

Tarea 1. Conceptos básicos

1. Conceptos generales:

- a. Explica cuatro diferencias entre un sistema de **procesamiento de archivos** y un **SMBD**.

Refiriéndonos al procesamiento de archivos existen estas diferencias contra el SMBD:

- El almacenamiento de datos: El sistema de archivos limita el tamaño de éstos, además los datos pueden perderse cuando el sistema se cae y la seguridad es insuficiente.
- La consulta o actualización: Se necesita escribir un nuevo programa para cada nueva consulta y es necesario implementar políticas para lograr un buen rendimiento.
- La concurrencia proporciona una protección limitada: Los usuarios no pueden interferir entre sí y no se puede proporcionar diferentes vistas de los mismos datos.
- Al cambiar el esquema se requiere cambiar casi todas las aplicaciones.

- b. ¿Qué **desventajas** encuentras al trabajar con una **base de datos**?

- Se requieren mayores conocimientos para su uso.
- Mayores costos tanto para su manejo como el hardware que se necesita.
- Gente especializada en el tema.

- c. Explica las diferencias entre independencia de datos **física** y **lógica**. ¿Cuál es más difícil de lograr? Justifica tu respuesta.

El nivel lógico o conceptual es la descripción básica, única y global, que sirve de referencia para el resto de los esquemas, es decir cómo se almacenan, mantienen y acceden los datos. Mientras el nivel físico, lo denominamos esquema interno.

Sabiendo lo anterior tenemos que la independencia física es cuando los cambios en la organización física de la BD no afectan al mundo exterior. Mientras la independencia lógica es cuando los usuarios no se ven afectados por los cambios en el nivel lógico tales como cambios en el esquema conceptual o cambios en los esquemas externos.

El más difícil de lograr es la independencia lógica, ya que el almacenamiento en disco es más fácil de modificar.

- d. ¿Cuál es la importancia del **diccionario de datos**?

Nos provee de un repositorio centralizado de información sobre datos tales como significado, relación con otros datos, origen, uso y formato.

Los cuales pueden ser usados para las siguientes razones:

- i. Para manejar los detalles en sistemas grandes.
- ii. Para comunicar un significado común para todos los elementos del sistema.
- iii. Para documentar las características del sistema.
- iv. Para facilitar el análisis de los detalles con la finalidad de evaluar las características y determinar donde efectuar cambios en el sistema.
- v. Localizar errores y omisiones en el sistema.

e. ¿Qué es un **modelo de datos** y qué características tiene?

Es un conjunto de conceptos que puede usarse para describir la estructura de una BD. Además proporciona tres tipos de herramientas:

- Estructuras de datos para construir la BD.
- Restricciones de integridad.
- Operaciones para trabajar con los datos.

f. Indica las características principales de los **SMBD**

Es un software que facilita los siguientes procesos:

- Definir: involucra especificar los tipos, estructuras y restricciones de los datos almacenados en la BD.
- Construir: Almacena los datos en algún medio de almacenamiento controlado por el SABD.
- Manipular: incluye funciones tales como consultas, recuperación y actualización de datos.
- Compartir: permite el acceso a más de un usuario.

Además su funcionalidad es:

- Lenguaje para definición de datos DDL.
- Lenguaje para manipulación de datos DML.
- Lenguaje para consulta de datos QL.
- Administración de almacenamiento.
- Administración de consultas.
- Administración de transacciones, tal como control de concurrencia o recuperación.

g. Explica cuál(es) de los siguientes conceptos juegan un papel importante en la representación de los datos acerca de un problema real en una base de datos y porqué: **lenguaje para definición de datos, lenguaje para manipulación de datos, el modelo de datos, manejo de transacciones.**

El que representa un papel importante es el modelo de datos, ya que nos permite describir la estructura de una Base de Datos. En otras palabras poder definir el problema a resolver.

- h. Supón que una pequeña compañía desea almacenar su información en una base de datos. Desea comprar la que tenga la menor cantidad de características posibles, se desea ejecutar la aplicación en una sola computadora personal y no se planea compartir la información con nadie. Para cada una de las siguientes características explica por qué se debería o no incluir en la base de datos que se desea comprar (suponiendo que se pueden comprar por separado): **Seguridad, Control de concurrencia, Recuperación en caso de fallas, lenguaje de consulta, mecanismo de vistas, manejo de transacciones.**

- Seguridad: Es necesario ya que se necesita proteger los datos contra usos no autorizados.
- Control de concurrencia: No es necesario ya que no se requiere compartir información con nadie.
- Recuperación en caso de fallas: Siempre es necesario tener un medio de recuperación.
- Lenguaje de consulta: Es necesario ya que se requiere consultar la información almacenada.
- Mecanismo de vistas: No es necesario ya que con consultas es suficiente.
- Manejo de transacciones: Es necesario, ya que si llega a haber algún error queremos mantener la integridad de los datos.

2. Investigación

- a. ¿Qué se entiende por **Calidad de Datos** y **Gobierno de Datos**?

Se deben cumplir las siguientes seis dimensiones que son puntos clave para asegurar nuestros procesos de limpieza y calidad.

- Completitud: Los datos que no están son irrelevantes, pero cuando se vuelven necesarios para un proceso, estos se vuelven críticos.
- Conformidad: Los datos que están en los campos de la tabla, deben estar en un formato estándar y legible.
- Consistencia: Al hacer el cruce de información con los registros, se debe evitar la información contradictoria.
- Precisión/Exactitud: Si los datos no son precisos, estos no pueden ser utilizados. En este sentido, para detectar si estos son precisos, se compara el dato con una fuente de referencia.
- Duplicación: Es importante saber si se tiene la misma información en formatos iguales o similares dentro de la tabla.
- Integridad: Otra dimensión de calidad importante radica en el hecho de saber si toda la información relevante de un registro está presente de forma que se pueda utilizar.

El gobierno de datos consiste en la capacidad de una organización para gestionar el conocimiento que tiene sobre su información de forma que pueda responder a preguntas tales como ¿qué sabemos sobre nuestra información?, ¿de dónde provienen esos datos?, ¿están estos datos alineados con nuestra política de empresa?. El gobierno de datos proporciona un enfoque holístico para administrar, mejorar y

aprovechar la información de forma que pueda ayudarnos a ganar percepción y generar confianza en decisiones y operaciones empresariales.

- b. Especifica las características más importantes de las bases de datos **NoSQL**, indica el modelo de datos que utilizan y principales proveedores.

NoSQL es un término que describe las bases de datos no relacionales de alto desempeño. Las bases de datos NoSQL utilizan varios modelos de datos, incluidos los de documentos, gráficos, claves-valores y columnas. Las bases de datos NoSQL son famosas por la facilidad de desarrollo, el desempeño escalable, la alta disponibilidad y la resiliencia.

El modelo que usan estas bases de datos no relacionales (NoSQL) no suelen contener un esquema. Se suele utilizar una clave de partición para recuperar valores, conjuntos de columnas o documentos JSON o XML semiestructurados, así como otros documentos que contengan atributos de elementos relacionados.

Los principales proveedores son MongoDB, DynamoDB, Cassandra y Couchbase.

- c. Indica las características de una **base de datos Objeto – Relacional**.

Es una extensión de la base de datos relacional tradicional, a la cual se le proporcionan características de la programación orientada a objetos.

Se le agregan las características de:

- Relaciones anidadas
- Datos complejos
- Herencia
- Funciones