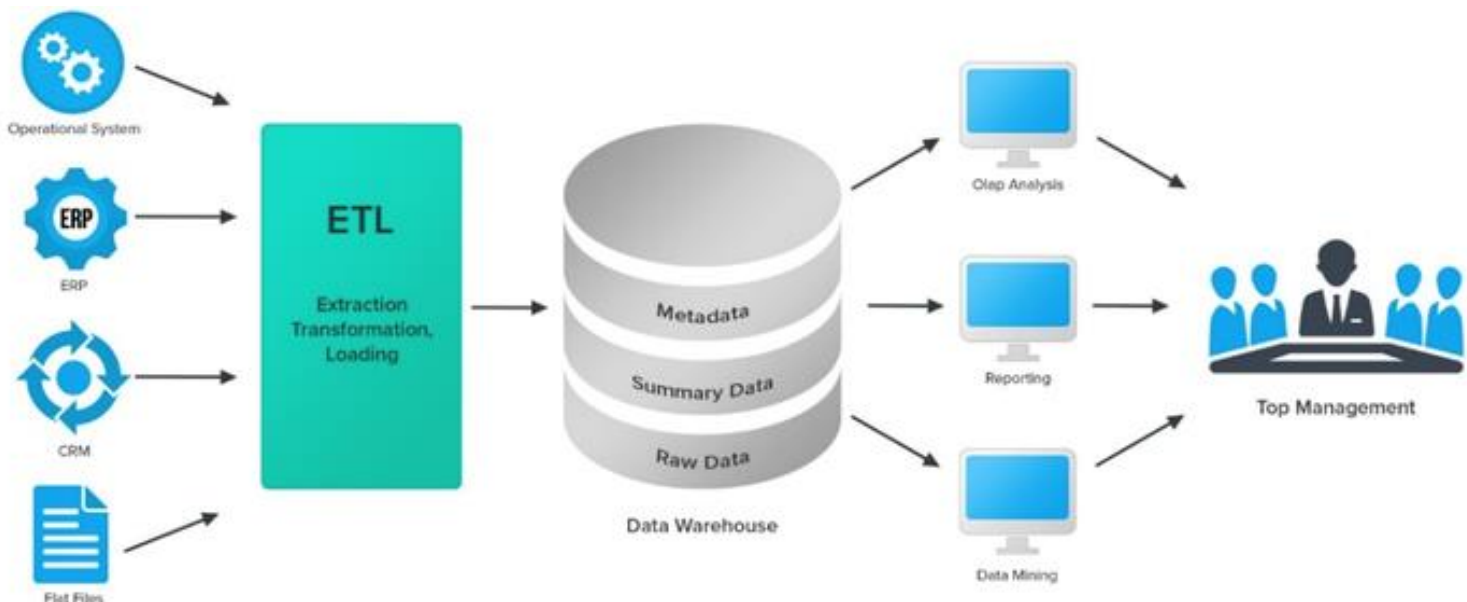
VTADTOFTE1583519483456_3876	
-----------------------------	--

## Información



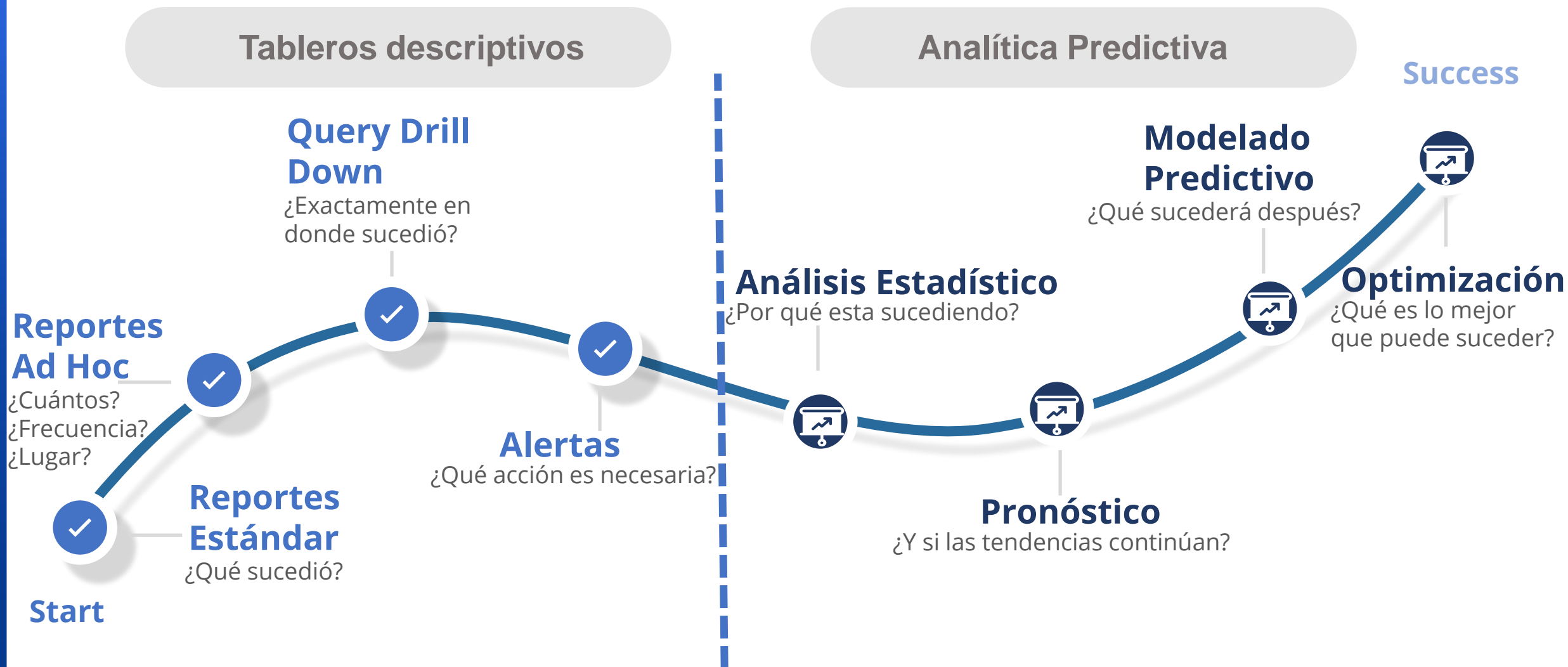
## Conocimiento

*“El crecimiento de la empresa esta estrechamente vinculado a la ejecución de nuevas campañas personalizadas y un mejor entrenamiento de nuestra fuerza de ventas”*





# Niveles de análisis en la toma de decisiones

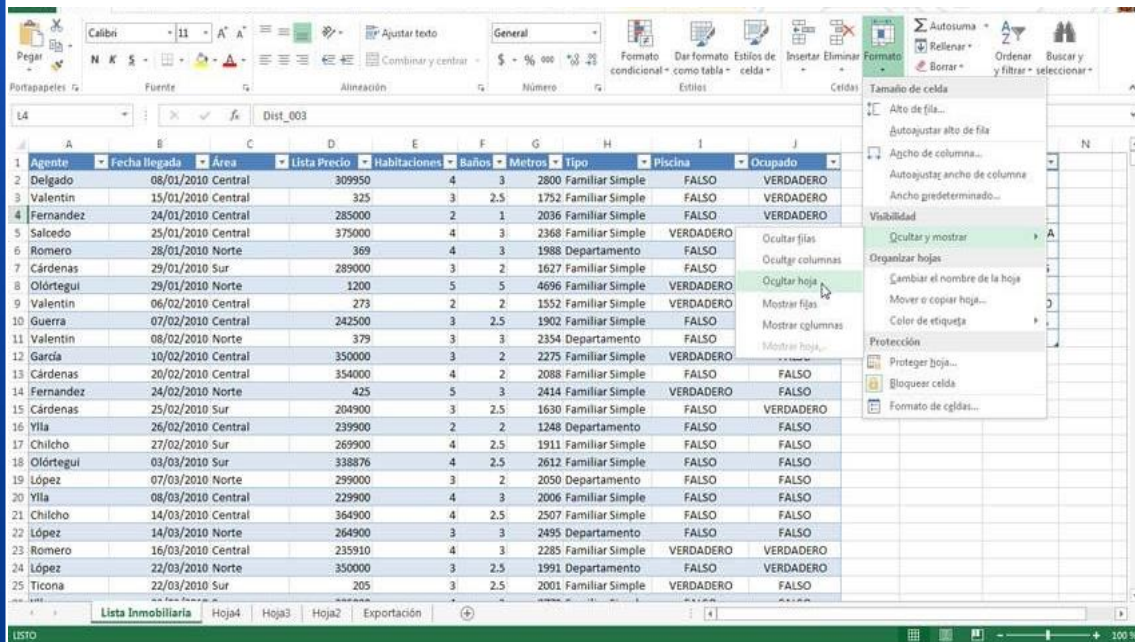




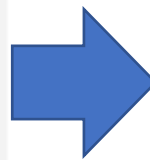
# Definición y arquitectura de bases de datos

- Una base de datos, a fin de ordenar la información de manera lógica, posee un orden que debe ser cumplido para acceder a la información de manera coherente. Cada base de datos contiene una o más tablas, que cumplen la función de contener los campos.

*Como podemos responder diferentes preguntas de lo que esta pasando:  
¿ Cuanto vendimos?, ¿ Que vendimos ?, ¿ A quienes le vendimos?*



Agente	Fecha llegada	Area	Lista Precio	Habitaciones	Baños	Metros	Tipo	Piscina	Ocupado
Delgado	08/01/2010	Central	309950	4	3	2800	Familiar Simple	FALSO	VERDADERO
Valentin	15/01/2010	Central	325	3	2.5	1752	Familiar Simple	FALSO	VERDADERO
Fernandez	24/01/2010	Central	285000	2	1	2036	Familiar Simple	FALSO	VERDADERO
Salcedo	25/01/2010	Central	375000	4	3	2368	Familiar Simple	VERDADERO	
Romero	28/01/2010	Norte	369	4	3	1988	Departamento	FALSO	
Cárdenas	29/01/2010	Sur	289000	3	2	1627	Familiar Simple	FALSO	
Olórtégui	29/01/2010	Norte	1200	5	5	4696	Familiar Simple	VERDADERO	
Valentin	06/02/2010	Central	273	2	2	1552	Familiar Simple	VERDADERO	
Guerra	07/02/2010	Central	242500	3	2.5	1902	Familiar Simple	FALSO	
Valentin	08/02/2010	Norte	379	3	3	2354	Departamento	FALSO	
García	10/02/2010	Central	350000	3	2	2275	Familiar Simple	VERDADERO	
Cárdenas	20/02/2010	Central	354000	4	2	2088	Familiar Simple	FALSO	
Fernandez	24/02/2010	Norte	425	5	3	2414	Familiar Simple	VERDADERO	
Cárdenas	25/02/2010	Sur	204900	3	2.5	1630	Familiar Simple	FALSO	VERDADERO
Ylla	26/02/2010	Central	239900	2	2	1248	Departamento	FALSO	FALSO
Chilcho	27/02/2010	Sur	269900	4	2.5	1911	Familiar Simple	FALSO	FALSO
Olórtégui	03/03/2010	Sur	338876	4	2.5	2612	Familiar Simple	FALSO	FALSO
López	07/03/2010	Norte	299000	3	2	2050	Departamento	FALSO	FALSO
Ylla	08/03/2010	Central	229900	4	3	2006	Familiar Simple	FALSO	FALSO
Chilcho	14/03/2010	Central	364900	4	2.5	2507	Familiar Simple	FALSO	FALSO
López	14/03/2010	Norte	264900	3	3	2495	Departamento	FALSO	FALSO
Romero	16/03/2010	Central	235910	4	3	2285	Familiar Simple	VERDADERO	VERDADERO
López	22/03/2010	Norte	350000	3	2.5	1991	Departamento	FALSO	VERDADERO
Ticona	22/03/2010	Sur	205	3	2.5	2001	Familiar Simple	VERDADERO	FALSO



Results

Messages

	BusinessEntityID	FirstName	LastName	JobTitle	BirthDate	MaritalStatus	Gender	HireDate
1	285	Syed	Abbas	Pacific Sales Manager	1975-01-11	M	M	2013-03-14
2	38	Kim	Abercrombie	Production Technician - WC60	1966-12-14	M	F	2010-01-16
3	211	Hazem	Abolrous	Quality Assurance Manager	1977-10-26	S	M	2009-02-28
4	121	Pilar	Ackerman	Shipping and Receiving Supervisor	1972-09-09	S	M	2009-01-02
5	67	Jay	Adams	Production Technician - WC60	1976-02-11	S	M	2009-03-05
6	287	Amy	Alberts	European Sales Manager	1957-09-20	M	F	2012-04-16
7	207	Greg	Alderson	Production Technician - WC45	1970-10-18	S	M	2008-12-02
8	216	Sean	Alexander	Quality Assurance Technician	1976-03-06	S	M	2008-12-28
9	227	Gary	Altman	Facilities Manager	1971-02-18	M	M	2009-12-02
10	43	Nancy	Anderson	Production Technician - WC60	1988-11-19	M	F	2009-01-02
11	280	Pamela	Ansman-Wolfe	Sales Representative	1974-12-06	S	F	2011-05-31
12	217	Zainal	Arfin	Document Control Manager	1976-01-30	M	M	2009-01-04
13	269	Dan	Bacon	Application Specialist	1987-05-26	M	M	2009-01-11
14	41	Bryan	Baker	Production Technician - WC60	1973-08-27	S	M	2009-01-21
15	104	Mary	Baker	Production Technician - WC10	1986-09-19	M	F	2009-12-25





# Definición y arquitectura de bases de datos

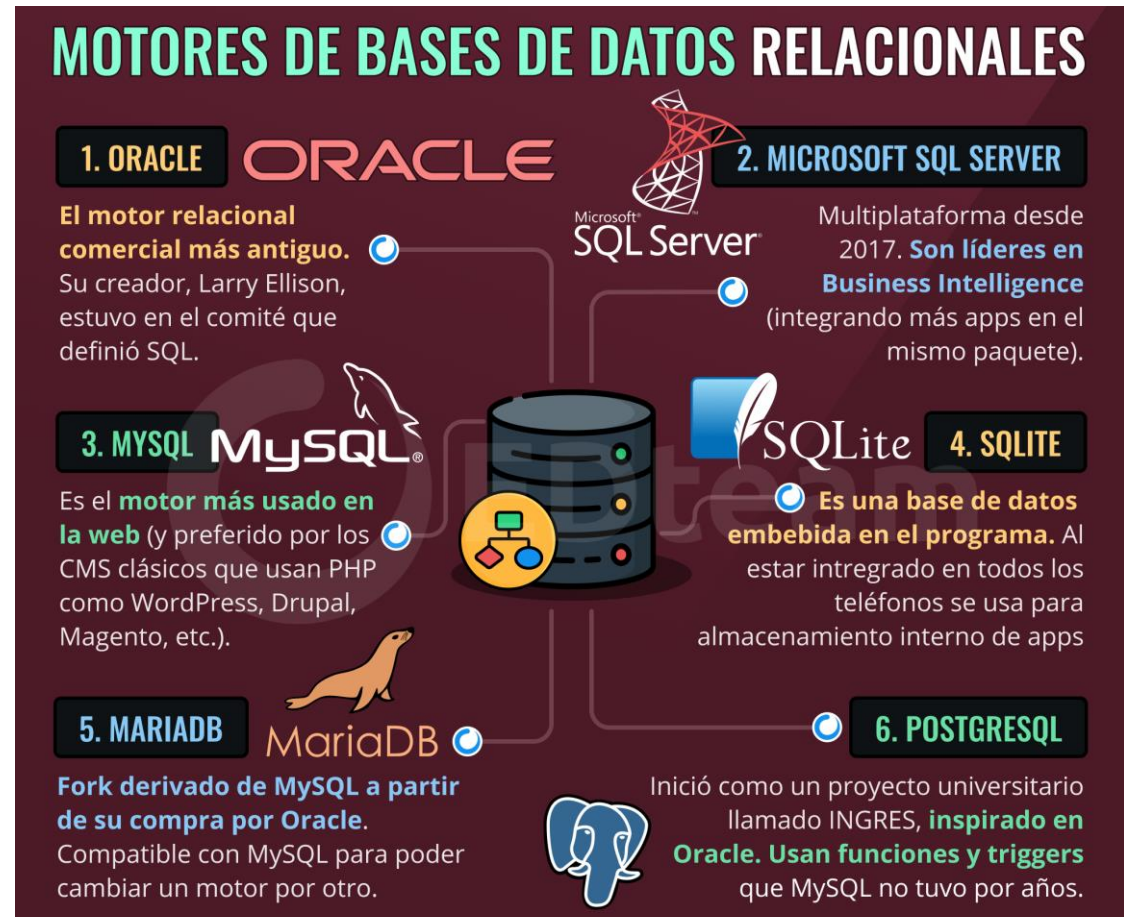
## Definición de base de datos

Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

## Características

Entre las principales características de los sistemas de base de datos podemos mencionar:

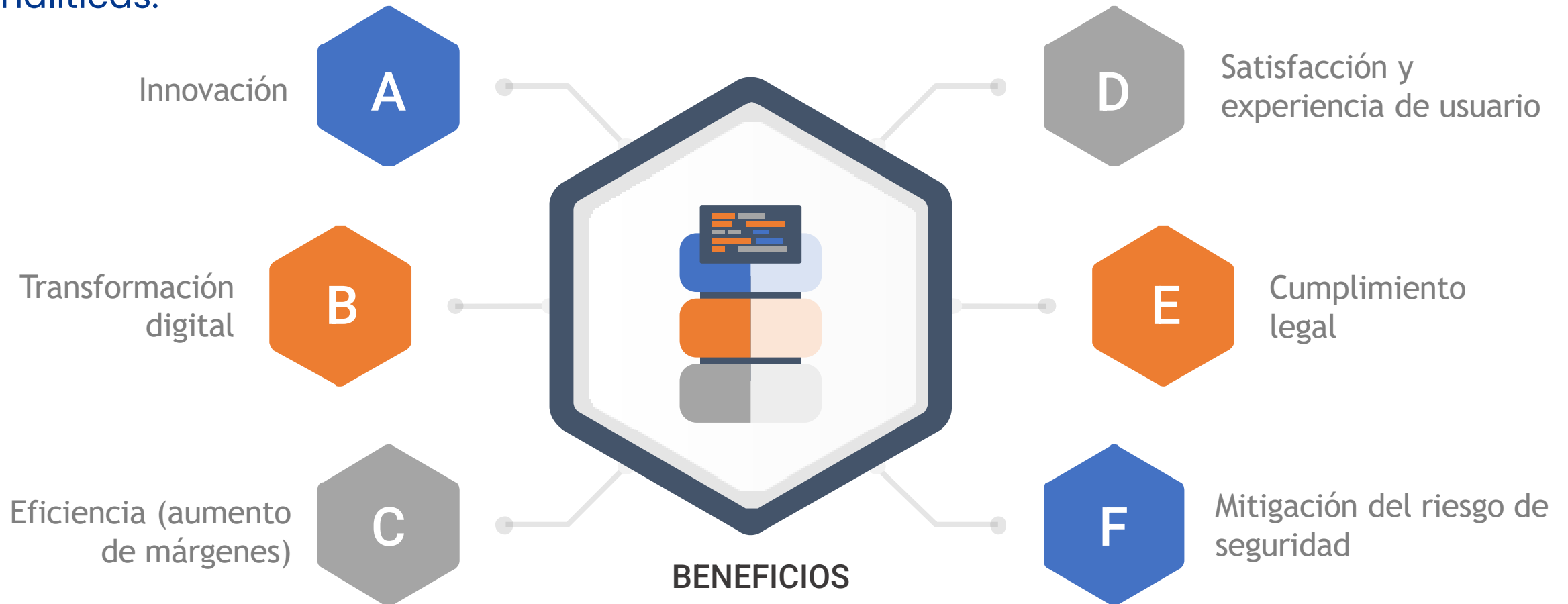
- Acceso concurrente por parte de múltiples usuarios.
- Integridad de los datos.
- Consultas complejas optimizadas.
- Seguridad de acceso y auditoría.
- Acceso a través de lenguajes de programación estándar (Transact-SQL)





# Empresa Data Driven

*Data Driven Organization* es una organización que ha implantando cultura de utilización de los datos y desarrollado los procesos correctos para tomar decisiones críticas que producen impacto significativo en el negocio utilizando tecnologías analíticas.

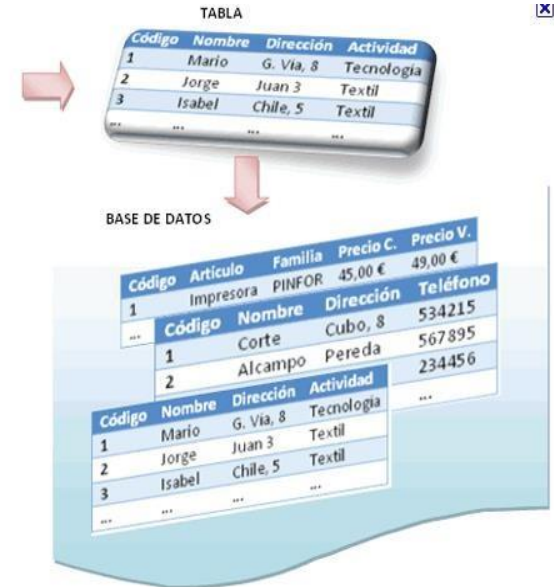




# Conceptos Básicos: Base de datos

## Base de Datos

Se define una base de datos como una serie de datos organizados en tablas relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.



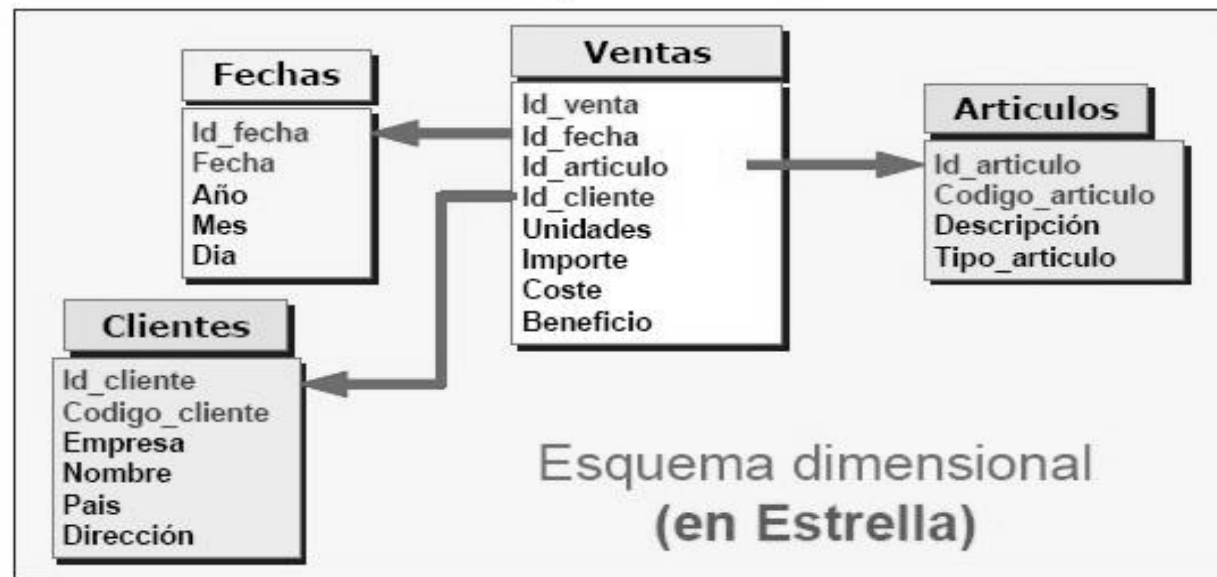
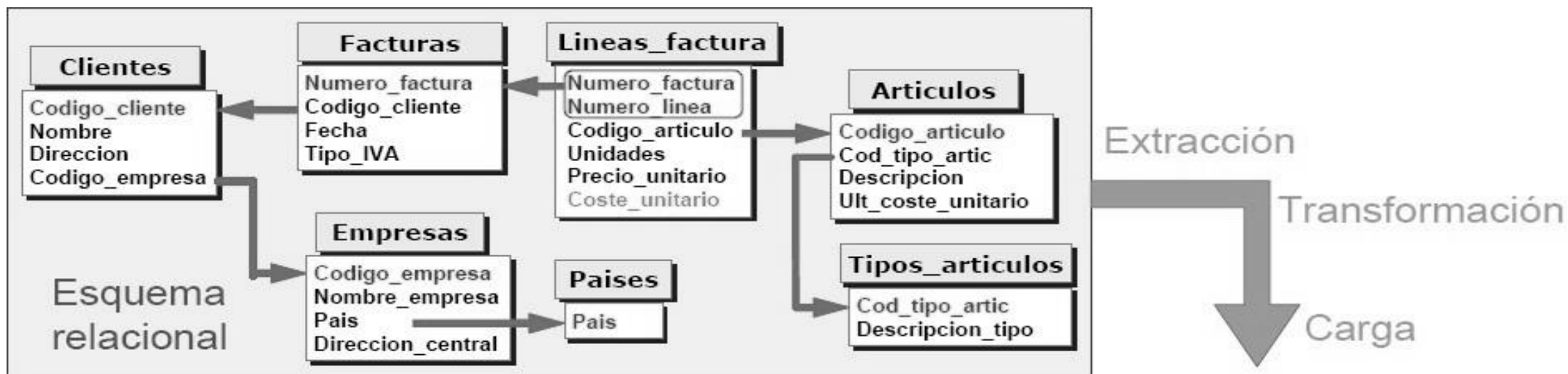
## Tablas

En una base de datos la información se organiza en **tablas**, que son filas y columnas similares a las de los libros contables o a las de las hojas de cálculo, contienen columnas con diversos atributos o variables y registros.



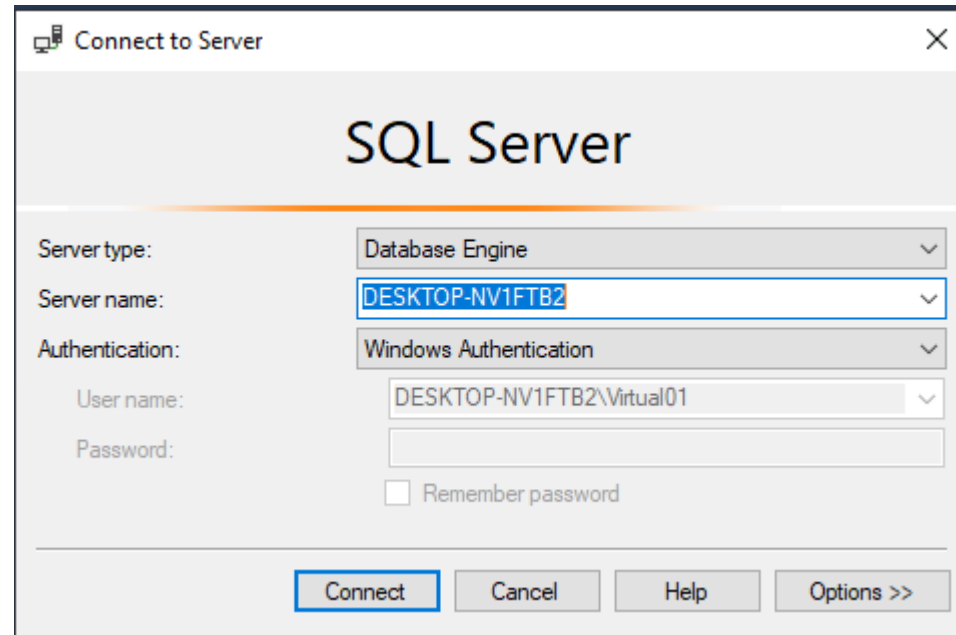
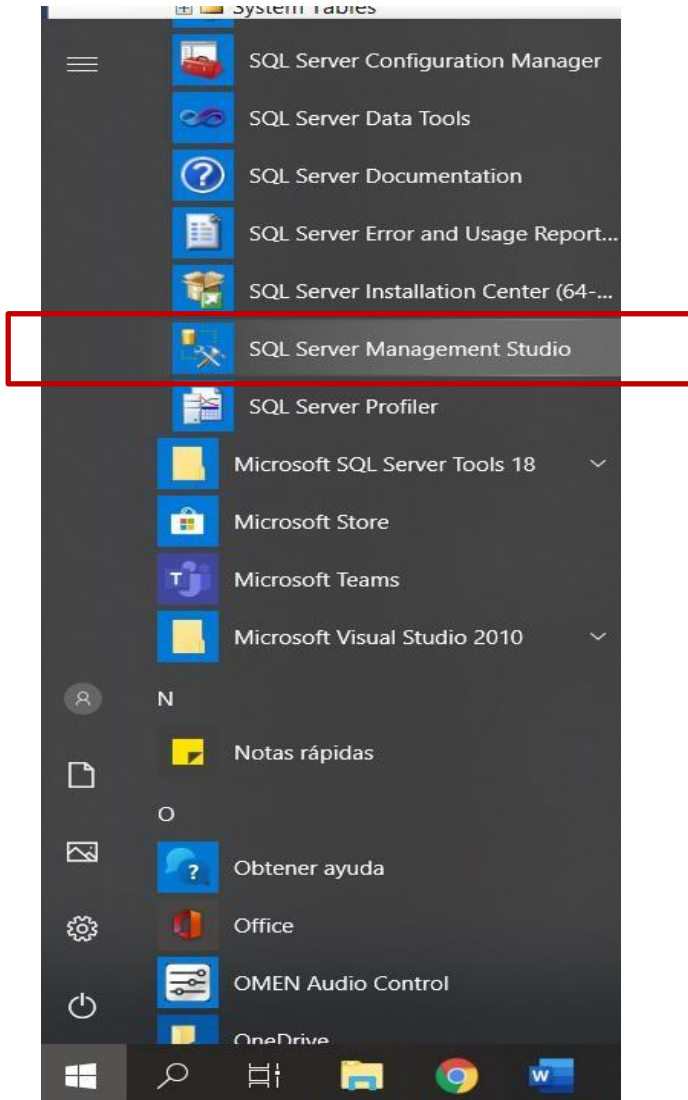


# Definición y arquitectura de bases de datos





# Iniciando con Microsoft SQL Server





# Entorno de Microsoft SQL Server

Explorar los  
objetos de la  
Base de  
datos Bases  
de datos y  
componentes  
de la BD

The screenshot displays the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure for 'BD\_PRACTICAS', including tables like 'AGENCIAS', 'CLIENTE\_PERFIL', and 'CODIGO'. The Query Editor in the center contains the SQL command: `USE BD_PRACTICAS; SELECT* FROM [dbo].[CLIENTE_PERFIL]`. The Results grid at the bottom shows the output of the query, which is a table with 13 columns and 19 rows of data. A red box highlights the Object Explorer and the Query Editor, with a red arrow pointing to the 'Ejecutar' (Execute) button. Another red box highlights the Results grid, with a red arrow pointing to the 'Resultados del query' (Query Results) label.

**Ejecutar sintaxis y comandos SQL sobre la BD**

**Resultados del query**

	CODIGO	MES	AGENCIA	SALDO_ACTI...	SALDO_PASI...	RENTABILID...	FH_NACIMIENTO	ECVIL	SEXO	FH_ALTA	UBIGEO	TIENE_TELEFO...	TIENE_CORR...	USO_AGENCI...
1	069237015925	201105	0183	0.00	80.00	0.00	1992-11-04	S	F	2011-03-08	0101001	N	S	S
2	969208905805	201105	0713	0.00	549.00	8.00	1954-03-19	D	F	2000-09-07	0101033	S	N	N
3	269216211291	201105	0288	0.00	250.00	1.00	1957-06-29	S	M	2004-08-14	1501001	N	S	N
4	069236057337	201105	0231	0.00	0.00	0.00	1988-09-04	S	M	2010-12-27	1803002	S	N	N
5	669230688218	201105	0301	0.00	1020.00	29.00	2009-07-22			2009-08-28	1601004	S	S	N
6	069213082222	201105	0133	0.00	0.00	0.00	1966-04-04	S	F	2002-06-22	0101035	S	S	N
7	369227326016	201105	0276	0.00	0.00	0.00	1982-08-03	S	M	2008-09-07	0102000	N	N	N
8	969224982662	201105	0267	0.00	0.00	0.00	1977-05-05	S	F	2008-02-08	2001011	N	N	N
9	269233248778	201105	0484	0.00	30.00	0.00	1988-11-08	S	M	2010-05-18	0101015	N	S	N
10	369233341121	201105	0211	0.00	0.00	0.00	1973-01-29	S	M	2010-05-24	2001004	N	N	N
11	369228337367	201105	0137	0.00	48.00	0.00	1979-01-15	S	F	2011-05-31	0101009	N	N	S
12	469215431981	201105	0831	0.00	0.00	0.00	1972-06-15	S	F	2004-01-28	2001009	S	S	N
13	569234566136	201105	0222	0.00	128.00	0.00	1984-04-07	S	M	2010-09-08	0601001	S	N	N
14	269225233873	201105	0701	1410.00	0.00	13.00	1950-12-31	S	F	2008-02-28	0101031	N	S	N
15	369222308148	201105	0126	0.00	0.00	0.00	1983-07-30	S	F	2007-05-19	0101017	S	S	N
16	969216981561	201105	0341	63810.00	0.00	501.00	1969-08-08	S	M	2005-03-09	0101030	N	S	N
17	769228724514	201105	0142	0.00	0.00	0.00	1977-12-19	S	F	2009-01-29	0101006	S	N	N
18	269219238258	201105	0249	5000.00	5872.00	183.00	1967-08-11	S	M	2006-06-12	1401010	N	S	N
19	669087634374	201105	0195	0.00	386.00	8.00	1922-07-26	C	F	1991-10-17	0101018	N	S	N

Query executed successfully. DESKTOP-3R0MBV6 (11.0 RTM) DESKTOP-3R0MBV6\dvelas... BD\_PRACTICAS 00:00:03 200000 rows



# Lenguaje Transact SQL

## DDL (Lenguaje de Definición de datos)

- Se utiliza para definir y administrar objetos de la BD, tales como Bases, tablas, y vistas.
- Usualmente las mas usadas son CREATE TABLE, ALTER TABLE, DROP TABLE. S
- Se utilizan para crear tablas, modificar (agregar o borrar columnas, modificar, etc), y eliminar tablas respectivamente.

## DML (Lenguaje de Manipulación de datos)

- Se utiliza para manipular información de las BD, para ello utilizaremos instrucciones como INSERT, SELECT, CASE, DATE, UPDATE, DELETE y otros.
- Estas instrucciones nos permiten seleccionar filas, filtrar, insertar nuevas filas, modificar las filas existentes y eliminar datos no deseados.



# Orden de ejecución de una sintaxis

Reporte de agencias cuyo suma de saldo pasivo total de sus clientes superen los S/. 100,000 soles, no considerar los clientes del segmento corporativo; el reporte tiene que ser por código agencia mostrando el numero de clientes y el monto total del pasivo.

## Clausula SELECT

```
SELECT A.COD_AGENCIA, COUNT(A.CODIGO),  
SUM(B.SALDO_PASIVO)
```

## Clausula FROM

```
FROM TB_CLIENTE_PERFIL A  
INNER JOIN TB_CLIENTE_RENTABILIDAD B ON A.CODIGO=B.CODIGO
```

## Clausula WHERE

```
WHERE COD_SEGMENTO NOT IN ('288011','288022')
```

## Clausula GROUP BY

```
GROUP BY COD_AGENCIA
```

## Clausula HAVING

```
HAVING SUM(B.SALDO_PASIVO)>100000
```

## Clausula ORDER BY

```
ORDER BY A.COD_AGENCIA ASC
```



# Cláusulas

Son condiciones de modificación utilizadas para definir los datos que se desea seleccionar o manipular

Clausula	Descripción
SELECT	Utilizado para consultar registros de la base de datos que satisfacen un criterio determinado.
FROM	Utilizado para especificar la tabla de la cual se va seleccionar los registros
WHERE	Utilizado especificar las condiciones que deben de reunir los registros que se van a seleccionar
GROUP BY	Utilizado para separar los registros seleccionados en grupos específicos
HAVING	Utilizado para expresar la condición que debe de satisfacer cada grupo
ORDER BY	Utilizado para ordenar los registros seleccionados de acuerdo con un orden específico.





**Ciencia de Datos  
Ecuador**

# Gracias por participar

Síguenos en nuestras redes sociales

Facebook: @CienciaDeDatosEcuador

Instagram: @datos\_ecuador

LinkedIn: @datosecuador

[www.cienciadedatosec.com](http://www.cienciadedatosec.com)

+593 99 614 9457