

# Estrategias de Alta Disponibilidad

La tarea fundamental para mantener tus sistemas en línea

Los términos disponibilidad y alta disponibilidad pueden tener significados muy diferentes en función del contexto en el que se utilicen y del público al que se dirijan. Se pueden usar para describir una serie de objetivos empresariales y requisitos técnicos: desde objetos de disponibilidad sólo para hardware, hasta objetivos de importancia vital del servicio de mensajería en su conjunto.

Por norma general, es relativamente fácil que las organizaciones tengan expectativas poco adecuadas en cuanto a los objetivos de disponibilidad y, además, que exijan niveles de disponibilidad mayores a los que están dispuestos a pagar antes de que se conozcan los costos implicados.

Las implicaciones de costo de las soluciones de mayor disponibilidad son las siguientes (aunque sin limitarse a ellas):

- Hardware
- Infraestructura de red
- Capacidad de Servicios
- Software
- Formación
- Costos de Explotación

Por capacidad de servicio se entienden los acuerdos firmados con otros proveedores de servicios o los acuerdos a nivel operativo realizados con divisiones de tecnología de la información de su organización, para proporcionar o mantener componentes o servicios de tecnología de información.

## Disponibilidad

La disponibilidad se refiere a un nivel de servicio proporcionado por aplicaciones, servicios o sistemas. Los sistemas de alta disponibilidad tienen un tiempo de inactividad mínimo, ya sea previsto o imprevisto. La disponibilidad se suele expresar como el porcentaje del tiempo que un servicio o un sistema está disponible; por ejemplo, el 99,9 por ciento para un servicio que no está disponible durante 8,75 horas al año.

Para mejorar la disponibilidad, deben implementarse mecanismos de tolerancia a errores que enmascaren o minimicen el impacto de las averías de los componentes y las dependencias del servicio. La tolerancia a errores se logra implementando un sistema de redundancia en componentes de puntos únicos de error.

A la hora de diseñar la disponibilidad de Microsoft Exchange, tenga en cuenta todos los componentes que forman parte de la infraestructura de mensajería. Algunos componentes podrían ser también otros servicios que tienen subcomponentes. La disponibilidad del servicio de mensajería se determina por medio de la disponibilidad de cada componente que forma parte de la infraestructura.

## Definición de requisitos de disponibilidad

La disponibilidad de un servicio es un tema complicado que abarca muchas disciplinas. Se pueden usar muchos métodos diferentes para ofrecer los niveles de disponibilidad requeridos, cada uno con sus propios costos.

Sin embargo, con frecuencia, el cliente puede expresar los requisitos de disponibilidad en términos relativamente simples y sin comprender del todo las implicaciones. Esta situación puede derivar en malentendidos entre el cliente y la organización de tecnología de la información, en niveles de inversión inadecuados y, por último, en la insatisfacción del cliente debida a la definición de expectativas inadecuadas.

Un requisito expresado para una disponibilidad del 99,5 por ciento puede ser diferente a otro requisito para esa misma disponibilidad. Un requisito se puede referir a la disponibilidad de la plataforma de hardware exclusivamente, mientras que el otro, a la disponibilidad del servicio completo de un extremo a otro. Incluso la definición de la disponibilidad del servicio completo de un extremo a otro puede variar enormemente. Es importante saber exactamente cómo se va a medir cualquier requisito de disponibilidad. Por ejemplo:

- Si todo el hardware y el software del servidor principal funciona correctamente y las conexiones del usuario están listas para ser aceptadas por la aplicación, ¿se considera que la solución está disponible al 100 por cien?
- Si hay 100 usuarios, pero el 25 por ciento de ellos no se puede conectar debido a un error de la red local, ¿la solución se sigue considerando disponible al 100 por cien?
- Si sólo un usuario de 100 se puede conectar y procesar el trabajo, ¿sólo hay un 1 por ciento de disponibilidad?
- Si los 100 usuarios se pueden conectar, pero el servicio se degrada hasta tal punto que sólo hay disponibles dos de tres transacciones del cliente, o el rendimiento es bajo, ¿en qué afecta esto a las mediciones de disponibilidad?

El período a lo largo del cual se mide la disponibilidad también puede tener un efecto importante en la definición de la misma. Si la disponibilidad requerida a lo largo de un año es del 99,9 por ciento, se permiten 8,75 horas de inactividad. Si la disponibilidad requerida durante un período continuado de cuatro semanas es del 99,9 por ciento, sólo se permiten 40 minutos de inactividad en cada período.

También es necesario identificar y negociar períodos de inactividad para las actividades de mantenimiento previstas, las actualizaciones de Service Pack y las actualizaciones de software. La cantidad de tiempo de inactividad previsto que se puede tolerar tiene un efecto importante en la definición de los requisitos de disponibilidad.

El lanzamiento de la versión RTM (para fabricantes) de Microsoft Exchange Server 2007 incluye nuevas funciones que pueden reducir los costos e incrementar la disponibilidad:

- Replicación continua local La replicación continua local (LCR) es una solución de

servidor único que utiliza tecnología integrada para crear y mantener una copia de un grupo de almacenamiento en un segundo conjunto de discos que están conectados al mismo servidor que el grupo de almacenamiento de producción. LCR proporciona transporte de registros asincrónico, reproducción de registros y cambio manual rápido a una copia de los datos. Para obtener más información acerca de la replicación continua local (LCR).

- Replicación continua en clúster La replicación continua en clúster (CCR) combina las funciones de replicación y reproducción de Exchange 2007 con las funciones de conmutación por error de los servicios de clústeres de Microsoft. La replicación continua en clúster es una solución que se puede implementar sin errores puntuales en un solo centro de datos o entre dos centros de datos. Para obtener más información acerca de la replicación continua en clúster. La replicación continua en clúster tiene varias ventajas con respecto a las agrupaciones en clúster de las versiones anteriores de Exchange Server y los clústeres de copia única de Exchange 2007. Para conocer los detalles de estas ventajas.
- Clústeres de copia única Los clústeres de copia única (SCC), conocidos como clústeres de almacenamiento compartido en las versiones anteriores de Exchange Server, están presentes en Exchange 2007, aunque con algunos cambios y mejoras significativas. Para obtener más información acerca de SCC

Microsoft Exchange Server 2007 Service Pack 1 (SP1) agrega una función adicional diseñada para proporcionar fiabilidad de sitios:

- Replicación continua en espera Réplica continua en espera (SCR) es una nueva característica introducida en Exchange 2007 SP1. Tal como indica su nombre, SCR ha sido diseñado para los casos en los que se usan o se habilita el uso de servidores de recuperación de espera. SCR amplía las características de replicación continua ya existente y habilita nuevos casos de disponibilidad de datos para los servidores de buzones Exchange 2007. SCR utiliza la misma tecnología de reproducción y de trasvase de registros que utilizan LCR y CCR, para proporcionar opciones y configuraciones de implementación añadidas. SCR se puede usar para replicar datos de servidores de buzones autónomos y servidores de buzones en clúster. Para obtener más información acerca de SCR.

## Selección de la solución de disponibilidad adecuada

Se pueden usar varias configuraciones para mejorar la disponibilidad de una implementación de Exchange 2007. Un paso adelante importante hacia la selección de la solución de disponibilidad correcta requiere un análisis de un conjunto seleccionado de opciones a fin de determinar qué solución es la que mejor se ajusta a los objetivos de su negocio y a los requisitos de disponibilidad. Una forma de hacerlo consiste en crear una tabla con un apartado para cada tipo de error. En cada apartado de la tabla, utilice una fila para identificar cada una de las soluciones que ofrezca una estrategia de recuperación en caso de error coherente con los requisitos de disponibilidad. Documente los factores importantes de la solución en las columnas. Los factores habituales son:

- Tiempo hasta la recuperación
- Costos de hardware y software asociados
- Probabilidad del evento
- Riesgos de complejidad
- Ventajas
- Consecuencias de la recuperación en los datos
- Costos de recursos asociados
- Implicaciones para el negocio
- Soluciones de otros fabricantes
- Inconvenientes

Después de completar estas tablas, seleccione varias soluciones para analizar los costos. Por cada solución seleccionada, elabore una estimación de costos por buzón que también se puede representar en una tabla. En la tabla de costos, no olvide incluir una fila que represente la calidad de la solución con respecto a los objetivos del negocio. Compruebe que evalúa varias opciones. Seleccione como mínimo una solución que cumpla los requisitos, pero que sea diferente de la solución habitual que su empresa hubiera elegido.

Un requisito expresado para una disponibilidad del 99,5 por ciento puede ser diferente a otro requisito para esa misma disponibilidad. Un requisito se puede referir a la disponibilidad de la plataforma de hardware exclusivamente, mientras que el otro, a la disponibilidad del servicio completo de un extremo a otro. Incluso la definición de la disponibilidad del servicio completo de un extremo a otro puede variar enormemente. Es importante saber exactamente cómo se va a medir cualquier requisito de disponibilidad.

Por último, revise los objetivos del negocio, los requisitos de disponibilidad, las soluciones posibles y el análisis de costos para seleccionar la solución. Durante este proceso, tenga en cuenta las claves siguientes para tomar una decisión acertada:

- **Replicación continua local** La replicación continua local (LCR) es una solución de servidor único que utiliza tecnología integrada para crear y mantener una copia de un grupo de almacenamiento en un segundo conjunto de discos que están conectados al mismo servidor que el grupo de almacenamiento de producción. LCR proporciona transporte de registros asincrónico, reproducción de registros y cambio manual rápido a una copia de los datos. Para obtener más información acerca de la replicación continua local (LCR).
- **Poner en duda las verdades establecidas**, que tal vez ya no sean aplicables Aproveche todo el potencial de Exchange 2007 en las etapas de diseño y evaluación. La experiencia demuestra que las soluciones más rentables pueden necesitar nuevos enfoques en cuanto a copias de seguridad, almacenamiento y operaciones.
- **Examine los errores puntuales de su sistema de mensajería** Sólo con que haya una única copia de datos del buzón almacenada en una red de área de almacenamiento (SAN), los datos no estarán completamente protegidos contra daños y errores. En esa única copia, los datos se pueden dañar o se pueden producir en ellos errores de varias formas, independientemente de la cantidad de redundancia que proporcione la SAN. Con los clústeres de copia única, un error de la SAN puede suponer horas de pérdida de datos y días de interrupción de las actividades. La replicación continua en clúster es una solución de disponibilidad que puede perder datos cuando se produce un error del servidor,

pero que guarda dos copias de los datos. CCR mitiga la mayoría de las pérdidas de datos cuando se produce un error del servidor al usar una función del servidor Transporte de concentradores denominada volcado de archivos de transporte. Como consecuencia, los datos de buzón se conservan en la mayoría de las veces en que se produce un error del hardware.

- Explorar la gama de opciones de almacenamiento con que cuenta para cada solución La replicación continua en clúster ofrece a las organizaciones la posibilidad de usar una amplia gama de soluciones de almacenamiento como el almacenamiento directo conectado. La replicación continua en clúster no requiere ninguna infraestructura de SAN, que reduce la complejidad y los costos. El almacenamiento directo conectado, ya sea en una SAN o una solución de almacenamiento de bajo costo, resulta más fácil de implementar y usar.
- La replicación continua en clúster y la replicación continua local permiten cambiar la estrategia de copia de seguridad y pasar de una copia completa diaria habitual a una copia completa menos frecuente y una estrategia incremental diaria La replicación continua en clúster y la replicación continua local también pueden admitir un acuerdo de nivel de servicio (SLA) más breve para la recuperación tras el primer error. El SLA para la recuperación después de un error doble (las dos copias tiene errores o están dañadas) puede necesitar una prolongación para abarcar el SLA de recuperación actual. Los cambios de este tipo pueden reducir drásticamente el costo total de propiedad (TCO), ya que los costos de copia de seguridad suelen ser un componente importante del TCO. Además, al cambiar a una estrategia de copia de seguridad basada en disco, se pueden reducir también los costos de copia de seguridad.
- Investigue el uso de la tecnología de replicación continua en Exchange 2007 para crear una solución CCR hace que la tecnología de replicación de otros fabricantes sea innecesaria.
- Crea tablas que identifiquen el comportamiento de recuperación y los costos asociados a las diversas opciones En el caso de la tabla de costos, compruebe que contenga varias opciones que cuestionen sus prácticas existentes.

## Establecimiento de los objetivos de disponibilidad y requisitos del SLA

Alcanzar niveles altos de disponibilidad comienza por la implementación de productos y componentes de buena calidad. No obstante, no es probable que estos productos y componentes por sí solos proporcionen los niveles sostenidos de disponibilidad necesarios. Es recomendable tener en cuenta los objetivos de disponibilidad en el proceso de diseño, en la primera fase posible del desarrollo. De esta forma se evita la posibilidad de incurrir en mayores costos debidos a la necesidad de volver a realizar el trabajo, las actualizaciones no planeadas necesarias para satisfacer la disponibilidad exigida, las herramientas no planeadas para supervisar la infraestructura, los gastos no planeados para eliminar errores puntuales de la infraestructura, de la capacidad de mantenimiento y de la capacidad de servicio.

Uno de los primeros pasos para obtener una alta disponibilidad es revisar el SLA establecido en la organización. Tras establecer un SLA, es posible determinar las configuraciones de implementación y del servidor de Exchange 2007 más adecuadas para el acuerdo.

A continuación se incluyen los aspectos fundamentales para obtener una disponibilidad elevada relacionados con la recuperación ante desastres:

- **Tiempo de inactividad permitido** Tenga en cuenta el tiempo de inactividad permitido máximo aceptable para su organización según la definición de la disponibilidad del servicio de Exchange de la misma. Es posible que pueda cumplir el SLA de su organización mediante una estrategia de recuperación de tono de marcado de mensajería, dependiendo de la definición de tiempo de inactividad de su organización. Una estrategia de recuperación de tono de marcado de mensajería proporciona buzones temporales a los usuarios para que puedan enviar y recibir mensajes de correo electrónico inmediatamente después de que se produzca un desastre. Esta estrategia restaura rápidamente el servicio de correo electrónico antes de recuperar datos históricos de los buzones. Normalmente, la recuperación se hará combinando finalmente datos históricos y temporales de los buzones
- **Tiempo de recuperación permitido** Tenga en cuenta el tiempo máximo permitido para cada tipo de operación de recuperación ante desastres. Por ejemplo, es necesario especificar el período de tiempo aproximado necesario para recuperar un buzón, una sola base de datos o todo un servidor que este ejecutando Exchange 2007.
- **Tolerancia a pérdida de datos** Tenga en cuenta la tolerancia de su organización a la pérdida de datos temporal o permanente de Exchange. Por ejemplo, es posible que su organización tolere la pérdida temporal de los datos de los buzones desde la copia de seguridad anterior durante un período de 24 horas, siempre que los usuarios puedan enviar y recibir mensajes en un período de 4 horas. En otros casos, es posible que prefiera un requisito más estricto, como solicitar que todos los datos de Exchange hasta el momento del error se restauren en 4 horas.

## Establecimiento de acuerdos de nivel de servicio con sus proveedores

Muchas empresas que dan importancia a las soluciones de alta disponibilidad utilizan los servicios de proveedores terceros para lograr sus objetivos de alta disponibilidad. En estos casos, para lograr un sistema de mensajería de alta disponibilidad, se requieren los servicios de proveedores externos de hardware y software. Unos proveedores que no responden o unos empleados del proveedor mal entrenados pueden reducir la disponibilidad del sistema de mensajería.

Asegúrese de negociar un SLA con cada uno de sus proveedores más importantes. El establecimiento de los SLA con sus proveedores ayuda a garantizar que el sistema de mensajería funciona de acuerdo con las especificaciones, admite el crecimiento necesario y está disponible según un estándar específico. La ausencia de un SLA puede aumentar considerablemente el tiempo que el sistema de mensajería no está disponible. Además de establecer un SLA con sus principales proveedores, también debe probar periódicamente los procedimientos de escalado mediante la realización de pruebas de solicitud de soporte técnico. Para confirmar que dispone de la información de contacto más reciente, asegúrese de probar también los localizadores (pagers) y los árboles telefónicos.

## Consideraciones acerca de la disponibilidad

Recomendamos tener en cuenta los siguientes aspectos para determinar los requisitos de almacenamiento:

- Comprenda la vulnerabilidad a los errores del diseño de infraestructura propuesto. Se debe asegurar de que no haya errores puntuales. Un error puntual es cualquier componentes de la infraestructura de mensajería que no tiene capacidad de redundancia y pueda afectar al usuario cuando se produzca el error. El diseño técnico propuesto para la solución debería abarcar toda la configuración de extremo a extremo.
- Tenga en cuenta los niveles mínimos de disponibilidad del servicio de mensajería que necesita el negocio así como los niveles mínimos de fiabilidad, capacidad de mantenimiento y de servicio de cada uno de los componentes de la infraestructura de mensajería.
- Tenga en cuenta la capacidad de probar o simular nuevos componentes para garantizar que satisfacen los requisitos especificados. Para evaluar si los nuevos componentes del diseño pueden satisfacer los requisitos establecidos, es importante que el régimen de pruebas que promueve se cerciore de que es posible proporcionar la disponibilidad esperada. También se deben realizar pruebas cuando se llevan a cabo tareas de mantenimiento en los componentes. Se deberán tener muy en cuenta las herramientas de simulación para generar la demanda de usuario prevista del nuevo servicio de tecnología de la información, con el fin de garantizar que los componentes seguirán funcionando en condiciones de volumen y sobrecarga.

## Consideraciones acerca de la alta disponibilidad

Una solución de mensajería de alta disponibilidad requiere la inversión y la implementación de una solución de supervisión, de procesos de administración de mantenimiento, de herramientas de administración del sistema y de redundancia. En las implementaciones de alta disponibilidad es importante que no existan errores puntuales en esta configuración de extremo a extremo. El diseño de alta disponibilidad debe tener en cuenta la eliminación de los errores puntuales y la disposición de componentes alternativos que posibiliten la interrupción mínima del funcionamiento del negocio en el caso de error de un componente. El diseño también debe eliminar o reducir al mínimo los efectos en el funcionamiento del negocio de los tiempos de inactividad planeados normalmente necesarios para las actividades de mantenimiento, como la implementación de cambios en la infraestructura. Los criterios de recuperación deberían definir la recuperación y el restablecimiento del servicio como un objetivo clave en la fase de diseño dedicada a la recuperación.

Al desarrollar un plan de implementación para la solución de mensajería debe identificar los objetivos de la solución. Esto es especialmente importante, ya que debe diseñar las características de disponibilidad de la solución. A menudo, los objetivos empresariales dan lugar a contradicciones. Por ejemplo, sus objetivos de disponibilidad podrían incluir un 100% de disponibilidad y necesitar que se apliquen las actualizaciones de seguridad más

recientes en una semana desde que están disponibles. Los costos suelen ser otro factor difícil en los planes de implementación. Adoptar una metodología de planeamiento que identifique todos los requisitos empresariales y evalúe las opciones disponibles teniendo en cuenta dichos requisitos es el enfoque óptimo para encontrar la mejor solución para su negocio.

Para lograr una alta disponibilidad, se necesita un enfoque continuo y constante en las prácticas operativas de la organización. Es preciso entender todas las causas de las interrupciones. En el caso de interrupciones no programadas que se pueden evitar con cambios en el proceso, hay que iniciar los cambios en el proceso adecuados.

Otro factor clave para maximizar la disponibilidad es la supervisión proactiva del entorno de Exchange. Mediante la supervisión proactiva, las áreas problemáticas en el sistema se pueden identificar antes de que produzcan errores e interrupciones no programadas. Además, la supervisión puede alertar al personal de operaciones de los problemas no recuperados automáticamente por el sistema. En estos casos, una respuesta a tiempo puede acortar la duración de la interrupción no programada e incrementar así la disponibilidad.

La disponibilidad de Exchange 2007 depende de la disponibilidad de otros servicios en la infraestructura de tecnología de la información. Los servicios como Active Directory y la conexión en red deben estar funcionando para que Exchange esté operativo. La disponibilidad de estos servicios repercute directamente en la disponibilidad de Exchange. Por consiguiente, debe asegurarse de que los requisitos de disponibilidad de Exchange no sean superiores a los de sus dependencias. A continuación, mostramos una lista de dependencias típica:

- Active Directory
- Sistema de nombres de dominio
- Red TCP/IP
- Subsistema de almacenamiento
- Servicios de copia de seguridad
- Servicios de supervisión
- Infraestructura de centro de datos (electricidad y aire acondicionado)

Después de establecer los objetivos empresariales y sus SLA para las dependencias de Exchange, recomendamos desarrollar una lista inicial de requisitos de disponibilidad para servicios de mensajería. Esta lista debería incluir cada clase general de error y el objetivo de tiempo de recuperación (RTO) esperado. En el caso de los errores relacionados con datos, esta lista debería incluir una indicación del impacto de los errores en los datos. Esto se puede especificar indicando un objetivo de punto de recuperación (RPO). Un RPO identifica el impacto de los datos especificando un tiempo que define un nivel de datos que estará disponible después de la recuperación.