

CASO Ingeniero de Datos

Se le presentan una serie de preguntas relacionadas con la construcción de bases de datos y relaciones entre ellas. Favor responder dejando evidencia del proceso y/o código utilizado.

1. Dada esta base de datos, ¿cuál es el nivel de granularidad más fino en la reportería que se puede tener? Explique su razonamiento (el identificador de gestión no se usa en reportería).

Base de datos pregunta 1			
Id_gestión	ID_cliente	Cuenta	Tarjeta
0001	A	01	T1
0002	A	01	T1
0003	B	02	T2
0004	B	02	T2
0005	B	03	T3
0006	C	04	T4
0007	C	04	T4
0008	C	04	T5
0009	C	05	T6
0010	C	05	T6
0011	A	01	T1
0012	A	01	T1
0013	D	06	T7
0014	D	06	T7
0015	D	06	T8
0016	D	06	T8
0017	D	06	T7
0018	D	06	T7
0019	D	06	T9
0020	D	06	T7

2. Si se tienen las siguientes bases de datos, haga el código SQL que permita unir estas dos tablas de la forma más eficiente posible. Explique su razonamiento.

Base de datos pregunta 2.1			
Cuenta	Tarjeta	Fecha (año-mes)	Monto
01	A1	202102	324.27
01	A1	202102	948.15
01	A1	202102	590.57
01	A2	202102	503.55
02	B1	202102	685.66
02	B1	202102	502.23
02	B1	202102	855.73
03	C1	202102	699.11
03	C1	202102	784.01
04	D1	202102	990.07
04	D1	202102	723.55
04	D1	202102	585.19
04	D2	202102	993.21

Base de datos pregunta 2.2			
Cuenta	Tarjeta	Fecha (año-mes)	Monto
01	A1	202101	497.05
01	A1	202101	802.77
02	B1	202101	0.41
02	B1	202101	753.65
02	B1	202101	342.07
02	B1	202101	574.72
03	C1	202101	621.92
03	C2	202101	317.52
04	D1	202101	681.45
04	D2	202101	479.97
05	E1	202101	419.34
05	E1	202101	944.46
05	E1	202101	692.59

3. Dada la estructura de esta base de datos, ¿qué variables (mínimo 1 y máximo 2) usaría para crearle un índice? Explique su razonamiento.

Resumen de base de datos pregunta 3				
Campo	Descripción	tipo de dato	Ejemplo	Cantidad únicos
ID_cliente	Identificador del cliente	string	CLIENTE1	113,456
Fecha	Año y mes de los datos	numérico	202103	36
Cuenta	Número de cuenta que hizo la transacción	string	1234567890	185,678
Tipo_cuenta	Cuenta es de crédito o débito	string	CRE - DEB	2
plan_lealtad	plan de lealtad de la cuenta	string	Millas Plus	12
Color	color de la tarjeta de crédito	string	Black	4
Monto_trx	Monto de la transacción	numérico	117.56	81,456,662
Afiliado	Comercio donde se hizo la transacción	string	30405060	25,678
			total filas	113,567,890

4. ¿Cuál es la diferencia entre un Data Warehouse y un Data Mart? Si el jefe comercial de seguros le solicita un repositorio de datos, ¿cuál de estos 2 sería la mejor solución? Explique su razonamiento.

5. Se tienen varios exceles con la información que se muestra a continuación. Sugiera qué estructura tiene que tener una base de datos de SQL que unifique todas estas tablas. Explique su razonamiento (puede apoyarse creando un prototipo de cómo será la base de datos resultante). Cada fin de mes se irán agregando registros nuevos. Aparte de los 3 ejemplos hay 27 tablas más para unificar.

Base de datos pregunta 5.1		
ID_Cliente	Fecha	Trx_total
CLIENTE01	202012	447.96
CLIENTE02	202012	585.15
CLIENTE03	202012	560.09
CLIENTE04	202012	79.03
CLIENTE05	202012	226.61
CLIENTE06	202012	250.05
CLIENTE07	202012	294.66
CLIENTE08	202012	293.62
CLIENTE09	202012	425.01
CLIENTE10	202012	447.77

Base de datos pregunta 5.2		
ID_Cliente	Fecha	Trx_total
CLIENTE01	202101	494.87
CLIENTE03	202101	649.52
CLIENTE05	202101	531.75
CLIENTE06	202101	3.98
CLIENTE08	202101	499.74
CLIENTE09	202101	26.11
CLIENTE10	202101	939.41
CLIENTE14	202101	278.27
CLIENTE22	202101	28.83
CLIENTE77	202101	613.60

Base de datos pregunta 5.3		
ID_Cliente	Fecha	Trx_total
CLIENTE01	202102	494.87
CLIENTE03	202102	278.27
CLIENTE04	202102	531.75
CLIENTE05	202102	3.98
CLIENTE07	202102	499.74
CLIENTE08	202102	26.11
CLIENTE09	202102	939.41
CLIENTE11	202102	28.83
CLIENTE23	202102	649.52
CLIENTE62	202102	613.60

6. Considere la base de datos que se muestra a continuación. A esta base de datos se le hizo una tabla índice con las siguientes variables: V1 y V2 pero no está funcionando como se esperaba. Encuentre el error.

Base de datos pregunta 6				
V1	V2	V3	V4	V5
202001	Z	628.96	203.89	0.22
202001	Y	407.45	187.96	0.60
202001	X	23.29	91.17	0.86
202005	X	86.49	150.19	0.47
202005	Y	576.18	55.58	0.89
202005	Z	40.32	22.65	0.88
202101	Z	37.19	8.36	0.59
202101	Y	179.92	134.58	0.23
202101	X	109.67	25.65	0.40
201911	X	365.30	28.64	0.82
201911	Y	259.54	208.05	0.38
201911	Z	812.41	169.31	0.09
202102	X	414.65	196.99	0.26
202102	Y	439.23	147.63	0.94
202102	Z	100.51	129.22	0.57
202103	X	184.81	79.09	0.59
202103	Y	946.74	183.17	0.91
202103	Z	645.04	58.19	0.12

7. Se tienen las 2 bases de datos que se muestran a continuación. A la base de datos A se le quiere agregar un campo que se encuentra en la tabla B. Sugiera 2 formas de hacer esta operación, una usando joins y la otra sin usarlos.

Base de datos pregunta 7.A			Base de datos pregunta 7.B	
Cliente_ID	Tipo_cliente	Cant_cuentas	Cliente_ID	Facturacion
CLIENTE01	CRE	1	CLIENTE01	147.02
CLIENTE02	CRE	2	CLIENTE02	215.09
CLIENTE03	CRE	3	CLIENTE03	711.89
CLIENTE04	CRE	1	CLIENTE04	672.37
CLIENTE05	CRE	2	CLIENTE05	596.86
CLIENTE06	CRE	2	CLIENTE06	658.09
CLIENTE07	CRE	3	CLIENTE07	554.31
CLIENTE08	CRE	1	CLIENTE08	394.17
CLIENTE09	CRE	4	CLIENTE09	433.99
CLIENTE10	CRE	1	CLIENTE10	961.11
CLIENTE11	CRE	1	CLIENTE11	274.90
CLIENTE12	CRE	2	CLIENTE12	455.11
CLIENTE13	CRE	1	CLIENTE13	689.29
CLIENTE14	CRE	2	CLIENTE14	498.56

8. Se tienen las siguientes 3 bases de datos. Escriba el código SQL que va a permitir tener en una sola tabla la información de Cuenta, plan de lealtad, calificación por mes y facturación total (por mes). Trate de hacer el query lo más eficiente posible.

BD pregunta 8.A	
Cuenta	Plan_lealtad

01	Millas Plus
02	Cashback
03	Cashback
04	Millas Plus
05	Millas Plus
06	Cashback

BD pregunta 8.B		
Cuenta	Fecha	Calificacion

01	202101	A
01	202102	A
01	202103	A
02	202101	A
02	202102	C
02	202103	D
03	202101	E
03	202102	E
03	202103	E
04	202101	A
04	202102	A
04	202103	B
05	202101	E
05	202102	E
05	202103	E
06	202101	G
06	202102	C
06	202103	C