

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. ¿Qué desafío principal resuelve la arquitectura Data Mesh?

- a)* La centralización excesiva en equipos de datos
 - b)* La necesidad de procesamiento batch exclusivamente
 - c)* La falta de seguridad en los datos
 - d)* La imposibilidad de usar machine learning
- 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. En el contexto de protección de datos, ¿qué estrategia es más efectiva para datos sensibles en uso?
 - a)* Solo encriptación en reposo
 - b)* Únicamente control de acceso
 - c)* Backup diario
 - d)* Encriptación en tránsito + Enmascaramiento dinámico
- 8. Para mejorar la calidad de datos en tiempo real, ¿qué enfoque es más efectivo?
 - a)* Reglas automatizadas + Monitoreo continuo + Alertas
 - b)* Validación manual periódica
 - c)* Solo documentación

- d)* Revisión mensual
- 9. En una arquitectura moderna de datos, ¿qué característica es esencial para garantizar el linaje de datos?
 - a)* Solo logs de acceso
 - b)* Metadata activa + Tracking de transformaciones
 - c)* Únicamente documentación
 - d)* Backup semanal
- 10. Para implementar DataOps efectivamente, ¿qué conjunto de prácticas es más importante?
 - a)* Solo testing manual
 - b)* Únicamente documentación
 - c)* Automatización + CI/CD + Monitoreo + Colaboración
 - d)* Reuniones diarias

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. ¿Qué desafío principal resuelve la arquitectura Data Mesh?

- a) **La centralización excesiva en equipos de datos**
- b) La necesidad de procesamiento batch exclusivamente
- c) La falta de seguridad en los datos
- d) La imposibilidad de usar machine learning

Data Mesh NO resuelve batch (b), seguridad (c) o ML (d). Su foco es descentralizar equipos de datos mediante dominios autónomos.

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.

7. En el contexto de protección de datos, ¿qué estrategia es más efectiva para datos sensibles en uso?

- a) Solo encriptación en reposo
- b) Únicamente control de acceso
- c) Backup diario
- d) **Encriptación en tránsito + Enmascaramiento dinámico**

Encriptación en reposo (a), control de acceso (b) o backup (c) NO protegen datos EN USO. Tránsito + enmascaramiento sí.

8. Para mejorar la calidad de datos en tiempo real, ¿qué enfoque es más efectivo?

- a) **Reglas automatizadas + Monitoreo continuo + Alertas**
- b) Validación manual periódica
- c) Solo documentación
- d) Revisión mensual

Validación manual (b), documentación (c) o revisión mensual (d) NO detectan problemas en tiempo real. Automatización + monitoreo sí.

9. En una arquitectura moderna de datos, ¿qué característica es esencial para garantizar el linaje de datos?

- a) Solo logs de acceso
- b) **Metadata activa + Tracking de transformaciones**
- c) Únicamente documentación
- d) Backup semanal

Logs (a), documentación (c) o backup (d) NO rastrean transformaciones. Metadata activa + tracking sí registra origen y cambios.

10. Para implementar DataOps efectivamente, ¿qué conjunto de prácticas es más importante?

- a) Solo testing manual
- b) Únicamente documentación
- c) **Automatización + CI/CD + Monitoreo + Colaboración**
- d) Reuniones diarias

Testing manual (a), documentación (b) o reuniones (d) NO escalan.
DataOps requiere automatización, CI/CD y monitoreo para agilidad.