

---

## Tarea 01

---

### Alumnos:

Castañon Maldonado Carlos Emilio  
Chávez Zamora Mauro Emiliano  
Gallegos Diego Cristian Ricardo  
Navarro Santana Pablo César  
Nepomuceno Escarcega Arizdelcy Lizbeth



**Facultad de  
Ciencias**  
UNAM

## 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 tendría sentido utilizar una hoja de cálculo?**

♦ Principales características del enfoque de bases de datos:

- 1) **Reducción de redundancia e inconsistencia:** Los sistemas de bases de datos evitan la duplicación de datos y la inconsistencia, ya que la información se almacena de manera centralizada y gestionada en una sola ubicación. Esto elimina la posibilidad de que diferentes copias de los mismos datos se desvíen.
- 2) **Acceso eficiente a los datos:** Los sistemas de bases de datos permiten realizar consultas y búsquedas de manera eficiente, lo que facilita la obtención de la información necesaria. A diferencia de los sistemas de archivos, donde el acceso a los datos puede ser difícil y poco flexible, los sistemas de bases de datos están diseñados para recuperar datos de manera conveniente.
- 3) **Integridad de datos:** Las restricciones de integridad proporcionan un medio de asegurar que las modificaciones hechas a la base de datos por los usuarios autorizados no provoquen la pérdida de la consistencia de los datos. Por tanto, las restricciones de integridad protegen a la base de datos contra los daños accidentales.
- 4) **Atomicidad y control de concurrencia:** Los sistemas de bases de datos están diseñados para garantizar la atomicidad de las operaciones, lo que significa que las transacciones se ejecutan en su totalidad o no se ejecutan en absoluto. También manejan el acceso concurrente de múltiples usuarios, evitando problemas de inconsistencia causados por actualizaciones simultáneas.
- 5) **Seguridad:** Los sistemas de bases de datos permiten establecer permisos y restricciones de acceso a los datos, asegurando que solo los usuarios autorizados puedan acceder a la información específica. Entonces, esta seguridad se refiere a la variedad de herramientas, controles y medidas diseñadas para establecer y preservar la confidencialidad, integridad y disponibilidad de la base de datos.

♦ Enfoque Basado en Hojas de Cálculo:

- 1) **Redundancia e inconsistencia de datos:** Pueden contener datos duplicados, lo que puede llevar a inconsistencias si no se actualizan correctamente en todas las ubicaciones.
- 2) **Acceso limitado a los datos:** Pueden ser menos eficientes para realizar consultas y búsquedas complejas. Los usuarios pueden tener dificultades para acceder a datos específicos de manera rápida y eficiente.
- 3) **Problemas de integridad:** Pueden carecer de mecanismos robustos para garantizar la integridad de los datos. Las restricciones y reglas deben ser implementadas manualmente, lo que puede llevar a errores.
- 4) **Anomalías de acceso concurrente:** Las hojas de cálculo no están diseñadas para manejar el acceso concurrente de manera eficiente. Actualizaciones simultáneas pueden resultar en problemas de integridad.
- 5) **Problemas de seguridad:** Las hojas de cálculo pueden carecer de características avanzadas de seguridad, lo que dificulta el control preciso de quién puede acceder y modificar los datos.

♦ ¿En qué casos tendría sentido utilizar una hoja de cálculo?

- Las hojas de cálculo podrían ser ocupadas para realizar cosas más simples, como cálculos matemáticos, operaciones rápidas, seguimiento de gastos personales, listas de tareas, entre otras. Tendría sentido utilizarse en problemas donde la complejidad de los datos es limitada.

- b) **¿Qué ventajas y desventajas encuentras al trabajar con una base de datos?**

Recordando que el propósito de una base de datos es el almacenar de forma electrónica un conjunto de datos a los cuales acceder y consultar de forma rápida y sencilla, encontraremos que sus ventajas serán:

- ♦ **Organización:** Las bases de datos nos permitirán organizar los datos de forma estructurada, lo cual hará que la búsqueda y el análisis de los mismos sea algo fácil de hacer.
- ♦ **Rendimiento:** Las bases de datos nos pueden proporcionar un rendimiento extraordinariamente superior a las hojas de cálculo para realizar consultas y análisis complejos, además de que con las bases de datos no estaremos limitados por límites como los de una hoja de cálculo, ya que por ejemplo, estas no son infinitas y al estar cada vez más cerca de llenarse, el rendimiento de estas decae a niveles muy grandes.

- ♦ Resistencia: Las bases de datos pueden proporcionar una mayor fiabilidad que las hojas de cálculo, ya que los datos se almacenan de forma centralizada y están protegidos por mecanismos de seguridad.
- ♦ Escalabilidad: Las bases de datos pueden escalarse fácilmente para adaptarse al crecimiento de los datos y las necesidades de las organizaciones.

c) **¿Qué es la independencia de datos? Indica si sería posible alcanzar independencia de datos lógica y/o física, al trabajar con datos almacenados en un sistema de archivos.**

- **Independencia de datos:** Se refiere a la capacidad de modificar la estructura y organización de la base de datos sin afectar las aplicaciones que utilizan esos datos. En otras palabras, es la capacidad de cambiar la forma en que los datos están almacenados, sin que los programas y aplicaciones que acceden a esos datos tengan que ser modificados.

Existen dos tipos principales de independencia de datos:

- **Independencia Lógica:** Esto significa que los cambios en la estructura lógica de la base de datos (como agregar, eliminar o modificar tablas y relaciones) no afectarán a las aplicaciones que utilizan los datos. Los usuarios y las aplicaciones seguirán accediendo a los datos de la misma manera, incluso si la estructura interna de la base de datos cambia.
- **Independencia Física:** Se refiere a la capacidad de cambiar la forma en que los datos se almacenan físicamente en el sistema (por ejemplo, cambiar de un sistema de almacenamiento a otro) sin afectar las aplicaciones que acceden a los datos. Los programas y las aplicaciones siguen interactuando con los datos de la misma manera, independientemente de los cambios en el almacenamiento físico.

Posiblemente se pueda lograr cierto grado de independencia de datos, tanto a nivel lógico como físico, sin embargo, sería mucho más complejo de lograr, ya que en un sistema de archivos por lo regular los datos suelen estar dispersos (en archivos y programas) y las modificaciones serían muchas, así como en las distintas ubicaciones.

d) **Investiga cuáles serían los costos potenciales de implementar un sistema de bases de datos en alguna organización que trabaja sus datos con hojas de cálculo.**

Los costos potenciales van a depender de una serie de distintos factores, los cuales harán que los costos se disparen o se amortigüen un poco, los principales factores a considerar son; el tamaño de la organización, la complejidad de los datos y el tipo de base de datos en el que se desee implementar lo anterior, ya que en base a esto, podremos visualizar los costos de hardware/software, capacitación de empleados, migración y mantenimiento de la nueva base de datos, si bien esto podría representar un costo relativamente alto, los beneficios a largo plazo como lo son el rendimiento, fiabilidad, escalabilidad y flexibilidad que nos ofrece una base de datos hará que valga la pena la inversión.

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.**

Las responsabilidades de un SDBD y los problemas respectivos de no cumplirlas son las siguientes:

- Definir: Al tener que especificar los tipos, estructuras y restricciones de los datos podemos hacerlos más sencillo de manipular, ya que tenemos conceptos concretos, cuyos alcances están bien delimitados. No tener datos bien definidos podría generar asociaciones sin sentido, como que una edad sea negativa o que no sea sencilla de asociar porque nos falta una estructura, un dato aislado no nos funciona sino hasta que podemos relacionarlo a través de algún tipo de estructura. Hasta que la data tiene cierto orden, cierta estructura podemos hablar de tener información.
- Construir: como necesitamos genera memoria, no nos sirve una base de datos que no tenga ningún tipo de data almacenada. Una base de datos puramente teórica no tiene ninguna utilidad, y para existir necesita ser alojada en algún tipo de memoria que permita su consulta, el medio físico es lo que permite hacer tangible el acceso por lo que la Base de Datos necesita ser construida.
- Manipular: Sin las consultas, la recuperación y la actualización de base de datos es equivalente a que no la tengamos en primer lugar. Operar una Base de datos permite que la organización/empresa/escuela/etc. tome decisiones o realice sus actividades con normalidad, de no poder manipularla todo el esfuerzo por diseñarla o construirla no tiene efectos prácticos.
- Compartir: Asumiendo que ya está definida, que fue construida y que es posible manipularla; hay que poder compartir el acceso a múltiples usuarios, las bases de datos surgen por la necesidad de gestionar datos de manera colaborativa ya que muchas veces esta colectividad es la misma que los genera. Debe ser posibles para múltiples personas tener acceso a ella para ejecutar sus tareas, compartir la data genera más data y es esencial para articularla en acciones colaborativas.

**f) Investiga qué es la redundancia de datos. ¿Cuál sería la diferencia entre redundancia de datos controlada y no controlada?**

La diferencia consiste en que la redundancia de datos controlada será aquella que permite de forma deliberada la misma pieza de datos en dos o más lugares separados para mejorar la disponibilidad o la integridad de los datos, en cambio, la redundancia de datos no controlada ocurre indeliberadamente ya que esta no estaba contemplada, esta puede causar problemas como inconsistencia de datos y duplicaciones de datos no contempladas, lo cual puede provocar el aumento del costo y la complejidad de la base de datos.

**g) Investiga qué papel juegan los analistas de bases de datos, diseñadores y desarrolladores de bases de datos en la construcción de un sistema de bases de datos.**

Los analistas de bases de datos, diseñadores y desarrolladores de bases de datos juegan un papel fundamental en la construcción de un sistema de bases de datos.

Los analistas de bases de datos son responsables de recopilar los requisitos del negocio y traducirlos en un modelo de datos, cabe resaltar que el analista trabaja mayormente con los usuarios finales para comprender sus necesidades y luego utiliza esa información para crear un modelo de datos que satisfaga esas necesidades.

Los diseñadores de bases de datos son responsables de implementar el modelo de datos en una base de datos física, también debemos mencionar que el diseñador trabaja con el analista de bases de datos para comprender el modelo de datos y luego utilizar esa información para crear la base de datos física.

Los desarrolladores de bases de datos son responsables de escribir el código que permite a los usuarios acceder a la base de datos, como con los dos anteriores, el desarrollador trabaja con el diseñador de bases de datos para comprender la base de datos física y luego utilizar esa información para escribir el código que permitirá a los usuarios realizar operaciones en la base de datos, como insertar, actualizar, eliminar y consultar datos.

**h) ¿Qué son los datos estructurados y los no estructurados? Proporciona algunos ejemplos para contrastarlos. ¿Cuál tipo es más prevaleciente en un ambiente de negocios típico?**

Los datos estructurados tienen un formato estandarizado que permiten tanto al software como a las personas acceder a estos de forma eficaz, son fácilmente búscables y normalmente categorizados como datos cuantitativos, cumplen con un modelo de datos y siguen un orden persistente.

Este tipo de datos generalmente se almacena en una base de datos y el Lenguaje de Consulta Estructurada (SQL) es el lenguaje utilizado para gestionar datos estructurados.

Un ejemplo, es una tabla estructurada de datos de clientes que contenga columnas (nombre, dirección, número de tarjeta de crédito, número de teléfono y otros datos que pueden ser fácilmente buscados utilizando SQL).

Los datos no estructurados, normalmente categorizados como datos cualitativos, son información sin un modelo de datos establecido o son datos que no están ordenados de una manera predefinida, es decir; no se ajustan a ningún otro modelo y no tienen una estructura fácilmente identificable.

Dado que los datos no estructurados no tienen un modelo de datos predefinido, se administran mejor en bases de datos no Relacionales (NoSQL). Otra forma de gestionar datos no estructurados es utilizar lagos de datos para conservarlos sin procesar.

Algunos ejemplos comunes de datos no estructurados son: archivos de texto, archivos de vídeo, informes, correo electrónico, imágenes, documentos de Word, etc.

En un ambiente de negocios típico, los datos estructurados suelen ser mas prevalentes. Las empresas recopilan y almacenan grandes cantidades de datos estructurados en bases de datos y sistemas organizados. Dado que los datos estructurados son mucho mas accesibles que los no estructurados, son de gran valor para las empresas.

Sin embargo, los datos no estructurados ofrecen información valiosa sobre tendencias del mercado, satisfacción del cliente y otros aspectos para la toma de decisiones empresariales. En la actualidad aproximadamente el 80-90 % de los datos totales son no estructurados.

**i) Asumiendo que una base de datos es un lugar donde se almacenan datos de forma sistemática y que la información se obtiene al consultar los datos entonces, un diccionario puede considerarse como una base de datos. Imagina que vas a buscar el significado de la palabra Luminiscencia, indica cómo efectuarías la búsqueda y los problemas que enfrentarías con:**

Primero, necesitamos entrar en contexto, en los 3 siguientes casos es como si buscáramos una palabra en un diccionario común de la vida real para saber su significado. En programación necesitamos de un ID para almacenar y buscar un elemento. Dicho esto, resolveremos lo siguiente:

- ★ Un diccionario con palabras desordenadas.

Si tenemos un diccionario con palabras desordenadas, significa que para este problema nuestro ID sera la palabra a buscar su significado.

Si y solo si vamos a realizar la búsqueda del significado de la palabra "Luminiscencia" la manera mas eficaz para resolver este problema, es recorrer el diccionario desordenado con dos punteros, e ir comparando palabra con palabra, es decir recorrer el diccionario desde la cabeza y cola, o lo que es equivalente a decir, del inicio y final hasta encontrar la palabra "Luminiscencia". Sin embargo, la complejidad de tiempo es  $O(\frac{n}{2})$ , que es lo mismo a  $O(n)$ , ya que por contener a  $\frac{1}{2}$  como constante; nos queda como  $O(\frac{n}{2}) = O(n)$ .

Con lo anterior hemos buscado una solución eficiente solo si se va a realizar la búsqueda de una palabra y no mas palabras, Con lo cual, si seguimos buscando más palabras en el diccionario, a la larga, esta es la peor forma de almacenar tus datos; lo que nos obligaría a buscar más soluciones, como lo es ordenar el diccionario u ordenar el diccionario y tener un índice, sin embargo de estas soluciones voy a hablar a continuación.

- ★ Un diccionario con palabras ordenadas, pero sin índice.

En este inciso la manera mas eficiente para poder buscar una palabra como "Luminiscencia" es utilizar el enfoque de como usamos un diccionario de Español o Ingles en la vida real, ¿Qué es lo que hacemos generalmente para buscar una palabra en un diccionario? Mayoritariamente, lo que hacemos es tomar o partir el diccionario a la mitad y ver la primera palabra que sale en la página, si es la misma palabra que estamos buscando pues simplemente ya tenemos nuestra definición, sino sucede lo anterior, si la palabra se encuentra antes de ese pivote, pues volvemos a partir el diccionario a la mitad; pero de la mitad izquierda que ya tenemos, si la palabra estuviese después de ese pivote pues nos tomamos la mitad derecha, y así repetimos este mismo procedimiento hasta encontrar la palabra.

¿Cómo es que comparamos dos palabras? Simplemente tomamos carácter por carácter de ambas palabras y a hacemos una comparación entre esos dos caracteres siguiendo la jerarquía del abecedario, donde el carácter "a" es el menor y el "z" el mayor.

Básicamente estamos usando un enfoque de búsqueda binaria, que tiene una complejidad de  $O(\log(n))$ , que en comparación con la complejidad anterior ya bajo demasiado el tiempo para realizar una búsqueda de una palabra. Consideramos que es un buen tiempo para datos como el tamaño de un diccionario, sin embargo si tuviésemos muchos mas datos que el numero de palabras en un diccionario, esto todavía tardaría tiempo en realizar una búsqueda.

La otra solución es la siguiente.

- ★ Un diccionario con palabras ordenadas y con índice.

Aquí el enfoque que nosotros vemos, es que la computadora por así decirlo tenemos alojado en una parte de memoria un espacio asignado para el diccionario, y ese diccionario es administrado por un ente, y ese ente sabe y conoce la dirección de donde esta alojada cada palabra mediante el ID, un ejemplo burdo que se nos ocurre es el siguiente, el ente que administra esa parte de la memoria, gobierna o conoce donde esta cada palabra, algo así como si estuviésemos en un fraccionamiento, y ese ente sabe en que casa se encuentra cada palabra, a partir del ID de la casa.

Dando este ejemplo, nosotros tomaremos como ID la palabra a buscar, y si realizamos la búsqueda de la palabra "Luminiscencia" nos tomaría un tiempo de  $O(1)$ , que es el mejor tiempo para realizar una búsqueda, ya que ese ente cuando preguntemos por donde esta ese ID, el inmediatamente va acordarse de donde esta ubicado y nos va a dar la dirección de en que casa se encuentra el significado de esa palabra.

Algo que no tomamos en cuenta, es que sabemos que un diccionario como lo es el de Español es finito, y difícilmente se le van a agregar nuevas palabras, por lo cual aquí estamos considerando que el diccionario no tiene colisiones, pero ¿que tal si ese diccionario lo quiero para guardar datos de usuario de un banco?, enfrentaríamos el problema de colisiones, ya que estaríamos manejando un

gran volumen de datos, y nuestro diccionario es propenso a sufrir colisiones.

j) Supón que deseas crear un sitio de vídeos similar a YouTube. Considera cada uno de los puntos enumerados en el documento “Purpose of Database Systems”, como desventajas de administrar los datos en un sistema de procesamiento de archivos. Discute la relevancia de cada uno de los puntos indicados con respecto al almacenamiento de datos de los vídeos: título, el usuario que lo subió, la fecha de carga, las etiquetas, qué usuarios lo vieron, cantidad de “Me gusta”, entre otros.

- 1) Título: El título puede ser un candidato a redundancia e inconsistencia de data (data redundancy and inconsistency) ya que si tenemos varias listas de reproducción es muy posible que al momento de editar el nombre tengamos que hacerlo en múltiples archivos, lo cual es propenso a mostrar inconsistencias si no se recorre absolutamente todos los archivos que componen el sitio, es muy sencillo que esta operación de editar el título de un video no pueda ser atómica (atomicity problems), de ahí que también se muestre redundancia al tenerlo en múltiples listas de reproducción en vez de tener una forma de recuperar el mismo vídeo a través de una referencia única. En temas de buscador tenemos una problemática similar, para hacer búsquedas dentro del sitios los vídeos no sólo pueden tener un nombre igual (como en YouTube mismo), sino que también es muy complicado recuperar todos esos vídeos para hacer una simple búsqueda, esto es un ejemplo de dificultada para acceder a la data (difficulty in accesing data). La falta de filtros en sistemas de archivos hace que ediciones a los títulos del vídeo no sean complicados para usuarios que no sean los propietarios (security problems).
- 2) El usuario que lo subió: Este usuario al no tener mecanismos para restringir su data está vulnerable a la posibilidad de que su contenido pueda ser editado por cualquier persona con pocos o nulos estándares de seguridad (security problems). Además los nombres de usuarios deben ser únicos para poder llevar un registro único de a quién le pertenece cierto contenido (que incluso puede estar protegido por derechos de autor), el que no haya nombres únicos o que no se pueda verificar la unicidad de los nombres al tener que recuperarlos de múltiples archivos, puede generar problemas por tener dos usuarios con el mismo nombre o no reflejar cuando un usuario activo ya haya cambiado a un nuevo nombre, para que el anterior pase a estar disponible (data redundancy and inconsistency).
- 3) La fecha de carga: Las fechas y datos de carga nos permiten tener datos estadísticos del uso del sitio web, lo cual es importante para temas empresariales, pero sacar estadísticas de datos difíciles de recuperar no permiten que éstas puedan ser precisas si en determinado momento quisiéramos medir qué tanta actividad tiene el sitio web en ciertos periodos del año (difficulty in accesing data). Además, se podría editar con relativa facilidad ese dato (security problems), más todavía no tenemos filtros para establecer valores de fechas que no son posibles, como tener “fechas negativas” o que haya condiciones que no permitan que la fecha se comporte como no corresponda (integrity problems).
- 4) Las etiquetas: Una etiqueta tiene la finalidad de que un contenido de cierto tipo sea más sencillo de localizar, es un objetivo perdido si uno debe buscar cada uno de los archivos para hallarlas, el tiempo de espera será monstruoso entre más cantidad de usuarios tengamos en la plataforma (difficulty in accesing data). También las etiquetas posiblemente admitirían caracteres que no hacen más sencillo localizar data, ya que no permiten solamente caracteres alfanuméricos al no tener restricciones en el tipo de data (integrity problems).
- 5) Qué usuarios lo vieron: el vídeo tiene que estar disponible para múltiples usuarios al mismo tiempo, este tipo de concurrencia debe mostrar las ediciones hechas al vídeo en cualquier momento y debe editar la estadística de qué cantidad de vistas ha tenido, lo cual puede ser candidato a anomalías de acceso concurrente (concurrent-access anomalies) y en la sección de comentarios esta problemática también podría mostrarse si la única sección para comentar no se actualiza con las interacciones de las personas.
- 6) Cantidad de “Me gusta”: En la cantidad de estos elementos tenemos que la edición de la estadística debe ser concurrente, como se retira y se otorga un “me gusta”, podríamos tener problemas si el sistema no llega un adecuado registro concurrente que haga que se refleje un número incorrecto que ha cambiado por la interacción simultánea de dos usuarios, sobre todo en vídeos altamente virales, donde las interacciones concurrentes son más críticas.

## 2 Lectura de artículo

- a) Leer el artículo “The Economic Impact of Data: Why Data Is Not Like Oil” y realizar un resumen del documento, destacando los puntos que a su consideración sean los más relevantes (no más de una cuartilla).



## Resumen The Economic Impact of Data: Why Data Is Not Like Oil

El objetivo del documento es hacer notar de manera temprana por qué el autor considera que es importante delimitar las diferencias para medir el impacto económico de la data, además de crear medidas específicas (como legislación pertinente) que permitan aprovechar los beneficios de estos nuevos paradigmas, especialmente con miras al futuro sobre los posibles beneficios de recolectar data de hospitales, búsquedas de empleo, escuelas, servicios, etc.

Una de las formas en que más la legislación puede tener fricciones con el uso de la data es a través de las políticas de privacidad, ya que el autor aboga por leyes laxas cuando se trata de acceder a datos personales, las ve favorables sólo en términos de bienes escasos, de difícil producción o limitados, como el petróleo. Como la data se enriquece al conectarla y no se gasta con el uso, no ve sentido en tratar de tener condiciones estrictas para proteger su circulación o propiedad. A partir de aquí el artículo se divide en cuatro secciones en las que desarrollará los puntos: Por qué la data no es como el petróleo, Más data lleva a mejor productividad e incremento salarial, Una tipología de los beneficios económicos de la data e Implicaciones y principios.

- Por qué la data no es como el petróleo: Aquí habla de que no es posible hacer balances generales como con bienes físicos como maíz, trabajo, recursos naturales, petróleo, etc. Ya que no se merma la data por consumirla y no se paga directamente por servicios como Facebook, las matemáticas convencionales no permiten analizar el impacto económico de la data. A diferencia del petróleo, muestra en una tabla, la data crece todos los días; pues siempre se genera más con interacciones. La data crece en oferta de tipo y volumen, así que no puede ser monopolizada, además de que su valor se genera del tipo de análisis que se le haga, no de la data en sí. La data no tiene valor si se almacena y no se usa, pero otros bienes sí pueden conservar su valor almacenados. La data no es fungible, así que varias personas la pueden aprovechar sin importar cuántas personas ya la han usado. Da ejemplos de esto con cursos en línea, vídeos de cocina, instrucciones para ensamblar automóviles, etc. pero esos son temas en concreto y pequeña escala, también habla de conocimiento en tecnología que puede ser exportado sin que hacerlo genere un costo excesivo o pérdidas económicas a quienes produjeron esa data.
- Más data lleva a mejor productividad e incremento salarial: Aquí el análisis hace una diferenciación macroeconómica; al hablar de industrias digitales e industrias físicas puede notar que el sector digital (integrado por tecnología, telecomunicaciones, financieras, etc.) hace la mayor parte de inversión en tecnologías de la información, a pesar de ser sólo el 30 % de la producción del sector privado. Pero esta inversión parece traducirse en crecimiento de la productividad, ya que ha crecido en promedio un 2.7 % desde el año 2000, en comparativa con el 0.8 % anual de crecimiento del sector físico. Las horas de trabajo también han aumentado, lo cual se suele ver como un indicador de creación de empleos. Un ejemplo notable es el uso de data para comercio en línea, cuyos datos sólo parecen crecer en números de empleabilidad, salarios y transacciones, mientras minimiza el tiempo que demora en comprar un consumidor promedio.
- Una tipología de los beneficios económicos de la data: Aquí describe cómo sectores económicos importantes puede ser optimizados con uso de data; en la educación (la cual integra parte importante del PIB) puede darse la posibilidad de tener cursos más amplios pero también mejor personalizados para distintos tipos de aprendizaje, en la empleabilidad podemos tener un mejor registro de las habilidades particulares de cada aspirante mientras damos mejor promoción a las ofertas. En las compras la data puede ayudar a recomendar productos pero también a elegir mejor ya que contamos con reseñas de personas que hicieron la misma compra en otro momento. Las redes de capital social se pueden articular a través de las conexiones generadas por redes sociales. En temas de exportaciones la data ayuda a crear sistemas más óptimas que permitan que entre a mercados en el extranjero a empresas pequeñas que no tienen la estructura amplia que beneficia a empresas más grandes. También las decisiones gubernamentales se facilitan en pandemias en uso de data.
- Implicaciones y principios: Para todos estos objetivos de data el artículo enfatiza que tiene que ser capaz de recolectarla, por lo que las leyes deben facilitar esa labor. En términos de privacidad sugiere sólo utilizar consentimiento opcional en casos con riesgo alto, abogando beneficio económico a las empresas y menos trabajo de analizar políticas a los individuos, también pide hacer un proceso para no identificar a personas con la data para tener privacidad, sugiere el que la data no tenga limitaciones de frontera de manera que pueda ser compartida entre países y que se busque la innovación como parte de los objetivos más importantes.

El artículo concluye que por todos los puntos mencionados es importante que la industria movida por data y la privacidad vayan de la mano en pos de un crecimiento económico mayor.

- 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).

### **-Carlos Emilio Castañón Maldonado**

En un mundo tan globalizado y conectado como en el que vivimos actualmente, es realmente impresionante como es que todo lo que creíamos saber sobre economía e información ha quedado desactualizado debido a esta nueva era de la tecnología/información, pues antes como humanidad estábamos acostumbrados a encontrar algo con lo cual enriquecernos hasta el punto de monopolizarlo y que a su vez este tuviera un impacto directo en la economía, sin embargo, como dice el texto, nos hemos encontrado el que parecería ser el petróleo del siglo 21, este se trata de la información electrónica de todas las índoles, desde la industrial, gubernamental, escolar, económica, hasta la de las personas que usen un dispositivo electrónico (siendo esta una de las más reutilizables-importantes) sin embargo esta última tiene un pequeño gran detalle y es que al tratarse de la información de los usuarios, esta no se puede monopolizar (al igual que otros tipos de información), lo cual cambia radicalmente las reglas del juego al que estábamos acostumbrados, ya que si bien esta no puede ser monopolizada, esta sí puede ser comercializada, estas son precisamente ideas con las que juega el autor del texto, además de que también juega con ideas como el cómo debe haber algo detrás (una base de datos) que pueda almacenar estos datos y el cómo las industrias están jugando un papel importante en esto ya que cada una tiene una pequeña parte de todo este gran conjunto de datos electrónicos, el autor también hace mucho énfasis en cómo al no ser algo físico que puede ser explotado como el petróleo, y al ser algo digital (intangible), la información ha tomado un rumbo inédito en cuanto a cómo juega en la economía, sin embargo, cuando se comercializan datos como los datos personales de una persona, esto solo hace que la privacidad de la que las personas deberían gozar con su información personal digital, se vea nulificada (y esto es parte de la postura del autor, el establecer que aunque es difícil, es viable un camino en el que los intereses económicos de la información digital y la privacidad puedan convivir en uno solo) y es que contar con la información correcta (por ejemplo los gustos de un conjunto de personas) tiene un impacto directo en por ejemplo, como se va a comercializar un producto y si este va en caminado a vender varias unidades (y por ende tener un impacto en la economía), esto está estrechamente relacionado con la práctica profesional, pues todos estos datos deben ser almacenados (en una base de datos) y a su vez debe haber una persona que pueda manejar, manipular y modificar estos datos almacenados para que otras personas los puedan leer, y es aquí donde entramos nosotros como científicos de la computación.

### **-Mauro Emiliano Chávez Zamora**

El artículo me parece interesante en el sentido en que una de sus tesis principales es muy intuitiva una vez entendida, pero no muy fácil llegar a ella por la inercia del quehacer económico tradicional, la otra tesis me interesa más porque habla de las implicaciones en privacidad de la recolección de data. Creo que el objetivo del autor es motivar decisiones legislativas en un tema en la intersección de industria, gobierno e individuos; el mismo establece al final puntos concretos en las implicaciones de cómo interpreta la data de la que ya dispone. No es gratuito que la única vez que un texto se resalta en azul (además de cuanto da datos estadísticos o resalta información del texto) es para dar indicaciones de lo que espera que suceda en el futuro respecto a las leyes; el mismo formato del texto nos da indicios de lo que busca lograr con su persuasión.

Está íntimamente vinculado con la materia en el sentido de que vamos a generar en el futuro sistemas que puedan manipular toda esta data de manera eficiente, que tengan relaciones que sean relevantes para los objetivos que queremos cumplir dentro de una empresa, hospital, escuela, etc. Me parece que aunque el artículo usa estos sectores de entrada para abogar por el papel de las empresas, las cuales suelen tener lineamientos más estrictos para acceder a la data; las autorizaciones suelen ser más fáciles de otorgar si alguien necesita un proceso médico o acceder al sistema educativo. Como personas que trabajan con estos temas debemos estar conscientes de en qué medio o qué ambiente va a ser en la que se desarrolle la práctica profesional, sobre todo porque las bases de datos están en todas partes y gestionan muchos



sistemas, debemos saber cuál es la importancia de la materia en el medio laboral actual además del tipo de comportamiento que puede acarrear nuestra participación. Dentro de nuestra práctica incluso puede ser posible que se solicite nuestra opinión como profesionales si llegan a buscar información de cómo tomar decisiones ejecutivas o gubernamentales, por lo que es importante que no simplemente ejerzamos nuestra profesión, sino que sepamos qué implicaciones tiene en la sociedad ser activos gestores de la data, que nuestra opinión sea una opinión consciente e informada.

En mi opinión el artículo tiene una postura muy marcada desde el liberalismo democrático, el hacer una crítica a su contenido es criticar un modelo económico porque la visión está centrada en el desarrollo que toma el crecimiento económico como un indicador directo de calidad de vida de la población. Me queda claro que es muy importante considerar estas estimaciones de la data, en realidad me parecen muy buenas, pero creo que hay matices que me hacen pensar en si es posible que el beneficio mayoritario sea en temas de mercantilización de la data para empresas, a pesar de que el artículo mismo trata de usar ejemplos positivos como pensar en tratamientos para el cáncer que pueden verse beneficiados. Los argumentos que abogan por el beneficio me parecen muy similares a aquellos que se usaron en su momento para justificar el neoliberalismo económico, el cual las leyes pedían que se relajaran leyes que evitaran la circulación entre países de mercancías bajo la idea de que la disponibilidad de ofertas podría traer beneficios basados en la colaboración, sólo que esto aunque sí tuvo ventajas evidenció muchas otras desventajas. El artículo creo que hace un intento de explicar que su enfoque como consecuencia traería una reducción de las desigualdades en términos de educación y salarios; pero las mismas problemáticas del liberalismo podrían presentarse, en el sentido de que la educación muchas veces trató de alinearse con la actividad corporativa para medir su eficacia o valor monetario, pero el valor existía en el saber previo a la medición con cuentas o balances, los sistemas educativos como la Academia de Platón se han visto motivados por más cosas que hacer crecer el PIB (de muy reciente creación, criticado por invisibilizar desigualdades). También las desigualdades tienden a acrecentarse en espacios que no protegen el contexto histórico de las personas, no dudo que los salarios en el mundo de la tecnología sean más equitativos y que esas industrias tengan un enorme crecimiento, pero habría que cuestionar por qué esto es así más allá de creer que crecen por el sentido de tener data. Quienes pueden aprovechar este avance económico son personas que de entrada tienen acceso a una computadora, lo cual puede decirse que es mucho en sitios con gran falta de alfabetización digital o conexión, donde no hay luz o no hay infraestructura para que las personas puedan integrarse, tener una educación superior que permita usarlos para ofrecer servicios profesionales todavía puede ser más complicado. Si tenemos grupos relativamente pequeños de personas con distintos tipos de capital (cultural, social, etc) que les permiten estar en estos círculos, es entendible que puedan generar más ganancias, más capital dado que muchas veces son personas con necesidades básicas cubiertas que son capaces de utilizar ese dinero extra para generar ganancias, así que creo que habría que tener cuidado en pensar que la industria crece solamente por ser digital sin pensar en el tipo de personas que lo integran; aunque es cierto que facilita que ciertas transacciones o compras sean más sencillas, apelan a la acumulación de dinero que caracteriza a éste sistema económico porque puede ir más rápido en su lógica de funcionamiento. Me parece relevante pensar en quién enuncia estos discursos y quiénes somos nosotros, cuál es la posición en la que estamos para escucharlos.

### **-Cristian Ricardo Gallegos Diego**

El artículo principalmente se centra bajo la analogía sobre ¿Los datos son el nuevo petróleo? El autor trata de darnos como objetivo como los datos son un recurso clave para el crecimiento económico, y su uso extenso puede generar beneficios en diversos sectores. No solo eso, sino que también intenta convencernos de que compartir y analizar datos es fundamental para el desarrollo económico. Esto se relaciona con Fundamentos de Bases de Datos, ya que destaca como la gestión efectiva de datos es esencial para potenciar la innovación y la toma de decisiones para la construcción de sistemas de bases de datos eficientes.

La temática central del artículo gira en torno al impacto económico de los datos. Explora temas secundarios como la educación personalizada y el emparejamiento laboral. Estos temas laterales reflejan como la tecnología y los datos están transformando la forma en que las industrias funcionan y como se toman decisiones en campos como la educación, la selección de empleo, salud, etc. En mi campo esto se relaciona llevando a cabo la gestión y el análisis de datos, ya que son fundamentales para comprender el funcionamiento del mercado y para optimizar los procesos de las organizaciones y empresas. El autor también menciona la importancia de una legislación adecuada para equilibrar los intereses económicos y la privacidad de los individuos ya que implica una responsabilidad ética en el manejo de datos sensibles y la toma de decisiones sobre como se utilizan los datos.

En mi opinión considero que el autor busca persuadir a la audiencia sobre como el valor de la información supera la analogía del petróleo, influyendo en decisiones de gobierno y empresariales. Es importante preguntarnos quiénes se benefician en este cambio, aunque la tecnología y los datos pueden impulsar industrias y mercados, es necesario considerar quiénes tienen acceso a estas oportunidades. La línea digital y las desigualdades socio-económicas pueden limitar la participación equitativa en esta economía basada en datos. Además debemos considerar como esta nueva economía puede contemplar las desigualdades sino se implementan medidas para llevarlas a cabo.

#### **-Pablo César Navarro Santana**

En la actualidad, la gente comenta (Sin saber bien sobre el tema) sobre como "los datos" son el futuro, que la ciencia de datos es la carrera del futuro y que hay mucho dinero de por medio, pero justamente el artículo nos habla de porque no hay que ver a los datos como si fuera petróleo, ya que hay unas diferencias muy marcadas e importantes, y al no darnos cuenta de estas, podemos estar cometiendo el error de tratarlas como cosas similares.

La materia "Fundamentos de Bases de Datos" se relaciona con el artículo ya que en el artículo se comenta como la data puede darnos información muy valiosa, y en la materia, vemos como podemos ordenar toda esta data para poder analizar la data y así, obtener esa información que tanto habla el artículo.

El artículo nos habla sobre como la data es un recurso muy importante y diferente, ya que este no es un recurso "Físico" del todo, y por lo tanto no tiene tales limitantes como lo sería el petróleo. Tales diferencias son la posibilidad de que al usarla, esta no se gasta como el petróleo, no hay pérdidas al "exportarla" y que es posible duplicarla (Ya que es relativamente bajo el costo).

También nos cuenta sobre como influye la data en la economía, explicando que por si misma no tiene valor alguno, pero al analizarla, puede darnos información muy valiosa para varios sectores, tales como la educación, el comercio, mejoras en el desempeño del gobierno, etc, de tal forma que a la vez de impulsar estos sectores, impulsan también a la economía. Igualmente, habla sobre la importancia en la privacidad de los datos personales y la acumulación responsables de estos por partes de la empresa, ya que pueden ser muy beneficiosos para ciertos sectores como lo puede ser el sector de salud, y da el ejemplo que al obtener datos de pacientes, es posible obtener estadísticas que ayuden a investigaciones futuras contra enfermedades, pero que la recolección de estos datos deben de ser de una forma responsable y justa.

Todos estos temas, son muy importantes en nuestra práctica profesional como computólogos, ya que nos permite hacer un uso responsable y consciente sobre el uso de la data, de la cual usamos y analizamos a diario, ya sea en un proyecto muy importante con una carga social muy grande, o simplemente en alguna búsqueda de un navegador web.

Con respecto a mi opinión sobre el artículo, debo decir que es muy acertado en lo que habla, ya que muchas veces la gente se va simplemente por el lado monetario de los asuntos sin ver realmente las implicaciones que llevan. Y también es muy interesante el que explique como es que la data llega a afectar tanto a tantos sectores públicos. En general, podría decir que estoy de acuerdo con la mayoría de cosas que menciona, debido a que el uso de la data, realmente ayuda mucho a la sociedad y se puede ver claramente en la toma de decisiones de ciertas empresas al escuchar a sus consumidores y realizar decisiones en favor tanto a la empresa como a los consumidores. Aunque me hubiera gustado que hablase mas sobre los posibles peligros económicos que podría llegar a tener este crecimiento económico gracias al uso de la data, ya que se siente que es una burbuja que va creciendo cada vez mas y no se sabe si puede reventar.

#### **-Arizdelcy Lizbeth Nepomuceno Escarcega**

Me resulta interesante la postura del artículo, es decir, este muestra una perspectiva acertada sobre el papel que toman los datos en la economía y la relación entre la privacidad y el crecimiento económico. El autor argumenta que la idea de que "los datos son como el petróleo" es incorrecta y que en realidad los datos pueden impulsar el crecimiento económico de diversas maneras. La temática central del artículo, nos deja ver este impacto y beneficios económicos de los datos en temas relacionados con la educación, el trabajo, redes sociales, el consumismo, exportaciones, entre otros.

Considero que el objetivo principal del autor es que analicemos el papel crucial que desempeña el uso de datos en la economía actual, desde una percepción común de los datos, donde estos tienen que ser protegidos, asimismo, ver cómo la regulación de la privacidad de los datos puede afectar el crecimiento económico y la innovación.

En lo personal, considero que el autor aborda un tema bastante importante entre los datos y la privaci-

dad. La idea de que para proteger la privacidad de las personas se deben establecer restricciones sobre como se recopilan y almacenan los datos, podría limitar esa innovación del crecimiento económico y el argumento de que no necesariamente esto es correcto, pone en el centro este equilibrio entre ambas partes. Podemos ver como es que esto se podría relacionar con “Fundamentos de Bases de Datos” debido a la gestión, privacidad de los datos y análisis de los datos. Resaltaría el uso de datos para la “innovación”, ya que es uno de los temas que el artículo resalta y podríamos verlo en cómo las organizaciones pueden aprovechar los datos almacenados para obtener información, en la mejora de productos y servicios; otro de estos a resaltar es la protección de datos personales, el artículo nos menciona la necesidad de proteger los datos y estos a su vez sean usados para la innovación, las bases de datos deben incluir medidas de seguridad y regulaciones que eviten acceso que no sea autorizado a datos.

Desde una perspectiva profesional, pienso que es importante entender cómo se gestionan y se utilizan los datos de una manera responsable y sobretodo ética, centrándonos en este requerimiento de los datos para impulsar la innovación y el crecimiento económico, tener en cuenta y saber que los datos juegan un papel fundamental en la toma de decisiones; tener en cuenta que las bases de datos son herramientas poderosas y encontrar el equilibrio adecuado para sociedad.

Analizando todo el panorama, realmente beneficiaría al aspecto económico y a ambos lados, es decir la privacidad y la innovación, estos, no son necesariamente excluyentes; si se encuentra un equilibrio, se podrían aprovechar los datos de una forma ética y responsable. Sin embargo, sería algo complejo de lograr ya que involucra cuestiones mas allá de solo ética y responsabilidad, sino cuestiones legales y políticas que podrían obstaculizar el progreso tecnológico y la innovación o a la vez no.

## Bibliografía y Referencias:

- AWS Amazon. (s.f.). What is Structured Data? <https://aws.amazon.com/es/what-is/structured-data/>
- IBM Cloud Education. (2021, 29 de junio). Structured vs. Unstructured Data: What’s the Difference? <https://www.ibm.com/blog/structured-vs-unstructured-data/>
- Coursera. (2023, 16 de junio). Structured vs. Unstructured Data: What’s the Difference? <https://www.coursera.org/art-vs-unstructured-data>
- TIBCO. (s.f.). What is Structured Data? <https://www.tibco.com/reference-center/what-is-structured-data>
- ¿Qué es una base de datos? (s/f). Oracle.com. Recuperado el 1 de septiembre de 2023, de <https://www.oracle.com/mx/what-is-database/>