



# **INDUSTRY 4.0** for **VET**

## **3. BIG DATA**

### 3.1 Concepto

#### Introducción

¿Sabes que un ser humano tardaría unos 181 millones de años en descargar todos los datos de Internet? Estas grandes cantidades de datos que están disponibles hoy en día, y la forma en que se procesan, se llama **Big Data**.

Como verás en esta unidad, nos enfrentamos al Big Data casi a diario, a menudo sin nuestro conocimiento. Aprenderás las **ventajas**, pero también los **peligros que supone** el Big Data y por qué el **manejo correcto** de estas grandes cantidades de datos es a menudo más importante que los datos en sí.



#### La relevancia práctica – Para esto necesitas el conocimiento y las habilidades

No sólo los especialistas en informática, casi todo el mundo se encuentra con las llamadas grandes cantidades de datos en **situaciones cotidianas**, como cuando se visita al médico, se navega por las redes sociales como Facebook e Instagram, se hace una búsqueda en Google o en un vehículo de la red.

Saber cómo se utilizan grandes cantidades de datos y qué **oportunidades**, pero también **peligros**, se asocian a ello puede ser relevante tanto para el **uso personal** de Internet como para la **vida profesional**, tal vez en una empresa que analiza grandes cantidades de datos.

#### Objetivos y competencias de aprendizaje

Esta unidad de aprendizaje te proporcionará conocimientos básicos sobre el Big Data. Conocerás el **modelo 3-V** y aprenderás cómo se recogen y analizan grandes cantidades de datos. También aprenderás sobre los propósitos para los que se utiliza el **conocimiento** obtenido de las grandes cantidades de datos y los **riesgos** que implican. Comprenderás por qué la **protección de datos** se ha vuelto cada vez más **importante** en los últimos años y verás cómo el **manejo** de grandes cantidades de Big Data plantea grandes desafíos tanto para las empresas como para los particulares.

#### Objetivos de aprendizaje

Entender y describir el término Big Data.

Saber cómo utilizar Big Data.

Entender y explicar cómo se seleccionan las grandes cantidades de información.

Entender los retos y los riesgos del Big Data.

## 3.2 ¿Qué es el Big Data?

¿Sabías que alrededor del 90 por ciento de todos los datos disponibles hoy en día en el mundo se generaron en los últimos años? Debido a las numerosas nuevas tecnologías de la información y comunicación, el **volumen de datos** en todo el mundo ha crecido increíblemente y ofrece posibilidades hasta ahora desconocidas. El **Big Data** representa este **volumen de datos** estructurados y no estructurados, que no pueden ser procesados con software o hardware convencional debido a su tamaño.



Estos volúmenes de datos se crean, entre otras cosas, con cada uno de nuestros **clics en Internet**. Esto puede ser, por ejemplo, una compra en Amazon, una consulta de búsqueda en Google, actividad en redes sociales como Instagram o Facebook, etc.

Sin embargo, las grandes cantidades de datos por sí solas no constituyen el Big Data. Sólo el **análisis y el procesamiento** de estos volúmenes de datos, por ejemplo, por una empresa, constituye el Big Data. En 2001, el analista **Doug Lane** definió el concepto de Big Data, estableciendo un **modelo 3-V** que sigue vigente hoy en día. Según Lane, el Big Data tiene las siguientes tres características:

- **Volumen:** Las empresas reúnen grandes volúmenes de datos de diversas fuentes. Estas incluyen dispositivos inteligentes (IO) como teléfonos móviles, videos, medios sociales, etc. En el pasado, no habría sido posible almacenar estos grandes volúmenes de datos; hoy en día, existen plataformas de almacenamiento para ello.
- **Velocidad:** Las empresas procesan flujos de datos a velocidades sin precedentes y a grandes velocidades.
- **Variedad:** Los datos recogidos son diversos y tienen una gran variedad de formatos: los datos numéricos, que están disponibles en forma estructurada y almacenados en bases de datos ordinarias, pueden formar parte del Big Data, así como los documentos de texto no estructurados, los datos de las transacciones financieras o los correos electrónicos.

### Definición

#### Big Data

... representa una **gran cantidad** de **datos** disponibles que se **analizan** y **procesan** para un propósito específico. Según Doug Lane, el Big Data se caracteriza por su **volumen, velocidad y diversidad**.

## Big Data vs. Small Data



“Let’s shrink Big Data into Small Data ...  
and hope it magically becomes Great Data.”

A diferencia del Big Data, el Small Data se refiere a los datos en un volumen y formato accesible para los humanos. Los siguientes puntos muestran cómo se pueden distinguir el Big Data del Small Data:

- **Objetivos:** el Small Data se usa para un objetivo concreto; el uso del Big Data, a menudo se desarrolla sin organización.
- **Ubicación:** el Small Data se almacena generalmente en un lugar, por lo general en un solo archivo del ordenador, mientras que el Big Data suele estar repartidos en numerosos archivos en diferentes servidores ubicados en diferentes países.
- **Estructura de datos:** el Small Data está estructurado en línea recta, mientras que el Big Data puede no estar estructurado y puede tener muchos formatos de archivo de diferentes campos.
- **Preparación de los datos:** en la preparación del Small Data sólo suele participar un usuario final. Sin embargo, en el caso del Big Data, a menudo un grupo de personas prepara los datos, otro grupo los analiza y finalmente un tercer grupo los utiliza. Cada uno de estos grupos puede tener objetivos diferentes.
- **Durabilidad:** el Small Data generalmente se conserva durante un cierto período de tiempo después de la finalización de un proyecto. En el caso del Big Data, sin embargo, los datos permanecen almacenados por un período de tiempo ilimitado.
- **Origen:** el Small Data se almacena en poco tiempo y en unidades de medida específicas. El Big Data, por otro lado, se originan en diferentes lugares, países, compañías, organizaciones, etc.
- **Reproducibilidad:** el Small Data generalmente puede ser reproducidos completamente. El Big Data, por el contrario, se origina en muchas fuentes diferentes y está disponibles en muchas formas por lo que su reproducción es imposible.

- **Calidad:** los significados de los datos en un conjunto de Small Data son inequívocos, estos datos pueden por lo tanto describirse a sí mismos. El Big Data, por el contrario, son mucho más complejos y también pueden contener información no identificable que no tiene un significado específico. Esto puede reducir la calidad de los datos.
- **Análisis:** para el análisis del Small Data suele bastar un único proceso, ya que los datos se analizan a partir de un solo archivo informático. En el caso del Big Data, los datos deben ser extraídos, comprobados, reducidos, en un proceso que requiere mucho tiempo.

Como puedes ver por la distinción entre Big Data y Small Data, es a menudo, difícil de comprender.

### 3.3 Utilidad y oportunidades que ofrece el Big Data

El análisis de grandes cantidades de datos permite obtener **conocimientos**. Estos resultados pueden servir de base para decisiones, por ejemplo, sobre la **dirección estratégica** de la empresa. Las empresas, por ejemplo, quieren saber más sobre las preferencias de sus clientes para adaptar su gama de productos, la publicidad, etc. a ellas.

**Para el Deep Learning, o “aprendizaje profundo”** también se utiliza el Big Data: es un método especial de procesamiento de información y una subárea de aprendizaje automático. Una máquina "se alimenta" con grandes cantidades de datos, los cuales son analizados y utilizados para entrenar a la máquina. La máquina es capaz de vincular nueva información entre sí y sobre esta base puede hacer pronósticos y tomar sus propias decisiones. Sin embargo, el resultado es sólo tan bueno como los datos de los que la máquina ha "aprendido".



Un ejemplo es un sistema de traducción automática que "aprende" a traducir correctamente los términos técnicos en una empresa introduciendo datos (traducciones existentes).

Además, **las autoridades** y **los servicios secretos** utilizan el Big Data para detectar discrepancias y anomalías que podrían indicar actividades delictivas o terroristas. En **la ciencia**, se utilizan el Big Data para investigar **fenómenos naturales complejos** como el cambio climático o la aparición de terremotos y epidemias.

Sin embargo, el Big Data no siempre se manejan de manera **responsable**. Algunas empresas o instituciones no se adhieren a las normas de protección de datos, lo que significa que la información se da a conocer al

público. Esto puede ser trivial, pero en algunos casos también puede ser peligroso y dar lugar a **fraude y chantaje**.

<p><b>Ejemplo</b></p> <p>En 2015, el portal de aventuras de Ashley Madison, donde la gente que buscaba una aventura extramarital podía crear un perfil, fue víctima de un ataque de hackers. Como resultado, la información sobre las personas registradas en el portal se publicó en Internet. La información sobre las aventuras de los famosos y la información personal, así como los números de tarjetas de crédito se hizo pública. Además, se pidió por correo electrónico a los afectados que pagaran un rescate para que su pareja no se enterara del perfil en el portal de aventuras.</p>
<p><b>Recuerda</b></p> <p>Se pueden utilizar el Big Data para los siguientes <b>finés</b>, entre otros:</p> <ul style="list-style-type: none"> <li>• Orientación estratégica de las empresas.</li> <li>• Aprendizaje profundo.</li> <li>• La lucha contra la delincuencia y el terrorismo.</li> <li>• La investigación científica de los fenómenos naturales (por ejemplo, los terremotos y el cambio climático).</li> <li>• Evaluaciones ilegales que pueden conducir al chantaje o al fraude.</li> </ul> <p>El factor decisivo en relación con Big Data no es tanto los datos en sí como lo que les sucede.</p>

**Las empresas** en particular se benefician del análisis y la evaluación del Big Data. Tanto consciente como inconscientemente, generan y almacenan grandes cantidades de datos hoy en día. A continuación, aprenderás en detalle qué posibilidades ofrece a las empresas el análisis correcto del Big Data.

**Toma de decisiones**

Analizando la gran cantidad de datos generados en la empresa, las compañías pueden identificar patrones y filtrar la información. Esto permite a las empresas tomar mejores decisiones comerciales que aumentan el éxito de la empresa. Evaluando los datos de las máquinas, por ejemplo, es posible calcular en qué intervalos se avería una máquina. La compañía puede usar este conocimiento para dar servicio a la máquina antes de que falle. Big Data también se utiliza en la industria financiera y de seguros para calcular mejor los riesgos.

<p><b>Ejemplo</b></p> <p>La Sra. Schmidt tiene 47 años y le gustaría contratar un seguro médico privado. Al visitar a su corredor de seguros, se sorprende de los altos costes. Resulta que su proveedor analiza el Big Data para calcular mejor los costes del seguro individual. La empresa averigua, por ejemplo, qué riesgos de salud particulares corren las mujeres de esta edad que, como la Sra. Schmidt, son fumadoras, no tienen hijos y nunca hacen deporte.</p>
---

**Aumentar la eficiencia**

La competitividad es muy importante para las empresas. Para mantenerse al día con la competencia, las empresas deben diseñar estrategias para ahorrar costes sin comprometer el rendimiento. El análisis y la conexión del Big Data ayuda.

<p><b>Ejemplo</b></p> <p>¿Has oído que los conductores de la empresa de mensajería UPS casi siempre giran a la derecha? Eso es porque UPS ha descubierto, a través de un análisis de Big Data, que esto puede ahorrar unos 10 millones de dólares al año. Probablemente te preguntes cómo es posible: la fusión de varios conjuntos de datos, como las estadísticas de accidentes, los datos de consumo de combustible, etc., ha demostrado que</p>
---

los vehículos de UPS son mucho menos propensos a tener accidentes si no giran a la izquierda. Esto puede ahorrar mucho dinero, incluso si las rutas se vuelven más complicadas como resultado.

### La predicción en la investigación y el desarrollo

Al dar a conocer a los clientes actuales o potenciales su preferencia por determinados productos, la investigación puede identificar y predecir las tendencias, diseñar estrategias de comercialización adecuadas y desarrollar productos a medida. Con los métodos analíticos apropiados, también es posible, por ejemplo, predecir la seguridad de ruptura de un producto mientras se está desarrollando.

#### Ejemplo

El operador de una tienda online instala cookies y seguimiento online y rastrea los movimientos de sus visitantes. Puede determinar de dónde vienen los visitantes, en qué productos hacen clic, con qué frecuencia visitan el sitio, etc. Con la ayuda de estos datos, el operador puede adaptar el contenido del sitio y los productos ofrecidos a las preferencias de los visitantes y así aumentar su volumen de negocios.

### Servicio de atención al cliente personalizado

Al almacenar las decisiones de los clientes, las empresas pueden proporcionarles un servicio personalizado. Por ejemplo, si un usuario ve una película o serie determinada en Netflix, el sistema la guardará y la próxima vez que el usuario inicie sesión, las recomendaciones se basarán en las películas o series que el usuario ya haya visto. Pero esta oferta personalizada no siempre es acertada:

#### Ejemplo

Cuando el Sr. Maier se da cuenta de que sus viejas botas de montaña ya no sirven, busca en Google "botas de montaña nuevas para hombres". Se siente abrumado por las diferentes ofertas y también descubre que muchos productos no pueden ser entregados en su país de origen, Austria. El Sr. Maier decide obtener asesoramiento personal en una tienda especializada y también compra un par de botas de montaña. Sin embargo, en los días y semanas siguientes ve cada vez más publicidad de botas de montaña en Internet, ya que su consulta de búsqueda se ha guardado y analizado en Google. El Sr. Maier está irritado y se siente observado. Decide no hacer más consultas de búsqueda en Google en el futuro.



### Recapitemos:

#### Recuerda

Las empresas tienen numerosas oportunidades de utilizar Big Data para tener más éxito. Estas incluyen:

- **La toma de decisiones:**  
El análisis de Big Data permite a las empresas tomar mejores decisiones comerciales y evaluar mejor los riesgos.
- **Aumento de la eficiencia:**

El análisis y la vinculación de los datos (como los datos meteorológicos y de congestión con los precios del combustible) ayuda a las empresas a hacer más eficientes los procesos.

- **Previsión en el campo de la investigación y el desarrollo:**  
Con la ayuda de Big Data, se pueden hacer predicciones sobre tendencias, características de un producto, etc.
- **Servicio de atención al cliente personalizado**  
Al almacenar las decisiones tomadas por los clientes, las empresas pueden ofrecerles un servicio personalizado en su próxima visita.

### 3.4 ¿Cómo se analiza el Big Data?

Has aprendido cómo se define Big Data y qué opciones hay para utilizar las grandes cantidades de datos. En este capítulo, entraremos en más detalles y trataremos el **análisis** de Big Data. Este campo especializado se conoce como **Análisis de Big Data**.



#### Análisis de Big Data – Teoría

El primer paso es reunir **grandes cantidades de datos** de diferentes fuentes, que tienen diferentes formatos. Esto se hace a menudo mediante consultas de búsqueda. Luego los datos se preparan para su posterior procesamiento. Uno de los problemas suele ser que el Big Data está disponible en forma no estructurada y en formatos completamente diferentes y, por lo tanto, no puede ser capturado por los programas informáticos convencionales de las bases de datos.

Por lo tanto, Big Data Analytics utiliza **procesos complicados** para extraer y capturar los datos. Los datos son entonces analizados usando un software especial de Big Data. Finalmente, los resultados son procesados y presentados.

Es importante que el programa informático utilizado sea capaz de ejecutar rápidamente muchas solicitudes de búsqueda e importar y procesar rápidamente los diversos registros de datos. Para ser aún más potentes, muchos sistemas no utilizan el **espacio del disco duro** (como las aplicaciones de bases de datos convencionales) para el procesamiento de datos, sino la **memoria principal**, que suele ser mucho más rápida. De esta manera, se puede aumentar la velocidad de acceso y se pueden realizar análisis casi en tiempo real.

#### Recuerda

El **análisis** de Big Data puede dividirse a grandes rasgos en tres áreas diferentes:

- Obtención de datos de muchas y diversas fuentes mediante consultas de búsqueda.
- Evaluación y optimización de los datos recogidos.
- Análisis de datos y resumen y presentación de resultados.

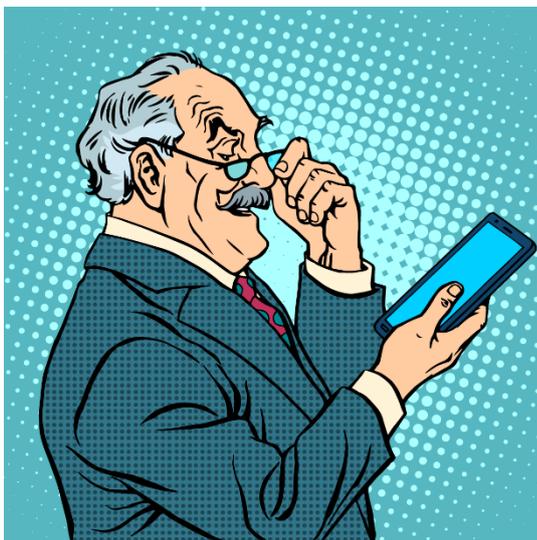
Un **software potente y adecuado** es muy importante para eso.

### Big Data Analytics – En la práctica

Es interesante observar que el Big Data Analytics está todavía en sus inicios en la mayoría de las empresas y las **oportunidades** que ofrece **todavía no se han agotado**. De media, las empresas analizan sólo un poco más de **un tercio** de los datos generados por el contacto digital con sus clientes (por ejemplo, a través de tiendas en línea o sitios web).

La razón son las estrictas regulaciones de **protección de datos** que hacen más difícil el análisis del Big Data. Las leyes y regulaciones que rigen la protección de datos se discuten con más detalle en el siguiente capítulo. Sin embargo, en realidad, en muchos aspectos las empresas todavía no están preparadas para utilizar eficazmente las grandes cantidades de datos para sí mismas. Las siguientes áreas juegan un papel importante:

En primer lugar, es aconsejable distribuir los resultados correctamente: las **fuentes de datos** deben proceder de diferentes áreas, los resultados deben ser utilizados en varias áreas de la empresa. También se requiere una **estrategia** adecuada: una empresa debe saber con qué propósito se analizan las grandes cantidades de datos. También es muy importante una **cultura empresarial** adecuada: las nuevas tecnologías, por ejemplo, no deben rechazarse en principio, sino considerarse de manera realista.



La mayoría de las empresas no tienen su propio departamento de análisis de datos. No obstante, algunos empleados deben traer consigo los **conocimientos necesarios** o adquirirlos en cursos de capacitación. Puede ser necesario contratar nuevos empleados. Las responsabilidades y autorizaciones también deben definirse dentro de la empresa.

Para el análisis se requiere una **tecnología eficiente**, en forma de **herramientas** adecuadas de **análisis de Big Data**. Sin embargo, las herramientas adecuadas dependen de la estrategia previamente definida o del propósito definido del análisis. Por último, pero no por ello menos importante, una estrategia adecuada de protección de datos también es esencial para garantizar que los datos personales de los individuos no se revelen al público. Un experto en protección de datos de la empresa se encarga de que el análisis de los datos cumpla con las leyes y reglamentos aplicables.

#### Recuerda

En resumen, los siguientes puntos son importantes para que el **Big Data Analytics** tenga **éxito** en una empresa:

- Una estrategia de Big Data - definiendo el propósito del análisis.
- Una cultura corporativa adecuada - apertura a las nuevas tecnologías.
- Personal con los conocimientos necesarios - formación o reclutamiento.
- Una tecnología poderosa - herramientas apropiadas de análisis de grandes datos.
- Una política de protección de datos adecuada - el cumplimiento de las leyes y reglamentos aplicables.

### 3.5 Retos y peligros del Big Data

En los capítulos anteriores ya has aprendido lo complejo que puede resultar analizar y utilizar el Big Data. Los desafíos y riesgos asociados al Big Data son igual de complejos. Probablemente el mayor **desafío** para las empresas en relación con los Big Data es la **protección de datos**:



Aunque las empresas han prestado más atención a la protección de datos en los últimos años, todavía hay problemas. Por ejemplo, los datos personales de los usuarios de Internet se utilizan sin su consentimiento y las personas afectadas pueden ser identificadas, controladas y, en el peor de los casos, chantajeadas.

#### Definición

#### Datos personales

... se refiere a los **datos** que se relacionan con una **persona** y permite sacar conclusiones sobre su **personalidad**. Esto incluye, por ejemplo, el número de matrícula de Werner Kogler, la fecha de nacimiento de su vecino o el saldo de la cuenta de Bill Gates.

Un ejemplo de **violación de la protección de datos** en relación con el Big Data es el caso del portal de aventuras de Ashley Madison, que ya se mencionó como ejemplo en el capítulo 2. En este caso, los **datos personales** se hicieron **públicos** y se utilizaron para **chantajear** a los propietarios de los datos.

Los reglamentos y leyes de protección de datos ayudan a proteger a los consumidores del abuso. La **base** de la **ley general de protección de datos** en la Unión Europea y en Austria es el **Reglamento general de protección de datos**, que entró en vigor el 25 de mayo de 2018.



### Excursus

## El Reglamento General de Protección de Datos

El Reglamento General de Protección de Datos, o GDPR por sus siglas en inglés, se denomina en su totalidad "Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, y por el que se deroga la Directiva 95/46/CE". Es directamente aplicable en Austria y se complementa con la Ley de protección de datos y la Directiva sobre protección de datos.

Este reglamento permite a los ciudadanos de la UE controlar mejor la recopilación y el uso de sus datos personales. Esto debería fortalecer la confianza de los consumidores en las empresas individuales. Los derechos existentes de los ciudadanos de la UE se consolidan en el RGPD, y también se establecen nuevos derechos. Los derechos establecidos en la RGPD incluyen:

- **El acceso simplificado a los datos personales**, lo que incluye el suministro de información completa, clara y comprensible sobre el procesamiento de los datos.
- Un nuevo **derecho a la transferencia de datos** – los datos personales serán transferidos de manera simplificada.
- Un **derecho más claro a la supresión ("derecho al olvido")** – los datos se suprimen si un ciudadano no está de acuerdo con que se procesen sus datos y no hay una razón legítima para conservarlos.
- Derecho a ser informado **sobre los datos personales pirateados** – las empresas y organizaciones informarán inmediatamente a las personas afectadas sobre las violaciones graves de la protección de datos personales. La autoridad de supervisión de protección de datos responsable también debe ser notificada.

En el caso de las empresas, la RGPD tiene por objeto crear más oportunidades de negocio y promover la innovación con numerosas medidas. Entre ellas se incluyen:

- La creación de **normas uniformes para toda la UE**, lo que supondrá un gran ahorro.
- **El nombramiento de un oficial de protección de datos** dentro de las autoridades y empresas que se ocupan de grandes conjuntos de datos.
  - **La designación de un único punto de contacto** en su propio país al que las empresas deben dirigirse.

- La creación de **normas de la UE para las empresas de terceros países** a las que las empresas de terceros países deben adherirse cuando ofrecen bienes o servicios o controlan el comportamiento de las personas
- La creación de **normas que promuevan la innovación** y garanticen que las normas de protección de datos se tengan en cuenta en una etapa temprana del desarrollo de los servicios o productos
- El uso de **técnicas compatibles con la protección de datos**, como la **seudonimización** (sustitución de los pasajes de un registro de datos que permiten identificar a la persona asociada) y la **encriptación** (los datos se encriptan para que sólo puedan ser leídos por personas autorizadas).
- La eliminación de las obligaciones de **presentación de informes** de las empresas a fin de promover la libre circulación de datos personales en la Unión Europea.
- La realización de **evaluaciones de impacto** cuando el procesamiento de los datos pueda amenazar los derechos y libertades de las personas.

El Reglamento General de Protección de Datos completo puede consultarse en <https://eurlex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A32016R0679>

Otro problema es que los empleados actuales de las empresas no siempre tienen la **experiencia** necesaria y no están abiertos a las posibilidades que ofrece a la empresa el análisis de grandes cantidades de datos.

A menudo se desperdician tiempo y recursos porque los implicados no tienen claro el objetivo de un proyecto de Big Data o qué infraestructura se requiere para el mismo. Encontrar y retener **empleados competentes** suele ser difícil para las empresas, ya que tienen una gran demanda.



Además, la **tecnología de Big Data** es diversa y **confusa** para los principiantes. ¿Has oído hablar de Spark, Hadoop MapReduce, Cassandra o Hbase? Estas son tecnologías de Big Data con diferentes características y beneficios.

Además, las tecnologías evolucionan a un ritmo rápido, por lo que las empresas a menudo simplemente no pueden seguir el ritmo de adopción. Por lo tanto, para las empresas que están considerando utilizar un análisis de Big Data, el **asesoramiento de expertos** es útil.

Otro punto es que los proyectos de Big Data son muy **costosos**. Esto se aplica tanto a las empresas que optan por un modelo de dominios locales como a las que prefieren un modelo de nube. La diferencia es que, con

un modelo de **dominios locales**, la empresa utiliza el software de Big Data en su propio centro de datos y es responsable de su operación y mantenimiento. En un **modelo de nubes**, en cambio, el software sólo es alquilado por la empresa y los datos permanecen en el proveedor.

#### Definición

##### Modelo de Dominios Locales

... se refiere a una solución en la que la empresa **compra** o arrienda el software de Big Data y lo despliega en su **propio centro de datos**. La compañía tiene que ocuparse del hardware en sí, y también se responsabiliza del uso del software y los datos se almacenan en la empresa.

#### Definición

##### Modelo de la Nube

... se refiere a una solución en la que una empresa adquiere el software de Big Data como un **servicio**; el proveedor asume la responsabilidad del mantenimiento y la operación. La empresa paga un precio de alquiler que incluye el hardware, la operación y los costes de mantenimiento. Con esta solución, los datos se almacenan en el proveedor.

Si una empresa decide utilizar el modelo bajo demanda, debe invertir en un nuevo hardware y contratar nuevos empleados para operar y mantener el sistema. En el caso de una solución en la nube, la empresa sólo necesita contratar empleados para operar y mantener el sistema, y la empresa debe pagar por los servicios de la nube.

Después de todo, la **calidad** de los datos suele ser deficiente, y las empresas se enfrentan al reto de armonizar los datos de diferentes fuentes de calidad variable. Por ejemplo, un comerciante en línea analiza datos de medios sociales, registros de sitios web, centros de llamadas y sitios web que tienen diferentes formatos.

Pero incluso cuando todos los problemas mencionados han sido resueltos, a menudo no es tan sencillo para las empresas obtener **conocimientos** útiles de las grandes cantidades de datos. Si la información se vincula **entre sí y se sacan conclusiones erróneas**, por ejemplo, esto puede ser peligroso.

Por ejemplo, una persona puede ser considerada no digna de crédito por un banco que realiza un análisis de Big Data porque vive en el mismo vecindario que muchas personas no dignas de crédito y conduce el mismo automóvil que muchas personas consideradas no dignas de crédito. El siguiente ejemplo también muestra por qué es crucial el uso correcto de las grandes cantidades de datos:

#### Ejemplo

Un minorista en línea se basa en Big Data Analytics, que se basa en datos históricos sobre el comportamiento de los clientes. Resulta que las personas que compran zapatillas negras a menudo añaden un par de calcetines negros. El minorista ajusta su gama para la primavera en consecuencia. Sin embargo, justo antes del comienzo de la primavera, un conocido rapero publica una foto suya con zapatillas negras y calcetines amarillos en Instagram. Por lo tanto, muchos jóvenes buscan calcetines amarillos para sus zapatillas negras, pero lamentablemente el vendedor en línea pronto se queda sin ellos porque no estaba preparado para esta demanda. El minorista simplemente utilizó la estrategia equivocada de Big Data, confiando sólo en los resultados históricos e ignorando otras fuentes de datos importantes como los medios sociales, las tiendas de los competidores, etc.

#### Recuerda

En resumen, estos son los principales **desafíos** que las empresas enfrentan cuando utilizan Big Data:

- Garantizar la **seguridad de los datos** – cumplimiento del Reglamento General de Protección de Datos (RGPD).
- **Competencia profesional** de los empleados – **uso competente** de la tecnología diversa y de rápido desarrollo de Big Data.
- **Los altos costes** de los proyectos de Big Data (hardware y software o costes de alquiler, personal, mantenimiento, etc.).
- **La mala calidad** de los datos, la estandarización de los datos en diferentes formatos y con diferente calidad.
- **Interpretación correcta** de los resultados.

Como habrás notado, el Big Data ofrece enormes posibilidades y oportunidades que las empresas ni siquiera se han acercado a explotar. Sin embargo, los grandes volúmenes de datos también se asocian con desafíos y riesgos que no deben ser subestimados y son inquietantes para muchas personas. Por lo tanto, el factor decisivo para asegurar que los grandes datos se utilicen con éxito sin causar daños a otras personas es el manejo **responsable** y **competente** de los grandes volúmenes de datos.



### 3.6 Resumen

**El Big Data hace referencia** a grandes cantidades de datos que ya no pueden procesarse con el software o el hardware convencional, y para los cuales el procesamiento y el análisis se realiza con un **propósito** específico. A diferencia del **Big Data**, el **Small Data** se refieren a datos que son accesibles para los humanos debido a su volumen y formato.

Encontramos el Big Data en situaciones cotidianas, por ejemplo, cuando navegamos en las redes sociales o hacemos una búsqueda en Google. Para definir mejor el término Big Data, el analista Doug Lane diseñó el **modelo 3-V**, que afirma que el Big Data se caracteriza por el **volumen, la velocidad y la variedad**.

Se puede utilizar Big Data, entre otras cosas, para mejorar la **orientación estratégica** de las empresas, para los **sistemas de aprendizaje profundo**, para **luchar contra la delincuencia y el terrorismo**, para la **investigación científica de fenómenos naturales** (por ejemplo, terremotos y cambio climático), pero también para **evaluaciones ilegales** que pueden dar lugar a chantaje o fraude. El factor decisivo no es tanto el gran volumen de datos en sí, sino lo que sucede con ellos.

**Las empresas** pueden utilizar Big Data para **aumentar su éxito comercial**. Entre otras cosas, Big Data Analytics permite a las compañías **tomar mejores decisiones de negocios** y **evaluar los riesgos** con mayor precisión. Además, la **eficiencia** de los procesos empresariales puede **aumentar** cuando los datos son analizados, evaluados y vinculados entre sí. Big Data ayuda a las empresas de investigación y desarrollo a hacer **predicciones** sobre las tendencias, las características de los productos, etc. Por último, los conocimientos adquiridos por Big Data también pueden utilizarse para ofrecer **un servicio de atención al cliente personalizado**.

Para que el análisis del Big Data tenga éxito, se requiere una estrategia apropiada, una **cultura corporativa** adecuada, **personal** con los **conocimientos** necesarios, **tecnología eficiente** y, por último, pero no menos importante, una **estrategia de protección de datos adecuada**. Pero el análisis y el procesamiento del Big Data no sólo ofrece oportunidades y posibilidades, sino que también plantea desafíos y riesgos.

Un gran desafío para las empresas es garantizar la **seguridad de los datos** y cumplir con el **Reglamento General de Protección de Datos**. Además, a menudo es difícil para las empresas encontrar y retener a los **profesionales adecuados** que puedan manejar la **compleja tecnología del Big Data**. Los grandes proyectos de datos también se asocian con **altos costes** y **la calidad de los datos** suele ser **deficiente**. Finalmente, se deben sacar las **conclusiones correctas de** los resultados del análisis de datos y se deben tomar las **decisiones adecuadas**.