

Bases de datos 2023-1

Tarea 1: Conceptos básicos

2 de septiembre de 2022

1. Conceptos generales

- a) Describe las principales características del enfoque de bases de datos y contrástalo con el enfoque basado en archivos. ¿En qué casos no tendría sentido utilizar una base de datos?

Algunas de las principales características del enfoque de bases de datos son:

- Almacenar grandes cantidades de datos (hasta cientos o miles de GB) por mucho tiempo.
- Evitar cualquier tipo de pérdida de los datos, ya sea protegiéndolos de desastres (al hardware principalmente) o
- protegiéndolos de usos no deseados de algunos usuarios.
- Permitir la creación de nuevos datos.
- Permitir la lectura de los datos, así como consultas específicas acerca de la información almacenada.
- Permitir la actualización o modificación de los datos.
- Permitir la eliminación de los datos.
- Permitir el acceso a cuantos usuarios se desee (con sus respectivas medidas de autorización).
- Permitir modificar el esquema de la misma base de datos.

Ahora contrastemos algunos aspectos de el Enfoque de Bases de Datos (EBD) con el Enfoque Basado en Archivos (EBA):

- En el EBA tenemos capacidad de almacenamiento muy limitada vs los cientos de GB que nos proporciona el EBD.
- En el EBA los datos pueden perderse dada una caída del sistema.
- En el EBA puede llegar a ser muy engorroso escribir un programa para cada tipo de consulta de la información, mientras que en el EBD esto se resuelve con queries.
- El EBA es muy propenso a alterar la realidad de los datos ya que múltiples usuarios pueden modificar de maneras distintas los mismos datos o información, lo que nos implica fragmentar la realidad y no tener consistencia en los datos. En el EBD no sucede esto gracias a los SMD.

Finalmente, los casos donde no tiene mucho sentido utilizar una base de datos es cuando no almacenaremos ni muchos datos ni mucha información. Cuando no necesitamos hacer muchas consultas ni modificar constantemente los datos y cuando no habrá múltiples usuarios leyendo y escribiendo los datos.

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

Las ventajas son la consulta, actualización y almacenamiento de datos, ya que de hecho ese es su principal propósito y pueden almacenar cantidades muy grandes de información y permiten la rápida consulta y actualización de esta, por otro lado también tenemos la validación de dichos datos, la seguridad y el tener todos los datos en un mismo sistema en vez de varios sistemas que no se puedan sincronizar, validar o cualquier problema de ese estilo.

Entre las principales desventajas se encuentran

- La falta de personal **capacitado** para manipular dichas bases de datos.
- El costo, ya que aparte del costo del desarrollo de estas también está el costo del servidor en el que se va a alojar la base de datos.
- El peso de estas puede llegar a ser muy grande ya que se busca tener los datos centralizados, y puede llegar a alentarse si ésta crece mucho (más de lo esperado inicialmente).
- requiere personal de mantenimiento.
- Hay que actualizar la base de datos según las necesidades lo requieran.

c) ¿Qué es la independencia de datos? ¿Cuál tipo de independencia de datos es más difícil de lograr? Justifica tu respuesta.

La independencia de datos se refiere a la inmunidad de un nivel a ser afectados por cambios en el que esta directamente debajo. Existen dos tipos, lógica y física.

La independencia de datos lógica se refiere a que la aplicación de la BDD no se vea afectada por cambios en el esquema de organización de la información, por ejemplo a agregar un nuevo atributo a una relación. Este tipo de independencia es el más complicado de lograr pues las aplicaciones dependen bastante del esquema lógico en el que están los datos.

La independencia física se refiere a la inmunidad del nivel lógico a cambios en el esquema físico como por ejemplo cambios en las unidades de almacenamiento.

d) Explica la diferencia entre los esquemas externo, interno y conceptual. ¿Cómo se relacionan estas diferentes capas de esquemas con los conceptos de independencia de datos lógica y física?

El esquema externo tiene que ver con las vistas que cada usuario tiene de los datos.

El esquema interno tiene que ver con la descripción física de los datos.

El esquema conceptual se encuentra, de cierta manera, "entre estos 2", pues es una descripción lógica (como el esquema externo), única y global (como el esquema interno) de nuestros datos.

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

La redundancia de datos se refiere a que los datos se encuentran en más de un lugar de la memoria que usamos (en el caso de bases de datos, tenemos datos que se repiten en la base). Esta redundancia puede ser útil o contraproducente dependiendo de la situación.

La redundancia controlada es cuando repetimos los datos intencionalmente, lo solemos hacer para protegernos de pérdidas o fallos en nuestros sistemas o para amplificar las capacidades de nuestras aplicaciones. Las copias de seguridad *offsite* por ejemplo nos protegen de pérdidas de datos causadas por accidentes o catástrofes. Las caches aprovechan la redundancia para acelerar el tiempo de acceso a recursos, por ejemplo una cache en un CDN nos deja servir archivos a muchos clientes sin tener que remitirnos al servidor original para satisfacer la petición (particularmente útil cuando los datos son estáticos, como una biblioteca de uso general).

La redundancia no controlada es cuando sin nuestra intención se repiten los datos, no podemos aprovecharnos de esto y terminan alentando o obstruyendo nuestras necesidades.

- f) 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.

Si. Dado a que algunas de las principales responsabilidades de un DBA son el crear y configurar las bases de datos relacionales así como producir diagramas de entidades relacionales, localización lógica y física de bases de datos y parámetros de tablas (según Wikipedia), es fundamental que el DBA entienda el modelo de datos lógico ya que debe ser capaz de diseñar el diagrama E/R y traducirlo al modelo relacional. Lo que lo hace capaz de diseñar el esquema de la base de datos, es decir, entiende el modelo de datos lógico de la misma.

- g) Entrevista a algún usuario de sistemas de bases de datos, ¿qué características de SMBD encuentran más útiles y por qué? ¿qué instalación(es) de SMBD encuentran más/menos complicada y por qué? ¿cuáles perciben estos usuarios que son las ventajas y desventajas de un SMBD?

Entrevistado: Fernando López Balcazar.

Algunas de las características mas importantes son:

- Los drivers: Que los drivers que tengan los lenguajes de programación con el SMBD y qué tan bien funcionan estos ya que al momento de desarrollar un sistema es importante decidir cuál es el lenguaje que más beneficios o facilidades aporta. También es importante que tengan facilidad al establecer conexiones.
- Sistemas de tipos: Es importante conocer la forma en que el SMBD va a guardar los atributos ya que es importante el uso de memoria.
- Soporte y comunidad: Es importante conocer el soporte que el SMBD que estemos utilizando así como qué tan actualizado está.

MariaDB ha sido el SMBD que mas me ha costado trabajo instalar debido a su incompatibilidad con MySQL en la distribución Linux Fedora pues son prácticamente el mismo sistema y trabajan muchas configuraciones iguales. El mas sencillo ha sido PostgreSQL pues fue de manera bastante intuitiva y mi sistema operativo no presento incompatibilidad.

Las principales ventajas e utilizar un SMBD es la versatilidad, seguridad y si estamos en un proyecto lo suficientemente robusto entonces el uso de un SMBD es indispensable.

- h) Supón que deseas crear un sitio de videos similar a TikTok. 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 videos: el usuario que lo subió, la fecha de carga, las etiquetas, qué usuarios comentaron, cantidad de “Me gusta”, entre otros.

Todos los puntos son bastante relevantes, en la inconsistencia y redundancia de datos, debido a las múltiples relaciones entre los datos, por ejemplo que un creador puede tener múltiples videos si se usaran archivos comunes para almacenar dicha información, el usuario del creador quedaría escrito múltiples veces.

Al momento de registrar las views o likes, Habrá problemas de acceso concurrente y seria difícil mantener registro de estos eventos frecuentes.

Por cuestiones similares, buscar videos por creador por ejemplo, se vuelve difícil. También seria difícil escalar o modificar las funciones de la aplicación pues el esquema no seria flexible. Igualmente la seguridad estaría comprometida pues hay mucha información que no debe ser accesible al usuario.

- i) 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 tradicionales.

La llegada del “big data” y el requerimiento de procesar cantidades gigantescas de datos lleva a compañías como amazon (SimpleDB, DynamoDB) y google (Bigtable, MapReduce) a desarrollar sistemas manejadores de bases de datos que sean mas rápidos, altamente paralelizables y flexibles.

- j) 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: un diccionario con palabras desordenadas, un diccionario con palabras ordenadas (sin índice) y un diccionario con palabras ordenadas (con índice).

Para el diccionario con palabras desordenadas no tenemos de otra más que ir revisando una por una las palabras hasta encontrar Luminiscencia pues no tiene estructura que podamos aprovechar. Esto puede ser particularmente malo si Luminiscencia resulta ser la última palabra del diccionario.

Si tenemos las palabras ordenadas podemos usar un algoritmo de búsqueda binaria: Abrimos el diccionario a la mitad y vemos si la palabra es mayor o menor en orden lexicográfico a la primer palabra de la página, si es mayor hacemos recursión con el lado izquierdo del diccionario si es menor con la derecha. Al final debemos de terminar entre dos páginas y de ahí seguimos la búsqueda binaria ahora sobre la cantidad de palabras.

Finalmente con el índice solo tenemos que buscar la letra que buscamos y saltar directamente a ese capítulo, de ahí hacemos búsqueda binaria sobre las todas las palabras que empiezan con la letra L para encontrar Luminiscencia.

2. Lectura

- a) Leer el resumen ejecutivo ¿Por qué son tan importantes los datos? y realizar un resumen del documento (páginas 1 a 16), destacando los puntos que a su consideración sean los más relevantes (no más de una cuartilla).

Resumen

Nos encontramos al frente de una revolución. Los avances tecnológicos recientes han impulsado un avance secundario en si mismo, el de los datos masivos. Hemos pasado de ser capaces de almacenar unos cuantos archivos de texto en una computadora a poder guardar millones de libros sin esfuerzo alguno. Este cambio en almacenamiento y procesamiento de la información digital ha traído consigo una revolución en como hacemos negocios y tomamos decisiones.

La velocidad a la que producimos datos, ha crecido enormemente, se estima que hasta el 90 por ciento de todos los datos en existencia se crearon en los últimos dos años y esta abundancia ha transformado los datos en un nuevo y valioso recurso que además es renovable.

Para poder participar en esta riqueza, se propone que debemos de pasar por 4 etapas:

Primero, conseguir los datos. Esto que era tarea complicada hace algunos años ya no lo es, cientos de sensores cubren cada posible área de interés y su conectividad con internet significa que recogerlos.^{es} cuestión de milisegundos.

Segundo, estos datos necesitamos almacenarlos en algún lugar, pero de nuevo el progreso de tecnologías de almacenamiento significa que el costo representa una fracción minúscula comparada a la de inicios del siglo.

Tercero, la parte complicada: Hemos de analizar estos datos y transformarlos en información que nos permita tomar mejores decisiones. Para esto nos tendremos que apoyar de nuestra creatividad e ingenio para extraer hasta la última gota de valor. Aquí también el progreso en algoritmo y técnicas de análisis nos ayudarán.

Finalmente, usando lo que aprendimos podremos decidir en formas nuevas e innovadoras. Aprovechando avances como la inteligencia artificial y la toma automática de decisiones.

En conclusión, el cambio en magnitud del acceso a los datos se ha vuelto un cambio en tipo. Han desaparecido los límites de la humanidad. Ahora podemos predecir muchísimas cosas de nuestro futuro y explorar millones de posibilidades sin arriesgar nada. Este campo será nuestra guía hacia el futuro.

- 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 :...

Angel:

Considero que el texto leído es un intento mal realizado de persuasión acerca de el valor de los datos. No estoy en desacuerdo con que los datos son verdaderamente valiosos, pase meses en la lista de espera de DALL-E, sé de que son capaces los datos masivos, pero siento que en la lectura no se toman en serio los nuevos retos que nos han traído los datos.

En pocas palabras, "Por que son importantes los datos.^{es} un texto hablando de los beneficios que nos han traído los datos y como el pequeño rol que jugaban en el pasado se esta volviendo una verdadera oportunidad de progreso, se tocan las áreas de recopilación de datos, internet de las cosas, algoritmos, análisis de datos y computación en la nube.

La lectura se enfoca en las nuevas bondades que han traído estos datos a nuestra vida diaria, desde mejores decisiones en los negocios hasta beneficios en la medicina. A la vez se ignoran los cientos de nuevos problemas que nos han traído los datos, algoritmos de recomendación que nos desconectan de nuestros seres queridos, inteligencia artificial enfocada al daño de seres vivos y la extracción de hasta nuestro último centavo por parte de las grandes empresas de comercio.

La explosión cambrica en los negocios que han causado los datos masivos es una espada de doble filo, este cambio paradigmático en como nos relacionamos con el mundo no necesariamente será positivo a largo plazo, en lugar de la actitud temeraria y sobre optimista debemos de ser escépticos y cuidadosos, no sería la primera vez que un cambio en nuestra tecnología terminará por mordernos la mano.

Ian

El texto nos informa sobre el gran potencial que tienen los datos, además del gran valor que ya están agregando .actualmente" (2015?) a varios sectores, entre los cuales podemos encontrar el de la medicina, el transporte, la energía, las finanzas, la producción y la agricultura.

En general el protagonista principal del texto son los datos, pasando primero por todos los puntos de su ciclo de vida, su obtención, almacenamiento, análisis y aprovechamiento. para después centrarse en todo lo que estos datos pueden traer a nuestro mundo, desde ahorros económicos, aumentos a la productividad, mejoras en eficiencia y eficacia, e inclusive cosas que hace no mucho tiempo considerábamos imposibles, como vislumbrar un poco hacia lo que nos depara el futuro.

Aunque me siento optimista con el papel que jugaran los datos a futuro, la realidad es que no podemos seguir pensando que el crecimiento de todas las cosas puede ser exponencial indefinidamente, hace casi 2 décadas vimos el final del escalamiento de dennard, de la misma manera estamos cerca del fin de la ley de moore, y tarde o temprano también nos toparemos con otros límites físicos, como el principio de Landauer (cota inferior del consumo de energía por cálculo), o con el límite teórico máximo de datos que pueden ser transmitidos por una fibra óptica.

No por esto los datos presentan una oportunidad menos interesante, son la conclusión lógica de todo el progreso que hemos tenido en IT, además de ser una herramienta muy valiosa en todos los aspectos de la vida. Los retos que estos presentan, más que ser un motivo para desalentarnos, deben ser lo que nos motive a seguir adelante, buscando soluciones a los problemas para los que actualmente parece no haber solución alguna.

Rodrigo

El autor principalmente nos quiere dar a entender que los datos son y seguirán siendo un elemento clave en el desarrollo tecnológico y económico de nuestra sociedad. Nos invita a conocer la manera en que los datos son recabados y utilizados para la solución de una amplia cantidad de problemas e incluso menciona que el estudio de los mismos se está convirtiendo día a día en una disciplina indispensable para todos los sectores económicos. Concuerdo totalmente con esta postura, pues como bien dice el autor, cantidades enormes de datos se están produciendo diariamente; estos datos contienen información vital para conocer mejor el panorama/las dinámicas que ocurren en algún contexto (sector financiero, agricultura, medicina, producción, etc...) Por lo que aquella empresa incapaz de tomar provecho de estas nuevas tecnologías, simplemente estaría perdiendo la información que necesita para mantenerse vigente.

El artículo se relaciona con la materia porque su tema principal es justamente el objeto de estudio de esta, ya que sobre nosotros cae la responsabilidad del impacto que podrían tener los datos en los ámbitos sociales, económicos, etc, si no sabemos cómo organizarlos, es decir, nosotros tenemos que aprender a analizarlos y a manejarlos para así poder diseñar una base de datos eficiente en donde podamos guardarlos de manera que estos estén seguros, sea fácil poder acceder a ellos, podamos almacenarlos por mucho tiempo, podamos actualizarlos y la forma en la se guarden este organizada y optimizada, y así poder brindarle a las personas una buena herramienta con la que puedan trabajar y que esta sea fácil de usar.

Estoy de acuerdo con la postura del autor, principalmente en la parte de que se pueden utilizar los datos de una manera amigable con el medio ambiente, como bien pueden ser los edificios de energía positiva o de los sensores telemáticos para la captura de datos del rendimiento del motor, con lo que se ahorran millones de galones de combustible, por estas y otras razones se están ahorrando enormes cantidades de energía, y si se hace un mejor uso de los datos y se impulsa más, se pueden llegar a mejorar el medio ambiente a la vez que las empresas ahorran. Otro punto a considerar es que la economía es impulsada por los datos, ya que con un mejor uso de los datos se pueden innovar ciertas áreas económicas e impulsar el empleo, y de hecho se hizo una estimación de que el PIB mundial tendía a crecer por el uso de los datos. Pero estos no son los únicos sectores afectados de manera positiva por los datos, también se encuentran la medicina, la agricultura y ganadería, las TI, entre otros. Creo que el buen uso de los datos nos puede llevar a ser una sociedad más consciente con el medio ambiente y puede impulsarnos como sociedad para crecer tanto económicamente como en calidad de vida.

Fernando:

Bajo mi punto de vista el texto trata de convencernos de que los datos tienen y tendrán muchísimo poder en nuestra sociedad, que este poder se esta ganando exponencialmente pero que desafortunadamente hay una carencia de información (aunque abundancia de datos), y es aquí donde entra en papel nuestra materia ya que en Fundamentos de Bases de Datos nos enfocaremos en administrar de manera correcta, limpia y legible los datos para poder evitar algunos de los problemas que se presentan en el texto. Pudiendo sacar el máximo provecho que nos otorgan las tecnologías desarrolladas y en desarrollo teniendo a la mano bases de datos con muchísimo conocimiento valioso y tomar buenas decisiones en nuestras vidas (y/o negocios).

El texto gira al rededor del poder de los datos y sin duda nos hace ver que en este campo de los datos hay muchísima, pero muchísima demanda laboral, ya sea en la creación de las bases de datos, en la limpieza de los datos, en el análisis de los datos o en una mezcla de estas. Nos hace ver la necesidad que hay en el mundo de poder analizar, almacenar y consultar datos para el avance en el desarrollo de la sociedad y la economía.

A pesar de que la lectura se siente un tanto optimista, creo que solo se ven los lados buenos (que estoy totalmente de acuerdo, creo que con estas cantidades masivas de información

se pueden hacer maravillas como sociedad). No me gustan los superheroes, pero todo gran poder conlleva una gran responsabilidad, y el poder que nos pueden dar los datos creo que es inigualable, por lo que también se debe tener una enorme responsabilidad (la cual veo que no sucede siempre empresas como Meta). Hay muchísimo potencial aquí, pero si solo se ve para fines lucrativos de algunas empresas siento que se esta desperdiciando algo que puede resultar maravilloso para todos, y si solo se quieren datos para saber que producto eres mas propenso a comprar creo que no es el camino correcto.

A pesar de todo soy optimista y pienso que la evolución de estas tecnologías adquirirá un poder suficiente para que sin importar a donde nos lleve, sera un buen camino al final del día. Bajo mi punto de vista el texto trata de convencernos de que los datos tienen y tendrán muchísimo poder en nuestra sociedad, que este poder se esta ganando exponencialmente pero que desafortunadamente hay una carencia de información (aunque abundancia de datos), y es aquí donde entra en papel nuestra materia ya que en Fundamentos de Bases de Datos nos enfocaremos en administrar de manera correcta, limpia y legible los datos para poder evitar algunos de los problemas que se presentan en el texto. Pudiendo sacar el máximo provecho que nos otorgan las tecnologías desarrolladas y en desarrollo teniendo a la mano bases de datos con muchísimo conocimiento valioso y tomar buenas decisiones en nuestras vidas (y/o negocios).

Josue:

En general, creo que ya se escucha mucho esta visión optimista y benévola de los datos y su minería, es la norma, por lo que creo que deberíamos de poner mucho más énfasis en cómo podemos hacer un uso responsable y justo de ellos.

Los datos son un recurso extremadamente valioso y las herramientas que podemos crear con ello son muy poderosas, por eso mismo creo que necesitamos conscientizar mas a la población general en su uso y la gran industria que hay detrás, para que de esta forma, ambas partes informadas, puedan velar por sus intereses.

La presentación me parece solo propaganda corporativa, fuertemente sesgada, poniendo demasiado benevolente a quienes las usan.

- c) Mitos sobre los datos (páginas 17 a 27). Cada integrante del equipo deberá seleccionar al menos 2 mitos que le hayan parecido curiosos y/o interesantes sobre los datos e indicar qué pensaban antes de leerlos y cómo cambio su perspectiva después de la lectura.

Angel

Las personas no tienen control alguno sobre sus datos.

Siento que no se desmiente correctamente este mito, es verdad que algunas empresas te permitan no ser rastreado muchas no lo hacen y este es un caso donde un solo actor malo puede tener impactos enormes sobre la vida de las personas. En particular tracking a través del internet: un experimento acerca de rastreo digital

Voy a admitir que si hay un control que todos tenemos sobre nuestros datos y es en decidir no crearlos, lo que no digitalizas no existe en ninguna base de datos.

La única forma de que los datos puedan estar protegidos es que los gobiernos intervengan para exigir su protección.

Mi opinión acerca de este mito no cambia, fuera de tratarse de detener el progreso de la tecnología aislar los datos puede ser una medida para proteger los derechos de los ciudadanos de los países. Yo no puedo votar acerca de la legislación de datos en Estados Unidos pero si en mi propio país, por que quería dejar mi información en manos de quienes no están obligados a hacerme caso ?

Rodrigo

Los datos se utilizarán como una herramienta de exclusión a partir de la capacidad de reforzar los obstáculos que enfrentan las comunidades desfavorecidas y de bajos ingresos.

Si bien no creo que los datos son usados para mantener opresión sobre comunidades desfavorecidas, me parece que éstos raramente son usados para ayudar a dichas comunidades. Desde mi punto de vista, muchos de los beneficios que las ciencias de datos brindan únicamente son accesibles para personas con un cierto privilegio socio-económico. Según el artículo, los datos usados de manera *responsable* son una herramienta para estudiar las causas y las soluciones de problemáticas sociales como la marginación, la pobreza, la falta de acceso a la educación, entre otras. Desgraciadamente, los datos siguen sin ser usados de la manera adecuada; y en muchos países (incluyendo el nuestro) todas estas situaciones persisten.

No se puede confiar en las empresas que utilizan los datos.

Con controversias como fue el caso de Facebook, acusado de vender datos personales de sus usuarios a mas de 150 empresas, muchas personas (incluyéndome) tenemos la sensación de que nuestros datos personales son comerciados deliberadamente entre empresas con el objetivo de crear publicidad y promover la cultura del consumismo. El artículo nos dice que esto no sucede y que de hecho a las empresas les interesa cuidar la privacidad de sus usuarios. Personalmente no creo que este sea el caso de la mayor parte de las grandes vendedoras de servicios, pero es bueno saber que algunas empresas como Microsoft están adoptando políticas más estrictas para proteger la privacidad de sus clientes.

Fernando

A las empresas no les preocupa la protección de los datos personales.

Antes de leer el mito, creía que era verdad, pero después de leer su sección Realidad” me quedo en las mismas. Entiendo que hay empresas que si les interesa proteger nuestros datos, pero creo que no se puede generalizar. Lo único que nos puede garantizar que no darán un uso indebido a nuestros datos personales es si leemos los términos y condiciones de todo lo que firmamos, y no creo que decir que muchas empresas siguen buenas practicas sea suficiente para no dar un mal uso de nuestros datos personales (aunque tengan nuestro consentimiento ignorante).

Los datos nunca se pueden desidentificar por completo. La desidentificación de los datos no es efectiva.

Nunca había escuchado el termino ”desidentificar”, por lo cual este punto llamo mi atención. Fue algo interesante y agradable de saber que esto sea posible, ya que creo que puede haber muchas personas que de saber esto y en caso de ser posible la desidentificación, accederían de forma voluntaria y consciente a ceder sus datos.

Josue

-Los datos benefician únicamente al sector de la TI y no a otros sectores de la economía.

Me parece importante reconocer el valor y la riqueza que se puede generar a partir de los datos. Ciertamente, tienen un gran impacto positivo (aunque a veces también negativo) en la población general. Por ejemplo, existe gran potencial en la creación de herramientas que ayuden en el diagnostico médico, eso o creaciones similares abaratarían mucho servicios necesarios, lo cual los haría mas accesibles para la población general.

-La única forma de que los datos puedan estar protegidos es que los gobiernos intervengan para exigir su protección.

En este, también estoy de acuerdo. Me parece que definitivamente hay cierto rol para instituciones sociales como el gobierno regulen esta industria, pero en general creo que la competencia, la divulgación de información y el libre mercado son los mecanismos que mejor encontraran alternativas mejores para todos.

Ian

Una mayor cantidad de datos equivale siempre a mejores datos.

En el contexto de los datos es fácil pensar que más es mejor, usualmente pensamos en encontrar tendencias, patrones y cosas que en general es más sencillo obtener mientras más datos poseamos, pero en retrospectiva es cierto que tener datos limpios y relevantes aún en pequeñas cantidades es muy importante para la toma informada de decisiones.

Los países en desarrollo no están listos aún para sacar provecho de los análisis de datos

Yo pensaba que serían los países desarrollados los que gracias a su gran penetración en materia de IT recibirían la gran mayoría de los beneficios de los datos, pero ahora puedo ver que los datos van a permitir que los países que actualmente se encuentran en desarrollo sean mejores para superar los obstáculos en el camino que los países que se desarrollaron antes del advenimiento de los datos.