

Instrucciones:

- Si el candidato es empleado: 12 días de plazo para entregar.
- Si el candidato no lo es: 7 días.
- Publicar la prueba en un repositorio público.
- Cualquier duda o inquietud sobre la prueba, escribir al correo sebastian.granda@jikkosoft.com o diego.trujillo@jikkosoft.com

Sección 1: Conocimiento teórico

1. Explique las diferencias entre los sistemas OLTP y OLAP. Proporcione ejemplos de escenarios de uso y cómo optimizar cada tipo de sistema.
2. Describa el concepto de normalización de datos. Incluya ejemplos de problemas que pueden surgir al aplicar diferentes niveles de normalización y cómo resolverlos.
3. Describa el concepto de desnormalización de datos. Incluya ejemplos de problemas que pueden surgir al aplicar diferentes niveles de normalización y cómo resolverlos.
4. ¿Qué es un esquema de estrellas y cómo se diferencia de un esquema de copo de nieve? Diseñe un ejemplo concreto y compare las implicaciones de rendimiento y mantenimiento.
5. Explique el concepto de un almacén de datos y cómo se diferencia de una base de datos. Describa cómo manejaría una migración de una base de datos transaccional a un almacén de datos.
6. Describa el teorema de CAP y sus implicaciones para las bases de datos distribuidas.
7. ¿Cuáles son las propiedades de ACID en un sistema de base de datos? Proporcione ejemplos de cómo estas propiedades se aplican y se garantizan en bases de datos distribuidas.
8. Explique el término “ETL” y “ELT”.

9. ¿Qué es un lago de datos y cómo se diferencia de un almacén de datos? Proporcione ejemplos de casos de uso.

Sección 2: Habilidades prácticas

A. Escritura de consultas SQL:

1. Escriba una consulta SQL para encontrar los 5 principales clientes por ingresos del último año en la tabla de ventas.
2. Escriba una consulta SQL para recuperar el segundo salario más alto de la tabla de empleados.
3. Optimice una consulta SQL y explique las mejoras de rendimiento.

B. Diseño de la base de datos:

1. Diseñe un esquema de base de datos normalizado para una librería en línea. Incluya tablas para libros, autores, editores, clientes y pedidos. Nota: se aprecia en gran medida incluir requisitos funcionales adicionales como manejo de stock y seguimiento de envíos.
2. Explique sus decisiones de diseño y cree un diagrama de ER para el esquema diseñado.

C. Modelado de datos:

1. Desarrolle un esquema de estrellas para un almacén de datos de ventas al por menor. Identifique la tabla de hechos y las tablas de dimensiones, incluyendo múltiples tablas de hechos y dimensiones compartidas.

D. Transformación de datos:

1. Proporcione un flujo de trabajo de muestra de ETL para extraer los datos de los clientes de una base de datos transaccional, transformarlos mediante la limpieza y la agregación, y cargarlos en un almacén de datos.

Sección 3: Estudio de caso

***Estudio de caso 1: Integración de datos** Se le asigna la tarea de integrar datos de múltiples fuentes (por ejemplo, bases de datos transaccionales, API de terceros y archivos planos) en un almacén de datos central. Describa los pasos que tomaría para garantizar la calidad, la consistencia y la integridad de los datos durante todo el proceso. Incluya las herramientas y tecnologías que usaría. Nota: se aprecia en gran medida incluir un plan detallado con diagramas de flujo de datos.

***Estudio de caso 2: Optimización del rendimiento** Una empresa está experimentando un rendimiento de consulta lento en su almacén de datos. Describa su enfoque para diagnosticar y resolver los problemas de rendimiento. ¿Qué técnicas y herramientas utilizaría para optimizar el rendimiento del almacén de datos?

Sección 4: Preguntas basadas en escenarios

1. Su empresa está migrando de un almacén de datos local a una solución basada en la nube. ¿Qué factores tendría en cuenta al elegir una plataforma de almacenamiento de datos en la nube? Proporcione una comparación de al menos dos soluciones populares de almacenamiento de datos en la nube. Incluya requisitos específicos, como cumplimiento de normativas, costos y planes de contingencia.
2. Necesita diseñar una arquitectura de datos para un sistema de análisis en tiempo real que supervise los dispositivos IoT. Describa la arquitectura que propondría, incluidos los componentes de ingestión, almacenamiento, procesamiento y visualización de datos.
Nota: se aprecia en gran medida incluir detalles sobre el volumen de datos y latencia requerida.
3. Explique cómo manejaría la privacidad y la seguridad de los datos en una arquitectura de datos que incluya información confidencial de los clientes. ¿Qué medidas pondría en marcha para proteger los datos? Incluya un plan de cumplimiento.
4. Una unidad de negocio quiere implementar una nueva herramienta de BI que actualmente no cuenta con el apoyo del departamento de TI. ¿Cómo evaluaría y potencialmente integraría esta herramienta en la arquitectura de datos existente?

Sección 5: Habilidades blandas y liderazgo

1. Describa un momento en el que tuvo que dirigir un equipo multifuncional para entregar un proyecto de datos. ¿A qué desafíos se enfrentó y cómo los superó?
2. ¿Cómo se mantiene actualizado con las últimas tendencias y tecnologías en arquitectura de datos y gestión de datos? Explique cómo aplicó una tecnología o tendencia reciente en un proyecto real.

3. Proporcione un ejemplo de un problema complejo que resolvió en su papel anterior como arquitecto de datos. Explique su enfoque y el resultado, incluyendo un problema técnico específico relacionado con arquitectura de datos.

Sección 6: Codificación y secuencias de comandos

1. Escriba un script de Python para conectarse a una base de datos y recuperar datos de una tabla. Maneje errores y asegure conexiones. Incluir operaciones adicionales, como inserción y actualización de datos.
2. Cree un script ETL simple utilizando un lenguaje de su elección (por ejemplo, Python, Java) para extraer datos de un archivo CSV, transformarlo eliminando duplicados y cargarlo en una base de datos. Incluya transformaciones más complejas, como mapeo de campos y agregaciones.