Las opciones de conectividad de red están disponibles AWS para las ofertas de SaaS

AWS Guía prescriptiva

AWS Guía prescriptiva: Las opciones de conectividad de red están disponibles AWS para las ofertas de SaaS

Table of Contents

Las opciones de conectividad de red están disponibles AWS para las ofertas de SaaS

Tomas Sykora y Luca Schumann, Amazon Web Services

Septiembre de 2025 (historia del documento)

Esta guía explora escenarios comunes para conectar aplicaciones de consumo con proveedores de software como servicio (SaaS). Describe cómo conectarse a recursos que se encuentran en las instalaciones, en las Nube de AWS nubes de otros proveedores de servicios en la nube (CSP) o en arquitecturas híbridas. Entre estos escenarios se incluyen los siguientes:

- Exponer los servicios web a través de HTTPS
- Exponer los servicios basados en TCP
- Uso AWS AppSyncpara implementar publish-subscribe (Pub/Sub) y GraphQL APIs
- Uso de AWS recursos para exponer aplicaciones en tiempo real WebSockets
- Permite el acceso bidireccional para la comunicación interactiva de los servicios

Al alinearse con las mejores prácticas incluidas en esta guía, los proveedores de SaaS pueden impulsar la confianza de los clientes y respaldar el acceso escalable, seguro y flexible a las ofertas de SaaS.

Esta guía también incluye criterios de autoevaluación para ayudarlo a evaluar con qué éxito cumple con los requisitos de redes de consumo para su oferta de SaaS. Además de los patrones de conectividad, encontrará comparaciones exhaustivas de los servicios de AWS red, diagramas de arquitectura de alto nivel para diversos escenarios de implementación y orientación práctica sobre cómo seleccionar el enfoque adecuado en función de su contexto empresarial específico. La guía analiza las consideraciones de seguridad de cada opción de red, analiza los errores más comunes que se deben evitar y proporciona recomendaciones de implementación que equilibran los requisitos técnicos con la eficiencia operativa. Además, encontrará marcos estratégicos para alinear sus decisiones de red con su modelo de negocio, sus objetivos de crecimiento y sus necesidades de conformidad normativa.

Destinatarios previstos

Esta guía está destinada a los proveedores de SaaS. Ayuda a los arquitectos de la nube, los gerentes de productos y los ingenieros de redes que diseñan, implementan y optimizan la conectividad de red para las ofertas de SaaS en el. Nube de AWS Para comprender los conceptos y las recomendaciones de esta guía, debe estar familiarizado con AWS los fundamentos, los conceptos básicos de SaaS y los principios de redes de alto nivel.

Objetivos

Esta guía analiza las opciones de arquitectura de red y las mejores prácticas probadas en campo que ayudan a los consumidores a optimizar el acceso a las ofertas de SaaS. La implementación de las recomendaciones de esta guía respalda lo siguiente:

- Facilidad de integración: ofrezca a sus clientes un recorrido sencillo, desde la incorporación hasta la producción, de modo que pueda acelerar el tiempo de creación de valor de sus clientes y acortar su ciclo de reconocimiento de ingresos.
- Adaptabilidad: intégrese sin problemas con las infraestructuras de red existentes de sus clientes adaptándose a sus necesidades cambiantes. Esto mejora la propuesta de valor de su producto.
- Costo total de propiedad: estandarice el acceso a la red para reducir los costos de cambio y los costos por inquilino. Al mejorar la coherencia de la implementación, también puede reducir el tiempo necesario para realizar el análisis o la reparación de la causa raíz.
- Administración de dependencias: comprenda las dependencias, las implicaciones a largo plazo y las ventajas y desventajas de las distintas opciones de acceso a la red. Esto ayuda a los líderes de producto a tomar decisiones bien informadas sobre los productos.
- Capacidad de composición y ampliabilidad: separe el desarrollo de la funcionalidad principal de la infraestructura operativa. Esto ayuda a los equipos de desarrollo a avanzar más rápido y a centrarse en crear valor para sus clientes.
- Fomente la confianza: al proporcionar un acceso resiliente, tolerante a errores, seguro y escalable a las ofertas de SaaS, puede reducir los riesgos regulatorios y ganarse la confianza en su capacidad para respaldar el crecimiento de sus clientes.

Evaluación de las decisiones de acceso a la red para las ofertas de SaaS

Comprenda su mercado

Las decisiones que tome ahora sobre las redes determinarán si la propuesta de valor de su producto SaaS puede entregarse a sus clientes. A pesar de la importancia estratégica de estas decisiones, proporcionar acceso a su oferta de SaaS suele percibirse como un tema puramente tecnológico. El riesgo que conlleva esta percepción incluye ciclos prolongados de reconocimiento de ingresos, ineficiencias operativas y desajustes con la estrategia empresarial. Por ejemplo, si la expansión rápida es un objetivo empresarial estratégico, una guía para su proceso de toma de decisiones debería ser si las soluciones que está considerando son lo suficientemente escalables y flexibles como para respaldar la expansión. Incluso si logra hacer crecer su negocio, los gastos operativos no deben convertirse en un obstáculo para el crecimiento futuro, y una estructura de costos desalineada podría consumir todos sus beneficios.

Por ejemplo, considere cómo afectan las siguientes consideraciones de mercado a los aspectos técnicos del producto, como las redes:

- Si su modelo de negocio se basa en suscripciones, es probable que sus clientes prefieran soluciones con costes predecibles y recurrentes en lugar de grandes inversiones iniciales.
- Si su estrategia empresarial se dirige a clientes de alto valor a nivel empresarial, los criterios de seguridad, gobierno y cumplimiento normativo determinan si se tendrá en cuenta su oferta de SaaS.
- Si su mercado objetivo está compuesto principalmente por empresas emergentes, es probable que la facilidad de integración, la rentabilidad y la adaptabilidad sean factores importantes. Las startups suelen priorizar la velocidad y la agilidad. Debido a que necesitan construir una marca y generar beneficios rápidamente, es probable que prefieran soluciones que sean rápidas y fáciles de integrar, que puedan ampliarse de forma rentable, que reduzcan la dependencia de los expertos y que no acumulen ciclos valiosos.
- Algunas empresas requieren un acceso estable, de alto rendimiento y de baja latencia. Esto incluye la industria del entretenimiento y los medios de comunicación, la fabricación y el procesamiento de transacciones financieras. Si estos son sus clientes objetivo, la confiabilidad es su principal preocupación.

En todos estos casos, los clientes podrían percibir una oferta de SaaS que, por lo demás, sería buena si el acceso a la red no es fluido. Si la creación de redes se convierte en un obstáculo, esto no respalda su estrategia empresarial. Si sus clientes no pueden acceder de manera confiable a los servicios que ofrece, la propuesta de valor de sus ofertas de SaaS es nula.

Entender su función

Tu función a la hora de apoyar los objetivos empresariales depende de quién seas, cuáles sean tus objetivos individuales y de equipo específicos, quiénes son tus clientes y qué es importante para ellos. Incluso si no forma parte de un equipo que normalmente interactúa con los clientes, debe preocuparse por quiénes son y qué necesitan. Los equipos de ingeniería y desarrollo también deben preocuparse por sus clientes internos, especialmente por aquellos con los que interactúan de forma habitual. Por lo general, se trata de los equipos de operaciones y de satisfacción de los clientes.

Si forma parte de una organización de ventas, es esencial que se comunique con los equipos de producto e ingeniería sobre las redes, aunque se trate de un tema aparentemente exclusivamente tecnológico. Comparta información sobre la estructura del mercado objetivo. Comunique los puntos débiles y las necesidades de sus clientes y socios actuales y potenciales. Comparta datos y anécdotas sobre las oportunidades perdidas, el crecimiento previsto por segmento y los eventos. Haga preguntas que pongan a prueba la capacidad de su organización para respaldar el crecimiento empresarial. Esto aumenta el número de oportunidades y mejora la rentabilidad a largo plazo de su empresa. En última instancia, esto ayuda a su organización a financiar la expansión y el desarrollo futuros.

Si forma parte de la organización de ingeniería, comprenda la estrategia empresarial de su organización antes de intentar elaborar una solución. La alineación con la estrategia empresarial le ayuda a elegir las métricas correctas para evaluar las diferentes opciones de acceso a la red. También puede evitar un costoso rediseño de la red a gran escala a medida que su organización crece. La alineación empresarial ayuda a su equipo a proteger y retener los recursos necesarios para los desafíos futuros. La plantilla de su equipo, el presupuesto para el desarrollo profesional o el acceso a tecnología de vanguardia dependerán de su capacidad para demostrar la alineación empresarial. Lo ideal es que puedas demostrar cómo tus decisiones contribuyeron al éxito empresarial de la organización. Por lo tanto, le sugerimos que capture el proceso de toma de decisiones, incluidos los criterios de selección de métricas. Revise periódicamente sus métricas para confirmar que se alinean con los objetivos empresariales. Esto puede ayudar a tu equipo a obtener el reconocimiento que se merece. Las revisiones periódicas también ayudan a confirmar que tu equipo no toma decisiones basándose en suposiciones o en motivos históricos obsoletos.

La lista de métricas de las siguientes secciones es relevante para el acceso a la red:

- Métricas comerciales y de productos
- Métricas de ingeniería que influyen en las decisiones de red

Esta guía utiliza un subconjunto de estas métricas en su totalidad para ayudarlo a identificar los enfoques de acceso a la red óptimos para sus ofertas de SaaS. Elija las métricas que sean más importantes y relevantes para su empresa y, a continuación, evalúe los enfoques en función de esas métricas.

Métricas comerciales y de productos que influyen en las decisiones de creación de redes

Los equipos comerciales y de productos utilizan criterios de éxito para evaluar si están cumpliendo los objetivos empresariales. En esta sección se describen las métricas comerciales o de productos que pueden verse influidas positiva o negativamente por las decisiones de acceso a la red que tome su organización.

Utilice estas métricas y estas preguntas de autoevaluación para evaluar la forma en que su enfoque de acceso a la red se alinea con el posicionamiento empresarial y la estrategia de mercado. Esta evaluación le ayuda a determinar si sus decisiones actuales en materia de redes respaldan la diferenciación de su empresa en el mercado, las ventajas competitivas y las necesidades del público objetivo.

Esta sección contiene métricas y preguntas de autoevaluación sobre los siguientes temas:

- Modelo de negocio y posicionamiento en el mercado
- Mercado total accesible, tasa de adquisición de nuevos clientes, crecimiento y escalabilidad
- Experiencia y retención de los clientes
- Eficiencia y rendimiento financiero
- Cumplimiento normativo y gestión de riesgos
- ¿Estrategia de socios

Modelo de negocio y posicionamiento en el mercado

Estas métricas se relacionan con la posición de su empresa en el mercado, incluida la diferenciación competitiva, el alcance del mercado y la percepción de la marca. Es fundamental que evalúe la alineación entre el enfoque de acceso a la red y el modelo empresarial. Realice una evaluación independientemente de si se basa en suscripciones, en el uso, de forma gratuita, escalonada, basada en el mercado, basada en API o de marca blanca. Asegúrese de que el modelo respalde los objetivos de la organización y los objetivos de los clientes.

Criterios de puntuación alta

El enfoque de acceso a la red se alinea perfectamente con el modelo de negocio. Facilita la adopción y la prestación del servicio. Apoya la viabilidad financiera a largo plazo del modelo de negocio y la estructura de costes es compatible con el crecimiento esperado. Minimiza cualquier fricción para los clientes o socios a la hora de adoptar la oferta. Esto mejora la experiencia del usuario y fomenta una mayor aceptación del servicio.

Indicadores de puntuación baja

El enfoque de acceso a la red seleccionado no está alineado con el modelo de negocio que debería respaldar. La estructura de costos y el plazo de implementación representan un obstáculo para su adopción en el mercado objetivo. Los costos operativos y de infraestructura actuales inhiben cualquier beneficio potencial. Esto impide el crecimiento empresarial y dificulta su funcionamiento a la escala prevista. Como alternativa, las propiedades del enfoque de acceso a la red podrían impedir que los clientes consideren el servicio por motivos reglamentarios.

Preguntas de autoevaluación

- ¿Cuáles son las implicaciones financieras del enfoque de acceso a la red seleccionado para la implementación inicial y la entrega continua? ¿Cuáles son los costos fijos y variables del enfoque?
- ¿El enfoque de acceso a la red puede ampliarse de manera eficaz y eficiente para satisfacer las demandas de crecimiento del modelo empresarial? Tenga en cuenta el tamaño de cada inquilino individual y el número de inquilinos incorporados.
- ¿El enfoque de acceso a la red impone alguna limitación técnica u operativa que pueda limitar la flexibilidad o la adaptabilidad del modelo empresarial?
- En cuanto al enfoque de acceso a la red, ¿cómo se alinean el plazo de implementación con la velocidad de comercialización que exige el modelo empresarial?

Mercado total accesible, tasa de adquisición de nuevos clientes, crecimiento y escalabilidad

Es fundamental que evalúe el impacto de las decisiones sobre redes en la capacidad de la organización para expandirse a nuevos mercados, adquirir clientes de forma eficaz y mantener la escalabilidad operativa. Estos factores afectan a las tasas de conversión. También influyen en si el enfoque de acceso a la red permite la expansión a segmentos de mercado importantes o si lo limita a atender únicamente a tipos de clientes específicos.

Criterios de puntuación alta

El enfoque de acceso a la red ayuda a la organización a llegar a una parte importante del mercado objetivo, o se puede combinar eficazmente con otros enfoques de red para ampliar el alcance del mercado. Este enfoque debería requerir un esfuerzo de integración adicional mínimo. El enfoque permite plazos de entrega cortos para el despliegue, una entrada rápida en el mercado y la expansión. Permite un gran número de despliegues paralelos. La integración es sencilla para los clientes, lo que reduce las barreras a la adopción y mejora la experiencia del cliente. El enfoque minimiza los gastos operativos, preserva la capacidad operativa y respalda las proyecciones de crecimiento.

Indicadores de puntuación baja

El enfoque de acceso a la red solo apoya a una pequeña parte del mercado objetivo o se adapta principalmente a segmentos especializados que no tienen prioridad en la estrategia empresarial. No complementa eficazmente otros enfoques de acceso a la red que ya son compatibles. Los plazos de despliegue están a la zaga de las exigencias del mercado, lo que limita la expansión del mercado y la adquisición de nuevos clientes. El modelo de despliegue es secuencial, lo que aumenta los riesgos de que se produzcan cuellos de botella en los servicios a medida que aumenta la demanda. Los complejos procesos de integración disuaden a los clientes potenciales, lo que afecta negativamente a la tasa de adquisición y a las tasas de conversión. Los gastos operativos importantes disminuyen la capacidad operativa de la organización. Esto se convierte en un obstáculo para el crecimiento proyectado.

Para estos indicadores, evalúe si la introducción de un nuevo enfoque de acceso a la red puede ayudar a la organización a alcanzar sus objetivos comerciales estratégicos. Considere si el nuevo enfoque de acceso a la red podría crear nuevas dependencias entre los productos o consumir recursos operativos sin ofrecer los resultados deseados.

Preguntas de autoevaluación

- ¿Existen lagunas en el enfoque actual que le impidan llegar a segmentos más amplios del mercado objetivo?
- ¿Cuál es el conjunto mínimo de enfoques de acceso a la red estandarizados y no superpuestos que debería admitir para cubrir entre el 70 y el 90% del mercado objetivo?
- ¿Qué alcance permite cada enfoque de acceso a la red y cuáles son los aumentos relacionados en métricas importantes, como los costos de infraestructura, los ciclos operativos y la dependencia de los expertos?
- ¿Cómo se alinean las capacidades de despliegue y los límites de servicio de la infraestructura de red con las expectativas de crecimiento de sus mercados objetivo?
- ¿La integración de la red crea alguna barrera de entrada para nuevos clientes? ¿Cómo se pueden abordar estos problemas para mejorar las tasas de conversión?
- ¿Cómo afecta la sobrecarga operativa de la administración de la red a su capacidad de crecimiento y escalabilidad?
- ¿Qué estrategias puede implementar para reducir los plazos de entrega de la red y mejorar la expansión del mercado y la captación de clientes?
- ¿Existe alguna dependencia de los recursos de expertos que pueda retrasar la implementación o la integración con los ecosistemas de clientes?

Experiencia y retención de los clientes

Las métricas de esta sección le ayudan a comprender la capacidad de su organización para adquirir y, lo que es más importante, retener clientes. Comprender la relación entre los enfoques de acceso a las redes y la satisfacción de los clientes puede ayudar a los equipos de productos e ingeniería a tomar decisiones basadas en los datos.

Criterios de puntuación alta

El enfoque de acceso a la red es confiable y fácil de administrar. Contribuye a aumentar la satisfacción de los clientes (CSAT) y a obtener una puntuación neta de promotores (NPS). Estas puntuaciones son indicativas de una sólida reputación de marca y de la fidelidad de los clientes. Gracias a la perfecta integración con los ecosistemas existentes de sus clientes, las dificultades de adopción son bajas y la dependencia de los expertos es baja. Su organización cumple constantemente los acuerdos de nivel de servicio (SLAs), lo que refuerza la confianza de los clientes

y las obligaciones contractuales. Como los clientes disfrutan de servicios estables y confiables, usted tiene una alta retención de clientes.

Indicadores de puntuación baja

La difícil integración y el acceso incoherente a los servicios suelen provocar frustración y comentarios negativos por parte de los clientes. Esto daña la reputación de la marca. Los nuevos clientes no logran pasar de planes gratuitos o de prueba a servicios de pago debido a que dependen de expertos o a que los tiempos de incorporación e integración son prolongados. Los incumplimientos frecuentes se SLAs traducen en sanciones financieras y en una pérdida de credibilidad, lo que podría reducir las tasas de retención de clientes.

Preguntas de autoevaluación

- ¿Cómo afecta directamente el rendimiento de la red (como la velocidad, el tiempo de actividad y la latencia) a los resultados de CSAT y NPS? ¿Qué mejoras específicas de la red podrían impulsar estos puntajes más altos?
- ¿Cómo afectan las métricas actuales de latencia y tiempo de actividad de la red a la experiencia inicial del usuario y a las tasas de adopción? ¿Qué mejoras específicas en el rendimiento de la red se requieren para optimizar estas métricas?
- ¿Hay algún problema recurrente en la configuración de la red o en los ajustes de seguridad que complique la integración para los nuevos clientes? ¿Cómo puede agilizar estos procesos?
- ¿Cómo afecta la facilidad de configuración del acceso a la red a la experiencia de incorporación de los nuevos usuarios? ¿Existen puntos de acceso a la red o plazos de entrega específicos que puedan optimizarse para mejorar las impresiones iniciales de los usuarios?
- ¿Cuáles son los desafíos a la hora de automatizar el aprovisionamiento de servicios de red para nuevos clientes? ¿Cómo puede ajustar este proceso para mejorar la escalabilidad y la confiabilidad?
- Analice las causas fundamentales de las recientes infracciones de los SLA. ¿Estaban relacionados con la configuración de la red, la planificación de la capacidad o problemas de proveedores externos?
- ¿Con qué frecuencia los problemas de red hacen que no cumpla con sus compromisos de SLA? ¿Cuáles son los fallos más frecuentes relacionados con la red?
- ¿Qué mejoras en el rendimiento de la red han tenido el impacto positivo más significativo en la satisfacción del cliente en el pasado?

Eficiencia y rendimiento financiero

Esta categoría evalúa los aspectos financieros y de rentabilidad de su empresa, como la rentabilidad, la viabilidad a largo plazo, la rentabilidad, el retorno de la inversión (ROI) y el coste total de propiedad (TCO). Al optimizar las operaciones de red mediante la estandarización, puede reducir los gastos generales operativos y los costos de mantenimiento. Esto respalda los objetivos de crecimiento de su organización.

Criterios de puntuación alta

La estructura de costos del enfoque de acceso a la red está bien alineada con el modelo de negocio. Apoya el crecimiento sostenible y los importantes ahorros de costos que se logran aumentan la rentabilidad. El acceso eficiente a la red permite una rápida incorporación de los clientes, lo que reduce el tiempo necesario para generar valor y acelera la penetración en el mercado. Esto acorta directamente el ciclo de reconocimiento de ingresos.

Indicadores de puntuación baja

Los clientes recurren a su competencia para acelerar la entrega de sus aplicaciones y servicios. Su organización ha aumentado los costes operativos asociados a las complejas y variadas configuraciones de red y a la ampliación de los plazos de entrega. La estructura de costos y el modelo de negocio están desalineados, lo que podría provocar altos costos iniciales para los servicios basados en suscripciones. Los engorrosos procesos de incorporación reducen la penetración en el mercado y posponen el reconocimiento de los ingresos.

Preguntas de autoevaluación

- ¿Cuáles son los plazos de entrega actuales para la implementación de nuevos servicios y cómo afectan al tiempo de comercialización y al reconocimiento de los ingresos?
- ¿Con qué eficacia las operaciones de red estandarizadas reducen los gastos generales y los costos de mantenimiento?
- ¿Se necesitan recursos expertos para completar con éxito la integración inicial, operar a diario, solucionar problemas o implementar cambios?
- ¿Cuán sostenibles son las inversiones actuales en redes en términos de avances tecnológicos? ¿Está invirtiendo en tecnologías preparadas para el futuro que se alineen con la evolución proyectada del mercado?
- ¿Con qué eficacia asigna y realiza un seguimiento de los costos relacionados con el tráfico y el uso de la red por parte de los inquilinos individuales?

Cumplimiento normativo y gestión de riesgos

Es fundamental validar el cumplimiento de las normas relacionadas con la red. Esto confirma que opera legalmente y que puede mantener la confianza de los clientes. La estandarización de las operaciones de la red simplifica el proceso de cumplimiento y promueve la coherencia en diversas jurisdicciones y geografías. Estas medidas le ayudan a ampliar sus servicios.

Criterios de puntuación alta

Las operaciones de la red se adhieren constantemente a las normas legales sin complicaciones, lo que contribuye a la expansión del mercado, reduce las dificultades de adopción y mejora la confianza de los clientes. El cumplimiento demostrado de los marcos normativos fundamentales, como la Ley de Resiliencia Operativa Digital (DORA) y el Instituto Nacional de Estándares y Tecnología (NIST), le ayuda a conseguir clientes sensibles al cumplimiento de las normas. La visibilidad continua de su estado de cumplimiento reduce el tiempo necesario para completar una auditoría.

Indicadores de puntuación baja

Las brechas en el cumplimiento de la red provocan grandes problemas de adopción, demoras en el lanzamiento del servicio, impugnaciones legales y posibles multas. Estos desafíos provocan el retraso o la cancelación de los planes de expansión a nuevos mercados. Es difícil mantener las prácticas de cumplimiento estándar en las diferentes jurisdicciones, y esto afecta a la eficiencia operativa y a la reputación en el mercado.

Preguntas de autoevaluación

- ¿En qué medida se alinean las operaciones de su red con las directrices reglamentarias o industriales aplicables? ¿Qué reveló su auditoría de cumplimiento más reciente?
- ¿Cómo mantiene el cumplimiento de las nuevas normativas en los ámbitos de la seguridad digital y de redes?
- ¿Cuán eficaz es su proceso de documentación e informes para cumplir con los requisitos de los diferentes organismos reguladores?
- ¿Qué estrategias de gestión de riesgos tiene implementadas para identificar y abordar los posibles riesgos de cumplimiento antes de que den lugar a impugnaciones legales?
- ¿Qué nivel de formación y conocimiento sobre el cumplimiento necesitan sus equipos de administración de redes para respaldar sus enfoques de acceso a la red?

¿Estrategia de socios

Evalúe en qué medida el enfoque de acceso a la red se alinea con un ecosistema de socios, plataformas y mercados reconocidos. Esto es esencial, especialmente si su estrategia de crecimiento depende de la expansión a través de socios.

Criterios de puntuación alta

El enfoque de acceso a la red está integrado en todo su ecosistema de socios. Su estructura de costes se alinea bien con los modelos de negocio de sus socios clave. Los socios poseen las habilidades de red necesarias para una integración perfecta de sus ofertas de SaaS y pueden ofrecer un acceso y una funcionalidad sostenidos.

Indicadores de puntuación baja

El enfoque de acceso a la red seleccionado exige habilidades, recursos o equipos especializados que son escasos o difíciles de adquirir. Se diferencia de los protocolos de acceso a la red estándar que suelen utilizar las plataformas y los mercados. Esto se traduce en una estructura de costes impredecible que es difícil de conciliar. El enfoque de acceso a la red no está alineado con los modelos de negocio de sus socios clave.

Preguntas de autoevaluación

- ¿Cuáles son las implicaciones financieras del enfoque de acceso a la red para los socios? ¿Cómo se alinean estos costos con sus modelos de negocio? ¿Qué lado de la integración asume la mayor parte de los costos y cuántos ciclos operativos se deben invertir?
- En cuanto al enfoque de acceso a la red, ¿existen barreras para la integración o el mantenimiento que puedan afectar a las relaciones con los socios o a la escalabilidad del ecosistema?
- ¿Cómo se puede optimizar el enfoque de acceso a la red para mejorar la compatibilidad y la facilidad de integración en todo el ecosistema?

Métricas de ingeniería que influyen en las decisiones de red

Al igual que los equipos comerciales y de productos, los equipos de ingeniería también utilizan criterios de éxito para evaluar si cumplen los objetivos empresariales. Sin embargo, estas métricas son diferentes y se centran en la capacidad del equipo para desarrollar, operar y cumplir los

requisitos de seguridad y conformidad. En esta sección se describen las métricas de ingeniería que pueden verse influidas positiva o negativamente por las decisiones de acceso a la red que tome su organización.

Utilice estas métricas y estas preguntas de autoevaluación para evaluar su enfoque actual de acceso a la red en comparación con los requisitos empresariales y las capacidades técnicas. Esta evaluación le ayuda a identificar las brechas en su arquitectura y a priorizar las mejoras que se ajusten a sus objetivos estratégicos. Al revisar estos criterios con regularidad, puede asegurarse de que su estrategia de acceso a la red sigue respaldando las necesidades de sus clientes y los planes de crecimiento de su organización.

Esta sección contiene métricas y preguntas de autoevaluación para las siguientes categorías y temas:

Métricas de desarrollo relacionadas con el acceso a la red para las ofertas de SaaS

Esta sección contiene las siguientes métricas:

- La calidad del código y el rendimiento del equipo de ingeniería
- Reducción técnica de la deuda
- Escalabilidad, capacidad y rendimiento

Frecuencia de despliegue, tiempo de despliegue y velocidad de sprint

Para optimizar la eficiencia del ciclo de desarrollo, es esencial que comprenda la influencia del aprovisionamiento de pilas de red en la velocidad de los sprints.

Criterios de puntuación alta

El aprovisionamiento de la pila de redes está optimizado y automatizado, y requiere una intervención manual mínima. No afecta significativamente a la velocidad del sprint. Cualquier miembro del equipo puede realizar el aprovisionamiento y la redistribución de la pila de redes. Esto reduce los cuellos de botella y la dependencia de los recursos especializados.

Indicadores de puntuación baja

Se necesita una gran cantidad de argumentos para aprovisionar la red. Esto sugiere un proceso complejo y lento que dificulta el desarrollo de nuevas funciones. La redistribución frecuente de la red implica gastos generales considerables de tiempo y costes. Las tareas de aprovisionamiento de redes requieren conocimientos de ingeniería especializados, lo que crea cuellos de botella y ralentiza el ciclo de desarrollo.

Preguntas de autoevaluación

- Qué pasos manuales, si los hay, están involucrados en el proceso de implementación. ¿Cómo afectan a la frecuencia y el tiempo de despliegue?
- ¿Cómo se gestionan las reversiones en caso de errores en la implementación? ¿Cuál es su impacto en la frecuencia de implementación y el tiempo de recuperación?
- ¿Cuántos argumentos se necesitan para aprovisionar la red cuando se configuran nuevos entornos?
- ¿Cuáles son los costes adicionales y la sobrecarga de tiempo que conlleva la redistribución frecuente de la pila de red durante el proceso de desarrollo?
- ¿El aprovisionamiento de la red depende de la experiencia en ingeniería especializada o se trata de una tarea que puede gestionar cualquier miembro del equipo?

Flexibilidad y entrega de funciones

El enfoque de acceso a la red puede influir en la capacidad del equipo de ingeniería para innovar e implementar nuevas funciones de manera eficiente.

Criterios de puntuación alta

El enfoque de acceso a la red ofrece la flexibilidad necesaria para un despliegue de funciones rápido y fluido. Es compatible con una amplia gama de protocolos de comunicación, comunicaciones unidireccionales y bidireccionales y tamaños de mensajes. No impone restricciones significativas a los procesos de desarrollo ni a la innovación.

Indicadores de puntuación baja

El enfoque de acceso a la red restringe la capacidad del equipo para implementar nuevas funciones debido a la falta de protocolos de comunicación compatibles, a la inflexibilidad en el tamaño de los mensajes o a la dependencia de tecnologías específicas y recursos especializados relacionados. Esto puede provocar ciclos de desarrollo más lentos y obstaculizar la evolución del servicio.

Preguntas de autoevaluación

- ¿Cómo afecta el enfoque de acceso a la red a la agilidad del equipo a la hora de desarrollar e implementar nuevas funciones?
- ¿Existen limitaciones en el enfoque de acceso a la red que restrinjan la compatibilidad con determinados protocolos o tecnologías de comunicación?
- ¿Cómo facilita o limita este enfoque la integración de nuevas tecnologías e innovaciones en el servicio?
- ¿Cómo afecta el enfoque de acceso a la red a los plazos de desarrollo y a la hoja de ruta del producto?

Cambie la tasa de fallas

El enfoque de acceso a la red que elija puede afectar a la tasa de errores de los cambios al implementar nuevos servicios o funciones. Un mayor control a menudo implica una mayor flexibilidad, pero también aumenta la posibilidad de que se produzcan errores de configuración, como cuando se administra una configuración de enrutamiento compleja.

Criterios de puntuación alta

Puede implementar cambios en la red con un riesgo mínimo de fallo. Existen suficientes mecanismos de prueba, mecanismos de reversión eficientes y una supervisión eficaz le ayuda a identificar y resolver los problemas rápidamente.

Indicadores de puntuación baja

El enfoque de acceso a la red es propenso a fallar durante los cambios. Las opciones de prueba son limitadas, las estrategias de implementación son complicadas o las capacidades de supervisión y solución de problemas son insuficientes. Se requiere la participación de varias partes en las sesiones de solución de problemas. Esto puede provocar un aumento del tiempo de inactividad y reducir la disponibilidad de la oferta de SaaS.

Preguntas de autoevaluación

- ¿Qué medidas se han adoptado para mitigar el riesgo de que los cambios no se produzcan al actualizar el conjunto de redes?
- ¿Existen procesos exhaustivos de prueba y validación?
- ¿Con qué rapidez puede el sistema recuperarse de un cambio fallido? ¿Existe un proceso de reversión eficiente?
- ¿Existen sistemas proactivos de monitoreo y alerta para detectar y abordar los problemas con rapidez durante y después de los cambios en la red?
- ¿Cuál es la tasa de errores por cambios históricos en las implementaciones de pilas de red? ¿Qué lecciones se han aprendido de los incidentes pasados?
- ¿Cómo facilita o limita el enfoque de acceso a la red la implementación de los cambios? ¿El enfoque minimiza la interrupción del servicio?
- ¿Cuál es el riesgo de afectar a la disponibilidad de la oferta de SaaS en el entorno de producción cuando se implementan cambios que implican el enfoque de acceso a la red?

La calidad del código y el rendimiento del equipo de ingeniería

Los enfoques de acceso a la red pueden afectar indirectamente a la calidad del código de las ofertas de SaaS. La falta de estandarización en el acceso a la red puede obligar al equipo de ingeniería a admitir múltiples enfoques de integración, lo que puede llevar a una base de código sobrecargada. Esto, a su vez, puede dificultar la capacidad del equipo para desarrollar la profundidad

y el control sobre la calidad del código necesarios para mantener un alto rendimiento de los equipos de ingeniería.

Criterios de puntuación alta

El equipo de ingeniería se mantiene concentrado gracias a la modularidad del código y a la reutilización en todos los enfoques de acceso a la red compatibles. Los enfoques de acceso a la red son compatibles con los procesos de despliegue y las estrategias de pruebas automatizadas existentes.

Indicadores de puntuación baja

El rendimiento del equipo de ingeniería se reduce debido a la sobrecarga asociada a la integración y el mantenimiento de demasiados enfoques de acceso a la red. Algunos enfoques aumentan significativamente la complejidad, generan deuda tecnológica o requieren el desarrollo de soluciones alternativas para abordar las capacidades faltantes o insuficientes.

Preguntas de autoevaluación

- ¿Cómo gestiona el enfoque de acceso a la red la variabilidad de la red?
- ¿Necesita desarrollar un código adicional para gestionar las interrupciones en la conectividad?
- ¿Un nuevo enfoque de acceso a la red se integra perfectamente con los enfoques existentes o requiere un desarrollo personalizado significativo?
- ¿Cuál es el alcance del cambio necesario para adoptar un nuevo enfoque de acceso a la red? ¿Se pueden utilizar eficazmente la base de código existente y las pruebas automatizadas?
- ¿Qué tan fácil o difícil es implementar o volver a implementar el servicio con el enfoque de acceso a la red seleccionado? ¿Se puede hacer esto con frecuencia? ¿Existe alguna dependencia de los recursos de expertos?
- ¿El enfoque de acceso a la red facilita o complica el cumplimiento de los estándares de codificación y las mejores prácticas?
- ¿Cómo afecta este enfoque a time-to-market las nuevas funciones o correcciones?

Reducción técnica de la deuda

Una evaluación del impacto de un enfoque de acceso a la red en la deuda técnica debería considerar sus capacidades de escalabilidad, observabilidad y seguridad.

Criterios de puntuación alta

El enfoque agiliza de manera efectiva la administración de la infraestructura a medida que la base de clientes se expande. Ofrece sólidas capacidades de observabilidad. out-of-the-box Esto promueve una supervisión y un mantenimiento eficientes.

Indicadores de puntuación baja

El enfoque de acceso a la red protege inadecuadamente los canales de comunicación y carece de herramientas suficientes para la observación métrica cualitativa. También puede requerir un desarrollo adicional para la administración de la infraestructura a medida que aumenta la base de clientes, o puede necesitar soluciones alternativas para los problemas de confiabilidad.

Preguntas de autoevaluación

- ¿Cómo influye el enfoque de acceso a la red en la escalabilidad a largo plazo de la infraestructura? ¿Facilita un crecimiento continuo con una inversión adicional mínima?
- ¿Qué tan completas son las herramientas de observabilidad incluidas? ¿Permiten una supervisión proactiva y la resolución de problemas?
- ¿Cuál es el impacto previsto del enfoque de acceso a la red en el mantenimiento y la evolución del código base a lo largo del tiempo?
- ¿El enfoque se integra bien con la infraestructura existente y planificada? ¿Requiere cambios o adiciones importantes?

Escalabilidad, capacidad y rendimiento

Para determinar la idoneidad de un enfoque de acceso a la red para una oferta de SaaS, es esencial analizar cómo mantiene un rendimiento óptimo a medida que aumenta la demanda.

Criterios de puntuación alta

El enfoque de acceso a la red facilita la expansión sin problemas. Mantiene una latencia baja durante el procesamiento de las solicitudes y gestiona de forma eficiente los picos de tráfico. Proporciona un rendimiento uniforme independientemente del aumento de los niveles de tráfico y no impone límites operativos al crecimiento.

Indicadores de puntuación baja

El enfoque de acceso a la red no se escala de manera eficaz, posiblemente debido a las limitaciones inherentes del ancho de banda o a una capacidad de infraestructura insuficiente. El

aprovisionamiento y la administración de recursos aumentan la complejidad o crean dependencias. El rendimiento del servicio se deteriora debido al aumento de la latencia, la fluctuación y la variabilidad del rendimiento, especialmente en condiciones de red congestionadas.

Preguntas de autoevaluación

- ¿Cómo se adapta el enfoque de acceso a la red a un número cada vez mayor de inquilinos y sus volúmenes de datos?
- ¿Es inherentemente escalable para cumplir con las demandas del futuro?
- ¿Qué medidas existen para garantizar que el rendimiento sea uniforme, incluso durante los períodos de mayor tráfico o los eventos de escalamiento rápido?
- ¿Cómo gestiona el enfoque la latencia y la fluctuación de la red? ¿Existen mecanismos para optimizar el rendimiento de los datos y minimizar los retrasos?
- ¿Puede el enfoque de acceso a la red adaptarse a las diferentes condiciones de la red? ¿Puede proporcionar una experiencia de usuario único para cada cliente?
- ¿Cuál es el impacto del enfoque de acceso a la red en la infraestructura subyacente? ¿Requiere actualizaciones o cambios significativos en los sistemas existentes?

Métricas de excelencia operativa relacionadas con el acceso a la red para las ofertas de SaaS

Esta sección contiene las siguientes métricas:

- Resiliencia operativa y recuperación ante desastres
- Supervisión del rendimiento de los servicios y las aplicaciones

Resiliencia operativa y recuperación ante desastres

El enfoque de acceso a la red debería ayudar a la oferta de SaaS a resistir varios tipos de interrupciones y a recuperarse rápidamente de cualquier desastre.

Criterios de puntuación alta

Los planes de recuperación ante desastres establecidos y probados muestran de forma sistemática que el enfoque de acceso a la red cumple con los requisitos de recuperación ante desastres. El enfoque de acceso a la red admite configuraciones de alta disponibilidad y admite mecanismos de conmutación por error automáticos, rápidos y confiables.

Indicadores de puntuación baja

El enfoque de acceso a la red dificulta la creación de una estrategia coherente de recuperación ante desastres. Se observan tiempos de recuperación prolongados después de las interrupciones. Los frecuentes fallos operativos de la infraestructura de red están afectando a la prestación de servicios.

Preguntas de autoevaluación

- ¿Cuándo se llevó a cabo el último simulacro de recuperación ante desastres y cuáles fueron los resultados?
- ¿Cuánto tiempo se tarda en recuperar los servicios críticos después de una interrupción? ¿Qué parte de la infraestructura de red se debe volver a implementar?
- ¿Qué mejoras se pueden realizar en la infraestructura de red para agilizar sus planes de recuperación ante desastres?
- ¿Existen redundancias para los componentes de red más importantes?
- ¿Ha automatizado la posible redistribución de la infraestructura de red tras una interrupción crítica?
- ¿Cómo contribuye el enfoque de acceso a la red a la fiabilidad y la tolerancia a los errores? ¿Existen mecanismos integrados para gestionar las interrupciones de la red y mantener la integridad de los datos?

Supervisión del rendimiento de los servicios y las aplicaciones

El enfoque de acceso a la red puede afectar a las herramientas de supervisión del rendimiento que se utilizan para validar el funcionamiento óptimo y el tiempo de actividad del servicio. Según el servicio, es posible que tenga acceso a métricas de bajo nivel (como las tasas de entrega de paquetes) o métricas de nivel superior (como la duración de la sesión). Las métricas de bajo nivel proporcionan información técnica detallada sobre el comportamiento de la red, pero su interpretación puede resultar compleja. Por el contrario, las métricas de nivel superior suelen ofrecer una forma más directa y sencilla de evaluar la experiencia general del usuario. Esto se debe a que agregan el impacto de las condiciones subyacentes de la red en indicadores claros de la calidad del servicio.

Criterios de puntuación alta

Están fácilmente disponibles herramientas de monitoreo integrales que proporcionan información casi en tiempo real. Dispone de sistemas de alerta y respuesta automatizados que abordan los problemas de rendimiento. Puede predecir posibles cuellos de botella o fallos en el servicio antes de que afecten a los usuarios.

Indicadores de puntuación baja

Las interrupciones frecuentes del servicio o los problemas de rendimiento se producen sin que se observen ni se tomen medidas al respecto. La falta de visibilidad del rendimiento del servicio provoca una respuesta lenta a los cuellos de botella en el rendimiento. Se requieren equipos multipartitos para solucionar los problemas de infraestructura de la red.

Preguntas de autoevaluación

- ¿Qué herramientas de monitoreo y métricas de infraestructura de red están disponibles actualmente? ¿Cuán eficaces son a la hora de detectar anomalías en el servicio?
- ¿Con qué rapidez puede identificar y resolver los problemas de rendimiento?
- ¿Cuenta con mecanismos que predicen posibles problemas de rendimiento?
- ¿Qué mejoras puede realizar para mejorar las capacidades de observabilidad?

Métricas de seguridad y gobierno relacionadas con el acceso a la red para las ofertas de SaaS

Esta sección contiene las siguientes métricas:

- Gestión de la seguridad, el cumplimiento y las vulnerabilidades

Gestión de la seguridad, el cumplimiento y las vulnerabilidades

Es fundamental que evalúe los aspectos de seguridad del enfoque de acceso a la red, incluido el cumplimiento de las normas de seguridad y la gestión de las vulnerabilidades.

Criterios de puntuación alta

El enfoque de acceso a la red ayuda a su equipo a cumplir con los marcos de seguridad, como la Organización Internacional de Normalización (ISO) 27001, los Controles de Sistemas y Organizaciones 2 (SOC 2) o el NIST. Facilita la realización de auditorías de seguridad periódicas. Existen sólidos mecanismos de cifrado y autenticación. Las redes están aisladas y solo los recursos necesarios están expuestos a la infraestructura del cliente. Puede detectar anomalías en las redes prácticamente en tiempo real, sin sobrecargas excesivas.

Indicadores de puntuación baja

El enfoque de acceso a la red es propenso a sufrir brechas de seguridad o vulnerabilidades recurrentes y no cumple con las principales normas de seguridad. Con frecuencia, se observan retrasos en la detección y la respuesta a los incidentes de seguridad.

Preguntas de autoevaluación

- ¿Hay alguna violación de seguridad reciente relacionada con un enfoque de acceso a la red seleccionado y qué hemos aprendido de ella?
- ¿Cómo cumple su enfoque de acceso a la red con los estándares de seguridad globales?
- ¿Cuánto tiempo se tarda en detectar las amenazas a la seguridad y responder a ellas? ¿Cómo ayuda o limita el acceso a la red a esta capacidad?
- ¿Con qué frecuencia se realizan evaluaciones de seguridad en los enfoques de acceso a la red? ¿Puede utilizar las herramientas habituales para evaluar la seguridad del enfoque de acceso a la red o se necesita un software especializado?
- ¿Qué nivel de seguridad es inherente al enfoque de acceso a la red y cómo se ajusta a las mejores prácticas del sector y a los requisitos normativos?

Descripción general de los servicios de AWS red para las ofertas de SaaS

En esta sección se analizan los servicios AWS de red a los que se hace referencia en esta guía. También compara sus capacidades y describe las consideraciones de seguridad de cada servicio.

Esta sección contiene los siguientes temas:

- AWS servicios de red
- Comparación de las capacidades de los servicios
- Características y consideraciones de seguridad

AWS servicios de red

Los siguientes son los Servicios de AWS que se analizan de forma coherente en esta guía.

AWS PrivateLink

AWS PrivateLinkes un servicio nativo de la nube que puede proporcionar acceso a su oferta de SaaS si sus clientes ya operan en el. Nube de AWS Su cliente se conecta a la oferta de SaaS a través de un punto final de VPC de interfaz. Se trata de una interfaz de red de punto final que se aprovisiona en una o más subredes de la del cliente. Cuenta de AWS En los escenarios de esta guía, el tráfico viaja a través del punto final de la VPC de la interfaz y llega a un Network Load Balancer de su cuenta. El Network Load Balancer reenvía el tráfico a la aplicación SaaS, que ha registrado como servicio de punto final. A través de los puntos finales de VPC de recursos, también AWS PrivateLink puede ayudarlo a acceder a otros recursos, como las bases de datos.

Amazon VPC Lattice

Amazon VPC Lattice es un servicio de redes de aplicaciones que ayuda a los proveedores de SaaS a ofrecer sus servicios de forma segura y eficiente a los clientes que operan en múltiples y. VPCs Cuentas de AWS Los clientes acceden a su oferta de SaaS a través de VPC Lattice, que ofrece una conectividad de red uniforme, controles de acceso sólidos y una gestión avanzada del tráfico. En estos escenarios, el tráfico fluye a través de VPC Lattice hacia los servicios de aplicaciones registrados. Proporciona una comunicación escalable y segura, independientemente del servicio informático que utilice.

Emparejamiento de VPC

El emparejamiento de VPC es una conexión de red entre dos nubes privadas virtuales (VPCs) que enruta el tráfico entre ellas mediante direcciones o IPv4 direcciones privadas. IPv6 La interconexión de VPC se suele utilizar entre entidades de confianza, como las que están dentro de la misma organización. Su cliente crea una solicitud de interconexión para una de las suyas. VPCs Si lo aceptas, el tráfico puede fluir entre ambos VPCs en cualquier dirección. Este enfoque de conexión destaca por su singularidad, ya que implica la comunicación directa entre dos VPCs sin ningún servicio intermediario o infraestructura que gestionar.

AWS Transit Gateway

AWS Transit Gatewayes un centro de tránsito de red centralizado que puede conectar VPCs conexiones de red privada virtual (VPN), AWS Direct Connect puertas de enlace, dispositivos virtuales de terceros en una VPC y otras puertas de enlace de tránsito. Una pasarela de tránsito puede tener una tabla de rutas diferente para cada adjunto. Esto proporciona la máxima flexibilidad de enrutamiento y le ayuda a aislar las redes. A menudo se usa para VPCs conectar muchas de ellas o para una inspección centralizada.

AWS Site-to-Site VPN

AWS Site-to-Site VPNpuede usar la tecnología de protocolo de seguridad de Internet (IPsec) para establecer conexiones entre redes locales, oficinas remotas, fábricas, otros proveedores de servicios en la nube y la red AWS global. La conexión se establece desde una puerta de enlace privada virtual o una puerta de enlace de tránsito en una VPC en la Nube de AWS hasta una puerta de enlace de cliente física o basada en software, que puede estar en las Nube de AWS instalaciones o en la nube de otro CSP. La conexión puede realizarse a través de Internet o mediante una conexión física. AWS Direct Connect También es posible tener una conexión Site-to-Site VPN acelerada mediante el uso de AWS Global Accelerator. Una conexión acelerada dirige el tráfico a una ubicación AWS perimetral y ofrece una latencia reducida y un rendimiento mejorado.

AWS Direct Connect

AWS Direct Connectestablece una conexión privada de alta velocidad entre un centro de datos local y el Nube de AWS. Al evitar la Internet pública, Direct Connect proporciona una conexión de baja latencia más confiable, segura y consistente al. Nube de AWS Los clientes se conectan a una de las Direct Connect ubicaciones y, a continuación, eligen una conexión alojada o dedicada a AWS. Si

bien esta es una opción de arquitectura poco común para las ofertas de SaaS, puede ser adecuada para los proveedores de SaaS que tienen pocos consumidores pero son grandes empresas.

Comparación de las capacidades de los servicios

En la siguiente tabla se describen las capacidades compatibles Servicios de AWS que se describen en esta guía. A continuación se describen las capacidades incluidas en esta tabla:

- Rangos de CIDR superpuestos: puede conectar dos o más redes con los mismos rangos de CIDR o superpuestos
- Comunicación bidireccional: puede admitir un canal de comunicación bidireccional para que el consumidor de SaaS pueda exponer los recursos internos, como una base de datos, al proveedor de SaaS
- IPv6— Puede admitir una o dos IPv6 pilas
- Trama gigante: puede admitir tramas gigantes con un tamaño de trama de hasta 8.500 bytes
- Nube híbrida: admite una conexión con una red local
- Nube múltiple: puede admitir una conexión entre redes de diferentes proveedores de servicios en la nube

Servicio o enfoque	Intervalos de CIDR superpuestos	Comunicación bidireccional	IPv6	Marco jumbo	Nube híbrida	Nube múltiple
Interconexión con VPC	No	Sí	Sí	Sí [5]	No	No
AWS PrivateLink	Sí	Sí [1]	Sí	Sí	No [6]	Número [6]
Amazon VPC Lattice	Sí	Sí [1]	Sí	Sí	Número [6]	Número [6]

AWS Transit Gateway	No	Sí	Sí	Sí	Sí [3]	Sí [3]
AWS Site-to-Site VPN	No	Sí	Sí	No	Sí	Sí
AWS Direct Connect	No	Sí	Sí	Sí[2]	Sí	Sí
Acceso público a internet [4]	No aplicable	No	Sí	Sí	Sí	Sí

1. Con recursos de VPC en Amazon VPC Lattice
2. Solo para interfaces virtuales privadas y de tránsito
3. Con Site-to-Site VPN o AWS Direct Connect archivos adjuntos
4. Como término general para AWS los recursos que hacen que una aplicación sea accesible públicamente, como Application Load Balancer
5. Solo para interconectar conexiones dentro de una Región de AWS
6. Es posible mediante una conexión de capa 3 preexistente entre los entornos

Características y consideraciones de seguridad

En la siguiente tabla se describen las características de seguridad Servicios de AWS que se describen en esta guía.

- Medios de autenticación: cómo puede asegurarse de que solo sus clientes puedan conectarse a su servicio. Por lo general, todavía se requiere otro nivel de autenticación para las solicitudes entrantes, especialmente en los entornos de inquilinos compartidos.
- Cifrado en tránsito: describe si el cifrado en tránsito se proporciona de forma predeterminada. El cifrado nativo describe el cifrado que AWS cubre todo el tráfico dentro VPCs VPCs, a través o

entre los centros de datos. El cifrado complementario describe el cifrado que usted controla y que el servicio correspondiente puede detener.

Servicio o enfoque	Medios de autenticación	Cifrado en tránsito
Interconexión con VPC	Usted inicia una solicitud de intercambio entre pares Cuenta de AWS y la VPC de su cliente o acepta una solicitud que él inicie. Consulte Aceptar o rechazar una conexión de emparejamiento de VPC.	Solo cifrado nativo
AWS PrivateLink	Usted elige cuáles Cuentas de AWS están autorizados a crear puntos de conexión para su servicio. Estas cuentas se conocen como principales permitidas. Consulte Aceptar o rechazar solicitudes de conexión.	Solo cifrado nativo
Amazon VPC Lattice	Compartes un servicio o una red de servicios de VPC Lattice con tus clientes. Cuentas de AWS Consulte Compartir sus entidades de VPC Lattice.	Cifrado nativo y cifrado TLS complementario
AWS Transit Gateway	Su cliente crea una solicitud de adjunto entre pares a partir de su Cuenta de AWS cliente o usted inicia la solicitud. Consulte los archivos adjuntos de emparejamiento de Transit	Cifrado nativo y IPsec cifrado complementario con un adjunto de VPN

	Gateways en Amazon VPC Transit Gateways.	
AWS Site-to-Site VPN	Utiliza claves IPsec previamente compartidas o un certificado privado en el dispositivo del cliente. Consulte las opciones de autenticación de AWS Site-to-Site VPN túnel.	IPsec Cifrado complementario
AWS Direct Connect	Su cliente crea una solicitud de interfaz virtual desde su Cuenta de AWS. Consulte las interfaces Direct Connect virtuales y las interfaces virtuales alojadas.	Es posible el cifrado de capa 2 adicional en sitios seleccionados. Consulte Direct Connect las ubicaciones.
Acceso público a internet [1]	Se requiere una autenticación personalizada.	Es posible el cifrado TLS adicional

1. Como término general para AWS los recursos que hacen que una aplicación sea accesible públicamente, como Application Load Balancer

Evaluación de las opciones de acceso a la red para las ofertas de SaaS

Las métricas que son importantes para su organización dependerán de quiénes sean sus clientes, su estrategia empresarial y los objetivos de su organización. Esta guía presenta métricas que puede usar para elegir un enfoque de acceso a la red, pero debe priorizar aquellas que cumplan con los requisitos únicos de su caso de uso.

Esta sección contiene los siguientes temas:

Métricas de evaluación

Algunas métricas son consistentes en todas las organizaciones y casos de uso, y estas son las métricas que podemos ayudarlo a calificar. Estas métricas son las siguientes:

- Facilidad de integración: ¿con qué rapidez y facilidad puede incorporar nuevos clientes?
- Coste total de propiedad (TCO): ¿Cuál es la estructura de costes? Más allá de los costos de infraestructura fijos y variables, existen importantes consideraciones de costos adicionales relacionadas con los gastos operativos, la dependencia de los expertos, el costo de implementar los cambios y el cumplimiento. Para obtener más información, consulte la sección Costo total de propiedad.
- Escalabilidad: ¿su enfoque de acceso a la red puede ampliarse para respaldar el crecimiento de su empresa? Ampliar su base de clientes implica importantes consideraciones arquitectónicas y organizativas. Considere cómo podría escalar para dar cabida a entre 5 y 100 veces más clientes de los que atiende en la actualidad.
- Adaptabilidad: ¿puede implementar cambios fácilmente? Los cambios pueden incluir una nueva aplicación, una nueva capacidad, una plataforma diferente o una red diferente.
- Aislamiento de la red: ¿qué parte de la infraestructura de red está exponiendo a sus clientes? ¿Está proporcionando el grado de acceso justo o está exponiendo redes enteras? Si aísla los recursos de la red desde el principio, será más fácil garantizar la seguridad, la privacidad y el cumplimiento en el futuro.

- Observabilidad: ¿Cuál es su capacidad para detectar una falla o degradación del servicio? ¿Qué tan fácil y rápido es identificar el problema? ¿Con qué rapidez (y con qué gastos generales) puede ayudar a sus clientes a comprender sus puntos de falla y ayudarlos a resolverlos?
- Tiempo de reparación: ¿cuánto tiempo transcurre entre la detección de un fallo o una degradación del servicio y la reanudación de las operaciones? ¿Cuáles son los factores que afectan a esta capacidad?

Otras métricas son exclusivas de su organización u oferta porque se relacionan con las operaciones, la estrategia o los objetivos de su empresa. Solo usted puede calificar estas métricas. Estas son las siguientes métricas:

- Alineación del modelo de negocio: ¿Cuál es su modelo de negocio y en qué medida se ajustan a él los enfoques de acceso individual?
- Mercado total direccionable (TAM): ¿Cuál es su mercado actual y futuro y qué tan bien lo cubre el enfoque de acceso a la red?
- Retorno de la inversión (ROI): ¿Qué mejoras espera en términos de rentabilidad y márgenes? ¿Los beneficios financieros esperados son suficientes para satisfacer sus necesidades de acceso a los servicios flexible y adaptable?
- Cumplimiento normativo: ¿qué tipo de requisitos reglamentarios se aplican y en qué mercado?
- Acuerdos de nivel de servicio (SLAs): ¿Los clientes necesitan que su oferta de SaaS tenga una alta disponibilidad? ¿Qué tipo de compromisos está obligado contractualmente a cumplir?

Costo total de propiedad

En esta sección se analiza el costo total de propiedad (TCO), que es una de las métricas de evaluación utilizadas para comparar los enfoques de acceso a la red. El TCO es una métrica compuesta que consta de los costos de infraestructura fijos y variables, los gastos operativos, la dependencia de los especialistas, el costo del cambio y los costos de cumplimiento.

La calificación del TCO de cada enfoque de acceso a la red puede variar según el caso de uso. Por ejemplo, el costo del cambio para un proveedor de SaaS con un servicio web simple y cinco inquilinos difiere del de un proveedor de SaaS con una cartera de productos compleja e interconectada y cientos o miles de inquilinos. Además, no todos los componentes tienen el mismo peso. Por ejemplo, contratar a un especialista en redes suele ser más caro que los costes de

infraestructura necesarios para la implementación individual de su servicio. Utilice los valores de la siguiente tabla como orientación inicial y como punto de referencia para un análisis más detallado.

Método de acceso	Costos de infraestructura fijos	Costes de infraestructura variables	Gastos generales operativos	Dependencia de especialistas	Coste del cambio	Costes de cumplimiento
Interconexión con VPC	Ninguno	Ninguno	Alto	Bajo	Alto	Medio
AWS PrivateLink	Bajo	Bajo	Bajo	Ninguno	Bajo	Bajo
Amazon VPC Lattice	Medio	Medio	Bajo	Bajo	Bajo	Bajo
AWS Transit Gateway	Medio	Medio	Bajo	Bajo	Bajo	Medio
AWS Site-to-Site VPN	Medio	Alto	Alto	Medio	Medio	Bajo
AWS Direct Connect	Alto	Medio	Medio	Alto	Alto	Bajo
Acceso público a Internet	Bajo	Alto	Medio	Bajo	Bajo	Alto

Costes de emparejamiento de VPC

No hay ningún coste de infraestructura directo asociado a una conexión de emparejamiento de VPC. Cuando el tráfico permanece dentro de la misma zona de disponibilidad, no hay ningún cargo por transferencia de datos. Sin embargo, la sobrecarga operativa puede ser significativa porque la administración y la complejidad aumentan exponencialmente con cada conexión interconectada adicional. Para configurar una conexión entre pares basta con tener algunos conocimientos básicos sobre las redes, pero los cambios en la red son difíciles de implementar con más de un puñado de conexiones entre pares. Los costos de cumplimiento son ligeramente más altos porque ambas partes exponen una VPC completa entre sí, en lugar de a servicios individuales.

AWS PrivateLink costes

AWS PrivateLink suele ser una solución rentable con una pequeña sobrecarga operativa. Esto se debe a que el proveedor de SaaS solo debe administrar un Network Load Balancer y el consumidor solo debe administrar los puntos finales de la VPC. Puede realizar cambios en ambos lados de forma transparente, lo que reduce la colaboración entre organizaciones, que es costosa y requiere muchos recursos. Los costos de cumplimiento suelen ser bajos porque el proveedor de SaaS expone solo los servicios que desea y no toda la red.

Costos de Amazon VPC Lattice

Amazon VPC Lattice ofrece una estructura de costes equilibrada con costes de infraestructura fijos y variables moderados. Al tratarse de una red de servicios totalmente gestionada, reduce considerablemente la sobrecarga operativa al automatizar la detección de servicios, la gestión del tráfico y los controles de acceso en múltiples áreas. VPCs Esto simplifica tanto la implementación inicial como la administración continua en comparación con las configuraciones de red manuales. Puede implementar cambios mediante controles basados en políticas sin complejas actualizaciones de enrutamiento, lo que reduce la dependencia de los especialistas en redes. Los costos de cumplimiento suelen ser más bajos que los enfoques de redes tradicionales porque VPC Lattice proporciona controles de acceso detallados y una visibilidad integral a través de capacidades integradas de monitoreo y registro. Esto puede facilitar la demostración del cumplimiento normativo.

AWS Transit Gateway costes

AWS Transit Gateway tiene cargos por hora y procesamiento de datos más altos que AWS PrivateLink, pero tiene una sobrecarga operativa similar. Debe tener un conocimiento más profundo

del AWS Transit Gateway servicio y el AWS enrutamiento para poder configurar correctamente todas las tablas de rutas. Los cambios en la infraestructura pueden requerir actualizaciones de enrutamiento o DNS. Los costos de cumplimiento son similares a los de la interconexión de VPC, ya que ambas partes podrían exponer las subredes o la totalidad VPCs entre sí. AWS Transit Gateway Las tablas de enrutamiento también deben gestionarse con cuidado, ya que son compartidas por varios consumidores y no se debe permitir ningún tráfico entre ellas.

AWS Site-to-Site VPN costes

Debido a que la Site-to-Site VPN básicamente envía tráfico a Internet, el costo variable es más alto en comparación con los cargos por transferencia de datos. Si bien es un servicio de red privada virtual (VPN) gestionada, conlleva una sobrecarga operativa significativa, especialmente en la pasarela de clientes. El aprovisionamiento y las operaciones requieren conocimientos avanzados de redes, y los cambios suelen requerir la acción de ambas partes. Los costes de conformidad suelen ser bajos porque los equipos de seguridad suelen aprobar previamente los IPsec túneles sin necesidad de una revisión adicional.

AWS Direct Connect costes

AWS Direct Connect tiene el mayor costo de infraestructura fija porque se trata de una conexión física privada directamente al Nube de AWS. Se requieren conocimientos especializados para configurar y operar una sesión del Border Gateway Protocol (BGP) (si es necesario), operar una conexión VPN y realizar tareas de ingeniería de tráfico. Este servicio reduce el esfuerzo de los equipos de seguridad porque combina la conectividad privada con la opción de contar, además, con el control de acceso a los medios, la seguridad (MACsec) y el IPsec cifrado.

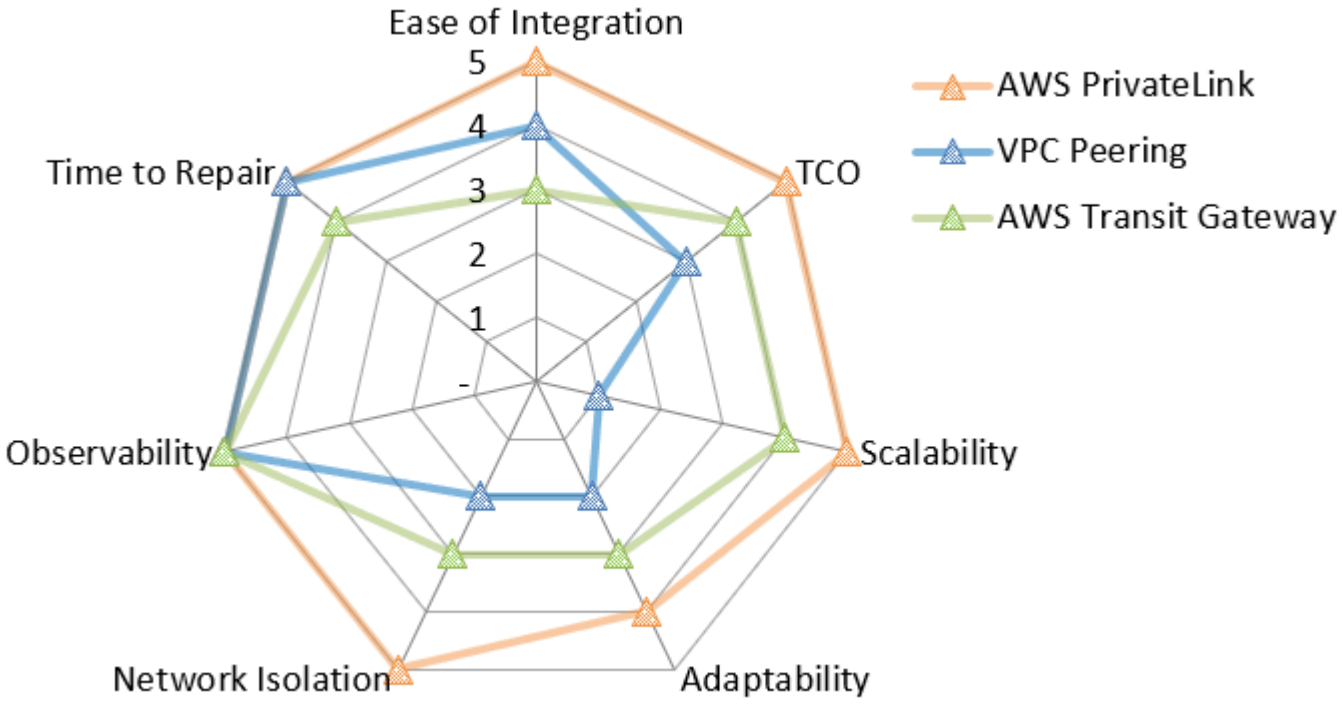
Costos de acceso público a Internet

El acceso público a Internet se refiere a los AWS recursos que puede usar para hacer que una aplicación sea accesible públicamente, como un Application Load Balancer. Para este enfoque, existen costos variables relacionados con la prestación de acceso a tus servicios, incluidos los cargos por la transferencia de datos a Internet. Los gastos operativos y de cumplimiento pueden ser importantes porque se expone el servicio a Internet y se requieren mecanismos de seguridad y autenticación adicionales. Sin embargo, no implica un enrutamiento complejo y ninguna de las partes tiene que conocer los detalles de la infraestructura de la otra parte.

Mapa de valores de redes

Para ayudarlo a ver el panorama general y tomar decisiones informadas, esta guía incluye un mapa de valores de redes para cada escenario. Como las calificaciones varían de un escenario a otro, es posible que el mismo servicio obtenga una puntuación diferente en dos escenarios. Los mapas de valores son gráficos radiales, en los que la puntuación perfecta hipotética sería de cinco en todas las categorías.

Por ejemplo, en la siguiente imagen se muestra un ejemplo de gráfico radial. Incluye solo las métricas que podemos ayudar a evaluar. Le recomendamos que cree su propio mapa de valores que incluya las métricas adicionales que solo usted puede evaluar.



Escenarios de acceso a redes para ofertas de SaaS en el Nube de AWS

En esta sección se describen las diferentes opciones de acceso a la red para sus ofertas de SaaS en. Nube de AWS Analiza los enfoques desde la perspectiva de su consumidor, que podría tener necesidades de conectividad dentro Nube de AWS, desde centros de datos locales o desde otros proveedores de servicios en la nube ()CSPs. Además, es posible que necesite admitir el acceso desde varios tipos de entornos de consumo.

Comprender los requisitos de conectividad de red en estos diversos entornos es esencial para crear una estrategia de acceso integral. Sus decisiones de arquitectura deben tener en cuenta los diferentes modelos de seguridad, las expectativas de rendimiento y las limitaciones técnicas, al tiempo que mantienen la eficiencia operativa. El enfoque correcto proporciona una conectividad segura y confiable que se adapta al crecimiento de su empresa y minimiza tanto la complejidad de la implementación como la sobrecarga de administración continua.

Al evaluar las opciones de acceso a la red, tenga en cuenta cómo afecta cada enfoque al costo total de propiedad, incluidos no solo los costos de infraestructura, sino también los gastos operativos y los requisitos de cumplimiento. Algunos enfoques destacan por su escalabilidad, pero pueden introducir complejidad, mientras que otros dan prioridad a la facilidad de integración en detrimento del aislamiento de la red. Las capacidades y los recursos técnicos de sus consumidores también desempeñan un papel importante a la hora de determinar la solución más adecuada.

Para los consumidores extranjeros Nube de AWS, estos servicios AWS PrivateLink ofrecen importantes ventajas en materia de seguridad y escalabilidad. Los consumidores locales podrían beneficiarse de AWS Direct Connect un rendimiento uniforme o de una Site-to-Site VPN para una conectividad rentable. Los escenarios multinube suelen requerir una consideración cuidadosa de los desafíos de interoperabilidad, y puede utilizar arquitecturas de VPC de tránsito para estandarizar los patrones de acceso. En todos los casos, su diseño debe anticipar el crecimiento futuro de los consumidores y el tráfico para que su arquitectura de red siga siendo resiliente y adaptable a medida que evolucione su oferta de SaaS.

Esta sección contiene los siguientes escenarios:

- Los consumidores de SaaS que operan en AWS
- Consumidores de servicios que operan en las instalaciones
- Consumidores de SaaS que operan en otros proveedores de servicios en la nube

- Compatible con entornos híbridos

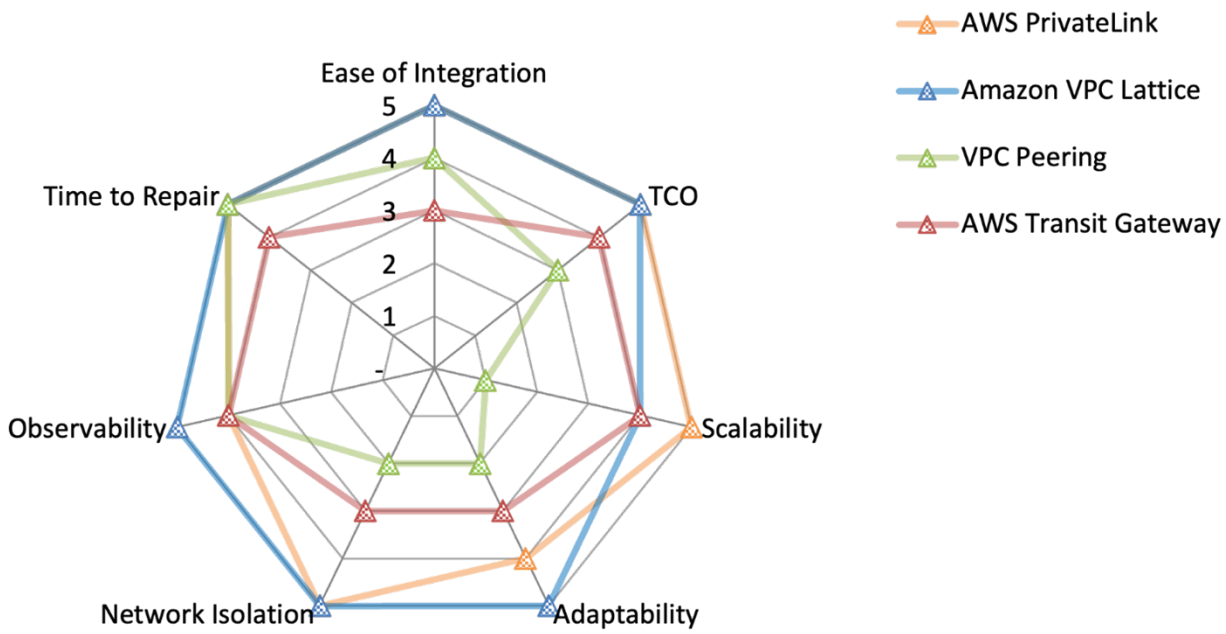
Los consumidores de SaaS que operan en AWS

En esta sección se analizan las opciones de conectividad si tanto usted como sus consumidores operan en el. Nube de AWS Este escenario ofrece la mayor flexibilidad porque muchas se integran de Servicios de AWS forma nativa y porque ambas partes tienen acceso a toda la Servicio de AWS cartera.

En esta sección se analizan los siguientes enfoques de acceso a la red:

- Integrating AWS PrivateLink with
- Compartir un servicio de Amazon VPC Lattice
- Creación de conexiones de emparejamiento de VPC
- Conectarse VPCs con AWS Transit Gateway

El siguiente mapa de valores de red resume la puntuación de cada una de estas opciones para cada métrica de evaluación. Para obtener más información sobre las métricas de evaluación, consulte las métricas de evaluación en esta guía. En el mapa, un cinco representa la mejor puntuación, por ejemplo, el menor TCO, el mejor aislamiento de la red o el menor tiempo de reparación. Para obtener más información sobre cómo leer este gráfico radial, consulte Mapa de valores de redes esta guía.



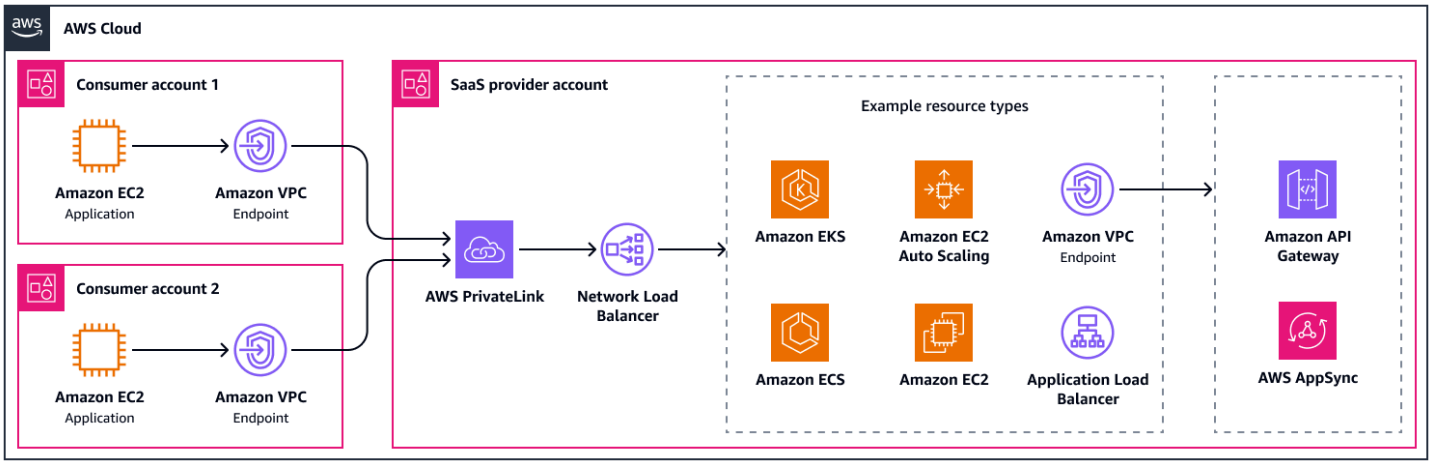
El gráfico radial muestra los siguientes valores.

Métrica de evaluación	AWS PrivateLink	Amazon VPC Lattice	Emparejamiento de VPC	AWS Transit Gateway
Facilidad de integración	5	5	4	3
TCO	5	5	3	4
Escalabilidad	5	4	1	4
Adaptabilidad	4	5	2	3
Aislamiento de red	5	5	2	3
Observabilidad	4	5	4	4
Es hora de reparar	5	5	5	4

Integrating AWS PrivateLink with

AWS PrivateLinkes la forma más nativa de la nube de integrar una oferta de SaaS. Los proveedores de SaaS pueden alojar sus aplicaciones o bien detrás de un Network Load Balancer. El balanceador de carga de red se integra directamente con un balanceador de carga de aplicaciones, grupos de Amazon Elastic Container Service (Amazon ECS), Amazon Elastic Kubernetes Service (Amazon EKS) y Auto Scaling. También es posible enrutar el tráfico desde el Network Load Balancer a los puntos finales de la VPC de interfaz en la cuenta del proveedor de SaaS. Esto le ayuda a usar una API para llegar a las aplicaciones, por ejemplo, a través de Amazon API Gateway o AWS AppSync. Si tu aplicación requiere acceso a recursos del entorno del cliente que no tienen un equilibrio de carga, como una base de datos, puedes usar puntos finales de VPC de recursos.

AWS PrivateLink admite un ancho de banda de hasta 100 Gbps por zona de disponibilidad. El siguiente diagrama muestra una configuración básica con algunas posibles integraciones. Conecta dos cuentas de consumidor a la cuenta del proveedor de SaaS a través de. AWS PrivateLink Hay puntos finales de servicio en las cuentas de los consumidores y un Network Load Balancer en la cuenta del proveedor de SaaS.



Los siguientes son beneficios de este enfoque:

- Facilidad de integración: no es necesario cambiar la tabla de enrutamiento
- Facilidad de integración: puede ofrecer servicios de punto final a través de AWS Marketplace
- Facilidad de integración: los puntos de conexión de VPC admiten nombres DNS fáciles de entender
- Escalabilidad: puede ampliarse a miles de consumidores de SaaS
- Adaptabilidad: Support para rangos CIDR superpuestos

- Adaptabilidad: Support for IPv6
- Adaptabilidad: soporte interregional
- TCO: AWS PrivateLink es un servicio totalmente gestionado, por lo que requiere menos esfuerzo operativo
- Aislamiento de la red: beneficio de seguridad para el consumidor de SaaS porque el tráfico no puede iniciarse desde el proveedor de SaaS
- Aislamiento de la red: beneficio de seguridad para el proveedor de SaaS porque no expone una subred o VPC completa

Los inconvenientes de este enfoque son los siguientes:

- Adaptabilidad: el proveedor de SaaS debe usar las mismas zonas de disponibilidad que el consumidor
- Adaptabilidad: Support solo para conexiones iniciadas por el cliente y se requieren puntos finales de VPC de recursos para la comunicación iniciada por el servicio
- Adaptabilidad: Network Load Balancer es la única integración directa para AWS PrivateLink

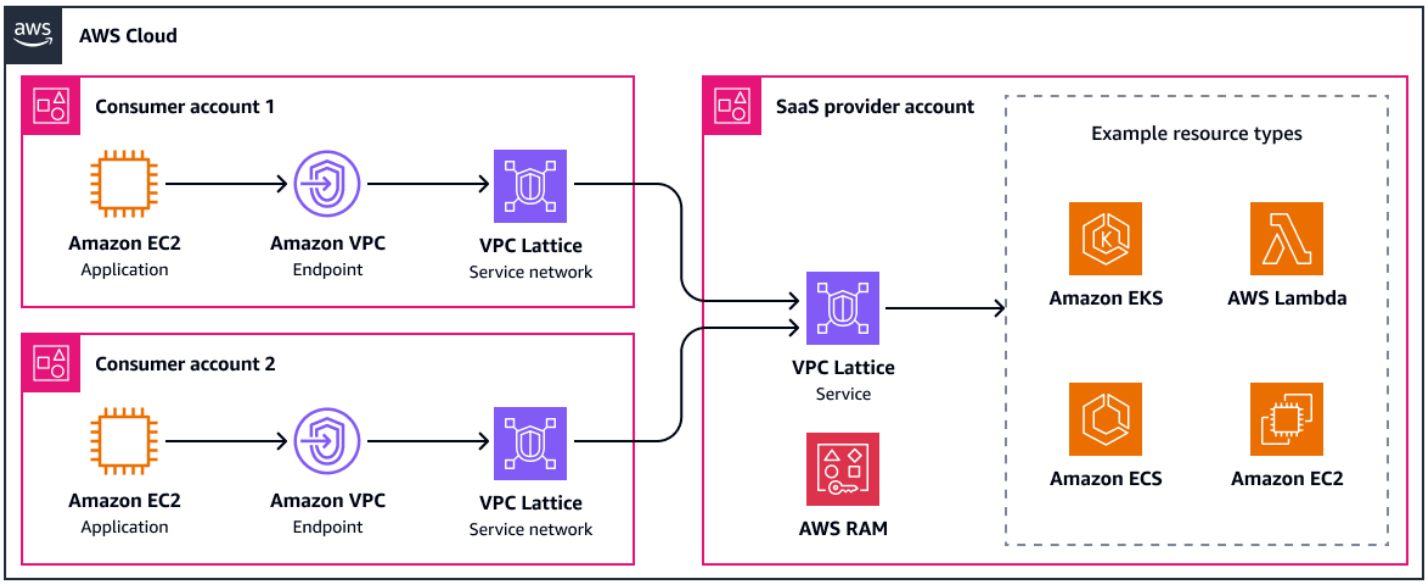
Compartir un servicio de Amazon VPC Lattice

Para utilizar Amazon VPC Lattice como opción de conectividad para su aplicación SaaS, primero debe crear uno o más servicios de VPC Lattice que representen los componentes de la aplicación SaaS. Usted configura los oyentes y las reglas de enrutamiento para dirigir el tráfico a sus destinos de backend, como instancias, contenedores o funciones de Amazon EC2. AWS Lambda Para obtener más información, consulte Conexión de servicios SaaS dentro de una red de servicios de VPC Lattice AWS (entrada del blog). Desde el punto de vista conceptual, es casi lo mismo que configurar un Application Load Balancer. Luego, comparte su servicio SaaS de forma segura con los clientes Cuentas de AWS o las organizaciones mediante AWS Resource Access Manager (AWS RAM), especificando los permisos que tienen. Una vez que los clientes acepten el recurso compartido, pueden asociar su servicio SaaS a sus redes de servicios VPC Lattice existentes o recién creadas para permitir la comunicación. service-to-service

Cada servicio de VPC Lattice puede admitir hasta 10 Gbps y 10 000 solicitudes por segundo por zona de disponibilidad. Al implementar políticas de autenticación, sus clientes pueden tener un control detallado sobre qué servicios y recursos pueden acceder a la aplicación SaaS. Puede usar pasarelas de recursos para acceder a los recursos que requieren una conexión TCP. Por ejemplo,

puede ser un clúster de Amazon EKS que usted administre o un recurso administrado por el cliente al que su aplicación necesite acceder. Para obtener más información sobre el uso de pasarelas de recursos para las ofertas de SaaS, consulte Ampliar las capacidades de SaaS AWS PrivateLink al Cuentas de AWS uso del soporte para los recursos de VPC (entrada del blog).AWS

El siguiente diagrama muestra una configuración de entramado de VPC de alto nivel con algunos ejemplos de integraciones. Utiliza redes de servicios gestionadas por el cliente para acceder a la aplicación SaaS.



Los siguientes son beneficios de este enfoque:

- Facilidad de integración: no es necesario cambiar la tabla de enrutamiento
- Facilidad de integración: detección de servicios lista para usar
- Escalabilidad: puede ampliarse a miles de consumidores de SaaS
- Adaptabilidad: Support para rangos CIDR superpuestos
- Adaptabilidad: Support for IPv6
- Adaptabilidad: se integra con cualquier servicio de AWS cómputo como un servicio de VPC Lattice
- TCO: VPC Lattice es un servicio totalmente gestionado, por lo que requiere menos esfuerzo operativo
- TCO: equilibrio de carga integrado con enrutamiento de tráfico avanzado
- Aislamiento de la red: autorización detallada con políticas de autenticación
- Aislamiento de la red: beneficio de seguridad para el consumidor de SaaS porque el tráfico no puede iniciarse desde el proveedor de SaaS

- Aislamiento de la red: beneficio de seguridad para el proveedor de SaaS porque no se expone una subred o VPC completa

Los inconvenientes de este enfoque son los siguientes:

- Adaptabilidad: Support solo para conexiones iniciadas por el cliente y se requieren pasarelas de recursos para la comunicación iniciada por el servicio
- Adaptabilidad: no es compatible con otras regiones

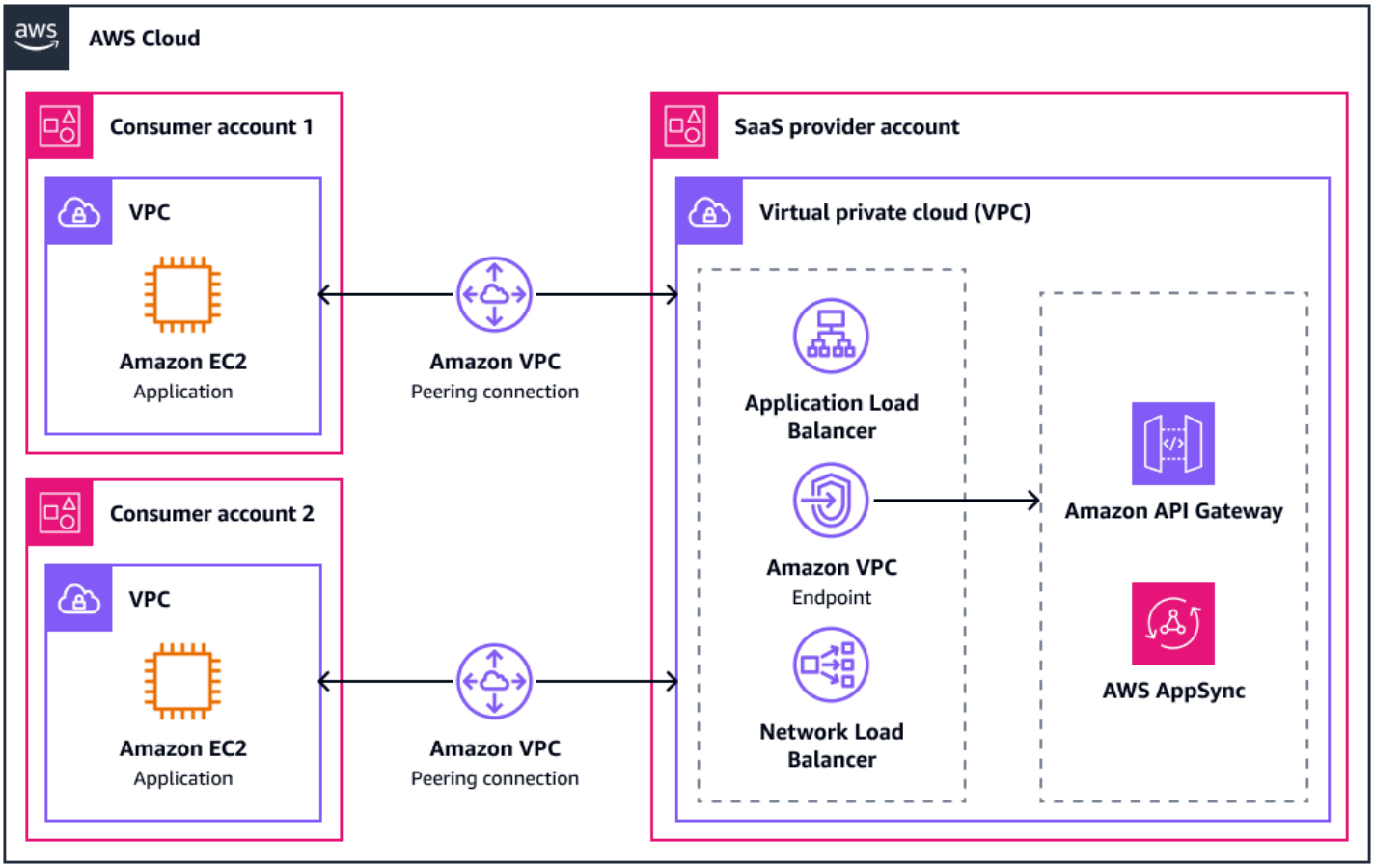
Creación de conexiones de emparejamiento de VPC

Cuando se utiliza la interconexión de VPC para conectar la VPC del proveedor de SaaS con la VPC del consumidor, ambas partes pueden iniciar las conexiones. Esto requiere una configuración adecuada de los grupos de seguridad, los firewalls y las listas de control de acceso a la red () en ambas cuentas. NACLs De lo contrario, el tráfico no deseado podría entrar en la red a través de la conexión de interconexión. Puede usar los grupos de seguridad para hacer referencia a los grupos de seguridad desde los grupos de seguridad interconectados VPCs. Esto puede ayudarle a controlar el acceso a su aplicación, ya que los grupos de seguridad que incluyen listas de usuarios permitidos proporcionan un control de acceso más explícito y detallado en comparación con las direcciones IP que sí están permitidas.

Con la interconexión de VPC, se puede acceder a la oferta de SaaS a través de un servicio o recurso implementado en la VPC. La mayoría de las aplicaciones SaaS se basan en un Application Load Balancer o Network Load Balancer. AWS AppSync private APIs o Amazon API Gateway private APIs son otros puntos de entrada comunes a las aplicaciones SaaS, ya que pueden ser un objetivo a través de una conexión de pares a través de los puntos de enlace de la interfaz de VPC.

Tras establecer una conexión de emparejamiento, debe actualizar las tablas de rutas de ambas cuentas para definir VPCs la conexión de emparejamiento como el siguiente salto del rango CIDR correspondiente. Esta solución se recomienda solo para los proveedores de SaaS que tienen pocos consumidores, ya que la gestión de varias conexiones entre pares se vuelve demasiado compleja rápidamente.

El siguiente diagrama muestra una configuración básica con algunas posibles integraciones. VPCs en dos cuentas de consumidores, tienen una conexión de emparejamiento con una VPC en la cuenta del proveedor de SaaS.



Los siguientes son beneficios de este enfoque:

- Tiempo de reparación: no existe un punto único de fallo en la comunicación
- Escalabilidad: sin limitaciones de ancho de banda en comparación con la interconexión de VPC
- TCO: la conexión de emparejamiento o el tráfico a través de la conexión de emparejamiento dentro de la misma zona de disponibilidad son gratuitos
- TCO: no hay infraestructura que administrar
- Adaptabilidad: Support for IPv6
- Adaptabilidad: se admite la interconexión interregional

Los inconvenientes de este enfoque son los siguientes:

- Adaptabilidad: no se admite el enrutamiento transitivo
- Adaptabilidad: no se admiten rangos de CIDR superpuestos
- Escalabilidad: escalabilidad limitada (máximo 125 conexiones de emparejamiento por VPC)

- TCO: la complejidad crece exponencialmente con cada conexión de emparejamiento adicional
- TCO: gastos generales derivados de la administración de las tablas de rutas, las propias conexiones entre pares, las reglas de los grupos de seguridad y la inspección del tráfico
- Aislamiento de la red: se requieren controles de seguridad estrictos porque ambas partes VPCs están expuestas en su totalidad

Conectarse VPCs con AWS Transit Gateway

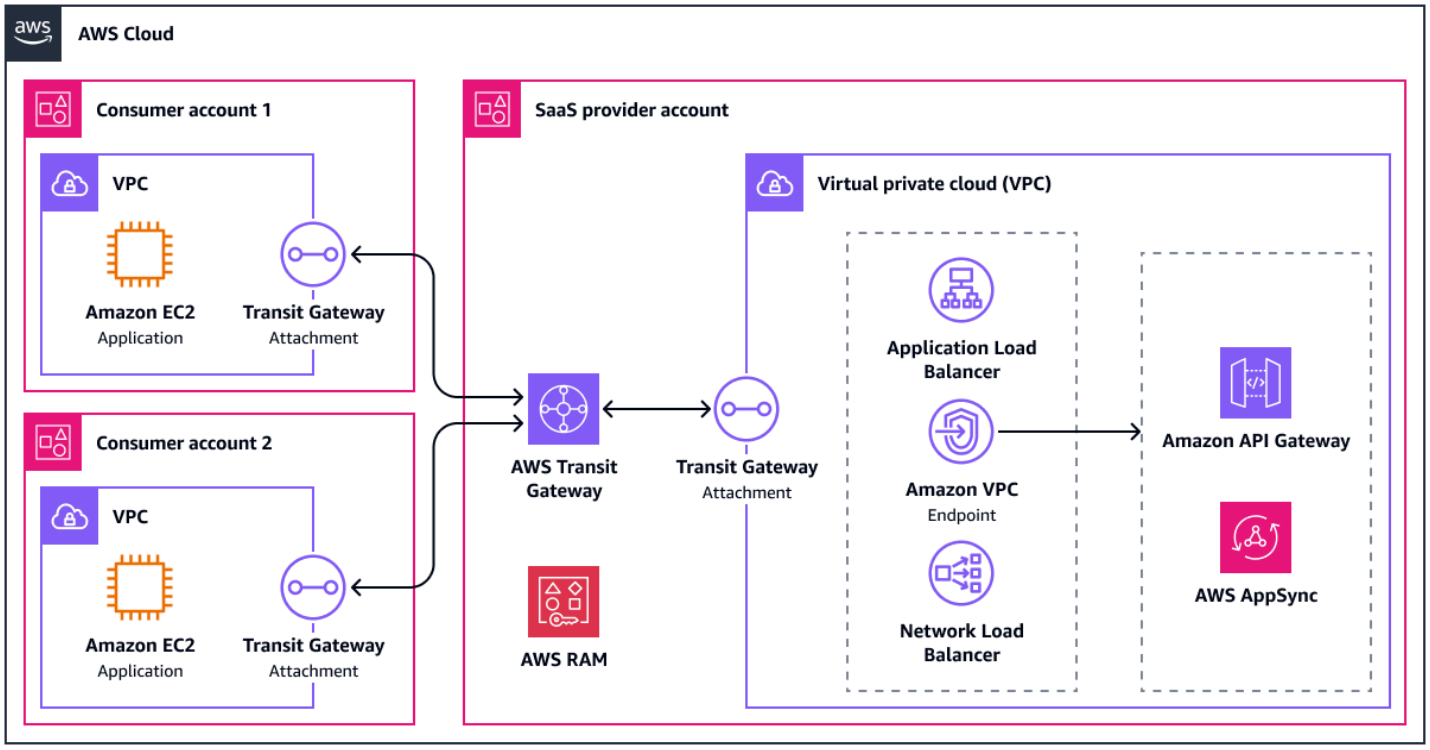
Cuando te conectas VPCs AWS Transit Gateway, crea adjuntos de VPC e implementa interfaces de red en las subredes de cada zona de disponibilidad que deberían enrutar el tráfico hacia y desde la VPC. Se recomienda tener una `/28` subred dedicada en cada zona de disponibilidad para el adjunto de la VPC. Para obtener más información, consulte las prácticas recomendadas de diseño de Amazon VPC Transit Gateways. VPCs Necesitan una tabla de rutas actualizada para enviar el tráfico a través de la interfaz de red implementada y las tablas de rutas de Transit Gateway deben actualizarse en consecuencia. En una configuración de varios inquilinos, desea que la VPC del proveedor de SaaS tenga una ruta a la de todos los consumidores. VPCs El consumidor VPCs debe tener una ruta únicamente hacia la VPC del proveedor de SaaS.

Transit Gateway tiene un diseño de alta disponibilidad. Admite la supervisión con registros de flujo de VPC y el ancho de banda máximo para un adjunto de Transit Gateway es de 100 Gbps por zona de disponibilidad. Al igual que la interconexión de VPC, este enfoque permite hacer referencia a grupos de seguridad entre VPC, lo que simplifica el control de acceso entre los entornos.

Hay dos opciones principales para conectar a los consumidores con su oferta de SaaS con Transit Gateway.

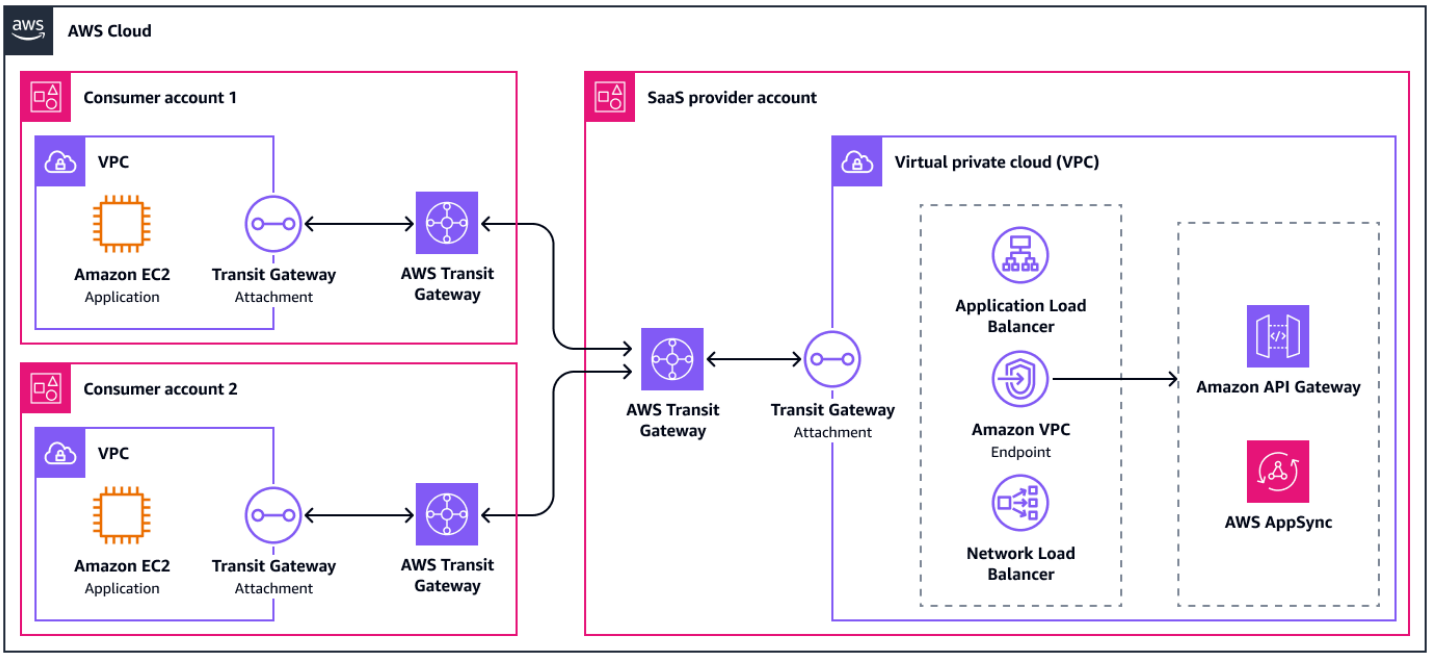
Opción 1: usar RAM

En la primera opción, el proveedor de servicios comparte la Transit Gateway con los consumidores mediante AWS Resource Access Manager (AWS RAM). Esto permite a los consumidores implementar los archivos adjuntos de la VPC en sus propias cuentas. El siguiente diagrama muestra esta opción en un nivel superior.



Opción 2: pasarelas de tránsito interconectadas

La segunda opción es vincular tu pasarela de transporte público con una pasarela de transporte público en las cuentas de los consumidores. Esto proporciona a los consumidores más flexibilidad, ya que ahora pueden controlar completamente las tablas de rutas dentro de su pasarela de transporte público. Por ejemplo, podrían configurar una inspección centralizada entre el servicio y sus cargas de trabajo. Una desventaja de esta opción es que solo se admite el enrutamiento estático entre las pasarelas de tránsito. El siguiente diagrama muestra esta opción en un nivel alto.



Los siguientes son beneficios de este enfoque:

- Escalabilidad: Support para hasta 5000 archivos adjuntos
- Escalabilidad: un solo lugar para administrar y monitorear todas las conexiones VPCs
- Adaptabilidad: Transit Gateway también se puede conectar a VPNs Direct Connect pasarelas y dispositivos SD-WAN de terceros
- Adaptabilidad: arquitectura flexible, como añadir una VPC de inspección
- Adaptabilidad: Support for transitive routing
- Adaptabilidad: ¿Pueden emparejarse las pasarelas de tránsito intrarregionales e interregionales
- Adaptabilidad: Support for IPv6
- TCO: AWS Transit Gateway es un servicio totalmente gestionado, por lo que requiere menos esfuerzo operativo
- TCO: el TCO crece de forma lineal con cada conexión adicional a la pasarela de tránsito

Los inconvenientes de este enfoque son los siguientes:

- Facilidad de integración: la configuración del enrutamiento requiere conocimientos avanzados de redes
- Adaptabilidad: no se admiten rangos de CIDR superpuestos

- TCO: sobrecarga derivada de la administración de las entradas de las tablas de rutas, las reglas de los grupos de seguridad y la inspección del tráfico
- Seguridad: se requieren controles de seguridad estrictos porque ambas partes están expuestas en su totalidad VPCs

Consumidores de servicios que operan en las instalaciones

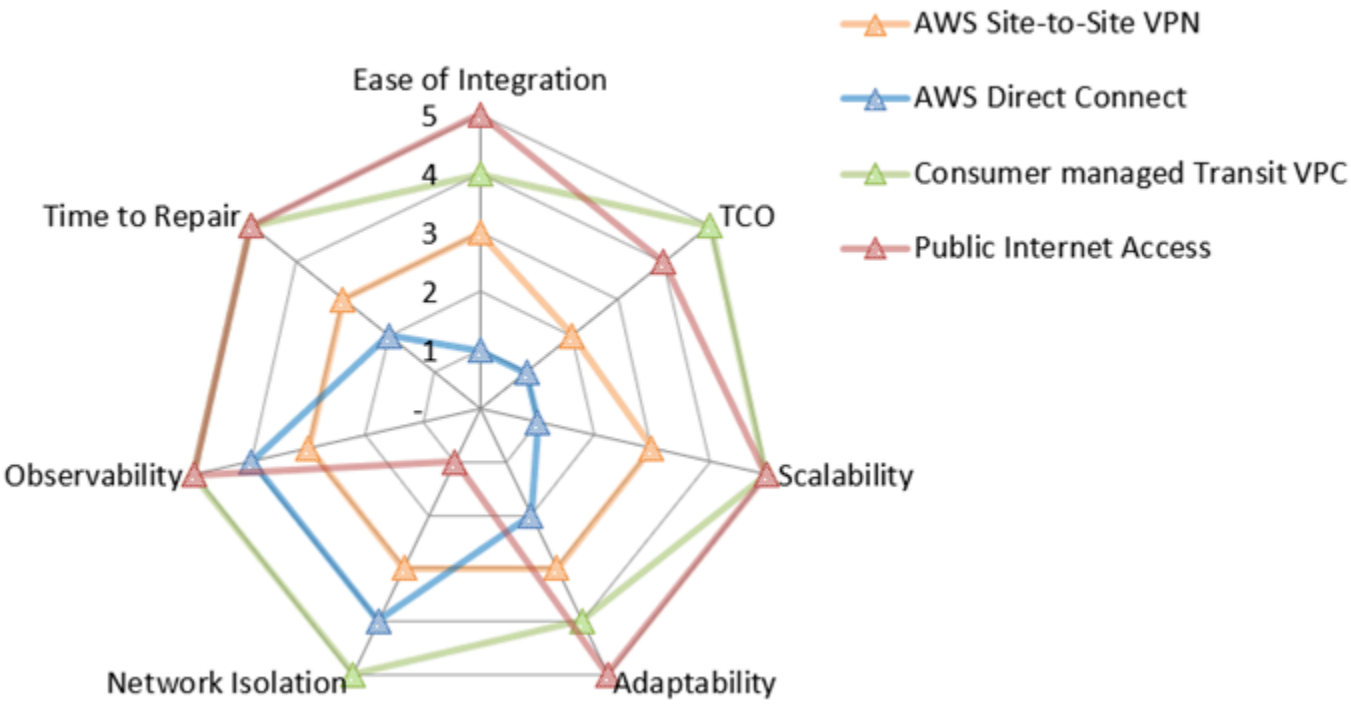
En esta sección, se analizan las opciones de conectividad entre las cargas de trabajo de SaaS en los centros de datos Nube de AWS locales y los locales. Muchos consumidores con requisitos locales, especialmente a nivel empresarial, ven la nube como una extensión de su red física y quieren que eso se refleje en su arquitectura. Esto significa conectividad privada a la oferta de SaaS en la nube, ya sea a través de túneles lógicos o incluso a través de una conexión física privada. Otros consumidores aceptarán la conectividad a través de Internet pública, algo que también se analiza en esta sección.

En esta sección se analizan los siguientes enfoques de acceso a la red:

- Conectando con AWS Site-to-Site VPN
- Conectarse con AWS Direct Connect
- Conexión con una arquitectura de VPC de tránsito
- Conexión a través de la Internet pública

El siguiente mapa de valores de red resume la puntuación de cada una de estas opciones para cada métrica de evaluación. Para obtener más información sobre las métricas de evaluación, consulte las métricas de evaluación en esta guía. En el mapa, un cinco representa la mejor puntuación, por ejemplo, el menor TCO, el mejor aislamiento de la red o el menor tiempo de reparación. Para obtener más información sobre cómo leer este gráfico radial, consulte Mapa de valores de redes esta guía.

> **Note**
>
> Se excluye la opción de VPC de transporte gestionada por el proveedor porque las puntuaciones dependen en gran medida de los servicios que se operen.



El gráfico radial muestra los siguientes valores.

Métrica de evaluación	AWS Site-to-Site VPN	AWS Direct Connect	VPC de tránsito gestionada por el consumidor	Acceso público a Internet
Facilidad de integración	3	1	4	5
TCO	2.	1	5	4
Escalabilidad	3	1	5	5
Adaptabilidad	3	2	4	5
Aislamiento de red	3	4	5	1
Observabilidad	3	4	5	5

Es hora de reparar	3	2	5	5

Conectando con AWS Site-to-Site VPN

AWS Site-to-Site VPNlas conexiones pueden terminar en una puerta de enlace privada virtual o en una puerta de enlace de tránsito. Una puerta de enlace privada virtual es el punto final de la VPN situado en el AWS lateral de la conexión Site-to-Site VPN que se puede conectar a una sola VPC. Una puerta de enlace de tránsito es un centro de tránsito que se puede usar para interconectar redes múltiples VPCs y locales. También se puede usar como punto final de VPN para el AWS lado de la conexión Site-to-Site VPN. En esta sección se analizan ambas opciones.

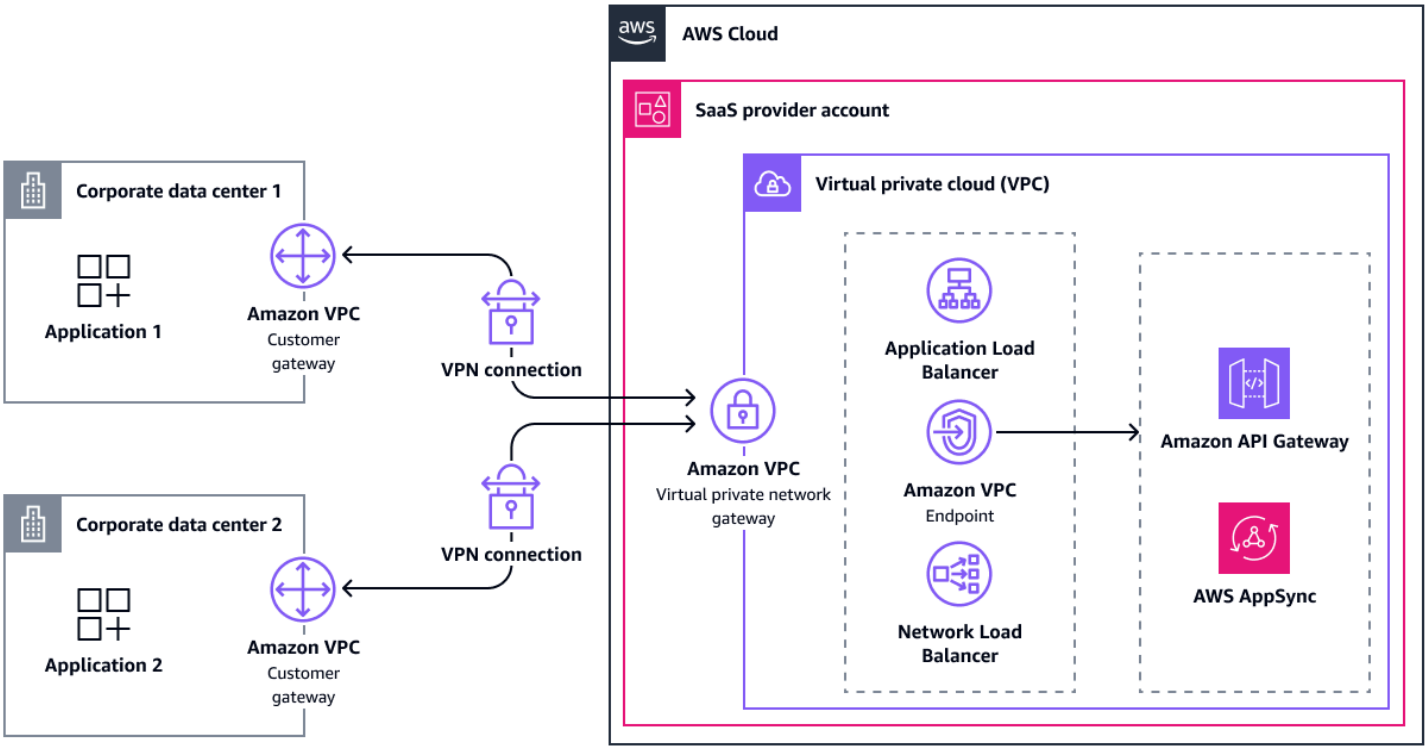
Conexión a través de una puerta de enlace privada virtual

Después de crear una puerta de enlace privada virtual, debe adjuntarla a la VPC que contiene su oferta de SaaS. A continuación, habilita la propagación de rutas para propagar las rutas de la VPN a la tabla de rutas de la VPC. Esas rutas pueden ser rutas estáticas o dinámicas anunciadas por BGP.

Para una alta disponibilidad, una conexión Site-to-Site VPN tiene dos túneles VPN que terminan en dos zonas de disponibilidad laterales. AWS Si uno deja de estar disponible, el segundo túnel puede tomar el control. Un único túnel permite un ancho de banda máximo de 1,25 Gbps. Como las puertas de enlace privadas virtuales no admiten el enrutamiento de rutas múltiples (ECMP) de igual costo, solo puede usar un túnel a la vez.

Para aumentar la tolerancia a los errores, puede configurar una segunda conexión VPN a una segunda puerta de enlace física para el cliente. Una vez establecida la conexión, el consumidor puede acceder a los recursos de la VPC del proveedor de SaaS.

El siguiente diagrama muestra esta arquitectura.



Los siguientes son beneficios de este enfoque:

- Tiempo de reparación: se gestionó la conmutación por error al túnel VPN secundario
- Observabilidad: integración para la supervisión activa gestionada mediante Network Synthetic Monitor
- Facilidad de integración: soporte de enrutamiento dinámico a través de BGP
- Adaptabilidad: compatibilidad con la mayoría de los equipos de red locales
- Adaptabilidad: soporte IPv6
- TCO: AWS Site-to-Site VPN es un servicio totalmente gestionado, por lo que requiere menos esfuerzo operativo
- TCO: las pasarelas virtuales no tienen costo alguno, aunque hay cargos por las dos direcciones públicas IPv4 de cada una
- Aislamiento de la red: permite una comunicación privada segura a través de Internet

Los inconvenientes de este enfoque son los siguientes:

- Facilidad de integración: el consumidor debe configurar su pasarela de clientes

- Escalabilidad: la falta de compatibilidad con ECMP limita el ancho de banda a 1,25 Gbps por puerta de enlace virtual
- Escalabilidad: escalabilidad limitada debido al aumento de la complejidad de la red y de la sobrecarga operativa
- Adaptabilidad: solo IPv6 admite las direcciones IP internas de los túneles VPN
- Adaptabilidad: sin enrutamiento transitivo
- TCO: sobrecarga operativa para mantener, administrar y configurar numerosas conexiones VPN para el proveedor de SaaS

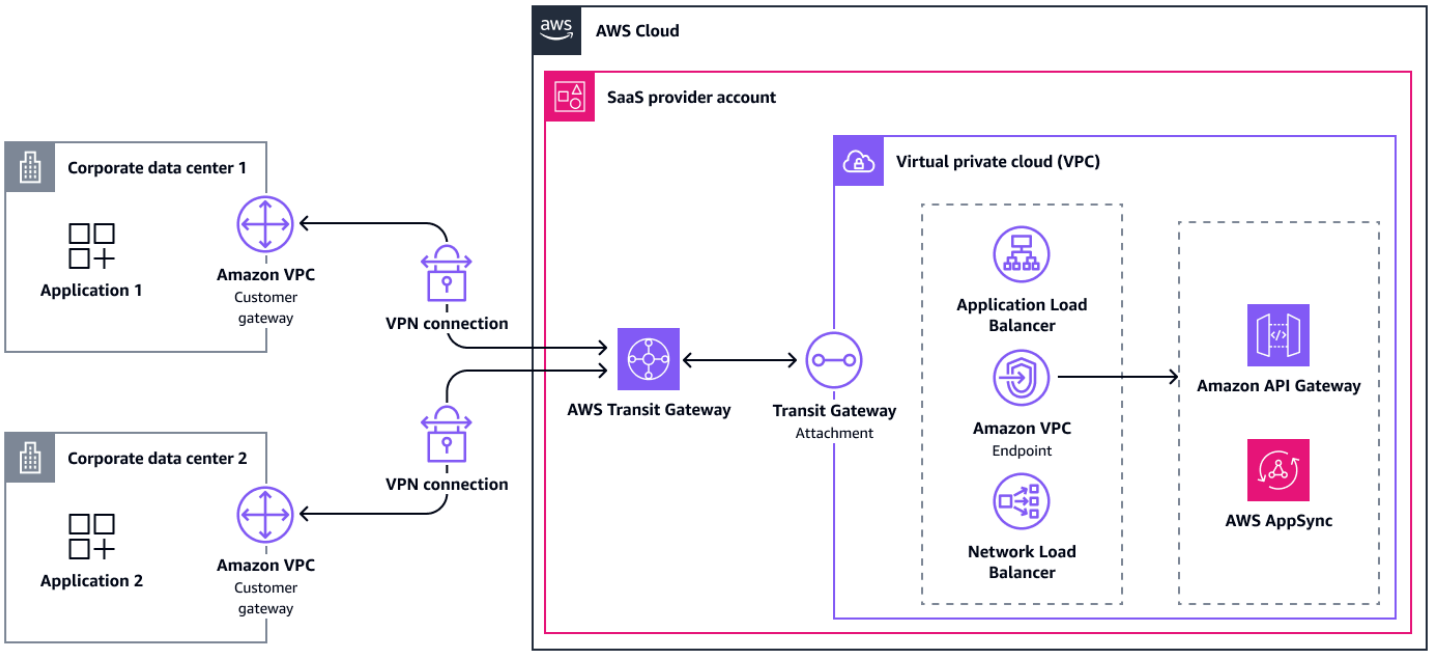
Conexión a través de una pasarela de tránsito

Las conexiones a través de las pasarelas de tránsito son similares a las pasarelas virtuales. Sin embargo, hay algunas diferencias que hay que tener en cuenta.

En primer lugar, las rutas del adjunto de la VPN se pueden propagar automáticamente dentro de la tabla de rutas de la pasarela de tránsito, pero hay que añadir manualmente las rutas a la adjunta VPCs.

En comparación con una puerta de enlace virtual, Transit Gateway es compatible con ECMP. Si la pasarela del cliente es compatible con el ECMP, puede utilizar ambos túneles para lograr un rendimiento máximo total de 2,5 Gbps. Puede establecer varias conexiones entre la misma red local y la puerta de enlace de tránsito. Con este enfoque, puede aumentar el ancho de banda máximo hasta 2,5 Gbps por conexión.

El siguiente diagrama muestra esta arquitectura.



Los siguientes son beneficios de este enfoque:

- Tiempo de reparación: se gestionó la conmutación por error al túnel VPN secundario
- Observabilidad: integración para la supervisión activa gestionada mediante Network Synthetic Monitor
- Facilidad de integración: soporte de enrutamiento dinámico a través de BGP
- Escalabilidad: la compatibilidad con ECMP permite escalar el rendimiento de la VPN para satisfacer los requisitos de un gran ancho de banda
- Escalabilidad: una única puerta de enlace de tránsito admite un gran número de conexiones VPN (hasta casi 5000)
- Escalabilidad: un lugar para administrar y monitorear todas las conexiones VPN
- Adaptabilidad: compatibilidad con la mayoría de los equipos de red locales
- Adaptabilidad: soporte IPv6
- Adaptabilidad: hereda la flexibilidad de AWS Transit Gateway
- TCO: AWS Transit Gateway es un servicio totalmente gestionado, por lo que requiere menos esfuerzo operativo
- TCO: las pasarelas virtuales no tienen costo alguno, aunque hay cargos por las dos direcciones públicas IPv4 de cada una
- Aislamiento de la red: permite una comunicación privada segura a través de Internet

Los inconvenientes de este enfoque son los siguientes:

- Facilidad de integración: el consumidor debe configurar su pasarela de clientes
- Escalabilidad: escalabilidad limitada debido al aumento de la complejidad de la red y de la sobrecarga operativa
- Adaptabilidad: solo IPv6 admite las direcciones IP internas de los túneles VPN
- TCO: sobrecarga operativa para mantener, administrar y configurar numerosas conexiones VPN para el proveedor de SaaS
- TCO: cargos adicionales por el uso de AWS Transit Gateway
- TCO: complejidad adicional de administrar las tablas de rutas de las pasarelas de tránsito

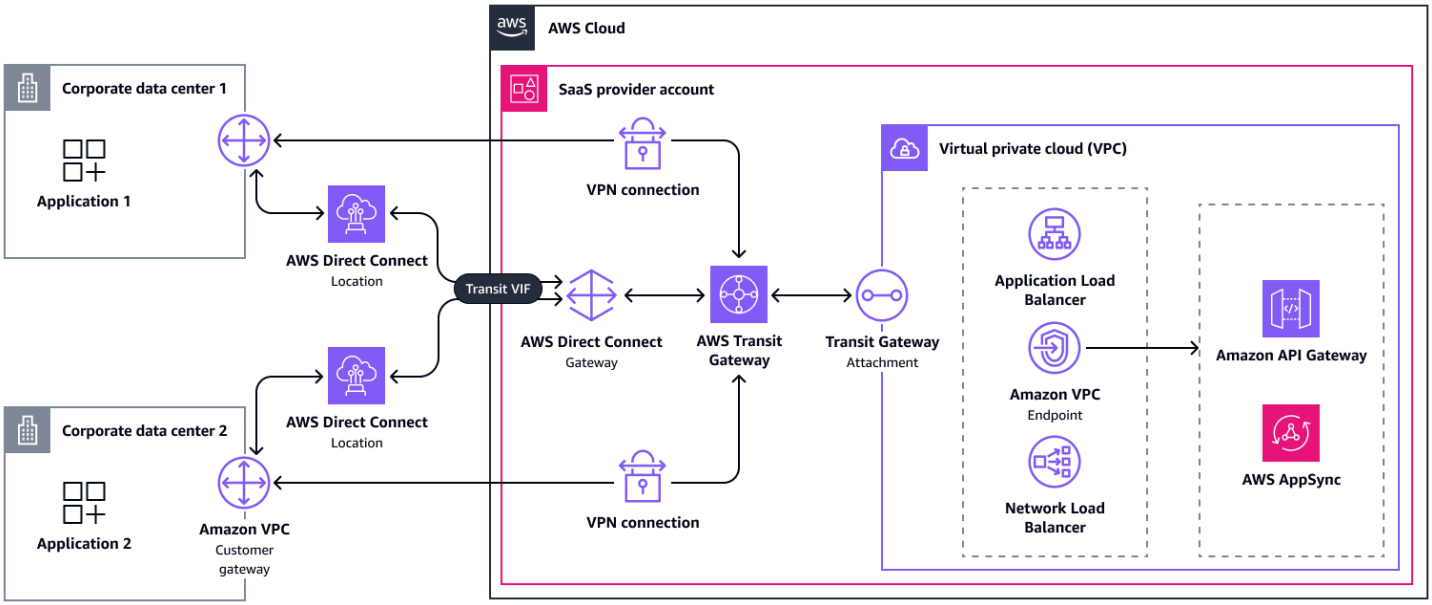
Conectarse con AWS Direct Connect

AWS Direct Connectconecta su red interna a una Direct Connect ubicación a través de un cable de fibra óptica Ethernet estándar. A diferencia de las otras opciones de arquitectura, no se puede establecer una conexión dedicada en unos minutos. En cambio, este proceso puede tardar varios días si se cumplen todos los requisitos. Si no es así, puede que tarde más. Por lo tanto, le sugerimos que se ponga en contacto con su equipo de AWS cuentas o AWS Support pida ayuda con este enfoque. Si lo desea, puede elegir una conexión alojada proporcionada por un AWS socio y compartida con otros clientes. En cualquier caso, la arquitectura es la misma. Puede elegir Direct Connect porque reduce la latencia, mejora el ancho de banda o cumple con los requisitos reglamentarios.

Para usar la Direct Connect conexión, los consumidores deben crear una interfaz virtual pública, privada o de tránsito. Hay diferentes opciones de arquitectura disponibles. La más flexible para conectar múltiples ubicaciones locales Nube de AWS es una interfaz virtual de tránsito conectada a una Direct Connect puerta de enlace. Una Direct Connect puerta de enlace es un componente lógico global que permite al proveedor de servicios conectar hasta seis puertas de enlace de tránsito a ella. Además, puede conectar hasta 30 interfaces virtuales a la puerta de enlace. Para aumentar la escala, puede crear Direct Connect pasarelas adicionales. En la cuenta del proveedor de SaaS, las pasarelas de tránsito se conectan entonces a la VPCs, como se describió anteriormente.

Los consumidores pueden conectarse mediante una o cuatro Direct Connect conexiones desde un total de una o dos Direct Connect ubicaciones, según el nivel de resiliencia deseado. Para obtener más información, consulte Configurar Direct Connect para obtener la máxima resiliencia. Una AWS Site-to-Site VPN conexión a través de Internet también puede servir como ruta de respaldo de menor

costo para una Direct Connect conexión. Las conexiones Direct Connect dedicadas compatibles se pueden utilizar MACsecpara cifrar el enlace de la capa 2 entre la Direct Connect ubicación y el centro de datos. Es habitual disponer de una conexión Site-to-Site VPN para aumentar la confidencialidad de los datos. La conexión Site-to-Site VPN puede terminar en la pasarela de tránsito mediante un adjunto VPN normal. El siguiente diagrama muestra esta arquitectura.



Los siguientes son beneficios de este enfoque:

- Observabilidad: integración para la supervisión activa gestionada mediante Network Synthetic Monitor
- Escalabilidad: Support para aumentar el rendimiento del ancho de banda
- Adaptabilidad: soporte IPv6
- TCO: posibilidad de reducir la transferencia de datos
- TCO: experiencia de red uniforme
- Aislamiento de la red: conectividad privada que puede cumplir con los requisitos reglamentarios

Los inconvenientes de este enfoque son los siguientes:

- Facilidad de integración: la configuración requiere tiempo y esfuerzo manual
- Escalabilidad: escalabilidad limitada más allá de decenas de Direct Connect conexiones porque hay varias cuotas que rastrear

- Adaptabilidad: las opciones de configuración dependen de las ubicaciones disponibles Direct Connect
- TCO: el Direct Connect mantenimiento programado puede provocar un tiempo de inactividad que requiere la adopción de medidas

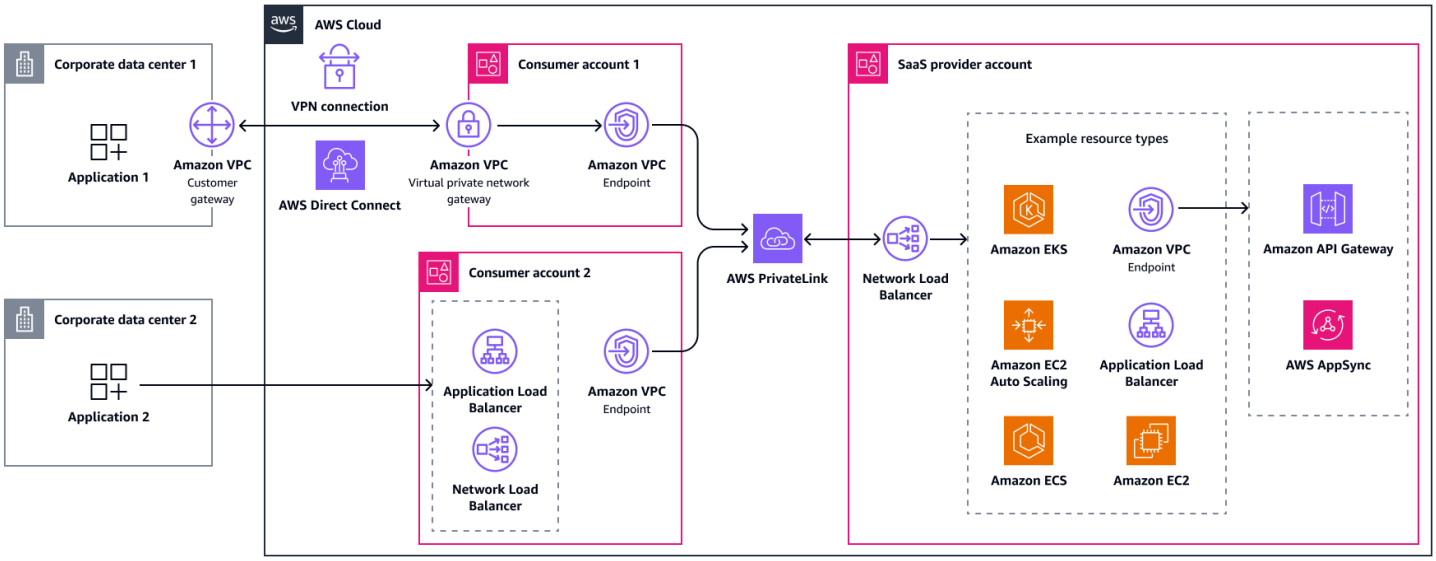
Conexión con una arquitectura de VPC de tránsito

Transit VPC es una opción de arquitectura que brinda flexibilidad a los consumidores en cuanto a la forma de conectarse y permite a AWS los proveedores de SaaS beneficiarse de un acceso unificado a su servicio mediante. AWS PrivateLink El consumidor se conecta desde las instalaciones a una VPC de tránsito que contiene solo un punto de entrada (como una puerta de enlace privada virtual) y un punto final de VPC de interfaz, que es un recurso. AWS PrivateLink El transporte VPCs debe ser propiedad del proveedor de SaaS o de los consumidores. En esta sección se analizan ambas opciones.

Puede crear la VPC de tránsito y las subredes con rangos de CIDR que sean compatibles con el centro de datos local. Si requieren conectividad privada, los consumidores pueden conectarse a esa VPC a través AWS Direct Connect de o. AWS Site-to-Site VPN También puedes configurar el acceso a la cuenta de tránsito desde la Internet pública mediante un Application Load Balancer o Network Load Balancer que apunte al punto final de la VPC.

VPC de tránsito gestionada por el consumidor

En este enfoque, el proveedor de SaaS deja la gestión del tránsito en VPCs manos de los consumidores. Desde un punto de vista técnico, la arquitectura del proveedor de SaaS es la misma que cuando se conecta directamente con los Nube de AWS consumidores. AWS PrivateLink Desde el punto de vista de las ventas y los productos, se trata de un esfuerzo adicional porque algunos consumidores Cuentas de AWS aún no lo han hecho. Es posible que duden en abrir y operar una cuenta. El proveedor de SaaS debe orientar a sus consumidores sobre cómo crear Cuentas de AWS y conectar su centro de datos local. El siguiente diagrama muestra una combinación de acceso público y privado, en la que los consumidores son los dueños del transporte. VPCs



Los siguientes son beneficios de este enfoque:

- Tiempo de reparación: los gastos operativos se transfieren en gran medida a los consumidores de SaaS
- Adaptabilidad: los consumidores de SaaS pueden elegir entre diferentes opciones de acceso
- Adaptabilidad: no hay conflictos de rango de CIDR, incluso cuando se utiliza una VPN o Site-to-Site Direct Connect
- Todos los indicadores: el proveedor de servicios hereda los beneficios AWS PrivateLink

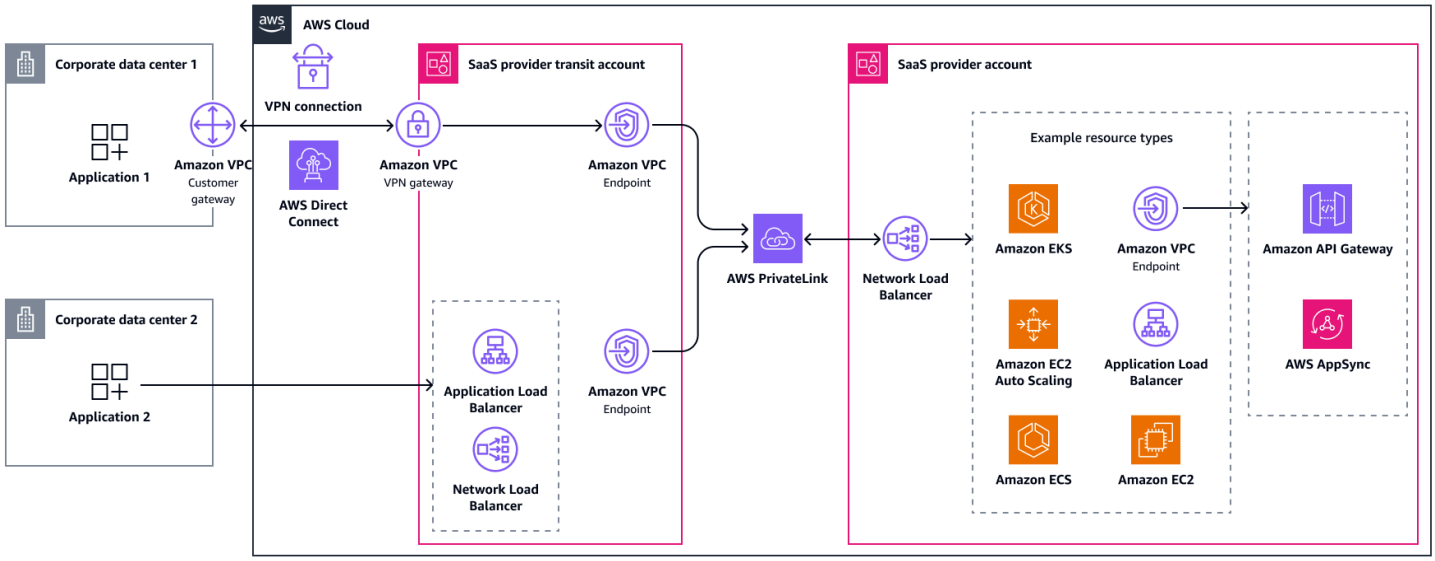
Los inconvenientes de este enfoque son los siguientes:

- Facilidad de integración: los consumidores de SaaS necesitan al menos un Cuenta de AWS
- TCO: una VPC de tránsito es una arquitectura, no un servicio totalmente gestionado, por lo que requiere un mayor esfuerzo operativo

VPC de tránsito gestionada por el proveedor

Este enfoque utiliza las mismas tecnologías, pero los límites y las responsabilidades de las cuentas cambian. En este caso, el proveedor de SaaS es el propietario del tránsito VPCs, preferiblemente en una cuenta separada de la oferta de SaaS. Esta disociación reduce los costos, reduce los riesgos y permite que la cuenta de tránsito se amplíe de forma independiente. Para los entornos que requieren un alto grado de aislamiento, puede crear una separación adicional entre los inquilinos mediante una subred o creando una VPC de tránsito independiente para cada consumidor. Los consumidores

pueden entonces elegir cómo conectarse a la VPC de tránsito. Este enfoque ofrece más opciones para expandir el mercado total accesible, pero tiene un mayor TCO para el proveedor de SaaS debido a la necesidad de operar y monitorear componentes arquitectónicos adicionales.



Los siguientes son beneficios de este enfoque:

- Adaptabilidad: los consumidores de SaaS pueden elegir entre diferentes opciones de acceso
- Adaptabilidad: los consumidores de SaaS no necesitan tener un Cuenta de AWS
- Adaptabilidad: no hay conflictos de rango de CIDR, incluso cuando se utiliza una VPN o Site-to-Site Direct Connect

Los inconvenientes de este enfoque son los siguientes:

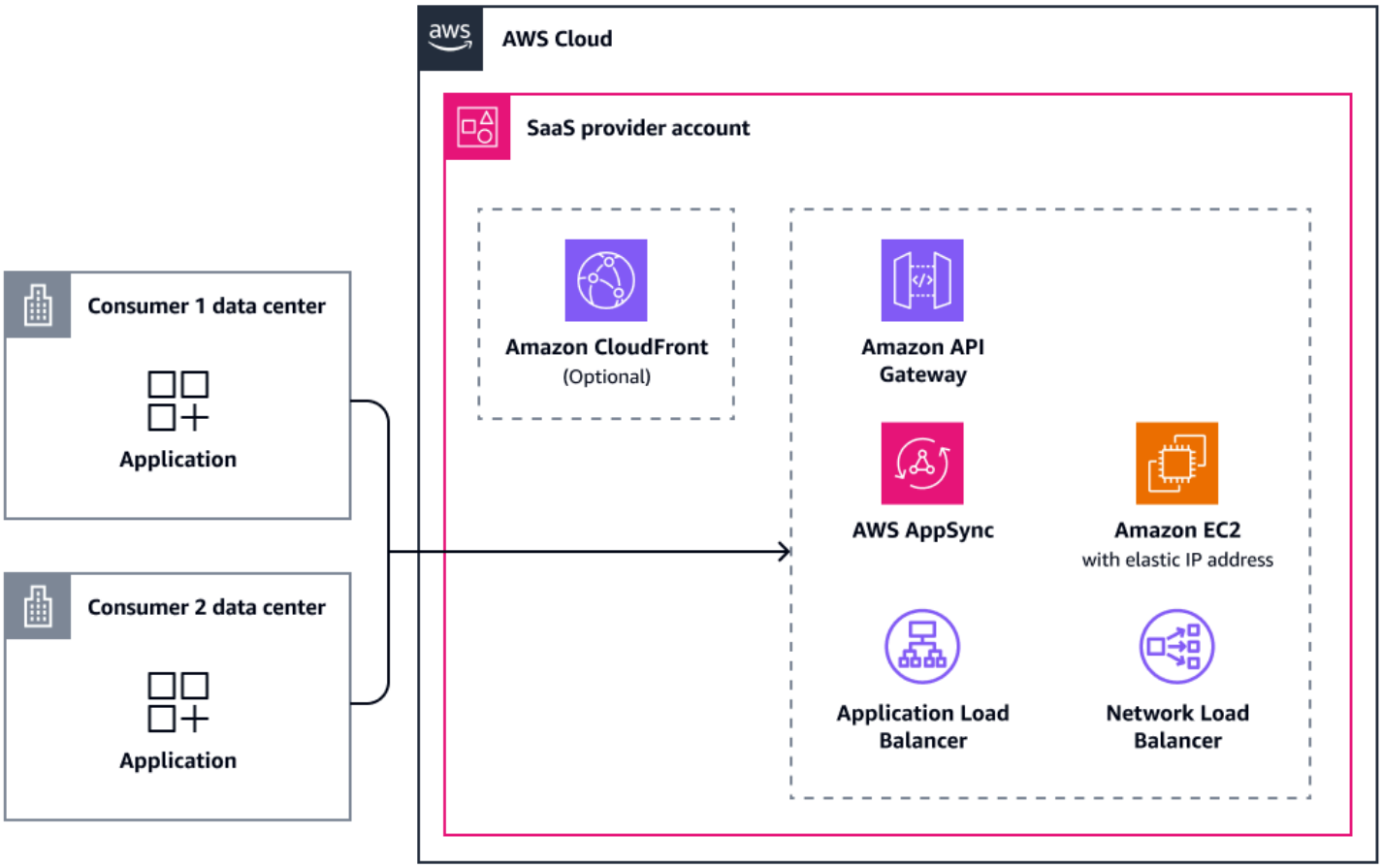
- TCO: una VPC de tránsito es una arquitectura, no un servicio totalmente gestionado, por lo que requiere un mayor esfuerzo operativo
- TCO: el proveedor de SaaS necesita operar y monitorear componentes arquitectónicos adicionales

Conexión a través de la Internet pública

El acceso público a Internet también es una opción válida para proporcionar acceso a una oferta de SaaS, aunque no ofrece conectividad privada en el sentido tradicional. Es posible que algunos consumidores sigan prefiriendo un enfoque de acceso público porque no requiere una infraestructura de red adicional entre ellos y el proveedor de SaaS. Reduce la complejidad, el coste y el tiempo de integración a cambio de una mayor superficie de ataque. Los sólidos mecanismos de autenticación

y autorización pueden ayudar a mitigar el aumento del nivel de amenaza, y siempre se debe cifrar el tráfico. Aun así, se recomienda disponer de un nivel de seguridad adicional en este escenario, por ejemplo, mediante AWS WAFel uso de.

La arquitectura en este escenario es sencilla. El consumidor se conecta a un servidor público (el proveedor de SaaS) a través de Internet. La aplicación se puede alojar directamente en una instancia pública de Amazon Elastic Compute Cloud (Amazon EC2) con una dirección IP elástica. La opción preferida es alojarlo detrás de un Application Load Balancer o un servicio similar. Para obtener un mejor rendimiento y almacenar en caché los activos estáticos, puede utilizar una red de entrega de contenido, como Amazon CloudFront. Para ofrecer una aplicación con una latencia mínima en dos direcciones IP Anycast estáticas globales, puede colocarla AWS Global Acceleratordelante de una instancia de Amazon EC2, Network Load Balancer o Application Load Balancer. Además CloudFront, los balanceadores de carga de aplicaciones y Amazon API Gateway se integran con AWS WAF. AWS AppSync El siguiente diagrama proporciona una descripción general de las opciones de conectividad del acceso público a Internet.



En la siguiente tabla se describen los protocolos e integraciones compatibles en este escenario.

Servicio o recurso	IPv6	AWS WAF integration	Puede ser un punto final de Global Accelerator
Amazon CloudFront	Soportado	compatible	No compatible
Amazon API Gateway	Soportado	compatible	No compatible
AWS AppSync	Compatible parcialmente	compatible	No compatible
Amazon EC2 con una dirección IP elástica	compatible	No compatible	compatible
Equilibrador de carga de aplicación	Soportado	Soportado	compatible
Network Load Balancer	compatible	No compatible	compatible

Los siguientes son beneficios de este enfoque:

- Facilidad de integración: simplicidad y accesibilidad
- Escalabilidad: escala ilimitada
- Adaptabilidad: no es posible que surjan conflictos de rango CIDR
- Adaptabilidad: soporte CloudFront

Los inconvenientes de este enfoque son los siguientes:

- Aislamiento de la red: no hay conectividad privada
- Aislamiento de la red: se requieren fuertes medidas de seguridad

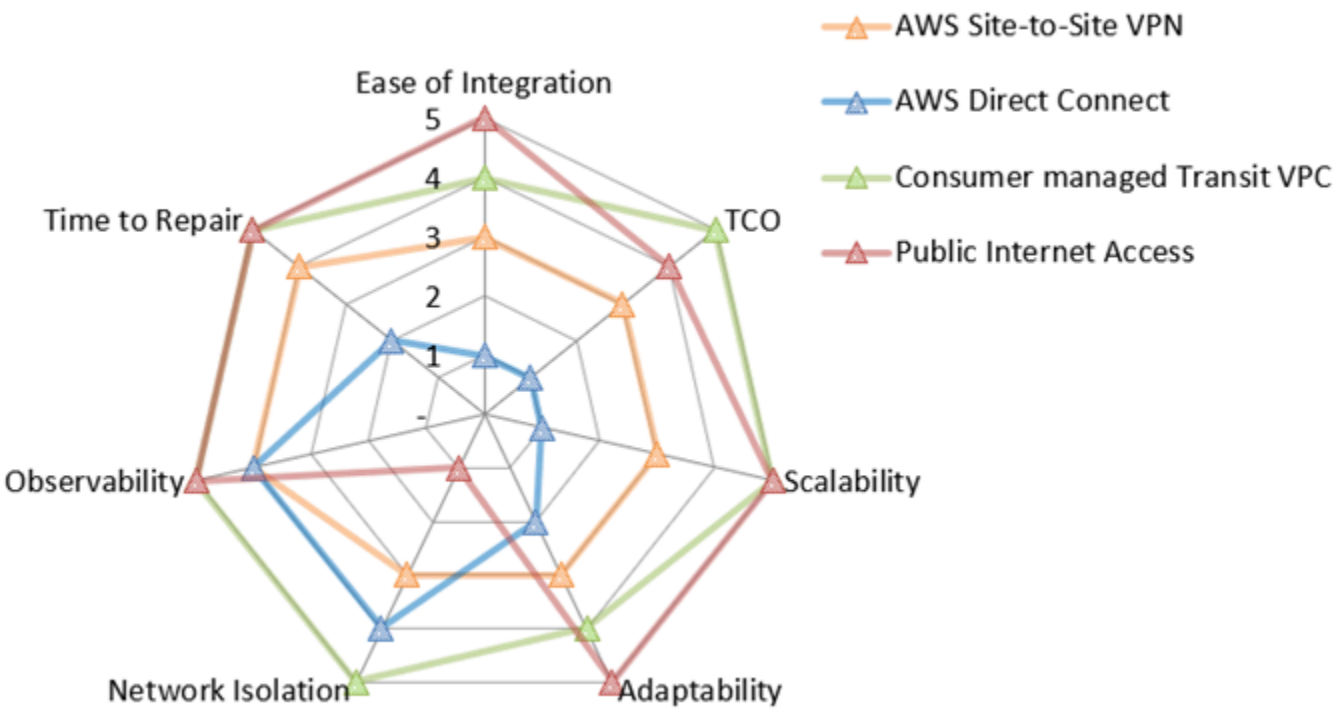
Existen otros beneficios e inconvenientes, según los servicios que elija.

Consumidores de SaaS que operan en otros proveedores de servicios en la nube

En este escenario se describen las soluciones para los consumidores de otros proveedores de servicios en la nube (CSPs). Este escenario comparte algunos puntos en común con las conexiones a los centros de datos locales. De hecho, todas las opciones de conectividad para entornos locales son igual de válidas para los consumidores que en otras CSPs, incluso una conexión privada AWS Direct Connect es posible en algunas de ellas. CSPs La mayoría CSPs ofrece documentación y asistencia sobre cómo conectarse a Nube de AWS través de AWS Site-to-Site VPN o AWS Direct Connect.

Al elegir Site-to-Site una VPN, los consumidores pueden beneficiarse de las pasarelas gestionadas o de recursos similares de sus respectivos CSP. Los consumidores no necesariamente tienen que configurarlas ellos mismos, como en el caso de las instalaciones locales. Esto influye en algunas de las métricas de la Site-to-Site VPN, como las mejoras en el tiempo de reparación y la observabilidad. Esto se debe a que ahora se administran ambos extremos de la conexión.

El siguiente mapa de valores de red resume la puntuación de cada una de estas opciones para cada métrica de evaluación. Es muy similar al mapa de valores de red para las conexiones locales, aunque los valores de las Site-to-Site VPN son diferentes. Para obtener más información sobre las métricas de evaluación, consulte Métricas de evaluación esta guía. En el mapa, un cinco representa la mejor puntuación, por ejemplo, el menor TCO, el mejor aislamiento de la red o el menor tiempo de reparación. Para obtener más información sobre cómo leer este gráfico radial, consulte Mapa de valores de redes esta guía.



El gráfico radial muestra los siguientes valores.

Métrica de evaluación	AWS Site-to-Site VPN	AWS Direct Connect	VPC de tránsito gestionada por el consumidor	Acceso público a Internet
Facilidad de integración	3	1	4	5
TCO	3	1	5	4
Escalabilidad	3	1	5	5
Adaptabilidad	3	2	4	5
Aislamiento de red	3	4	5	1
Observabilidad	4	4	5	5

Es hora de reparar	4	2	5	5

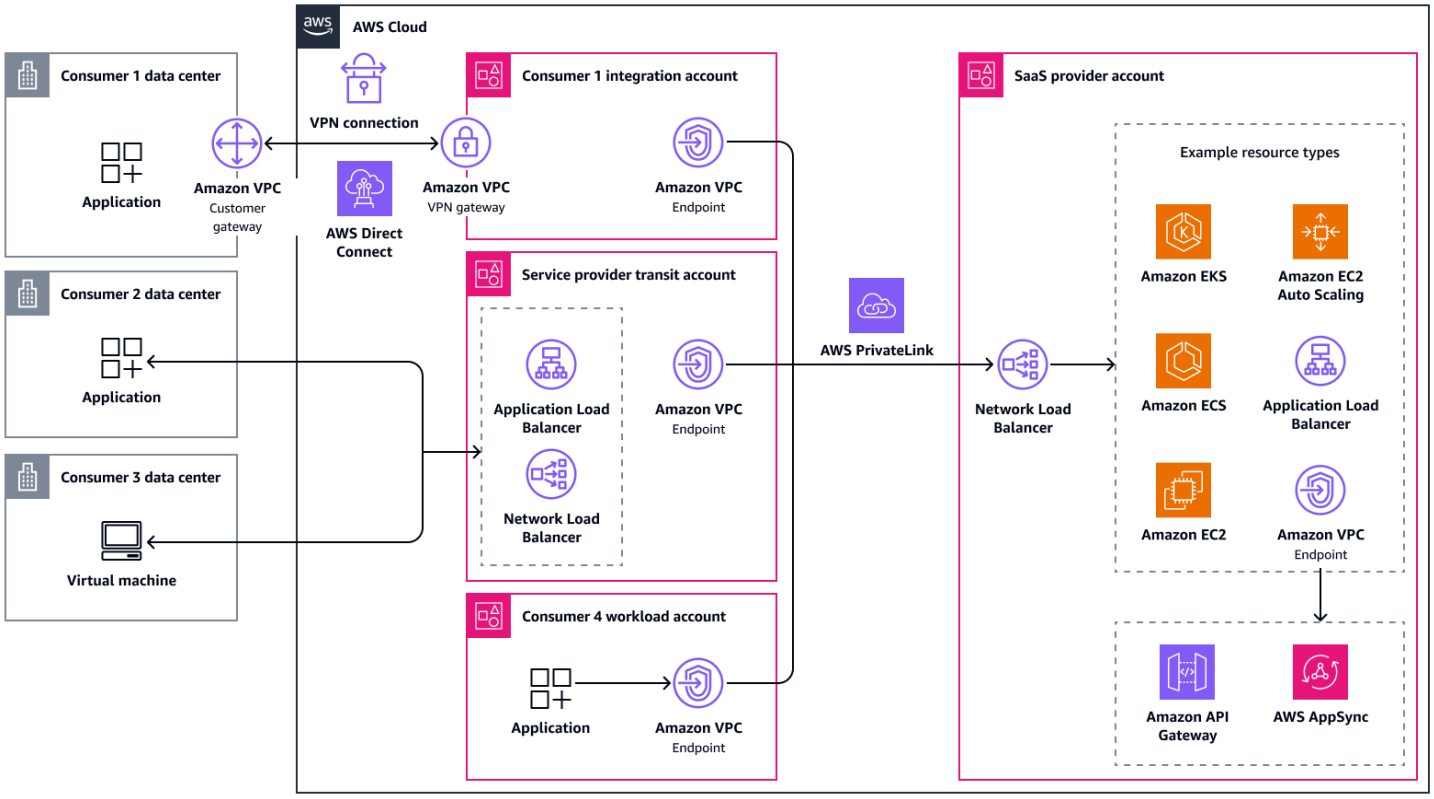
Compatible con entornos híbridos

Es habitual que los consumidores procedan de entornos diferentes, cada uno con sus propias limitaciones técnicas y de seguridad. Es posible que algunos clientes operen exclusivamente desde centros de datos locales que requieren una conectividad segura a través de Internet o mediante enlaces de red dedicados. Es posible que otros ya estén ejecutando cargas de trabajo en ellas AWS y esperen utilizar rutas de red privadas de baja latencia. Un tercer grupo podría depender de otro CSPs, donde la conectividad debe unir diferentes redes en la nube.

En cualquier caso, debe aspirar a un acceso de red estandarizado a su aplicación SaaS para simplificar su arquitectura y reducir la complejidad operativa. Dos de los enfoques presentados anteriormente (el acceso público a Internet y el tránsito VPCs) funcionan bien en estos escenarios. El acceso público a Internet ofrece la ruta de incorporación más rápida con una configuración mínima para sus clientes. Transit VPCs ofrece un acceso más controlado y privado, que a menudo utiliza AWS PrivateLink.

Al diseñar su oferta de SaaS, puede adoptar un único modelo de acceso a la red o combinar varios enfoques en una oferta escalonada. Por ejemplo, puede ofrecer un nivel de implementación de acceso público para los clientes que priorizan la facilidad de conexión y la rápida incorporación, y puede ofrecer un nivel de implementación de acceso privado para los clientes que tienen requisitos estrictos de cumplimiento o control de seguridad. Estos niveles tienen distintos perfiles de coste, rendimiento y riesgo. También es posible combinar ambos enfoques en una sola arquitectura. En ese caso, asegúrese de contar con medidas de seguridad sólidas para que las rutas públicas y privadas permanezcan aisladas.

El siguiente diagrama muestra un enfoque de acceso híbrido, en el que los consumidores tienen la opción de conectarse de forma privada desde su centro de datos o CSP, de forma pública o directamente AWS PrivateLink (si tienen cargas de trabajo en él). Nube de AWS


AWS Cloud
Consumer 1 data center
Application
Amazon VPC
Customer gateway
VPN connection
AWS Direct Connect
Amazon VPC
VPN gateway
Consumer 1 integration account
Amazon VPC
Endpoint
Consumer 2 data center
Application
Consumer 3 data center
Virtual machine
Service provider transit account
Application Load Balancer
Network Load Balancer
Amazon VPC
Endpoint
Consumer 4 workload account
Application
Amazon VPC
Endpoint
AWS PrivateLink
SaaS provider account
Network Load Balancer
Example resource types
Amazon EKS
Amazon EC2 Auto Scaling
Amazon ECS
Application Load Balancer
Amazon EC2
Amazon VPC
Endpoint
Amazon API Gateway
AWS AppSync

Escenarios de acceso a redes avanzados para las ofertas de SaaS en el Nube de AWS

Las arquitecturas que se describen en la Escenarios de acceso a redes para ofertas de SaaS en el Nube de AWS sección deberían ayudarle a encontrar una solución para la mayoría de los casos de uso. Sin embargo, hay algunos escenarios que tienen requisitos técnicos específicos. Muchos están fuera del alcance de esta guía.

En esta sección se analizan las siguientes consideraciones y requisitos técnicos avanzados:

- Comunicación bidireccional
- TCP, UDP y protocolos propietarios

Comunicación bidireccional

En algunos casos, las aplicaciones requieren tráfico bidireccional para funcionar según lo esperado. Los casos de uso más comunes son los webhooks o los servicios de notificación. Por lo general, esto se consigue mediante una WebSocket conexión entre el servidor y el cliente. Esta conexión mantiene abierta la sesión TCP y permite a ambos participantes enviar tráfico a través de la conexión. La mayoría de los servicios descritos en esta guía son compatibles de forma nativa WebSocket, incluidos los balanceadores de carga de red, los balanceadores de carga de aplicaciones, Amazon API Gateway y AWS AppSync (a través de AWS PrivateLink puntos de enlace privados en tiempo real).

En otros casos, una aplicación del lado del proveedor de SaaS podría necesitar acceso a recursos del lado del consumidor, como una base de datos. Cuando te conectas a través de canales bidireccionales, como una AWS Site-to-Site VPN conexión, eso no supone ningún problema.

Por otro lado, AWS PrivateLink Elastic Load Balancing solo admite tráfico unidireccional. Si utiliza estos servicios, debe configurar otra ruta de red para el tráfico que se inicia desde su oferta de SaaS. Por ejemplo, puede tratarse de una AWS PrivateLink conexión adicional que vaya en la dirección contraria.

TCP, UDP y protocolos propietarios

Muchas aplicaciones se sirven a través de HTTP o HTTPS, pero no todas. Algunos pueden usar otros protocolos de capa 7 además del TCP, como Message Queuing Telemetry Support (MQTT). Otros podrían incluso utilizar el UDP para atender a los consumidores. En raras ocasiones, los servicios utilizan protocolos propietarios que deben transmitirse dentro de paquetes (capa 3). Para estos escenarios, es importante entender qué servicios respaldan su oferta de SaaS.

Para los servicios de capa 3, puede usar AWS PrivateLink balanceadores de carga de red, los cuales admiten todo el tráfico TCP y UDP.

Para los servicios de capa 7, Application Load Balancers y Amazon CloudFront admiten HTTP WebSocket, HTTPS y Google Remote Procedure Calls (gRPC). Del mismo modo, Amazon API Gateway y AWS AppSync cada uno admiten HTTP, HTTPS y WebSocket. Amazon CloudFront es el único servicio que actualmente admite HTTP/3.

Puede usar Amazon VPC Lattice para conectar aplicaciones de capa 7 y recursos de capa 3. Es compatible con la transferencia HTTP, HTTPS, gRPC, TCP y TLS.

Si la aplicación solo puede atender el tráfico a través de la capa 3, es fundamental que utilice los servicios de AWS red principales, como AWS Transit Gateway, AWS Direct Connect AWS Site-to-Site VPN, y el emparejamiento de VPC. Luego, el tráfico debe enrutarse directamente desde el consumidor de SaaS a la capa de cómputo de la oferta de SaaS.

Antipatrones para el acceso a la red en el Nube de AWS

Un antipatrón es una solución que se utiliza con frecuencia para un problema recurrente en el que la solución es contraproducente, ineficaz o menos eficaz que una alternativa. Las opciones de diseño que se mencionan en esta sección suelen funcionar, pero presentan desventajas importantes. Si es posible, deben evitarse porque hay mejores alternativas disponibles.

En esta sección se analizan los siguientes obstáculos y desafíos:

- La zona de disponibilidad no coincide con AWS PrivateLink
- AWS Site-to-Site VPN conexiones entre Cuentas de AWS

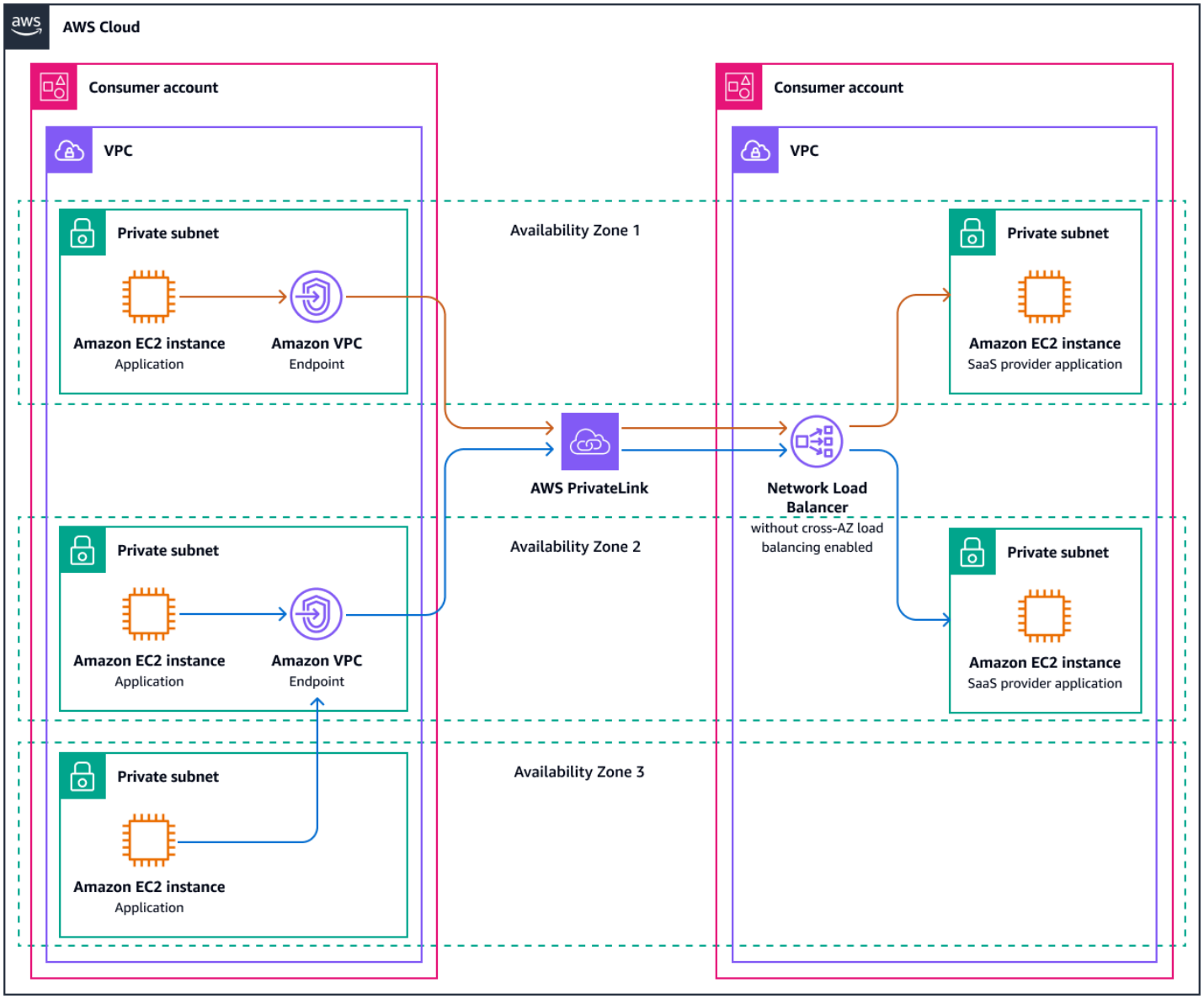
La zona de disponibilidad no coincide con AWS PrivateLink

Al proporcionar acceso a una aplicación AWS PrivateLink, los consumidores de SaaS pueden crear puntos de enlace de VPC de interfaz solo en las zonas de disponibilidad en las que se implementa la aplicación. Por ejemplo, si la aplicación se implementa en `use1-az1` y`use1-az2`, el consumidor no puede implementar un punto final de VPC en. `use1-az3` Le recomendamos que implemente la oferta de SaaS en todas las zonas de disponibilidad. La mayoría Regiones de AWS tienen tres zonas de disponibilidad, aunque algunas tienen más. Para obtener una lista completa, consulte Regiones y zonas de disponibilidad. Tenga en cuenta el número de zonas de disponibilidad al elegir una Región de AWS.

> **Note**
>
> Los nombres de las zonas de disponibilidad son diferentes de los de las zonas de disponibilidad IDs. Para obtener más información, consulte Zona de disponibilidad IDs para ver sus AWS recursos.

Si un proveedor de SaaS decide no realizar la implementación en todas las zonas de disponibilidad, hay algunas consecuencias. Suponga que la oferta de SaaS se implementa en `use1-az1` y`use1-az2`, pero el consumidor utiliza las tres zonas de disponibilidad, incluidas. `use1-az3` Los puntos finales de la VPC de la interfaz se implementan en el lado del consumidor `use1-az1` y`use1-az2`, ahora, la aplicación `use1-az3` necesita acceder a uno de estos puntos de enlace. En primer lugar, se debe permitir el tráfico desde las subredes de las zonas de disponibilidad no coincidentes hacia los puntos finales de VPC respectivos. El consumidor puede decidir usar el nombre AWS

PrivateLink DNS regional, que se puede resolver en cualquiera de los extremos de la VPC y que distribuye uniformemente el tráfico entre los dos. O bien, el consumidor puede optar por enviar el tráfico directamente a un punto final, por ejemplo. `use1-az2` Esto se traduce en que el 67% del tráfico llega al proveedor `use1-az2` y el 33% entra`use1-az1`. En la siguiente figura se muestra este escenario.



Con un número significativo de consumidores y una distribución desigual del tráfico, una carga de trabajo puede tener problemas de capacidad en una zona de disponibilidad y estar insuficiente en otra. Para solucionar este problema, el proveedor de SaaS puede decidir equilibrar de manera uniforme la carga del tráfico de su lado habilitando el equilibrio de carga entre zonas en el Network Load Balancer. Esto conlleva cargos adicionales.

Si el proveedor de servicios solo coincide con una zona de disponibilidad, todo el tráfico entrará por un único punto final. Esto crea un desequilibrio aún mayor. Como resultado, la oferta de SaaS ya no está muy disponible para el consumidor. Al consumidor no le importa si la aplicación se entrega en zonas de disponibilidad adicionales que no esté utilizando él mismo. En el peor de los casos, es posible que un proveedor de SaaS no pueda atender a un consumidor que no utilice ninguna de las mismas zonas de disponibilidad.

En el raro caso de que el proveedor de SaaS no tenga una opción viable para aprovisionar su aplicación en todas las zonas de disponibilidad, también es posible crear una subred solo en las zonas de disponibilidad que faltan y, a continuación, extender el servicio a esas zonas de disponibilidad vacías. El equilibrio de carga entre zonas permite entonces distribuir el tráfico entrante entre los puntos finales de las aplicaciones reales de las demás zonas de disponibilidad.

AWS Site-to-Site VPN conexiones entre Cuentas de AWS

Las empresas que migran de entornos locales a la nube a veces intentan impulsar y cambiar toda la red. Esto puede causar problemas porque existen diferencias significativas entre las prácticas de red locales y en la nube. Si este cambio de mentalidad no se produce, pueden ocurrir cosas como AWS Site-to-Site VPN las conexiones de una VPC a otra. Este enfoque no aprovecha los servicios de red diseñados específicamente para tal fin Nube de AWS, que simplifican la administración y mejoran el rendimiento. La adaptación a los diseños nativos de la nube ayuda a reducir la sobrecarga operativa y da como resultado una conectividad más confiable y escalable entre ellos. VPCs

Si está pensando en ofrecer esta opción de conectividad como proveedor de SaaS, pregúntese a sí mismo o al consumidor por qué AWS Site-to-Site VPN debería utilizarla. Luego, analice esos requisitos para encontrar una mejor opción de conectividad. La sección de comparación de las capacidades de los servicios de esta guía contiene una matriz que puede utilizar para identificar las opciones. A continuación, puede revisar las secciones pertinentes de esta guía para encontrar un enfoque arquitectónico que aborde su caso de uso.

Siguientes pasos

Esta guía describe varios enfoques de acceso a la red en diferentes escenarios y describe las ventajas y desventajas de cada arquitectura. Debe comprender por qué la elección de un enfoque de acceso a la red no debe ser una cuestión meramente tecnológica. La alineación entre la empresa y la tecnología es esencial. Los siguientes pasos y recomendaciones pueden ayudarlo a evaluar y estandarizar su estrategia de arquitectura de red mediante la evaluación de las capacidades actuales, el análisis de las necesidades del mercado y la implementación de controles de gobierno.

Esta sección contiene los siguientes temas:

- Evaluar la arquitectura y las capacidades actuales
- Análisis de mercados y clientes
- Alineación estratégica
- Normalización
- Gobernanza
- Repetición

Evaluar la arquitectura y las capacidades actuales

Revise la arquitectura de red actual comparándola con las fuentes de datos pertinentes, como el marco de autoevaluación de esta guía, los requisitos reglamentarios actuales y el estado actual del mercado (tanto en términos de clientes como de un análisis de la competencia). Por ejemplo, considere la posibilidad de utilizar el marco AWS Well-Architected, que se basa en décadas de experiencia en el funcionamiento de sistemas de producción a escala en el. Nube de AWS

Revise las posibles excepciones, las situaciones puntuales y las decisiones históricas sobre productos. Sea curioso, desafíelas y no asuma automáticamente su validez. Es posible que los requisitos de los clientes de hace años ya no sean válidos. Las suposiciones desafiantes crean la oportunidad de simplificar y reducir la complejidad de su arquitectura.

En términos sencillos, documente las observaciones para que las distintas funciones de su organización puedan acceder a ellas y comprenderlas. Capture en qué se diferencia el estado actual del estado objetivo, cuál es el estado objetivo, el impacto y cuándo se realizaron las observaciones. Registrar esta información ayuda a sus organizaciones a tomar decisiones basadas en datos recientes.

Análisis de mercados y clientes

Recopile información sobre las tendencias del mercado. ¿Cuál es la forma preferida actualmente por los consumidores de acceder a ofertas de SaaS como la suya? ¿Sigues reuniéndote con tus clientes donde están? ¿Cambiaron las cohortes o el comportamiento de los clientes? ¿Sus ejecutivos dirigieron el barco hacia un nuevo mercado, una zona geográfica con requisitos normativos específicos o un nuevo nivel de clientes? ¿Cambió su negocio o modelo operativo? Por ejemplo, ¿está pensando en etiquetar sus servicios con una etiqueta blanca? ¿Su plan de crecimiento incluye trabajar con socios para que su servicio esté disponible para los clientes cuando se pongan en contacto con esos socios?

Alineación estratégica

Cuando comprenda sus capacidades actuales, su arquitectura actual, su mercado y sus clientes, convoque una reunión de alineación estratégica. Con las partes interesadas pertinentes en materia de productos, negocios y tecnología, cuestione qué requisitos siguen siendo válidos y qué requisitos nuevos deben tenerse en cuenta. Encuentre oportunidades para reducir la complejidad eliminando los requisitos que ya no son necesarios. No se trata de un diseño realizado por un comité; el equipo de ingeniería debe preparar y gestionar los detalles reales de la arquitectura y la implementación. Sin embargo, esta reunión debería aclarar por qué este es el conjunto de requisitos que maximiza los beneficios para sus clientes y su organización.

Normalización

Para atraer clientes, puede resultar tentador dejar que cada uno elija libremente cómo conectarse a su servicio. Al fin y al cabo, cualquier solución puede funcionar técnicamente, y es posible que también tengas los conocimientos y los recursos para gestionarlas y utilizarlas todas. Esto puede funcionar bien hasta cierto punto, pero a medida que su empresa crece, se hace más difícil de administrar. Su conjunto de observabilidad debe respaldar las métricas de varias soluciones, y los ingenieros de confiabilidad de sus sitios también deben poder entenderlas. Necesita up-to-date documentación para cada enfoque de conectividad. Los cambios principales en su aplicación deben evaluarse en función de cada enfoque de acceso que ofrezca. Debe escribir y mantener las automatizaciones y la infraestructura como código (IaC) para cada enfoque de acceso. La sobrecarga adicional que supone no estandarizar el acceso a su servicio debe sopesarse con la flexibilidad que desea ofrecer a sus clientes.

Si necesita una estrella polar que guíe su toma de decisiones, le sugerimos que la estandarice. La estandarización de la forma en que sus clientes interactúan con los servicios que presta suele ser la acción más impactante que puede tomar para mejorar muchos de los indicadores de éxito de su organización. La estandarización facilita a los equipos de productos la comprensión de la estructura de costes de sus servicios y la toma de decisiones sobre los productos basadas en los datos. A los equipos de operaciones les resulta más fácil solucionar problemas y automatizar partes del proceso de solución de problemas en un entorno que se desarrolla, implementa y opera de acuerdo con estándares predefinidos. Puede ayudarle a detectar anomalías, comportamientos inesperados o acciones de un actor malintencionado. La estandarización también reduce la deuda técnica. Los equipos de ingeniería tardan menos ciclos en probar e implementar los cambios en la producción. También puede aumentar su velocidad de comercialización, mejorar el éxito de la incorporación del autoservicio y reducir el riesgo regulatorio.

Por lo tanto, le sugerimos que revise también cualquier oferta única que esté vigente en la actualidad. Cuantifique la cantidad de ciclos operativos que dedica a atender a los clientes actuales. Compare sus resultados con los datos históricos y evalúe si su enfoque actual se adapta a los próximos años. Siempre que sea necesario desviarse de los estándares, cuestione los requisitos en los que se basan esas solicitudes. Evalúe el impacto y equilibre los beneficios inmediatos con los compromisos a largo plazo.

En los casos en que la personalización sea inevitable pero entre en conflicto con sus estándares, considere un modelo de responsabilidad compartida. En este modelo, tus productos están en gran medida protegidos de los cambios solicitados y la personalización se realiza en un entorno minimalista y específico. Para ver un ejemplo, consulte la Conexión con una arquitectura de VPC de tránsito sección.

Gobernanza

Para cumplir con los requisitos reglamentarios y con sus propios estándares internos, la gobernanza es esencial. Con una gobernanza adecuada, puede controlar dónde y cómo hacer cumplir las normas. También debe establecer controles para detectar las divergencias con respecto a las normas e informar a los propietarios de los recursos sobre las medidas correctivas necesarias. AWS Organizations, AWS ConfigAWS CloudTrail, y AWS Control Towerson algunos de los muchos Servicios de AWS que pueden ayudarle a gestionar y controlar sus cargas de trabajo en el. Nube de AWS

Repetición

Utilizando lo aprendido de sus esfuerzos iniciales, establezca un proceso ligero y repetible para mantenerse alineado en el futuro. Defina las funciones de las que necesita información, con qué frecuencia, qué precisión deben tener los datos, cómo se compartirán y quién actuará en consecuencia.

Recursos

AWS documentación

- Integración de servicios de terceros en la Nube de AWS (Guía AWS prescriptiva)
- Autorización de SaaS para múltiples inquilinos y control de acceso a la APIAWS (guía prescriptiva)
- Administre a los inquilinos en varios productos SaaS en un solo plano de control (orientaciónAWS prescriptiva)
- ¿Qué es? AWS Direct Connect (Direct Connect documentación)
- ¿Qué es AWS PrivateLink? (documentación de Amazon VPC)
- ¿Qué es? AWS Site-to-Site VPN (AWS Site-to-Site VPN documentación)
- ¿Qué es AWS Transit Gateway? (documentación de Amazon VPC)
- ¿Qué es una interconexión de VPC? (documentación de Amazon VPC)

Otros recursos AWS

- Opciones de conectividad de Amazon Virtual Private Cloud (AWS documento técnico)
- AWS re:Invent 2021: cómo elegir el balanceador de cargas adecuado para sus cargas de trabajo () AWS YouTube
- ¿Qué es SaaS? (AWS sitio web)
- AWS Programa SaaS Factory (programa)AWS Partner
- Guía para arquitecturas multiusuario en (Biblioteca de soluciones AWS)AWS

Historial de documentos

En la siguiente tabla, se describen cambios significativos de esta guía. Si quiere recibir notificaciones de futuras actualizaciones, puede suscribirse a las notificaciones RSS.

Cambio	Descripción	Fecha
Publicación inicial	—	12 de septiembre de 2025

AWS Glosario de orientación prescriptiva

Los siguientes son términos de uso común en las estrategias, guías y patrones proporcionados por la Guía AWS prescriptiva. Para sugerir entradas, utilice el enlace Enviar comentarios al final del glosario.

Números

Las 7 R

Siete estrategias de migración comunes para trasladar aplicaciones a la nube. Estas estrategias se basan en las 5 R que Gartner identificó en 2011 y consisten en lo siguiente:

- Refactor/re-architect — Mueva una aplicación y modifique su arquitectura aprovechando al máximo las funciones nativas de la nube para mejorar la agilidad, el rendimiento y la escalabilidad. Por lo general, esto implica trasladar el sistema operativo y la base de datos. Ejemplo: migre su base de datos Oracle local a la PostgreSQL-Compatible edición Amazon Aurora.
- Redefinir la plataforma (transportar y redefinir): traslade una aplicación a la nube e introduzca algún nivel de optimización para aprovechar las capacidades de la nube. Ejemplo: Migrar la base de datos Oracle en las instalaciones a Amazon Relational Database Service (Amazon RDS) para Oracle en la nube de Nube de AWS.
- Recomprar (readquirir): cambie a un producto diferente, lo cual se suele llevar a cabo al pasar de una licencia tradicional a un modelo SaaS. Ejemplo: migre su sistema de gestión de relaciones con los clientes (CRM) a Salesforce.com.
- Volver a alojar (migrar mediante lift-and-shift): traslade una aplicación a la nube sin hacer cambios para aprovechar las funcionalidades de la nube. Ejemplo: Migrar la base de datos de Oracle en las instalaciones a Oracle en una instancia de EC2 en la Nube de AWS.
- Reubicar: (migrar el hipervisor mediante lift and shift): traslade la infraestructura a la nube sin comprar equipo nuevo, reescribir aplicaciones o modificar las operaciones actuales. Los servidores se migran de una plataforma en las instalaciones a un servicio en la nube para la misma plataforma. Ejemplo: migrar una Microsoft Hyper-V aplicación a AWS.
- Retener (revisitar): conserve las aplicaciones en el entorno de origen. Estas pueden incluir las aplicaciones que requieren una refactorización importante, que desee posponer para más adelante, y las aplicaciones heredadas que desee retener, ya que no hay ninguna justificación empresarial para migrarlas.

- Retirar: retire o elimine las aplicaciones que ya no sean necesarias en un entorno de origen.

A

A2A () Agent-to-Agent

Un protocolo completo para la colaboración entre agentes que facilita la delegación de tareas y la transferencia de estados.

ABAC

Consulte control de acceso basado en atributos.

servicios abstractos

Consulte servicios administrados.

ACID

Consulte atomicidad, consistencia, aislamiento, durabilidad.

migración activa-activa

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas (mediante una herramienta de replicación bidireccional o mediante operaciones de escritura doble) y ambas bases de datos gestionan las transacciones de las aplicaciones conectadas durante la migración. Este método permite la migración en lotes pequeños y controlados, en lugar de requerir una transición única. Es más flexible, pero requiere más trabajo que una migración activa-pasiva.

migración activa-pasiva

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas, pero solo la de origen gestiona las transacciones de las aplicaciones conectadas, mientras los datos se replican en la de destino. La base de datos de destino no acepta ninguna transacción durante la migración.

Agente

Un sistema de IA que puede razonar, planificar y tomar medidas de forma autónoma utilizando herramientas para alcanzar los objetivos.

Agent Ops

Prácticas operativas para crear, probar, implementar y ejecutar agentes de IA en producción a escala.

función de agregación

Función SQL que actúa en un grupo de filas y calcula un único valor de devolución para el grupo. Entre los ejemplos de funciones de agregación se incluyen `SUM` y `MAX`.

IA

Consulte inteligencia artificial.

AIOps

Consulte operaciones de inteligencia artificial

anonimización

El proceso de eliminar permanentemente la información personal de un conjunto de datos. La anonimización puede ayudar a proteger la privacidad personal. Los datos anonimizados ya no se consideran datos personales.

antipatrones

Una solución que se utiliza con frecuencia para un problema recurrente en el que la solución es contraproducente, ineficaz o menos eficaz que una alternativa.

control de aplicaciones

Enfoque de seguridad que permite usar de manera exclusiva aplicaciones aprobadas para ayudar a proteger un sistema contra el malware.

cartera de aplicaciones

Recopilación de información detallada sobre cada aplicación que utiliza una organización, incluido el costo de creación y mantenimiento de la aplicación y su valor empresarial. Esta información es clave para el proceso de detección y análisis de la cartera y ayuda a identificar y priorizar las aplicaciones que se van a migrar, modernizar y optimizar.

inteligencia artificial (IA)

El campo de la informática que se dedica al uso de tecnologías informáticas para realizar funciones cognitivas que suelen estar asociadas a los seres humanos, como el aprendizaje, la resolución de problemas y el reconocimiento de patrones. Para más información, consulte ¿Qué es la inteligencia artificial?

operaciones de inteligencia artificial (AIOps)

El proceso de utilizar técnicas de machine learning para resolver problemas operativos, reducir los incidentes operativos y la intervención humana, y mejorar la calidad del servicio. Para obtener más información sobre cómo se utiliza AIOps en la estrategia de migración de AWS , consulte la Guía de integración de operaciones.

cifrado asimétrico

Algoritmo de cifrado que utiliza un par de claves, una clave pública para el cifrado y una clave privada para el descifrado. Puede compartir la clave pública porque no se utiliza para el descifrado, pero el acceso a la clave privada debe estar sumamente restringido.

atomicidad, consistencia, aislamiento, durabilidad (ACID)

Conjunto de propiedades de software que garantizan la validez de los datos y la fiabilidad operativa de una base de datos, incluso en caso de errores, cortes de energía u otros problemas.

control de acceso basado en atributos (ABAC)

La práctica de crear permisos detallados basados en los atributos del usuario, como el departamento, el puesto de trabajo y el nombre del equipo. Para obtener más información, consulte ABAC AWS en la documentación AWS Identity and Access Management (IAM).

origen de datos fidedigno

Ubicación en la que se almacena la versión principal de los datos, que se considera la fuente de información más fiable. Puede copiar los datos del origen de datos autorizado a otras ubicaciones con el fin de procesarlos o modificarlos, por ejemplo, anonimizarlos, redactarlos o seudonimizarlos.

Zona de disponibilidad

Una ubicación distinta dentro de una Región de AWS que está aislada de los fallos en otras zonas de disponibilidad y que proporciona una conectividad de red económica y de baja latencia a otras zonas de disponibilidad de la misma región.

AWS Marco de adopción de la nube (AWS CAF)

Un marco de directrices y mejores prácticas AWS para ayudar a las organizaciones a desarrollar un plan eficiente y eficaz para migrar con éxito a la nube. AWS CAF organiza la orientación en seis áreas de enfoque denominadas perspectivas: negocios, personas, gobierno, plataforma, seguridad y operaciones. Las perspectivas empresariales, humanas y de gobernanza se centran en las habilidades y los procesos empresariales; las perspectivas de plataforma, seguridad y

operaciones se centran en las habilidades y los procesos técnicos. Por ejemplo, la perspectiva humana se dirige a las partes interesadas que se ocupan de los Recursos Humanos (RR. HH.), las funciones del personal y la administración de las personas. Desde esta perspectiva, AWS CAF proporciona orientación para el desarrollo, la formación y la comunicación de las personas a fin de preparar a la organización para una adopción exitosa de la nube. Para obtener más información, consulte la Página web de AWS CAF y el Documento técnico de AWS CAF.

AWS Marco de calificación de la carga de trabajo (AWS WQF)

Herramienta que evalúa las cargas de trabajo de migración de bases de datos, recomienda estrategias de migración y proporciona estimaciones de trabajo. AWS WQF se incluye con AWS Schema Conversion Tool ().AWS SCT Analiza los esquemas de bases de datos y los objetos de código, el código de las aplicaciones, las dependencias y las características de rendimiento y proporciona informes de evaluación.

B

bot malicioso

Bot destinado a causar interrupciones o daños a personas u organizaciones.

BCP

Consulte planificación de la continuidad del negocio.

gráfico de comportamiento

Una vista unificada e interactiva del comportamiento de los recursos y de las interacciones a lo largo del tiempo. Puede utilizar un gráfico de comportamiento con Amazon Detective para examinar los intentos de inicio de sesión fallidos, las llamadas sospechosas a la API y acciones similares. Para obtener más información, consulte Datos en un gráfico de comportamiento en la documentación de Detective.

sistema big-endian

Un sistema que almacena primero el byte más significativo. Consulte también endianidad.

clasificación binaria

Un proceso que predice un resultado binario (una de las dos clases posibles). Por ejemplo, es posible que su modelo de ML necesite predecir problemas como “¿Este correo electrónico es spam o no es spam?” o “¿Este producto es un libro o un automóvil?”.

filtro de floración

Estructura de datos probabilística y eficiente en términos de memoria que se utiliza para comprobar si un elemento es miembro de un conjunto.

blue/green despliegue

Estrategia de implementación en la que se crean dos entornos separados, pero idénticos. La versión actual de la aplicación se ejecuta en un entorno (azul) y la nueva versión de la aplicación se ejecuta en el otro entorno (verde). Esta estrategia lo ayuda a hacer reversiones rápidas con un impacto mínimo.

bot

Aplicación de software que ejecuta tareas automatizadas a través de Internet y simula la actividad o interacción humana. Algunos bots son útiles o beneficiosos, como los rastreadores web que indexan la información de Internet. Otros bots, conocidos como bots maliciosos, tienen como objetivo causar interrupciones o daños a personas u organizaciones.

botnet

Redes de bots infectadas por malware y que están bajo el control de una sola parte, conocida como pastor de bots u operador de bots. Las botnets son el mecanismo más conocido para escalar los bots y su impacto.

branch

Área contenida de un repositorio de código. La primera rama que se crea en un repositorio es la rama principal. Puede crear una rama nueva a partir de una rama existente y, a continuación, desarrollar características o corregir errores en la rama nueva. Una rama que se genera para crear una característica se denomina comúnmente rama de característica. Cuando la característica se encuentra lista para su lanzamiento, se vuelve a combinar la rama de característica con la rama principal. Para obtener más información, consulte Acerca de las sucursales (GitHub documentación).

acceso de emergencia

En circunstancias excepcionales y mediante un proceso aprobado, es una forma rápida de que un usuario pueda acceder a un Cuenta de AWS sitio al que normalmente no tiene permisos de acceso. Para obtener más información, consulte el indicador de implementación de procedimientos rompe-cristales en la AWS Well-Architected guía.

estrategia de implementación sobre infraestructura existente

La infraestructura existente en su entorno. Al adoptar una estrategia de implementación sobre infraestructura existente para una arquitectura de sistemas, se diseña la arquitectura en función de las limitaciones de los sistemas y la infraestructura actuales. Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de implementación desde cero.

caché de búfer

El área de memoria donde se almacenan los datos a los que se accede con más frecuencia.

capacidad empresarial

Lo que hace una empresa para generar valor (por ejemplo, ventas, servicio al cliente o marketing). Las arquitecturas de microservicios y las decisiones de desarrollo pueden estar impulsadas por las capacidades empresariales. Para obtener más información, consulte la sección Organizado en torno a las capacidades empresariales del documento técnico Ejecutar microservicios en contenedores en AWS.

planificación de la continuidad del negocio (BCP)

Plan que aborda el posible impacto de un evento disruptivo, como una migración a gran escala en las operaciones y permite a la empresa reanudar las operaciones rápidamente.

C

CAF

Consulte AWS Cloud Adoption Framework.

implementación canario

Lanzamiento lento e incremental de una versión para los usuarios finales. Cuando tenga mayor confianza en la nueva versión, la implementa y reemplaza la versión actual en su totalidad.

CCoE

Consulte Centro de excelencia en la nube.

CDC

Consulte captura de datos de cambios.

captura de datos de cambio (CDC)

Proceso de seguimiento de los cambios en un origen de datos, como una tabla de base de datos, y registro de los metadatos relacionados con el cambio. Puede utilizar los CDC para diversos fines, como auditar o replicar los cambios en un sistema de destino para mantener la sincronización.

ingeniería del caos

Introducción intencionada de fallos o eventos disruptivos para poner a prueba la resiliencia de un sistema. Puedes usar AWS Fault Injection Service (AWS FIS) para realizar experimentos que estresen tus AWS cargas de trabajo y evalúen su respuesta.

CI/CD

Consulte integración continua y entrega continua.

clasificación

Un proceso de categorización que permite generar predicciones. Los modelos de ML para problemas de clasificación predicen un valor discreto. Los valores discretos siempre son distintos entre sí. Por ejemplo, es posible que un modelo necesite evaluar si hay o no un automóvil en una imagen.

Desarrollador ciudadano

Un usuario empresarial que crea aplicaciones de IA utilizando plataformas sin code/low código sin conocimientos técnicos especializados.

cifrado del cliente

Cifrado de datos localmente, antes de que el objetivo los Servicio de AWS reciba.

Centro de excelencia en la nube (CCoE)

Equipo multidisciplinario que impulsa los esfuerzos de adopción de la nube en toda la organización, incluido el desarrollo de las prácticas recomendadas en la nube, la movilización de recursos, el establecimiento de plazos de migración y la dirección de la organización durante las transformaciones a gran escala. Para obtener más información, consulte las publicaciones de CCoE en el blog de estrategia Nube de AWS empresarial.

computación en la nube

La tecnología en la nube que se utiliza normalmente para la administración de dispositivos de IoT y el almacenamiento de datos de forma remota. La computación en la nube suele estar relacionada con la tecnología de computación de periferia.

modelo operativo en la nube

En una organización de TI, el modelo operativo que se utiliza para crear, madurar y optimizar uno o más entornos de nube. Para obtener más información, consulte Creación de su modelo operativo de nube.

etapas de adopción de la nube

Las siguientes son las cuatro fases por las que suelen pasar las empresas cuando migran a la Nube de AWS:

- Proyecto: ejecución de algunos proyectos relacionados con la nube con fines de prueba de concepto y aprendizaje
- Fundamento: realización de inversiones fundamentales para escalar la adopción de la nube (p. ej., crear una zona de aterrizaje, definir un CCoE, establecer un modelo de operaciones)
- Migración: migración de aplicaciones individuales
- Re-invention — Optimizar los productos y servicios e innovar en la nube

Stephen Orban definió estas etapas en la entrada del blog The Journey Toward Cloud-First & the Stages of Adoption del blog Nube de AWS Enterprise Strategy. Para obtener información sobre su relación con la estrategia de AWS migración, consulte la guía de preparación para la migración.

CMDB

Consulte base de datos de administración de configuración.

repositorio de código

Una ubicación donde el código fuente y otros activos, como documentación, muestras y scripts, se almacenan y actualizan mediante procesos de control de versiones. Algunos repositorios en la nube comunes son GitHub o Bitbucket Cloud. Cada versión del código se denomina rama. En una estructura de microservicios, cada repositorio se encuentra dedicado a una única funcionalidad. Una sola CI/CD canalización puede utilizar varios repositorios.

caché en frío

Una caché de búfer que está vacía no está bien poblada o contiene datos obsoletos o irrelevantes. Esto afecta al rendimiento, ya que la instancia de la base de datos debe leer desde la memoria principal o el disco, lo que es más lento que leer desde la memoria caché del búfer.

datos fríos

Datos a los que se accede con poca frecuencia y que suelen ser históricos. Al consultar este tipo de datos, normalmente se aceptan consultas lentas. Trasladar estos datos a niveles o clases de almacenamiento de menor rendimiento y menos costosos puede reducir los costos.

visión artificial (CV)

Campo de la IA que utiliza el machine learning para analizar y extraer información de formatos visuales, como imágenes y videos digitales. Por ejemplo, Amazon SageMaker AI proporciona algoritmos de procesamiento de imágenes para CV.

deriva de configuración

En el caso de una carga de trabajo, un cambio en la configuración con respecto al estado esperado. Podría provocar que la carga de trabajo deje de cumplir las normas y, por lo general, es gradual e involuntaria.

base de datos de administración de configuración (CMDB)

Repositorio que almacena y administra información sobre una base de datos y su entorno de TI, incluidos los componentes de hardware y software y sus configuraciones. Por lo general, los datos de una CMDB se utilizan en la etapa de detección y análisis de la cartera de productos durante la migración.

paquete de conformidad

Un conjunto de AWS Config reglas y medidas correctivas que puede reunir para personalizar sus controles de conformidad y seguridad. Puede implementar un paquete de conformidad como una entidad única en una región Cuenta de AWS y, o en una organización, mediante una plantilla YAML. Para obtener más información, consulta los paquetes de conformidad en la documentación. AWS Config

integración y entrega continuas () CI/CD

El proceso de automatización de las etapas de origen, creación, prueba, puesta en escena y producción del proceso de publicación del software. CI/CD se describe comúnmente como una canalización. CI/CD puede ayudarlo a automatizar los procesos, mejorar la productividad, mejorar

la calidad del código y entregar más rápido. Para obtener más información, consulte Beneficios de la entrega continua. CD también puede significar implementación continua. Para obtener más información, consulte Entrega continua frente a implementación continua.

CV

Consulte visión artificial.

D

datos en reposo

Datos que están estacionarios en la red, como los datos que se encuentran almacenados.

clasificación de datos

Un proceso para identificar y clasificar los datos de su red en función de su importancia y sensibilidad. Es un componente fundamental de cualquier estrategia de administración de riesgos de ciberseguridad porque lo ayuda a determinar los controles de protección y retención adecuados para los datos. La clasificación de los datos es un componente del pilar de seguridad del AWS Well-Architected Framework. Para obtener más información, consulte Clasificación de datos.

deriva de datos

Una variación significativa entre los datos de producción y los datos que se utilizaron para entrenar un modelo de machine learning, o un cambio significativo en los datos de entrada a lo largo del tiempo. La deriva de datos puede reducir la calidad, la precisión y la imparcialidad generales de las predicciones de los modelos de machine learning.

datos en tránsito

Datos que se mueven de forma activa por la red, por ejemplo, entre los recursos de la red.

malla de datos

Marco de arquitectura que proporciona una propiedad de datos distribuida y descentralizada con una administración y una gobernanza centralizadas.

minimización de datos

El principio de recopilar y procesar solo los datos estrictamente necesarios. Practicar la minimización de los datos Nube de AWS puede reducir los riesgos de privacidad, los costos y la huella de carbono de la analítica.

perímetro de datos

Un conjunto de barreras preventivas en su AWS entorno que ayudan a garantizar que solo las identidades confiables accedan a los recursos confiables desde las redes esperadas. Para obtener más información, consulte Crear un perímetro de datos sobre. AWS

preprocesamiento de datos

Transformar los datos sin procesar en un formato que su modelo de ML pueda analizar fácilmente. El preprocesamiento de datos puede implicar eliminar determinadas columnas o filas y corregir los valores faltantes, incoherentes o duplicados.

procedencia de los datos

El proceso de rastrear el origen y el historial de los datos a lo largo de su ciclo de vida, por ejemplo, la forma en que se generaron, transmitieron y almacenaron los datos.

titular de los datos

Persona cuyos datos se recopilan y procesan.

almacenamiento de datos

Sistema de administración de datos que respalda la inteligencia empresarial, como los análisis. Los almacenes de datos suelen contener grandes cantidades de datos históricos y, por lo general, se utilizan para las consultas y los análisis.

lenguaje de definición de datos (DDL)

Instrucciones o comandos para crear o modificar la estructura de tablas y objetos de una base de datos.

lenguaje de manipulación de datos (DML)

Instrucciones o comandos para modificar (insertar, actualizar y eliminar) la información de una base de datos.

DDL

Consulte lenguaje de definición de bases de datos.

conjunto profundo

Combinar varios modelos de aprendizaje profundo para la predicción. Puede utilizar conjuntos profundos para obtener una predicción más precisa o para estimar la incertidumbre de las predicciones.

aprendizaje profundo

Un subcampo del ML que utiliza múltiples capas de redes neuronales artificiales para identificar el mapeo entre los datos de entrada y las variables objetivo de interés.

defensa en profundidad

Un enfoque de seguridad de la información en el que se distribuyen cuidadosamente una serie de mecanismos y controles de seguridad en una red informática para proteger la confidencialidad, la integridad y la disponibilidad de la red y de los datos que contiene. Al adoptar esta estrategia AWS, se añaden varios controles en diferentes capas de la AWS Organizations estructura para ayudar a proteger los recursos. Por ejemplo, un enfoque de defensa en profundidad podría combinar la autenticación multifactor, la segmentación de la red y el cifrado.

administrador delegado

En AWS Organizations, un servicio compatible puede registrar una cuenta de AWS miembro para administrar las cuentas de la organización y gestionar los permisos de ese servicio. Esta cuenta se denomina administrador delegado para ese servicio. Para obtener más información y una lista de servicios compatibles, consulte Servicios que funcionan con AWS Organizations en la documentación de AWS Organizations .

Implementación

El proceso de hacer que una aplicación, características nuevas o correcciones de código se encuentren disponibles en el entorno de destino. La implementación abarca implementar cambios en una base de código y, a continuación, crear y ejecutar esa base en los entornos de la aplicación.

entorno de desarrollo

Consulte entorno.

control de detección

Un control de seguridad que se ha diseñado para detectar, registrar y alertar después de que se produzca un evento. Estos controles son una segunda línea de defensa, ya que lo advierten sobre los eventos de seguridad que han eludido los controles preventivos establecidos. Para obtener más información, consulte Controles de detección en Implementación de controles de seguridad en AWS.

asignación de flujos de valor para el desarrollo (DVSM)

Proceso que se utiliza para identificar y priorizar las restricciones que afectan negativamente a la velocidad y la calidad en el ciclo de vida del desarrollo de software. DVSM amplía el proceso de asignación del flujo de valor diseñado originalmente para las prácticas de fabricación ajustada. Se centra en los pasos y los equipos necesarios para crear y transferir valor a través del proceso de desarrollo de software.

gemelo digital

Representación virtual de un sistema del mundo real, como un edificio, una fábrica, un equipo industrial o una línea de producción. Los gemelos digitales son compatibles con el mantenimiento predictivo, la supervisión remota y la optimización de la producción.

tabla de dimensiones

En un esquema en estrella, tabla más pequeña que contiene los atributos de datos sobre los datos cuantitativos en una tabla de hechos. Los atributos de la tabla de dimensiones suelen ser campos de texto o números discretos que se comportan como texto. Estos atributos se suelen utilizar para restringir consultas, filtrarlas y etiquetar los conjuntos de resultados.

desastre

Un evento que impide que una carga de trabajo o un sistema cumplan sus objetivos empresariales en su ubicación principal de implementación. Estos eventos pueden ser desastres naturales, fallos técnicos o el resultado de acciones humanas, como una configuración incorrecta involuntaria o un ataque de malware.

recuperación de desastres (DR)

Estrategia y proceso que utiliza para minimizar el tiempo de inactividad y la pérdida de datos a causa de un desastre. Para obtener más información, consulte Recuperación de cargas de trabajo ante desastres en AWS: Recuperación en la nube en el AWS Well-Architected marco.

DML

Consulte lenguaje de manipulación de bases de datos.

diseño basado en el dominio

Un enfoque para desarrollar un sistema de software complejo mediante la conexión de sus componentes a dominios en evolución, o a los objetivos empresariales principales, a los que sirve cada componente. Eric Evans introdujo este concepto en su libro Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston: Addison-Wesley Professional, 2003). Para

obtener información sobre cómo utilizar el diseño basado en dominios con el patrón de higos estranguladores, consulte Modernización gradual de los servicios web antiguos de ASP.NET Microsoft (ASMX) mediante contenedores y Amazon API Gateway.

DR

Consulte recuperación ante desastres.

Detección de desviaciones

Seguimiento de las desviaciones con respecto a una configuración con línea de base. Por ejemplo, puedes usarlo AWS CloudFormation para detectar desviaciones en los recursos del sistema o puedes usarlo AWS Control Tower para detectar cambios en tu landing zone que puedan afectar al cumplimiento de los requisitos de gobierno.

DVSM

Consulte asignación de flujos de valor para el desarrollo.

E

EDA

Consulte análisis de datos de tipo exploratorio.

EDI

Consulte intercambio electrónico de datos.

computación en la periferia

La tecnología que aumenta la potencia de cálculo de los dispositivos inteligentes en la periferia de una red de IoT. En comparación con la computación en la nube, la computación de periferia puede reducir la latencia de la comunicación y mejorar el tiempo de respuesta.

intercambio electrónico de datos (EDI)

Intercambio automatizado de documentos comerciales entre organizaciones. Para más información, consulte ¿Qué es el intercambio electrónico de datos?

cifrado

Proceso de computación que transforma datos de texto plano, que son legibles por humanos, en texto cifrado.

clave de cifrado

Cadena criptográfica de bits aleatorios que se genera mediante un algoritmo de cifrado. Las claves pueden variar en longitud y cada una se ha diseñado para ser impredecible y única.

endianidad

El orden en el que se almacenan los bytes en la memoria del ordenador. Big-endian los sistemas almacenan primero el byte más significativo. Little-endian los sistemas almacenan primero el byte menos significativo.

punto de conexión

Consulte punto de conexión de servicio.

servicio de punto de conexión

Servicio que puede alojar en una nube privada virtual (VPC) para compartir con otros usuarios. Puede crear un servicio de punto final con AWS PrivateLink entidades principales Cuentas de AWS o AWS Identity and Access Management (de IAM) y conceder permisos a ellas. Estas cuentas o entidades principales pueden conectarse a su servicio de punto de conexión de forma privada mediante la creación de puntos de conexión de VPC de interfaz. Para obtener más información, consulte Creación de un servicio de punto de conexión en la documentación de Amazon Virtual Private Cloud (Amazon VPC).

planificación de recursos empresariales (ERP)

Sistema que automatiza y administra los procesos empresariales clave (como la contabilidad, MES y la administración de proyectos) de una empresa.

cifrado de sobre

El proceso de cifrar una clave de cifrado con otra clave de cifrado. Para obtener más información, consulte el cifrado de sobres en la documentación de AWS Key Management Service (AWS KMS).

entorno

Una instancia de una aplicación en ejecución. Los siguientes son los tipos de entornos más comunes en la computación en la nube:

- entorno de desarrollo: instancia de una aplicación en ejecución que solo se encuentra disponible para el equipo principal responsable del mantenimiento de la aplicación. Los

entornos de desarrollo se utilizan para probar los cambios antes de promocionarlos a los entornos superiores. Este tipo de entorno a veces se denomina entorno de prueba.

- entornos inferiores: todos los entornos de desarrollo de una aplicación, como los que se utilizan para las compilaciones y pruebas iniciales.
- entorno de producción: instancia de una aplicación en ejecución a la que pueden acceder los usuarios finales. En un CI/CD proceso, el entorno de producción es el último entorno de implementación.
- entornos superiores: todos los entornos a los que pueden acceder usuarios que no sean del equipo de desarrollo principal. Esto puede incluir un entorno de producción, entornos de preproducción y entornos para las pruebas de aceptación por parte de los usuarios.

epopeya

En las metodologías ágiles, son categorías funcionales que ayudan a organizar y priorizar el trabajo. Las epopeyas brindan una descripción detallada de los requisitos y las tareas de implementación. Por ejemplo, las epopeyas AWS de seguridad de CAF incluyen la gestión de identidades y accesos, los controles de detección, la seguridad de la infraestructura, la protección de datos y la respuesta a incidentes. Para obtener más información sobre las epopeyas en la estrategia de migración de AWS , consulte la Guía de implementación del programa.

ERP

Consulte planificación de recursos empresariales.

análisis de datos de tipo exploratorio (EDA)

El proceso de analizar un conjunto de datos para comprender sus características principales. Se recopilan o agregan datos y, a continuación, se realizan las investigaciones iniciales para encontrar patrones, detectar anomalías y comprobar las suposiciones. El EDA se realiza mediante el cálculo de estadísticas resumidas y la creación de visualizaciones de datos.

F

tabla de hechos

Tabla central de un esquema en estrella. Almacena datos cuantitativos sobre operaciones empresariales. Por lo general, una tabla de hechos contiene dos tipos de columnas: las que contienen medidas y las que contienen una clave externa para una tabla de dimensiones.

Fail Fast

Filosofía que utiliza pruebas frecuentes e incrementales para reducir el ciclo de vida del desarrollo. Es una parte fundamental de los enfoques ágiles.

límite de aislamiento de errores

En el Nube de AWS, un límite, como una zona de disponibilidad Región de AWS, un plano de control o un plano de datos, que limita el efecto de una falla y ayuda a mejorar la resiliencia de las cargas de trabajo. Para más información, consulte AWS Fault Isolation Boundaries.

rama de característica

Consulte rama.

características

Los datos de entrada que se utilizan para hacer una predicción. Por ejemplo, en un contexto de fabricación, las características pueden ser imágenes que se capturan periódicamente desde la línea de fabricación.

importancia de las características

La importancia que tiene una característica para las predicciones de un modelo. Por lo general, esto se expresa como una puntuación numérica que se puede calcular mediante diversas técnicas, como las explicaciones aditivas de Shapley (SHAP) y los gradientes integrados. Para obtener más información, consulte Interpretabilidad del modelo de aprendizaje automático con. AWS

transformación de funciones

Optimizar los datos para el proceso de ML, lo que incluye enriquecer los datos con fuentes adicionales, escalar los valores o extraer varios conjuntos de información de un solo campo de datos. Esto permite que el modelo de ML se beneficie de los datos. Por ejemplo, si divide la fecha del “27 de mayo de 2021 00:15:37” en “jueves”, “mayo”, “2021” y “15”, puede ayudar al algoritmo de aprendizaje a aprender patrones matizados asociados a los diferentes componentes de los datos.

peticiones con pocos pasos

Proporcionar a un LLM una pequeña cantidad de ejemplos que demuestren la tarea y el resultado deseado antes de pedirle que lleve a cabo una tarea similar. Esta técnica es una aplicación del aprendizaje contextual, en el que los modelos aprenden a partir de ejemplos (tomas) integrados en las instrucciones. Few-shot Las indicaciones pueden ser eficaces para tareas que requieren

un formato, un razonamiento o un conocimiento del dominio específicos. Consulte también peticiones desde cero.

FGAC

Consulte control de acceso detallado.

control de acceso preciso (FGAC)

El uso de varias condiciones que tienen por objetivo permitir o denegar una solicitud de acceso.

migración relámpago

Método de migración de bases de datos que utiliza la replicación continua de datos mediante la captura de datos de cambio para migrar los datos en el menor tiempo posible, en lugar de utilizar un enfoque gradual. El objetivo es reducir al mínimo el tiempo de inactividad.

FM

Consulte modelo fundacional.

Modelo fundacional (FM)

Gran red neuronal de aprendizaje profundo que se entrenó con conjuntos de datos masivos de datos generalizados y no etiquetados. Los FM pueden hacer una amplia variedad de tareas generales, como comprender el lenguaje, generar texto e imágenes y conversar en lenguaje natural. Para más información, consulte ¿Qué son los modelos fundacionales?

Puerta de enlace FM

Un intermediario centralizado que controla y normaliza el acceso a los modelos básicos. También se conoce como puerta de enlace LLM.

G

IA generativa

Subconjunto de modelos de IA que se entrenaron con grandes cantidades de datos y que pueden utilizar una simple petición de texto para crear contenido y artefactos nuevos, como imágenes, videos, texto y audio. Para más información, consulte ¿Qué es la IA generativa?

bloqueo geográfico

Consulte restricciones geográficas.

restricciones geográficas (bloqueo geográfico)

En Amazon CloudFront, una opción para impedir que los usuarios de países específicos accedan a las distribuciones de contenido. Puede utilizar una lista de permitidos o bloqueados para especificar los países aprobados y prohibidos. Para obtener más información, consulta Restringir la distribución geográfica del contenido en la CloudFront documentación.

Flujo de trabajo de Gitflow

Un enfoque en el que los entornos inferiores y superiores utilizan diferentes ramas en un repositorio de código fuente. El flujo de trabajo de Gitflow se considera heredado, mientras que el flujo de trabajo basado en enlaces troncales es el enfoque moderno preferido.

imagen dorada

Instantánea de un sistema o software que se usa como plantilla para implementar nuevas instancias de ese sistema o software. Por ejemplo, en la fabricación, una imagen dorada se puede utilizar para aprovisionar software en varios dispositivos y ayuda a mejorar la velocidad, la escalabilidad y la productividad de las operaciones de fabricación de dispositivos.

estrategia de implementación desde cero

La ausencia de infraestructura existente en un entorno nuevo. Al adoptar una estrategia de implementación desde cero para una arquitectura de sistemas, puede seleccionar todas las tecnologías nuevas sin que estas deban ser compatibles con una infraestructura existente, lo que también se conoce como implementación sobre infraestructura existente. Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de implementación desde cero.

barrera de protección

Una regla de alto nivel que ayuda a regular los recursos, las políticas y la conformidad en todas las unidades organizativas (OU). Las barreras de protección preventivas aplican políticas para garantizar la alineación con los estándares de conformidad. Se implementan mediante políticas de control de servicios y límites de permisos de IAM. Las barreras de protección de detección detectan las vulneraciones de las políticas y los problemas de conformidad, y generan alertas para su corrección. Se implementan mediante Amazon AWS Config AWS Security Hub CSPM GuardDuty AWS Trusted Advisor, Amazon Inspector y AWS Lambda cheques personalizados.

barandas (AI)

Mecanismos de seguridad que filtran, validan y restringen las entradas y salidas de los agentes para ayudar a garantizar un comportamiento responsable y seguro de la IA.

H

HA

Consulte alta disponibilidad.

migración heterogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que utilice un motor de base de datos diferente (por ejemplo, de Oracle a Amazon Aurora). La migración heterogénea suele ser parte de un esfuerzo de rediseño de la arquitectura y convertir el esquema puede ser una tarea compleja. AWS ofrece AWS SCT, lo cual ayuda con las conversiones de esquemas.

alta disponibilidad (HA)

La capacidad de una carga de trabajo para funcionar de forma continua, sin intervención, en caso de desafíos o desastres. Los sistemas de alta disponibilidad están diseñados para realizar una conmutación por error automática, ofrecer un rendimiento de alta calidad de forma constante y gestionar diferentes cargas y fallos con un impacto mínimo en el rendimiento.

modernización histórica

Un enfoque utilizado para modernizar y actualizar los sistemas de tecnología operativa (TO) a fin de satisfacer mejor las necesidades de la industria manufacturera. Un histórico es un tipo de base de datos que se utiliza para recopilar y almacenar datos de diversas fuentes en una fábrica.

datos de reserva

Parte de los datos históricos etiquetados que se ocultan de un conjunto de datos que se utiliza para entrenar un modelo de machine learning. Puede utilizar los datos de reserva para evaluar el rendimiento del modelo mediante la comparación de las predicciones del modelo con los datos de reserva.

human-in-the-loop (HiTL)

Un patrón de flujo de trabajo en el que la ejecución de los agentes se detiene para su revisión y aprobación por parte de una persona en los puntos de decisión críticos.

migración homogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que comparte el mismo motor de base de datos (por ejemplo, Microsoft SQL Server a Amazon RDS para SQL Server).

La migración homogénea suele formar parte de un esfuerzo para volver a alojar o redefinir la plataforma. Puede utilizar las utilidades de bases de datos nativas para migrar el esquema.

datos recientes

Datos a los que se accede con frecuencia, como datos en tiempo real o datos traslacionales recientes. Por lo general, estos datos requieren un nivel o una clase de almacenamiento de alto rendimiento para proporcionar respuestas rápidas a las consultas.

hotfix

Una solución urgente para un problema crítico en un entorno de producción. Debido a su urgencia, una revisión suele realizarse fuera del flujo de trabajo habitual de las DevOps versiones.

periodo de hiperatención

Periodo, inmediatamente después de la transición, durante el cual un equipo de migración administra y monitorea las aplicaciones migradas en la nube para solucionar cualquier problema. Por lo general, este periodo dura de 1 a 4 días. Al final del periodo de hiperatención, el equipo de migración suele transferir la responsabilidad de las aplicaciones al equipo de operaciones en la nube.

I

IaC

Consulte infraestructura como código.

políticas basadas en identidades

Política asociada a uno o más directores de IAM que define sus permisos en el entorno. Nube de AWS

aplicación inactiva

Aplicación que utiliza un promedio de CPU y memoria de entre 5 y 20 por ciento durante un periodo de 90 días. En un proyecto de migración, es habitual retirar estas aplicaciones o mantenerlas en las instalaciones.

IIoT

Consulte Internet de las cosas industrial.

infraestructura inmutable

Modelo que implementa una nueva infraestructura para las cargas de trabajo de producción en lugar de actualizar o modificar la infraestructura existente o aplicarle revisiones. Las infraestructuras inmutables son de manera intrínseca más coherentes, fiables y predecibles que las infraestructuras mutables. Para obtener más información, consulte las mejores prácticas del Framework para implementar con una infraestructura inmutable. AWS Well-Architected

VPC entrante (de entrada)

En una arquitectura de AWS cuentas múltiples, una VPC que acepta, inspecciona y enruta las conexiones de red desde fuera de una aplicación. La Arquitectura de referencia de seguridad de AWS recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

migración gradual

Estrategia de transición en la que se migra la aplicación en partes pequeñas en lugar de realizar una transición única y completa. Por ejemplo, puede trasladar inicialmente solo unos pocos microservicios o usuarios al nuevo sistema. Tras comprobar que todo funciona correctamente, puede trasladar microservicios o usuarios adicionales de forma gradual hasta que pueda retirar su sistema heredado. Esta estrategia reduce los riesgos asociados a las grandes migraciones.

Industria 4.0

Un término que Klaus Schwab introdujo en 2016 para referirse a la modernización de los procesos de fabricación mediante avances en la conectividad, los datos en tiempo real, la automatización, el análisis y. AI/ML

infraestructura

Todos los recursos y activos que se encuentran en el entorno de una aplicación.

infraestructura como código (IaC)

Proceso de aprovisionamiento y administración de la infraestructura de una aplicación mediante un conjunto de archivos de configuración. La IaC se ha diseñado para ayudarlo a centralizar la administración de la infraestructura, estandarizar los recursos y escalar con rapidez a fin de que los entornos nuevos sean repetibles, fiables y consistentes.

Internet de las cosas industrial (IIoT)

El uso de sensores y dispositivos conectados a Internet en los sectores industriales, como el productivo, el eléctrico, el automotriz, el sanitario, el de las ciencias de la vida y el de la

agricultura. Para obtener más información, consulte Creación de una estrategia de transformación digital del Internet de las cosas industrial (IIoT).

VPC de inspección

En una arquitectura de AWS cuentas múltiples, una VPC centralizada que gestiona las inspecciones del tráfico de red entre las VPC (iguales o Regiones de AWS diferentes), Internet y las redes locales. La Arquitectura de referencia de seguridad de AWS recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

Internet de las cosas (IoT)

Red de objetos físicos conectados con sensores o procesadores integrados que se comunican con otros dispositivos y sistemas a través de Internet o de una red de comunicación local. Para obtener más información, consulte ¿Qué es IoT?.

interpretabilidad

Característica de un modelo de machine learning que describe el grado en que un ser humano puede entender cómo las predicciones del modelo dependen de sus entradas. Para obtener más información, consulte Interpretabilidad del modelo de aprendizaje automático con. AWS

IoT

Consulte Internet de las cosas.

biblioteca de información de TI (ITIL)

Conjunto de prácticas recomendadas para ofrecer servicios de TI y alinearlos con los requisitos empresariales. La ITIL proporciona la base para la ITSM.

administración de servicios de TI (ITSM)

Actividades asociadas con el diseño, la implementación, la administración y el soporte de los servicios de TI para una organización. Para obtener información sobre la integración de las operaciones en la nube con las herramientas de ITSM, consulte la Guía de integración de operaciones.

ITIL

Consulte biblioteca de información de TI.

ITSM

Consulte administración de servicios de TI.

L

control de acceso basado en etiquetas (LBAC)

Una implementación del control de acceso obligatorio (MAC) en la que a los usuarios y a los propios datos se les asigna explícitamente un valor de etiqueta de seguridad. La intersección entre la etiqueta de seguridad del usuario y la etiqueta de seguridad de los datos determina qué filas y columnas puede ver el usuario.

zona de aterrizaje

Una landing zone es un AWS entorno multicuenta bien diseñado, escalable y seguro. Este es un punto de partida desde el cual las empresas pueden lanzar e implementar rápidamente cargas de trabajo y aplicaciones con confianza en su entorno de seguridad e infraestructura. Para obtener más información sobre las zonas de aterrizaje, consulte Configuración de un entorno de AWS seguro y escalable con varias cuentas.

modelo de lenguaje de gran tamaño (LLM)

Modelo de IA de aprendizaje profundo que se entrenó previamente con una gran cantidad de datos. Un LLM puede llevar a cabo varias tareas, como responder preguntas, resumir documentos, traducir textos a otros idiomas y completar oraciones. Para más información, consulte ¿Qué es un LLM (modelo de lenguaje de gran tamaño)?

migración grande

Migración de 300 servidores o más.

LBAC

Consulte control de acceso basado en etiquetas.

privilegio mínimo

La práctica recomendada de seguridad que consiste en conceder los permisos mínimos necesarios para realizar una tarea. Para obtener más información, consulte Aplicar permisos de privilegio mínimo en la documentación de IAM.

migrar mediante lift-and-shift

Consulte Las 7 R.

sistema little-endian

Un sistema que almacena primero el byte menos significativo. Consulte también endianidad.

LLM

Consulte modelo de lenguaje de gran tamaño.

entornos inferiores

Consulte entorno.

M

machine learning (ML)

Un tipo de inteligencia artificial que utiliza algoritmos y técnicas para el reconocimiento y el aprendizaje de patrones. El ML analiza y aprende de los datos registrados, como los datos del Internet de las cosas (IoT), para generar un modelo estadístico basado en patrones. Para más información, consulte Machine learning.

rama principal

Consulte rama.

malware

Software diseñado para comprometer la seguridad o la privacidad de la computadora. El malware podría interrumpir los sistemas informáticos, filtrar información confidencial u obtener acceso no autorizado. Algunos ejemplos de malware son los virus, los gusanos, el ransomware, los troyanos, el spyware y los registradores de pulsaciones de teclas.

Servicios administrados

Servicios de AWS en el que AWS opera la capa de infraestructura, el sistema operativo y las plataformas, y se accede a los puntos finales para almacenar y recuperar datos. Amazon Simple Storage Service (Amazon S3) y Amazon DynamoDB son ejemplos de servicios administrados. También se conocen como servicios abstractos.

sistema de ejecución de fabricación (MES)

Sistema de software para seguir, supervisar, documentar y controlar los procesos de producción que convierten las materias primas en productos acabados en la zona de producción.

MAP

Consulte Programa de aceleración de la migración.

MCP

Consulte Model Context Protocol.

Protocolo de contexto para modelos (MCP)

Un protocolo sin estado para la comunicación entre el agente y la herramienta.

Servidor MCP

Un servicio que expone una o más herramientas a través del protocolo Model Context.

mecanismo

Proceso completo mediante el que se crea una herramienta, se impulsa su adopción y, a continuación, se inspeccionan los resultados para hacer ajustes. Un mecanismo es un ciclo que se refuerza y mejora por sí mismo a medida que funciona. Para obtener más información, consulte Creación de mecanismos en el AWS Well-Architected marco.

cuenta de miembro

Todas las Cuentas de AWS demás cuentas, excepto la de administración, que forman parte de una organización AWS Organizations. Una cuenta no puede pertenecer a más de una organización a la vez.

MES

Consulte sistema de ejecución de fabricación.

Message Queuing Telemetry Transport (MQTT)

Un protocolo de comunicación ligero de máquina a máquina (M2M), basado en el publish/subscribepatrón, para dispositivos de IoT con recursos limitados.

microservicio

Un servicio pequeño e independiente que se comunica a través de API bien definidas y que, por lo general, es propiedad de equipos pequeños e independientes. Por ejemplo, un sistema de seguros puede incluir microservicios que se adapten a las capacidades empresariales, como las de ventas o marketing, o a subdominios, como las de compras, reclamaciones o análisis. Los beneficios de los microservicios incluyen la agilidad, la escalabilidad flexible, la facilidad de implementación, el código reutilizable y la resiliencia. Para obtener más información, consulte Integrar microservicios mediante servicios sin servidor. AWS

arquitectura de microservicios

Un enfoque para crear una aplicación con componentes independientes que ejecutan cada proceso de la aplicación como un microservicio. Estos microservicios se comunican a través de una interfaz bien definida mediante API ligeras. Cada microservicio de esta arquitectura se puede actualizar, implementar y escalar para satisfacer la demanda de funciones específicas de una aplicación. Para obtener más información, consulte Implementación de microservicios en. AWS

Programa de aceleración de la migración (MAP)

Un AWS programa que proporciona soporte de consultoría, formación y servicios para ayudar a las organizaciones a crear una base operativa sólida para migrar a la nube y para ayudar a compensar el costo inicial de las migraciones. El MAP incluye una metodología de migración para ejecutar las migraciones antiguas de forma metódica y un conjunto de herramientas para automatizar y acelerar los escenarios de migración más comunes.

migración a escala

Proceso de transferencia de la mayoría de la cartera de aplicaciones a la nube en oleadas, con más aplicaciones desplazadas a un ritmo más rápido en cada oleada. En esta fase, se utilizan las prácticas recomendadas y las lecciones aprendidas en las fases anteriores para implementar una fábrica de migración de equipos, herramientas y procesos con el fin de agilizar la migración de las cargas de trabajo mediante la automatización y la entrega ágil. Esta es la tercera fase de la estrategia de migración de AWS.

fábrica de migración

Cross-functional equipos que agilizan la migración de las cargas de trabajo mediante enfoques ágiles y automatizados. Los equipos de las fábricas de migración suelen estar compuestos por analistas y propietarios de operaciones, ingenieros de migración, desarrolladores y DevOps profesionales que trabajan a pasos agigantados. Entre el 20 y el 50 por ciento de la cartera de aplicaciones empresariales se compone de patrones repetidos que pueden optimizarse mediante un enfoque de fábrica. Para obtener más información, consulte la discusión sobre las fábricas de migración y la Guía de fábricas de migración a la nube en este contenido.

metadatos de migración

Información sobre la aplicación y el servidor que se necesita para completar la migración. Cada patrón de migración requiere un conjunto diferente de metadatos de migración. Algunos ejemplos de metadatos de migración son la subred de destino, el grupo de seguridad y AWS la cuenta.

patrón de migración

Tarea de migración repetible que detalla la estrategia de migración, el destino de la migración y la aplicación o el servicio de migración utilizados. Ejemplo: rehospede la migración a Amazon EC2 AWS con Application Migration Service.

Migration Portfolio Assessment (MPA)

Herramienta en línea que proporciona información a fin de validar los argumentos comerciales necesarios para migrar a la Nube de AWS. La MPA ofrece una evaluación detallada de la cartera (adecuación del tamaño de los servidores, precios, comparaciones del costo total de propiedad, análisis de los costos de migración), así como una planificación de la migración (análisis y recopilación de datos de aplicaciones, agrupación de aplicaciones, priorización de la migración y planificación de oleadas). La herramienta MPA (requiere iniciar sesión) está disponible de forma gratuita para todos los AWS consultores y consultores de los socios de APN.

Evaluación de la preparación para la migración (MRA)

Proceso que consiste en obtener información sobre el estado de preparación de una organización para la nube, identificar sus puntos fuertes y débiles y elaborar un plan de acción para cerrar las brechas identificadas mediante el AWS CAF. Para obtener más información, consulte la Guía de preparación para la migración. La MRA es la primera fase de la estrategia de migración de AWS.

estrategia de migración

Enfoque utilizado para migrar una carga de trabajo a la Nube de AWS. Para más información, consulte la entrada Las 7 R de este glosario y también Mobilize your organization to accelerate large-scale migrations.

ML

Consulte machine learning.

modernización

Transformar una aplicación obsoleta (antigua o monolítica) y su infraestructura en un sistema ágil, elástico y de alta disponibilidad en la nube para reducir los gastos, aumentar la eficiencia y aprovechar las innovaciones. Para más información, consulte Strategy for modernizing applications in the Nube de AWS.

evaluación de la preparación para la modernización

Evaluación que ayuda a determinar la preparación para la modernización de las aplicaciones de una organización; identifica los beneficios, los riesgos y las dependencias; y determina qué

tan bien la organización puede soportar el estado futuro de esas aplicaciones. El resultado de la evaluación es un esquema de la arquitectura objetivo, una hoja de ruta que detalla las fases de desarrollo y los hitos del proceso de modernización y un plan de acción para abordar las brechas identificadas. Para más información, consulte Evaluating modernization readiness for applications in the Nube de AWS.

aplicaciones monolíticas (monolitos)

Aplicaciones que se ejecutan como un único servicio con procesos estrechamente acoplados. Las aplicaciones monolíticas presentan varios inconvenientes. Si una característica de la aplicación experimenta un aumento en la demanda, se debe escalar toda la arquitectura. Agregar o mejorar las características de una aplicación monolítica también se vuelve más complejo a medida que crece la base de código. Para solucionar problemas con la aplicación, puede utilizar una arquitectura de microservicios. Para obtener más información, consulte Descomposición de monolitos en microservicios.

MPA

Consulte Migration Portfolio Assessment.

MQTT

Consulte Message Queuing Telemetry Transport.

clasificación multiclase

Un proceso que ayuda a generar predicciones para varias clases (predice uno de más de dos resultados). Por ejemplo, un modelo de ML podría preguntar “¿Este producto es un libro, un automóvil o un teléfono?” o “¿Qué categoría de productos es más interesante para este cliente?”.

infraestructura mutable

Modelo que actualiza y modifica la infraestructura actual para las cargas de trabajo de producción. Para mejorar la coherencia, la confiabilidad y la previsibilidad, el AWS Well-Architected Marco recomienda el uso de una infraestructura inmutable como práctica recomendada.

O

OAC

Consulte control de acceso de origen.

OAI

Consulte identidad de acceso de origen.

OCM

Consulte administración del cambio organizacional.

migración fuera de línea

Método de migración en el que la carga de trabajo de origen se elimina durante el proceso de migración. Este método implica un tiempo de inactividad prolongado y, por lo general, se utiliza para cargas de trabajo pequeñas y no críticas.

OI

Consulte integración de operaciones.

OLA

Consulte acuerdo de nivel operativo.

migración en línea

Método de migración en el que la carga de trabajo de origen se copia al sistema de destino sin que se desconecte. Las aplicaciones que están conectadas a la carga de trabajo pueden seguir funcionando durante la migración. Este método implica un tiempo de inactividad nulo o mínimo y, por lo general, se utiliza para cargas de trabajo de producción críticas.

OPC-UA

Consulte Open Process Communications: arquitectura unificada.

Comunicaciones de proceso abierto: arquitectura unificada () OPC-UA

Un protocolo de comunicación de máquina a máquina (M2M) para la automatización industrial. OPC-UA proporciona un estándar de interoperabilidad con esquemas de cifrado, autenticación y autorización de datos.

acuerdo de nivel operativo (OLA)

Acuerdo que aclara lo que los grupos de TI operativos se comprometen a ofrecerse entre sí, para respaldar un acuerdo de nivel de servicio (SLA).

revisión de la preparación operativa (ORR)

Lista de comprobación de preguntas y prácticas recomendadas asociadas que son útiles para comprender, evaluar, prevenir o reducir el alcance de los incidentes y posibles errores. Para

obtener más información, consulte las revisiones de preparación operativa (ORR) en el AWS Well-Architected marco.

tecnología operativa (TO)

Sistemas de hardware y software que funcionan con el entorno físico para controlar las operaciones, los equipos y la infraestructura industriales. En el sector de la fabricación, la integración de los sistemas de TO y tecnología de la información (TI) es un enfoque clave para las transformaciones de la industria 4.0.

integración de operaciones (OI)

Proceso de modernización de las operaciones en la nube, que implica la planificación de la preparación, la automatización y la integración. Para obtener más información, consulte la Guía de integración de las operaciones.

registro de seguimiento organizativo

Un registro creado por y AWS CloudTrail que registra todos los eventos Cuentas de AWS de una organización AWS Organizations. Este registro de seguimiento se crea en cada Cuenta de AWS que forma parte de la organización y realiza un seguimiento de la actividad en cada cuenta. Para obtener más información, consulte Crear un registro para una organización en la CloudTrail documentación.

administración del cambio organizacional (OCM)

Marco para administrar las transformaciones empresariales importantes y disruptivas desde la perspectiva de las personas, la cultura y el liderazgo. La OCM ayuda a las empresas a prepararse para nuevos sistemas y estrategias y a realizar la transición a ellos, al acelerar la adopción de cambios, abordar los problemas de transición e impulsar cambios culturales y organizacionales. En la estrategia de AWS migración, este marco se denomina aceleración de personal, debido a la velocidad de cambio que requieren los proyectos de adopción de la nube. Para obtener más información, consulte la Guía de OCM.

control de acceso de origen (OAC)

En CloudFront, una opción mejorada para restringir el acceso y proteger el contenido del Amazon Simple Storage Service (Amazon S3). El OAC admite todos los buckets de S3 Regiones de AWS, el cifrado del lado del servidor con AWS KMS (SSE-KMS) y `DELETE` las solicitudes `PUT` y dinámicas al bucket de S3.

identidad de acceso de origen (OAI)

En CloudFront, una opción para restringir el acceso y proteger el contenido de Amazon S3. Cuando utiliza OAI, CloudFront crea un principal con el que Amazon S3 puede autenticarse. Los directores autenticados solo pueden acceder al contenido de un bucket de S3 a través de una distribución específica. CloudFront Consulte también el OAC, que proporciona un control de acceso más detallado y mejorado.

ORR

Consulte revisión de la preparación operativa.

OT

Consulte tecnología operativa.

VPC saliente (de salida)

En una arquitectura de AWS cuentas múltiples, una VPC que gestiona las conexiones de red que se inician desde una aplicación. La Arquitectura de referencia de seguridad de AWS recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

P

límite de permisos

Una política de administración de IAM que se adjunta a las entidades principales de IAM para establecer los permisos máximos que puede tener el usuario o el rol. Para obtener más información, consulte Límites de permisos en la documentación de IAM.

información de identificación personal (PII)

Información que, vista directamente o combinada con otros datos relacionados, puede utilizarse para deducir de manera razonable la identidad de una persona. Algunos ejemplos de información de identificación personal son los nombres, las direcciones y la información de contacto.

PII

Consulte información de identificación personal.

manual de estrategias

Conjunto de pasos predefinidos que capturan el trabajo asociado a las migraciones, como la entrega de las funciones de operaciones principales en la nube. Un manual puede adoptar la forma de scripts, manuales de procedimientos automatizados o resúmenes de los procesos o pasos necesarios para operar un entorno modernizado.

PLC

Consulte controlador lógico programable.

PLM

Consulte administración del ciclo de vida del producto.

policy

Objeto que puede definir permisos (consulte política basada en identidad), especificar las condiciones de acceso (consulte política basada en recursos) o definir los permisos máximos para todas las cuentas de una organización de AWS Organizations (consulte política de control de servicio).

persistencia políglota

Elegir de forma independiente la tecnología de almacenamiento de datos de un microservicio en función de los patrones de acceso a los datos y otros requisitos. Si sus microservicios tienen la misma tecnología de almacenamiento de datos, pueden enfrentarse a desafíos de implementación o experimentar un rendimiento deficiente. Los microservicios se implementan más fácilmente y logran un mejor rendimiento y escalabilidad si utilizan el almacén de datos que mejor se adapte a sus necesidades.

evaluación de cartera

Proceso de detección, análisis y priorización de la cartera de aplicaciones para planificar la migración. Para obtener más información, consulte la Evaluación de la preparación para la migración.

predicate

Condición de consulta que devuelve `true` o `false`. En general, se encuentra en una cláusula `WHERE`.

inserción de predicados

Técnica de optimización de consultas en bases de datos que filtra los datos de la consulta antes de transferirlos. Esta técnica reduce la cantidad de datos de la base de datos relacional que se tienen que recuperar y procesar. Además, mejora el rendimiento de las consultas.

control preventivo

Un control de seguridad diseñado para evitar que ocurra un evento. Estos controles son la primera línea de defensa para evitar el acceso no autorizado o los cambios no deseados en la red. Para obtener más información, consulte Controles preventivos en Implementación de controles de seguridad en AWS.

entidad principal

Una entidad AWS que puede realizar acciones y acceder a los recursos. Esta entidad suele ser un usuario raíz para un Cuenta de AWS rol de IAM o un usuario. Para obtener más información, consulte Entidad principal en Términos y conceptos de roles en la documentación de IAM.

Privacidad desde el diseño

Enfoque de ingeniería de sistemas que tiene en cuenta la privacidad durante todo el proceso de desarrollo.

zonas alojadas privadas

Contenedor que aloja información acerca de cómo desea que responda Amazon Route 53 a las consultas de DNS de un dominio y sus subdominios en una o varias VPC. Para obtener más información, consulte Uso de zonas alojadas privadas en la documentación de Route 53.

control proactivo

Control de seguridad que se diseñó para evitar la implementación de recursos que no cumplan con la normativa. Estos controles analizan los recursos antes de aprovisionarlos. Si el recurso no cumple con los requisitos del control, no se aprovisiona. Para obtener más información, consulte la guía de referencia de controles en la AWS Control Tower documentación y consulte Controles proactivos en Implementación de controles de seguridad en AWS.

administración del ciclo de vida del producto (PLM)

Administración de los datos y los procesos de un producto a lo largo de todo su ciclo de vida, desde el diseño, el desarrollo y el lanzamiento, pasando por el crecimiento y la madurez, hasta la reducción de su uso y su retirada.

entorno de producción

Consulte entorno.

controlador lógico programable (PLC)

En el sector de la fabricación, computadora adaptable y altamente fiable que supervisa las máquinas y automatiza los procesos de fabricación.

encadenamiento de peticiones

Uso de la salida de una petición de LLM como entrada para la siguiente petición a fin de generar mejores respuestas. Esta técnica se utiliza para dividir una tarea compleja en tareas secundarias o para refinar o ampliar de forma iterativa una respuesta preliminar. Ayuda a mejorar la precisión y la relevancia de las respuestas de un modelo y permite obtener resultados más detallados y personalizados.

seudonimización

El proceso de reemplazar los identificadores personales de un conjunto de datos por valores de marcadores de posición. La seudonimización puede ayudar a proteger la privacidad personal. Los datos seudonimizados siguen considerándose datos personales.

publish/subscribe (pub/sub)

Patrón que permite establecer comunicaciones asíncronas entre microservicios para mejorar la escalabilidad y la capacidad de respuesta. Por ejemplo, en un MES basado en microservicios, un microservicio puede publicar mensajes de eventos en un canal al que se pueden suscribir otros microservicios. El sistema puede agregar nuevos microservicios sin cambiar el servicio de publicación.

Q

plan de consulta

Serie de pasos, como instrucciones, que se utilizan para acceder a los datos de un sistema de base de datos relacional SQL.

regresión del plan de consulta

El optimizador de servicios de la base de datos elige un plan menos óptimo que antes de un cambio determinado en el entorno de la base de datos. Los cambios en estadísticas,

restricciones, configuración del entorno, enlaces de parámetros de consultas y actualizaciones del motor de base de datos PostgreSQL pueden provocar una regresión del plan.

R

Matriz RACI

Consulte responsable, fiable, consultada e informada (RACI).

RAG

Consulte generación aumentada por recuperación.

ransomware

Software malicioso que se ha diseñado para bloquear el acceso a un sistema informático o a los datos hasta que se efectúe un pago.

Matriz RASCI

Consulte responsable, fiable, consultada e informada (RACI).

RCAC

Consulte control de acceso por filas y columnas.

réplica de lectura

Una copia de una base de datos que se utiliza con fines de solo lectura. Puede enrutar las consultas a la réplica de lectura para reducir la carga en la base de datos principal.

rediseñar

Consulte Las 7 R.

objetivo de punto de recuperación (RPO)

La cantidad de tiempo máximo aceptable desde el último punto de recuperación de datos. Esto determina qué se considera una pérdida de datos aceptable entre el último punto de recuperación y la interrupción del servicio.

objetivo de tiempo de recuperación (RTO)

La demora máxima aceptable entre la interrupción del servicio y el restablecimiento del servicio.

refactorizar

Consulte Las 7 R.

Region

Conjunto de AWS recursos en un área geográfica. Cada uno Región de AWS está aislado e independiente de los demás para proporcionar tolerancia a las fallas, estabilidad y resiliencia. Para más información, consulte Specify which Regiones de AWS your account can use.

regresión

Una técnica de ML que predice un valor numérico. Por ejemplo, para resolver el problema de "¿A qué precio se venderá esta casa?", un modelo de ML podría utilizar un modelo de regresión lineal para predecir el precio de venta de una vivienda en función de datos conocidos sobre ella (por ejemplo, los metros cuadrados).

volver a alojar

Consulte Las 7 R.

versión

En un proceso de implementación, el acto de promover cambios en un entorno de producción.

reubicar

Consulte Las 7 R.

redefinir la plataforma

Consulte Las 7 R.

recomprar

Consulte Las 7 R.

resiliencia

Capacidad de una aplicación para resistir interrupciones o recuperarse de ellas. Al planificar la resiliencia en la Nube de AWS, la alta disponibilidad y la recuperación ante desastres son consideraciones comunes. Para más información, consulte Resiliencia en la Nube de AWS.

política basada en recursos

Una política asociada a un recurso, como un bucket de Amazon S3, un punto de conexión o una clave de cifrado. Este tipo de política especifica a qué entidades principales se les permite el acceso, las acciones compatibles y cualquier otra condición que deba cumplirse.

matriz responsable, confiable, consultada e informada (RACI)

Una matriz que define las funciones y responsabilidades de todas las partes involucradas en las actividades de migración y las operaciones de la nube. El nombre de la matriz se deriva de los tipos de responsabilidad definidos en la matriz: responsable (R), contable (A), consultado (C) e informado (I). El tipo de soporte (S) es opcional. Si incluye el soporte, la matriz se denomina matriz RASCI y, si la excluye, se denomina matriz RACI.

control receptivo

Un control de seguridad que se ha diseñado para corregir los eventos adversos o las desviaciones con respecto a su base de seguridad. Para obtener más información, consulte Controles receptivos en Implementación de controles de seguridad en AWS.

retain

Consulte Las 7 R.

retirar

Consulte Las 7 R.

Generación aumentada de recuperación (RAG)

Tecnología de IA generativa mediante la que un LLM hace referencia a un origen de datos autorizado que se encuentra fuera de sus orígenes de datos de entrenamiento antes de generar una respuesta. Por ejemplo, un modelo de RAG podría hacer una búsqueda semántica en la base de conocimientos o en los datos personalizados de una organización. Para más información, consulte ¿Qué es RAG (generación aumentada por recuperación)?

rotación

Proceso mediante el que periódicamente se actualiza un secreto para que resulte más difícil que un atacante pueda acceder a las credenciales.

control de acceso por filas y columnas (RCAC)

El uso de expresiones SQL básicas y flexibles que tienen reglas de acceso definidas. El RCAC consta de permisos de fila y máscaras de columnas.

RPO

Consulte objetivo de punto de recuperación.

RTO

Consulte objetivo de tiempo de recuperación.

manual de procedimientos

Conjunto de procedimientos manuales o automatizados necesarios para realizar una tarea específica. Por lo general, se diseñan para agilizar las operaciones o los procedimientos repetitivos con altas tasas de error.

S

SAML 2.0

Un estándar abierto que utilizan muchos proveedores de identidad (IdPs). Esta función permite el inicio de sesión único (SSO) federado, de modo que los usuarios pueden iniciar sesión Consola de administración de AWS o llamar a las operaciones de la AWS API sin tener que crear un usuario en IAM para todos los miembros de la organización. Para obtener más información sobre la federación basada en SAML 2.0, consulte Acerca de la federación basada en SAML 2.0 en la documentación de IAM.

SCADA

Consulte control de supervisión y adquisición de datos.

SCP

Consulte política de control de servicio.

secreta

En AWS Secrets Manager, información confidencial o restringida, como una contraseña o credenciales de usuario, que se almacena de forma cifrada. Se compone del valor del secreto y de sus metadatos. El valor del secreto puede ser binario, una sola cadena o varias cadenas. Para más información, consulte What's in a Secrets Manager secret? en la documentación de Secrets Manager.

seguridad desde el diseño

Enfoque de ingeniería de sistemas que tiene en cuenta la seguridad durante todo el proceso de desarrollo.

control de seguridad

Barrera de protección técnica o administrativa que impide, detecta o reduce la capacidad de un agente de amenazas para aprovechar una vulnerabilidad de seguridad. Existen cuatro tipos de controles de seguridad principales: preventivos, de detección, de respuesta y proactivos.

refuerzo de la seguridad

Proceso de reducir la superficie expuesta a ataques para hacerla más resistente a los ataques. Esto puede incluir acciones, como la eliminación de los recursos que ya no se necesitan, la implementación de prácticas recomendadas de seguridad consistente en conceder privilegios mínimos o la desactivación de características innecesarias en los archivos de configuración.

sistema de información sobre seguridad y administración de eventos (SIEM)

Herramientas y servicios que combinan sistemas de administración de información sobre seguridad (SIM) y de administración de eventos de seguridad (SEM). Un sistema de SIEM recopila, monitorea y analiza los datos de servidores, redes, dispositivos y otras fuentes para detectar amenazas y brechas de seguridad y generar alertas.

automatización de la respuesta de seguridad

Acción predefinida y programada que está diseñada para responder automáticamente a un evento de seguridad o corregirlo. Estas automatizaciones sirven como controles de seguridad preventivos o adaptables que le ayudan a implementar las mejores prácticas AWS de seguridad. La modificación de un grupo de seguridad de VPC, la aplicación de revisiones a una instancia de Amazon EC2 o la rotación de credenciales son algunos ejemplos de acciones de respuesta automatizadas.

cifrado del servidor

Cifrado de los datos en su destino, por parte de Servicio de AWS quien los recibe.

política de control de servicio (SCP)

Una política que proporciona un control centralizado de los permisos de todas las cuentas de una organización en AWS Organizations. Las SCP definen barreras de protección o establecen límites a las acciones que un administrador puede delegar en los usuarios o roles. Puede utilizar las SCP como listas de permitidos o rechazados, para especificar qué servicios o acciones se encuentra permitidos o prohibidos. Para obtener más información, consulte las políticas de control del servicio en la AWS Organizations documentación.

punto de enlace de servicio

La URL del punto de entrada de un Servicio de AWS. Para conectarse mediante programación a un servicio de destino, puede utilizar un punto de conexión. Para obtener más información, consulte Puntos de conexión de Servicio de AWS en Referencia general de AWS.

acuerdo de nivel de servicio (SLA)

Acuerdo que aclara lo que un equipo de TI se compromete a ofrecer a los clientes, como el tiempo de actividad y el rendimiento del servicio.

indicador de nivel de servicio (SLI)

Medición de un aspecto del rendimiento de un servicio, como la tasa de errores, la disponibilidad o el rendimiento.

objetivo de nivel de servicio (SLO)

Métrica objetivo que representa el estado de un servicio medido mediante un indicador de nivel de servicio.

modelo de responsabilidad compartida

Un modelo que describe la responsabilidad con AWS la que compartes la seguridad y el cumplimiento de la nube. AWS es responsable de la seguridad de la nube, mientras que usted es responsable de la seguridad en la nube. Para obtener más información, consulte el Modelo de responsabilidad compartida.

Shadow AI

Aplicaciones de IA no autorizadas creadas o utilizadas fuera de los canales regulados dentro de una organización.

SIEM

Consulte sistema de administración de eventos e información de seguridad.

único punto de error (SPOF)

Error en un único componente crítico de una aplicación que puede interrumpir el sistema.

SLA

Consulte acuerdo de nivel de servicio.

SLI

Consulte indicador de nivel de servicio.

SLO

Consulte objetivo de nivel de servicio.

modelo de dividir y sembrar

Un patrón para escalar y acelerar los proyectos de modernización. A medida que se definen las nuevas funciones y los lanzamientos de los productos, el equipo principal se divide para crear nuevos equipos de productos. Esto ayuda a ampliar las capacidades y los servicios de su organización, mejora la productividad de los desarrolladores y apoya la innovación rápida. Para más información, consulte Phased approach to modernizing applications in the Nube de AWS.

SPOF

Consulte único punto de error.

esquema en estrella

Estructura organizativa de una base de datos que utiliza una tabla de hechos de gran tamaño para almacenar datos transaccionales o medidos y una o varias tablas dimensionales más pequeñas para almacenar los atributos de los datos. Esta estructura está diseñada para utilizarse en un almacén de datos o con fines de inteligencia empresarial.

patrón de higo estrangulador

Un enfoque para modernizar los sistemas monolíticos mediante la reescritura y el reemplazo gradual de las funciones del sistema hasta que se pueda desmantelar el sistema heredado. Este patrón utiliza la analogía de una higuera que crece hasta convertirse en un árbol estable y, finalmente, se apodera y reemplaza a su host. El patrón fue presentado por Martin Fowler como una forma de gestionar el riesgo al reescribir sistemas monolíticos. Para ver un ejemplo de cómo aplicar este patrón, consulte Modernización gradual de los servicios web antiguos de Microsoft ASP.NET (ASMX) mediante contenedores y Amazon API Gateway.

subred

Un intervalo de direcciones IP en la VPC. Una subred debe residir en una sola zona de disponibilidad.

control de supervisión y adquisición de datos (SCADA)

En el sector de la fabricación, sistema que utiliza hardware y software para supervisar los activos físicos y las operaciones de producción.

cifrado simétrico

Un algoritmo de cifrado que utiliza la misma clave para cifrar y descifrar los datos.

pruebas sintéticas

Prueba de un sistema de manera que simule las interacciones de los usuarios para detectar posibles problemas o supervisar el rendimiento. Puede usar Amazon CloudWatch Synthetics para crear estas pruebas.

petición del sistema

Técnica para proporcionar contexto, instrucciones o pautas a un LLM para dirigir su comportamiento. Las peticiones del sistema ayudan a establecer el contexto y las reglas para las interacciones con los usuarios.

T

etiquetas

Key-value pares que actúan como metadatos para organizar sus AWS recursos. Las etiquetas pueden ayudar a administrar, identificar, organizar, buscar y filtrar recursos de . Para obtener más información, consulte Etiquetado de los recursos de AWS.

variable de destino

El valor que intenta predecir en el ML supervisado. Esto también se conoce como variable de resultado. Por ejemplo, en un entorno de fabricación, la variable objetivo podría ser un defecto del producto.

lista de tareas

Herramienta que se utiliza para hacer un seguimiento del progreso mediante un manual de procedimientos. La lista de tareas contiene una descripción general del manual de procedimientos y una lista de las tareas generales que deben completarse. Para cada tarea general, se incluye la cantidad estimada de tiempo necesario, el propietario y el progreso.

entorno de prueba

Consulte entorno.

entrenamiento

Proporcionar datos de los que pueda aprender su modelo de ML. Los datos de entrenamiento deben contener la respuesta correcta. El algoritmo de aprendizaje encuentra patrones en los

datos de entrenamiento que asignan los atributos de los datos de entrada al destino (la respuesta que desea predecir). Genera un modelo de ML que captura estos patrones. Luego, el modelo de ML se puede utilizar para obtener predicciones sobre datos nuevos para los que no se conoce el destino.

herramienta

Una función o API que un agente puede invocar para realizar operaciones en sistemas externos.

puerta de enlace de tránsito

Centro de tránsito de red que puede utilizar para interconectar las VPC y las redes en las instalaciones. Para obtener más información, consulte Qué es una pasarela de tránsito en la AWS Transit Gateway documentación.

flujo de trabajo basado en enlaces troncales

Un enfoque en el que los desarrolladores crean y prueban características de forma local en una rama de característica y, a continuación, combinan esos cambios en la rama principal. Luego, la rama principal se adapta a los entornos de desarrollo, preproducción y producción, de forma secuencial.

acceso de confianza

Otorgar permisos a un servicio que especifique para realizar tareas en su organización AWS Organizations y en sus cuentas en su nombre. El servicio de confianza crea un rol vinculado al servicio en cada cuenta, cuando ese rol es necesario, para realizar las tareas de administración por usted. Para obtener más información, consulte AWS Organizations Utilización con otros AWS servicios en la AWS Organizations documentación.

ajuste

Cambiar aspectos de su proceso de formación a fin de mejorar la precisión del modelo de ML. Por ejemplo, puede entrenar el modelo de ML al generar un conjunto de etiquetas, incorporar etiquetas y, luego, repetir estos pasos varias veces con diferentes ajustes para optimizar el modelo.

equipo de dos pizzas

Un DevOps equipo pequeño al que puedes alimentar con dos pizzas. Un equipo formado por dos integrantes garantiza la mejor oportunidad posible de colaboración en el desarrollo de software.

U

incertidumbre

Un concepto que hace referencia a información imprecisa, incompleta o desconocida que puede socavar la fiabilidad de los modelos predictivos de ML. Hay dos tipos de incertidumbre: la incertidumbre epistémica se debe a datos limitados e incompletos, mientras que la incertidumbre aleatoria se debe al ruido y la aleatoriedad inherentes a los datos.

tareas indiferenciadas

También conocido como tareas arduas, es el trabajo que es necesario para crear y operar una aplicación, pero que no proporciona un valor directo al usuario final ni proporciona una ventaja competitiva. Algunos ejemplos de tareas indiferenciadas son la adquisición, el mantenimiento y la planificación de la capacidad.

entornos superiores

Consulte entorno.

V

succión

Una operación de mantenimiento de bases de datos que implica limpiar después de las actualizaciones incrementales para recuperar espacio de almacenamiento y mejorar el rendimiento.

control de versión

Procesos y herramientas que realizan un seguimiento de los cambios, como los cambios en el código fuente de un repositorio.

Emparejamiento de VPC

Conexión entre dos VPC que permite enrutar el tráfico mediante direcciones IP privadas. Para obtener más información, consulte ¿Qué es una interconexión de VPC? en la documentación de Amazon VPC.

vulnerabilidad

Defecto de software o hardware que pone en peligro la seguridad del sistema.

W

caché caliente

Un búfer caché que contiene datos actuales y relevantes a los que se accede con frecuencia. La instancia de base de datos puede leer desde la caché del búfer, lo que es más rápido que leer desde la memoria principal o el disco.

datos templados

Datos a los que el acceso es infrecuente. Al consultar este tipo de datos, normalmente se aceptan consultas moderadamente lentas.

función de ventana

Función SQL que hace un cálculo en un grupo de filas que se relacionan de alguna manera con el registro actual. Las funciones de ventana son útiles para las tareas de procesamiento, como calcular una media móvil o acceder al valor de las filas en función de la posición relativa de la fila actual.

carga de trabajo

Conjunto de recursos y código que ofrece valor comercial, como una aplicación orientada al cliente o un proceso de backend.

flujo de trabajo

Grupos funcionales de un proyecto de migración que son responsables de un conjunto específico de tareas. Cada flujo de trabajo es independiente, pero respalda a los demás flujos de trabajo del proyecto. Por ejemplo, el flujo de trabajo de la cartera es responsable de priorizar las aplicaciones, planificar las oleadas y recopilar los metadatos de migración. El flujo de trabajo de la cartera entrega estos recursos al flujo de trabajo de migración, que luego migra los servidores y las aplicaciones.

WORM

Consulte escritura única y lectura múltiple.

WQF

Consulte AWS Workload Qualification Framework.

escritura única y lectura múltiple (WORM)

Modelo de almacenamiento que escribe los datos una sola vez y evita que se eliminen o modifiquen. Los usuarios autorizados pueden leer los datos tantas veces como sea necesario, pero no los pueden cambiar. Esta infraestructura de almacenamiento de datos se considera inmutable.

Z

ataque de día cero

Ataque, normalmente de malware, que se aprovecha de una vulnerabilidad de día cero.

vulnerabilidad de día cero

Un defecto o una vulnerabilidad sin mitigación en un sistema de producción. Los agentes de amenazas pueden usar este tipo de vulnerabilidad para atacar el sistema. Los desarrolladores suelen darse cuenta de la vulnerabilidad a raíz del ataque.

peticiones desde cero

Proporcionar a un LLM instrucciones para llevar a cabo una tarea, pero sin ejemplos (pasos) que puedan ayudar a guiarlo. El LLM debe usar los conocimientos del entrenamiento previo para llevar a cabo la tarea. La eficacia de la petición desde cero depende de la complejidad de la tarea y de la calidad de la petición. Consulte también peticiones con pocos pasos.

aplicación zombi

Aplicación que utiliza un promedio de CPU y memoria menor al 5 por ciento. En un proyecto de migración, es habitual retirar estas aplicaciones.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la version original de inglés, prevalecerá la version en inglés.