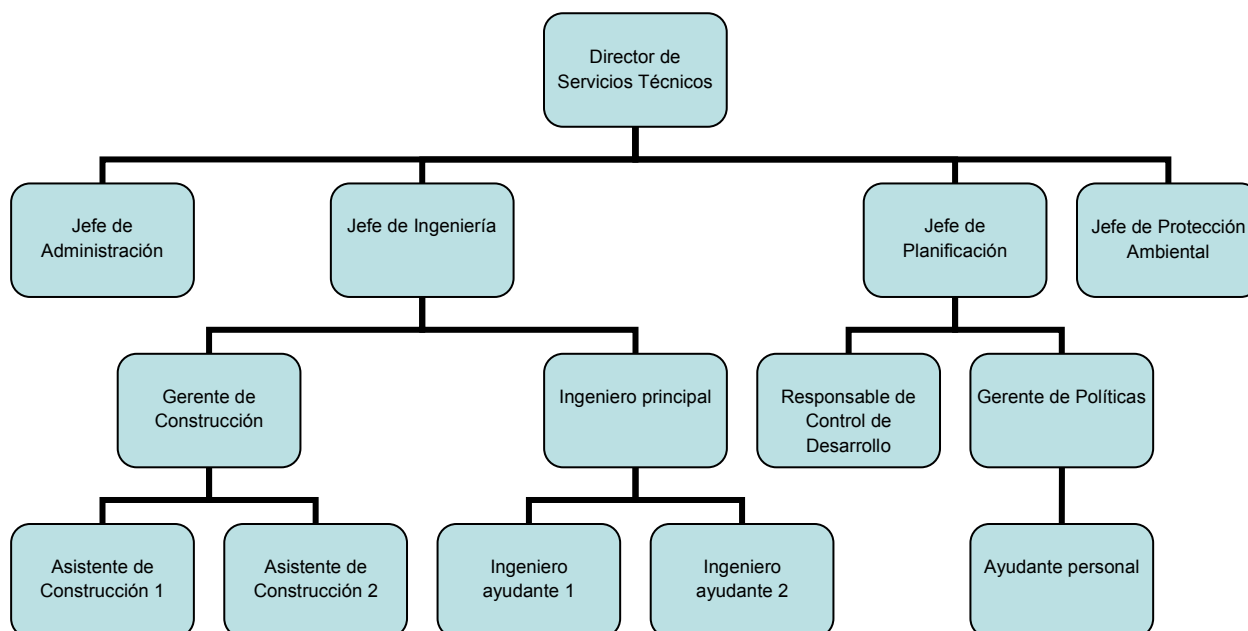


Prueba 2: Opción A: Bases de datos

Temas troncales del NM y el NS

A1. Una compañía tiene una división llamada Servicios Técnicos. Esta división está compuesta por departamentos. El diagrama siguiente muestra cómo se organizan los distintos miembros del personal.



El departamento de Recursos Humanos mantiene una base de datos con archivos planos para almacenar la información de salarios, vacaciones y promoción de todos los miembros del personal. A continuación se muestra un ejemplo de la información almacenada:

Número	Apellido	Nombre	Puesto	Gerente de línea	Salario
948	Kent	McKenzie	Jefe de Planificación	Adrienne Mathis	90875
156	Barker	Rae	Jefe de Administración	Adrienne Mathis	30465
815	Riddle	Kevyn	Responsable de Control de Desarrollo	Kent McKenzie	34768
580	Figueroa	Rina	Gerente de Políticas	Kent McKenzie	45078
871	Mathis	Adrienne	Director de Servicios Técnicos	Russell Z. Harrell	45800
457	Neal	Paul	Jefe de Protección Ambiental	Adrienne Mathis	80670
297	Oliver	Ralph	Ingeniero principal	Daria Gilmore	50796
51	Gilmore	Daria	Jefe de Ingeniería	Adrienne Mathis	90655

El departamento de Recursos Humanos está sufriendo muchos errores al recopilar informes de personal y se le ha sugerido que cambie a una base de datos relacional.

(a) (i) Defina el término *base de datos*.

[1 punto]

Una base de datos es un almacén de información digital que se mantiene de forma organizada para recuperar datos de manera sencilla.

(ii) Describa la diferencia entre una *base de datos de archivos planos* y una *base de datos relacional*.

[2 puntos]

En una base de datos relacional, los datos están organizados en tablas que están relacionadas con otras tablas que le dan el significado completo. En una base de datos de archivos planos, toda la información relacionada con una fila está almacenada en una tabla, en la misma fila.

(b) (i) Identifique **dos** problemas generados por la *redundancia de datos*.

[2 puntos]

La redundancia de datos se produce cuando los datos se duplican en una base de datos. Hace que los errores sean más probables, ya que es posible dejar los datos en un estado incoherente al actualizar o borrar registros.

(ii) Esboce, con ejemplos, cómo la redundancia de datos en esta base de datos puede generar errores.

[2 puntos]

En esta base de datos, el nombre del gerente de línea se almacena en cada fila. Si se modifica el nombre de un gerente de línea, p. ej. porque contrae matrimonio o se sustituye al empleado, habrá que actualizar el nombre para cualquiera que esté bajo la supervisión de ese gerente de línea, y es posible que el nombre no se cambie en todas las instancias y, por tanto, algún empleado tenga un gerente de línea incorrecto.

(c) Los departamentos suelen tener varios proyectos en curso que mantienen ocupados a los miembros del equipo. El departamento de Recursos Humanos necesita generar listas que muestren quién está trabajando en los proyectos actuales.

Cuentan con la siguiente información:

Proyecto: Proyecto_Ref, Proyecto_Nombre, Presupuesto

Departamento: Ref_Dept, Dept_Name,

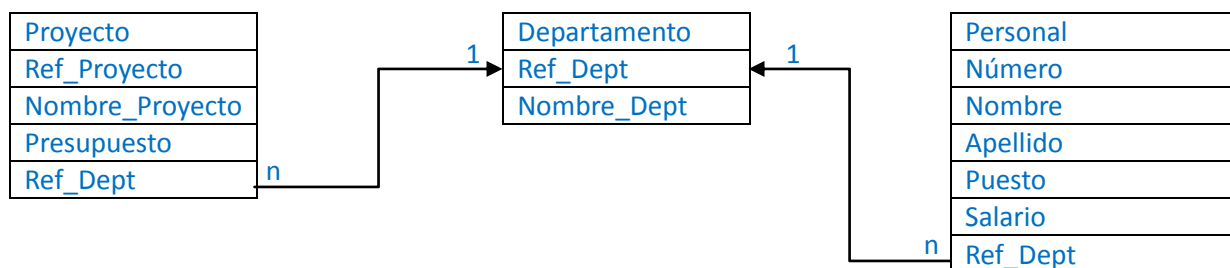
Personal: Número, Nombre, Apellido, Puesto, Salario

Cada proyecto pertenece a un departamento.

Cada miembro del equipo es miembro de un departamento.

(i) Construya un Diagrama de Entidad Relación para mostrar cómo se podría diseñar una base de datos relacional con la mínima probabilidad de que se produzcan errores ocasionados por la redundancia de datos.

[4 puntos]



(ii) Identifique los pasos necesarios para crear una consulta que se pueda utilizar para generar una lista con los miembros del personal cuyo salario sea superior a \$50.000 y que trabajen en el proyecto “Relocation_2012”.

[4 puntos]

Seleccionar los campos siguientes: Nombre, Apellido y Salario de la tabla **Personal**, y Proyecto Nombre de la tabla **Proyecto**

Relacionar las tablas: **Personal** con **Departamento**, y **Departamento** con **Proyecto**

Y que se cumplan los criterios: Salario > \$50000, Proyecto Nombre = “Relocation_2012”

A2. Un banco quiere transferir dinero de una cuenta a otra. Ambas cuentas están almacenadas en el sistema gestor de bases de datos (SGBD) del banco. El SGBD de cuentas realiza una serie de operaciones independientes para completar esta transacción. Entre éstas, se incluyen el retiro de dinero de una cuenta y su depósito en otra. Si se interrumpe este proceso (si, por ejemplo, se produce un corte de energía eléctrica), se puede producir un error. Esto puede poner en peligro la integridad de la base de datos de cuentas.

(a) (i) Identifique **dos** características de una *transacción* en términos de una base de datos. [2 puntos]

Una acción indivisible realizada en una base de datos y que se ejecuta independientemente de otras transacciones.

(ii) Distinga entre el *esquema* y el *estado* de una base de datos. [2 puntos]

Un esquema de base de datos es la definición de la estructura de la base de datos, mientras que el estado de la base de datos incluye los datos que contiene esa base de datos, así como cualquier restricción definida sobre los datos.

(iii) Explique por qué la interrupción descrita anteriormente podría comprometer la integridad de la base de datos de cuentas. [2 puntos]

El estado de la base de datos incluye la cantidad total de dinero, especificado en todas las cuentas. Si se extrae dinero de una cuenta y no se ingresa en la otra cuenta correspondiente, se deja a la base de datos en un estado incoherente debido a que la cantidad total de dinero ha cambiado.

(b) Identifique los pasos que el software debe seguir para mantener la integridad de la base de datos de cuentas. [4 puntos]

El estado de la base de datos debe comprobarse antes de que el software realice los cambios.

Entonces, se debería realizar la transacción.

El software debería, de nuevo, comprobar el estado de la base de datos.

Si los cambios han dejado un estado incoherente, deberían deshacerse.

Si esto ocurre, se debería volver a intentar realizar la modificación o anotarlo para que intervenga una persona.

- (c) Los datos de la cuenta bancaria son compartidos por varias aplicaciones, como el sistema de contabilidad interna del banco y el sistema de transacciones en línea. Explique cómo pueden surgir problemas de concurrencia y cómo se pueden evitar. *[6 puntos]*

Si los datos están duplicados entre varias aplicaciones, en algún momento dichos datos deben sincronizarse entre las aplicaciones. Habrá alguna latencia en esta sincronización y por ello las distintas aplicaciones tendrán versiones de datos ligeramente diferentes. Si una transacción tiene lugar antes de actualizar los datos, se podría retirar de una cuenta dinero que ya se ha retirado de otra cuenta. Además, si se producen dos transacciones a la vez, se genera una condición de carrera y el valor final de los datos depende de qué transacción se ha producido en último lugar.

Para evitar problemas de concurrencia, todas las transacciones deberían realizarse tan pronto como fuese posible para que las futuras transacciones usaran los valores correctos actualizados. Además, mientras que se realiza una transacción se debería bloquear los datos para que otros procesos no puedan modificarlos u obtener valores incorrectos.

Los datos deberían almacenarse en una única ubicación y no permitir que distintas aplicaciones tengan sus propias copias locales, ya que la existencia de copias locales favorece la posibilidad de que existan distintas copias de los datos y, como muchas transacciones se basarán en los valores actualizados de los datos, esto hará que se usen valores incorrectos como base para las transacciones.

A3. (a) (i) Identifique **dos** funciones de un *sistema gestor de bases de datos* (SGBD). [2 puntos]

El sistema gestor de bases de datos es el encargado de crear y usar una base de datos o un conjunto de bases de datos.

(ii) Enumere **dos** herramientas que suelen proporcionar los SGBD. [2 puntos]

Una herramienta para generar consultas
Herramienta para importar/exportar datos

(b) Los sistemas gestores de bases de datos permiten compartir datos y tener diferentes vistas.
Explique por qué estas características son importantes para una organización que use bases de datos. [4 puntos]

Dentro de una organización es probable que haya muchos empleados que necesiten acceder a los mismos datos, lo que implica que compartir datos es muy importante para que los empleados puedan acceder con facilidad y simultáneamente a la información necesaria. Es importante tener varias vistas para que cada persona que accede a los datos vea solo la información que le concierne. Probablemente la base de datos contenga información que es irrelevante o confidencial para algunos empleados y esta información debería ocultarse a aquellos que no la necesitan o no deben ver esos datos.

(c) Explique por qué es importante que los datos sean independientes del software de aplicación que los manipule. [6 puntos]

Se podrían usar los mismos datos en ubicaciones distintas dentro de una organización, y así se podrían usar distintas aplicaciones para ver y editar los datos. Si los datos son solo compatibles con un software concreto, toda la interacción con los datos se debería producir con ese software. Además, el software puede tener una vida útil limitada y si se acaba la asistencia técnica para un software concreto o se necesita una nueva funcionalidad, podría ocurrir que no fuera posible añadirla si hay una gran dependencia entre el software y los datos.

No obstante, si los datos se almacenan independientemente seguirán siendo válidos aunque se usen distintos tipos de software. Además, lo contrario también es correcto: si se modifica la estructura de los datos (por ejemplo, si se añade otra columna a una tabla), la independencia de los datos y el software garantizará que el cambio no hará que el software falle.

Ampliación del NS

A4. Meubles de France es una gran compañía que fabrica y vende muebles. Tiene una gran fábrica y oficinas en la Provenza y más de cien puntos de venta minorista. Emplea cientos de trabajadores. Posee grandes cantidades de datos en sistemas informáticos independientes que gestionan:

- Pedidos de clientes
- Empleados
- Datos de ventas
- Datos de producción
- Finanzas
- Presupuestos

Todos estos sistemas se han comprado a distintos proveedores de software. La gerencia de Meubles de France ha recibido el consejo de que un almacén de datos sería de gran utilidad para el negocio.

(a) (i) Identifique **dos** características de un almacén de datos.

[2 puntos]

Un almacén de datos ofrece una forma de agregar información desde distintas fuentes dentro de una organización. También ofrece vistas relevantes e informes basados en los datos.

(ii) Identifique **cuatro** características del negocio de Meubles de France que indiquen la necesidad de contar con un almacén de datos.

[4 puntos]

Varios orígenes de datos.

Datos de distintos sistemas informáticos que podrían contener datos relacionados pero no vinculados.

Como los sistemas en funcionamiento proceden de distintos proveedores de software, no es probable que se pueda combinar los datos sin la ayuda de un almacén de datos que interprete estos datos.

Un almacén de datos también ofrecería una ubicación central para generar informes sobre el negocio con el objetivo de ofrecer mejores indicaciones sobre el estado de la compañía.

(iii) Identifique **cuatro** transformaciones que pueden ser necesarias para poder producir datos usables en un almacén de datos. *[4 puntos]*

Hay que convertir datos de distintos sistemas para que toda la información simbólica tenga la misma forma.

Habría que fusionar algunas partes de los datos.

Habría que eliminar todas las columnas duplicadas para dejar de tener datos redundantes.

Es posible que los datos también deban validarse.

(b) Meubles de France usa la minería de datos para descubrir patrones de gastos de sus clientes que le permitan optimizar su marketing para futuros productos y servicios.

Compare el uso de la asociación y el análisis de clústeres con este objetivo. *[4 puntos]*

La asociación tiene en cuenta cómo se conectan los eventos. Por ejemplo, podría tener en cuenta los productos que se suelen comprar juntos e informar a los gerentes de Meubles de France sobre métodos de promocionar esos productos en el futuro. Por ejemplo, se descubre que los clientes que compran muebles para el jardín suelen comprar también plantas.

El análisis de los grupos puede organizar los clientes de Meubles de France en grupos concretos o clústeres del conjunto de datos total guardado en el almacén de datos, en función de distintas variables. Por ejemplo, se descubre que hay un grupo de gente que compra en el centro de jardinería que tiene más de 50 años y son los que proporcionan la mayor parte de los ingresos del restaurante en el horario de la comida los días de semana. Los gestores pueden usar estas agrupaciones para acercar el restaurante al centro de jardinería.

Muchas organizaciones como Meubles de France usan la minería de datos para generar un perfil complejo de sus clientes.

(c) Explique por qué a los grupos defensores de las libertades civiles les podría preocupar que Meubles de France tuviera una información tan detallada. *[6 puntos]*

La minería de datos permite que Meubles de France construya un perfil complejo de los clientes mediante los datos de estos. No obstante, una vez que se han obtenido los datos, otras compañías podrían usarlos de una manera que los clientes de Meubles de France no pensaron cuando dieron los datos inicialmente. A los grupos defensores de las libertades civiles también les puede preocupar que otras compañías que tengan acceso a estos datos los usen con objetivos que podrían causar problemas inesperados a los clientes de Meubles de France. Por ejemplo, si un tercero cede a otras organizaciones información que pudiera producir un aumento en el costo de los seguros sanitarios, o si la cede a agencias gubernamentales que pudieran infringir la privacidad de los clientes.