

Introducción

La generación de datos a través de los dispositivos electrónicos ha sido un factor determinante a partir del siglo XXI; aunado a esto, se han desarrollado unidades de almacenamiento cada vez con mayor capacidad, permitiendo que la acumulación de la información sea masiva. Sin embargo, no todo se puede considerar Big Data. En este tema se abordará el primer modelo importante de la analítica de datos, que se refiere a sus propiedades, características o dimensiones, también llamadas *tres v*, que son volumen, velocidad y variedad.

Según el sitio web Grupo Bit (s.f.), cada minuto las personas norteamericanas usan 3,138,420 Gb de datos de internet, y esta cifra sigue en aumento (*volumen*). Como ya sabes, las personas generan información por medio de sus dispositivos móviles segundo tras segundo (*velocidad*); generando imágenes, videos, datos de clima, señales GPS y muchas más formas (*variedad*). Sin duda este milenio puede considerarse como la era de la información, ya que nunca en la historia de la humanidad se había generado tanta *data* como en los últimos 20 años y además esta sigue en aumento.



¿Para qué sirve tanta información?, ¿qué beneficios tiene en nuestra vida? La información es relevante para todas las áreas del conocimiento, ya que esta ayuda a tomar mejores decisiones, a generar estrategias de publicidad, campañas de mercadotecnia, y en general, a lograr una comprensión profunda del consumidor y de tu mercado meta.

Incluso la información te puede permitir prevenir cuándo, cómo y dónde realizar el lanzamiento de un nuevo producto, así como la probabilidad de tener éxito. ¿En realidad es posible prever tal suceso con la información? Los hechos actuales indican que sí, que mientras exista una serie de datos recopilados adecuadamente es posible, a través de los modelos indicados del análisis de datos, poder prevenir situaciones y tomar decisiones para asegurar un resultado positivo con menor margen de error.

Antes de iniciar con la explicación, piensa en el siguiente caso:

Una mujer decide buscar en internet, sin reflexionar demasiado en sus acciones, todo tipo de artículos para bebé: ropa, muebles, decoración, etc. No está embarazada, o al menos no lo sabe, pero basada en intuición y/o en deseos personales, busca en internet y deja una huella clara. Enseguida de sus acciones empieza a recibir en su correo y redes sociales promociones de artículos relacionados a bebés. Todas esas señales de navegación que dejó en la red le permitieron rastrear sus pasos a alguien dispuesto a recopilarlas y analizarlas, e incluso adelantarse de tal forma que parece haber prevenido un evento que sucede en la realidad semanas después, cuando ella efectivamente descubre que será madre. El uso de datos personales y cada día más íntimos, puede resultar escalofriante y todo un reto a utilizar en tu área de trabajo. Es momento de seguir avanzando en este maravilloso mundo de la analítica de datos.

Explicación

Dimensiones del Big Data

De acuerdo con Indracompany.com. (s.f.), de Indra Software Labs, desde los inicios del concepto, IBM y Gartner, líderes reconocidos en la ciencia de datos, plantearon tres dimensiones del Big Data con el fin de definirla como "el conjunto de herramientas que trabajan conjuntamente y permiten captar, almacenar y gestionar un gran *volumen* de información *variada* a una alta *velocidad*, permitiendo obtener datos para la toma de decisiones". Estas tres dimensiones, propiedades o características, representan a su vez retos para las empresas que se enfrentan a problemáticas relacionadas con cada una de ellas.

De manera más reciente, es común poder encontrar diferentes fuentes como la empresa de software Zendesk, donde se habla ya de *cinco v*, añadiendo *veracidad* y *valor*. (Da Silva, 2021).

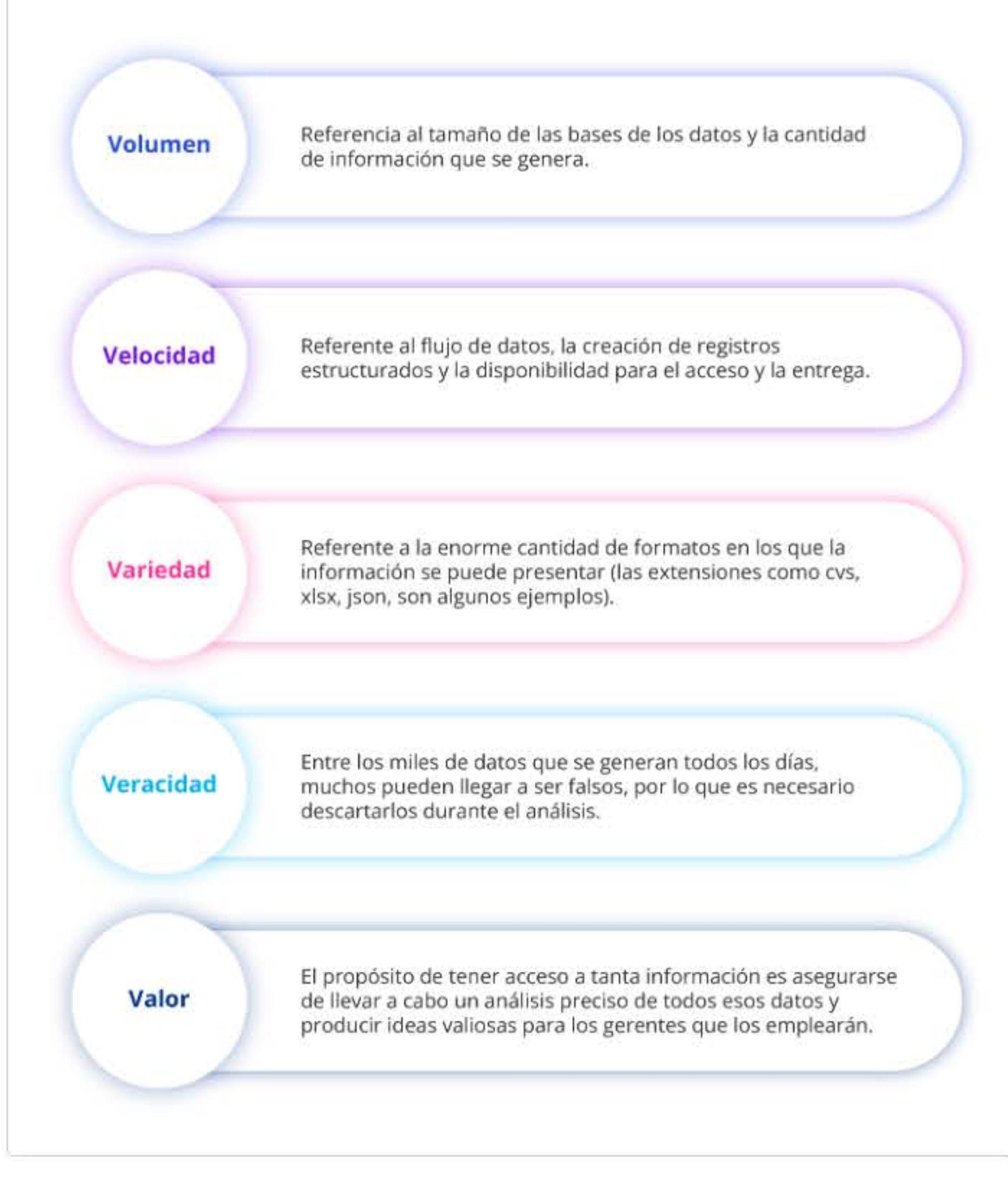


Figura 1. Cinco v de la analítica de datos.

Volumen

Al analizar una definición formal de la primera *v* de *volumen*, se refiere a la cantidad de datos electrónicos que son recopilados y validados en una base de datos. Dicha cantidad generada por segundo en el mundo digital es mayúscula; así que es necesario desarrollar nuevas tecnologías y formas de almacenamiento. ¿Te has preguntado cómo eran los discos duros hace 50 años?, ¿cómo te imaginas que son ahora? Bien, pues una imagen dice más que mil palabras. En la siguiente línea del tiempo, basada en Teitelroit (2021), puedes observar cómo han evolucionado las formas de almacenamiento de la información hasta hace algunos años.

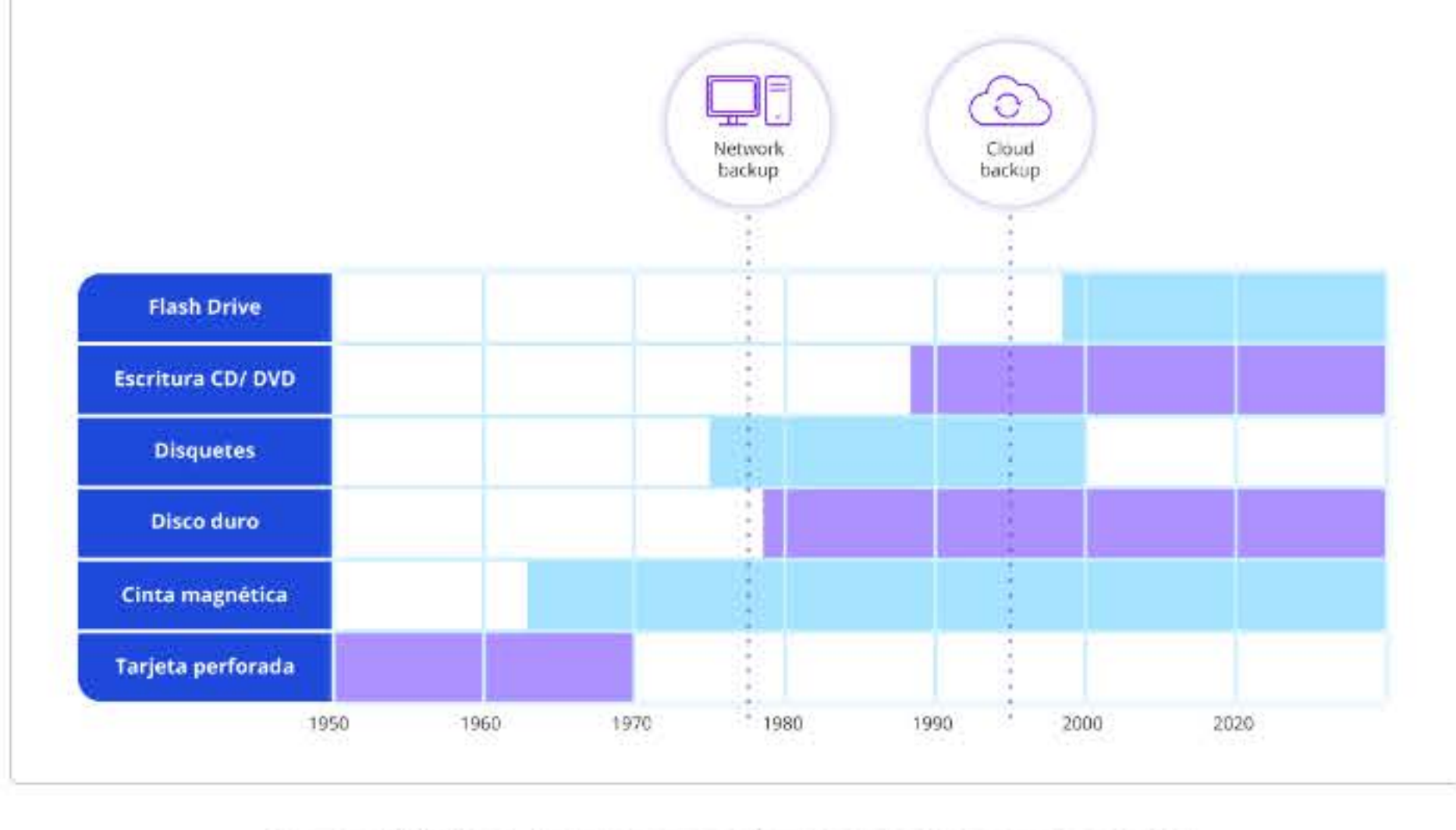


Figura 2. Historia del almacenamiento de la información digital.

La nanotecnología ha permitido la reducción significativa del tamaño del hardware de almacenamiento en las últimas décadas y se sigue trabajando en la creación de nuevas posibilidades con menor tamaño y mayor capacidad.

De acuerdo con el sitio Iberdrola (s.f.), empresa española dedicada a la producción, distribución y comercialización de energía, la nanotecnología es un campo innovador, se centra en manipular la estructura molecular de los materiales para alterar sus propiedades intrínsecas y crear aplicaciones revolucionarias. Un ejemplo destacado es el grafeno, una forma modificada de carbono que supera al acero en dureza, es más ligero que el aluminio y logra ser prácticamente transparente. Las nanopartículas también encuentran amplio uso en diversas áreas como la electrónica, la biomedicina y la defensa.

La disponibilidad de la información ha cambiado a lo largo del tiempo, incluso hace algunos años, las empresas disponían de información sobre sus clientes de forma anual, o por semestre, trimestre o mes. Ahora, con la tecnología, se acortó el tiempo en que la información puede estar disponible para consulta, no solo a formato semanal o diario, sino por minuto y segundo. En algunos campos como la ciencia, específicamente la física, la información puede almacenarse en nanosegundos.



La pregunta a partir de esto sería: ¿es relevante disponer de la información a cada instante? Imagina que estás en una tienda comprando lo que necesitas, y mientras compras, junto con otra docena de clientes, tus decisiones son recopiladas, procesadas y analizadas de inmediato, de forma que, antes de que acabe el día, la gerencia ya tiene suficiente información para tomar decisiones, tales como iniciar una campaña de última hora para intentar alcanzar esa meta de ventas que parece inminente que no va a lograr. Sin duda, el análisis instantáneo de datos permite tomar decisiones de negocios que pueden cambiar el rumbo de resultados que podrían ser desastrosos.

La cuestión no es responder si la información es útil, sino cómo y dónde almacenarla. Piensa en toda la información que tienes acumulada en tu propia casa: música, fotografías, videos, documentos electrónicos, etc. ¿Cómo la almacenas y dónde? ¿Cuánto espacio ocupa tu información personal? ¿Son Gigabytes o Terabytes? ¿Utilizas un hardware o la nube?

Velocidad

En un primer acercamiento al término velocidad, se puede decir que es el espacio recorrido por un objeto (distancia), en determinado tiempo. Si recuerdas tu clase de física, la fórmula de velocidad es la distancia sobre el tiempo recorrido. Cuando se habla de información, se refiere al concepto de velocidad, pero relativo al flujo de datos, es decir, la velocidad a la que los datos se almacenan (cuánto tarda en crearse un registro nuevo) y posterior a eso, la velocidad a la que se puede acceder a ellos de forma regular. En otras palabras, la *velocidad* es la rapidez a la que se producen los datos, y la disponibilidad de estos para la toma de decisiones.



Piensa que perteneces a la gerencia nacional de marketing de una empresa del sector automotriz, y quieres revertir la tendencia a la baja en las ventas. Uno de los problemas que tendrías que solucionar es el relativo a la velocidad de acceso de la información para mejorar y acelerar el proceso de toma de decisiones y mejorar el servicio al cliente, por ejemplo, ¿con qué velocidad podrías disponer de la información del inventario? El beneficio de que la gerencia pueda tener acceso en tiempo real a los registros de ventas, o que el mismo personal de ventas pueda conocer el estatus de un coche al momento, es de mayor importancia para proporcionar un buen servicio y para que las decisiones que se tomen sean las acertadas.

Si cuentas con un buen sistema de información que esté actualizado en tiempo real, podrás evitar situaciones como alguna por las que seguramente has pasado en la que, después de decidir realizar tu compra, cuando ya imaginabas que contabas con el beneficio del producto necesitado, quien vende regresa diciéndote que no está disponible, aunque el sistema le indicaba que sí lo tenía en tienda. Definitivamente no todas las empresas cuentan con esta capacidad.

¿Cómo te sientes como persona que compra al enfrentar estas situaciones? ¿Qué pasó con los datos? ¿La velocidad de acceso y de procesamiento fue la adecuada? Esta es la realidad de la velocidad de los datos, y es sumamente interesante, porque los problemas más comunes en los sistemas de información son justamente la estandarización y la velocidad en la que puedes disponer de los datos. La solución dependerá de la importancia que cada empresa les dé a estos aspectos para que la información esté disponible y a tiempo.

Finalmente, cuando se habla de redes sociales, es común que algunas publicaciones se hagan virales, ¿puedes imaginar la rapidez con que se genera dicha información y se difunde? Es sorprendente, por ejemplo, que algunos videos sean vistos millones de veces en un solo día. ¿Cómo es posible procesar esa información para hacer una campaña de publicidad y monetizarla?

En el siguiente esquema se puede observar el proceso de la velocidad de los datos, entendiendo por velocidad a la capacidad de poner disponibles los datos para su análisis una vez recopilados y almacenados (García et al., 2018).

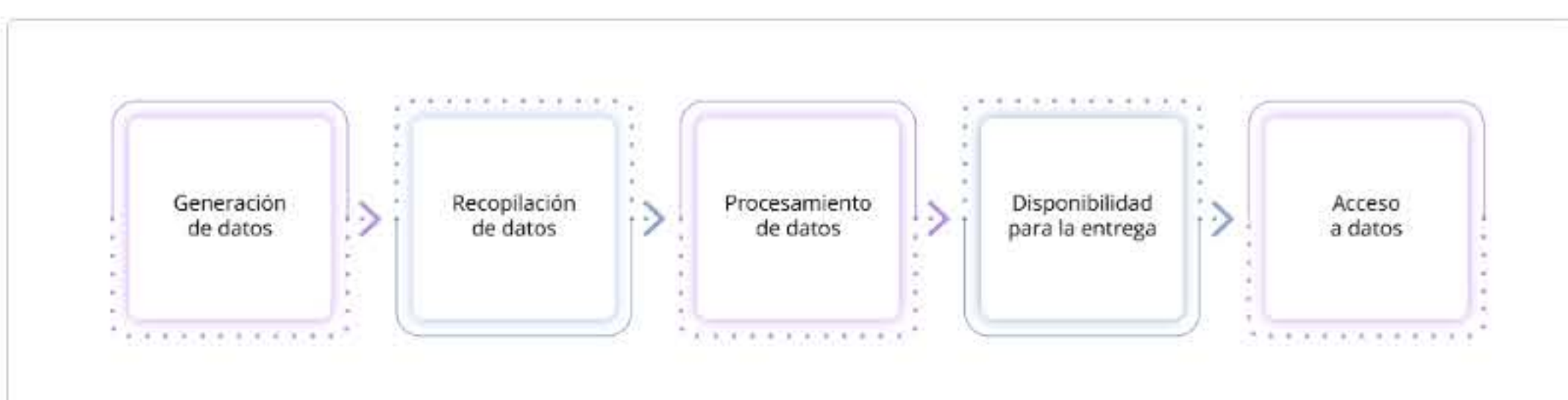


Figura 3. Proceso de la velocidad de los datos.

Variedad

Recuerda la última vez que hiciste una transacción bancaria por cualquiera de los medios disponibles para hacerlo. Reflexiona sobre la variedad tan amplia de datos que se generan en cada transacción; si estuviste en un cajero automático retirando efectivo, la información que es captada a través de la cámara de video, los datos de tu tarjeta, las claves de identificación, etc.; si lo hiciste por teléfono, desde la identificación de voz que utiliza el banco para reconocerte, hasta los números que generan para identificar la transacción. En caso de que utilices un sitio web, desde el navegador que utilizas, las cookies informáticas que se almacenan en tu equipo para identificarte en futuros accesos, etc. Se está hablando de video, imágenes, voz, datos y otros.



Toda esa información pertenece a una misma transacción y es necesario almacenarla y procesarla de forma agrupada, pero cuando con formatos muy diversos y esto puede ser complicado. Es lo equivalente a haber experimentado en tu computadora la dificultad de no poder abrir un archivo porque los programas que tienes no pueden asociar su formato. Mucha de la información que se genera derivada de transacciones de cualquier tipo en lo comercial, no son datos como tal, como es el caso de los audios, videos, imágenes, las coordenadas geográficas de una ubicación o el seguimiento de tu movimiento dentro de un centro comercial.

Piensa en los diferentes formatos que utilizas en tu ejercicio profesional. Bastará que abras tu procesador de texto preferido, intentes guardar un documento y revisar todos los formatos posibles en los que puedes hacerlo. La *variedad de datos* es toda aquella información que concierne a todos los lugares de donde se puede extraer datos, así como su almacenaje (Da Silva, 2021).

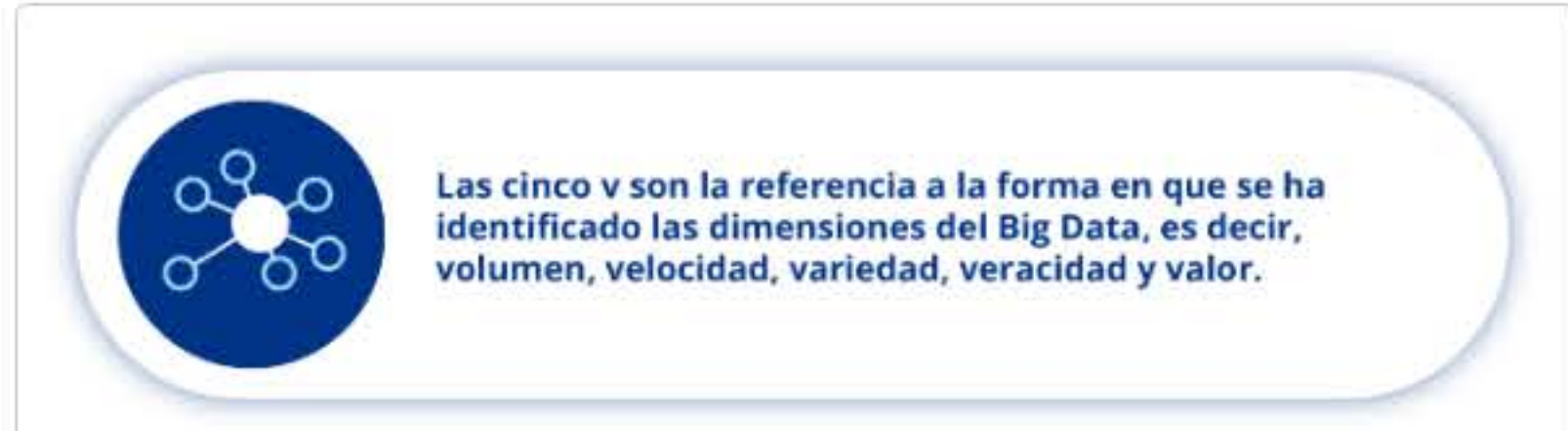
Veracidad

La *cuarta v* hace hincapié en la importancia de contar con información verdadera. Un ejemplo de la importancia de esto es el caso que todos conocemos del video, la foto, o incluso la publicación textual que nos llega al celular, pero que fue previamente modificada, ya sea para distorsionarla o para darle otro contexto. Como todos sabemos, la veracidad de la información es sumamente importante y debe ser una de las propiedades a considerar como parte de Big Data.

Valor

Da Silva (2021) incluso propone una *quinta v* como propiedad o cualidad del Big Data: el *valor*, en referencia a la importancia que tiene dicha información para la toma de decisiones de la empresa, ya que no toda la información tiene el mismo valor, o incluso alguno. ¿Qué tanto de la información que es almacenada en realidad es procesada y analizada para servir de apoyo a la toma de decisiones? Es necesario presente que el propósito máximo del Big Data es, mediante el análisis de información, generar información valiosa para la empresa.

Finalmente, la definición de García et al. (2018) dice que "Big Data ha de tener la capacidad de combinar una gran variedad de información digital en los diferentes formatos". Es decir, el reto fundamental del Big Data consistirá en transformar la información para que esté estructurada, organizada, para que tenga sentido y pueda analizarse para la adecuada toma de decisiones.



Cierre

En este espacio de aprendizaje has revisado con precisión y de una forma contextual la explicación de las *cinco v* de Big Data. Su importancia en la toma de decisiones de una empresa y las características que debe tener, ya que no todo gran volumen de información puede considerarse Big Data. De manera tal que las empresas, para determinar patrones de compra que les permitan identificar el perfil de sus clientes y establecer sus necesidades y motivaciones, deberán enfrentar los retos del Big Data y las dimensiones que lo caracterizan. Deberán encontrar la forma de resolver los retos de almacenar grandes *volumenes* de información, en sistemas que puedan procesar y analizar gran *variedad* de formatos, asegurando su *veracidad* y que todo esto se realice a gran *velocidad* para proporcionar información de *valor* para la toma de decisiones.

Es importante que la próxima vez que navegues por la web en tu computadora o en tu aplicación móvil favorita, reflexiones en lo siguiente: ¿Cuál es el volumen de mis datos?, ¿cuál es la velocidad a la que lo logro acceder a ellos y/o procesarlos?, ¿qué variedad de formatos tiene? En la medida que relaciones los temas con la vida real, tu aprendizaje será significativo y habrás logrado la meta.

Checkpoints

Asegúrate de:

- Comprender la relevancia de las *cinco v* de Big Data.
- Identificar el significado de volumen, velocidad, variedad y valor de los datos.
- Comprender la relevancia de las *cinco v* en la toma de decisiones de las empresas.
- Distinguir la diferencia entre volumen, velocidad, variedad, veracidad y valor de los datos.

Referencias

- Da Silva, D. (2021). *¿Qué es el big data y para qué sirve?* Recuperado de <https://www.zendesk.com.mx/blog/big-data-que-es/>
- García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, Á., y Padilla, W. (2018). *Ciencia de Datos, técnicas analíticas y aprendizaje estadístico*. México: Alfaomega.
- Grupo Bit. (s.f.). *¿Cuántos datos se producen en un minuto?* Recuperado de <https://business-intelligence.grupobit.net/blog/cuantos-datos-se-producen-en-un-minuto>
- Iberdrola. (s.f.). *Aplicaciones de la nanotecnología. Nanotecnología: una pequeña solución a los grandes problemas*. Recuperado de <https://www.iberdrola.com/innovacion/aplicaciones-nanotecnologia>
- indracompany.com. (s.f.). *¿Cuántas V debería tener el Big Data?* Recuperado de <https://www.indracompany.com/es/blog/que-cuantas-v-deberia-big-data>
- Teitelroit, L. (2021). *A brief history of data backup*. Recuperado de <https://itsupportguys.com/it-blog/a-brief-history-of-data-backup/>

Todos los derechos reservados @ Universidad Tecmilenio

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECMILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor. El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECMILENIO. Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECMILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste a la propiedad del material.