

# Guía de Preguntas - Gestión de Datos

Gestión y Arquitectura de Datos, Universidad de San Andrés

## Preguntas

1. En el contexto de calidad de datos, ¿cuál de las siguientes afirmaciones sobre la dimensión de “Consistencia” es correcta?
  - a) Se refiere únicamente a que los datos estén completos
  - b) Implica que los datos sean coherentes solo dentro de una misma base de datos
  - c) Significa que los datos son coherentes a través de diferentes sistemas y representaciones
  - 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) Lambda Architecture + Stream Processing
  - c) Solo Data Warehouse
  - d) Data Mart + Batch Processing
3. En el rol de Data Steward, ¿cuál de las siguientes NO es una responsabilidad principal?
  - a) Definir la estrategia general de datos de la empresa
  - b) Mantener los metadatos actualizados
  - c) Asegurar la calidad de los datos
  - d) Verificar el cumplimiento de políticas de datos
4. ¿Qué desafío principal resuelve la arquitectura Data Mesh?
  - a) La necesidad de procesamiento batch exclusivamente
  - b) La centralización excesiva en equipos de datos

- 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)* Solo en la fase de Análisis
  - c)* En la fase de Procesamiento (ETL)
  - 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)* Roles definidos + Políticas claras + Procesos documentados
  - c)* Únicamente automatización
  - 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)* Encriptación en tránsito + Enmascaramiento dinámico
  - c)* Únicamente control de acceso
  - d)* Backup diario
- 8. Para mejorar la calidad de datos en tiempo real, ¿qué enfoque es más efectivo?
  - a)* Validación manual periódica
  - b)* Reglas automatizadas + Monitoreo continuo + Alertas
  - 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)* Autom
  - c)* Únicamente documentación
  - 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) Se refiere únicamente a que los datos estén completos
  - b) Implica que los datos sean coherentes solo dentro de una misma base de datos
  - c) **Significa que los datos son coherentes a través de diferentes sistemas y representaciones**
  - d) Solo se aplica a datos numéricos

La consistencia implica que los datos sean coherentes a través de diferentes sistemas y representaciones. Esto es fundamental para mantener la integridad de la información en toda la organización.

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

La Lambda Architecture combinada con Stream Processing permite manejar tanto procesamiento batch como en tiempo real, ofreciendo una solución completa para diferentes necesidades de procesamiento.

3. En el rol de Data Steward, ¿cuál de las siguientes NO es una responsabilidad principal?
  - a) **Definir la estrategia general de datos de la empresa**
  - b) Mantener los metadatos actualizados
  - c) Asegurar la calidad de los datos
  - d) Verificar el cumplimiento de políticas de datos

Definir la estrategia general de datos es responsabilidad del Chief Data Officer (CDO), no del Data Steward, quien se enfoca en la calidad, metadatos y cumplimiento de políticas.

4. ¿Qué desafío principal resuelve la arquitectura Data Mesh?

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

Data Mesh aborda el problema de la centralización excesiva al proponer una arquitectura descentralizada basada en dominios, permitiendo mayor autonomía y escalabilidad.

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) Solo en la fase de Análisis
- c) **En la fase de Procesamiento (ETL)**
- d) Solo en la fase de Creación

La fase de Procesamiento (ETL) es crítica para la validación de calidad, ya que es donde se transforman y limpian los datos antes de su uso.

6. Para implementar una política efectiva de gobierno de datos, ¿qué combinación de elementos es más crítica?

- a) Solo tecnología y herramientas
- b) **Roles definidos + Políticas claras + Procesos documentados**
- c) Únicamente automatización
- d) Solo documentación técnica

La combinación de roles definidos, políticas claras y procesos documentados es esencial para un gobierno de datos efectivo, ya que abarca tanto el aspecto humano como el organizacional.

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

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

La combinación de encriptación en tránsito y enmascaramiento dinámico proporciona una protección completa para datos sensibles mientras están siendo utilizados.

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

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

Las reglas automatizadas, el monitoreo continuo y las alertas permiten detectar y corregir problemas de calidad en tiempo real de manera eficiente.

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

La metadata activa combinada con el tracking de transformaciones permite mantener un registro completo del origen y las transformaciones de los datos.

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

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

La combinación de automatización, CI/CD, monitoreo y colaboración es fundamental para una implementación efectiva de DataOps, permitiendo una gestión ágil y confiable de los datos.