



Creación de un sistema de rastreo web escalable para los datos de ESG en AWS

AWS Guía prescriptiva



AWS Guía prescriptiva: Creación de un sistema de rastreo web escalable para los datos de ESG en AWS

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas registradas que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, conectados o patrocinados por Amazon.

Table of Contents

Introducción	1
Destinatarios previstos	2
Resultados empresariales específicos	2
Arquitectura	4
Diseño y operaciones de los rastreadores web	5
Procesamiento de datos y procesamiento de datos	7
Construir el sistema	8
Preparar un conjunto de datos	8
Construir el rastreador web	10
Capturar y procesar el archivo robots.txt	10
Capturar y procesar el mapa del sitio	12
Diseñar el rastreador	14
Construir la infraestructura AWS	21
Prácticas recomendadas	23
Conformidad con Robots.txt	23
Limitación de la velocidad de rastreo	24
Transparencia entre los agentes de usuario	24
Rastreo eficiente	24
Enfoque adaptativo	24
Gestión de errores	24
Arrastrándose en lotes	24
Seguridad	25
Otras consideraciones	25
Preguntas frecuentes	26
¿Qué sucede si el archivo robots.txt no está disponible?	26
¿Qué sucede si el archivo sitemaps.xml no está disponible?	26
¿Puedo usar una solución sin servidor en lugar de Amazon EC2 o Amazon ECS?	27
¿Por qué el rastreador recibe un código de estado 403?	27
Próximos pasos y recursos	28
Recursos	28
Herramientas	29
Historial de documentos	30
Glosario	31
#	31

A	32
B	35
C	37
D	40
E	45
F	47
G	49
H	50
I	52
L	54
M	55
O	60
P	63
Q	66
R	66
S	69
T	73
U	75
V	75
W	76
Z	77
.....	lxxviii

Creación de un sistema de rastreo web escalable para datos de ESG en AWS

Vijit Vashishtha y Mansi Doshi, Amazon Web Services

[Enero de 2025 \(historial del documento\)](#)

Los factores ambientales, sociales y de gobierno (ESG) son consideraciones fundamentales para los inversores a la hora de evaluar las posibles inversiones:

- **Ambientales:** se centra en el impacto de una empresa en el mundo natural. Incluye factores como las emisiones de carbono, la gestión de los recursos y la eficiencia energética.
- **Social:** examina cómo una empresa gestiona las relaciones con los empleados, los proveedores, los clientes y las comunidades. Abarca aspectos como las prácticas laborales, la diversidad y la participación de la comunidad.
- **Gobernanza:** analiza el liderazgo, los controles internos y los derechos de los accionistas de una empresa. Incluye la composición del consejo de administración, la compensación de los ejecutivos y la ética empresarial.

Cada vez se considera que las empresas con prácticas ESG sólidas están mejor posicionadas para lograr la sostenibilidad y la rentabilidad a largo plazo. Los inversores demandan cada vez más información sobre ESG. Las empresas que pueden demostrar sus credenciales de sostenibilidad mediante datos ESG útiles y fiables están mejor posicionadas para atraer capital y seguir siendo competitivas. Las empresas publican los datos ESG a través de diversas fuentes, como noticias, artículos e informes anuales. Como esta información está dispersa, un rastreador web puede ayudarlo a recopilar estos datos de manera eficiente.

Esta guía completa muestra cómo utilizar [AWS Fargate](#), [Amazon Elastic Compute Cloud \(Amazon EC2\)](#) y [Amazon Simple Storage Service \(Amazon S3\)](#) para crear un canal de recopilación de datos sólido, escalable y responsable. [AWS Batch](#) En ella se analiza lo siguiente:

- Diseñar un sistema de rastreo escalable mediante lo siguiente: Servicios de AWS
 - Fargate o Amazon EC2 para ejecutar la aplicación de rastreo
 - AWS Batch para organizar de forma eficiente las tareas de rastreo a gran escala
 - Amazon S3 para un almacenamiento de datos seguro y duradero

- Implementar las mejores prácticas para un rastreo ético, que incluyen:
 - Respetar las políticas de robots.txt y del sitio web
 - Gestionar los límites de velocidad para evitar abrumar a los sitios objetivo
 - Garantizar la privacidad de los datos y el uso responsable de la información recopilada
- Desarrollando un Pythonrastreador basado en software optimizado para la infraestructura AWS
- Optimizar el rendimiento de los rastreadores sin dejar de mantener los estándares éticos

Destinatarios previstos

Esta guía está destinada a ingenieros de datos y arquitectos de nube que desean recopilar de manera eficiente grandes cantidades de datos up-to-date ESG de sitios web públicos. Es particularmente relevante para proyectos que implican análisis de mercado, evaluaciones financieras sostenibles o investigaciones financieras.

Resultados empresariales específicos

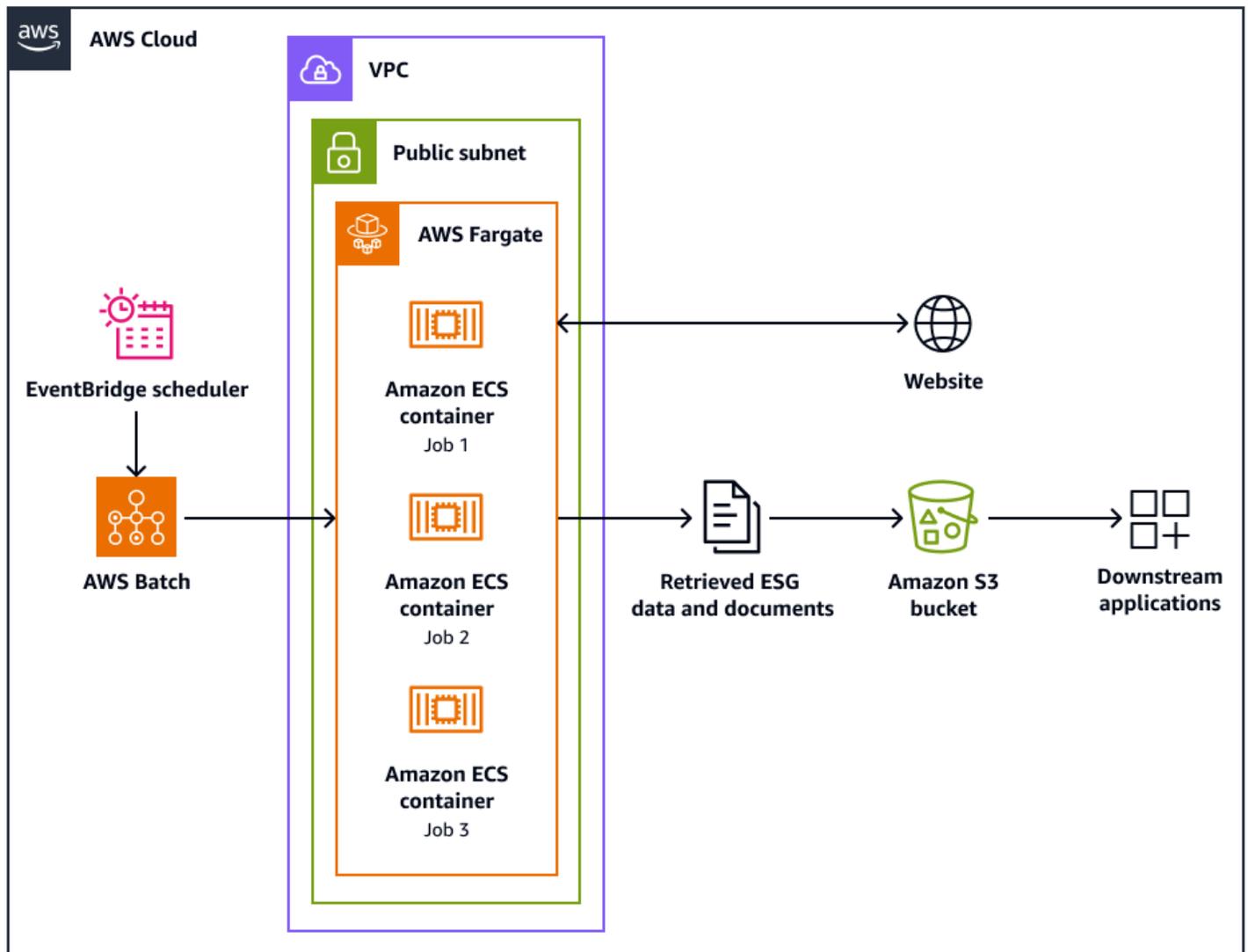
Las siguientes son las razones más comunes por las que las empresas utilizan los datos ESG:

- **Gestión de riesgos:** los datos ESG le ayudan a identificar y mitigar los posibles riesgos relacionados con cuestiones ambientales, sociales y de gobierno.
- **Atracción de inversores:** muchos inversores ahora tienen en cuenta los factores ESG a la hora de tomar decisiones de inversión. Consideran que las prácticas ESG sólidas son indicadores de sostenibilidad y rentabilidad a largo plazo.
- **Gestión de la reputación:** un buen desempeño en materia de ESG puede mejorar la reputación de una empresa entre los clientes, los empleados y el público en general.
- **Cumplimiento regulatorio:** a medida que aumentan las regulaciones relacionadas con los ESG, la adopción de prácticas de ESG ayuda a las empresas a mantenerse a la vanguardia de los requisitos de cumplimiento.
- **Innovación y eficiencia:** centrarse en los factores ESG puede impulsar la innovación en los productos, los servicios y las operaciones. Esto conduce a una mayor eficiencia y a un ahorro de costes.
- **Ventaja competitiva:** un sólido desempeño en materia de ESG puede diferenciar a una empresa de sus competidores y abrir nuevas oportunidades de mercado.

- Participación de las partes interesadas: las prácticas ESG ayudan a las empresas a interactuar mejor con las distintas partes interesadas y a cumplir sus expectativas, incluidos los empleados, los clientes y las comunidades locales.

Arquitectura para un sistema de rastreo web escalable en AWS

El siguiente diagrama de arquitectura muestra un sistema de rastreo web diseñado para extraer de manera ética datos ambientales, sociales y de gobierno (ESG) de los sitios web. Utilizas un Pythonrastreador basado en software optimizado para la AWS infraestructura. Solía organizar AWS Batch las tareas de rastreo a gran escala y utilizar Amazon Simple Storage Service (Amazon S3) para el almacenamiento. Las aplicaciones posteriores pueden ingerir y almacenar los datos del bucket de Amazon S3.



En el diagrama, se muestra el siguiente flujo de trabajo:

1. Amazon EventBridge Scheduler inicia el proceso de rastreo en un intervalo que usted programe.
2. AWS Batch gestiona la ejecución de las tareas del rastreador web. La cola de AWS Batch trabajos contiene y organiza los trabajos de rastreo pendientes.
3. Los trabajos de rastreo web se ejecutan en contenedores de Amazon Elastic Container Service (Amazon ECS) en AWS Fargate. Los trabajos se ejecutan en una subred pública de una nube privada virtual (VPC).
4. El rastreador web rastrea el sitio web de destino y recupera datos y documentos de ESG, como PDF, CSV u otros archivos de documentos.
5. El rastreador web almacena los datos recuperados y los archivos sin procesar en un bucket de Amazon S3.
6. Otros sistemas o aplicaciones ingieren o procesan los datos y archivos almacenados en el bucket de Amazon S3.

Diseño y operaciones de los rastreadores web

Algunos sitios web están diseñados específicamente para ejecutarse en ordenadores de sobremesa o dispositivos móviles. El rastreador web está diseñado para admitir el uso de un agente de usuario de escritorio o un agente de usuario móvil. Estos agentes le ayudan a realizar correctamente las solicitudes al sitio web de destino.

Una vez inicializado el rastreador web, realiza las siguientes operaciones:

1. El rastreador web llama al método `setup()`. Este método busca y analiza el archivo `robots.txt`.

Note

También puedes configurar el rastreador web para que busque y analice el mapa del sitio.

2. El rastreador web procesa el archivo `robots.txt`. Si se especifica un retraso de rastreo en el archivo `robots.txt`, el rastreador web extrae el retraso de rastreo para el agente de usuario de escritorio. Si no se especifica ningún retraso de rastreo en el archivo `robots.txt`, el rastreador web utiliza un retraso aleatorio.
3. El rastreador web llama al método `crawl()`, lo que inicia el proceso de rastreo. Si no URLs hay ninguno en la cola, añade la URL de inicio.

Note

El rastreador continúa rastreando hasta que alcance el número máximo de páginas o se quede sin páginas URLs para rastrear.

4. El rastreador procesa el. URLs Para cada URL de la cola, el rastreador comprueba si la URL ya se ha rastreado.
5. Si no se ha rastreado una URL, el rastreador llama al método de la siguiente manera: `crawl_url()`
 - a. El rastreador comprueba el archivo robots.txt para determinar si puede utilizar el agente de usuario de escritorio para rastrear la URL.
 - b. Si está permitido, el rastreador intentará rastrear la URL mediante el agente de usuario de escritorio.
 - c. Si no está permitido o si el agente de usuario de escritorio no puede rastrear, el rastreador comprueba el archivo robots.txt para determinar si puede utilizar el agente de usuario móvil para rastrear la URL.
 - d. Si está permitido, el rastreador intentará rastrear la URL mediante el agente de usuario móvil.
6. El rastreador llama al `attempt_crawl()` método, que recupera y procesa el contenido. El rastreador envía una solicitud GET a la URL con los encabezados adecuados. Si se produce un error en la solicitud, el rastreador utiliza la lógica de reintento.
7. Si el archivo está en formato HTML, el rastreador llama al método. `extract_esg_data()` Utiliza [Beautiful Soup](#) para analizar el contenido HTML. Extrae datos ambientales, sociales y de gobierno (ESG) mediante la coincidencia de palabras clave.

Si el archivo es un PDF, el rastreador llama al `save_pdf()` método. El rastreador descarga y guarda el archivo PDF en el bucket de Amazon S3.
8. El rastreador llama al método. `extract_news_links()` Busca y almacena enlaces a artículos de noticias, comunicados de prensa y publicaciones de blogs.
9. El rastreador llama al `extract_pdf_links()` método. Esto identifica y almacena los enlaces a los documentos PDF.
- 10 El rastreador llama al `is_relevant_to_sustainable_finance()` método. Esto comprueba si las noticias o los artículos están relacionados con las finanzas sostenibles mediante el uso de palabras clave predefinidas.

- 11 Después de cada intento de rastreo, el rastreador implementa un retraso mediante el `delay()` método. Si se especificó un retraso en el archivo `robots.txt`, utilizará ese valor. De lo contrario, utiliza un retraso aleatorio de entre 1 y 3 segundos.
- 12 El rastreador llama al `save_esg_data()` método para guardar los datos de ESG en un archivo CSV. El archivo CSV se guarda en el bucket de Amazon S3.
- 13 El rastreador utiliza el `save_news_links()` método para guardar los enlaces de noticias a un archivo CSV, incluida la información relevante. El archivo CSV se guarda en el bucket de Amazon S3.
- 14 El rastreador utiliza el `save_pdf_links()` método para guardar los enlaces del PDF a un archivo CSV. El archivo CSV se guarda en el bucket de Amazon S3.

Procesamiento de datos y procesamiento de datos

El proceso de rastreo se organiza y se realiza de forma estructurada. AWS Batch asigna los trabajos a cada empresa para que se ejecuten en paralelo, en lotes. Cada lote se centra en el dominio y los subdominios de una sola empresa, tal como los ha identificado en su conjunto de datos. Sin embargo, los trabajos del mismo lote se ejecutan de forma secuencial para no inundar el sitio web con demasiadas solicitudes. Esto ayuda a la aplicación a gestionar la carga de trabajo de rastreo de forma más eficiente y a garantizar que se recopilen todos los datos relevantes de cada empresa.

Al organizar el rastreo web en lotes específicos de la empresa, se agrupan los datos recopilados. Esto ayuda a evitar que los datos de una empresa se mezclen con los datos de otras empresas.

El procesamiento por lotes ayuda a la aplicación a recopilar datos de la web de manera eficiente, al tiempo que mantiene una estructura clara y una separación de la información en función de las empresas objetivo y sus respectivos dominios web. Este enfoque ayuda a garantizar la integridad y la usabilidad de los datos recopilados, ya que están perfectamente organizados y asociados a la empresa y los dominios adecuados.

Creación de un sistema de rastreo web escalable en AWS

En esta sección se describe cómo crear el rastreador web que se describe en la [Arquitectura](#) sección. Incluye un enfoque sistemático para crear un conjunto de datos sólido de empresas y sus propiedades web asociadas. Este conjunto de datos sirve de base para sus actividades de rastreo. A continuación, en esta sección se describe cómo crear un rastreador web ético en Python.

Temas

- [Preparar un conjunto de datos](#)
- [Construir el rastreador web](#)
- [Construir la infraestructura AWS](#)

Preparar un conjunto de datos

Si aún no lo ha hecho, prepare un conjunto de datos detallado de los sitios web de los que desea recopilar información. Este conjunto de datos debe incluir la URL del sitio web, los nombres de dominio y los nombres de subdominios relevantes. En esta sección se proporciona un step-by-step proceso para crear este conjunto de datos.

Para preparar un conjunto de datos

1. Defina el alcance: determine la industria o los sectores en los que se está centrando. Decida cuántas empresas quiere incluir. Y defina los criterios que desee recopilar sobre estas empresas, como el número de empleados, la ubicación o los ingresos.
2. Identifique las fuentes de datos: identifique qué fuentes de información puede utilizar para recopilar información sobre estas empresas. Algunos ejemplos son los directorios de empresas (como [Crunchbase](#), [Bloomberg](#) o [Forbes](#)), las bolsas de valores (como la Bolsa de Valores de Nueva York y el NASDAQ), las asociaciones o publicaciones específicas de un sector o las bases de datos gubernamentales (como las presentadas ante la SEC).
3. Cree una tabla: en su herramienta preferida, como Microsoft Excel, Google Sheets o un sistema de administración de bases de datos, cree una tabla para recopilar los criterios de cada empresa. Incluye una columna para cada criterio. Como mínimo, incluye columnas para el nombre de la empresa, el dominio principal, los subdominios, el sector, el tamaño y la ubicación.
4. Recopile la información inicial de la empresa: recopile la siguiente información sobre cada empresa e introdúzcala en la tabla que creó:

- Nombre de la empresa
 - Industria o sector
 - Tamaño de la empresa (número de empleados)
 - Revenue (Ingresos)
 - Ubicación de la sede de la empresa
5. Recopile información del dominio: para cada empresa, extraiga el nombre de dominio principal de la URL del sitio web principal, por ejemplo `example.com`. Puedes verificar la información del dominio mediante una herramienta de búsqueda de dominios de WHOIS.
 6. Recopile información sobre los subdominios: para cada empresa, investigue los subdominios registrados, por ejemplo `blog.example.com` [Puedes usar herramientas de enumeración de subdominios, como Sublist3r, OWASP Amass o Subfinder](#). Puedes realizar búsquedas en Google (mediante una búsqueda `site:example.com`), comprobar los registros de DNS mediante un `dig` comando o una herramienta de búsqueda de DNS, o puedes analizar los certificados SSL o TLS.
 7. Valida y limpia los datos: revisa, verifica y estandariza los datos que has recopilado. Por ejemplo, elimina las entradas duplicadas, elimina la información de URL innecesaria de los dominios y subdominios y comprueba que todos los dominios y subdominios estén activos.
 8. (Opcional) Clasifique los subdominios: puede clasificar los subdominios en tipos. A continuación, se muestran algunos ejemplos de categorías que puede encontrar:
 - Blogs, como `blog.example.com`
 - Support o ayuda, como `support.example.com` o `help.example.com`
 - Comercio electrónico, como `shop.example.com` o `store.example.com`
 - Recursos para desarrolladores, como `dev.example.com` o `api.example.com`
 - Regiones o ubicaciones, como `us.example.com` o `uk.example.com`
 9. (Opcional) Agregue metadatos relevantes: puede registrar cualquier metadato relevante en el conjunto de datos. Por ejemplo, puedes añadir la fecha de la última actualización, la fuente de información o tu puntuación de confianza para determinar la precisión del subdominio.
 10. Implemente el control de versiones: utilice un sistema de control de versiones, como Git, para realizar un seguimiento de los cambios en la tabla a lo largo del tiempo. Realice copias de seguridad del conjunto de datos con regularidad.

11. Mantenga la tabla: establezca un cronograma, por ejemplo trimestral, para actualizar la tabla. Estandarice e implemente un proceso para agregar nuevas empresas o eliminar las que ya no necesite. Cuando sea posible, automatice la detección de subdominios.

Construir el rastreador web

Como se describe en la [Arquitectura](#) sección, la aplicación se ejecuta en lotes, uno para cada empresa.

Temas

- [Capturar y procesar el archivo robots.txt](#)
- [Capturar y procesar el mapa del sitio](#)
- [Diseñar el rastreador](#)

Capturar y procesar el archivo robots.txt

Después de preparar el conjunto de datos, debe confirmar si el dominio tiene un archivo [robots.txt](#). En el caso de los rastreadores web y otros bots, el archivo robots.txt indica qué secciones del sitio web pueden visitar. Seguir las instrucciones de este archivo es una buena práctica importante para rastrear sitios web de forma ética. Para obtener más información, consulte [las prácticas recomendadas para rastreadores web éticos](#) en esta guía.

Para capturar y procesar el archivo robots.txt

1. Si aún no lo ha hecho, instale la `requests` biblioteca ejecutando el siguiente comando en una terminal:

```
pip install requests
```

2. Ejecute el siguiente script. Este script hace lo siguiente:
 - Define una `check_robots_txt` función que toma un dominio como entrada.
 - Construye la URL completa del archivo robots.txt.
 - Envía una solicitud GET a la URL del archivo robots.txt.
 - Si la solicitud se realiza correctamente (código de estado 200), existe un archivo robots.txt.

- Si la solicitud falla o devuelve un código de estado diferente, significa que el archivo robots.txt no existe o no está accesible.

```
import requests
from urllib.parse import urljoin
def check_robots_txt(domain):
    # Ensure the domain starts with a protocol
    if not domain.startswith(('http://', 'https://')):
        domain = 'https://' + domain
    # Construct the full URL for robots.txt
    robots_url = urljoin(domain, '/robots.txt')
    try:
        # Send a GET request to the robots.txt URL
        response = requests.get(robots_url, timeout=5)
        # Check if the request was successful (status code 200)
        if response.status_code == 200:
            print(f"robots.txt found at {robots_url}")
            return True
        else:
            print(f"No robots.txt found at {robots_url} (Status code:
{response.status_code})")
            return False
    except requests.RequestException as e:
        print(f"Error checking {robots_url}: {e}")
        return False
```

Note

Este script gestiona las excepciones por errores de red u otros problemas.

3. Si existe un archivo robots.txt, utilice el siguiente script para descargarlo:

```
import requests

def download(self, url):
    response = requests.get(url, headers=self.headers, timeout=5)
    response.raise_for_status() # Raise an exception for non-2xx responses
    return response.text

def download_robots_txt(self):
```

```
# Append '/robots.txt' to the URL to get the robots.txt file's URL
robots_url = self.url.rstrip('/') + '/robots.txt'
try:
    response = download(robots_url)
    return response
except requests.exceptions.RequestException as e:
    print(f"Error downloading robots.txt: {e}, \nGenerating sitemap using combinations...")
    return e
```

Note

Estos scripts se pueden personalizar o modificar según su caso de uso. También puede combinar estos scripts.

Capturar y procesar el mapa del sitio

A continuación, debe procesar el [mapa del sitio](#). Puedes usar el mapa del sitio para centrar el rastreo en las páginas importantes. Esto mejora la eficiencia del rastreo. Para obtener más información, consulte [las prácticas recomendadas para rastreadores web éticos](#) en esta guía.

Para capturar y procesar el mapa del sitio

- Ejecute el siguiente script. Este script define una `check_and_download_sitemap` función que:
 - Acepta una URL base, una URL de mapa del sitio opcional de robots.txt y una cadena de agente de usuario.
 - Comprueba varias ubicaciones posibles del mapa del sitio, incluida la de robots.txt (si se proporciona).
 - Intenta descargar el mapa del sitio desde cada ubicación.
 - Comprueba que el contenido descargado esté en formato XML.
 - Llama a la `parse_sitemap` función para extraer el URLs. Esta función:
 - Analiza el contenido XML del mapa del sitio.
 - Gestiona tanto los mapas de sitio normales como los archivos de índice de los mismos.
 - Busca de forma recursiva los submapas del sitio si encuentra un índice de mapa del sitio.

```
import requests
from urllib.parse import urljoin
import xml.etree.ElementTree as ET

def check_and_download_sitemap(base_url, robots_sitemap_url=None,
    user_agent='SitemapBot/1.0'):
    headers = {'User-Agent': user_agent}
    sitemap_locations = [robots_sitemap_url, urljoin(base_url, '/sitemap.xml'),
        urljoin(base_url, '/sitemap_index.xml'),
        urljoin(base_url, '/sitemap/'), urljoin(base_url, '/sitemap/sitemap.xml')]

    for sitemap_url in sitemap_locations:
        if not sitemap_url:
            continue
        print(f"Checking for sitemap at: {sitemap_url}")
        try:
            response = requests.get(sitemap_url, headers=headers, timeout=10)
            if response.status_code == 200:
                content_type = response.headers.get('Content-Type', '')
                if 'xml' in content_type:
                    print(f"Successfully downloaded sitemap from {sitemap_url}")
                    return parse_sitemap(response.text)
                else:
                    print(f"Found content at {sitemap_url}, but it's not XML.
Content-Type: {content_type}")
            except requests.RequestException as e:
                print(f"Error downloading sitemap from {sitemap_url}: {e}")
        print("No sitemap found.")
        return []

def parse_sitemap(sitemap_content):
    urls = []
    try:
        root = ET.fromstring(sitemap_content)
        # Handle both sitemap and sitemapindex
        for loc in root.findall('.//{http://www.sitemaps.org/schemas/
sitemap/0.9}loc'):
            urls.append(loc.text)

        # If it's a sitemap index, recursively fetch each sitemap
        if root.tag.endswith('sitemapindex'):
```

```
        all_urls = []
        for url in urls:
            print(f"Fetching sub-sitemap: {url}")
            sub_sitemap_urls = check_and_download_sitemap(url)
            all_urls.extend(sub_sitemap_urls)
        return all_urls
    except ET.ParseError as e:
        print(f"Error parsing sitemap XML: {e}")
    return urls

if __name__ == "__main__":
    base_url = input("Enter the base URL of the website: ")
    robots_sitemap_url = input("Enter the sitemap URL from robots.txt (or press
Enter if none): ").strip() or None
    urls = check_and_download_sitemap(base_url, robots_sitemap_url)
    print(f"Found {len(urls)} URLs in sitemap:")
    for url in urls[:5]: # Print first 5 URLs as an example
        print(url)
    if len(urls) > 5:
        print("...")
```

Diseñar el rastreador

A continuación, diseñe el rastreador web. El rastreador está diseñado para seguir las prácticas recomendadas que se describen [Mejores prácticas para rastreadores web éticos](#) en esta guía. En esta `EthicalCrawler` clase se muestran varios principios clave del rastreo ético:

- Obtener y analizar el archivo `robots.txt`: el rastreador busca el archivo `robots.txt` del sitio web de destino.
- Respetar los permisos de rastreo: antes de rastrear cualquier URL, el rastreador comprueba si las reglas del archivo `robots.txt` permiten rastrear esa URL. Si una URL no está permitida, el rastreador la omite y pasa a la siguiente URL.
- Respetar el retraso de rastreo: el rastreador comprueba si hay una directiva de retraso de rastreo en el archivo `robots.txt`. Si se especifica alguna, el rastreador utiliza este retraso entre las solicitudes. De lo contrario, utiliza un retraso predeterminado.
- Identificación del agente de usuario: el rastreador utiliza una cadena de agente de usuario personalizada para identificarse en los sitios web. Si es necesario, los propietarios de los sitios web pueden establecer reglas específicas para restringir o permitir el rastreo.

- Gestión de errores y degradación correcta: si el archivo robots.txt no se puede recuperar ni analizar, el rastreador sigue las reglas predeterminadas conservadoras. Gestiona los errores de red y las respuestas HTTP distintas de las 200.
- Rastreo limitado: para evitar sobrecargar el servidor, existe un límite en cuanto al número de páginas que se pueden rastrear.

El siguiente script es un pseudocódigo que explica cómo funciona el rastreador web:

```
import requests
from urllib.parse import urljoin, urlparse
import time

class EthicalCrawler:
    def __init__(self, start_url, user_agent='EthicalBot/1.0'):
        self.start_url = start_url
        self.user_agent = user_agent
        self.domain = urlparse(start_url).netloc
        self.robots_parser = None
        self.crawl_delay = 1 # Default delay in seconds

    def can_fetch(self, url):
        if self.robots_parser:
            return self.robots_parser.allowed(url, self.user_agent)
        return True # If no robots.txt, assume allowed but crawl conservatively

    def get_crawl_delay(self):
        if self.robots_parser:
            delay = self.robots_parser.agent(self.user_agent).delay
            if delay is not None:
                self.crawl_delay = delay
        print(f"Using crawl delay of {self.crawl_delay} seconds")

    def crawl(self, max_pages=10):
        self.get_crawl_delay()
        pages_crawled = 0
        urls_to_crawl = [self.start_url]
        while urls_to_crawl and pages_crawled < max_pages:
            url = urls_to_crawl.pop(0)
            if not self.can_fetch(url):
                print(f"robots.txt disallows crawling: {url}")
                continue
```

```
try:
    response = requests.get(url, headers={'User-Agent': self.user_agent})
    if response.status_code == 200:
        print(f"Successfully crawled: {url}")
        # Here you would typically parse the content, extract links, etc.
        # For this example, we'll just increment the counter
        pages_crawled += 1
    else:
        print(f"Failed to crawl {url}: HTTP {response.status_code}")
except Exception as e:
    print(f"Error crawling {url}: {e}")

# Respect the crawl delay
time.sleep(self.crawl_delay)

print(f"Crawling complete. Crawled {pages_crawled} pages.")
```

Para crear un rastreador web avanzado y ético que recopile datos de ESG

1. Copie el siguiente ejemplo de código para el rastreador web ético avanzado que se utiliza en este sistema:

```
import requests
from urllib.parse import urljoin, urlparse
import time
from collections import deque
import random
from bs4 import BeautifulSoup
import re
import csv
import os

class EnhancedESGCrawler:
    def __init__(self, start_url):
        self.start_url = start_url
        self.domain = urlparse(start_url).netloc
        self.desktop_user_agent = 'ESGEthicalBot/1.0'
        self.mobile_user_agent = 'Mozilla/5.0 (iPhone; CPU iPhone OS 14_0 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/14.0 Mobile/15E148 Safari/604.1'
        self.robots_parser = None
```

```
self.crawl_delay = None
self.urls_to_crawl = deque()
self.crawled_urls = set()
self.max_retries = 2
self.session = requests.Session()
self.esg_data = []
self.news_links = []
self.pdf_links = []

def setup(self):
    self.fetch_robots_txt() # Provided in Previous Snippet
    self.fetch_sitemap() # Provided in Previous Snippet

def can_fetch(self, url, user_agent):
    if self.robots_parser:
        return self.robots_parser.allowed(url, user_agent)
    return True

def delay(self):
    if self.crawl_delay is not None:
        time.sleep(self.crawl_delay)
    else:
        time.sleep(random.uniform(1, 3))

def get_headers(self, user_agent):
    return {'User-Agent': user_agent,
           'Accept': 'text/html,application/xhtml+xml,application/
xml;q=0.9,image/webp,*/*;q=0.8',
           'Accept-Language': 'en-US,en;q=0.5', 'Accept-Encoding': 'gzip,
deflate, br', 'DNT': '1',
           'Connection': 'keep-alive', 'Upgrade-Insecure-Requests': '1'}

def extract_esg_data(self, url, html_content):
    soup = BeautifulSoup(html_content, 'html.parser')
    esg_data = {
        'url': url,
        'environmental': self.extract_environmental_data(soup),
        'social': self.extract_social_data(soup),
        'governance': self.extract_governance_data(soup)
    }
    self.esg_data.append(esg_data)
    # Extract news links and PDFs
    self.extract_news_links(soup, url)
```

```
self.extract_pdf_links(soup, url)

def extract_environmental_data(self, soup):
    keywords = ['carbon footprint', 'emissions', 'renewable energy', 'waste
management', 'climate change']
    return self.extract_keyword_data(soup, keywords)

def extract_social_data(self, soup):
    keywords = ['diversity', 'inclusion', 'human rights', 'labor practices',
'community engagement']
    return self.extract_keyword_data(soup, keywords)

def extract_governance_data(self, soup):
    keywords = ['board structure', 'executive compensation', 'shareholder
rights', 'ethics', 'transparency']
    return self.extract_keyword_data(soup, keywords)

def extract_keyword_data(self, soup, keywords):
    text = soup.get_text().lower()
    return {keyword: len(re.findall(r'\b' + re.escape(keyword) + r'\b', text))
for keyword in keywords}

def extract_news_links(self, soup, base_url):
    news_keywords = ['news', 'press release', 'article', 'blog',
'sustainability']
    for a in soup.find_all('a', href=True):
        if any(keyword in a.text.lower() for keyword in news_keywords):
            full_url = urljoin(base_url, a['href'])
            if full_url not in self.news_links:
                self.news_links.append({'url': full_url, 'text':
a.text.strip()})

def extract_pdf_links(self, soup, base_url):
    for a in soup.find_all('a', href=True):
        if a['href'].lower().endswith('.pdf'):
            full_url = urljoin(base_url, a['href'])
            if full_url not in self.pdf_links:
                self.pdf_links.append({'url': full_url, 'text':
a.text.strip()})

def is_relevant_to_sustainable_finance(self, text):
    keywords = ['sustainable finance', 'esg', 'green bond', 'social impact',
'environmental impact',
```

```

        'climate risk', 'sustainability report', 'corporate
responsibility']
    return any(keyword in text.lower() for keyword in keywords)

def attempt_crawl(self, url, user_agent):
    for _ in range(self.max_retries):
        try:
            response = self.session.get(url,
headers=self.get_headers(user_agent), timeout=10)
            if response.status_code == 200:
                print(f"Successfully crawled: {url}")
                if response.headers.get('Content-Type', '').startswith('text/
html'):
                    self.extract_esg_data(url, response.text)
                elif response.headers.get('Content-Type',
'').startswith('application/pdf'):
                    self.save_pdf(url, response.content)
                return True
            else:
                print(f"Failed to crawl {url}: HTTP {response.status_code}")
        except requests.RequestException as e:
            print(f"Error crawling {url} with {user_agent}: {e}")

        self.delay()
    return False

def crawl_url(self, url):
    if not self.can_fetch(url, self.desktop_user_agent):
        print(f"Robots.txt disallows desktop user agent: {url}")
    if self.can_fetch(url, self.mobile_user_agent):
        print(f"Attempting with mobile user agent: {url}")
        return self.attempt_crawl(url, self.mobile_user_agent)
    else:
        print(f"Robots.txt disallows both user agents: {url}")
        return False

    return self.attempt_crawl(url, self.desktop_user_agent)

def crawl(self, max_pages=100):
    self.setup()

    if not self.urls_to_crawl:
        self.urls_to_crawl.append(self.start_url)

```

```
pages_crawled = 0
while self.urls_to_crawl and pages_crawled < max_pages:
    url = self.urls_to_crawl.popleft()
    if url not in self.crawled_urls:
        if self.crawl_url(url):
            pages_crawled += 1
            self.crawled_urls.add(url)
            self.delay()

print(f"Crawling complete. Successfully crawled {pages_crawled} pages.")
self.save_esg_data()
self.save_news_links()
self.save_pdf_links()

def save_esg_data(self):
    with open('esg_data.csv', 'w', newline='', encoding='utf-8') as file:
        writer = csv.DictWriter(file, fieldnames=['url', 'environmental',
'social', 'governance'])
        writer.writeheader()
        for data in self.esg_data:
            writer.writerow({
                'url': data['url'],
                'environmental': ', '.join([f"{k}: {v}" for k, v in
data['environmental'].items()]),
                'social': ', '.join([f"{k}: {v}" for k, v in
data['social'].items()]),
                'governance': ', '.join([f"{k}: {v}" for k, v in
data['governance'].items()])
            })
    print("ESG data saved to esg_data.csv")

def save_news_links(self):
    with open('news_links.csv', 'w', newline='', encoding='utf-8') as file:
        writer = csv.DictWriter(file, fieldnames=['url', 'text', 'relevant'])
        writer.writeheader()
        for news in self.news_links:
            writer.writerow({
                'url': news['url'],
                'text': news['text'],
                'relevant':
self.is_relevant_to_sustainable_finance(news['text'])
            })
```

```
print("News links saved to news_links.csv")

def save_pdf_links(self):
    # Code for saving PDF in S3 or filesystem

def save_pdf(self, url, content):
    # Code for saving PDF in S3 or filesystem

# Example usage
if __name__ == "__main__":
    start_url = input("Enter the starting URL to crawl for ESG data and news: ")
    crawler = EnhancedESGCrawler(start_url)
    crawler.crawl(max_pages=50)
```

2. Configure los distintos atributos, incluidos los agentes de usuario, las colecciones vacías y las URLs listas de almacenamiento de datos.
3. Ajusta las palabras clave y los criterios de relevancia del `is_relevant_to_sustainable_finance()` método para que se ajusten a tus necesidades específicas.
4. Asegúrese de que el archivo `robots.txt` permite rastrear el sitio web y de que utiliza el retardo de rastreo y el agente de usuario especificados en el archivo `robots.txt`.
5. Considere la posibilidad de realizar las siguientes personalizaciones en el script de rastreo web proporcionado, según sea necesario para su organización:
 - Implemente el `fetch_sitemap()` método para descubrir las URL de forma más eficiente.
 - Mejore el registro y la gestión de errores para su uso en producción.
 - Implemente un análisis de relevancia del contenido más sofisticado.
 - Añada controles de profundidad y amplitud para limitar el alcance del rastreo.

Construir la infraestructura AWS

Hay muchos Servicios de AWS que puede utilizar para crear la infraestructura de rastreo web. La sección de [arquitectura](#) de esta guía incluye una solución propuesta. Le recomendamos que considere utilizar lo siguiente Servicios de AWS para crear la infraestructura de soporte para su rastreador web:

- [Utilice Amazon Virtual Private Cloud \(Amazon VPC\) para crear la VPC y las subredes.](#)

- Inicie el proceso de rastreo mediante [Amazon EventBridge Scheduler](#).
- [Administre los trabajos del rastreador web mediante trabajos y colas AWS Batch de trabajos](#).
- Utilice una de las siguientes soluciones para ejecutar los trabajos del rastreador web:
 - Contenedores de Amazon Elastic Container Service (Amazon ECS) en [AWS Fargate](#)
 - [Instancias](#) de Amazon Elastic Compute Cloud (Amazon EC2)

 Note

Si su aplicación puede gestionar las interrupciones, considere la posibilidad de utilizar Amazon EC2 [Spot Instances](#) a través de [Spot Fleet](#). Las flotas de instancias puntuales pueden ayudarle a ahorrar considerablemente en costos de cómputo.

- AWS Lambda [funciones](#)
- Guarde los datos recuperados y los archivos sin procesar en un bucket de Amazon Simple Storage Service (Amazon [S3](#)).

Mejores prácticas para rastreadores web éticos

En esta sección, se analizan las mejores prácticas y las consideraciones éticas clave para crear una aplicación de rastreo web que recopile datos ambientales, sociales y de gobierno (ESG). Si sigue estas mejores prácticas, puede proteger su proyecto y su organización y contribuir a un ecosistema web más responsable y sostenible. Este enfoque le ayuda a acceder a datos valiosos y a utilizarlos para la investigación, los negocios y la innovación de una manera que respeta a todas las partes interesadas.

Conformidad con Robots.txt

El archivo robots.txt se utiliza en los sitios web para comunicarse con los rastreadores web y los bots sobre qué partes del sitio web deben o no deben accederse o rastrearse. Cuando un rastreador web encuentra un archivo robots.txt en un sitio web, analiza las instrucciones y ajusta su comportamiento de rastreo en consecuencia. Esto evita que el rastreador infrinja las instrucciones del propietario del sitio web y mantiene una relación de cooperación entre el sitio web y el rastreador. Por lo tanto, el archivo robots.txt ayuda al control de acceso, a la protección del contenido confidencial, a la gestión de la carga y al cumplimiento de la legislación.

Recomendamos que siga las siguientes prácticas recomendadas:

- Compruebe y respete siempre las reglas del archivo robots.txt.
- Antes de rastrear cualquier URL, compruebe las reglas de los agentes de usuario de escritorio y móviles.
- Si el sitio web solo permite los agentes de usuario móviles, utilice un encabezado de agente diferente, como un encabezado de agente móvil, para su solicitud.

La ausencia de un archivo robots.txt no significa necesariamente que no puedas o no debas rastrear un sitio web. El rastreo siempre debe realizarse de manera responsable, respetando los recursos del sitio web y los derechos implícitos del propietario. Las siguientes son las prácticas recomendadas cuando el archivo robots.txt no está presente:

- Suponga que está permitido rastrear, pero proceda con precaución.
- Implemente prácticas de rastreo educadas.
- Considere la posibilidad de pedir permiso al propietario del sitio web si tienes pensado realizar un rastreo exhaustivo.

Limitación de la velocidad de rastreo

Utilice una velocidad de rastreo razonable para evitar sobrecargar el servidor. Implemente los retrasos entre las solicitudes, ya sea según lo especificado en el archivo robots.txt o mediante un retraso aleatorio. En el caso de sitios web pequeños o medianos, podría ser adecuado realizar una solicitud cada 10 o 15 segundos. En el caso de sitios web más grandes o con permisos de rastreo explícitos, podría ser adecuado realizar entre 1 y 2 solicitudes por segundo.

Transparencia entre los agentes de usuario

Identifique su rastreador en el encabezado del agente de usuario. La información del encabezado HTTP está destinada a identificar el dispositivo que solicita el contenido. Por lo general, la palabra bot se incluye en el nombre del agente. Los rastreadores y otros bots a veces utilizan un campo importante en el encabezado para incluir la información de contacto.

Rastreo eficiente

Utiliza el mapa del sitio, desarrollado por el propietario del sitio web, para centrarte en las páginas importantes.

Enfoque adaptativo

Programe el rastreador para que cambie a un agente de usuario móvil si la versión de escritorio no funciona correctamente. Esto puede proporcionar acceso al rastreador y reducir la carga sobre el servidor del sitio web.

Gestión de errores

Asegúrese de que el rastreador gestione varios códigos de estado HTTP de forma adecuada. Por ejemplo, el rastreador debería hacer una pausa si encuentra un código de estado 429 («Demasiadas solicitudes»). Si el rastreador recibe 403 códigos de estado de forma continua («Prohibido»), considere la posibilidad de dejar de rastrear.

Arrastrándose en lotes

Le recomendamos que realice las siguientes acciones:

- En lugar de arrastrarlas todas URLs a la vez, divida la tarea en lotes más pequeños. Esto puede ayudar a distribuir la carga y reducir el riesgo de que surjan problemas, como tiempos de espera o limitaciones de recursos.
- Si se espera que la tarea de rastreo en general dure mucho tiempo, considere la posibilidad de dividirla en varias tareas más pequeñas y fáciles de gestionar. Esto puede hacer que el proceso sea más escalable y resiliente.
- Si el número de objetos URLs por rastrear es relativamente pequeño, considere la posibilidad de utilizar una solución sin servidor, como. AWS Lambda Las funciones Lambda pueden ser una buena opción para tareas de corta duración impulsadas por eventos, ya que escalan y gestionan automáticamente la administración de recursos.

Seguridad

Para las tareas informáticas de rastreo web, le recomendamos que configure el entorno para que solo permita el tráfico saliente. Esto ayuda a mejorar la seguridad al minimizar la superficie de ataque y reducir el riesgo de acceso entrante no autorizado. Al permitir únicamente las conexiones salientes, el proceso de rastreo permite comunicarse con los sitios web de destino y recuperar los datos necesarios, además de restringir cualquier tráfico entrante que pueda comprometer el sistema.

Otras consideraciones

Revise las siguientes consideraciones y prácticas recomendadas adicionales:

- Consulta las pautas de rastreo en las condiciones de servicio o la política de privacidad del sitio web.
- Busca meta etiquetas en el HTML que puedan proporcionar directivas de rastreo.
- Tenga en cuenta las restricciones legales de su jurisdicción en relación con la recopilación y el uso de datos.
- Prepárate para dejar de rastrear si así lo solicita el propietario del sitio web.

Preguntas frecuentes

¿Qué sucede si el archivo robots.txt no está disponible?

La ausencia de un archivo robots.txt no significa necesariamente que no puedas o no debas rastrear un sitio web. El rastreo siempre debe realizarse de manera responsable, respetando los recursos del sitio web y los derechos implícitos del propietario del sitio web.

¿Qué sucede si el archivo sitemaps.xml no está disponible?

Según el requisito, puede realizar una de las siguientes acciones:

- Busque mapas de sitio HTML: busque una página de mapa de sitio HTML que enumere las páginas importantes del sitio web. Suelen estar enlazados en el pie de página.
- Rastrea desde la página de inicio: comienza a rastrear desde la página de inicio y sigue los enlaces internos para descubrir otras páginas.
- Analice los patrones de URL: analice la estructura de URL del sitio web para identificar patrones y generar potencial mediante programación. URLs
- Revisa el archivo robots.txt: comprueba si hay páginas o directorios no permitidos en el archivo robots.txt. Estos pueden proporcionar pistas sobre la estructura del sitio.
- Revisa los puntos finales de la API: algunos sitios web ofrecen puntos finales de la API que se pueden utilizar para recuperar información de contenido y estructura.
- Comprueba los resultados de los motores de búsqueda: utiliza los motores de búsqueda para encontrar páginas indexadas del sitio web utilizando el [sitio: un operador de búsqueda](#), por ejemplo. `site:example.com`
- Analice los backlinks: analice los backlinks al sitio web para descubrir páginas importantes a las que otros sitios enlazan.
- Revisa los archivos web: consulta los archivos de Internet, como el [Wayback Machine](#), para ver si hay versiones anteriores del sitio que podrían haber tenido mapas de sitio o estructuras diferentes.
- Busca los patrones del sistema de administración de contenido (CMS): si puedes identificar el CMS, usa los patrones de URL comunes asociados a ese sistema.
- Confirma la JavaScript representación: si el sitio depende en gran medida de JavaScript él, asegúrate de que tu rastreador pueda renderizar JavaScript para detectar el contenido cargado

dinámicamente. En algunos sitios web, el archivo sitemap.xml se carga una vez que se habilita la JavaScript renderización.

¿Puedo usar una solución sin servidor en lugar de Amazon EC2 o Amazon ECS?

Sí. [AWS Lambda](#) las funciones de rastreo web pueden ser una opción viable, especialmente para tareas de rastreo más modulares o de menor escala. Sin embargo, para operaciones de rastreo a gran escala y de larga duración, podría ser más adecuado un enfoque más tradicional que utilice instancias de Amazon Elastic Compute Cloud (Amazon EC2) o Amazon Elastic Container Service (Amazon ECS). Es importante evaluar detenidamente los requisitos específicos y las ventajas y desventajas a la hora de elegir el servicio de cómputo adecuado para sus necesidades de rastreo web.

¿Por qué el rastreador recibe un código de estado 403?

El HTTP 403 es un código de estado HTTP que significa que está prohibido el acceso al recurso solicitado. Si la solicitud era correcta, el servidor entendió la solicitud y no la atenderá. Para evitar un código de estado 403, puede hacer lo siguiente:

- Limita tu velocidad de rastreo.
- Comprueba si el mapa del sitio o el archivo robots.txt permiten al rastreador acceder a la URL.
- Prueba con un agente de usuario móvil en lugar de un agente de usuario de escritorio.

Si nada de lo anterior funciona, debes respetar la decisión de los propietarios del sitio web y no rastrear la página.

Próximos pasos y recursos

Tras recopilar los datos ambientales, sociales y de gobierno (ESG) sin procesar, puede hacer lo siguiente para extraer información significativa de los datos:

1. Limpie los datos: los datos pueden incluir una cantidad significativa de información irrelevante que no esté relacionada con los factores ESG y los datos financieros. Es importante eliminar estos datos irrelevantes y conservar solo la información necesaria para realizar los análisis necesarios. Puede utilizar herramientas, como [yfinance](#), para ayudar a limpiar los datos.
2. Extraiga y transforme los datos: extraiga las características o variables relevantes de los datos sin procesar y transfórmelas en un formato adecuado para el análisis. Puede transformar los datos en formato tabular para mejorar la legibilidad y la claridad. Puede utilizar una biblioteca, como [pandas](#), para refinar los datos. También puede utilizar la ingeniería de características, la normalización de datos y las métricas derivadas para transformar los datos.
3. Realizar análisis: puede realizar diversas tareas analíticas. Esto podría incluir la generación de estadísticas descriptivas, la creación de visualizaciones de datos y la realización de análisis exploratorios de datos para obtener información sobre el desempeño de las empresas en materia de ESG.
4. Aplique el aprendizaje automático: puede usar los datos depurados y transformados para entrenar modelos de aprendizaje automático. Estos modelos pueden ayudarle a identificar las empresas que actualmente están demostrando sostenibilidad financiera y a proyectar su futuro desempeño en materia de sostenibilidad.

Al utilizar el rastreador web y este proceso de evaluación de datos, puede obtener de manera efectiva una comprensión integral de las prácticas de sostenibilidad y el desempeño financiero de las empresas que está evaluando. Puede utilizar esta información para fundamentar sus decisiones de inversión, realizar un seguimiento del progreso y respaldar las prácticas empresariales sostenibles.

Recursos

- [¿Qué es un rastreador web?](#) (Cloudflare sitio web)
- [Guía de inversión en ESG](#) (Investopedia sitio web)

Herramientas

- [Beautiful Soup](#) (Beautiful Soup documentación)
- [Manejo de datos tabulares y relacionales](#) (pandas sitio web)

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	—	6 de enero 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:

- **Refactorizar/rediseñar:** traslade una aplicación y modifique su arquitectura mediante el máximo aprovechamiento de las características nativas en 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 edición compatible con PostgreSQL de 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: migre su base de datos Oracle local a Amazon Relational Database Service (Amazon RDS) para Oracle en el. 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 realizar cambios para aprovechar las capacidades de la nube. Ejemplo: migre su base de datos Oracle local a Oracle en una EC2 instancia del. 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 local 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

ABAC

Consulte control de [acceso basado en atributos](#).

servicios abstractos

Consulte [servicios gestionados](#).

ACID

Consulte [atomicidad, consistencia, aislamiento y 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 la 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 base de datos de origen gestiona las transacciones de las aplicaciones conectadas, mientras los datos se replican en la base de datos de destino. La base de datos de destino no acepta ninguna transacción durante la migración.

función agregada

Función SQL que opera en un grupo de filas y calcula un único valor de retorno para el grupo. Algunos ejemplos de funciones agregadas incluyen SUM y MAX.

IA

Véase [inteligencia artificial](#).

AIOps

Consulte las [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.

antipatronos

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

Un enfoque de seguridad que permite el uso únicamente de 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 AIOps se utiliza en la estrategia de AWS migración, 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

Un bot malo

Un [bot](#) destinado a interrumpir o causar daño a personas u organizaciones.

BCP

Consulte la [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. Véase también [endianness](#).

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.

implementación azul/verde

Una estrategia de despliegue 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 en el otro entorno (verde). Esta estrategia le ayuda a revertirla rápidamente 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 información en Internet. Algunos otros bots, conocidos como bots malos, tienen como objetivo interrumpir o causar daños a personas u organizaciones.

botnet

Redes de [bots](#) que están infectadas por [malware](#) y que están bajo el control de una sola parte, conocida como pastor 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 con cristales rotos

En circunstancias excepcionales y mediante un proceso aprobado, un usuario puede acceder rápidamente a un sitio para el Cuenta de AWS que normalmente no tiene permisos de acceso. Para obtener más información, consulte el indicador [Implemente procedimientos de rotura de cristales en la guía Well-Architected AWS](#) .

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 el marco AWS de adopción de la nube.](#)

despliegue canario

El lanzamiento lento e incremental de una versión para los usuarios finales. Cuando está seguro, despliega la nueva versión y reemplaza la versión actual en su totalidad.

CCoE

Consulte [Cloud Center of Excellence](#).

CDC

Consulte la [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

Introducir intencionalmente 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 la [integración continua y la 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.

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 conectada a la tecnología de [computación perimetral](#).

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 cuatro fases por las que suelen pasar las organizaciones cuando migran a Nube de AWS:

- Proyecto: ejecución de algunos proyectos relacionados con la nube con fines de prueba de concepto y aprendizaje
- Fundamento: realizar inversiones fundamentales para escalar su adopción de la nube (p. ej., crear una landing zone, definir una CCoE, establecer un modelo de operaciones)

- Migración: migración de aplicaciones individuales
- Reinención: optimización de productos y servicios e innovación en la nube

Stephen Orban definió estas etapas en la entrada del blog [The Journey Toward Cloud-First & the Stages of Adoption en el](#) 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 la [base de datos de administración de la 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. Los repositorios en la nube más comunes incluyen 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 canalización de CI/CD 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 aprendizaje automático para analizar y extraer información de formatos visuales, como imágenes y vídeos digitales. Por ejemplo, Amazon SageMaker AI proporciona algoritmos de procesamiento de imágenes para CV.

desviación de configuración

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

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

Conjunto de AWS Config reglas y medidas correctivas que puede reunir para personalizar sus comprobaciones 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, compilación, prueba, puesta en escena y producción del proceso de publicación del software. CI/CD is commonly described as a pipeline. CI/CD puede ayudarlo a automatizar los procesos, mejorar la productividad, mejorar la calidad del código y entregar con mayor rapidez. 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

Vea la [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 datos es un componente del pilar de seguridad

del AWS Well-Architected Framework. Para obtener más información, consulte [Clasificación de datos](#).

desviación 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 desviación de los 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

Un marco arquitectónico que proporciona una propiedad de datos distribuida y descentralizada con una administración y un gobierno centralizados.

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

Un sistema de administración de datos que respalde la inteligencia empresarial, como el análisis. Los almacenes de datos suelen contener grandes cantidades de datos históricos y, por lo general, se utilizan para consultas y 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 el [lenguaje de definición de bases de datos](#) 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.

defense-in-depth

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 defense-in-depth enfoque podría combinar la autenticación multifactorial, 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 de 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 utilizan habitualmente para restringir consultas, filtrar y etiquetar 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)

La estrategia y el proceso que se utilizan para minimizar el tiempo de inactividad y la pérdida de datos ocasionados por un [desastre](#). Para obtener más información, consulte [Recuperación ante desastres de cargas de trabajo en AWS: Recovery in the Cloud in the AWS Well-Architected Framework](#).

DML

Consulte el 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. Este concepto lo introdujo Eric Evans en su libro, *Diseño impulsado por el dominio: abordando la complejidad en el corazón del 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 Microsoft ASP.NET \(ASMX\) mediante contenedores y Amazon API Gateway](#).

DR

Consulte [recuperación ante desastres](#).

detección de deriva

Seguimiento de las desviaciones con respecto a una configuración de referencia. 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 [el mapeo del flujo de valor del desarrollo](#).

E

EDA

Consulte el [análisis exploratorio de datos](#).

EDI

Véase [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 nube, la computación](#) perimetral puede reducir la latencia de la comunicación y mejorar el tiempo de respuesta.

intercambio electrónico de datos (EDI)

El intercambio automatizado de documentos comerciales entre organizaciones. Para obtener más información, consulte [Qué es el intercambio electrónico de datos](#).

cifrado

Proceso informático que transforma datos de texto plano, 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. Los sistemas big-endianos almacenan primero el byte más significativo. Los sistemas Little-Endian almacenan primero el byte menos significativo.

punto de conexión

[Consulte el punto final del 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 AWS PrivateLink y conceder permisos a otros directores

Cuentas de AWS o a AWS Identity and Access Management (IAM). 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)

Un sistema que automatiza y gestiona los procesos empresariales clave (como la contabilidad, el [MES](#) y la gestión 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 una canalización de CI/CD, 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 datos

La tabla central de un [esquema en forma de estrella](#). Almacena datos cuantitativos sobre las operaciones comerciales. Normalmente, 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.

fallan rápidamente

Una filosofía que utiliza pruebas frecuentes e incrementales para reducir el ciclo de vida del desarrollo. Es una parte fundamental de un enfoque ágil.

límite de aislamiento de fallas

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 obtener más información, consulte [Límites de AWS aislamiento de errores](#).

rama de característica

Consulte la [sucursal](#).

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.

indicaciones de unos pocos pasos

Proporcionar a un [LLM](#) un pequeño número de ejemplos que demuestren la tarea y el resultado deseado antes de pedirle que realice una tarea similar. Esta técnica es una aplicación del aprendizaje contextual, en el que los modelos aprenden a partir de ejemplos (planos) integrados en las instrucciones. Las indicaciones con pocas tomas pueden ser eficaces para tareas que requieren un formato, un razonamiento o un conocimiento del dominio específicos. [Consulte también el apartado de mensajes sin intervención](#).

FGAC

Consulte el 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 modificados](#) 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 el [modelo básico](#).

modelo de base (FM)

Una gran red neuronal de aprendizaje profundo que se ha estado entrenando con conjuntos de datos masivos de datos generalizados y sin etiquetar. FMs son capaces de realizar una amplia variedad de tareas generales, como comprender el lenguaje, generar texto e imágenes y conversar en lenguaje natural. Para obtener más información, consulte [Qué son los modelos básicos](#).

G

IA generativa

Un subconjunto de modelos de [IA](#) que se han entrenado con grandes cantidades de datos y que pueden utilizar un simple mensaje de texto para crear contenido y artefactos nuevos, como imágenes, vídeos, texto y audio. Para obtener más información, consulte [Qué es la IA generativa](#).

bloqueo geográfico

Consulta [las 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, y el [flujo de trabajo basado en enlaces troncales](#) es el enfoque moderno preferido.

imagen dorada

Instantánea de un sistema o software que se utiliza 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 el cumplimiento en todas las unidades organizativas (OUs). 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 GuardDuty AWS Trusted Advisor, Amazon Inspector y AWS Lambda cheques personalizados.

H

HA

Consulte la [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 retenidos

Parte de los datos históricos etiquetados que se ocultan de un conjunto de datos que se utiliza para entrenar un modelo de aprendizaje [automático](#). Puede utilizar los datos de reserva para evaluar el rendimiento del modelo comparando las predicciones del modelo con los datos de reserva.

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, las revisiones suelen realizarse fuera del flujo de trabajo habitual de las versiones.

DevOps

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

laC

Vea [la infraestructura como código](#).

políticas basadas en identidad

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

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

Un modelo que implementa una nueva infraestructura para las cargas de trabajo de producción en lugar de actualizar, aplicar parches o modificar la infraestructura existente. [Las infraestructuras inmutables son intrínsecamente más consistentes, fiables y predecibles que las infraestructuras mutables](#). Para obtener más información, consulte las prácticas recomendadas para [implementar con una infraestructura inmutable](#) en Well-Architected Framework AWS .

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 AWS de referencia de seguridad](#) recomienda configurar la cuenta de red con entradas, salidas e inspección VPCs para proteger la interfaz bidireccional entre la aplicación y el resto de Internet.

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.

I

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 la inteligencia artificial/aprendizaje automático.

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 de la Internet de las cosas \(IIoT\) industrial](#).

VPC de inspección

En una arquitectura de AWS cuentas múltiples, una VPC centralizada que gestiona las inspecciones del tráfico de red VPCs entre Internet y las redes locales (en una misma o Regiones de AWS diferente). La [arquitectura AWS de referencia de seguridad](#) recomienda configurar su cuenta de red con entrada, salida e inspección VPCs para proteger la interfaz bidireccional entre la 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 la [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 grande (LLM)

Un modelo de [IA](#) de aprendizaje profundo que se entrena previamente con una gran cantidad de datos. Un LLM puede realizar múltiples tareas, como responder preguntas, resumir documentos, traducir textos a otros idiomas y completar oraciones. [Para obtener más información, consulte Qué son. LLMs](#)

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

Ver [7 Rs](#).

sistema little-endian

Un sistema que almacena primero el byte menos significativo. Véase también [endianness](#).

LLM

Véase un modelo de lenguaje [amplio](#).

entornos inferiores

Véase [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

Ver [sucursal](#).

malware

Software diseñado para comprometer la seguridad o la privacidad de la computadora. El malware puede 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 gestionados

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

sistema de ejecución de fabricación (MES)

Un sistema de software para rastrear, monitorear, documentar y controlar los procesos de producción que convierten las materias primas en productos terminados en el taller.

MAP

Consulte [Migration Acceleration Program](#).

mecanismo

Un proceso completo en el que se crea una herramienta, se impulsa su adopción y, a continuación, se inspeccionan los resultados para realizar ajustes. Un mecanismo es un ciclo que se refuerza y mejora a sí mismo a medida que funciona. Para obtener más información, consulte [Creación de mecanismos](#) en el AWS Well-Architected Framework.

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 el [sistema de ejecución de la fabricación](#).

Transporte telemétrico de Message Queue Queue (MQTT)

[Un protocolo de comunicación ligero machine-to-machine \(M2M\), basado en el patrón de publicación/suscripción, para dispositivos de IoT con recursos limitados.](#)

microservicio

Un servicio pequeño e independiente que se comunica a través de una red bien definida APIs 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 AWS servicios sin servidor](#).

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 un uso ligero. APIs 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

Equipos multifuncionales que agilizan la migración de las cargas de trabajo mediante enfoques automatizados y ágiles. Los equipos de las fábricas de migración suelen incluir a analistas y propietarios de operaciones, empresas, 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: realoje la migración a Amazon EC2 con AWS Application Migration Service.

Migration Portfolio Assessment (MPA)

Una herramienta en línea que proporciona información para validar el modelo de negocio para migrar a 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 asociados 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

El enfoque utilizado para migrar una carga de trabajo a Nube de AWS Para obtener más información, consulte la entrada de las [7 R](#) de este glosario y consulte [Movilice a su organización para acelerar las migraciones a gran escala](#).

ML

[Consulte el aprendizaje automático.](#)

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 obtener más información, consulte [Estrategia para modernizar las aplicaciones en el 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 obtener más información, consulte [Evaluación de la preparación para la modernización de las aplicaciones en el 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](#).

MAPA

Consulte [la evaluación de la cartera de migración](#).

MQTT

Consulte [Message Queue Queue Telemetría](#) y Transporte.

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

Un modelo que actualiza y modifica la infraestructura existente para las cargas de trabajo de producción. Para mejorar la coherencia, la fiabilidad y la previsibilidad, el AWS Well-Architected Framework recomienda el uso [de una infraestructura inmutable](#) como práctica recomendada.

O

OAC

[Consulte el control de acceso de origen.](#)

OAI

Consulte la [identidad de acceso de origen](#).

OCM

Consulte [gestión 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

Véase el [acuerdo a 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 machine-to-machine (M2M) para la automatización industrial. El 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)

Una lista de preguntas y las mejores prácticas asociadas que le ayudan a comprender, evaluar, prevenir o reducir el alcance de los incidentes y posibles fallos. Para obtener más información, consulte [Operational Readiness Reviews \(ORR\)](#) en AWS Well-Architected Framework.

tecnología operativa (OT)

Sistemas de hardware y software que funcionan con el entorno físico para controlar las operaciones, los equipos y la infraestructura industriales. En la industria manufacturera, 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 el AWS CloudTrail que se registran todos los eventos para todos Cuentas de AWS los miembros 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 AWS KMS (SSE-KMS) y las solicitudes dinámicas PUT y DELETE dirigidas 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 la revisión de [la preparación operativa](#).

OT

Consulte la [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 AWS de referencia de seguridad](#) recomienda configurar la cuenta de red con entradas, salidas e inspección VPCs para proteger la interfaz bidireccional entre la 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 la [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 la [gestión del ciclo de vida del producto](#).

policy

Un objeto que puede definir los permisos (consulte la [política basada en la identidad](#)), especifique las condiciones de acceso (consulte la [política basada en los recursos](#)) o defina los permisos máximos para todas las cuentas de una organización AWS Organizations (consulte la política de control de [servicios](#)).

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. Para obtener más información, consulte [Habilitación de la persistencia de datos en los microservicios](#).

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

Una condición de consulta que devuelve true o false, por lo general, se encuentra en una cláusula. WHERE

pulsar un predicado

Técnica de optimización de consultas de bases de datos que filtra los datos de la consulta antes de transferirlos. Esto reduce la cantidad de datos que se deben recuperar y procesar de la base de datos relacional y 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

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

zonas alojadas privadas

Un contenedor que contiene información sobre cómo desea que Amazon Route 53 responda a las consultas de DNS de un dominio y sus subdominios dentro de uno o más VPCs. Para obtener más información, consulte [Uso de zonas alojadas privadas](#) en la documentación de Route 53.

control proactivo

Un [control de seguridad](#) diseñado para evitar el despliegue de recursos no conformes. Estos controles escanean los recursos antes de aprovisionarlos. Si el recurso no cumple con el control, significa que no está aprovisionado. 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.

gestión del ciclo de vida del producto (PLM)

La gestión 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 el rechazo y la retirada.

entorno de producción

Consulte [el entorno](#).

controlador lógico programable (PLC)

En la fabricación, una computadora adaptable y altamente confiable que monitorea las máquinas y automatiza los procesos de fabricación.

encadenamiento rápido

Utilizar la salida de una solicitud de [LLM](#) como entrada para la siguiente solicitud para generar mejores respuestas. Esta técnica se utiliza para dividir una tarea compleja en subtareas 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)

Un patrón que permite las 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 puedan suscribir otros microservicios. El sistema puede añadir 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

Véase [responsable, responsable, consultado, informado \(RACI\)](#).

RAG

Consulte [Retrieval Augmented Generation](#).

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

Véase [responsable, responsable, consultado, informado \(RACI\)](#).

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

Ver [7 Rs](#).

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

Ver [7 Rs.](#)

Región

Una colección 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 obtener más información, consulte [Regiones de AWS Especificar qué cuenta puede usar.](#)

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 [7 Rs.](#)

versión

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

trasladarse

Ver [7 Rs.](#)

redefinir la plataforma

Ver [7 Rs.](#)

recompra

Ver [7 Rs.](#)

resiliencia

La capacidad de una aplicación para resistir las interrupciones o recuperarse de ellas. [La alta disponibilidad](#) y la [recuperación ante desastres](#) son consideraciones comunes a la hora de planificar la resiliencia en el. Nube de AWS Para obtener más información, consulte [Nube de AWS Resiliencia](#).

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 [7 Rs](#).

jubilarse

Ver [7 Rs](#).

Generación aumentada de recuperación (RAG)

Tecnología de [inteligencia artificial generativa](#) en la que un máster [hace referencia](#) a una fuente de datos autorizada que se encuentra fuera de sus fuentes de datos de formación antes de generar una respuesta. Por ejemplo, un modelo RAG podría realizar una búsqueda semántica en la base de conocimientos o en los datos personalizados de una organización. Para obtener más información, consulte [Qué es](#) el RAG.

rotación

Proceso de actualizar periódicamente un [secreto](#) para dificultar el acceso de un atacante 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 el [objetivo del punto de recuperación](#).

RTO

Consulte el [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 AWS Management Console 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 el [control de supervisión y la adquisición de datos](#).

SCP

Consulte la [política de control de servicios](#).

secreta

Información confidencial o restringida, como una contraseña o credenciales de usuario, que almacene de forma cifrada. AWS Secrets Manager Se compone del valor secreto y sus metadatos. El valor secreto puede ser binario, una sola cadena o varias cadenas. Para obtener más información, consulta [¿Qué hay en un secreto de Secrets Manager?](#) en la documentación de Secrets Manager.

seguridad desde el diseño

Un 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 principales de controles de seguridad: [preventivos](#), [de detección](#), con [capacidad](#) 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

Una acción predefinida y programada que está diseñada para responder automáticamente a un evento de seguridad o remediarlo. Estas automatizaciones sirven como controles de seguridad [detectables](#) o [adaptables](#) que le ayudan a implementar las mejores prácticas AWS de seguridad. Algunos ejemplos de acciones de respuesta automatizadas incluyen la modificación de un grupo de seguridad de VPC, la aplicación de parches a una EC2 instancia de Amazon o la rotación de credenciales.

cifrado del servidor

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

política de control de servicio (SCP)

Política que proporciona un control centralizado de los permisos de todas las cuentas de una organización en AWS Organizations. SCPs defina barreras o establezca límites a las acciones que un administrador puede delegar en usuarios o roles. Puede utilizarlas SCPs como listas de permitidos o rechazados para especificar qué servicios o acciones están permitidos o prohibidos. Para obtener más información, consulte [las políticas de control de servicios](#) 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)

[Una 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 que compartes con respecto a la seguridad y AWS 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](#).

SIEM

Consulte [la información de seguridad y el sistema de gestión de eventos](#).

punto único de fallo (SPOF)

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

SLA

Consulte el acuerdo [de nivel de servicio](#).

SLI

Consulte el indicador de [nivel de servicio](#).

SLO

Consulte el objetivo de nivel de [servicio](#).

split-and-seed modelo

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 obtener más información, consulte [Enfoque gradual para modernizar las aplicaciones en el. Nube de AWS](#)

SPOF

Consulte el [punto único de falla](#).

esquema en forma de estrella

Estructura organizativa de una base de datos que utiliza una tabla de hechos grande para almacenar datos medidos o transaccionales y una o más tablas dimensionales más pequeñas para almacenar los atributos de los datos. Esta estructura está diseñada para usarse 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 dismantelar 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 con la aplicación de 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.

supervisión, control y adquisición de datos (SCADA)

En la industria manufacturera, un sistema que utiliza hardware y software para monitorear 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

Probar un sistema de manera que simule las interacciones de los usuarios para detectar posibles problemas o monitorear el rendimiento. Puede usar [Amazon CloudWatch Synthetics](#) para crear estas pruebas.

indicador del sistema

Una técnica para proporcionar contexto, instrucciones o pautas a un [LLM](#) para dirigir su comportamiento. Las indicaciones del sistema ayudan a establecer el contexto y las reglas para las interacciones con los usuarios.

T

etiquetas

Pares clave-valor que actúan como metadatos para organizar los recursos. AWS Las etiquetas pueden ayudarle a administrar, identificar, organizar, buscar y filtrar recursos. 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.

puerta de enlace de tránsito

Un centro de tránsito de red que puede usar para interconectar sus VPCs redes con las locales. 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. Para más información, consulte la guía [Cuantificación de la incertidumbre en los sistemas de aprendizaje profundo](#).

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

Ver [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

Una conexión entre dos VPCs que le 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 realiza un cálculo en un grupo de filas que se relacionan de alguna manera con el registro actual. Las funciones de ventana son útiles para procesar tareas, 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.

GUSANO

Mira, [escribe una vez, lee muchas](#).

WQF

Consulte el [marco AWS de calificación de la carga](#) de trabajo.

escribe una vez, lee muchas (WORM)

Un modelo de almacenamiento que escribe los datos una sola vez y evita que los datos se eliminen o modifiquen. Los usuarios autorizados pueden leer los datos tantas veces como sea necesario, pero no pueden cambiarlos. Esta infraestructura de almacenamiento de datos se considera [inmutable](#).

Z

ataque de día cero

Un ataque, normalmente de malware, que aprovecha 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.

aviso de tiro cero

Proporcionar a un [LLM](#) instrucciones para realizar una tarea, pero sin ejemplos (imágenes) que puedan ayudar a guiarla. El LLM debe utilizar sus conocimientos previamente entrenados para realizar la tarea. La eficacia de las indicaciones cero depende de la complejidad de la tarea y de la calidad de las indicaciones. [Consulte también las indicaciones de 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 versión original de inglés, prevalecerá la versión en inglés.