**Desarrollar**

1. Explicar los conceptos de Data Warehouse y Data Mining y el objetivo de utilizar cada uno de ellos.

*DW: son sistemas donde se almacenan y extraen datos de diversas fuentes para la toma de decisiones.*

*DM: es la extracción de información oculta y predecible de grandes BDs, ayuda a las compañías a concentrarse en la información más importantes de sus bases y predecir nuevas tendencias.*

1. Explique y desarrolle los diferentes niveles de aislamiento de una base de datos relacional.

*Niveles de aislamiento:*

* *Read uncommitted: no asegura lockeos por select, lo que mejora el rendimiento pero afecta la integridad porque hay lecturas sucias, lecturas no repetibles y lecturas fantasmas.*
  + - *no lockeos por select => mejor rendimiento*
    - *si lecturas sucias*
    - *si lectura no repetibles*
    - *si lectura fantasma fantasmas*
* *Read committed: asegura que no exista lecturas sucias pero no asegura lecturas repetibles, ya que una vez que leyó los datos, libera el lockeo. En una misma transacción puede tener dos llamados a un mismo select y este arrojar resultados distintos.*
  + - *no lecturas sucias*
    - *si lecturas no repetibles*
    - *si lectura fantasma*
    - *lee los datos y libera el lockeo*
    - *dos llamados en una misma transacción puede tener resultados distintos*
* *Repeatable read: asegura que no existan lecturas sucias y que las lecturas puedan ser repetibles, pero no evita lecturas fantasmas.*
  + - *no lecturas sucias*
    - *no lectura no repetible*
    - *si lectura fantasma*
* *Serializable read: asegura que no existan lecturas sucias, lecturas fantasmas y que las lecturas puedan ser repetibles. El problema es que se implementa un nivel de bloqueo que puede afectar a los demás usuarios.*
  + - *no lectura sucia*
    - *no lectura no repetible*
    - *no lectura fantasma*
    - *tantos bloqueos pueden afectar a los demás usuarios*

1. Explique las propiedades ACID de una BD y su relación con las transacciones.

*Las propiedades ACID garantizan que las transacciones se realicen de forma*

*confiable en una BD.*

* *Atomicidad: la transacción debe ejecutarse en su totalidad o no debe ejecutarse en absoluto.*
* *Consistencia: luego de la transacción , la BD debe quedar consistente satisfaciendo todas las restricciones de integridad.*
* *Aislamiento (isolation): cuando las transacciones se ejecutan concurrentemente, sus efectos deben ser aislados y no deben interferirse entre sí.*
* *Durabilidad: una vez que la transacción termina, su efecto no debe perderse aunque el sistema falle.*

1. Explique y desarrolle 3 objetos de BD que sirven para asegurar la Integridad de datos

*Objetos relacionados con la integridad:*

* *Constraints (unique, not null, check, default)*
* *Triggers*
* *Índices*
* *Vistas (con check options)*
* *Stores procedures*

1. Defina el concepto de transacción, cite un ejemplo de cuando el motor de BD ejecutaría un Rollback en una transacción.

*Transacción es un conjunto de instrucciones las cuales deben ejecutarse todas*

*o ninguna, manteniendo la integridad de los datos.*

*Un Rollback puede ejecutarse en el caso de que se quiera borrar una fila que*

*está relacionada con otra tabla por medio de una FK, por lo cual fallaría. Esto*

*en el caso que no esté configurada en “Delete on cascade”.*

1. Mencione dos objetos que tengan que ver con la seguridad, descríbalos e indique de qué modo puede utilizarlos para dicha funcionalidad.

*Objetos relacionados con la seguridad:*

* *Vistas: es una consulta que se presenta como una tabla (virtual). Se puede por ej, para algunos usuarios crear una vista de una tabla donde obtenga solo parte de las cantidad real de columnas que tiene la tabla.*
* *Triggers: son objetos que se relacionan a tablas, y permiten administrar mejor la BD. Se puede por ej, crear un trigger en una tabla que ante un delete de n filas, no haga caso al delete, o escriba un histórico.*
* *Stores procedures: es un conjunto de instrucciones que se almacenan y ejecutan en la BD. Se puede por ej, crear un SP para obtener un listado de los usuarios que accedieron a determinadas tablas, horarios, etc.*

1. Defina el concepto de lectura sucia, repetible y fantasma.

* *Lectura sucia: ocurre cuando se le permite a una transacción hacer una lectura de una fila que ha sido modificada por otra transacción concurrente, pero aún no ha sido confirmada (commiteada).*
* *Lectura repetible: ocurre cuando en el curso de la transacción se lee una fila dos veces, y los valores coinciden.*
* *Lectura no repetible: ocurre cuando en el curso de la transacción se lee una fila dos veces, y los valores no coinciden.*
* *Lectura fantasma: ocurre cuando durante una transacción se ejecutan dos consultas idénticas y los resultados de la segunda son distintos a los de la primera.*

1. Que es un Data Mart, explique por qué motivo posee datos desnormalizados.

*Un Data Mart es un subconjunto de un Data Warehouse, con el propósito de*

*ayudar en un área específica dentro del negocio para que se puedan tomar*

*mejores decisiones. Los datos en un Data Mart están desnormalizados para*

*disminuir la cantidad de tablas necesarias.*

1. Explique el concepto de transacción y su relación con las propiedades ACID.

*Una transacción es un conjunto de órdenes que se ejecutan atómicamente.*

*Las propiedades ACID garantizan que las transacciones se realicen de forma*

*confiable en una BD. (ver 3)*

1. Explique Algoritmo de Huffman, en que se basa para la compresión de datos.

*Es un algoritmo que puede ser usado para compresión o encriptación de datos. Por ej: se usa cuando la cantidad de espacio en disco es insuficiente o los tiempos de transmisión son prolongados con costos elevados.*

*Huffman se basa en asignar códigos de distinta longitud de bits a cada uno de los caracteres de un archivo. Si se asignan códigos más cortos a los caracteres que aparecen más a menudo se consigue una compresión del archivo.*

*Esta compresión es mayor cuando la variedad de caracteres diferentes que aparecen es menor. Ej: si el texto se compone únicamente de números o mayúsculas, se conseguirá una compresión mayor.*

1. Explique las Reglas de Integridad según el Modelo Relacional

* *Regla de integridad de la entidad: ninguno de los atributos que componen la clave primaria puede ser nulo.*
* *Regla de la integridad referencial: si en una relación hay una clave foránea, sus valores deben coincidir con los valores de la clave primaria.*

1. Explicar Clave Primaria, Foránea, Candidata y dominio de los datos.

* *Clave Primaria: es un campo o combinación de campos que identifica de forma única a cada fila de la tabla.*
* *Clave Foránea: identifica una columna o grupo de columnas en una tabla que se refiere a una columna o grupo de columnas en otra tabla.*
* *Clave Candidata: son claves que podrían servir como clave primaria.*
* *Dominio de los Datos: un dominio describe un conjunto de posibles valores para cierto atributo.*

1. Desarrolle el concepto de restore y recovery.

*Restore: es la acción de tomar un Back Up ya realizado y restaurar la*

*estructura y datos sobre una base dada.*

*Recovery: es un mecanismo provisto por los motores de base de datos que se ejecuta en cada inicio del motor de forma automática como dispositivo de tolerancia de fallas. Sus objetivos son , retornar el el motor al punto consistente más reciente y utilizar logs transaccionales para retornar al motor a un estado lógico consistente.*

1. Describa los diferentes conceptos que se relacionan directamente con la funcionalidad de concurrencia.

*Los motores permiten controlar elacceso concurrente a sus recursos a travez de loqueos y de niveles de aislamiento.*

1. Describa ventajas y desventajas de cada uno de los métodos de creación de índices, Arbol B y Tablas de Hashing.

*Hashing: busca establecer una relación directa entre el valor de los datos y el*

*valor de la clave con una función hash, que aplicada a una clave devuelve el*

*subíndice de la tabla.*

*Ventajas: acceso directo a los datos (más rápido que Árbol B para acceso a datos). No se utiliza espacio extra para su implementación.*

*Desventajas: el principal problema es cuando la función hash devuelve un valor igual para dos o más claves, lo que se llama colisión. Cuanto más datos, más colisiones. No es bueno para búsquedas secuenciales.*

*Árbol B: la estructura de Árbol B parte del concepto de árboles n-arios de búsqueda. Está pensado para disminuir la cantidad de accesos a disco y la posibilidad de mantener en memoria la parte que se está utilizando y el resto conservarlo en disco.*

*Ventajas: se utiliza para grandes volúmenes de datos y es mejor que hash para archivos secuenciales.*

*Desventajas: es más lento que Hashing para la creación de índices, dado que es necesario crear toda la estructura en memoria.*

1. Explicar objetos de BD que nos permiten la integridad referencial, reglas de integridad de entidad y la integridad semántica.

*Integridad semántica = Consistencia.*

1. Explique el concepto de Fact table y Dimensiones en un Data Warehouse.

*Fact table: es la tabla primaria en cada modelo dimensional, orientada a*

*contener reglas de negocio. Cada Fact table representa una relación de*

*muchos a muchos, y contiene un conjunto de dos o más foreings keys que*

*hacen referencia a sus respectivas tablas de dimensión.*

*Dimensiones en DW: llamamos dimensiones a aquellos datos que nos*

*permiten agrupar o filtrar información. Existe una Dimensión Table que*

*restringe los criterios de selección de la Fact table. Cada dimensión está*

*definida por su clave primaria que sirve como base para la integridad*

*referencial con la fact table.*

1. Explique la relación existente entre la Constraints UNIQUE y PRIMARY KEY y el Objeto índice.

*Constraint Unique: las restricciones son reglas que el motor aplica de forma*

*automática la Unique es para garantizar que todos los valores de una columna*

*sean únicos. La restricción Check valida que el dato de la columna esté dentro de un rango.*

*Primary Key:es un campo o combinación de campos que identifica de forma*

*única a cada fila de la tabla.*

*Objeto Índice: es una estructura de datos que mejora la velocidad de la*

*operaciones por medio de un identificador único de cada fila de la tabla,*

*permitiendo un rápido acceso a los registros de una BD.*

*Entonces => La PK es el identificador de la tabla que puede estar formado por*

*uno o más campos y no se puede repetir, la Constraint Unique se aplica sobre*

*una columna cualquiera para que no se repitan los datos y el índice crea ???*

1. Cómo implementaría la integridad referencial entre dos tablas de dos bases de datos en diferentes servidores.
2. Diferencias entre Store procedure, Función y Trigger.

*Los SPs son procedimientos almacenados en una BD, los cuales son ejecutados por un usuario o por otro proceso, y pueden realizar operaciones varias como altas, bajas y modificaciones de tablas. En cambio las Funciones no pueden hacer modificaciones de tablas, solo hacer consultas y retornar un valor. Los trigger se ejecutan automáticamente ante eventos (insert, delete, update) en las tablas y al igual que en los SPs, puede hacerce operaciones varias.*

1. Explique las principales características que diferencian a una Base de datos Operacional de un Datawarehouse.

**Preguntas más viejas**

1. Determine los componentes y las variables que intervienen en el diseño de un modelo OLAP.
   * Proporciona información fiable sobre los indicadores clave del funcionamiento de una organización para los sectores que toman decisiones.
   * A diferencia de los sistemas transaccionales, OLAP procesa información y aplica inteligencia.
   * Un sistema OLAP se alimenta de uno o varios OLTP que lo proveen de datos.
   * Todos los campos son calculados.
   * Se usa para tomar decisiones.
   * Se basa en patrones de interés.
   * OLTP → DTS → OLAP. El DTS se ejecuta periódicamente para generar información.
   * Se implementa con distintas tecnologías: MOLAP (bases de datos multidimensionales), ROLAP (bases de datos relacionales), HOLAP (bases de datos híbridas).
   * MOLAP: ya no se trabaja con el concepto de filas y columnas sino dimensiones. Técnicas de hipercubo (hay dispersión) y multicubo (se saltan los cubos que no contienen información).
   * ROLAP
2. Que características de funcionamiento tiene y que variantes existen del método quicksort de clasificación.

Es un método de fácil implementación y poco consumo de recursos.

Es recursivo pero se usan versiones iterativas para mejorar su rendimiento, también es in-situ, ya que usa solo una pila auxiliar.

El QuickSort está basado en la idea de divide y vencerás, en el cual un problema se soluciona dividiéndolo en dos o más subproblemas, resolviendo recursivamente cada uno de ellos para luego juntar sus soluciones para obtener la solución del original.

1. Indique por que un índice conformado por una clave numérica corta es más rápido en el acceso a los datos que un índice compuesto o conformado con un string largo.

Al ocupar más espacio cada nodo del árbol supera el tamaño de una página del File System, de esta forma el Sistema Operativo debe leer varias páginas para levantar un nodo, por eso es mas lento.

1. Desarrolle el concepto de trigger en cuanto a ejecución, instancias y funcionalidad.
   * Ejecución: un trigger es un procedimiento que se ejecuta ante un acontecimiento (INSERT, UPDATE, DELETE) sobre una tabla determinada.
   * Tipos: se puede aplicar a la fila que disparó el trigger o a todas las filas.
   * Before, After & Instead of (en lugar del evento que lo invocó).
   * Atomicidad: si un error ocurre cuando un trigger se está ejecutando, la operación que disparó el trigger falla, o sea que no se modifica la tabla.
   * Uso: se usan triggers cuando la integridad referencial y los constraints son insuficientes; reglas de consistencia (no provistas por el modelo relacional); replicación de datos; auditoría; acciones en cascada; autorización de seguridad; los triggers constituyen la herramienta más potente para el mantenimiento de la integridad de la base de datos, ya que pueden llevar a cabo cualquier acción que sea necesaria para mantener dicha integridad; un trigger puede modificar filas de una tabla que un usuario no puede modificar directamente; pueden llamar procedimientos y disparar otros triggers, pero no pueden llevar parámetros.
   * Ventaja: la principal ventaja es que permiten a los usuarios crear y mantener un conjunto de código más manejable para su empleo por todas las aplicaciones asociadas con las base de datos existentes y futuras.
   * Tablas virtuales: acceso a tablas virtuales de sólo lectura INSERTED y DELETED.
   * Recursividad: un trigger puede disparar una acción que a su vez, lance otro trigger y así sucesivamente.
2. Desarrollar el concepto de Data Warehouse, ejemplifique
   * Un Data Warehouse es una colección de datos orientada a sujetos, integrada, variante en el tiempo, no volátil, organizados para soportar necesidades empresariales.
   * Integración de bases de datos heterogéneas (relacionales, documentales, Geográficas, archivos, etc.).
   * Ejecución de consultas complejas no predefinidas visualizando el resultado en forma de gráfica y en diferentes niveles de agrupamiento y totalización de datos.
   * Acceso interactivo e inmediato a información estratégica de un área de negocios.
   * Análisis de problema en términos de dimensiones (por ejemplo el tiempo).
   * Control de calidad de datos para asegurar la relevancia de los datos en base a los cuales se toman las decisiones.
3. Desarrolle el concepto de índice asociado a una Base de Datos Relacional.
   * Los índices son estructuras físicas y no lógicas.
   * Mejoran la performance de la aplicación, aumentando la velocidad de acceso a los datos.
   * De forma similar al índice de un libro, se guardan parejas de elementos: el elemento que se desea indexar y su posición en la base de datos.
   * Se guardan por un lado los índices (header) ordenados y por otro los datos (cuerpo) en forma secuencial.
   * Se pueden utilizar dos métodos: *tablas de hashing y árbol-B*.
   * Tablas de hashing: se guarda en una tabla la clave (es por lo que se ordena), la posición relativa que ocupa en la parte de datos y un puntero a una lista de duplicados. Se indexa por los campos que son accedidos con más frecuencia. Las claves pueden ser compuestas, por ejemplo nombre y legajo, entonces ordeno por nombre y a igual nombre ordeno por legajo (un solo índice). Pero si necesito ordenar por DNI y legajo entonces se necesitan dos índices.
   * Árbol-B: el número máximo de claves por nodo constituye el orden del árbol-B y el número mínimo de claves permitidas en un nodo es normalmente la mitad del orden excepto para la raíz. En los nodos no terminales se guardan los datos y un apuntador a otro nodo con valores menores o iguales. Todas las hojas tienen la misma profundidad (pertenecen al mismo nivel). Las hojas poseen apuntadores a los registros en el fichero de datos, de forma que el nivel completo de las hojas forma un índice ordenado del fichero de datos.
4. Cual es el objetivo de un Trigger.  Describa un ejemplo práctico de su  
    utilización.

Es mantener la integridad referencial entre tablas.

1. Defina claramente el concepto de transacción, indicando su utilidad y  
    forma de uso.

Utilidad: Mantener la consistencia e integridad de los datos,  haciendo que  
 las instrucciones que la componen no finalicen en un estado intermedio  
 evitando, de esta manera, posibles fallos en el sistema donde se ejecutan.

Forma de uso:

BEGIN TRANSACTION

      UPDATE LEGAJOS

      SET LEGA\_APELLIDO = 'RIOS'

      WHERE LEGA\_LEGAJO = 1

COMMIT

1. Desde el punto de vista de la Performance, en que consiste el armado de un plan de consulta?

El plan de consulta es la forma en que el motor va a resolver una consulta enviada para su resolución, para ello existe un optimizador del propio DBMS que de varias opciones se queda con la mejor. Del mismo modo, quien ejecuta la consulta debe optimizarla previamente tratando de que por ejemplo utilice índices de acceso en forma parcial o total para que la consulta tenga mayor perfomance.

1. Que características distintivas brinda un DBMS?

Sistema de Administración de Bases de Datos (DBMS): Construido sobre sus capacidades en vez de las simples funciones de recuperación y almacenamiento de datos. Estos productos comúnmente incluyen grandes capacidades de back-up y recuperación de datos, ingreso al sistema, control de concurrencia y bloqueo, y mecanismos de seguridad.

1. A la hora de elegir un Motor de Bases de Datos determinado, que características tendría que analizar para su elección?

Costo, Volumen de datos, tiempo de respuesta, control de acceso concurrente y seguridad, redundancia y recuperación de datos.

1. Desde el punto de vista de la Performance, en que mejora el armado de un Modelo OLAP.

*OLAP: En el análisis multidimensional, los datos se presentan mediante dimensiones como producto, territorio y cliente. Por lo regular las dimensiones se relacionan en jerarquías, por ejemplo, ciudad, estado, región, país y continente, o estado, territorio y región.*

*Al procesamiento analítico o análisis multidimensional se le conoce también como procesamiento analítico en línea (OLAP). Se procesa en una visión multidimensional de los datos empresariales en el Data Warehouse y puede tener un motor de depósito de base de datos multidimensional. De esta forma, dentro de un Data Warehouse existen dos tecnologías complementarias, una relacional para consultas y una multidimensional para análisis. Este concepto nace ante la debilidad fundamental de las bases de datos relacionales para proveer capacidad de análisis y sistemas de soporte a las decisiones.*

*La siguiente definición muestra las mejoras que provee un Modelo Olap:*

*Fast (Rápido), significa que el sistema está orientado a tener tiempos de respuesta de 5 segundos, con respuestas de 1 segundo a consultas sencillas y, algunas más complejas, con respuestas de hasta 30 segundos . En situaciones típicas, si las respuestas toman más tiempo, el usuario perderá la concentración en lo que busca analizar.*

*Analysis (Análisis), lo que indica que al usuario debe proporcionársele suficiente funcionalidad para resolver sus problemas, sin la necesidad de contar con el apoyo de sistemas o de una pre-programación. De esta forma, se podrán realizar consultas no definidas, cálculos de diferencias, variaciones y tendencias, consolidar y llevar a cabo análisis de sensibilidad y de búsqueda de metas.*

*Shared (Compartida), significa que debe haber facilidad de acceso simultáneo, tanto de lectura como de escritura, aunado a un esquema de seguridad adecuado, con el objeto de guardar la confidencialidad (probablemente a nivel de celda).*

*Multidimensional, ha sido y continuará siendo un requisito, con la habilidad de manejar múltiples jerarquías y dimensiones.*

*Information (Información), son todos los datos y la información derivada, cuando y donde sea necesaria, en su contexto, tanto de información “suave”*

*como información “dura”, interna o externa.*

1. Que se entiende por Multicubo y Hypercubo y cual es la utilización y diferencias entre ambos.

Es una forma de almacenamiento en forma de array de los datos de las base Multidimensionales.

1. Cual es el concepto de Data Marts?  Para que se utilizan?.  Cual es la diferencia entre un Data Marts y un Datawarehouse?

Data Marts es el mismo concepto que Datwherehousing pero se utiliza en departamentos, son pequeños proyectos de Data wherehousing.  
La principal diferencia es el tamaño, dado que los Datawharehousing tienen grandes cantidades de información almacenada de toda la empresa, por otro lado el data mart solo tiene información de un departamento, esto no quiere decir que sea poca información, pero si lo es en relación, por otro lado, al ser mas pequeño, tiene diferencias, de costo, implementación, accesos, etc.

1. Qué se entiende por DTS?

Son herramientas y objetos que permiten mover datos entre varios orígenes de datos diferentes o iguales. Un DTS puede realizar las siguientes tareas:

- Importación y exportación de datos entre dos orígenes de datos.  
- Transformación de datos.  
- Copia de objetos de base de datos : tablas, índices, etc.  
- Ejecución de instrucciones de Transact-SQL.

DTS tiene una arquitectura OLE DB por lo que puede copiar y transformar múltiples orígenes de datos, entre ellos: SQL Server, Oracle.

1. En que consiste la Técnica de Data Mining? Para que se utiliza e indique alguna de las técnicas algorítmicas aplicadas en el mismo.

*Las computadoras son cargadas con mucha información acerca de una variedad de situaciones donde una respuesta es conocida y luego el software de Data Mining en la computadora debe correr a través de los datos y distinguir las características de los datos que llevarán al modelo. Una vez que el modelo se construyó, puede ser usado en situaciones similares donde no se conoce la respuesta.*

1. Indique las características y funcionalidad de una DataWarehouse.

* Un Data Warehouse es una colección de datos orientada a sujetos, integrada, variante en el tiempo, no volátil, organizados para soportar necesidades empresariales.
* Integración de bases de datos heterogéneas (relacionales, documentales, Geográficas, archivos, etc.).
* Ejecución de consultas complejas no predefinidas visualizando el resultado en forma de gráfica y en diferentes niveles de agrupamiento y totalización de datos.
* Acceso interactivo e inmediato a información estratégica de un área de negocios.
* Análisis de problema en términos de dimensiones (por ejemplo el tiempo).
* Control de calidad de datos para asegurar la relevancia de los datos en base a los cuales se toman las decisiones.

1. ¿Por qué se establece que las Reglas del Negocio deben estar en el Motor de Base de Datos y no en la aplicación cliente?

* Reglas de negocio: cada aplicación debe reflejar las restricciones que existen en el negocio dado, de modo que nunca sea posible realizar acciones no válidas.
* La información puede ser manipulada por muchos programas distintos que podrán variar de acuerdo a los departamentos de la organización, los cuales tendrán distintas visiones y necesidades pero que en cualquier caso, siempre deberán respetar las reglas de negocio.
* Es por lo anterior expuesto que las reglas del negocio deben estar en el motor de base de datos.

1. ¿En qué consisten, para qué se utilizan y cómo pueden implementarse las tecnologías OLAP? Ejemplifique.

* Proporciona información fiable sobre los indicadores clave del funcionamiento de una organización para los sectores que toman decisiones.
* A diferencia de los sistemas transaccionales, OLAP procesa información y aplica inteligencia.
* Un sistema OLAP se alimenta de uno o varios OLTP que lo proveen de datos.
* Todos los campos son calculados.
* Se usa para tomar decisiones.
* Se basa en patrones de interés.
* OLTP → DTS → OLAP. El DTS se ejecuta periódicamente para generar información.
* Se implementa con distintas tecnologías: MOLAP (bases de datos multidimensionales), ROLAP (bases de datos relacionales), HOLAP (bases de datos híbridas).
* MOLAP: ya no se trabaja con el concepto de filas y columnas sino dimensiones. Técnicas de hipercubo (hay dispersión) y multicubo (se saltan los cubos que no contienen información).
* ROLAP:

1. Explique el Modelo Star para el armado de una DataWarehouse. Ejemplifique.

* Es un esquema relacional orientado a la representación de dimensiones que convergen sobre un hecho común o un valor cuantificable.
* De esta forma se logra implementar en un sistema de base de datos tradicional el concepto de cubos multidimensionales.
* Existe una tabla central denominada “tabla de hechos” que representa el atributo cuantificable.
* Alrededor de esta tabla central se definen otras tablas ligadas a esta mediante claves foráneas, y se denominan tablas de dimensiones y son aquellas que definen cualidades del atributo representado por la tabla de hechos.

La clave primaria de la tabla de hechos será una clave compuesta por todas las claves foráneas que representan las claves primarias de cada una de las dimensiones asociadas.

1. Enumere y explique al menos 3 algoritmos de clasificación que conozca.

* Dicotómico o burbuja: funciona revisando cada elemento de la lista que va a ser ordenada con el siguiente, intercambiándolos de posición si están en el orden equivocado. Es necesario revisar varias veces la lista hasta que no se necesiten más intercambios, lo cual significa que la lista está ordenada.
* Quicksort: elegir un elemento de la lista a ordenar (pivote). Resituar los demás elementos de la lista a cada lado del pivote, de un lado los menores y del otro los mayores (en este momento el pivote está ordenado). Se repite el proceso de forma recursiva para las dos sublistas mientras éstas contengan más de un elemento. Una vez terminado el proceso todos los elementos estarán ordenados. Eficiencia promedio nxlog(n); en el peor de los casos n2.
* Heapsort: el algoritmo consiste en almacenar todos los elementos del vector a ordenar en un montículo (heap), y luego extraer el nodo que queda como nodo raíz del montículo (cima) en sucesivas iteraciones obteniendo el conjunto ordenado. Basa su funcionamiento en una propiedad de los montículos por la cual la cima contiene siempre el menor elemento (o el mayor, según se haya definido) de todos los elementos almacenados en él.

1. Enumere ventajas, desventajas y diferencias entre una Base de Datos Orientada a Objetos y una Base de Datos Relacional.

Ventajas BDOO:

* Simplifica la programación.
* Flexibilidad: uso de sub-clases que permiten la herencia de atributos, esto permite que una BDOO puede ajustarse a usar siempre el espacio de los campos que son necesarios, eliminando el espacio desperdiciado en registros con campos que no se usan.
* Manejo de datos complejos: manipula datos complejos de forma rápida y ágil. La estructura de la BDOO está dada por referencias (apuntadores lógicos) entre objetos.

Desventajas BDOO:

* Baja performance.
* Inmadurez del mercado de BDOO.
* Falta de estándares en la industria orientada a objetos.

1. ¿En qué consiste la técnica de Data Mining? ¿Para qué se utiliza? Indique alguna de las técnicas algorítmicas aplicadas en el mismo.

* Consiste en la extracción de información oculta y predecible de grandes bases de datos.
* Las herramientas de datamining predicen futuras tendencias y comportamientos, permitiendo tomar decisiones conducidas por un conocimiento acabado de la información (knowledge-driven).
* Predicción automatizada de tendencias y comportamientos.
* Descubrimientos automatizados de modelos previamente desconocidos.
* Las técnicas más comúnmente usadas son: redes neuronales, árboles de decisión, algoritmos genéticos, etc.

1. Desarrollar las ventajas y desventajas de crear mayor o menor cantidad de índices sobre las tablas de un modelo físico.
2. Si tuviera que elegir un método de creación de índices, entre Hashing y Árbol B, cuando usaría cada uno de ellos y por qué?
3. Indique las diferentes funcionalidades y características de un DBMS.
4. Desarrollo el concepto de FUNCION en PL-SQL. Indique ventajas y desventajas sobre otros objetos similares en un DBMS.
5. Que se entiende por DTS. Indique el concepto y su utilización.
6. Defina claramente el concepto de transacción, indicando su utilidad y forma de uso.
7. Cuáles son las diferencias entre una PK y una constraint de Unique ?
8. Explicar el modelo estrella de un datawarehouse.
9. Indique las características mandatarias y opcionales de un Motor de Base de Datos Relacional.
10. ¿Qué diferencia conceptual existe entre los índices de archivos de datos tradicionales y los índices creados en una Base de Datos Relacional?
11. ¿Desde el punto de vista de la performance, en qué mejora el armado de un modelo OLAP?
12. A la hora de elegir un Motor de Base de Datos determinado, ¿qué características tendría en que analizar para su elección?

¿Qué entiende por diccionario de datos, catálogo de datos o metadata?

* Para representar una base de datos relacional, necesitamos almacenar no sólo las relaciones mismas, sino también una cantidad de información adicional acerca de las relaciones.
* El conjunto de tal información se llama diccionario de datos o metainformación.
* El diccionario de datos debe contener información como: nombre de las relaciones en la base de datos; nombre de los atributos de cada relación; dominio de cada atributo; tamaño de cada relación; método de almacenamiento de cada relación; claves y otras restricciones de seguridad; nombres y definiciones de vistas.
* Además debe contener información sobre los usuarios de la base de datos y los poderes que éstos tienen: nombre de los usuarios; costo del uso efectuado por cada usuario; nivel de privilegio de cada usuario.
* Es de notar que el diccionario de datos no es más que una base datos acerca de la base de datos, y puede almacenarse y manejarse con los mismos métodos.

¿Qué beneficios brinda la aplicación de la normalización al diseño de un modelo de base de datos?

* Elimina la redundancia de datos.
* Mejora la performance.

Defina los conceptos de Data Warehouse y Datamart. Especialmente similitudes y diferencias.

* DataWarehouse: YA ESTÁ EXPLICADO.
* DataMart: podemos entender un Data Mart como un subconjunto de los datos del Data Warehouse con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica. La diferencia entre un DataWarehouse y un DataMart es que el data mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del data warehouse es la organización en su conjunto. Al igual que en un data warehouse, los datos están estructurados en modelos de estrella o copo de nieve y un data mart puede ser dependiente o independiente de un data warehouse. Los orígenes de datos son homogéneos a diferencia del DW.

Describa y cite al menos un ejemplo de una representación computacional que permite representar computacionalmente un grafo irrestricto.

* De manera estática: con una matriz de adyacencia nxn (n: cantidad de nodos del grafo), en donde un 1 en una posición ij representa la existencia de una flecha del nodo i hacia el nodo j. Presenta el problema de ocupar mucho espacio para representar lo que realmente existe ya que se reserva espacio para lugares que no tienen datos.
* De manera dinámica: utilizando una lista o vector de punteros donde se guardan los nodos que participan del grafo con sus correspondientes datos. Cada elemento contiene un puntero a una lista adicional que contiene “las flechas que salen” del nodo que representan. No se guarda el dato en la lista adicional, guarda los punteros a datos.

¿Cuál es la diferencia entre la utilización de Multicubos e Hypercubos para la implementación de bases de datos multidimensionales?

Hipecubos:

* Se guardan todas las dimensiones en un cubo.
* Cada intersección del cubo es otro cubo y así sucesivamente.
* Se pueden agregar dimensiones pero es estático porque no se puede variar la dimensión del cubo.
* Hay dispersión de datos porque los datos no están compactos por el hecho de que puede haber muchos datos vacíos.

Multicubos:

* Se genera una lista con los cubos que tienen información, es decir se saltea los cubos vacíos.

Desde el punto de vista de la performance, ¿en qué consiste el armado de un plan de consulta?

* El optimizador de consultas es el componente del sistema de gestión de base de datos que intenta determinar la forma más eficiente de ejecutar una consulta SQL.
* Se decide cuál de los posibles planes de ejecución de una consulta dada es el más eficiente.
* Los optimizadotes basados en costos asignan un costo (operaciones de E/S requeridas, CPU, etc.) a cada uno de esos planes y se elige el que tiene menor costo.
* El optimizador no puede ser accedido directamente por los usuarios, sino que, una vez enviadas las consultas al servidor, pasan primero por el analizador y recién entonces llegan al optimizador.
* Una consideración muy importante es el orden de los “join”.

¿Qué características distintivas brinda un DBMS? NO SÉ A QUÉ SE REFIERE

Describa los componentes básicos de un DBMS. Teniendo en cuenta la administración de datos, la interfaz con el usuario y el procesamiento cliente-servidor

* Dos lenguajes: DML y DDL.
* 3 capas: externa (usuario), lógica (conceptual) e interna (física).
* Disk manager (recibe peticiones del file manager y las envía al SO), file manager (decide que página contiene el registro deseado y la solicita el disk manager) y user manager (permite la interacción con el usuario).
* Arquitectura cliente-servidor: las aplicaciones corren en el cliente y generan solicitudes para y reciben respuestas del servidor. El servidor realiza el procesamiento de datos y aplica las reglas de negocio establecidas.
* Clustering: es la forma de agrupar la información. Hay dos formas, intra-file clustering (todas las tablas conviven con sus elementos) e inter-file clustering (se guarda cada objeto con sus relaciones).

¿Qué modelos de datos utilizaría para brindar información al área de marketing de su empresa? ¿Por qué?

OLAP?

¿A qué se denomina Plan de Ejecución de Consultas en un Motor de Base de Datos Relacional? YA ESTÁ RESPONDIDA

¿Por qué se recomienda que las Reglas de Negocio deben estar en el Motor de Base de Datos y no en la aplicación cliente? YA ESTÁ RESPONDIDA

1. A la hora de elegir un Motor de Bases de Datos determinado, que características tendría que analizar para su elección?
2. En que consiste un Plan de Consulta, cual es su utilización y la ventaja de su implementación.
3. Indique concepto, tipos, funcionalidades y cual es la mayor ventaja de los triggers.
4. Que se entiende por Multicubo y Hypercubo y cual es la utilización y diferencias entre ambos.
5. Cual es el concepto de Data Marts? Para que se utilizan?. Cual es la diferencia entre un Data Marts y un Datawarehouse?

*Los data marts se ajustan a las necesidades que tiene una parte específica de un negocio, más que a las de toda una empresa. Optimizan la*

*distribución de información útil para la toma de decisiones y se enfocan al manejo de datos resumidos o de muestras. No necesitan ser administrados centralmente por el departamento de sistemas de una organización, sino que pueden estar a cargo de un grupo específico dentro del área de la empresa que los utilice.*

*En ocasiones, los proyectos que comienzan como data warehouses evolucionan a data marts. Cuando las organizaciones acumulan grandes*

*cantidades de datos históricos para el apoyo de decisiones, que rara vez o nunca usan, pueden reducir la información guardada y convertir su data*

*warehouse en un data mart mejor enfocado.*

*Los sistemas que extraen y almacen datos de diversas fuentes para la toma de decisiones, se denominaban Data Warehouses. En fechas recientes se ha hecho una distinción entre los grandes sistemas para almacenar datos (data warehouses) y los sistemas más pequeños (data marts).*

1. En que consisten, para que se utilizan y como pueden implementarse las Tecnologías OLAP. Ejemplifique.

Responde con rapidez a las consultas, de modo que el proceso de

análisis no se interrumpe y la información no se desactualiza.

\_ Tiene un motor de depósito de datos multidimensional, que almacena los

datos en arreglos. Esos arreglos son una representación lógica de las

dimensiones empresariales.

La tecnología OLAP se aplica en muchas áreas funcionales de una empresa,

tales como producción, ventas y análisis de rentabilidad de la comercialización,

mezcla de manufacturas y análisis de logística; consolidaciones financieras,

presupuestos y pronósticos, planeación de impuestos y contabilidad de costos.

OLAP surge como un proceso para ser usado en el análisis de ventas y

mercadotecnia, para elaborar reportes administrativos y consolidaciones, para

presupuestación y planeación, para análisis de rentabilidad, reportes de calidad

y otras aplicaciones que requieren una visión flexible, de arriba a abajo, del

negocio. OLAP provee de reportes sumarios que los ejecutivos requieren para

tomar decisiones, así como la facilidad de elaborar cálculos complejos,

enfoques a detalles operativos y consultas no programadas. OLAP se alimenta

principalmente de los sistemas transaccionales y como tales, debe considerar

una eficiente administración de la base de datos y proveer un nivel adecuado

de seguridad.

1. Cual es el concepto, la funcionalidad y para que se utilizan los índices en una Base de Datos Relacional.
2. Desarrolle las ventajas y desventajas que aporta el uso de un DBMS.
3. Describa y cite al menos un ejemplo de una representación computacional de un grafo irrestricto.
4. Cual es la diferencia entre la utilización de Multicubos e Hypercubos, para la implementación de bases de datos multidimensionales.
5. Caracterice y describa el uso del Modelo STAR.
6. Describa los métodos de indexación que se aplican en un RDBMS, indicando las diferencias y similitudes entre ellos.
7. Defina simple y claramente el concepto de transacción.
8. Describa el funcionamiento del la técnica de Hash para la indexación de datos. Crear un ejemplo simple de implementación y nombrar un ejemplo de implementación real.
9. Defina os conceptos de Datawarehouse y Datamart. Especialmente Similitudes y diferencias.
10. Describa y cite al menos un ejemplo de una represtación computacional que permite representar computacionalmente un grafo irrestricto.
11. Enumere y explique al menos 3 algoritmos de clasificación que conozca.
12. Enumere ventajas, desventajas y diferencias entre una Base de Datos Orientada a Objetos y una Base de Datos Relacional.
13. Indique que ventajas brindan las siguientes características que posee un DBMS: atomicidad, consistencia, procesamiento distribuido y aislamiento.
14. Defina claramente el concepto y cuando implementaría una función (function), un procedimiento (store procedure) y/o disparador (trigger).
15. Que caracterísiticas de funcionamiento tiene y que variantes existen del método QuickSort de clasificación
16. Indique porque un índice conformado por una clave numérica corta es más rápido en el acceso a los datos que un índice compuesto o conformado con un string largo

2.a) En que consisten, para que se utilizan y como pueden implementarse las Tecnologías OLAP. Ejemplifique.

2.b) Cual es el concepto, la funcionalidad y para que se utilizan los índices en una Base de Datos Relacional.

Consultas del grupo de GDD

En la cursada se vió que la cantidad de filas que tiene una tabla de  
 Hashing es igual a la cantidad de punteros.  
 Cada nodo de un árbol B está formado por una tabla de Hashing.  
 Sin embargo en el apunte publicado en  este grupo y en varios apuntes  
 que hay en Internet muestran que la cantidad de índices que hay en un  
 árbol es igual a la cantidad de filas + 1.

1) ¿Que es lo que tenemos que tomar en cuenta?

Primero un nodo del árbol B no es una tabla de hashing es "similar", son dos  
 formas de implementación de un árbol B, puede tener n punteros con n  
 componentes de datos o n+1 punteros, según como se implemente.

2) ¿Un arbol B puede tener claves duplicadas?. Cómo actua un árbol B  
 en este caso.  
 Tengo entendido que en el caso de las Tablas de Hashing cuando se  
 producen repeticiones se asocia una lista enlazada en la que para cada  
 valor de una funcion hashing repetida se accede a cada nodo de la  
 lista hasta que se encuentra el valor deseado.

El árbol B puede administrar claves duplicadas si las hay, está una debajo  
 de la otra en la hoja que les corresponde apuntando cada una de ellas a la  
 componente de datos correspondiente.

1.      Es correcto decir que tanto las claves PK como las claves FK son  
 Check Constraints?

Si

2.      Se podría decir que las Check Constraints son restricciones al  
 modelo físico?

Si

2. Que diferencias existen entre decir Constraints y Check Constraints desde  
 el punto de vista conceptual?

Yo tengo entendido que las Check Constraints son un subconjunto de las  
 Constraints, pero no estoy muy seguro que sea correcto.

Es esto

1. Si tuviera que elegir un método de creación de índices, entre Hashing y  
 Árbol B, cuando usaría cada uno de ellos y por qué?

Que parámetros tendría que tener en cuenta para decidirme por algún método  
 en especial? Tengo entendido que el volumen de los datos influye en la  
 elección de los métodos pero me gustaría saber la respuesta correcta a esta  
 pregunta.

Influye el volumen de los datos, los tipos de clave, dado que Hashing genera  
 redundancia y el tipo de acceso, dado que Hashing está preparado para un  
 acceso directo y no secuencial

2. Determinar que índices implementaría sobre las tablas intervinientes en  
 la siguiente consulta, sabiendo que no existen claves implementadas  
 actualmente en el modelo, especificar por que:

SELECT usr\_nombre, usr\_clave, usr\_apellido

FROM  usuario, rol

WHERE ( user\_nombre='usuario001'  OR usr\_apellido LIKE '%Gonzalez' )

AND usuario.usr\_id = rol.usr\_id AND rol\_tipo=lower('Administrador');

Cuando se está hablando de índices, en este caso, sería erróneo contestar  
 que implementaría una clave PK y una clave FK, la clave PK asociada a la  
 columna usr\_id de la tabla usuario y la clave FK asociada a la columna  
 usr\_id de la tabla rol?

Pregunto esto, ya que al implementar estas claves PK se generan índices para  
 esas columnas.

Con respecto a la justificación:

Al hacer el join usuario.usr\_id = rol.usr\_id en el where y teniendo en  
 cuenta que ambas columnas son PK (tienen un indice asociado), el DBMS  
 resuelve la query de una manera mucho mas rápida y eficiente que sin  
 definirlas como PK.

Seria correcta esta respuesta? Haría falta agregar algo mas?

Es correcto, pero además se podría crear un índice por el campo user\_nombre  
 para que la busqueda sea más rápida, no por apellido, dado que el LIKE anula  
 la búsqueda por índice,