Documento técnico de AWS

Recuperación de desastres de cargas de trabajo en AWS: recuperación en la nube

Recuperación de desastres de cargas de trabajo en AWS: recuperación en la nube: Documento técnico de AWS

Table of Contents

Recuperación de desastres de cargas de trabajo en AWS: recuperación en la nube

Fecha de publicación: 12 de febrero de 2021 (Revisiones del documento)

Resumen

La recuperación de desastres es el proceso de prepararse para poder recuperar los datos después de un desastre. Se considera desastre cualquier evento que impida que una carga de trabajo o un sistema cumplan sus objetivos empresariales en la ubicación de implementación principal. En este documento se explican las prácticas recomendadas para planificar y probar la recuperación de desastres de cualquier carga de trabajo implementada en AWS. Además, se explican diferentes enfoques para mitigar riesgos y cumplir el objetivo de tiempo de recuperación (RTO) y el objetivo de punto de recuperación (RPO) para esa carga de trabajo.

Introducción

La carga de trabajo debe realizar la función prevista de manera correcta y coherente. Para lograrlo, debe diseñar una arquitectura resiliente. La resiliencia es la capacidad de una carga de trabajo de recuperarse de interrupciones en la infraestructura o el servicio, para incorporar dinámicamente recursos computacionales que satisfagan la demanda y para mitigar las interrupciones, como errores de configuración o problemas de red temporales.

La recuperación de desastres es una parte importante de la estrategia de resiliencia y se refiere a cómo responde la carga de trabajo cuando ocurre un desastre (un desastre es un evento que causa un impacto negativo grave en una empresa). Esta respuesta debe basarse en los objetivos empresariales de su organización, que especifican la estrategia de la carga de trabajo para evitar la pérdida de datos, lo que se conoce comoobjetivo de punto de recuperación (RPO), y en la reducción del tiempo de inactividad cuando la carga de trabajo no esté disponible para su uso, lo que se conoce como objetivo de tiempo de recuperación (RTO). Por lo tanto, debe implementar la resiliencia en el diseño de las cargas de trabajo en la nube para cumplir con los objetivos de recuperación (RPO y RTO) para un determinado desastre único. Este enfoque ayuda a su organización a mantener la continuidad del negocio como parte del plan de continuidad del negocio (BCP).

Este documento se centra en cómo planificar, diseñar e implementar arquitecturas en AWS que cumplan con los objetivos de recuperación de desastres de su empresa. Esta herramienta está dirigida a personas con puestos relacionados con la tecnología, como directores de tecnología (CTO), arquitectos, desarrolladores y miembros de los equipos de operaciones.

Disponibilidad y recuperación de desastres

La recuperación de desastres se puede comparar con la disponibilidad, que es otro componente importante de la estrategia de resiliencia. Mientras que la recuperación de desastres mide los objetivos para eventos puntuales, los objetivos de disponibilidad miden los valores medios durante un período de tiempo.

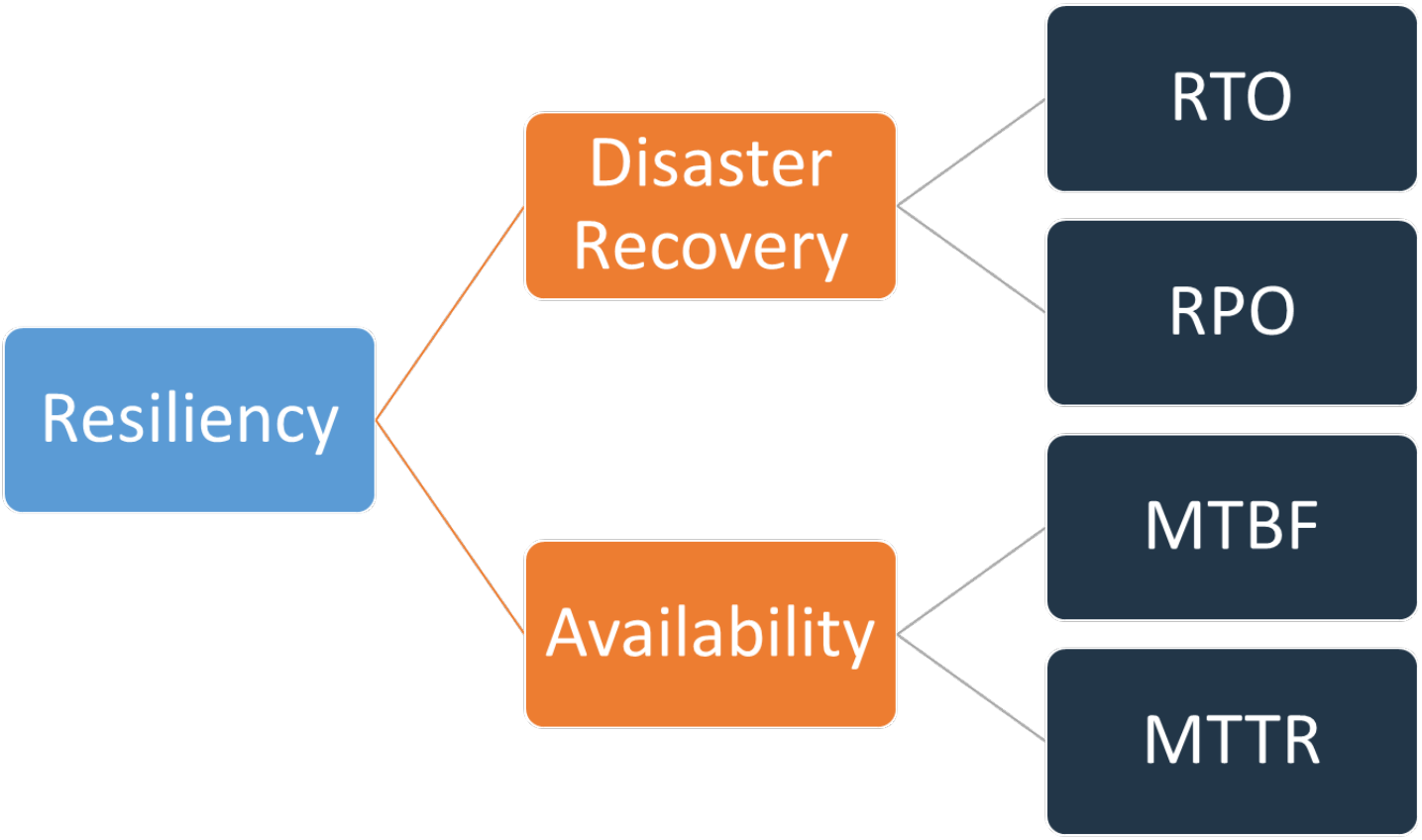

Ilustración 1: Objetivos de resiliencia

La disponibilidad se calcula dividiendo el tiempo medio entre errores (MTBD) entre el tiempo medio de recuperación (MTTR):

$$Availability = \frac{Available\ for\ Use\ Time}{Total\ Time} = \frac{MTBF}{MTBF + MTTR}$$

Este enfoque a menudo se denomina “nueves”, donde un objetivo de disponibilidad del 99,9 % se denomina “tres nueves”.

Para su carga de trabajo, puede ser más fácil contar las solicitudes correctas y erróneas en lugar de utilizar un enfoque basado en el tiempo. En este caso, se puede utilizar el siguiente cálculo:

$$Availability = \frac{Successful\ Responses}{Valid\ Requests}$$

La recuperación de desastres se centra en los desastres, mientras que la disponibilidad se centra en las interrupciones más comunes de menor escala, como errores de componentes, problemas de red y picos de carga. El objetivo de la recuperación de desastres es la continuidad del negocio, mientras que la disponibilidad se centra en maximizar el tiempo que una carga de trabajo está disponible para realizar la funcionalidad empresarial prevista. Ambos deben formar parte de su estrategia de resiliencia.

Modelo de responsabilidad compartida para la resiliencia

La resiliencia es una responsabilidad compartida entre AWS y usted, el cliente. Es importante que comprenda cómo operan la recuperación de desastres y la disponibilidad, como parte de la resiliencia, en este modelo compartido.

Responsabilidad de AWS: “resiliencia de la nube”

AWS es responsable de la resiliencia de la infraestructura en la que se ejecutan todos los servicios que se ofrecen en la nube de AWS. Esta infraestructura está compuesta por hardware, software, redes e instalaciones que ejecutan servicios en la nube de AWS. AWS realiza esfuerzos comercialmente razonables para que estos servicios en la nube de AWS estén disponibles, garantizando que la disponibilidad del servicio cumpla o supere los acuerdos de nivel de servicio (SLA) de AWS.

La infraestructura en la nube global de AWS está diseñada para permitir a los clientes crear arquitecturas de carga de trabajo altamente resilientes. Cada región de AWS está completamente aislada y está compuesta por varias zonas de disponibilidad, que son particiones de infraestructura aisladas físicamente. Las zonas de disponibilidad aíslan los errores que podrían afectar a la resiliencia de las cargas de trabajo, evitando que afecten a otras zonas de la región. Pero al mismo tiempo, todas las zonas de disponibilidad de una región de AWS están interconectadas con redes de ancho de banda alto y baja latencia, a través de una fibra metropolitana exclusiva totalmente redundante que proporciona una red de alto rendimiento y baja latencia entre las zonas. Todo el tráfico entre las zonas está cifrado. El rendimiento de la red es suficiente como para llevar a cabo la replicación sincrónica entre las zonas. Las zonas de disponibilidad simplifican el proceso de dividir las aplicaciones para obtener una alta disponibilidad.

Responsabilidad del cliente: “resiliencia en la nube”

Su responsabilidad vendrá determinada por los servicios en la nube de AWS que seleccione. Esto determinará la cantidad de trabajo de configuración que debe realizar como parte de sus responsabilidades de resiliencia. Por ejemplo, un servicio como Amazon Elastic Compute Cloud (Amazon EC2) requiere que el cliente lleve a cabo todas las tareas de administración y configuración de resiliencia necesarias. Los clientes que implementan instancias de Amazon EC2 son responsables de implementar instancias de EC2 en varias ubicaciones (como las zonas de disponibilidad de AWS), implementar la reparación automática mediante servicios como AWS

Auto Scaling, así como la aplicación de prácticas recomendadas de arquitectura de cargas de trabajo resilientes para las aplicaciones instaladas en las instancias. En el caso de los servicios administrados, como Amazon S3 y Amazon DynamoDB, AWS gestiona la capa de infraestructura, el sistema operativo y las plataformas, mientras que los clientes acceden a los puntos de conexión para recuperar y almacenar los datos. Usted es responsable de administrar la resiliencia de sus datos, incluidas las estrategias de copia de seguridad, control de versiones y replicación.

La implementación de la carga de trabajo en varias zonas de disponibilidad en una región de AWS forma parte de una estrategia de alta disponibilidad diseñada para proteger las cargas de trabajo aislando los problemas en una zona de disponibilidad. Aquí se utiliza la redundancia de las demás zonas de disponibilidad para seguir atendiendo solicitudes. La arquitectura Multi-AZ también forma parte de una estrategia de recuperación de desastres diseñada para que las cargas de trabajo estén mejor aisladas y protegidas de problemas como cortes de electricidad, rayos, tornados, terremotos y similares. Las estrategias de recuperación de desastres también pueden incluir usar varias regiones de AWS. Por ejemplo, en una configuración activa/pasiva, el servicio para la carga de trabajo conmutará por error desde la región activa a la región de recuperación de desastres si la región activa ya no puede atender solicitudes.

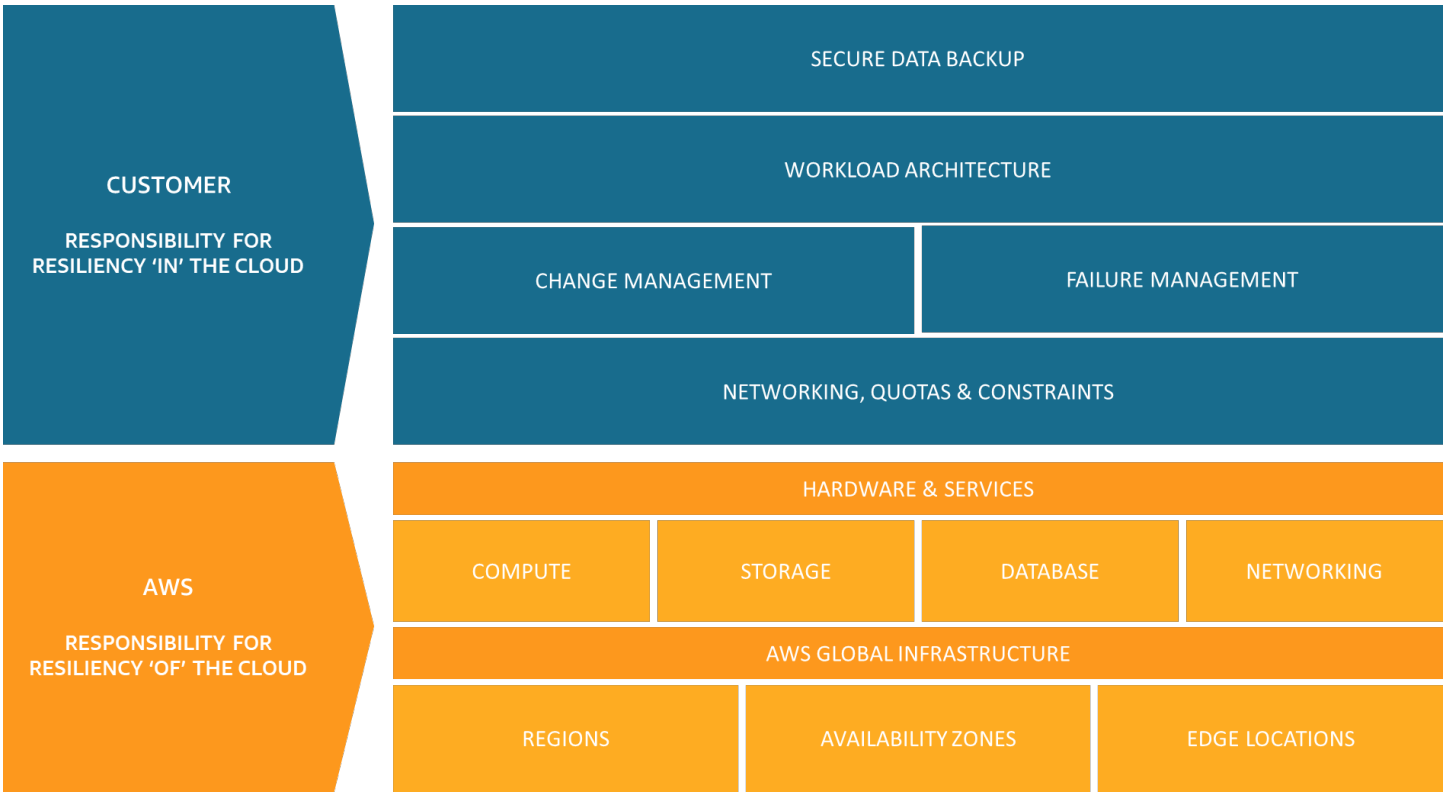

Ilustración 2: La resiliencia es una responsabilidad compartida entre AWS y el cliente

¿Qué es un desastre?

Al planificar la recuperación de desastres, deberá evaluar estas tres categorías principales de desastres:

- Desastres naturales, como terremotos o inundaciones
- Errores técnicos, como fallos de alimentación o conectividad de red
- Acciones del hombre, como una configuración incorrecta inadvertida o acceso o modificación no autorizados/externos

Cada uno de estos desastres potenciales también tendrá un impacto geográfico que puede ser local, regional, nacional, continental o mundial. Tanto la naturaleza del desastre como el impacto geográfico son importantes al diseñar la estrategia de recuperación de desastres. Por ejemplo, puede mitigar un problema de inundación local que provoque una interrupción del centro de datos empleando una estrategia Multi-AZ, ya que no afectaría a más de una zona de disponibilidad. Sin embargo, un ataque a los datos de producción requeriría aplicar una estrategia de recuperación de desastres de conmutación por error para hacer copias de seguridad de los datos en otra región de AWS.

La alta disponibilidad no es una recuperación de desastres

Tanto la disponibilidad como la recuperación de desastres se basan en algunas de las mismas prácticas recomendadas, como la supervisión de errores, la implementación en varias ubicaciones y la conmutación por error automática. Sin embargo, la disponibilidad se centra en los componentes de la carga de trabajo, mientras que la recuperación de desastres se centra en copias discretas de toda la carga de trabajo. La recuperación de desastres tiene objetivos diferentes a la disponibilidad, ya que mide el tiempo de recuperación después de los eventos a gran escala que se califican como desastres. Primero debe asegurarse de que su carga de trabajo cumpla los objetivos de disponibilidad, pues una arquitectura de alta disponibilidad le permitirá satisfacer las necesidades de los clientes en caso de que la disponibilidad afecte a los eventos. Su estrategia de recuperación de desastres requiere enfoques diferentes a los de disponibilidad, centrándose en la implementación de sistemas discretos en varias ubicaciones, de modo que pueda conmutar por error toda la carga de trabajo si es necesario.

Debe tener en cuenta la disponibilidad de su carga de trabajo en la planificación de la recuperación de desastres, ya que influirá en el enfoque que adopte. Una carga de trabajo que se ejecuta en una sola instancia de Amazon EC2 en una zona de disponibilidad no tiene alta disponibilidad. Si un problema de inundaciones local afecta a esa zona de disponibilidad, este escenario requiere conmutar por error a otra zona de disponibilidad para cumplir con los objetivos de recuperación de desastres. Compare este escenario con una carga de trabajo de alta disponibilidad activa/activa en varios sitios implementada, donde la carga de trabajo se implementa en varias regiones activas y todas las regiones atienden el tráfico de producción. En este caso, incluso en el improbable caso de que un desastre masivo acabe con una región entera, la estrategia de recuperación de desastres se logra dirigiendo todo el tráfico a las regiones restantes.

La forma en que se abordan los datos también difiere entre la disponibilidad y la recuperación de desastres. Considere una solución de almacenamiento que se replica continuamente en otro sitio para lograr una alta disponibilidad (como una carga de trabajo activa/activa en varios sitios). Si un archivo o archivos se eliminan o se corrompen en el dispositivo de almacenamiento principal, esos cambios destructivos se pueden replicar en el dispositivo de almacenamiento secundario. En este escenario, a pesar de la alta disponibilidad, la capacidad de conmutación por error en caso de eliminación o corrupción de datos se verá comprometida. En cambio, también se requiere una copia de seguridad en un momento dado como parte de una estrategia de recuperación de desastres.

Plan de continuidad del negocio (BCP)

Su plan de recuperación de desastres debe formar parte del plan de continuidad del negocio (BCP) de su organización, no debe ser un documento independiente. No tiene sentido mantener objetivos estrictos de recuperación de desastres para restaurar una carga de trabajo si los objetivos empresariales de esa carga de trabajo no se pueden alcanzar debido al impacto del desastre en elementos de la empresa que no sean la carga de trabajo. Por ejemplo, un terremoto podría impedirle transportar productos comprados a través de la aplicación de comercio electrónico; incluso si una recuperación de desastres efectiva mantuviera la carga de trabajo en funcionamiento, el BCP debería adaptarse a las necesidades de transporte. Su estrategia de recuperación de desastres debe basarse en los requisitos, las prioridades y el contexto de la empresa.

Análisis del impacto en la empresa y evaluación de riesgos

El análisis de impacto en la empresa debe cuantificar el impacto en la empresa de una interrupción en las cargas de trabajo. Debe identificar el impacto en los clientes internos y externos al no poder usar las cargas de trabajo y su efecto en el negocio. El análisis debería ayudar a determinar con qué velocidad puede ponerse a disposición la carga de trabajo y cuánta pérdida de datos se puede tolerar. Sin embargo, es importante tener en cuenta que los objetivos de recuperación no deben hacerse de forma aislada; la probabilidad de interrupción y el coste de la recuperación son factores clave que ayudan a explicar el valor empresarial de proporcionar recuperación de desastres para una carga de trabajo.

El impacto en la empresa puede depender del tiempo. Se recomienda incluir esto en la planificación de recuperación de desastres. Por ejemplo, es probable que la interrupción del sistema de nómina afecte en gran medida a la empresa justo antes de pagar las nóminas, pero puede tener un efecto poco perceptible si ocurre después de pagar las nóminas.

Una evaluación de riesgos del tipo de desastre y el impacto geográfico junto con información general sobre la implementación técnica de la carga de trabajo determinará la probabilidad de que se produzca una interrupción en cada tipo de desastre.

Para cargas de trabajo muy críticas, puede considerar la alta disponibilidad en varias regiones con copias de seguridad continuas implementadas para minimizar el impacto en el negocio. En el caso de cargas de trabajo menos críticas, quizá una estrategia válida sea no implementar ninguna recuperación de desastres. Y para algunos escenarios de desastres, también es válido no tener ninguna estrategia de recuperación de desastres implementada teniendo en cuenta la baja

probabilidad de que ocurra el desastre. Hay que recordar que las zonas de disponibilidad dentro de una región de AWS ya se han diseñado con una distancia significativa entre ellas y una planificación cuidadosa de la ubicación, para que los desastres más comunes solo afecten a una zona y no a las demás. Por lo tanto, es posible que una arquitectura Multi-AZ dentro de una región de AWS ya cubra las necesidades de mitigación.

Debe evaluarse el coste de las opciones de recuperación de desastres para garantizar que la estrategia de recuperación proporcione el nivel correcto de valor empresarial considerando el impacto y el riesgo en el negocio.

Con toda esta información, se puede documentar la amenaza, el riesgo, el impacto y el coste de los diferentes escenarios de desastre y las opciones de recuperación asociadas. Debe utilizar esta información para determinar sus objetivos de recuperación para cada una de las cargas de trabajo.

Objetivos de recuperación (RTO y RPO)

Al crear una estrategia de recuperación de desastres (DR), las organizaciones suelen planificar el objetivo de tiempo de recuperación (RTO) y el objetivo de punto de recuperación (RPO).

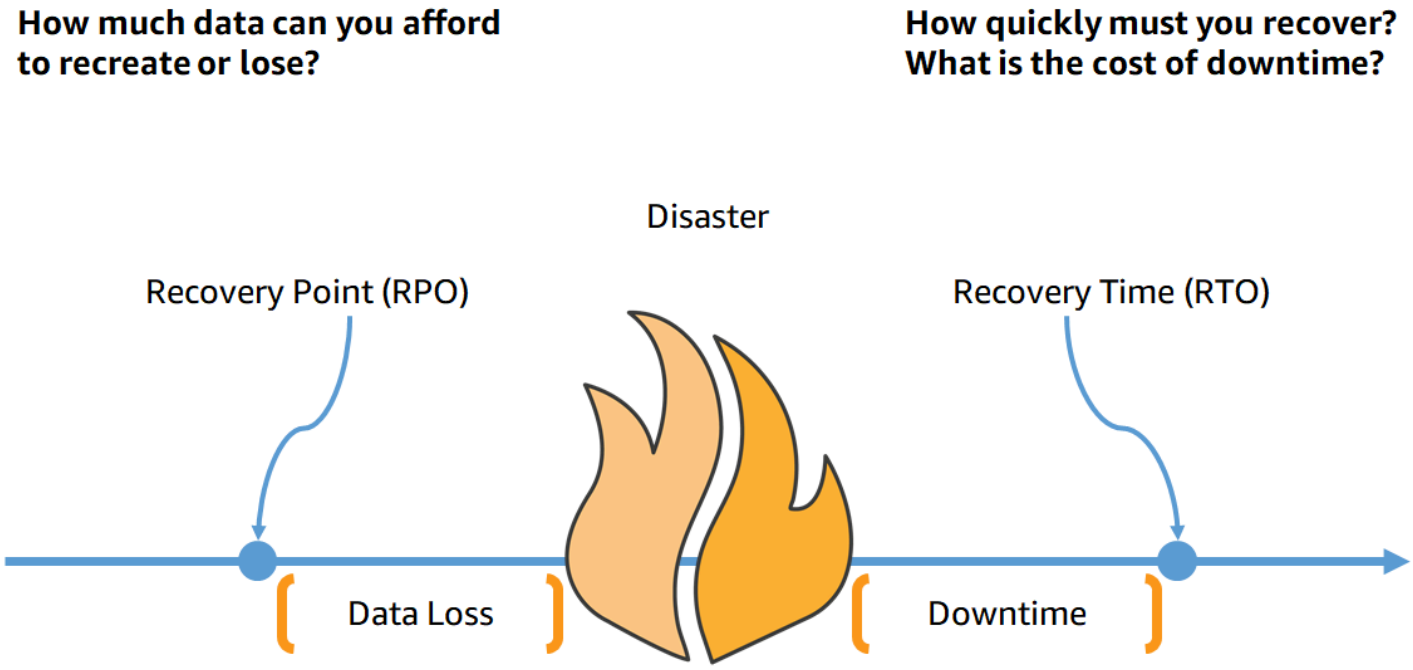

Ilustración 3: Objetivos de recuperación

El objetivo de tiempo de recuperación (RTO) es el retraso máximo aceptable entre la interrupción del servicio y la restauración del servicio. Este objetivo, que lo define la organización, determina lo que se considera una ventana de tiempo aceptable mientras el servicio no está disponible.

En este documento se analizan cuatro grandes estrategias de DR: copia de seguridad y restauración, luz piloto, espera semiactiva y activa/activa en varios sitios (consulte la sección Opciones de recuperación de desastres en la nube). En el siguiente diagrama, la empresa ha determinado su RTO máximo admisible, así como el límite de lo que puede gastar en la estrategia de restauración de servicios. Según los objetivos de la empresa, las estrategias luz piloto o espera semiactiva de DR cubrirían tanto el RTO como los criterios de coste.

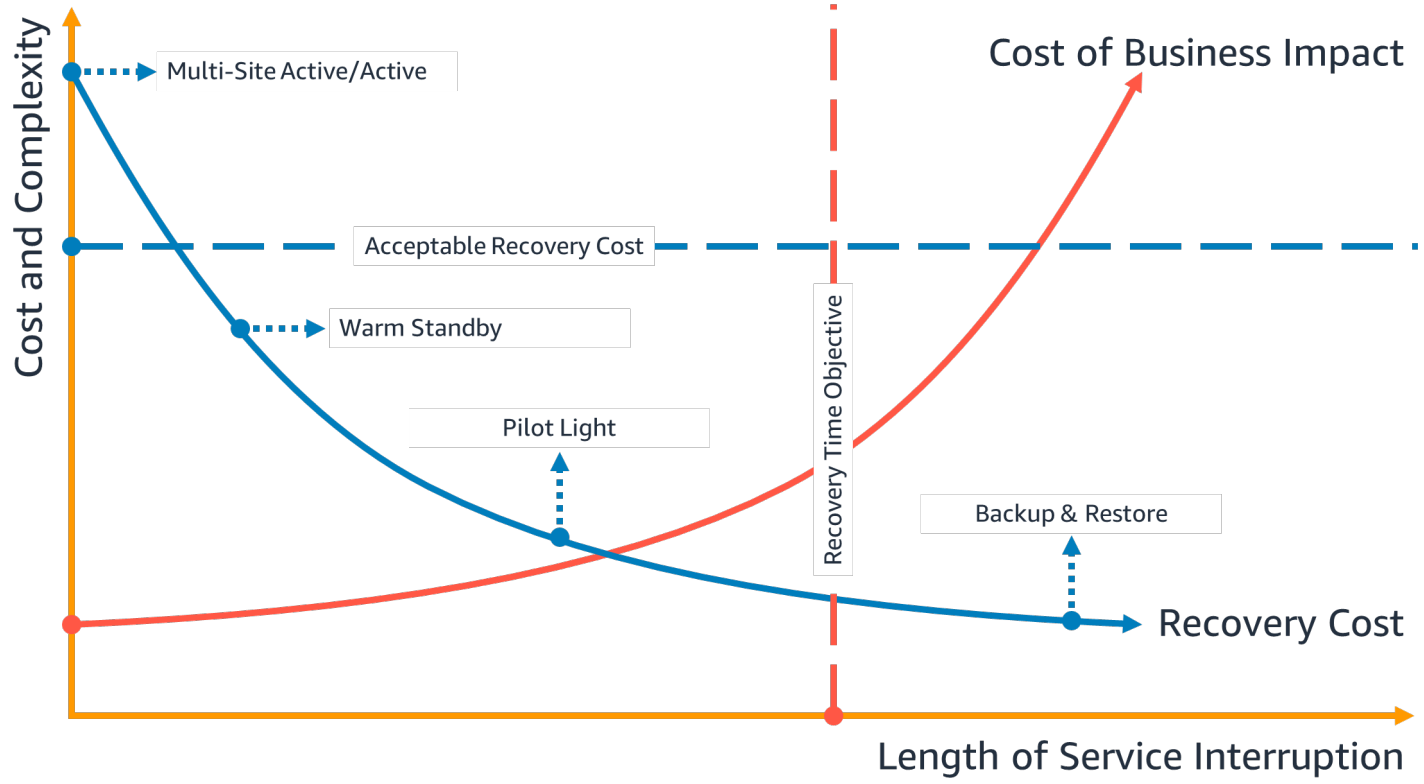

Ilustración 4: Objetivo de tiempo de recuperación

El objetivo de punto de recuperación (RPO) es la cantidad de tiempo máxima aceptable desde el último punto de recuperación de datos. Este objetivo, que lo define la empresa, determina lo que se considera una pérdida aceptable de datos entre el último punto de recuperación y la interrupción del servicio.

En el siguiente diagrama, la empresa ha determinado su RPO máximo admisible, así como el límite de lo que puede gastar en la estrategia de recuperación de datos. De las cuatro estrategias de DR, las estrategias luz piloto o espera semiactiva cumplen con los criterios de RPO y coste.

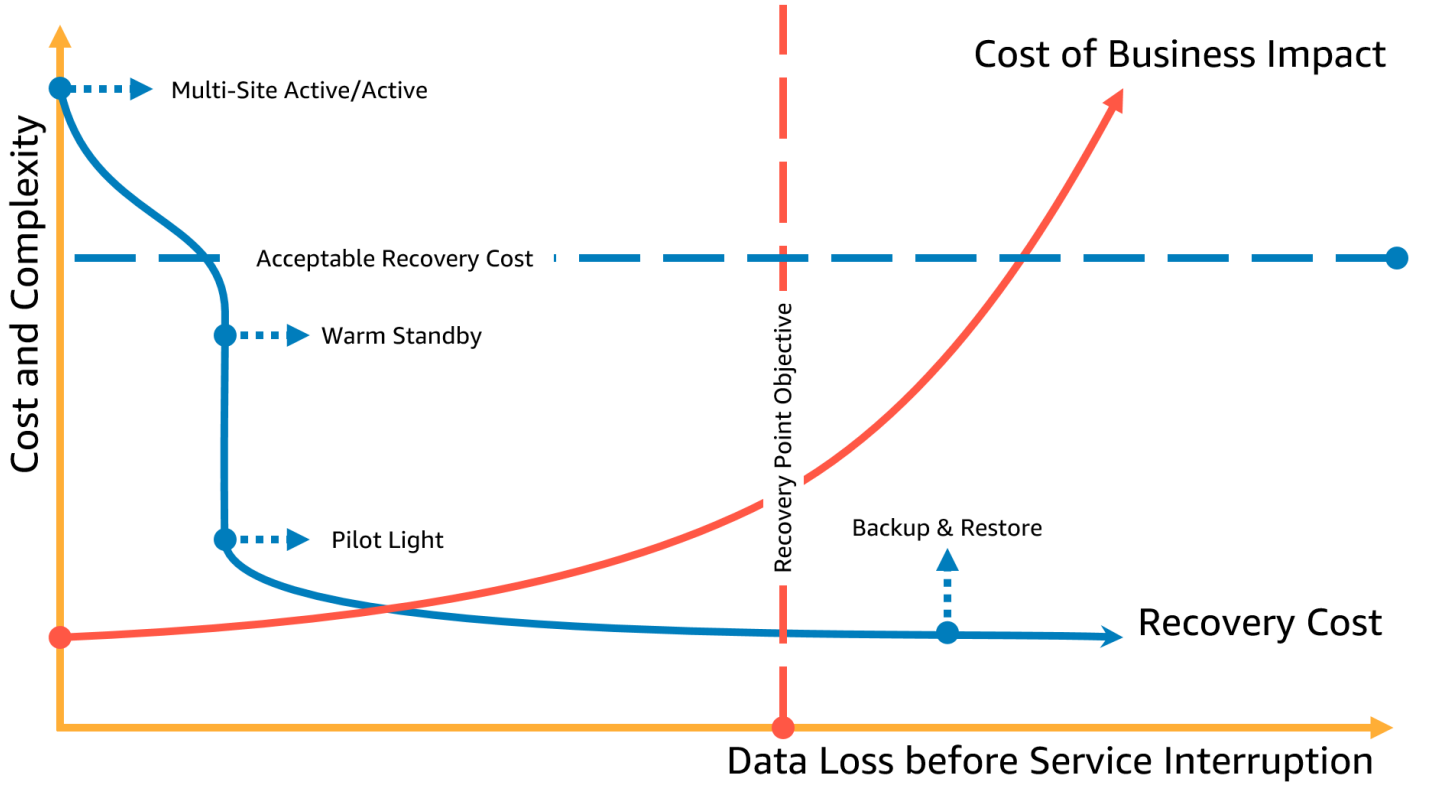

Ilustración 5: Objetivo de punto de recuperación

> **Note**
>
> Si el coste de la recuperación es mayor al del error o la pérdida, la opción de recuperación no debería implementarse a menos que haya un factor secundario, como los requisitos reglamentarios.

La recuperación de desastres es diferente en la nube

Las estrategias de recuperación de desastres evolucionan a la par que la innovación técnica. Un plan de recuperación de desastres local puede implicar el transporte físico de cintas o la replicación de datos a otro sitio. Su organización debe volver a evaluar el impacto empresarial, el riesgo y el coste de sus estrategias de recuperación de desastres anteriores para cumplir los objetivos de DR en AWS. La recuperación de desastres en la nube de AWS incluye las siguientes ventajas en comparación con los entornos tradicionales:

- Rápida recuperación ante un desastre de una complejidad reducida
- Unas pruebas simples y repetibles permiten realizar pruebas con mayor facilidad y frecuencia
- Una menor sobrecarga de gestión disminuye la carga operativa
- Las oportunidades para automatizar disminuyen las probabilidades de error y mejoran el tiempo de recuperación

AWS le permite intercambiar el gasto de capital fijo de un centro de datos de copia de seguridad físico por los gastos operativos variables de un entorno del tamaño correcto en la nube, lo que puede reducir significativamente los costes.

Para muchas organizaciones, la recuperación de desastres local se basaba en el riesgo de interrupción de una carga o varias cargas de trabajo en un centro de datos y la recuperación de datos de la copia de seguridad o replicados en un centro de datos secundario. Cuando las organizaciones implementan cargas de trabajo en AWS, pueden implementar una carga de trabajo bien diseñada y confiar en el diseño de la infraestructura en la nube global de AWS para ayudar a mitigar el efecto de tales interrupciones. Consulte el documento técnico AWS Well-Architected Framework - Pilar de fiabilidad para obtener más información sobre las prácticas de arquitectura recomendadas para diseñar y gestionar cargas de trabajo fiables, seguras, eficientes y rentables en la nube.

Si las cargas de trabajo están en AWS, no tiene que preocuparse por la conectividad del centro de datos (a excepción de la capacidad de acceder a él), el suministro eléctrico, el aire acondicionado, la extinción de incendios ni el hardware. Usted no tiene que administrar nada y tiene acceso a varias zonas de disponibilidad con aislamiento de fallos (cada una compuesta por uno o más centros de datos discretos).

Región de AWS única

Para un evento de desastre basado en la interrupción o pérdida de un centro de datos físico, la implementación de una carga de trabajo de alta disponibilidad en varias zonas de disponibilidad dentro de una sola región de AWS ayuda a mitigar los desastres naturales y técnicos y reduce el riesgo de amenazas de origen humano, como errores o actividades no autorizadas que pueden provocar pérdida de datos. Cada región de AWS está compuesta por varias zonas de disponibilidad y cada una de ellas está aislada de los fallos de las demás zonas. Además, cada zona de disponibilidad consta de varios centros de datos físicos. Para aislar mejor los problemas graves y lograr una alta disponibilidad, puede dividir las cargas de trabajo en varias zonas de la misma región. Las zonas de disponibilidad se diseñaron para ofrecer redundancia física y proveer resiliencia, lo que permite lograr un rendimiento continuo, incluso si se producen interrupciones en el suministro de electricidad, cortes en el servicio de Internet, inundaciones y otras catástrofes naturales. Consulte Infraestructura en la nube global de AWS para descubrir cómo lo hace AWS.

Al implementar en varias zonas de disponibilidad de una sola región de AWS, su carga de trabajo está mejor protegida contra los fallos de un solo centro de datos (o incluso de varios). Para disfrutar de mayor seguridad en la implementación de una sola región, puede hacer copias de seguridad de los datos y la configuración (incluida la definición de la infraestructura) en otra región. Esta estrategia reduce el alcance de su plan de recuperación de desastres e incluye únicamente la copia de seguridad y la restauración de datos. Aprovechar la resiliencia de varias regiones mediante la realización de copias de seguridad en otra región de AWS es sencillo y económico si comparamos con las demás opciones de varias regiones que se describen en la siguiente sección. Por ejemplo, hacer copias de seguridad en Amazon Simple Storage Service (Amazon S3) le permite recuperar sus datos de forma inmediata. Sin embargo, si su estrategia de recuperación de desastres para partes de sus datos tiene requisitos menos estrictos en cuanto a los tiempos de recuperación (de minutos a horas), el uso deAmazon S3 Glacier o Amazon S3 Glacier Deep Archive reducirá significativamente los costes de la estrategia de copia de seguridad y recuperación de los datos.

Algunas cargas de trabajo pueden tener requisitos regulatorios de residencia de datos. Si esto se aplica a la carga de trabajo en una localidad que actualmente solo tiene una región de AWS, además de diseñar cargas de trabajo Multi-AZ para una alta disponibilidad, como se mencionó anteriormente, también puede utilizar las AZ de esa región como ubicaciones discretas, lo que puede resultar útil para abordar los requisitos de residencia de datos aplicables a su carga de trabajo dentro de esa región. Las estrategias de recuperación de datos que se describen en las siguientes secciones utilizan varias regiones de AWS, pero también se pueden implementar en zonas de disponibilidad en lugar de regiones.

Varias regiones de AWS

En el caso de un desastre que incluya el riesgo de perder varios centros de datos con una distancia significativa entre sí, debería considerar las opciones de recuperación de desastres para mitigar los desastres naturales y técnicos que afecten a toda una región dentro de AWS. Todas las opciones descritas en las siguientes secciones se pueden implementar como arquitecturas de varias regiones para protegerse contra tales desastres.

Opciones de recuperación de desastres en la nube

Las estrategias de recuperación de desastres disponibles en AWS se pueden clasificar en términos generales en cuatro enfoques, que van desde un bajo coste y una complejidad menor al hacer copias de seguridad hasta estrategias más complejas que utilizan varias regiones activas. Es fundamental que pruebe de forma regular la estrategia de recuperación de desastres para que pueda aplicarla sin reparos, en caso de que sea necesario.

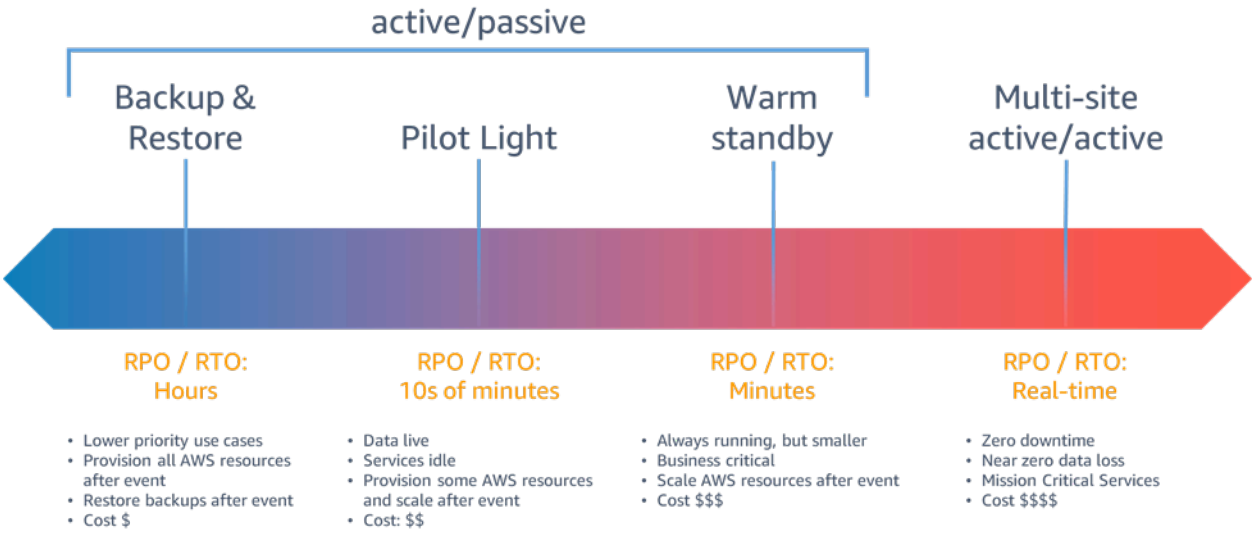

Ilustración 6: Estrategias de recuperación de desastres

En el caso de un desastre debido a la interrupción o pérdida de un centro de datos físico de una carga de trabajo de alta disponibilidad y buena arquitectura, es posible que solo necesite un enfoque de copia de seguridad y restauración para la recuperación de desastres. Si su definición de desastre va más allá de la interrupción o la pérdida de un centro de datos físico, es decir, más allá de una región o si está sujeto a requisitos regulatorios que lo exigen, debe considerar las opciones luz piloto, espera semiactiva y activa/activa en varios sitios.

Copia de seguridad y restauración

La copia de seguridad y la restauración son un enfoque adecuado para mitigar la pérdida o la corrupción de datos. Este enfoque también se puede utilizar para mitigar un desastre regional mediante la replicación de datos en otras regiones de AWS o para mitigar la falta de redundancia de las cargas de trabajo implementadas en una sola zona de disponibilidad. Además de los datos, debe volver a implementar la infraestructura, la configuración y el código de la aplicación en la región de recuperación. Para permitir que la infraestructura se vuelva a implementar rápidamente

sin errores, debe implementar siempre con la infraestructura como código (IaC) mediante servicios como AWS CloudFormation o el AWS Cloud Development Kit (AWS CDK). Sin IaC, puede resultar complejo restaurar las cargas de trabajo en la región de recuperación, lo que aumentará el tiempo de recuperación y posiblemente superará el RTO. Además de los datos del usuario, asegúrese de realizar también una copia de seguridad del código y de la configuración, incluidas las imágenes de máquina de Amazon (AMI) que utiliza para crear instancias de Amazon EC2. Puede utilizar AWS CodePipeline para automatizar la reimplementación del código y la configuración de la aplicación.

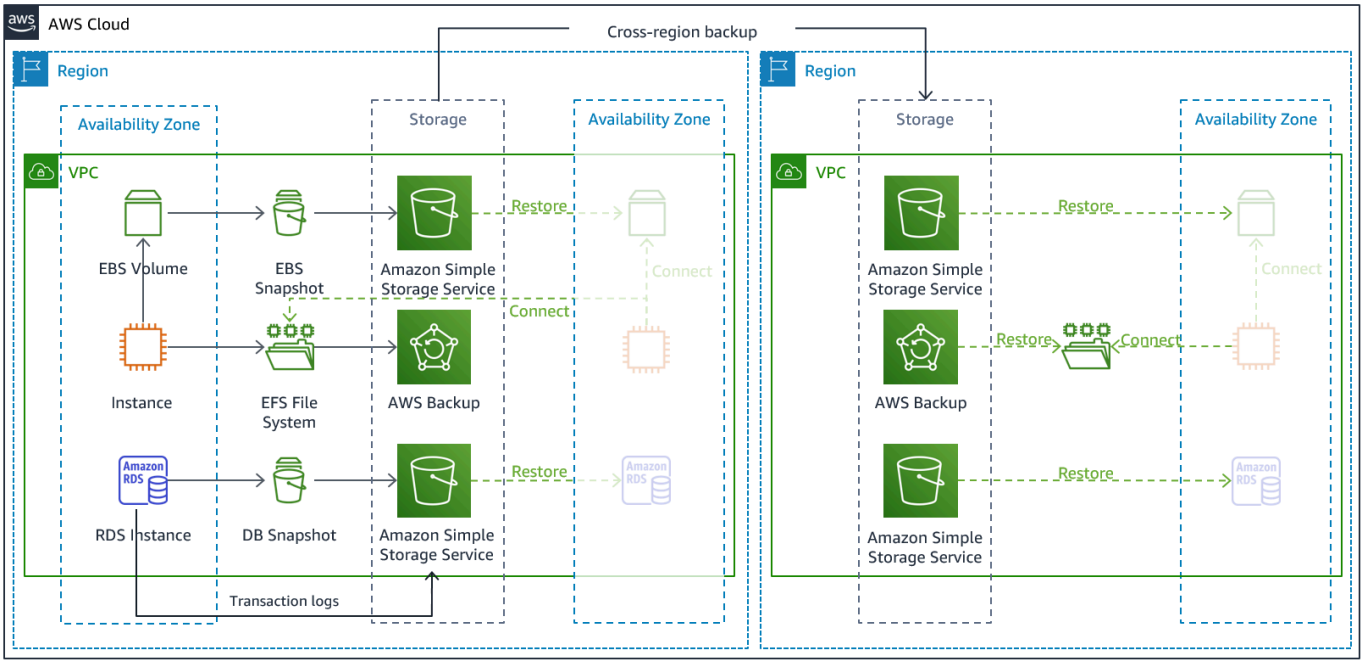

Ilustración 7: Arquitectura de copia de seguridad y restauración

Servicios de AWS

Los datos de la carga de trabajo requerirán una estrategia de copia de seguridad que se ejecute periódicamente o que sea continua. La frecuencia con la que ejecute la copia de seguridad determinará el punto de recuperación alcanzable (que debe alinearse para cumplir con su RPO). La copia de seguridad también debería ofrecer una forma de restaurarla al momento dado en el que se realizó. La copia de seguridad con recuperación a un momento dado está disponible a través de los siguientes servicios y recursos:

- Instantánea de Amazon Elastic Block Store (Amazon EBS)
- Copia de seguridad de Amazon DynamoDB
- Instantánea de Amazon RDS

- Instantánea de base de datos de Amazon Aurora
- Copia de seguridad de Amazon EFS (cuando se usa AWS Backup)
- Instantánea de Amazon Redshift
- Instantánea de Amazon Neptune

Para Amazon Simple Storage Service (Amazon S3), puede utilizar la replicación entre diferentes regiones (CRR) de Amazon S3 para copiar objetos de forma asíncrona en un bucket de S3 en la región de recuperación de desastres de forma continua, a la vez que proporciona el control de versiones de los objetos almacenados para que pueda elegir su punto de restauración. La ventaja de la replicación continua de datos es que las copias de seguridad de los datos son prácticamente inmediatas, pero es posible que no proteja contra desastres, como la corrupción o ataques maliciosos a los datos (por ejemplo, la eliminación no autorizada de datos), así como las copias de seguridad puntuales. La replicación continua se trata en la sección Luz piloto en los servicios de AWS.

AWS Backup proporciona una ubicación centralizada para configurar, programar y supervisar las capacidades de copia de seguridad de AWS para los siguientes servicios y recursos:

- Volúmenes de Amazon Elastic Block Store (Amazon EBS)
- Instancias de Amazon EC2
- Bases de datos de Amazon Relational Database Service (Amazon RDS) (incluidas las bases de datos de Amazon Aurora)
- Tablas de Amazon DynamoDB
- Sistemas de archivos de Amazon Elastic File System (Amazon EFS)
- Volúmenes de AWS Storage Gateway
- Amazon FSx for Windows File Server y Amazon FSx for Lustre

AWS Backup permite realizar copias de seguridad en todas las regiones, como, por ejemplo, en una región de recuperación de desastres.

Como estrategia de recuperación de desastres adicional para sus datos de Amazon S3, se recomienda que active el control de versiones de objetos de S3. El control de versiones de objetos protege sus datos en S3 de las consecuencias de las acciones de eliminación o modificación pues conserva la versión original anterior a la acción. El control de versiones de objetos puede ser una mitigación útil ante desastres provocados por errores humanos. Si utiliza la replicación de S3 para

realizar copias de seguridad de los datos en su región de recuperación de desastres, entonces, de forma predeterminada, cuando se elimina un objeto en el bucket de origen, Amazon S3 añade un marcador de eliminación solamente en el bucket de origen. Este enfoque protege los datos de la región de recuperación de desastres de eliminaciones maliciosas en la región de origen.

Además de los datos, también debe realizar una copia de seguridad de la configuración y la infraestructura necesarias para volver a implementar su carga de trabajo y cumplir con su objetivo de tiempo de recuperación (RTO). AWS CloudFormation proporciona infraestructura como código (IaC) y le permite definir todos los recursos de AWS en su carga de trabajo para que pueda implementar y volver a implementar de manera fiable en varias cuentas y regiones de AWS. Puede realizar copias de seguridad de las instancias de Amazon EC2 utilizadas por su carga de trabajo como imágenes de máquina de Amazon (AMI). La AMI se crea a partir de instantáneas del volumen raíz de la instancia y de cualquier otro volumen de EBS adjunto a la instancia. Puede usar esta AMI para lanzar una versión restaurada de la instancia de EC2. Una AMI se puede copiar dentro de las regiones o entre ellas. O bien, puede usar AWS Backup para copiar copias de seguridad en cuentas y en otras regiones de AWS. La capacidad de copia de seguridad entre cuentas ayuda a protegerse de los desastres, como las amenazas internas o la vulnerabilidad de las cuentas. AWS Backup también agrega capacidades adicionales para la copia de seguridad de EC2; además de los volúmenes de EBS individuales de la instancia, AWS Backup también almacena y rastrea los siguientes metadatos: tipo de instancia, nube privada virtual (VPC) configurada, grupo de seguridad, rol de IAM, configuración de supervisión y etiquetas. Sin embargo, estos metadatos adicionales solo se utilizan para restaurar la copia de seguridad de EC2 en la misma región de AWS.

Todos los datos almacenados en la región de recuperación de desastres en forma de copias de seguridad deben restaurarse en el momento de la conmutación por error. AWS Backup ofrece capacidad de restauración, pero actualmente no permite programar ni automatizar la restauración. Puede implementar la restauración automática en la región de recuperación de desastres mediante el SDK de AWS para llamar a las API AWS Backup. Puede configurarla como un trabajo recurrente o desencadenar la restauración cada vez que se complete una copia de seguridad. En la siguiente figura se muestra un ejemplo de restauración automática con Amazon Simple Notification Service (Amazon SNS) y AWS Lambda.

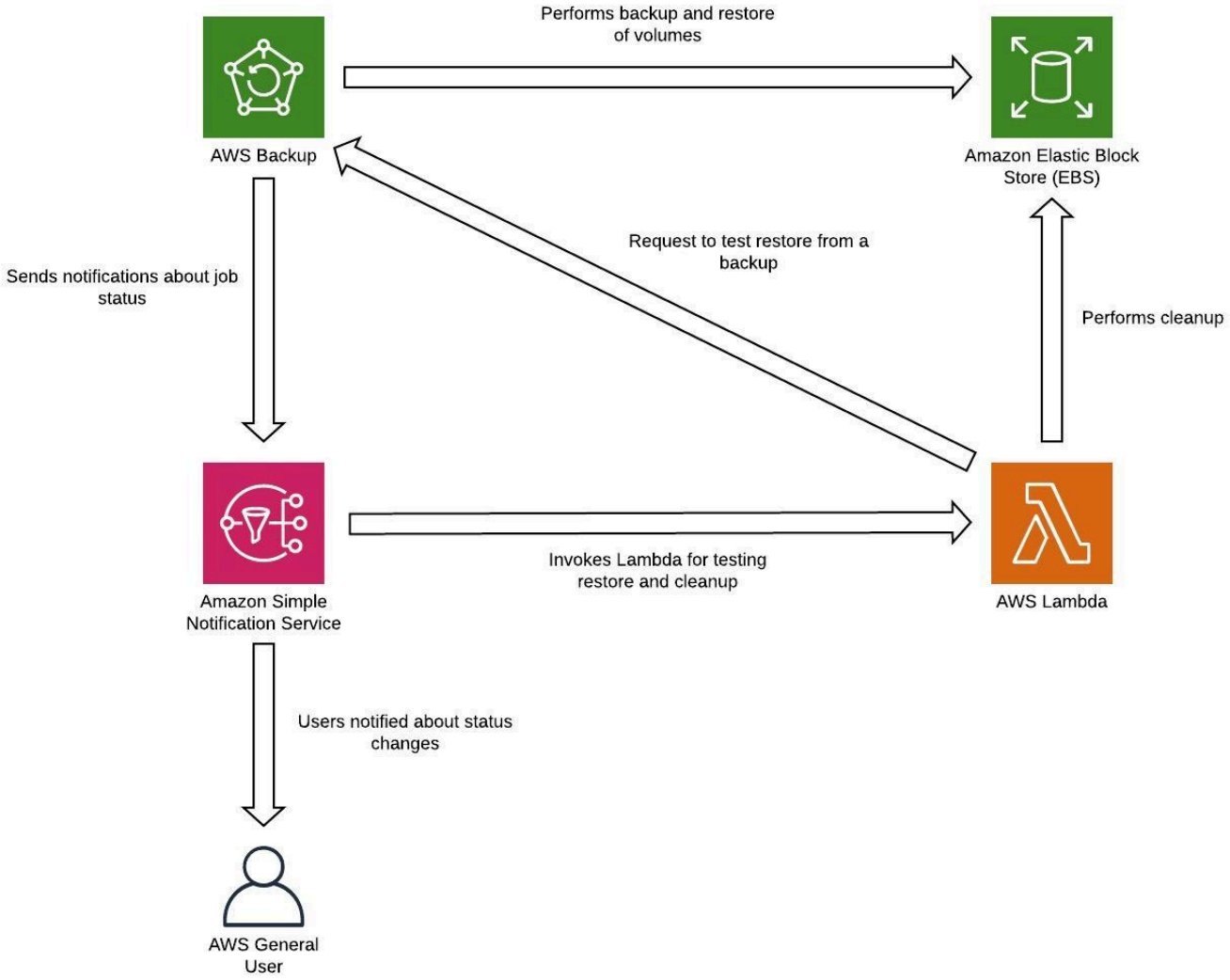

Ilustración 8: Restauración y pruebas de copias de seguridad

> **Note**
>
> Su estrategia de copia de seguridad debe incluir probar las copias de seguridad. Consulte la sección Probar la recuperación de desastres para obtener más información. Consulte el laboratorio AWS Well-Architected Lab: Testing Backup and Restore of Data para obtener una demostración práctica sobre las pruebas de copia de seguridad y restauración.

Luz piloto

Con el enfoque de luz piloto se replican los datos de una región a otra y se almacena una copia de la infraestructura de carga de trabajo principal. Los recursos necesarios para poder replicar y hacer una copia de seguridad de los datos, como las bases de datos y el almacenamiento de objetos, siempre están disponibles. Otros elementos, como los servidores de aplicaciones, se cargan con código de aplicación y configuraciones, pero se apagan y solo se usan durante las pruebas o cuando se invoca la conmutación por error de recuperación de desastres. A diferencia del enfoque de copia de seguridad y restauración, la infraestructura principal siempre está disponible y siempre tiene la opción de aprovisionar rápidamente un entorno de producción a gran escala al encender y escalar los servidores de aplicaciones horizontalmente.

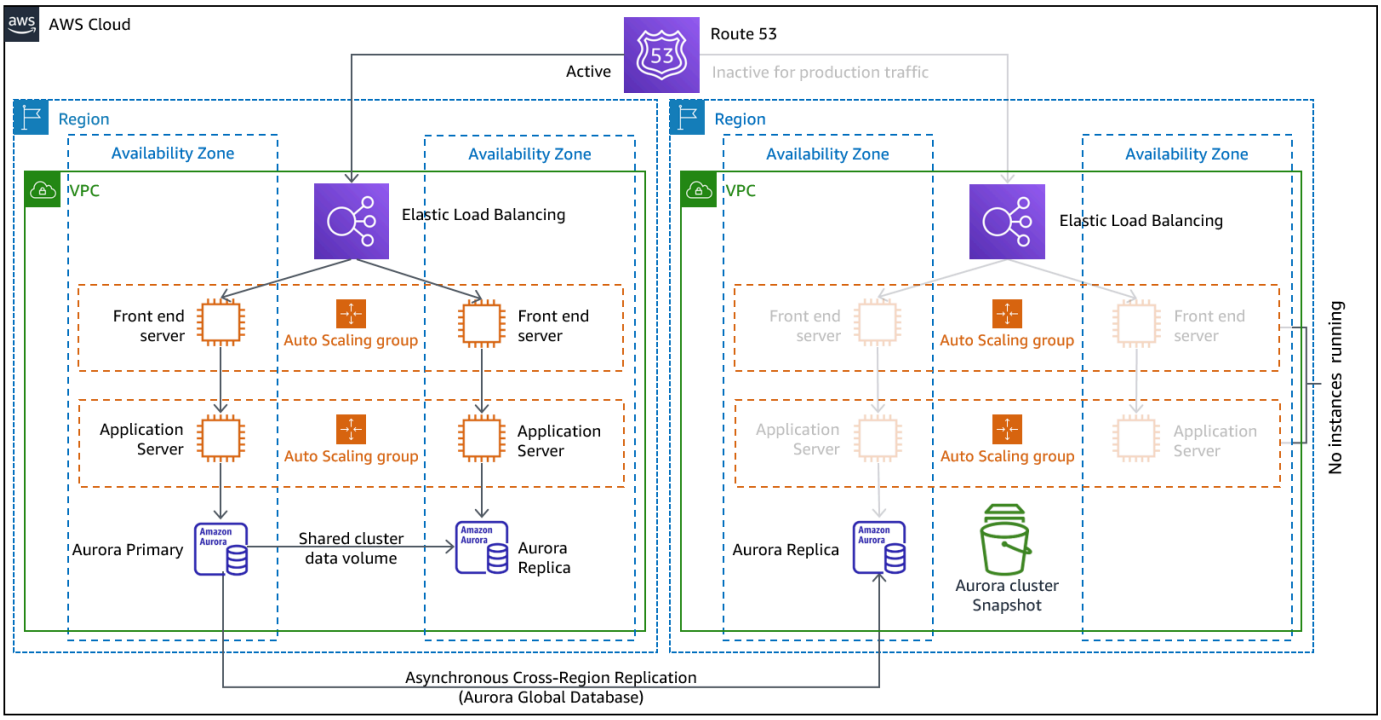

Ilustración 9: Arquitectura de luz piloto

El enfoque de luz piloto minimiza el coste continuo de la recuperación de desastres pues minimiza los recursos activos y simplifica la recuperación en el momento de un desastre porque todos los requisitos de la infraestructura central están en su lugar. Esta opción de recuperación requiere cambiar el enfoque de la implementación. Debe realizar cambios en la infraestructura central en cada región e implementar cambios en la carga de trabajo (configuración, código) simultáneamente en cada región. Este paso se puede simplificar automatizando las implementaciones y utilizando la infraestructura como código (IaC) para implementar la infraestructura en varias cuentas y

regiones (implementación completa de la infraestructura en la región principal e implementación de infraestructura reducida verticalmente/desconectada en regiones de recuperación de desastres). Se recomienda usar una cuenta diferente por región para proporcionar el nivel más alto de aislamiento de recursos y seguridad (en el caso de que las credenciales comprometidas también formen parte de los planes de recuperación de desastres).

Con este enfoque, también debe mitigar ante desastres relacionados con datos. La replicación continua de datos lo protege contra algunos tipos de desastres, pero es posible que no lo proteja contra la corrupción o la destrucción de datos, a menos que su estrategia también incluya el control de versiones de los datos almacenados u opciones para la recuperación a un momento dado. Puede hacer una copia de seguridad de los datos replicados en la región del desastre para crear copias de seguridad puntuales en esa misma región.

Servicios de AWS

Además de utilizar los servicios de AWS que se tratan en la sección Copia de seguridad y restauración para crear copias de seguridad puntuales, puede considerar también los siguientes servicios para la estrategia de luz piloto.

En esta estrategia, la replicación continua de datos en bases de datos en vivo y almacenes de datos en la región de recuperación de datos es el mejor enfoque para un RPO bajo (cuando se usa además de las copias de seguridad puntuales que hemos comentado anteriormente). AWS proporciona replicación de datos asíncrona, continua y entre regiones mediante los siguientes servicios y recursos:

- Replicación de Amazon Simple Storage Service (Amazon S3)
- Réplicas de lectura de Amazon RDS
- Amazon Aurora Global Databases
- Tablas globales de Amazon DynamoDB

Con la replicación continua, las versiones de sus datos están disponibles casi de inmediato en su región de recuperación de datos. Los tiempos de replicación reales se pueden supervisar mediante funciones de servicio como el control de tiempo de replicación de S3 (S3 RTC) para objetos de S3 y funciones de administración de Amazon Aurora Global Databases.

Si no logra ejecutar una carga de trabajo de lectura y escritura desde la región de recuperación de desastres, debe promover una réplica de lectura de RDS para que se convierta en la instancia principal. Para las instancias de base de datos distintas de Aurora, el proceso tarda unos minutos

en completarse y el reinicio forma parte del proceso. Para la replicación entre regiones (CRR) y la conmutación por error en RDS, el uso de la Amazon Aurora Global Databases proporciona varias ventajas. La base de datos global utiliza una infraestructura dedicada que deja sus bases de datos totalmente disponibles para servir a la aplicación. Además, puede replicarse en la región secundaria con una latencia típica de menos de un segundo (y dentro de una región de AWS es mucho menos de 100 milisegundos). Con Amazon Aurora Global Databases, si su región principal sufre una degradación o interrupción del rendimiento, puede promover una de las regiones secundarias para que asuma responsabilidades de lectura y escritura en menos de 1 minuto, incluso en el caso de una interrupción regional completa. La promoción puede ser automática y no se reinicia.

Se debe implementar una versión reducida verticalmente de la infraestructura de carga de trabajo principal con menos recursos o recursos más pequeños en su región de recuperación de desastres. AWS CloudFormation le permite definir la infraestructura e implementarla de manera coherente en las cuentas de AWS y en las regiones de AWS. AWS CloudFormation utiliza pseudoparámetros predefinidos para identificar la cuenta de AWS y la región de AWS en la que se implementa. Por lo tanto, puede implementar la lógica de condición en sus plantillas de CloudFormation para implementar solo la versión reducida verticalmente de su infraestructura en la región de recuperación de desastres. Para las implementaciones de instancias de EC2, una imagen de máquina de Amazon (AMI) proporciona información de la configuración del hardware y del software instalado. Puede implementar una canalización de Image Builder que cree las AMI que necesita y copiarlas tanto en su región principal como en la de copia de seguridad. Esto ayuda a garantizar que estas AMI doradas tengan todo lo que necesita para volver a implementar o escalar horizontalmente su carga de trabajo en una nueva región, en caso de desastre. Las instancias de Amazon EC2 se implementan en una configuración reducida verticalmente (menos instancias que en la región principal). Puede hibernar para detener las instancias de EC2, lo que significa que no paga gastos de EC2, solo paga por el almacenamiento utilizado. Para iniciar instancias de EC2, puede crear scripts mediante la interfaz de línea de comandos (CLI) o el SDK de AWS. Para escalar horizontalmente la infraestructura para admitir el tráfico de producción, consulte AWS Auto Scaling la sección Espera semiactiva.

Para una configuración activa/en espera, como luz piloto, todo el tráfico va inicialmente a la región principal y cambia a la región de recuperación de desastres si la región principal ya no está disponible. Al utilizar los servicios de AWS hay que tener en cuenta dos opciones de administración del tráfico. La primera opción es usar Amazon Route 53. Con Amazon Route 53, puede asociar varios puntos de conexión de IP en una o más regiones de AWS con un nombre de dominio de Route 53. A continuación, puede dirigir el tráfico al punto de conexión apropiado de ese nombre de dominio. Las comprobaciones de estado de Amazon Route 53 supervisan estos puntos de conexión.

Estas comprobaciones de estado le permiten configurar la conmutación por error de DNS para garantizar que el tráfico se envíe a puntos de conexión en buen estado.

La segunda opción es usar AWS Global Accelerator. Con AnyCast IP, puede asociar varios puntos de conexión a una o más regiones de AWS con las mismas direcciones IP estáticas. A continuación, AWS Global Accelerator dirige el tráfico al punto de conexión apropiado asociado a esa dirección. Las comprobaciones de estado de Global Accelerator supervisan los puntos de conexión. En función de estas comprobaciones de estado, AWS Global Accelerator comprueba automáticamente el estado de las aplicaciones y dirige el tráfico de los usuarios a un solo punto de conexión de la aplicación en buen estado. Global Accelerator ofrece latencias más bajas en el punto de conexión de la aplicación, ya que utiliza la extensa red de borde de AWS para colocar el tráfico en la red troncal de AWS lo antes posible. Global Accelerator también evita los problemas de almacenamiento en caché que pueden darse con los sistemas DNS (como Route 53).

CloudEndure Disaster Recovery

CloudEndure Disaster Recovery, disponible enAWS Marketplace, replica continuamente las aplicaciones alojadas en el servidor y las bases de datos alojadas en el servidor desde cualquier fuente en AWS mediante la replicación en el nivel de bloques del servidor subyacente. CloudEndure Disaster Recovery permite utilizar la nube de AWS como región de recuperación de desastres para una carga de trabajo local y su entorno. También se puede utilizar para la recuperación de desastres de cargas de trabajo alojadas en AWS si solo consisten en aplicaciones y bases de datos alojadas en EC2 (es decir, no en RDS). CloudEndure Disaster Recovery utiliza la estrategia de luz piloto, que conserva una copia de los datos y los recursos apagados en una Amazon Virtual Private Cloud (Amazon VPC) que se utiliza como área de almacenamiento provisional. Cuando se desencadena un evento de conmutación por error, los recursos por etapas se utilizan para crear automáticamente una implementación de capacidad completa en la Amazon VPC de destino que es la ubicación de recuperación.

Ilustración 10: Arquitectura de CloudEndure Disaster Recovery

Espera semiactiva

El enfoque de espera semiactiva implica garantizar que haya una reducción vertical, pero totalmente funcional, del entorno de producción en otra región. Este enfoque amplía el concepto de luz piloto y reduce el tiempo de recuperación porque la carga de trabajo siempre está activa en otra región. Este

enfoque también permite realizar pruebas más fácilmente o implementar pruebas continuas para aumentar la confianza en la capacidad para recuperarse de un desastre.

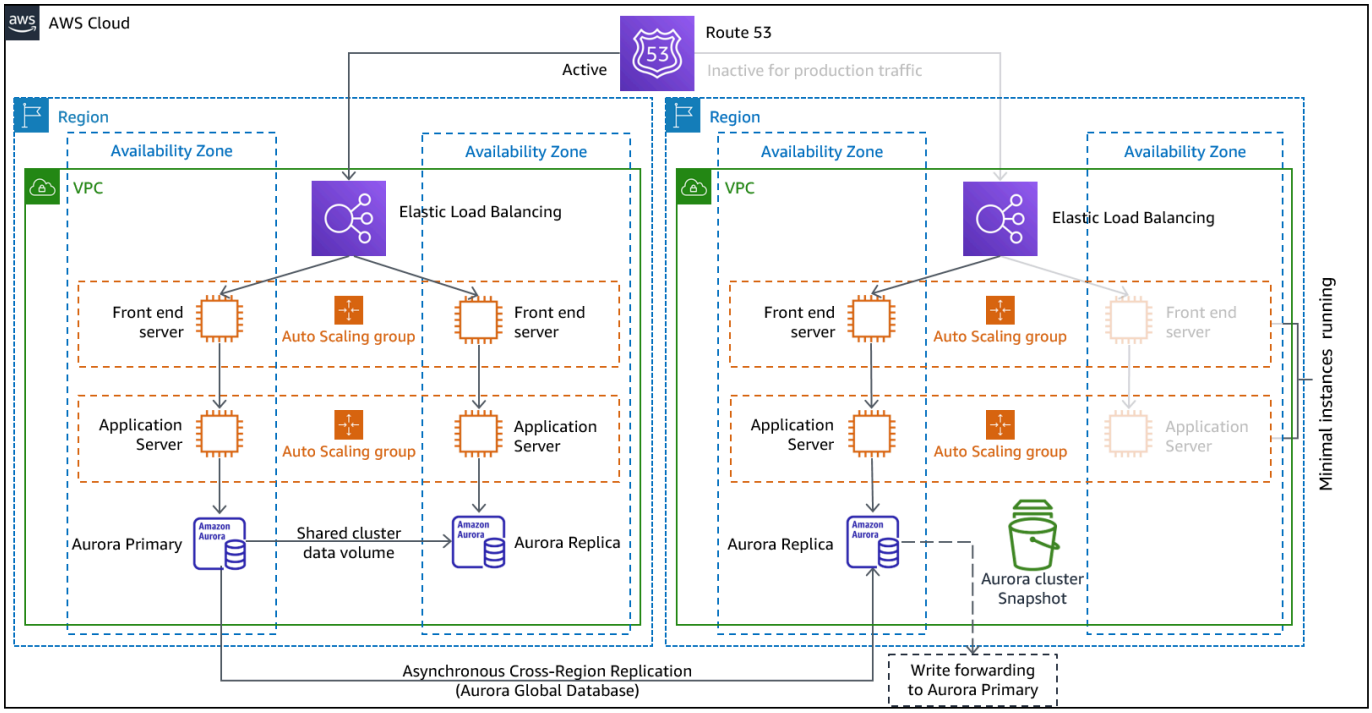

Ilustración 11: Arquitectura de espera semiactiva

Nota: La diferencia entre la estrategia de luz piloto y deespera semiactiva a veces puede resultar difícil de entender. Ambas incluyen un entorno en su región de recuperación de desastres con copias de los activos principales de su región. La diferencia es que la estrategia de luz piloto no puede procesar solicitudes sin que se tomen medidas adicionales primero, mientras que la espera semiactiva puede gestionar el tráfico de inmediato (en niveles de capacidad reducidos). El enfoque de luz piloto requiere que encienda los servidores, posiblemente implementando una infraestructura adicional (no central) y escalando verticalmente, mientras que el enfoque de espera semiactiva solo requiere que se escale verticalmente (todo ya está implementado y en ejecución). Piense en sus necesidades de RTO y RPO a la hora de elegir entre estos enfoques.

Servicios de AWS

Todos los servicios de AWS cubiertos por copia de seguridad y restauración y luz piloto también se utilizan en la espera semiactiva para la copia de seguridad de datos, la replicación de datos, el enrutamiento de tráfico activo/en espera y la implementación de infraestructura, incluidas las instancias de EC2.

AWS Auto Scaling se utiliza para escalar recursos, incluidas las instancias de Amazon EC2, las tareas de Amazon ECS, el rendimiento de Amazon DynamoDB y las réplicas de Amazon Aurora en una región de AWS. Amazon EC2 Auto Scaling escala la implementación de la instancia de EC2 en las zonas de disponibilidad dentro de una región de AWS, lo que proporciona resiliencia en esa región. Utilice Auto Scaling para escalar horizontalmente su región de recuperación de desastres hasta una capacidad de producción completa, como parte de las estrategias de luz piloto o espera semiactiva. Por ejemplo, para EC2, debe aumentar la configuración de capacidad deseada en el grupo de Auto Scaling. Puede ajustar esta configuración manualmente a través de AWS Management Console, automáticamente a través del SDK de AWS o mediante la reimplementación de la plantilla de AWS CloudFormation con el nuevo valor de capacidad deseado. Puede utilizar los parámetros de AWS CloudFormation para facilitar la reimplementación de la plantilla de CloudFormation. Asegúrese de que las Service Quotas de su región de recuperación de desastres sean lo suficientemente altas como para no limitar el escalado vertical hacia la capacidad de producción.

Activa/activa en varios sitios

Puede ejecutar la carga de trabajo simultáneamente en varias regiones como parte de una estrategia activa/activa en varios sitios o espera activa/pasiva. La estrategia activa/activa en varios sirve el tráfico desde todas las regiones en las que se implementa, mientras que la estrategia de espera activa solo sirve el tráfico de una sola región. Las demás regiones solo se utilizan para la recuperación de desastres. Con una estrategia activa/activa en varios sitios, los usuarios pueden acceder a la carga de trabajo en cualquiera de las regiones en las que se implemente. Este enfoque es el más complejo y caro para la recuperación de desastres, pero puede reducir el tiempo de recuperación a casi cero en la mayoría de desastres con las opciones de tecnología e implementación correctas (sin embargo, la corrupción de datos puede necesitar depender de copias de seguridad, lo que generalmente resulta en un punto de recuperación distinto de cero). El modo de espera activa utiliza una configuración activa/pasiva en la que los usuarios se dirigen a una sola región y las regiones de recuperación de desastres no aceptan el tráfico. La mayoría de los clientes considera que si van a tener un entorno completo en la segunda región, tiene sentido usar el modo activo/activo. Alternativamente, si no desea utilizar ambas regiones para gestionar el tráfico de usuarios, la opción de espera semiactiva es más económica y menos compleja desde el punto de vista operativo.

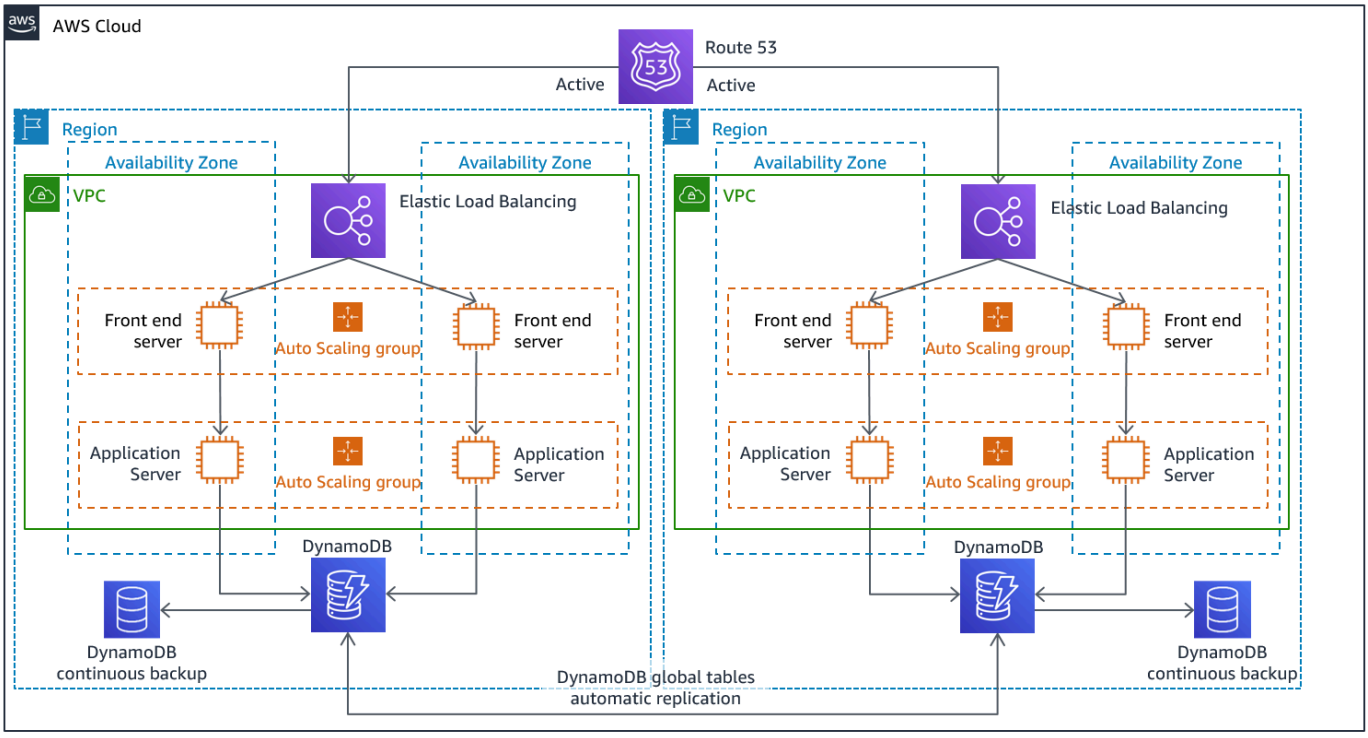

Ilustración 12: Arquitectura activa/activa en varios sitios (cambie una ruta activa por inactiva para el modo de espera activa)

En la estrategia activa/activa en varios sitios, dado que la carga de trabajo se ejecuta en más de una región, no existe la conmutación por error. En este caso, las pruebas de recuperación de desastres se centrarían en cómo reacciona la carga de trabajo ante la pérdida de una región: ¿el tráfico se dirige fuera de la región que ha fallado? ¿Las otras regiones pueden gestionar todo el tráfico? También deben realizarse pruebas para detectar un desastre de datos. La copia de seguridad y la recuperación siguen siendo necesarias y deben probarse con regularidad. También se debe tener en cuenta que los tiempos de recuperación de un desastre de datos que implique corrupción, eliminación u ofuscación de datos siempre serán mayores que cero y el punto de recuperación siempre estará en algún momento antes de que se descubriera el desastre. Si se requiere la complejidad y el coste adicionales de una estrategia activa/activa en varios sitios (o espera activa) para mantener tiempos de recuperación casi iguales a cero, entonces se deben realizar esfuerzos adicionales para mantener la seguridad y evitar errores humanos para mitigar los desastres humanos.

Servicios de AWS

Todos los servicios de AWS cubiertos en copia de seguridad y restauración, luz piloto y espera semiactiva también se utilizan aquí para copias de seguridad de datos de un período de tiempo concreto, replicación de datos, enrutamiento de tráfico activo/activo e implementación y escalado de infraestructura, incluidas las instancias de EC2.

Para los escenarios activos/pasivos analizados anteriormente (luz piloto y espera activa), tanto Amazon Route 53 como AWS Global Accelerator se pueden usar para dirigir el tráfico de red hacia la región activa. Para la estrategia activa/activa aquí, ambos servicios también permiten la definición de políticas que determinen qué usuarios van a qué punto de conexión regional activo. AWS Global Accelerator le permite establecer una señalización del tráfico para controlar el porcentaje de tráfico que se dirige a cada punto de conexión de la aplicación. Amazon Route 53 admite este enfoque porcentual y también muchas otras políticas disponibles, incluidas las basadas en geoproximidad y latencia. Global Accelerator aprovecha automáticamente la extensa red de servidores de borde de AWS para incorporar el tráfico a la red troncal de AWS lo antes posible, lo que reduce la latencia de las solicitudes.

La replicación de datos con esta estrategia permite un RPO casi igual a cero. Los servicios de AWS, así como Amazon Aurora Global Databases, usan una infraestructura dedicada que deja sus bases de datos completamente disponibles para servir a la aplicación, y puede replicarse en una región secundaria con una latencia típica de menos de un segundo. Con las estrategias activa/pasiva, las escrituras se producen solo en la región principal. La diferencia con activa/activa es la forma en que se diseña la gestión de las escrituras en cada región activa. En general, las lecturas de los usuarios se diseñan para que se les sirva desde la región más cercana, lo que se conoce como lectura local. Con las escrituras, existen varias opciones:

- Una estrategia de escritura global dirige todas las escrituras a una sola región. En caso de que falle la región, se promovería a otra región para que acepte escrituras. La base de datos global de Aurora es adecuada para la escritura global, ya que admite la sincronización con réplicas de lectura en todas las regiones, y puede promover una de las regiones secundarias para que asuma responsabilidades de lectura y escritura en menos de 1 minuto.
- Una estrategia de escritura local dirige las escrituras a la región más cercana (de igual modo que las lecturas). Las tablas globales de Amazon DynamoDB admiten una estrategia de este tipo, lo que permite la lectura y escritura desde todas las regiones en las que se implemente la tabla global. Las tablas globales de Amazon DynamoDB usan una reconciliación del tipo prevalece el último escritor entre las actualizaciones simultáneas.

- Una estrategia de partición de escritura asigna escrituras a una región específica en función de una clave de partición (como el ID de usuario) para evitar conflictos de escritura. En este caso se puede utilizar la replicación de Amazon S3 configurada bidireccionalmente. Actualmente admite la replicación entre dos regiones. Al implementar este enfoque, asegúrese de habilitar la sincronización de modificación de réplicas en los buckets A y B para replicar los cambios de metadatos de réplica, como las listas de control de acceso a objetos (ACL), las etiquetas de objetos o los bloqueos de objetos en los objetos replicados. También puede configurar si desea o no replicar marcadores de eliminación entre buckets en sus regiones activas. Además de la replicación, su estrategia también debe incluir copias de seguridad en un momento dado para proteger contra eventos de corrupción o destrucción de datos.

AWS CloudFormation es una poderosa herramienta que permite aplicar una infraestructura implementada de manera consistente en las cuentas de AWS en varias regiones de AWS. AWS CloudFormation StackSets amplía esta funcionalidad al permitirle crear, actualizar o eliminar pilas de CloudFormation en varias cuentas y regiones con una sola operación. Aunque AWS CloudFormation utiliza YAML o JSON para definir la infraestructura como código, AWS Cloud Development Kit (AWS CDK) le permite definir la infraestructura como código utilizando lenguajes de programación familiares. El código se convierte en CloudFormation, que luego se utiliza para implementar recursos en AWS.

Detección

Es importante saber lo antes posible que las cargas de trabajo no están dando los resultados empresariales que se espera. De esta manera, podrá declarar rápidamente un desastre y recuperarse de un incidente. En el caso de objetivos de recuperación más estrictos, este tiempo de respuesta junto con la información apropiada es fundamental para cumplir con los objetivos de recuperación. Si su objetivo de punto de recuperación es de una hora, debe detectar el incidente, notificar al personal apropiado, iniciar los procesos de derivación, evaluar la información (si la tiene) sobre el tiempo previsto de recuperación (sin ejecutar el plan de recuperación de desastres), declarar un desastre y recuperarse en una hora.

> **Note**
>
> Si las partes interesadas deciden no invocar la recuperación de desastres (DR) a pesar de que el RTO puede estar en riesgo, entonces debería reevaluar los planes y objetivos de recuperación de desastres. La decisión de no invocar los planes de DR puede deberse a que los planes son inadecuados o a que hay una falta de confianza en la ejecución.

Es fundamental tener en cuenta la detección, notificación, derivación, descubrimiento y declaración de incidentes en la planificación y los objetivos para, de este modo, establecer objetivos realistas y alcanzables que proporcionen valor empresarial.

AWS publica la información más actualizada sobre la disponibilidad del servicio en el Service Health Dashboard (Panel de estado del servicio). Puede consultarlo en cualquier momento para obtener información de estado actual o suscribirse a una fuente RSS para recibir notificaciones sobre interrupciones en cada servicio. Si tiene un problema operativo en tiempo real con uno de nuestros servicios que no aparece en el Service Health Dashboard, puede crear una solicitud de soporte.

El AWS Health Dashboard proporciona información sobre los eventos de AWS Health que pueden afectar a su cuenta. La información se presenta de dos formas: en un panel donde se muestran los eventos recientes y próximos organizados por categorías, y en un log de eventos que contiene todos los eventos de los últimos 90 días.

Para los requisitos de RTO más estrictos, puede implementar una conmutación por error automatizada basada en comprobaciones de estado. Diseñe comprobaciones de estado que representen la experiencia del usuario y se basen en indicadores clave de rendimiento. Las

comprobaciones de estado profundas son claves para la carga de trabajo y van más allá de las comprobaciones de estado superficiales. Utilice comprobaciones de estado profundas basadas en múltiples señales. Tenga cuidado con este enfoque pues puede activar falsas alarmas porque una conmutación por error no necesaria puede incurrir en riesgos de disponibilidad.

Probar la recuperación de desastres

Pruebe la implementación de la recuperación de desastres para validar la implementación y pruebe regularmente la conmutación por error en la región de recuperación de desastres de su carga de trabajo para garantizar que se cumplan los RTO y el RPO.

Un patrón que debe evitarse es el desarrollo de rutas de recuperación que se ejecuten pocas veces. Por ejemplo, puede tener un almacén de datos secundario que se utilice para consultas de solo lectura. Cuando escribe en un almacén de datos y el almacén principal falla, es posible que quiera conmutar por error al almacén de datos secundario. Si no se prueba frecuentemente esta conmutación por error, es posible que sus suposiciones sobre las capacidades del almacén de datos secundario sean incorrectas. Es posible que la capacidad del almacén secundario, que podría haber sido suficiente la última vez que la probó, ya no pueda tolerar la carga en este escenario o las cuotas de servicio en la región secundaria no sean suficientes.

Nuestra experiencia ha demostrado que la única forma de recuperación de errores que funciona es aquella que prueba constantemente. Por ello, es mejor tener un número reducido de rutas de recuperación.

Puede establecer patrones de recuperación y probarlos con frecuencia. Si tiene una ruta de recuperación compleja o crítica, debería ejecutar ese error en producción regularmente para asegurarse de que la ruta funciona correctamente.

Administre el cambio de configuración en la región de recuperación de desastres. Asegúrese de que la infraestructura, los datos y la configuración sean los necesarios en la región de recuperación de desastres. Por ejemplo, puede comprobar que las AMI y las cuotas de servicio estén actualizadas.

Puede utilizar AWS Config para supervisar continuamente y registrar sus configuraciones de recursos de AWS. AWS Config puede detectar desviaciones y activar AWSSystems Manager Automation para corregir la deriva y generar alarmas. AWS CloudFormation también puede detectar desviaciones en las pilas que haya implementado.

Conclusión

Los clientes son responsables de que las aplicaciones estén disponibles en la nube. Es importante definir el concepto de desastre y tener un plan de recuperación de desastres que refleje dicha definición y el impacto que pueda tener en los resultados de la empresa. Cree un objetivo de tiempo de recuperación (RTO) y un objetivo de punto de recuperación (RPO) que se basen en el análisis del impacto y las evaluaciones de riesgos. A continuación, elija la arquitectura adecuada para mitigar los desastres. Asegúrese de que se puedan detectar los desastres y que se haga a tiempo; es de suma importancia saber cuándo están en riesgo los objetivos. Asegúrese de tener un plan y valide el plan a base de pruebas. Los planes de recuperación de desastres que no se han validado corren el riesgo de no implementarse debido a la falta de confianza o al incumplimiento de los objetivos de recuperación de desastres.

Colaboradores

Han colaborado en la elaboración de este documento:

- Alex Livingstone, responsable principal de operaciones en la nube, AWS Enterprise Support
- Seth Eliot, arquitecto principal de soluciones de fiabilidad, Amazon Web Services

Documentación adicional

Para obtener información adicional, consulte:

- Pilar de fiabilidad, AWS Well-Architected Framework
- Lista de comprobación del plan de recuperación de desastres
- Implementación de las comprobaciones de estado
- Implementaciones de soluciones de AWS: Arquitectura de aplicación en múltiples regiones
- AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications (ARC209-R2)

Historial de revisión

Cambiar	Descripción	Fecha
Publicación inicial	Primera publicación.	12 de febrero de 2021

Para recibir notificaciones sobre las actualizaciones de este documento técnico, suscríbase a la fuente RSS.

Avisos

Los clientes son responsables de realizar sus propias evaluaciones de la información contenida en este documento. Este documento: (a) solo tiene fines informativos, (b) representa las prácticas y las ofertas de productos vigentes de AWS, que están sujetas a cambios sin previo aviso, y (c) no crea ningún compromiso ni garantía de AWS y sus empresas afiliadas, proveedores o concesionarios de licencias. Los productos o servicios de AWS se proporcionan “tal cual”, sin garantías, representaciones ni condiciones de ningún tipo, ya sean explícitas o implícitas. Las responsabilidades y obligaciones de AWS en relación con sus clientes se rigen por los acuerdos de AWS, y este documento no modifica ni forma parte de ningún acuerdo entre AWS y sus clientes.