

50 Preguntas de Data Modeling — Respuestas Completas

1. Diferencia entre modelo conceptual, lógico y físico

- **Conceptual:** visión de alto nivel del negocio; entidades y relaciones sin detalles técnicos.
 - **Lógico:** estructura detallada con atributos, claves, cardinalidades; independiente del motor.
 - **Físico:** implementación real en un motor (tipos de datos, particiones, índices, constraints).
-

2. ¿Qué es una entidad y qué es un atributo?

- **Entidad:** objeto o concepto del negocio (Cliente, Pedido).
 - **Atributo:** característica de la entidad (Nombre, Fecha).
-

3. ¿Qué es una relación en un modelo de datos?

Asociación entre entidades (por ejemplo, Cliente realiza Pedido).

4. ¿Qué es cardinalidad y cómo se representa?

Define cuántas ocurrencias de una entidad se asocian con otra.
Ej.: 1:1, 1:N, N:M.

5. Diferencia entre 1:1, 1:N y N:M

- **1:1:** un registro se relaciona solo con uno.
- **1:N:** un registro tiene muchos relacionados.
- **N:M:** muchos a muchos (requiere tabla intermedia).

6. ¿Qué es una clave primaria y por qué es importante?

Identifica un registro de forma única y garantiza integridad de entidad.

7. Clave natural vs clave sustituta

- **Natural:** proviene del negocio (DNI).
 - **Sustituta:** generada artificialmente (ID incremental).
-

8. Clave compuesta

Clave formada por varias columnas.

9. Atributos derivables

Atributos calculados a partir de otros (edad a partir de fecha de nacimiento).

10. Dependencias funcionales

Relación donde un atributo depende de otro: $A \rightarrow B$.

11. Normalización

Proceso para reducir redundancia y evitar anomalías.

12. 1NF, 2NF, 3NF y BCNF

- **1NF:** datos atómicos.
- **2NF:** sin dependencias parciales.

- **3NF**: sin dependencias transitivas.
 - **BCNF**: toda determinante debe ser clave candidata.
-

13. Desnormalización

Agregar redundancia para mejorar rendimiento en lecturas.

14. Granularidad del dato

Nivel de detalle al que se almacenan los datos.

15. Modelado dimensional vs relacional

- **Dimensional**: orientado a análisis; tablas de hechos + dimensiones.
 - **Relacional**: orientado a transacciones; normalizado.
-

16. Tabla de hechos

Registro de eventos del negocio con métricas numéricas.

17. Tabla de dimensiones

Contexto descriptivo de los hechos (productos, clientes, tiempo).

18. Surrogate key en dimensiones

Clave artificial para facilitar versionado, consistencia y rendimiento.

19. SCD (Slowly Changing Dimension)

Técnica para manejar cambios en dimensiones.

20. SCD1, SCD2 y SCD3

- **SCD1**: sobrescribe.
 - **SCD2**: historial (múltiples filas).
 - **SCD3**: guarda valores anteriores y actuales.
-

21. Dimensión degenerada

Atributos del hecho que funcionan como dimensión sin tabla separada (número de factura).

22. Dimensión junk

Agrupa atributos pequeños y de baja cardinalidad (flags, categorías).

23. Bridge table

Maneja relaciones N:M entre hechos y dimensiones.

24. Factless fact table

Tabla de hechos sin métricas; registra eventos o relaciones (asistencia a clase).

25. Snapshot

Captura periódica de un estado (inventario diario).

26. Definir granularidad en fact table

Identificar el nivel mínimo del evento medible (ej.: venta por línea de producto por día).

27. Esquema en estrella

Hechos en el centro con dimensiones desnormalizadas.

28. Esquema snowflake

Dimensiones normalizadas con tablas adicionales.

29. Ventajas del esquema en estrella

- Simple
 - Alto rendimiento
 - Fácil para analistas
-

30. Desventajas del esquema snowflake

- Más joins
 - Consultas complejas
 - Menor rendimiento
-

31. Data Vault

Modelo flexible basado en hubs, links y satellites para cambios, auditoría y escalabilidad.

32. Hubs, links y satellites

- **Hub:** entidades clave del negocio.
- **Link:** relaciones entre hubs.
- **Satellite:** atributos y cambios históricos.

33. Cuándo elegir Data Vault

- Alta volatilidad del negocio
 - Auditoría estricta
 - Grandes entornos distribuidos
 - Necesidad de escalabilidad
-

34. Modelo orientado a eventos

Registra eventos como entidad principal (Event-Driven Architecture, Event Sourcing).

35. Anchor modeling

Modelo altamente normalizado basado en anchors (entidades mínimas), attributes y ties.

36. Modelar datos jerárquicos

Métodos:

- Adjacency list
 - Nested sets
 - Path enumeration
 - Closure tables
-

37. Modelar relaciones N:M

Crear una tabla puente con claves foráneas de ambas entidades.

38. Matriz de trazabilidad

Documento que relaciona requerimientos → entidades → atributos → procesos.

39. Convertir requerimientos en entidades

Identificar sustantivos clave y sus propiedades; validar relaciones y eventos.

40. Validar un modelo de datos

- Reglas del negocio
 - Trazabilidad
 - Ejemplos concretos
 - Normalización
 - Revisión con stakeholders
-

41. Integridad referencial

Garantiza relaciones válidas entre tablas (FK válidas).

42. Integridad de dominio y de entidad

- **Dominio:** valores permitidos.
 - **Entidad:** unicidad de una fila (PK).
-

43. Garantizar calidad de datos desde el modelado

- Constraints
 - Tipos correctos
 - Catálogos
 - Reglas de negocio
 - Validación de formatos
-

44. Diccionario de datos

Documento que describe entidades, atributos, tipos y reglas.

45. Glosario de negocio

Define términos del negocio para evitar ambigüedades.

46. Metamodelo de datos

Modelo que describe los componentes del modelo (entidades, relaciones, reglas).

47. Criterios para nombrar entidades y atributos

- Claridad
 - Singular
 - Estándares consistentes
 - Sin abreviaturas confusas
 - Idioma único
-

48. Modelo canónico

Estructura estándar para integrar múltiples sistemas mediante un formato común.

49. Diferencias entre OLTP y OLAP

- **OLTP:** transaccional, normalizado, operaciones rápidas.
 - **OLAP:** analítico, desnormalizado/dimensional, consultas pesadas.
-

50. Errores comunes en modelado y cómo evitarlos

- Mala granularidad → definir correctamente.
- Sobrenormalización → balancear.

- Sin glosario → crear definiciones comunes.
- Nombres poco claros → estándares.
- No considerar crecimiento → partición y diseño escalable.