

Guía de Preguntas - Gestión de Datos

Gestión y Arquitectura de Datos, Universidad de San Andrés

Si encuentran algún error en el documento o hay alguna duda, mandenme un mail a rodriguezf@udesa.edu.ar y lo revisamos.

Preguntas

1. En el contexto de calidad de datos, ¿cuál de las siguientes afirmaciones sobre la dimensión de “Consistencia” es correcta?
 - a) Significa que los datos son coherentes a través de diferentes sistemas y representaciones
 - b) Se refiere únicamente a que los datos estén completos
 - c) Implica que los datos sean coherentes solo dentro de una misma base de datos
 - d) Solo se aplica a datos numéricos
2. Una empresa implementa una arquitectura de datos moderna. ¿Cuál de las siguientes combinaciones de componentes sería la más apropiada para procesamiento en tiempo real y batch?
 - a) Data Lake + Data Warehouse
 - b) Solo Data Warehouse
 - c) Lambda Architecture + Stream Processing
 - d) Data Mart + Batch Processing
3. En el rol de Data Steward, ¿cuál de las siguientes NO es una responsabilidad principal?
 - a) Mantener los metadatos actualizados
 - b) Asegurar la calidad de los datos
 - c) Verificar el cumplimiento de políticas de datos
 - d) Definir la estrategia general de datos de la empresa

4. En el ciclo de vida de los datos, ¿cuál es el objetivo principal de la fase de archivado y eliminación?
 - a) Garantizar que los datos se conserven solo el tiempo necesario y se eliminen de forma segura
 - b) Mejorar la visualización de los datos
 - c) Aumentar la velocidad de procesamiento en tiempo real
 - d) Automatizar la captura de datos desde sensores
5. En el ciclo de vida de los datos, ¿qué fase debería incluir necesariamente validación de calidad?
 - a) Solo en la fase de Archivado
 - b) En la fase de Procesamiento (ETL)
 - c) Solo en la fase de Análisis
 - d) Solo en la fase de Creación
6. Para implementar una política efectiva de gobierno de datos, ¿qué combinación de elementos es más crítica?
 - a) Solo tecnología y herramientas
 - b) Únicamente automatización
 - c) Roles definidos + Políticas claras + Procesos documentados
 - d) Solo documentación técnica
7. ¿Cuál de las siguientes dimensiones de calidad de datos se enfoca en la frecuencia de actualización de la información?
 - a) Precisión
 - b) Actualidad
 - c) Unicidad
 - d) Accesibilidad
8. En el contexto de arquitectura de datos, ¿qué tipo de procesamiento sería más adecuado para analizar datos históricos de ventas mensuales?
 - a) Procesamiento en streaming

- b)* Procesamiento por lotes (Batch)
 - c)* Procesamiento en tiempo real
 - d)* Procesamiento híbrido
- 9. ¿Cuál de las siguientes métricas de calidad de datos mide la frecuencia de duplicados en un conjunto de datos?
 - a)* Tasa de error
 - b)* Tasa de duplicación
 - c)* Tasa de actualización
 - d)* Tasa de completitud
- 10. En el ciclo de vida de los datos, ¿qué fase incluye principalmente la transformación y limpieza de datos?
 - a)* Creación y Captura
 - b)* Almacenamiento
 - c)* Procesamiento
 - d)* Análisis
- 11. ¿Qué rol en el gobierno de datos es responsable de mantener los datos actualizados y consistentes a nivel operativo?
- 12. ¿Cuáles son las tres dimensiones de calidad de datos más críticas para garantizar la confiabilidad de un sistema de información?
- 13. ¿En qué se diferencia un Data Lake de un Data Warehouse en términos de almacenamiento y uso?
- 14. ¿Por qué es importante implementar políticas de retención de datos en una organización?
- 15. ¿Cuál es la diferencia entre procesamiento por lotes y procesamiento en streaming, y cuándo se usa cada uno?

Respuestas

1. En el contexto de calidad de datos, ¿cuál de las siguientes afirmaciones sobre la dimensión de “Consistencia” es correcta?

- a) **Significa que los datos son coherentes a través de diferentes sistemas y representaciones**
- b) Se refiere únicamente a que los datos estén completos
- c) Implica que los datos sean coherentes solo dentro de una misma base de datos
- d) Solo se aplica a datos numéricos

La consistencia NO es completitud (b) ni se limita a una base (c). Requiere coherencia entre múltiples sistemas y representaciones, no solo datos numéricos (d).

2. Una empresa implementa una arquitectura de datos moderna. ¿Cuál de las siguientes combinaciones de componentes sería la más apropiada para procesamiento en tiempo real y batch?

- a) Data Lake + Data Warehouse
- b) Solo Data Warehouse
- c) **Lambda Architecture + Stream Processing**
- d) Data Mart + Batch Processing

Data Lake/Warehouse (a) o solo Warehouse (b) NO manejan tiempo real eficientemente. Lambda Architecture sí combina ambos procesamiento, no solo batch (d).

3. En el rol de Data Steward, ¿cuál de las siguientes NO es una responsabilidad principal?

- a) Mantener los metadatos actualizados
- b) Asegurar la calidad de los datos
- c) Verificar el cumplimiento de políticas de datos
- d) **Definir la estrategia general de datos de la empresa**

Data Steward SÍ hace metadatos (a), calidad (b) y cumplimiento (c). Estrategia general (d) es responsabilidad del CDO, no del Steward.

4. En el ciclo de vida de los datos, ¿cuál es el objetivo principal de la fase de archivado y eliminación?
 - a) **Garantizar que los datos se conserven solo el tiempo necesario y se eliminen de forma segura**
 - b) Mejorar la visualización de los datos
 - c) Aumentar la velocidad de procesamiento en tiempo real
 - d) Automatizar la captura de datos desde sensores

El objetivo de archivado y eliminación es asegurarse de que los datos no se guarden más tiempo del necesario y que se borren de manera segura. No tiene que ver con visualización, velocidad ni captura.

5. En el ciclo de vida de los datos, ¿qué fase debería incluir necesariamente validación de calidad?
 - a) Solo en la fase de Archivado
 - b) **En la fase de Procesamiento (ETL)**
 - c) Solo en la fase de Análisis
 - d) Solo en la fase de Creación

Archivado (a), Análisis (c) y Creación (d) NO transforman datos. ETL es donde se valida y limpia antes del uso final.

6. Para implementar una política efectiva de gobierno de datos, ¿qué combinación de elementos es más crítica?
 - a) Solo tecnología y herramientas
 - b) Únicamente automatización
 - c) **Roles definidos + Políticas claras + Procesos documentados**
 - d) Solo documentación técnica

Solo tecnología (a), automatización (b) o documentación (d) NO bastan. Gobierno necesita roles, políticas y procesos integrados para funcionar correctamente.

7. ¿Cuál de las siguientes dimensiones de calidad de datos se enfoca en la frecuencia de actualización de la información?

- a) Precisión
- b) **Actualidad**
- c) Unicidad
- d) Accesibilidad

La actualidad se refiere a qué tan recientes son los datos. Precisión (a) es exactitud, unicidad (c) es ausencia de duplicados, accesibilidad (d) es facilidad de acceso.

8. En el contexto de arquitectura de datos, ¿qué tipo de procesamiento sería más adecuado para analizar datos históricos de ventas mensuales?

- a) Procesamiento en streaming
- b) **Procesamiento por lotes (Batch)**
- c) Procesamiento en tiempo real
- d) Procesamiento híbrido

Datos históricos mensuales no requieren tiempo real. Batch es ideal para grandes volúmenes sin urgencia inmediata. Streaming (a) y tiempo real (c) son para datos inmediatos.

9. ¿Cuál de las siguientes métricas de calidad de datos mide la frecuencia de duplicados en un conjunto de datos?

- a) Tasa de error
- b) **Tasa de duplicación**
- c) Tasa de actualización
- d) Tasa de completitud

La tasa de duplicación mide específicamente la frecuencia de registros duplicados. Tasa de error (a) mide errores generales, actualización (c) mide frecuencia de cambios, completitud (d) no es una métrica estándar.

10. En el ciclo de vida de los datos, ¿qué fase incluye principalmente la transformación y limpieza de datos?

- a) Creación y Captura
- b) Almacenamiento
- c) **Procesamiento**
- d) Análisis

Creación (a) es captura inicial, almacenamiento (b) es guardado, análisis (d) es uso final.

11. ¿Qué rol en el gobierno de datos es responsable de mantener los datos actualizados y consistentes a nivel operativo?

→ El Data Steward es el responsable operativo que mantiene los datos actualizados y consistentes, a diferencia del CDO que define estrategia o el Data Owner que tiene responsabilidad general.

12. ¿Cuáles son las tres dimensiones de calidad de datos más críticas para garantizar la confiabilidad de un sistema de información?

→ Precisión asegura exactitud, completitud evita datos faltantes, y consistencia garantiza coherencia entre sistemas.

13. ¿En qué se diferencia un Data Lake de un Data Warehouse en términos de almacenamiento y uso?

→ Data Lake almacena datos en formato original (estructurados y no estructurados) para análisis exploratorio. Data Warehouse almacena datos estructurados y procesados para análisis histórico y reportes.

14. ¿Por qué es importante implementar políticas de retención de datos en una organización?

→ Evita almacenar datos innecesarios, cumple leyes de privacidad y reduce riesgos de seguridad.

15. ¿Cuál es la diferencia entre procesamiento por lotes y procesamiento en streaming, y cuándo se usa cada uno?

→ Batch procesa grandes volúmenes sin urgencia (datos históricos). Streaming procesa datos inmediatamente (sensores, transacciones). Batch es más eficiente para análisis, streaming para tiempo real.