Guía de prácticas recomendadas

Continuidad empresarial para aplicaciones vitales

Prácticas recomendadas para sobrevivir a los desastres en el centro de datos

Hewlett Packard Enterprise

Imagine que va conduciendo de camino al mercado. Por el rabillo del ojo, algo lo distrae durante un instante. En ese momento, el coche que va delante frena y tiene un accidente que provoca daños en ambos coches. Esto ya es bastante traumático de por sí pero, ¿y si no tiene contratado un seguro o no dispone del dinero suficiente para la reparación? ¿Qué ocurrirá a continuación?

A nivel personal, rara vez nos cuestionamos la necesidad de tener un seguro que nos proteja ante los accidentes, los robos u otras pérdidas. Las empresas aplican un enfoque similar y en ocasiones invierten millones de dólares al año en evaluar y mitigar los riesgos. En muchos sectores, las organizaciones de TI no escatiman en gastos para proteger los datos contra el acceso no autorizado o la exposición. Cuando se trata de habilitar la disponibilidad continua de las aplicaciones que procesan datos o de evitar las pérdidas provocadas por fallos del sistema o el centro de datos, algunas organizaciones pueden estar trabajando con una idea anticuada de qué es lo que tiene que estar "siempre activo". Como consecuencia, estas organizaciones a menudo son reacias a dedicar los recursos necesarios para garantizar un nivel aceptable de continuidad empresarial.

El riesgo es real

IDC estima que el coste medio del tiempo de inactividad es aproximadamente de 1,7 millones de dólares por hora en todos los sectores, y en algunos casos, la cifra puede llegar hasta 10 millones de dólares por hora. Una incidencia de tiempo de inactividad suele durar 90 minutos de media, aunque algunas duran más de 24 horas.[3]

Pregúntese: ¿cuál es el coste por pérdida de ingresos si un sistema orientado al cliente no está disponible? ¿Cuál es el impacto, para la empresa o sus clientes, de perder transacciones de clientes en curso? Es más, ¿cuál es el coste para su empresa en términos de productividad perdida si falla el sistema de correo electrónico o de comunicaciones unificadas y colaboración? ¿Qué ocurriría en caso de que se produjera un incendio en el centro de datos, un corte de suministro eléctrico o si un terremoto produjera daños y dejara el centro de datos desconectado? ¿Qué pasa después?

Estas cosas suceden más a menudo de lo que se imagina. El 95 % de las empresas han sufrido al menos una interrupción del centro de datos no planificada en los últimos 24 meses; y no sólo de un sistema, sino de todo el centro de datos. Una empresa de servicios financieros media sufrió 1,8 interrupciones completas de sus centros de datos en los últimos 24 meses. En el sector sanitario, la media es de tres interrupciones en los últimos 24 meses.[1]

IDC estima que el coste medio del tiempo de inactividad es aproximadamente de 1,7 millones de dólares por hora en todos los sectores, y en algunos casos, la cifra puede llegar hasta 10 millones de dólares por hora.[2] La incidencia media dura unos 90 minutos, y algunas pueden alargarse hasta 24 horas. Junto con la pérdida de ingresos, el coste real del tiempo de inactividad puede incluir daños a la reputación, pérdida de confianza y lealtad de los clientes, pérdida de competitividad e incluso exposición al incumplimiento reglamentario. En el mundo basado en las redes sociales de hoy en día, la noticia de una interrupción no tarda en volverse viral, y pueden pasar años hasta que se recupere de los daños.

Objetivos de recuperación

Si el servicio de una aplicación falla, ¿cuál es el plazo aceptable de recuperación? Este objetivo de tiempo de recuperación (RTO) varía de una aplicación a otra: puede ser de cinco segundos, cinco minutos o cinco días. El coste del tiempo de inactividad suele ser el factor que condiciona la configuración del objetivo de tiempo de recuperación (RTO) de una aplicación. Algunas aplicaciones tienen un RTO de cero, lo que significa que el cliente o el usuario final no debe conocer la existencia de un fallo.

Si falla el servicio de una aplicación, ¿cuál es el máximo de datos que se pueden perder? Este objetivo de punto de recuperación (RPO) puede fijarse en función de la naturaleza de los datos y de su importancia para la empresa, el valor de los datos para la empresa (es decir, el coste en términos de pérdida de ingresos, incluidas las posibles responsabilidades legales y por incumplimiento reglamentario) o de una combinación de ambas opciones.

[1] "Fingers Crossed? Or What is Your Business Continuity Plan for the Inevitable" (¿Toca madera? O ¿cuál es su plan de continuidad empresarial ante lo inevitable?), Gravic, Inc., 2015 (fuente original: Ponemon Institute)

[2, 3] "**High-Value Business Applications on x86: The Need for True Fault-Tolerant Systems**" (Aplicaciones empresariales de alto valor sobre x86: la necesidad de contar con auténticos sistemas tolerantes a fallos), Peter Rutten, IDC, mayo de 2015

RTO: objetivo de tiempo de recuperación; el tiempo máximo aceptable para la recuperación de una interrupción

RPO: objetivo de punto de recuperación; la cantidad máxima de pérdida de datos aceptable provocada por una interrupción del sistema

A menudo, los objetivos de punto de recuperación (RPO) se basan en el valor medio de una transacción perdida. Si bien puede sonar razonable a simple vista, si se profundiza un poco, el enfoque no siempre tiene sentido. Por ejemplo, para una institución financiera, la transferencia de fondos electrónica (EFT) media puede ser de 1.000 dólares, pero la más grande puede ser de varios millones.

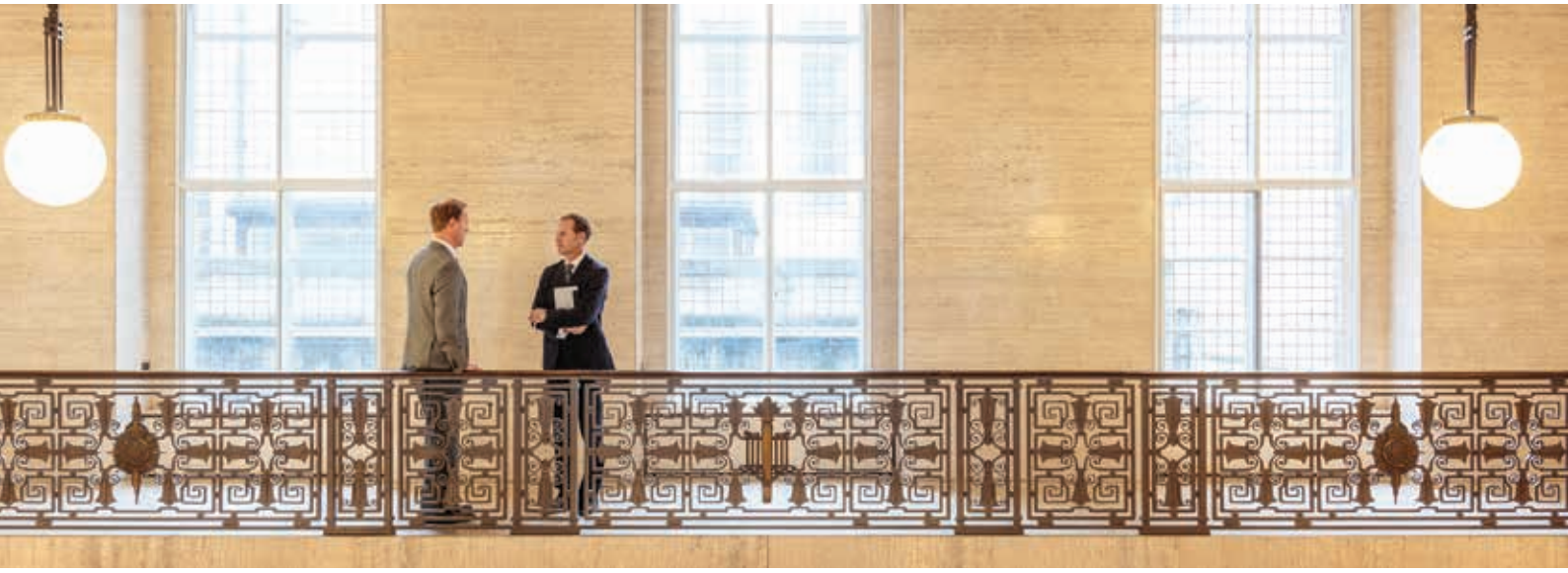
Como no se puede predecir qué transacciones pueden fallar, el coste potencial real de una interrupción es el de los datos más valiosos que podrían perderse (por ejemplo, la mayor transacción de transferencia de fondos electrónica, el mayor pedido de ventas, la transacción comercial de mayor volumen, la mayor responsabilidad legal por pérdida de datos o la mayor cuenta). Ése es el valor que debe determinar los objetivos de punto de recuperación (RPO).

Continuidad empresarial

Una conversación sobre transacciones, ingresos o productividad perdidos suele ser el desencadenante. Este tipo de problemas empresariales lleva a las organizaciones a empezar a pensar en la continuidad empresarial como un imperativo estratégico. No es una cuestión de si los sistemas fallarán o si se producirá un evento catastrófico, sino de cuándo. Por lo que la pregunta es: ¿Qué pasa después? La continuidad empresarial es la práctica de asegurar que su empresa sea capaz de seguir operando, pase lo que pase.

Desde una perspectiva práctica, la continuidad empresarial requiere el análisis (y la repetición de dicho examen, cada uno o dos años, según las prácticas recomendadas) de los objetivos de tiempo y punto de recuperación (RTO, RPO) de cada sistema necesario para almacenar o servir a las aplicaciones vitales y sus datos. En algunos casos, su determinación puede resultar relativamente simple. Por ejemplo, un sistema de historiales médicos debe estar disponible siempre y no puede haber pérdida de datos. La salud de los pacientes depende de ello. Si se produce un fallo, el sistema debe ser inmediatamente recuperable hasta el punto de que los usuarios nunca sean conscientes de que ha ocurrido un problema.

En otros casos, los RTO y RPO tolerables pueden ser algo más complicados de determinar. Por ejemplo, ¿qué valor tiene su tienda en línea, teléfono VoIP o sistema de gestión de relaciones con los clientes (CRM)? ¿Qué pasa si uno de ellos sufre una interrupción? ¿Es de misión crítica o crítico para la empresa? ¿Podría sobrevivir gran parte de su empresa mientras estos sistemas permanecen inactivos? ¿Cuánto tiempo puede permitirse dedicarse a recuperar datos perdidos? ¿Cuál es el impacto si los clientes no pueden llegar hasta usted? Este tipo de examen a menudo revela que muchos sistemas son más vitales para su empresa de lo que pensaba. Estos sistemas requieren un soporte para la continuidad empresarial más sólido, con objetivos de tiempo y punto de recuperación (RTO, RPO) más estrictos que los que actualmente estén en vigor. Si se profundiza en el análisis, estas aplicaciones o servicios pueden incluso resultar de misión crítica, lo que significa que cualquier interrupción que sufran ejercerá un impacto significativo en la empresa. Tenga en cuenta que lo que no es de misión crítica hoy puede serlo mañana.



Defina qué es "crítico"

Hemos establecido que las aplicaciones y los datos "críticos para la empresa" son necesarios para su funcionamiento y que las aplicaciones y los datos "de misión crítica" son tan valiosos que cualquier interrupción sería catastrófica. Ahora, pregúntese lo siguiente: ¿Cuál es el impacto si un sistema de misión crítica no está disponible o si se pierden datos de transacciones? ¿Cómo afecta a sus clientes? ¿Y a los ingresos de su empresa? ¿La pérdida de datos genera potenciales problemas legales o de conformidad?

La compatibilidad con la continuidad empresarial debe adoptar un proceso que clasifique la importancia de las aplicaciones y los datos desde la misión crítica hasta lo más básico. La evaluación de sus aplicaciones y sistemas en función de este proceso debe basarse en las necesidades de sus clientes y sus objetivos de ingresos. Una aplicación con una tolerancia muy limitada por su disponibilidad (objetivo de tiempo de recuperación cero o cercano a cero) o por la cantidad de datos que puede perder (objetivo de punto de recuperación cero o cercano a cero) debe considerarse de misión crítica.

Deben considerarse de misión crítica las aplicaciones que no pueden sufrir interrupciones o que deben volver a estar en funcionamiento tan rápido que nadie note el problema.

Algunos ejemplos de aplicaciones y servicios vitales:

- **Servicios financieros:** procesamiento de pagos, prevención del fraude, negociación de alta frecuencia
- **Telecomunicaciones:** gestión de redes móviles; máquina a máquina, servicio al cliente en tiempo real
- **Comercio minorista:** punto de venta, comercio electrónico, transacciones en línea y procesamiento de pedidos
- **Fabricación:** procesos de control de producción continuos y distribución multicanal
- **Sanidad:** datos de pacientes y laboratorio en tiempo real, recuperación de información de proveedores
- **Transporte:** reservas, billetes, programación

En un mundo siempre activo, el tiempo de inactividad de las cargas de trabajo complejas, interconectadas y de cara al cliente no suele ser una opción.

Enfoques de continuidad empresarial

La auténtica continuidad empresarial requiere un nivel de dispersión geográfica (o distancia) para sobrevivir tanto a eventos localizados (un incendio en el centro de datos) como a fallos regionales (un colapso de la red eléctrica regional). Existen tres enfoques básicos para crear una infraestructura de continuidad empresarial distribuida geográficamente, cada una con perfiles de objetivos de tiempo y puntos de recuperación (RTO, RPO) distintos:

- **Asíncrono activo/pasivo:** éste es un escenario clásico de recuperación en caso de desastre donde todas las transacciones y los datos se replican asíncronamente a un nodo de copia de seguridad pasivo. En caso de fallo, las aplicaciones deben iniciarse en el nodo de seguridad, lo que puede generar retardos y producir un objetivo de tiempo de recuperación (RTO) más largo. Los procedimientos de conmutación por recuperación para reiniciar estas aplicaciones son a menudo difíciles de probar o ejecutar, y el riesgo de fallo es alto.
- **Asíncrono activo/casi activo:** conocido también como Sizzling-Hot-Takeover (traspaso al rojo vivo) o Sizzling-Hot-Standby (en espera al rojo vivo), este enfoque es parecido a la arquitectura activa/pasiva con replicación, salvo que el nodo de copia de seguridad está preparado para empezar a procesar transacciones inmediatamente con la copia de la base de datos de aplicaciones ya abierta para permitir el acceso de lectura o escritura. Es básicamente una arquitectura activa/activa, a excepción de que todas las transacciones de los usuarios se dirigen al nodo principal. Ello mejora enormemente las probabilidades de éxito cuando se produce una conmutación por recuperación. Además, ofrece un objetivo de tiempo de recuperación (RTO) mucho mejor y repetible que la recuperación en caso de desastre tradicional.
- **Asíncrono activo/activo:** en una arquitectura tolerante a desastres, el procesamiento de producción se divide entre varios nodos. Cada uno de estos nodos cuenta con una copia de la base de datos, que se sincroniza utilizando replicación de datos bidireccional. Si falla un nodo, su tráfico puede enrutarse automáticamente a otros nodos activos. Los usuarios conectados a los nodos supervivientes no son conscientes de que se ha producido una interrupción. La conmutación por recuperación se convierte en un proceso sencillo que puede probarse y practicarse con facilidad, porque se sabe que todos los nodos están operando permanentemente.

Para aplicaciones de misión crítica, una arquitectura de varios nodos tolerante a desastres es la mejor alternativa y ofrece los mejores objetivos de tiempo y punto de recuperación (RTO, RPO). Naturalmente, existen varias tecnologías que ayudan a crear soluciones de continuidad empresarial, desde la replicación de datos transaccionales basada en software hasta la agrupación en clústeres basada en hardware y las tecnologías RAID. Cada una se caracteriza por tener límites en cuanto a los objetivos de tiempo y punto de recuperación (RTO, RPO) que puede proporcionar.

Otra consideración es la atomicidad, uno de los cuatro conceptos de atomicidad, homogeneidad, aislamiento y durabilidad (ACID, *Atomicity, Consistency, Isolation, Durability*) aplicables al diseño de bases de datos y la arquitectura de las aplicaciones. La atomicidad define una regla de "todo o nada" para el procesamiento de transacciones: una transacción empieza en un momento determinado, y si algo sale mal antes de su finalización, se deshace hasta llegar al principio, como si nunca se hubiera realizado. Los enfoques de continuidad empresarial deben diseñarse para observar este principio, especialmente para aplicaciones de misión crítica.

Coste total de propiedad

Si sufre un fallo, evento o desastre que provoca la interrupción de una aplicación durante más tiempo del aceptable para sus clientes, significa que debería haber dispuesto de una arquitectura tolerante a desastres para evitar la pérdida de oportunidades o el daño a su reputación. En la era de las redes sociales, los clientes harán pública su frustración en segundos y el problema se puede volver viral en cuestión de minutos. Para empresas con estrictas restricciones de conformidad con respecto al tiempo de inactividad, una arquitectura tolerante a desastres mitigará las sanciones y denuncias oficiales. Por ejemplo, si un sistema sufre una interrupción y tarda tres horas en recuperarse completamente, y si aplicamos los costes medios del tiempo de inactividad del estudio de IDC citado anteriormente, el coste sería superior a 5 millones de dólares. El mismo fallo puede durar tan solo unos segundos o incluso ser invisible para una aplicación que resida en un entorno de continuidad empresarial correctamente diseñado, y ello supone unas pérdidas mucho menores.

Las arquitecturas de recuperación en caso de desastre activas/pasivas presentan un alto coste total de propiedad como consecuencia del tiempo de inactividad.

En general:

- Cuanta mayor sea la disponibilidad, mayores son la complejidad y los costes de implementación
- Cuanta mayor sea la disponibilidad, menores son los costes de una interrupción

El resultado neto es que a medida que aumentan los costes de implementación, se reduce el coste total de propiedad global, pero a una velocidad mucho mayor.[4] En otras palabras, puede comprar mucha tolerancia a fallos con lo que cuesta una sola interrupción.

[4] "Fingers Crossed? Or What is Your Business Continuity Plan for the Inevitable" (¿Toca madera? O ¿cuál es su plan de continuidad empresarial ante lo inevitable?), Gravic, Inc., 2015 (fuente original: Ponemon Institute)

Soluciones de continuidad empresarial

Servicios de HPE
Para ayudar a los clientes con el centro de datos crítico para la empresa de hoy en día, Hewlett Packard Enterprise ofrece profesionales capacitados y con una amplia trayectoria que presten una gama completa de servicios de asesoramiento, diseño, implementación y gestión para arquitecturas tolerantes a desastres.

HPE Integrity NonStop X
Los sistemas HPE Integrity NonStop X se han diseñado de forma exclusiva para sectores que nunca se detienen. Ofrecen los mayores niveles de disponibilidad, seguridad para el sistema completo, escalabilidad masiva y el menor coste total de propiedad de su categoría. Según la definición del máximo nivel de disponibilidad (AL4) de IDC, AL4 es la combinación de varios componentes de hardware y software que permite una conmutación por recuperación prácticamente instantánea en recursos alternativos con el fin de que el procesamiento empresarial continúe sin interrupciones, como antes del fallo.[5] HPE Integrity NonStop X, combinado con el software HPE NonStop Shadowbase, ofrece tolerancia a fallos de nivel AL4 en ubicaciones completas, con un tiempo de inactividad planificado o no planificado cero, lo que elimina prácticamente cualquier posibilidad de que se produzca una interrupción de las aplicaciones. Con HPE NonStop X, podrá volver a diseñar la computación de misión crítica ajustada para obtener la máxima disponibilidad, escalabilidad e integridad de los datos.

Almacenamiento HPE XP
HPE XP7 se ha diseñado para el almacenamiento Flash híbrido y resulta idóneo para aplicaciones de misión crítica que requieren disponibilidad, escalabilidad y rendimiento continuo de los datos. La tecnología de virtualización basada en matrices permite implementar la virtualización, la replicación y la gestión en varios sitios y matrices con el fin de aumentar la disponibilidad, evitar desastres y mejorar el uso de los recursos al facilitar la eliminación de los silos de almacenamiento. HPE XP7 es la matriz SAN de mayor disponibilidad que ofrece HPE, con una serie de soluciones de software que ayudan a obtener los máximos objetivos de recuperación, y ofrece además replicación remota y funcionalidades de recuperación en caso de desastre.

[5] "Worldwide and U.S. High-Availability Server 2014–2018 Forecast and Analysis" (Predicción y análisis de servidores de alta disponibilidad a nivel internacional y en EE. UU, 2014-2018), IDC, Doc #250565

Conclusión

La cuestión no es si un evento catastrófico afectará a un sistema de misión crítica, sino cuándo. Y cuando ocurra, ¿qué sucederá a continuación? Una solución de continuidad empresarial diseñada para una tolerancia a los desastres continua puede ayudar a minimizar los daños al proporcionar el mejor tiempo de recuperación disponible, así como funcionalidades de punto de recuperación con un coste total de propiedad razonable para su empresa.[6]

Obtenga más información en

hpe.com/info/nonstop

hpe.com/storage/xp

hpe.com/info/dcm

[5] "**High-Value Business Applications on x86: The Need for True Fault-Tolerant Systems**" (Aplicaciones empresariales de alto valor sobre x86: la necesidad de contar con auténticos sistemas tolerantes a fallos), Peter Rutten, IDC, mayo de 2015

Regístrese y reciba las actualizaciones

Hewlett Packard Enterprise

© Copyright 2016 Hewlett Packard Enterprise Development LP. La información contenida en este documento está sujeta a cambios sin previo aviso. Las únicas garantías de los productos y servicios de Hewlett Packard Enterprise figuran en las declaraciones expresas de garantía incluidas en los mismos. Ninguna información contenida en este documento debe interpretarse como una garantía adicional. Hewlett Packard Enterprise no se responsabilizará de los errores u omisiones técnicos o editoriales que pudiera contener el presente documento.

4AA6-5326ESE, mayo de 2016