**PREGUNTA 1 (5 puntos)**

**Describa el ciclo de vida de la ingeniería de datos. Sea lo más detallado que pueda.**

El ciclo de vida de la ingeniería de datos es el proceso de crear datos y convertirlos en una forma útil para personas como los científicos de datos. El ciclo de vida tiene cinco etapas: captura, almacenamiento, ingesta, transformación y servicio de datos.

**Captura**

En esta etapa, los datos se obtienen de fuentes externas, como los sensores. Los datos pueden ser generados por la organización, sus clientes o terceros. Estos datos son valiosos para mejorar el servicio al cliente y realizar tareas de manera más efectiva. Los datos pueden ser creados o recolectados.

**Almacenamiento**

Para trabajar con datos, primero deben ser almacenados. Al elegir un sistema de almacenamiento, es importante considerar la disponibilidad y la preservación de los datos a lo largo del tiempo.

**Ingesta**

En esta etapa, se preparan los datos para su análisis. Se utilizan métodos para asegurar que los analistas no tengan que preocuparse por la calidad de los datos. Estos métodos incluyen la compresión para almacenar datos de manera más eficiente y el cifrado para proteger la privacidad de los datos.

**Transformación**

En esta etapa, los datos se convierten a un formato adecuado para análisis, aprendizaje automático y otras tareas. Esto incluye corregir tipos de datos, manejar datos faltantes y normalizar los datos.

**Servicio**

La última etapa es usar los datos para tareas prácticas como análisis y aprendizaje automático. Los datos solo tienen valor cuando se utilizan activamente. Almacenar datos sin usarlos es un riesgo para las empresas. Es importante entender el propósito final de los datos en cada etapa.

**Procesos transversales**

En cada etapa del ciclo de vida, hay procesos que aseguran la calidad, seguridad y buen manejo de los datos, como por ejemplo asegurarse de que los datos sean precisos, completos, consistentes y actuales. Esto incluye controles de calidad, validación y limpieza de datos para eliminar errores. Otro proceso es proteger los datos contra accesos no autorizados, pérdida o daño. Esto incluye cifrado de datos, control de acceso y registro de actividades para asegurar que solo personas autorizadas accedan a los datos.

Además**,** se establecen reglas y procedimientos para manejar los datos de manera efectiva. Esto incluye definir roles y responsabilidades, gestionar metadatos, cumplir con regulaciones y supervisar el uso y la calidad de los datos.

Incluso se puede mantener un registro detallado de todos los procesos, fuentes de datos, transformaciones y flujos de datos para asegurar que todo pueda ser revisado, y también usar herramientas y tecnologías para automatizar procesos repetitivos y monitorear el desempeño del sistema. Esto incluye pipelines de datos automatizados, alertas en tiempo real y paneles de control para el seguimiento continuo de los datos.

Estos procesos transversales son esenciales para asegurar que el ciclo de vida de los datos sea eficiente y cumpla con los requisitos de la organización y las regulaciones.