

## Tarea 1: Conceptos básicos

**Integrantes:** Arcos Morales Ramón : 319541478  
 Casarrubias Casarrubias Victor Manuel : 421003581  
 Castillo Hernández Eduardo : 420003557  
 López Asano Miguel Akira : 320219089

**Fecha:** 23/08/2024

### Ejercicio 1. *Conceptos Generales:*

- a. Describe las principales características del enfoque de bases de datos y contrástalo con el enfoque basado en hojas de cálculo. ¿En qué casos no tendría sentido utilizar una hoja de cálculo? ¿En qué condiciones resultaría mejor opción utilizar las hojas de cálculo?

Bases de Datos	Hojas de Cálculo
Organizado en estructuras de datos	Celdas, ordenadas en filas y columnas
Manejo de grandes volúmenes de datos	Limitaciones por memoria y recursos
Visualización de datos prediseñada	Gráficas y visualización de datos configurables
Opciones limitadas de cálculo	Cálculos y formulas matemáticas
Múltiples usuarios simultáneamente	Un usuario a la vez
Apariencia formalizada	Apariencia personalizable
Análisis relacional	Análisis de datos simple
Sistema de seguridad	Seguridad limitada
Integridad de los datos	Poca o nula restricción en el tipo de dato
Consistencia de datos	Propenso a duplicación de información

Tabla 1: Enfoque de bases de datos contra enfoque basado en hojas de cálculo. [Prado, 2020]

Las hojas de cálculo y los sistemas manejadores de bases de datos son dos herramientas para el registro y análisis de datos, a pesar de ser distintos en concepto y diseño, en los últimos años han ido adoptado funcionalidades entre si para adecuarse a las necesidades modernas.

Muchas personas usan las hojas de cálculo por su alta disponibilidad en computadoras, ya que es común que vengan integradas como parte de las aplicaciones de oficina, además la visualización de información tiene un formato más atractivo para el usuario casual. Con toda esta familiaridad, poder y facilidad de uso, las hojas de cálculo es la herramienta por preferencia para una gran cantidad de personas en una variedad de situaciones, pero por estas mismas razones, las personas se ven tentadas a forzar el uso de hojas de cálculo en situaciones donde no dan la talla.

Estos son algunos de los casos en los que utilizar hojas de cálculo pueden causar dificultades. [Gordon, 1999]

#### **Problema en el ingreso de datos.**

Si todos los datos se han recopilado antes de ingresarlos a la computadora, entonces una hoja de cálculo puede ser la mejor opción, sin embargo, si se tiene un conjunto de datos dinámico, donde nuevos datos siguen añadiéndose, o las relaciones entre los datos existentes cambian de manera frecuente, entonces se necesitaran de herramientas más complejas. Con un sistema manejador de datos, se puede diseñar un formulario de entrada de datos, para hacer la tarea de ingreso de datos lo más eficiente y preciso posible, esta misma tarea con una hoja de cálculo se vuelve propensa a muchos errores, ya que a pesar de que las hojas de cálculo modernas también soportan formularios de entrada de datos, uno como usuario tiene que apegarse a usarlo de la manera adecuada en vez de introducir directamente los datos en las celdas de la hoja.

#### **Problema en la consistencia de datos.**

Las hojas de cálculo están diseñadas con una escritura de datos libre y dinámica, por defecto, cualquier celda puede contener ya sea números o texto, se pueden imponer algunas restricciones sobre el tipo de dato ingresado en una celda, pero de igual forma es fácil cambiar o anular estas restricciones. En una hoja de cálculo moderna, las restricciones pueden ser tan sofisticadas como un sistema manejador de bases de datos, no solo pueden restringir los datos a un solo tipo particular, sino también a un conjunto determinado de valores permitido dentro de ese tipo. Por ejemplo, una fecha puede estar restringida al mes actual, o a un campo de texto que solo permite las letras S o N. Ya sea que se este usando una hoja de datos o un sistema manejador de bases de datos, se tiene que invertir tiempo y esfuerzo para establecer las restricciones.

**Problema en la búsqueda y recuperación de datos.**

Frecuentemente es útil observar un conjunto de datos desde diferentes perspectivas, o reorganizar los datos en una variedad de subconjuntos. En una hoja de cálculo, este objetivo a menudo se logra ordenando de diferentes maneras, reordenar no solo cambia la manera en que los datos son mostrados sino también como se almacenan. Si la hoja de cálculo contiene operaciones que dependen de datos particulares que se encuentran en celdas particulares, al ordenar se pueden destruir los resultados de los cálculos. Si uno inserta o elimina filas o columnas, la hoja de cálculo es lo suficientemente inteligente para cambiar las formulas en seguimiento de los datos, pero en el uso común es más probable que este movimiento de datos sea muy complejo y termine alterando la información.

**Problema en el intercambio de datos y colaboración.**

Las hojas de cálculo son diseñadas para ser mantenidas por una sola persona, se puede poner un archivo de hoja de cálculo en una red de computadoras, de modo que varias personas puedan acceder a los datos, sin embargo, la gestión del intercambio de datos no es tan sofisticado como lo es en un sistema manejador de bases de datos. Los sistemas manejadores de bases de datos son diseñados para mantener la integridad de los datos mientras que permite actualizaciones de mas de un usuario, además cuentan con sistemas de seguridad, que pueden ser usados para identificar a los usuarios por nombre y contraseña para selectivamente permitir o denegar acceso a distintas áreas de la base de datos.

- b. ¿Qué ventajas y desventajas encuentras al trabajar con un Sistema de Bases de Datos considerando que se planea implantar este sistema en una empresa de telemarketing?

Una empresa de telemarketing es aquella que ofrece o vende un producto a través de medios como el correo electrónico, la televisión y el teléfono. Para esto, es necesario que la empresa cuente con la información de los clientes, como sus formas de contacto, sus compras, el inventario que la empresa posee, además de información acerca de sus empleados.

Utilizar un Sistema de Bases de Datos brinda diversas ventajas, ya que permite guardar grandes cantidades de datos acerca de los clientes. Además, la posibilidad de relacionar clientes con los empleados que realizaron el contacto permite crear familiaridad. Otra clara ventaja del uso de este sistema es la facilidad para organizar los datos y poder realizar búsquedas de manera más eficiente. Por otra parte, si la empresa no es muy grande y no cuenta con personal con experiencia en Sistemas de Bases de Datos, estos pueden llegar a ser complicados de usar. Otra desventaja es el costo que conlleva la creación y mantenimiento de un Sistema de Bases de Datos.

- c. ¿Qué es la independencia de datos y por qué se carece de ella en los sistemas de archivos? ¿Cuál tipo de independencia de datos es más difícil de lograr? Justifica tu respuesta.

Primero recordemos los niveles de una base de datos, primero está el externo (la vista de los usuarios finales), luego el conceptual (el diseño de la base de datos) y por último el físico (la forma de almacenar la BD).

La independencia de datos es la capacidad de modificar un nivel de la base sin hacer cambios en los superiores. En el caso de la independencia física, se puede modificar la forma de almacenamiento de una base de datos sin tener que cambiar la parte conceptual de esta. Por otra parte, la independencia lógica permite cambiar el esquema conceptual sin afectar el nivel externo.

Un sistema de archivos (o de ficheros) es una herramienta para usuarios finales que controla el almacenamiento y recuperación de datos. La estructura física de este sistema se encuentra codificada, por lo que es difícil modificarla. Para lograrlo, se deben de modificar las aplicaciones que trabajan con los datos que se cambiarían. Es por ello que el sistema de archivos tiene una falta de independencia lógica-física.

La independencia más difícil de lograr es la lógica. Si queremos cambiar el lugar de almacenamiento de ciertos datos de una base, estos seguirán teniendo la misma relación con los demás datos. De forma diferente, si modificamos algo a nivel conceptual, la vista de los datos puede cambiar. Por ejemplo, si eliminamos una columna de datos en nuestra BD que un usuario final podía ver, la vista de este cambia, rompiendo la independencia.

- d. Describe el papel que tienen los Sistemas Manejadores de Bases de Datos (SMBD) en el enfoque de bases de datos. ¿Por qué consideras que es importante (o no) que un administrador de bases de datos (DBA) conozca las características de un SMBD?

Un Sistema de gestión de bases de datos (SMBD), como su nombre indica, es software que nos permite administrar una base de datos, facilitando el definir, construir, manipular y compartir sus datos. Con un SMBD y una BD podemos tener un Sistema de Bases de Datos. Los SMBD son muy importantes a la hora de manejar una base de datos, ya que nos permite hacer acciones en esta de forma segura y sencilla (incluso no expertos podrían usarlos hasta cierto punto). Es por ello que los SMBD son muy usados por las

Un DBA debería conocer las características y el funcionamiento de un SMBD para cumplir con su papel adecuadamente. Un administrador debería de aprender a usar un SMBD, pues se puede mantener y diseñar de una forma más sencilla, además de que estos sistemas son muy usados por las corporaciones para manejar bases de datos. Igualmente, si un administrador tiene un buen conocimiento en bases de datos, puede aprovechar en su trabajo las ventajas que permite tener un Sistema de Bases de Datos: independencia de los datos respecto a los tratamientos y viceversa, reducción de la redundancia, mejor disponibilidad de los datos para los usuarios, mayor valor informativo, mejor eficiencia para recibir datos, normalización de la documentación de la información, etc.

- e. Indica las responsabilidades que tiene un Sistema Manejador de Bases de Datos y para cada responsabilidad, explica los problemas que surgirían si dicha responsabilidad no se cumpliera.

Algunas de las responsabilidades más importantes de un sistema manejador de bases de datos son: [Europea, 2022] [Yasar, 2024] [Javatpoint, 2020]

- **Definición de los datos.** Define todos los objetos de la base de datos, de manera que se pueden almacenar según su tipo o estructura. Si esta responsabilidad no se cumpliera, los datos tendrían más libertad de tipo y sería más común tener errores de inconsistencia de datos.
- **Manipulación de los datos.** Responde a las solicitudes del usuario para realizar operaciones de eliminación, actualización y extracción de datos, también permite modificar el esquema físico o lógico de la base de datos sin realizar cambios en las aplicaciones que la utilizan. Por este mismo motivo si el sistema manejador de bases de datos no fuera el intermediario, se pondría en juego la integridad y consistencia de los datos al permitir al usuario modificar los datos directamente de la base de datos.
- **Almacenamiento de los datos.** El sistema manejador de bases de datos, provee un almacenamiento de datos eficiente, además crea una estructura para los conjuntos de datos complejos de tal forma que los usuarios pueden acceder y manipular los datos de manera sencilla. Algunos sistemas modernos no solo proveen de almacenamiento para los datos sino que también almacenan y administran los metadatos. De no ser posible cumplir esta responsabilidad cualquier operación de consulta tardaría demasiado o simplemente sería incapaz de concretarla. Los datos tendrían la posibilidad de corromperse, sobrescribirse o en peor de los casos perderse.
- **Independencia de los datos.** Ofrece independencia física y lógica para proteger a usuarios y las aplicaciones de la ubicación de los datos o de preocuparse por cambios en la estructura física de estos. En ausencia de esta responsabilidad los desarrolladores tendrían que modificar los programas cada vez que haya un cambio interno en la base de datos.
- **Seguridad de los datos.** Ante una petición, el sistema manejador de bases de datos aplica las medidas de seguridad definidas para evitar ataques o acceso de usuarios no autorizados. También se asegura que las operaciones leer, escribir y eliminar pueden ser ejecutadas en datos específicos. Es de vital importancia esta responsabilidad, ya que cualquier filtración, manipulación o incluso visualización no deseada de los datos y por ende de la información derivada puede ocasionar pérdida de activos, fiabilidad y datos sensibles.

- **Integridad de los datos.** Se asegura de que los datos están correctamente guardados en la base de datos, procurando la consistencia y minimizando la redundancia. El problema que se generaría al no cumplir con esta responsabilidad es similar a lo que pasa con las hojas de cálculo, las cuales al tener una entrada de tipos datos más libre, es usual cometer errores en el formato del mismo conjunto de dato.
  - **Recuperación y restauración de los datos.** Ante un fallo o actualización no deseada, los sistemas manejadores de bases de datos brindan la posibilidad de recuperar y restaurar la información a través de copias de seguridad. Esto protege la información de cualquier error de hardware, software, o algún evento inesperado.
- f. Investiga qué es la redundancia de datos. ¿Cuál sería la diferencia entre redundancia de datos controlada y no controlada?
- La redundancia de datos es cuando un dato es guardado multiples veces en varios lugares de la base de datos.[Silberschatz et al., 2020] Por ejemplo:

Id	Nombre	Edad	Promedio
01	Otis	19	8
02	Leon	22	9
02	Leon	22	7
03	Carlos	21	10
04	Myriam	20	6

Tenemos una redundancia de datos, pues el alumno con Id 02, se repite dos veces. Lo que puede causar inconsistencias a la hora de realizar una búsqueda.

La diferencia entre la redundancia controlada y la no controlada radica en el hecho de que en la controlada, el SMBD asegura que las múltiples copias son consistentes, por otra parte, en el no controlado el sistema no tiene ningún control sobre dichas inconsistencias. [M.Hoda, ], [Royle, ]

- g. Investiga cuáles son las responsabilidades de un DBA. Si asumimos que el DBA nunca está interesado en ejecutar sus propias consultas, ¿necesita entender y/o conocer el modelo de datos lógico de la base de datos? Justifica tu respuesta.
- h. Investiga por qué surgieron los sistemas NoSQL en la década de 2000 y compara a través de una tabla sus características vs. los sistemas de bases de datos relacionales.
- El origen de los sistemas NoSQL surge de la combinación entre la drástica reducción de costos de almacenamiento, el aumento en volumen, variedad de datos estructurados y no estructurados, y el crecimiento en el uso de sistemas distribuidos con distintos servidores. Causando que se necesitara de almacenamiento más flexible y fácil de escalar. [R.John, 2020]

NoSQL	Bases de Datos Relacionales
Fácil de escalar	Difícil de escalar
Sin normalización	Con normalización
Flexible	Propiedades ACID
Sin esquemas	Con esquemas
Variedad de modelos	Un solo modelo
Sin lenguaje estándar	Lenguaje SQL
Datos dinámicos	Datos estructurados
Alta disponibilidad	Alta disponibilidad

[J.Montemagno, 2022][MongoDB, ] [A.Nayak, 2013]

- i. Investiga tres noticias relacionadas con las hojas de cálculo, en donde se muestre que hay errores en el procesamiento/manejo/distribución de datos y realiza una discusión breve de cada una de ellas, contrastándolas con el enfoque de las bases de datos. ¿Por qué consideras que a pesar de los ejemplos mostrados, las hojas de cálculo siguen siendo la herramienta más prevaeciente en un ambiente de negocios típico?
- "The Norwegian sovereign wealth fund's 92mn Excel error." [Wigglesworth, 2024]
- En 2023 el fondo soberano de Noruega reveló que había perdido 980 millones de coronas

noruegas, aproximadamente 1756 millones de pesos mexicanos, por un error relacionado con la referencia de una fecha. El empleado que causó el incidente reveló que irónicamente después de tomar un curso sobre errores y como evitarlos, se le fue comisionada la tarea de arreglar un error de cálculo del índice de referencia del ministro de hacienda. Su idea fue apuntarlo en una nota para no olvidarlo, meses después se encargó de hacerlo, pero su peor pesadilla se cumplió cuando cometió una equivocación en la fecha e introdujo 1ro diciembre en vez de 1ro de noviembre, generando que el balance general no cuadrara.

A pesar de ser un error minúsculo, este se acumuló con el tiempo, problemas de este tipo son comunes en hojas de cálculo donde el tipo de dato de una celda es libre o tiene muy pocas restricciones. En una base de datos, con un simple formulario para la entrada de datos y que el mes se visualizara como una cadena en vez de un número hubiera sido suficiente para evitar este error millonario.

- "PSNI Faces £750,000 Data Breach Fine After Spreadsheet Leak." [Muncaster, 2024]

El Servicio de Policía de Irlanda del Norte (PSNI) se habría enfrentado a una multa abrumadora de 5.6 millones de libras esterlinas por una grave filtración de datos el año pasado si la entidad reguladora no hubiera adoptado una nueva política hacia los organismos del sector público, según la Oficina del Comisionado de Información (ICO). En cambio, el organismo de control de la protección de datos impuso una multa de 750,000 libras esterlinas al PSNI por no proteger información altamente confidencial sobre su fuerza laboral. Un error humano llevó a la publicación en línea de una hoja de cálculo que contenía el apellido, las iniciales, el rango y la función de los 9483 funcionarios y personal del PSNI en servicio. Fundamentalmente, esto incluyó los detalles de personas que trabajan en áreas sensibles como vigilancia e inteligencia, lo que generó preocupaciones sobre la seguridad de los oficiales y sus familias.

Un problema de seguridad perfectamente evitable con una base de datos, incluso la forma de seguridad más básica como la de un acceso por credenciales pudo haber evitado este error.

- ¿Y el ganador es? Un error lleva a los socialdemócratas austriacos a elegir un candidato equivocado. [AP, 2023]

Hans Peter Doskozil fue nombrado nuevo líder del partido, dos días después descubrieron un error en el escrutinio. Los socialdemócratas austriacos tienen un nuevo líder, después de que un error en el escrutinio del fin de semana les llevara a anunciar primero un ganador equivocado. Andreas Babler, alcalde de la ciudad de Traiskirchen, a las afueras de Viena, se enfrentaba a Hans Peter Doskozil, gobernador de la provincia sudoriental de Burgenland, en la votación para ser el nuevo líder del partido. Doskozil es una figura más conocida por la opinión pública austriaca, y en la convención del partido celebrada fue declarado vencedor por un estrecho margen. Sin embargo, el jefe de la comisión electoral del partido anunció que en realidad era Babler y no Doskozil quien había ganado, por un margen de 317 votos a favor y 280 en contra. El lunes se realizó un recuento después de que los responsables del partido descubrieran que faltaba un voto en el total original. En el proceso, descubrieron que el resultado del sábado se había invertido debido a un error que se produjo al introducir los votos en una hoja de cálculo.

De igual forma que la noticia anterior, es muy común que las personas cometan errores en el ingreso de datos, o que al momento de hacer expresiones matemáticas entre filas o columnas se interpolen datos que no deberían, arrojando un mal resultado. Qué algo tan importante como unas elecciones queden a merced de un sistema propenso a fallos, habla mucho sobre como se pierde fiabilidad por facilidad de uso.

Un negocio típico no tiene los recursos o el conocimiento para crear su propio sistema manejador de bases de datos, y aunque los tuviera a veces los retornos de esa inversión de tiempo y dinero son mínimos. Como se comentaba en el inciso a, las hojas de cálculo están prácticamente disponibles en todas las computadoras modernas y además son relativamente más fáciles de aprender a manejar. Recordemos que no todas las personas son expertas en computación y aún se aferran al uso de hojas físicas, así que cuando la obsolescencia los alcanza es más probable que comiencen a utilizar una hoja de cálculo a una base de datos. Incluso en una compañía mediana, la búsqueda de nuevos empleados se enfoca en un perfil que ya sea experto en el manejo de hojas de cálculo tipo excel, porque evita la capacitación en el uso de la herramienta y pueden desenvolverse en su trabajo lo más pronto posible.

- j. Supón que desees crear una red social orientada al uso empresarial, negocios y empleo, similar a LinkedIn. Considera cada una de las desventajas indicadas en el documento "Purpose of Database

Systems”, cuando se administran los datos en un sistema de archivos. Discute la relevancia de cada uno de los puntos indicados, con respecto al almacenamiento de datos de los perfiles profesionales: el usuario que lo subió, las fechas de los empleos que ha tenido, sus habilidades, sus postulaciones y los usuarios que vieron su perfil, cantidad de publicaciones realizadas, cantidad de reacciones, cantidad de compartidos, entre otros.

Siguiendo las desventajas asociadas a la administración de datos mediante un sistema de archivos indicadas por [Silberschatz et al., 2020], se tiene:

- **Redundancia e inconsistencia de datos:** Supóngase que se tiene una aplicación para cada carrera; es decir, se desea guardar y desplegar los perfiles profesionales de egresados de cierta carrera. En dado caso, cierto usuario puede ser guardado más de una vez, si este tiene más de alguna profesión. Dígase, busca empleado como médico y como músico, puesto tiene ambas licenciaturas. Se puede tener inconsistencia, si consigue empleo de alguna profesión, sin que se notifique dicha modificación al archivo de la otra carrera.
- **Dificultad al acceder a los datos:** La problemática asociada es similar a la que describe [Silberschatz et al., 2020], pues se tiene que filtrar la lista de perfiles profesionales o crear una nueva aplicación dado cierto requerimiento. Además, se da el caso de si un usuario desea consultar sus publicaciones realizadas acerca de cierta empresa, se necesita o desplegar todas o crear una aplicación que lo maneje.
- **Aislamiento de datos:** Se puede tener datos en archivos divididos por habilidades, por publicaciones en tendencia, ofertas de trabajo, etc. Los cuáles fácilmente pueden provocar que los datos queden aislados. Por ejemplo, se puede tener ciertas habilidades asociadas a sus usuarios sin tener las fechas de sus empleos o de las publicaciones realizadas.
- **Problemas de integridad:** Supóngase que se necesitan desplegar o consultar publicaciones con cierto número de vistas o usuarios con gran cantidad de compartidos, o se necesita configurar ofertas de empleo que tengan un salario en particular. Con lo que se tiene dificultad al adaptar las restricciones al agregar aplicaciones o realizar modificaciones en las mismas.
- **Problemas de Atomicidad** Al igual que lo descrito en [Silberschatz et al., 2020], si se tiene el sistema a la hora de aceptar alguna oferta, se tiene que tener la seguridad de haya sido efectuados o no. Que se despliegue su confirmación después de aceptar la oferta o que no sucedan las dos.
- **Anomalías en el acceso concurrente:** Se puede presentar algún problema a la hora de hacer diversas publicaciones o modificar perfiles empresariales si diversos usuarios modifican la información o detalles de los perfiles.
- **Problemas de Seguridad:** Es necesario que los usuarios solo puedan modificar la información únicamente de sus perfiles. Que solo puedan ver y no modificar los perfiles de otros usuarios.

**Ejercicio 2. Lectura de artículo**

- a. Leer el artículo *Data Is the New Oil–Sort of: A View on Why This Comparison Is Misleading and Its Implications for Modern Data Administration* y realizar un resumen del documento, destacando los puntos que a su consideración sean los más relevantes (no más de dos cuartillas).

El artículo comienza comparando los datos con el petróleo, enumerando los puntos en los que se parecen, pero inmediatamente da sus diferencias en cuanto a ciertos productos como el petróleo, mencionando 10 características únicas de los datos, que definen como se manejan, recolectan e intercambian, estas características son: los datos no se pueden consumir, se pueden duplicar, se generan a gran velocidad, son heterogéneos, su refinamiento se define en base a su fuente y a su uso previsto, su valor económico es incierto, puede ser manipulado de forma indiscernible (es decir, puede ser modificado por varios agentes), puede estar sujeta a restricciones especiales y que requiere nuevos conceptos de intercambio e infraestructura. Estas 10 características se toman en cuenta a la hora de diseñar y manejar diferentes sistemas que hagan uso de datos. Las anteriores características se engloban en 3 áreas de estudio que son la administración de datos, la seguridad de los datos y la privacidad de los datos.

La administración de datos es el área que se enfoca en todos los pasos necesarios para adquirir conocimiento de datos crudos, esta área se divide a su vez en 4 características, los cuales son: adquisición de datos, que consiste en elegir datos relevantes de una gran cantidad de fuentes, para después juntarlos en una arquitectura de manejo de datos central; preparación de datos, que se enfoca en el limpiamiento de datos, transformación de datos; agregar metadatos y armonizar esquemas; el manejo de datos el cual su nombre es auto explicativo, este concepto actualmente se maneja con implementaciones que hacen uso de lagos de datos (Data Lakes) en los cuales los datos crudos se almacenan casi sin modificar y la preparación de los datos toma lugar dinámicamente dependiendo del caso de uso; aprovisionamiento de datos el cual se encarga de otorgar un acceso simple a los datos.

La seguridad de los datos se divide igualmente en 4 características, los cuales son: confidencialidad, el cual protege en contra de la revelación información sensible mediante acercamientos criptográficos; integridad, característica que se refiere a que los datos no puedan ser modificados por agentes externos; disponibilidad, el cual explica cómo manejar el almacenamiento y disponibilidad de los datos de forma segura, regularmente con llaves para encriptación y desencriptación; autenticidad, que similar a la característica anterior, requiere el manejo de llaves para poder identificar usuarios, usando regularmente contraseñas. Estas características se manejan a través de implementaciones técnicas.

La privacidad de datos se basa en derechos que los usuarios tienen y se tienen que manejar de igual manera en el sistema, los cuales son el derecho a estar informado, es decir, cuales datos se recolectan de un usuario y cómo se manejan; el derecho a la restricción de procesamiento, el cual aboga porque los usuarios puedan elegir que información no dar de ellos y que no se les pueda identificar; y el derecho a ser olvidado, que trata sobre que el usuario tiene derecho a solicitar que sus datos sean eliminados sin posibilidad de recuperación. Estos derechos se toman en cuenta a la hora de diseñar arquitecturas acerca del manejo de datos, usando implementaciones como técnicas de obfuscación para no poder identificar usuarios concretos.

A pesar de que dichas características parecen ser independientes al hacer una implementación que las trate, en la práctica requiere de un modelo o arquitectura que tome en cuenta las 3 principales características para poder tener un sistema completo, por lo cual, el artículo introduce su propuesta, la cual es la plataforma de refinamiento (Refinery Platform). La Plataforma de Refinamiento se divide en 4 etapas principales: la adquisición de datos, preparación de datos, manejo de datos y la provisión de datos; este modelo integra varias de las procedimientos descritos en los puntos anteriores para tratarlos y manejar de forma general todas las características antes mencionadas, ejemplo de esto son los lagos de datos para el almacenamiento o la obfuscación de datos para la privacidad, o la aplicación de blockchain para asegurar la integridad de los datos.

En la última sección analiza y realiza pruebas de dicho modelo a un modelo práctico, utilizando varias herramientas, los cuales muestran que dicho modelo funciona bien a baja y a mediana escala.

- b. Realizar un ensayo donde expresen sus comentarios (cada integrante del equipo deberá indicar este punto de forma individual en el documento que redacten) sobre la lectura, considerando los siguientes puntos:

- Deberán indicar cuál es el objetivo que quiso plantear el autor: qué intenta decir, de qué intenta persuadirnos y/o convencernos, ¿cómo se relaciona con la materia de Fundamentos de Bases de Datos?
- Deberán indicar cuál es la temática central del artículo y se deben señalar el tema o los temas laterales que desarrolla el mismo y cómo estos tienen relación con tu práctica profesional.
- Consideraciones personales: deben indicar una postura ante las ideas planteadas en el artículo, proporcionar argumentos a favor o en contra (propios).

El texto trata sobre los datos como un producto, el cómo se maneja y cuáles son los problemas y algunas posibles soluciones, sobre las ventajas o desventajas de utilizar ciertos acercamientos u otros. Parece ser que el autor quiere convencernos que los datos son incluso más importantes que productos como el petróleo, pues comienza con una comparación entre ambos, listando en que se parecen, pero inmediatamente después comienza a mencionar aspectos que solo son únicos a los datos, por lo cuál, parecería querer implicar que los datos son un producto aun más especial que otros productos de otras revoluciones industriales en comparación a su época o contexto.

A su vez, este artículo se relaciona con la materia de Fundamentos de Bases de Datos, pues se toca partes de las que se dedica una base de datos, como por ejemplo, la estructuración de datos y asignación de metadatos, es decir, poder generar información de los datos crudos y que después pueda ser utilizado por los consumidores de datos, además, toca temas de seguridad y privacidad que las bases de datos deben de tomar en cuenta en su diseño, pues la disponibilidad y la integridad de los datos es algo que una base de datos necesita manejar.

El artículo se centra en el manejo de datos, desde que se recolecta hasta que son disponibles para los consumidores, por lo que en dicho proceso, se tocan varios procedimientos en los que un programador tiene que tomar participación como por ejemplo en la recolección de datos de forma segura, privada y rápida, en la creación de una base de datos que pueda manejar y estructurar dichos datos recolectados, en optimizaciones que puedan hacerse para mejorar la velocidad de estos procesos, e implementar tecnologías como blockchain para poder manejar y almacenar dichos datos. Estas herramientas o técnicas se aplican en base a características que los datos tienen para su manejo, descritas en el artículo.

Coincido con que los datos sean quizá el producto más importante en la era digital (o la cuarta revolución industrial), pues todos los productos que actualmente se utilizan hacen uso de datos, ya sea en mayor o menor medida, nuestros sistemas actuales de casi toda industria o gobierno se basan en datos, incluso nuestra propia salud depende de sistemas que hacen uso exhaustivo de datos, por lo cuál coincido en ese punto con el autor, aunque la comparación inicial con el petróleo parece más una excusa para poder convencernos que los datos son más especiales. El modelo de un sistema que maneje todas estas características parece ser bueno en verdad, que muestra buenos resultados en pruebas asiladas, aunque, a reserva de que no tengo el conocimiento necesario para juzgar correctamente dicho diseño, a escala real, pienso que podría ser un poco idealizado y en realidad podría sufrir de algunos problemas de los que inicialmente quiere evitar con tal diseño.

- Victor Manuel Casarrubias Casarrubias

**"Data is the new oil"** es el término que se está utilizando en la actualidad para referirse a los datos, con el fin de resaltar la importancia y desarrollo que están causando los datos en la cuarta revolución industrial. El autor Christoph Stach, a lo largo de su artículo **"Data Is the New Oil—Sort of: A View on Why This Comparison Is Misleading and Its Implications for Modern Data Administration"**, trata de contrastar el petróleo con los datos, además de explicar las ideas erróneas que se han formado con la popularización de los datos, proponiendo una plataforma como posible solución a las problemáticas actuales. Durante este contraste, el autor menciona las diferencias clave entre estos dos recursos, como su vida útil, su forma de extracción, sus aplicaciones en la industria, el manejo que requieren, entre otros. Un punto que cabe recalcar, es la mención de varios temas relacionados con áreas de Ciencias de la Computación, como Sistemas Distribuidos, Criptografía, Fundamentos de Bases de Datos, entre otros. El autor, en múltiples ocasiones, menciona cómo tener un sistema de almacenamiento que se pueda adaptar a las necesidades del manejo de grandes cantidades de datos, así como garantizar la seguridad e integridad de los mismos, es crucial para el desarrollo de estas tecnologías.

Puntos importantes a destacar, incluyen las fases del manejo de datos, donde se requieren múltiples pasos y campos para poder funcionar, en los cuales miles de profesionales trabajan actualmente



y que las corporaciones empiezan a priorizar. Dichos campos son los siguientes: administración de datos, donde se incluye la adquisición, preparación y organización de los mismos; la seguridad de los datos, donde se abordan temas de integridad, confidencialidad, autenticidad, disponibilidad; y, finalmente, la privacidad de los datos, donde se incluyen algunos derechos para poder tener cierto control sobre el manejo de los mismos. Otro punto que cabe destacar es la mención de la plataforma de refinoría, la cual es un acercamiento a un manejo eficiente de datos. Para poder realizar una plataforma como la mencionada en el artículo, se requiere de diversos conocimientos y áreas, las cuales pueden ser un posible camino para que egresados de la carrera u otras relacionadas puedan aportar y desarrollar de manera exitosa.

Las opiniones y consideraciones que nos aporta Stach son bastante valiosas, y coincido con muchos puntos que son difíciles de rechazar, como el aumento en el uso de datos, el cual se puede ver en el desarrollo de nuevas tecnologías que mejoran la experiencia de los usuarios, así como algunas características de los datos que son de suma importancia resaltar, ya que sin estas consideraciones, el manejo de datos puede ser ineficiente e incluso inseguro. Por otra parte, la forma en la que el autor propone mantener la privacidad y autenticidad de los datos me parece un tanto utópica, pues tener el derecho de restringir y olvidar cierta información podría llegar a ser explotado, causando que la información creada presente ciertos problemas. Otro punto a considerar es el elemento que priorizan las empresas, donde en muchos casos se opta por el capital sobre cualquier otro aspecto, creando la posibilidad de un mal uso de los datos.

Como conclusión, el desarrollo de las formas de manejo de los datos y su distribución serán un punto crítico para diversas áreas y sectores en un futuro no muy lejano. La creación de nuevas oportunidades que mejoren la calidad de vida será altamente bienvenidas; sin embargo, la intriga y miedo a la forma en la que estas nuevas tecnologías pueden ser utilizadas serán obstáculos que requerirán tiempo y preparación para ser mitigados.

-Miguel Akira López Asano

El artículo propone las características únicas que poseen los datos al definirlos como el nuevo petróleo. Busca convencer de la importancia y que tienen los datos y la marcada diferencia que tienen con respecto al petróleo; si bien ambos han funcionado como catalizadores de revoluciones industriales. Los datos pueden tener ciertas propiedades y formas de explotación similares al petróleo y a otros bienes materiales, estos poseen diferencias y particularidades, que se exploran en 3 grandes áreas de investigación. Propone una plataforma capaz de atender los problemas relacionados a dichas áreas. Analiza como es se consiguen, pulen y administran dichos datos para conseguir información relevante. Lo cual impacta directamente a las Bases de datos, no solo porque la plataforma propuesta las utiliza, porque estas trabajan con datos para conseguir información; así la recolección y almacenamiento de los mismos es de su total interés.

La temática central del artículo gira entorno a las características únicas de los datos. Las tres áreas creadas por dichas características traen consigo cuestiones a ser estudiadas. Da la actualidad en cada una de ellas y que problemas tiene cada una. Concluye analizando que, normalmente, se buscan soluciones aisladas a los problemas; mientras que, una solución general o que abarque a un conjunto más grande puede traer mejores resultados. Siendo lo anterior, propone la Plataforma de Refinería.

Los tópicos y problemas referidos en las áreas de investigación, afectan puntualmente a las prácticas profesionales de la computación. Se tienen problemas con la organización de los datos y su recolección, se necesitan expertos capaces de refinar los datos. En cuestiones de seguridad se necesitan conocimientos de criptografía para optimizar las encriptaciones. Se requieren conocimientos de Block Chain. Un ejemplo, en la Plataforma de Refinería propuesta por el artículo, el cómputo distribuido tiene gran importancia en ella. Así, cualquier profesional del área está involucrado con el manejo de los datos.

Finalmente, concuerdo con el autor. Pues, en el mundo actual, los datos tienen un valor mayúsculo; no solo para las empresas, sino para uno mismo. Resulta de importancia para las empresas saber qué productos promocionar y a quiénes; en cambio, como usuario me interesa recibir ofertas y recomendaciones, según mi gusto personal. Sin embargo, hay problemas importantes a resolver, siendo la privacidad el mayor relevancia; pues se deben de respetar los derechos que cada usuario tiene con sus datos. Dado el caso de una filtración, no solo se expone una compañía, también todos sus clientes. Actualmente es imposible obviar la importancia de los datos en la vida cotidiana,

en todos lados se recolectan y se utiliza, por ello eficientar sus aspectos es de vital importancia, requiriendo inversión e investigación.

- Arcos Morales Ramón

Christoph Stach, nos invita a reflexionar sobre el concepto de que los datos son el nuevo petróleo en la actual revolución industrial, esta frase se ha popularizado por diversos medios en los últimos años pero ¿realmente es el caso? de ser así, ¿qué similitudes presentan los datos y el petróleo? y ¿qué enfoque podríamos utilizar para aprovechar al máximo este recurso?.

El nos plantea desde el comienzo que las diferencias son enormes; los datos no se consumen, pueden ser duplicados fácilmente, son generados a gran velocidad, son volátiles, heterogéneos, su refinamiento tiene que estar en concordancia con la fuente de los datos y uso intencionado, su valor económico es incierto, puede ser manipulado, puede estar sujeto a restricciones especiales, requiere de nuevos conceptos para el intercambio junto a nuevas infraestructuras. Esto en teoría difiere totalmente con las principales características que consideramos en un recurso no renovable como lo es el petróleo, pero es importante tener en cuenta estas diferencias. Si nos ponemos de acuerdo con la idea planteada y consideramos que los datos no son lo mismo a cualquier otro recurso, el siguiente paso que se tiene que hacer es idear una forma de procesar y refinar eficientemente los datos.

El proceso de manejos de datos se compone de varias áreas separadas, las tres más importantes son la administración, seguridad y privacidad. Cada una cumple con su rol asignado delegando responsabilidades a procesos o etapas específicas pero hay un problema. El acercamiento individual a cada uno de estos procesos interfiere en el gran esquema de los datos, por ejemplo, si uno se preocupa por ofrecer la mejor privacidad al usuario, se pondrán a disposición herramientas para dejar de compartir los datos con la entidad y al no haber datos, no hay nada a lo que hacer refinamiento. Se parte con la idea principal del artículo, la cual es crear una manera más acoplada a lo que en realidad representan los datos, en palabras del autor, una versión más holística, esto es, considerar a los datos como un sistema en donde sus propiedades deben ser analizados en su conjunto y no solo a través de las partes que los componen. Se propone así su plataforma de refinamiento en donde explica parte por parte como deben ser manejados los datos para su propósito final. Estas partes son: la obtención de datos, la preparación de datos, el manejo de datos y el suministro de datos. Este esquema puede ser útil para la materia de fundamentos de datos, ya que al seguir o inspirarse en este método se puede afinar los datos que uno dispone, sin irse a la fuerza bruta y no solo perder datos importantes sino que además no mantener una consistencia entre diversas consultas.

No soy ningún experto en bases de datos como para opinar sobre el modelo propuesto por el autor, pero sin duda me llevo la idea de que la mejor forma de analizar información no solo viene de hacerlo por hacerlo, es de tener una metodología que considere aquellos detalles intrínsecos para ofrecer los mejores resultados de ser posibles.

- Eduardo Castillo Hernández.

- c. Imagina que eres el director de TI de una cadena de agencias de autos a nivel nacional. Redacta un informe al presidente ejecutivo explicando las oportunidades, beneficios y ventajas que se pueden tener si toma en cuenta, lo expresado en el artículo proporcionado.

### **Agencia de Autos a Nivel Nacional.**

#### **Informe sobre una nueva oportunidad:**

En el siguiente informe se presenta y describe los beneficios de invertir para el desarrollo de una Plataforma de Refinamiento, según se expone en [Stach, 2023]. Contrastando los beneficios de los datos, siendo considerados como en nuevo petróleo. Así, se exponen las aplicaciones que puede tener un sistema en esta compañía.

Los datos son considerados como el nuevo petróleo pues tienen un rol significativo en el día a día de las personas. Se prevé que recopilar y usar datos en beneficio de nuestros usuarios puede traer altas utilidades. Conocer los datos e información de nuestros clientes, permitiría no solo enfocarse en las características más favorables de nuestros vehículos; sino, poder expandirnos a nuevos compradores. Podemos explotar nuestros atributos y mejorar las fallas que alienan a nuevos usuarios.

Es por ello que se propone en invertir en el desarrollo e investigación de una aplicación basada en la Plataforma de Refinamiento. Dados los casos de estudio.[Stach, 2023] Una aplicación que conecte

a los usuarios con sus vehículos, capaz de desplegar y recopilar datos trae beneficios tanto para los clientes como para la compañía.

Sin embargo, la Plataforma de Refinamiento aún necesita de desarrollo; pues el manejo efectivo de los datos presenta dificultades. En especial, es importante hacer un énfasis en la seguridad y privacidad de nuestros usuarios. Si bien los datos son valiosos, estos requieren un tratamiento efectivo para ser recopilados, almacenados, administrados. Se tienen diversos retos en cada una de las áreas; no obstante, a diferencia de otros modelos, la Plataforma de Refinamiento trata de abarcar los problemas de manera conjunta y tiene resultados prometedores. De momento dicha plataforma sigue en etapa de concepto, con lo que se necesita inversión para investigación; para poder tener una aplicación que la implemente.

Si bien los datos son considerados como el nuevo petróleo, estos tienen características únicas que los diferencian de los bienes materiales. Dichas diferencias se engloban en tres áreas activas de estudio: administración, seguridad y privacidad. La Plataforma de Refinamiento, pretende solucionar las dificultades de esas tres áreas, sin perder los beneficios que las características únicas de los datos. Permitiendo construir una aplicación vinculada al automóvil, que permita recopilar y usar datos valiosos de nuestros clientes.

## Referencias

- [A.Nayak, 2013] A.Nayak, A.Poriya, D. (2013). Type of nosql databases and its comparison with relational databases.
- [AP, 2023] AP (2023). ¿y el ganador es? un error lleva a los socialdemócratas austriacos a elegir un candidato equivocado. *EuroNews*.
- [Europea, 2022] Europea, U. (2022). ¿para qué sirve un gestor de base de datos? Accedido en Agosto del 2024.
- [Gordon, 1999] Gordon, K. (1999). Spreadsheet or database: Which makes more sense? *Computing in Higher Education*, 10:111–116.
- [Javatpoint, 2020] Javatpoint (2020). Functions of dbms. Accedido en Agosto del 2024.
- [J.Montemagno, 2022] J.Montemagno, H. (2022). Sql vs. nosql data.
- [M.Hoda, ] M.Hoda. Csi2132-2012.
- [MongoDB, ] MongoDB. Relational vs. non-relational databases.
- [Muncaster, 2024] Muncaster, P. (2024). Psni faces £750,000 data breach fine after spreadsheet leak. *Infosecurity Magazine*.
- [Prado, 2020] Prado, I. (2020). Bases de datos vs hojas de cálculo. Accedido en Agosto del 2024.
- [R.John, 2020] R.John (2020). Brief history of nosql databases.
- [Royle, ] Royle, G. Databases - redundancy.
- [Silberschatz et al., 2020] Silberschatz, A., Korth, H. F., and Sudarshan, S. (2020). Database system concepts.
- [Stach, 2023] Stach, C. (2023). Data is the new oil—sort of: A view on why this comparison is misleading and its implications for modern data administration. *future internet* 2023, 15, 71.
- [Wigglesworth, 2024] Wigglesworth, R. (2024). The norwegian sovereign wealth fund’s 92mn excel error. *Financial Times*.
- [Yasar, 2024] Yasar, K. (2024). Database management system. Accedido en Agosto del 2024.