Ontrack®

Recuperación luego de pérdidas de datos en máquinas virtuales

Abril de 2020

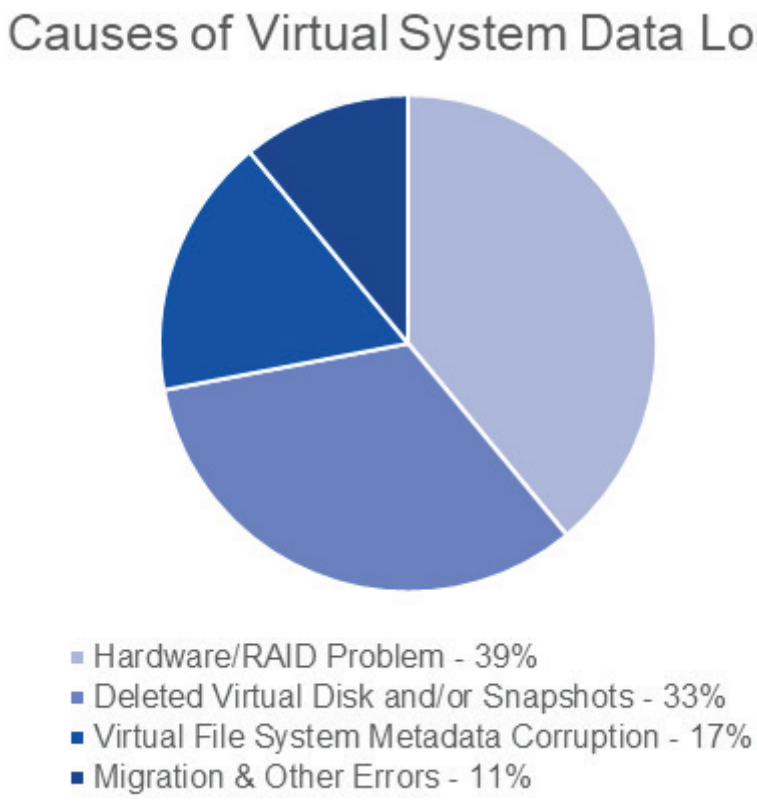
Introducción: Pérdidas de datos en máquinas virtuales

La tecnología de virtualización domina el entorno empresarial. Según Gartner, la mayor parte de las compañías indican que cuentan con un 75% o más de virtualización. La mejora de los hipervisores ha reducido la complejidad de la configuración y el mantenimiento de los servidores físicos, ha mejorado inmensamente la utilización de los servidores y ha aumentado la flexibilidad de la IT y su capacidad de respuesta ante las necesidades de las compañías. No sorprende que la mayoría de los sistemas de IT modernos estén virtualizados.

Aunque existe una potencial desventaja en la virtualización, ya sea que utilices VMware, Hyper-V, Citrix, Oracle o cualquiera de los demás hipervisores. Para poder transformar un servidor físico en muchas máquinas virtuales diferentes (VM), se agrega una capa adicional de software. La simplificación de la experiencia de usuario de admin aumenta la complejidad general del entorno de IT, ya que el hardware subyacente resulta ofuscado, lo que hace más difícil para los administradores saber cuál es el sistema físico que están ejecutando sus VM o cuál es el almacenamiento utilizado para una máquina en particular, en caso de ocurrir una pérdida de datos. Ya que son menos las personas requeridas para mantener y monitorear una cantidad mayor de máquinas virtuales (en comparación a los servidores físicos), aumentan las probabilidades de que ocurran problemas y pérdidas de datos.

Los datos recopilados en todo el mundo por Ontrack Data Recovery revelan que los incidentes de pérdidas de datos en entornos virtualizados se producen por variadas causas. Las principales razones de pérdidas de datos en máquinas virtuales son el error del usuario, el ransomware, las fallas de hardware y el deterioro de la RAID.

El propósito de este paper es el de exponer qué es lo que lleva a una pérdida de datos virtual y explicar de qué manera los proveedores de recuperación de datos globales son capaces de resolver un alto porcentaje de las situaciones de pérdida de datos en entornos virtualizados, aún en los casos más difíciles.



Fuente: Resultados del Laboratorio de Ontrack Data Recovery 2009-2019

Principales causas de pérdida de datos en máquinas virtuales

El cuadro anterior muestra un resumen de las principales causas de pérdida de datos en sistemas virtuales.

Problemas de hardware/RAID

Para evitar las pérdidas de datos, los sistemas modernos con frecuencia utilizan alguna forma de replicación de datos en múltiples unidades físicas (HDD o SSD), que luego se consolidan en una única unidad lógica. Esta protección de datos puede ocurrir a partir de una solución basada en software o en hardware. La RAID combina múltiples discos rígidos o bandas de datos para aumentar la redundancia, mejorar la confiabilidad de los datos e impulsar el desempeño de I/O (entrada/salida). La RAID fragmenta efectivamente los datos en muchos discos y los ensambla a partir de una solicitud del usuario o de una necesidad del sistema. Es necesario contar con un robusto sistema de RAID para monitorear todo y gestionar los datos.

La **principal** causa de las pérdidas de datos en máquinas virtuales son los problemas de hardware/RAID.

Los problemas de hardware que deben resolver los sistemas virtuales son básicamente los mismos que los de los sistemas físicos, como los fallos en las unidades, las fallas de los controladores, el deterioro de los componentes de los servidores y los problemas de alimentación eléctrica. Aunque el deterioro de la RAID es un problema de más difícil solución en una VM debido a la naturaleza de la virtualización.

Lamentablemente, no es infrecuente sufrir una pérdida de datos en un almacenamiento RAID. A la complejidad del software y hardware de las RAID modernas se le suma la desduplicación y la compresión.

Si agregamos la capa de virtualización adicional, la probabilidad de que ocurra una falla aumenta. Los controladores RAID son los responsables de mapear la localización de toda la información en todos los discos disponibles. Pero si se deteriora la configuración de RAID, no será posible reconstruir los archivos. Cuando eso ocurre, la interconectividad de los múltiples sistemas puede provocar pérdidas de datos y salidas de servicio significativas

Problemas de formateo/software

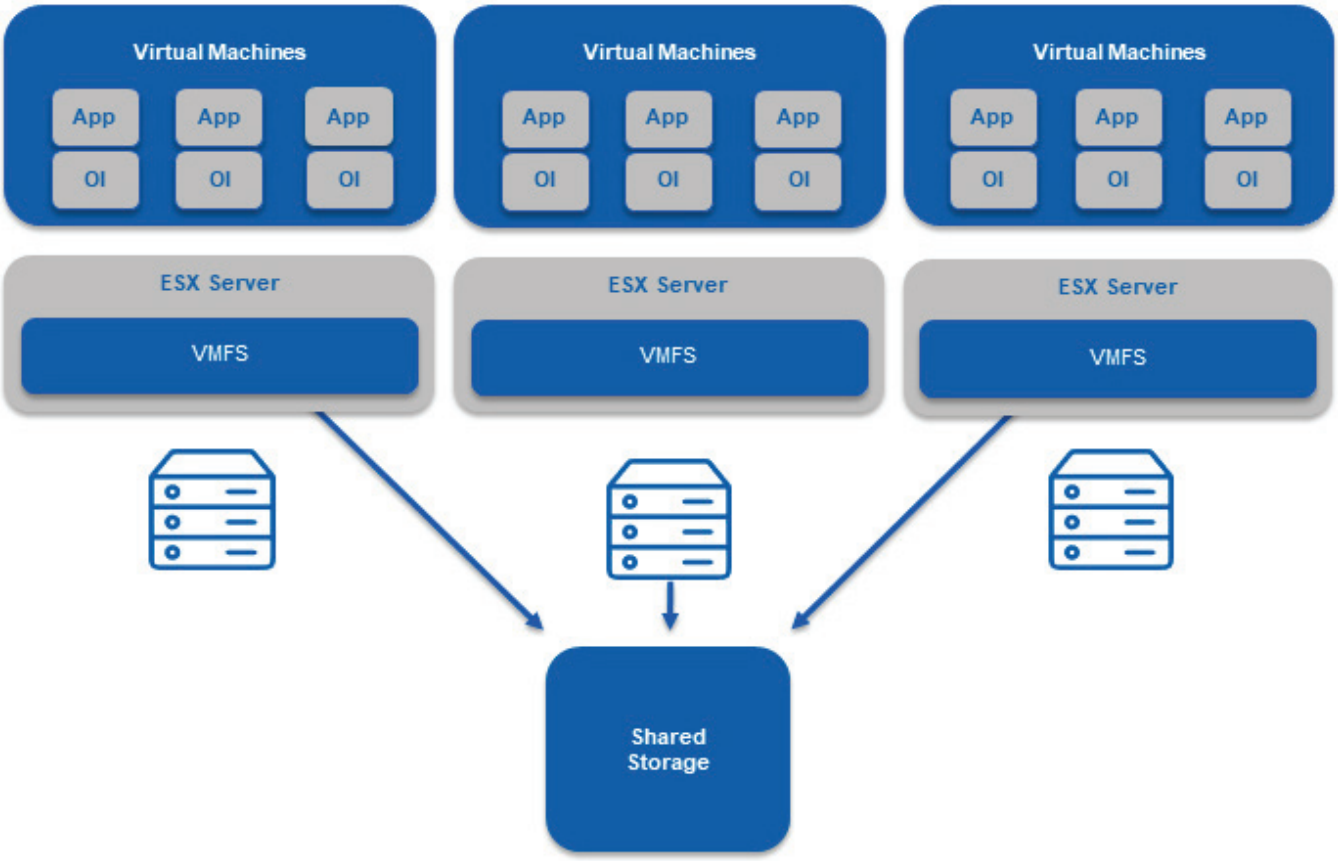
El formateo de un disco, disco virtual, matriz, LUN, disco virtual, volumen, etc (o cualquier otro tipo de dispositivo de almacenamiento) y la instalación de software nuevo son otras de las causas de pérdidas de datos en entornos virtualizados. Específicamente en las VM, por ejemplo, es posible reformatear a nivel de Guest o de Host.

El deterioro puede ocurrir también a partir de la implementación de actualizaciones y parches defectuosos si no se cuenta con una copia de respaldo offline, de una planificación deficiente de la implementación del nuevo software, de problemas de integración y del deterioro de las bases de datos. Estos problemas también pueden provocar el deterioro del archivo host y daños al sistema de archivos guest.

También tenemos que tener en cuenta las pérdidas de datos por aprovisionamiento dinámico. En lugar de asignarle todos los datos que necesitará la VM y posicionar las estructuras de sistemas de archivos en sus ubicaciones físicas específicas, el aprovisionamiento dinámico solamente ofrece la cantidad de espacio necesaria de forma inmediata y agrega bloques adicionales al disco virtual a medida que este último crece. Esto puede resultar en un disco o entorno virtual más complejo y más fragmentado. Si los indicadores de metadatos de los datos están dañados o faltan, será difícil localizar los distintos fragmentos y reconstruir el disco virtual. Por otro lado, la capa de mapeo dentro del disco virtual puede haber sufrido daños o haber sido sobrescrita, lo que tornará extremadamente difícil el reensamblado.

Deterioro de los metadatos del sistema de archivos virtuales

Otra causa más de pérdida de datos es el deterioro de los metadatos. Los metadatos cobran a un mayor importancia en los entornos virtualizados debido a la cantidad de capas y VM que existen. Un problema leve en los datos de VMFS puede tener repercusiones serias en la disponibilidad de los datos



Error del Usuario

Es posible clasificar a muchas de estas causas de pérdida de datos como errores de usuario por parte de los administradores. Los privilegios de acceso les permiten a los administradores borrar VM, aún de forma involuntaria. De todos modos, incluso en caso de gestionar correctamente los accesos, los errores siguen ocurriendo frecuentemente.

Un porcentaje sorprendentemente alto de fallas se debe a la eliminación involuntaria de discos virtuales, la sobrescritura de VM o la reasignación de su espacio. También es posible que ocurra un deterioro de la cadena de instantáneas (snapshots), por ejemplo, si una instantánea de la serie de instantáneas se deteriora, es eliminada o resulta inaccesible por alguna otra razón. Esto puede hacer defectuosas a las copias de respaldo y dificultar la recuperación de datos.

Irónicamente, la facilidad de uso de los hipervisores modernos está provocando que las organizaciones inviertan menos en capacitación. Se le asigna la responsabilidad de gestionar grandes entornos virtualizados y en constante crecimiento a personal con poca experiencia. Los medianos y pequeños proveedores de servicio gestionado (MSP) tal vez no cuenten con la cantidad suficiente de personal experimentado para monitorear frecuentemente los entornos virtuales, de manera de detectar problemas cuando se están desarrollando. En algunos casos, los administradores de IT tal vez no apliquen las medidas de seguridad adecuadas de una base de datos u omitan documentar los cambios. Si está activado el cifrado y se borra un volumen, por ejemplo, resultará difícil recuperar los datos.

Otra fuente de inconvenientes es el recambio de personal. El/la empleado/a nuevo/a tal vez no sea capaz de resolver las dificultades de las arquitecturas virtualizadas. Entonces tal vez borre algunas VM involuntariamente o realice cambios que provoquen pérdidas de datos. En otros casos, es posible que el archivo plano original esté almacenado pero que nadie pueda localizarlo cuando ocurre una pérdida de datos. Otra causa frecuente de pérdida de datos virtuales en el descuido de las copias de respaldo.

¿Y qué ocurre cuando los distintos equipos de guest, hipervisor y almacenamiento trabajan aislados unos de otros? Es posible que un equipo cree un volumen, otro adjunte el hipervisor y un administrador de guest luego configure la máquina virtual. Este tipo de estructura organizativa ofrece oportunidades para que ocurran errores y fisuras. Resulta más fácil entonces que ocurran borrados, sobrescrituras y reformateos involuntarios.

¿Qué es lo que pueden hacer las compañías cuando sufren una pérdida de datos en un entorno virtualizado? No hay un botón “Deshacer” o “Volver”. Si se borra una VM, se pierde. ¿Copias de respaldo? Con frecuencia están deterioradas o incompletas. Afortunadamente, con frecuencia es posible recuperar los datos a través de proveedores de servicio de recuperación de datos globales.

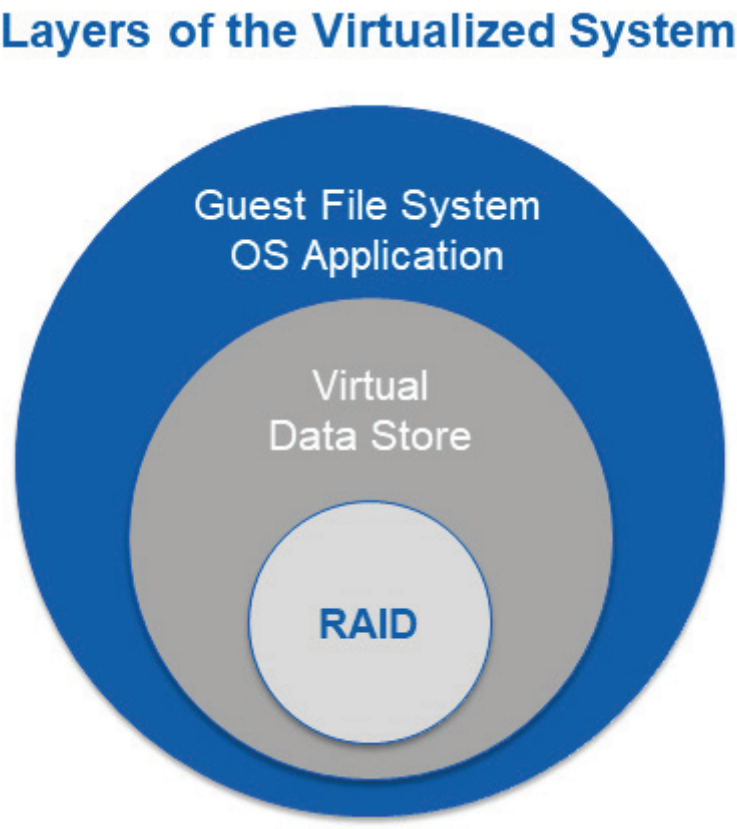
Opciones de recuperación

La buena noticia es que existen muchas maneras de recuperar una parte o, en muchos casos, la totalidad de los datos virtuales perdidos. El primer punto de acceso es a nivel del almacenamiento. En algunos casos es posible recuperar datos directamente desde las unidades físicas, creando una imagen de las unidades y leyendo los datos en crudo que estén disponibles en el disco.

La siguiente opción es intentar la recuperación de datos a partir de los volúmenes lógicos (LUN) o RAID. Si está disponible el controlador RAID, será posible utilizarlo para rastrear las distintas secciones de datos distribuidas en los discos virtuales. Al determinar cuál debe ser la configuración, los ingenieros pueden reconstruir virtualmente la matriz y acceder al almacenamiento. Si el controlador RAID está deteriorado, tal vez sea necesario emular el controlador RAID y reconstruir las partes faltantes.

El siguiente nivel, ya que cada uno de los niveles representa un grado mayor de dificultad para la recuperación, es el nivel de sistema de archivos host. En VMware esto sería VMFS, y en Hyper-V, NTFS o ReFS. En muchos casos, los datos no están disponibles directamente a nivel del almacenamiento. Pero si se utilizan las herramientas correctas, los expertos en recuperación pueden rastrear los datos a partir de los bloques de datos de almacenamiento básicos, mapearlos hasta el nivel de host y compilarlos de nuevo.

Si ese proceso no ofrece una recuperación adecuada, es posible utilizar herramientas adicionales para ingresar más profundamente dentro nivel del sistema de archivos de guest. Algunas veces, los especialistas en recuperación de datos pueden encontrar datos al investigar el sistema de archivos virtuales, los cuales se perderían de otro modo. Finalmente, es posible ingresar a nivel de archivos de guest y acceder a los datos residentes en archivos de aplicación como SQL, Exchange, SharePoint, Oracle, archivos de Office, archivos ZIP y otros.



Para ello, es necesario contar con una comprensión de cada nivel y el conocimiento de qué es lo que puede estar disponible y en dónde. Los especialistas con buena experiencia en arquitectura de almacenamiento pueden rastrear datos aparentemente perdidos, buscando porciones de esos datos en un nivel y otras porciones en otro nivel.

Esto se ve claramente en un ejemplo de RAID. Es un hecho que todas las unidades en algún momento dejarán de funcionar. Si se está utilizando RAID 1 o posterior, es posible instalar una nueva unidad y reconstruir el mapa de almacenamiento de datos sin que ocurra ninguna pérdida de datos. Pero, ¿qué es lo que ocurre si la falla de la unidad supera la capacidad de redundancia de RAID? En este caso, para recuperar los datos normalmente es necesario saltear cualquier falla física que pueda haber ocurrido, reconstruir el sistema de archivos de la RAID y evaluar las distintas capas y complejidades de cualquier arquitectura virtualizada presente. Con frecuencia esto hace que la recuperación sea extremadamente dificultosa y prolongada. Sin embargo, si contactamos al proveedor correcto, será posible alcanzar el éxito de los esfuerzos de recuperación en la mayoría de los casos. Asegúrate de que el proveedor cuente con las herramientas y la experiencia correctas, como así también de que trabaje en colaboración directa con las compañías de almacenamiento.

Recuperación de datos virtuales en Ontrack

A partir de sus más de 30 años de experiencia global en gestión de datos, recuperación de datos, borrado seguro de datos, e-discovery y computación forense, Ontrack ha recuperado datos virtualizados de miles de compañías.

Los ingenieros crean imágenes y leen datos en crudo de los discos, definen cuál debe ser la configuración y luego reconstruyen virtualmente la matriz para acceder al almacenamiento. Para ello, Ontrack ha desarrollado herramientas como las que realizan la emulación del controlador de RAID y reconstruyen las partes faltantes. La compañía también ha desarrollado una gran gama de herramientas adicionales para impedir la subsiguiente escritura de datos en los volúmenes, resolver las complejidades de los sistemas de archivos actualizados y otras.

El equipo de desarrollo de Ontrack actualiza continuamente sus herramientas para que estas sean compatibles con las más recientes plataformas de virtualización y entornos de almacenamiento. Gracias a sus conocimientos de los distintos medios de almacenamiento, sistemas operativos y arquitecturas de almacenamiento subyacente, Ontrack ofrece servicios exhaustivos de recuperación de datos, como así también servicios de seguimiento relacionados a las copias de respaldo y gestión de datos inteligente.

Examinemos algunos ejemplos:

Borrado accidental de un sistema de NetApp

Uno de los clientes de un Proveedor de Servicio Gestionado (MSP) en Corea utilizaba un sistema NetApp FAS8060 con 161 x HDD SAS de 900 GB. Estaban distribuidos en dos acumulados separados (68 unidades + 93 unidades). El cliente tenía tres LUN de Canal de Fibra de 468 GB conectadas de cada acumulado a un servidor de producción Sybase. En total, se combinaron seis LUN en un solo conjunto de discos con tres volúmenes lógicos salientes del conjunto.

Un ingeniero de MSP intentó aplicar cambios de configuración al archivador de NetApp. Sin embargo, ejecutó involuntariamente un comando de borrado en algunas de las LUN, borrando efectivamente 45 GB de datos del servidor Sybase. A consecuencia de ello, el MSP podría haber perdido el contrato con el cliente y posiblemente sufrido costos legales.

Se convocó a Ontrack mediante una consulta telefónica. Dentro de las 12 horas siguientes al evento de pérdida de datos, se le indicó al MSP llevar offline a los acumulados para evitar cualquier nueva sobrescritura. Se le indicó al cliente presentar los 161 HDD de ambos acumulados a una única máquina Windows. Se conectó entonces este sistema al servidor de Recuperación de Datos Remotos de Ontrack. Ya que ambos acumulados tenían el mismo nombre, no resultó fácil reconstruirlos. Como resultado de esto, fue necesario clasificar las unidades en grupos de acumulados y reconstruirlas manualmente hasta el punto en el tiempo lo más cercano posible al momento del implementación del comando de borrado incorrecto.

En ese punto, surgió otro problema. Los volúmenes lógicos estaban sido utilizados como almacenamiento en crudo por el servidor Sybase. Esta configuración imposibilitó la extracción de los datos internos de forma directa. La solución fue la de extraer las seis LUN como archivos planos y coordinar con el soporte de NetApp para presentar estas LUN al servidor Sybase. Los volúmenes lógicos recuperados superaron verificaciones de integridad y fue posible volverlos operativos sin pérdidas de datos.

Pérdidas de Datos Debidas al Reformateo de VMware

El equipo de IT de una compañía de producción de alimentos con sede en Singapur eliminó involuntariamente una LUN de almacén de datos VMFS del host ESXi de Vmware y la adjuntó a un servidor Windows. Eso provocó que la LUN fuera reformateada al sistema de archivos NTFS. Esta acción deterioró metadatos de front-end de VMFS, lo que provocó la pérdida de todas las máquinas virtuales del almacén de datos. La compañía llamó a Ontrack. Los ingenieros pudieron reconstruir las estructuras de VMFS para volver a acceder a las VM almacenadas en el sistema. Se recuperaron distintas VM intactas, mientras que fue necesario en algunas otras reparar los sistemas de archivo de guest internos y extraer los datos resultantes.

Eliminación de VM

Un proveedor de servicios del cuidado de la salud en Australia borró accidentalmente siete VM con aprovisionamiento dinámico de un almacén de datos de producción. Debido a naturaleza sensible de los datos perdidos, la compañía llamó inmediatamente a Ontrack y solicitó que nuestros ingenieros se trasladaran directamente a sus instalaciones. Una vez que llegaron al centro de datos, pudieron recuperar todas las VM, aunque evidenciaban algunos daños. En el momento, se realizaron reparaciones adicionales a los sistemas de archivos de guest de cada VM, utilizando las herramientas bajo derechos de propiedad intelectual de Ontrack. Este proceso permitió extraer a un almacenamiento externo datos internos más críticos. A pesar de que se perdieron algunos datos, fue posible recuperar la mayor parte de los datos contenidos en las VM.

Conclusión

La virtualización puede ahorrar tiempo y eliminar complejidades desde el punto de vista del usuario. Aunque esto ofrece un conjunto único de desventajas, una de las cuales es la creciente incidencia del deterioro y de las pérdidas de datos. Sin importar si es deterioro de volúmenes, borrado de volúmenes, ransomware, borrado o deterioro de copias de respaldo virtuales, fallas del hardware o de RAID o archivos borrados o deteriorados dentro de los sistemas de almacenamiento virtualizados, la pérdida de datos es una realidad que enfrenta todo aquel que gestiona sistemas virtuales. Las copias de respaldo son necesarias para salvaguardar los datos empresariales, pero están lejos de ser infalibles.

En su condición de proveedor líder de recuperación de datos en el mundo, Ontrack está listo para ofrecer exhaustivos servicios de recuperación de datos para almacenamiento y servidores virtuales. Los Servicios de Recuperación de Datos de Ontrack pueden:

Recover datos de virtualmente cualquier tipo de dispositivo de almacenamiento de datos, desde discos rígidos, unidades SSD y flash hasta servidores, sistemas virtuales, de cinta, SAN y NAS.

Minimizar los tiempos fuera de servicio al ofrecer una implementación rápida, opciones de servicio de emergencia y un servicio de recuperación de datos remotos con calidad de laboratorio único en el mundo y en la industria.

Informar acerca de todos los archivos recuperables y la condición de cada uno de ellos como parte de la evaluación, antes de que el cliente pague por el servicio de recuperación.

Tus expertos en datos

Ontrack es la empresa de destrucción de datos y de recuperación de datos de KLDiscovery, un proveedor global de soluciones de recuperación de datos, gobernanza de información y E-discovery.

Ontrack ofrece una premiada suite de software y servicios impulsados por tecnología para ayudar a las empresas, a los proveedores de servicio y a las entidades gubernamentales, como así también a los usuarios individuales, a gestionar, recuperar, buscar, destruir y migrar datos de forma eficiente y económica.

Experiencia que cuenta

A partir de sus más de 30 años de experiencia, Ontrack ha desarrollado sólidas raíces y tiene un futuro innovador e interesante. Nuestra capacidad de creación de excelentes tecnologías y de entrega de servicios de la más alta calidad sienta las bases para la producción de excepcionales soluciones de recuperación de datos para nuestros clientes.

Soluciones de gestión de datos

Contamos con instalaciones de sala limpia en todo el mundo y experiencia de ingeniería en todas las regiones globales principales. Puedes contar con nosotros para resolver todas tus necesidades de recuperación, restauración y destrucción de datos.

Pioneros e innovadores

Aprovechando la combinación de nuestra extensa experiencia en la industria y nuestro legado tecnológico, Ontrack te ofrece las mejores herramientas y capacidades bajo derechos de propiedad intelectual de todo el espectro de la recuperación de datos.

Servicio 24/7/365

Las pérdidas de datos pueden ocurrir en cualquier momento, por lo que nos enorgullecemos de ofrecer nuestros servicios de recuperación de datos en cualquier momento del día. Sin importar el horario o la situación, puedes confiar en nosotros para ayudarte con tus datos.

Para obtener más información acerca de cómo puede Ontrack ayudar a tu organización a resolver sus necesidades de recuperación de datos, por favor ponte en contacto con:

BRO DR ES_DR Virtualization_Apr 2020

Ontrack®

C. Anabel Segura 7
28108 Alcobendas, Madrid

24/7 900.112.012
www.ontrack.com/es