



Creando soluciones de recuperación y generación aumentada para la AWS atención médica

# AWS Guía prescriptiva



# AWS Guía prescriptiva: Creando soluciones de recuperación y generación aumentada para la AWS atención médica

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
Atención al paciente y productividad .....	2
Gestión del talento .....	2
Oportunidades y desafíos .....	4
Oportunidades para las aplicaciones de IA generativa en el sector sanitario .....	4
Análisis avanzado de imágenes .....	5
¿Cuáles son los desafíos de la industrialización de las soluciones .....	5
Caso de uso: creación de una aplicación de inteligencia médica .....	6
Información general de la solución .....	6
Paso 1: Descubrimiento de datos .....	8
Paso 2: Crear un gráfico de conocimientos médicos .....	9
Paso 3: Crear agentes de recuperación de contexto .....	15
Agentes de Amazon Bedrock .....	15
LangChain agentes .....	17
Paso 4: Crear una base de conocimientos .....	18
Uso del OpenSearch servicio .....	19
Crear una arquitectura RAG .....	20
Paso 5: Generar respuestas .....	23
Alineación con el marco de AWS Well-Architected .....	25
Caso de uso: predecir las tasas de reingreso .....	26
Información general de la solución .....	26
Paso 1: Predecir los resultados de los pacientes .....	29
Paso 2: Predecir el comportamiento del paciente .....	31
Paso 3: Predecir el reingreso del paciente .....	33
Paso 4: Calcular la puntuación de propensión .....	36
Alineación con el marco de AWS Well-Architected .....	39
Caso de uso: gestión del talento .....	40
Información general de la solución .....	41
Paso 1: Crear un perfil de habilidades .....	43
Paso 2: Descubrir role-to-skill la relevancia .....	44
Paso 3: Recomendar la formación .....	45
Alineación con el marco de AWS Well-Architected .....	46
Desarrollando soluciones .....	48
Amazon Q Developer .....	48

---

Diseño RAG con múltiples recuperadores .....	49
ReAct agentes .....	51
Evaluación de soluciones .....	53
Evaluación de la extracción de información .....	53
Evaluación de varios recuperadores .....	54
Uso de un LLM .....	54
Recursos .....	56
AWS documentación .....	56
AWS publicaciones de blog .....	56
Otros recursos .....	56
Colaboradores .....	57
Creación .....	57
Revisando .....	57
Redacción técnica .....	57
Historial de documentos .....	58
Glosario .....	59
# .....	59
A .....	60
B .....	63
C .....	65
D .....	68
E .....	73
F .....	75
G .....	77
H .....	78
I .....	80
L .....	82
M .....	83
O .....	88
P .....	91
Q .....	94
R .....	94
S .....	97
T .....	101
U .....	103
V .....	103

---

---

W .....	104
Z .....	105
.....	cvi

# Creando soluciones de recuperación y generación aumentada AWS para el cuidado de la salud

Amazon Web Services, Accenture, y Cadiem ([colaboradores](#))

Marzo de 2025 ([historial del documento](#))

Antes de los grandes modelos lingüísticos (LLMs) y la IA generativa, la tarea de desarrollar aplicaciones automatizadas y de alta precisión en el sector de la salud era todo un desafío. Los métodos tradicionales se basaban en gran medida en la entrada y el análisis manuales de los datos. La complejidad de analizar las imágenes médicas y los registros de los pacientes requería una amplia intervención humana, lo que a menudo daba lugar a flujos de trabajo fragmentados e ineficientes. El avance de las tecnologías de IA le ayuda a crear aplicaciones hiperpersonalizadas a escala. Las aplicaciones sanitarias ahora pueden integrarse con las bases de conocimientos médicos, interpretar las imágenes de diagnóstico con mayor precisión y pronosticar los resultados de los pacientes mediante modelos predictivos.

Esta guía analiza cómo LLMs están revolucionando la asistencia sanitaria mediante aplicaciones de generación aumentada de recuperación con las que puede crear aplicaciones. Servicios de AWSLa generación aumentada de recuperación (RAG) es una tecnología de IA 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. Las aplicaciones de RAG basan los resultados del modelo en el conocimiento del mundo real, lo que reduce las alucinaciones y aumenta la relevancia de la respuesta. En el sector de la salud, el RAG se puede utilizar para proporcionar información up-to-date médica y precisa, lo que garantiza que los proveedores de atención médica tengan acceso a las investigaciones y directrices clínicas más recientes. Al transformar los datos en información procesable y automatizar procesos complejos, estas tecnologías ayudan a mejorar la atención a los pacientes, agilizar las operaciones y mejorar la productividad de los profesionales de la salud.

En [Amazon Bedrock](#), puede ajustar los LLMs e integrarlos con agentes inteligentes para crear soluciones sanitarias avanzadas. Al destacar la sinergia entre [Amazon OpenSearch Service](#) y [Amazon Neptune](#), la guía demuestra cómo estos servicios mejoran las soluciones RAG mediante una mayor relevancia de búsqueda y una recuperación avanzada de datos de múltiples fuentes. Puede organizar soluciones integrales de Amazon Bedrock que utilicen agentes de Amazon Bedrock y [LangChain](#) para coordinar sin problemas las interacciones entre diversos repositorios de datos. Esta

integración demuestra el poder de combinar servicios especializados para crear sistemas impulsados por la IA más eficaces y eficientes.

## Atención al paciente y productividad

Esta guía presenta dos casos de uso reales para la atención y la productividad de los pacientes: el [aumento de los datos de los pacientes](#) y la [predicción](#) de los riesgos de reingreso. Proporciona planes estratégicos para implementar estas soluciones a escala, ofreciendo a las organizaciones de atención médica un camino claro hacia la industrialización de los procesos impulsados por la IA. Gracias a estos conocimientos, las instituciones sanitarias pueden utilizar tecnologías de IA avanzadas para crear flujos de trabajo más eficientes e inteligentes.

## Gestión del talento

Esta guía también describe las estrategias para volver a capacitar y capacitar a los trabajadores de la salud para que integren sin problemas la IA generativa en sus rutinas diarias. Esto puede mejorar tanto la productividad como la calidad de la atención al paciente. Al dotar a sus empleados de las habilidades necesarias para utilizar eficazmente las herramientas avanzadas de IA, las organizaciones sanitarias pueden maximizar el retorno de la inversión e impulsar la innovación en la atención a los pacientes.

Esta [solución de gestión del talento](#) basada en la IA incluye las siguientes características clave:

- **Analizador inteligente de currículums de talentos:** al utilizar el avanzado LLMs disponible en Amazon Bedrock, esta herramienta extrae y analiza de manera eficiente las habilidades y atributos fundamentales de los talentos de los currículums. Esta herramienta puede agilizar el proceso de contratación.
- **Base de conocimientos sobre talentos:** con la tecnología de Amazon Neptune, esta base de datos dinámica proporciona información en tiempo real sobre los niveles de personal, la distribución de habilidades y las tendencias del sector. Esto le ayuda a tomar decisiones basadas en datos sobre la administración de la fuerza laboral.
- **Motor de recomendaciones de aprendizaje:** esta herramienta impulsada por la IA identifica las brechas de habilidades dentro de la organización y recomienda programas de capacitación personalizados para el personal médico. Esta herramienta promueve el desarrollo profesional continuo y ayuda a su fuerza laboral a adaptarse a las tecnologías sanitarias en evolución.

En conjunto, estas funciones impulsadas por la IA ayudan a optimizar el rendimiento de la fuerza laboral, revolucionando la gestión del talento con una mayor inteligencia y eficiencia.

## Oportunidades y desafíos

Amazon Bedrock puede proporcionar una mayor productividad, escalabilidad, rentabilidad e información basada en datos. Amazon Bedrock permite a las organizaciones sanitarias utilizar LLMs eficazmente diversos casos de uso, desde la creación de contenido y el análisis de datos hasta la toma de decisiones automatizada. Esta guía proporciona enfoques para superar los desafíos generativos comunes de la IA, como los problemas de calidad de los datos, la escalabilidad de la infraestructura, el mantenimiento del rendimiento del modelo y los requisitos de mejora continua durante la transición de la prueba de concepto a la producción.

## Oportunidades para las aplicaciones de IA generativa en el sector sanitario

El sector de la salud está preparado para un cambio transformador, impulsado por las oportunidades que ofrecen las aplicaciones de IA generativa. La IA generativa tiene el potencial de mejorar la atención a los pacientes, agilizar las operaciones y acelerar la investigación médica. Mediante el uso de modelos avanzados de IA, los proveedores de atención médica pueden automatizar el aumento de los registros médicos. Los historiales completos y de los up-to-date pacientes facilitan diagnósticos y planes de tratamiento más precisos. El análisis de imágenes impulsado por la IA, como la interpretación de ecografías y otras imágenes médicas, puede proporcionar información rápida y precisa, lo que reduce la carga de trabajo de los profesionales médicos y minimiza el riesgo de errores humanos.

Más allá del diagnóstico y el tratamiento, la IA generativa puede desempeñar un papel fundamental en el análisis predictivo. El análisis predictivo ayuda a las organizaciones sanitarias a anticipar los resultados de los pacientes y a personalizar los planes de atención en consecuencia. Esta tecnología también puede optimizar los procesos administrativos, desde la gestión de los datos de los pacientes hasta la agilización de la comunicación entre los proveedores y los pacientes. Al integrar las soluciones de IA generativa con los sistemas de salud existentes, las instituciones médicas pueden lograr una mayor eficiencia, reducir los costos y, en última instancia, ofrecer una atención de mayor calidad. La integración de la IA en la atención médica no es solo una mejora, sino un cambio fundamental hacia una atención más inteligente, receptiva y centrada en el paciente.

## Análisis avanzado de imágenes

La combinación de Amazon Bedrock con almacenes de datos, como Amazon Neptune y OpenSearch Amazon Service, puede ayudarle a abordar las complejidades del análisis avanzado de imágenes en el sector sanitario. Las soluciones de recuperación de información pueden aumentar el proceso de descubrimiento de enfermedades y mejorar la precisión de la interpretación mediante la evaluación de las imágenes de diagnóstico y la interpretación de las ecografías. La solución puede integrar los datos de la evaluación visual y textual con la revisión manual de la evaluación del paciente por parte de los médicos.

## ¿Cuáles son los desafíos de la industrialización de las soluciones

Los principales obstáculos que hay que abordar a la hora de industrializar las soluciones de IA en el sector sanitario son la calidad y la disponibilidad de los datos. Los datos sanitarios suelen estar en formatos fragmentados e incoherentes. Asegurarse de que los modelos de IA tengan acceso a datos limpios, estructurados y representativos es crucial para mantener el rendimiento en escenarios del mundo real. La escalabilidad de la infraestructura puede convertirse en un desafío debido a los entornos de producción. Estos entornos deben gestionar grandes volúmenes de datos de pacientes en tiempo real y, al mismo tiempo, proporcionar tiempos de respuesta rápidos y mantener el cumplimiento de las normas de privacidad de datos, como la Ley de Portabilidad y Responsabilidad de los Seguros de Salud (HIPAA). Además, dado que la información médica emergente y los datos de los pacientes evolucionan con el tiempo, los modelos de IA deben reentrenarse y actualizarse para mantener su relevancia y ofrecer recomendaciones precisas. Por último, la integración de estas soluciones de IA en los sistemas de salud existentes puede resultar compleja debido a los problemas de interoperabilidad y a la necesidad de alinearlas con los flujos de trabajo clínicos actuales. Esta integración requiere cambios tanto técnicos como operativos.

# Caso de uso: creación de una aplicación de inteligencia médica con datos aumentados de pacientes

La IA generativa puede ayudar a aumentar la atención a los pacientes y la productividad del personal al mejorar las funciones clínicas y administrativas. El análisis de imágenes impulsado por la IA, como la interpretación de ecografías, acelera los procesos de diagnóstico y mejora la precisión. Puede proporcionar información fundamental que respalde las intervenciones médicas oportunas.

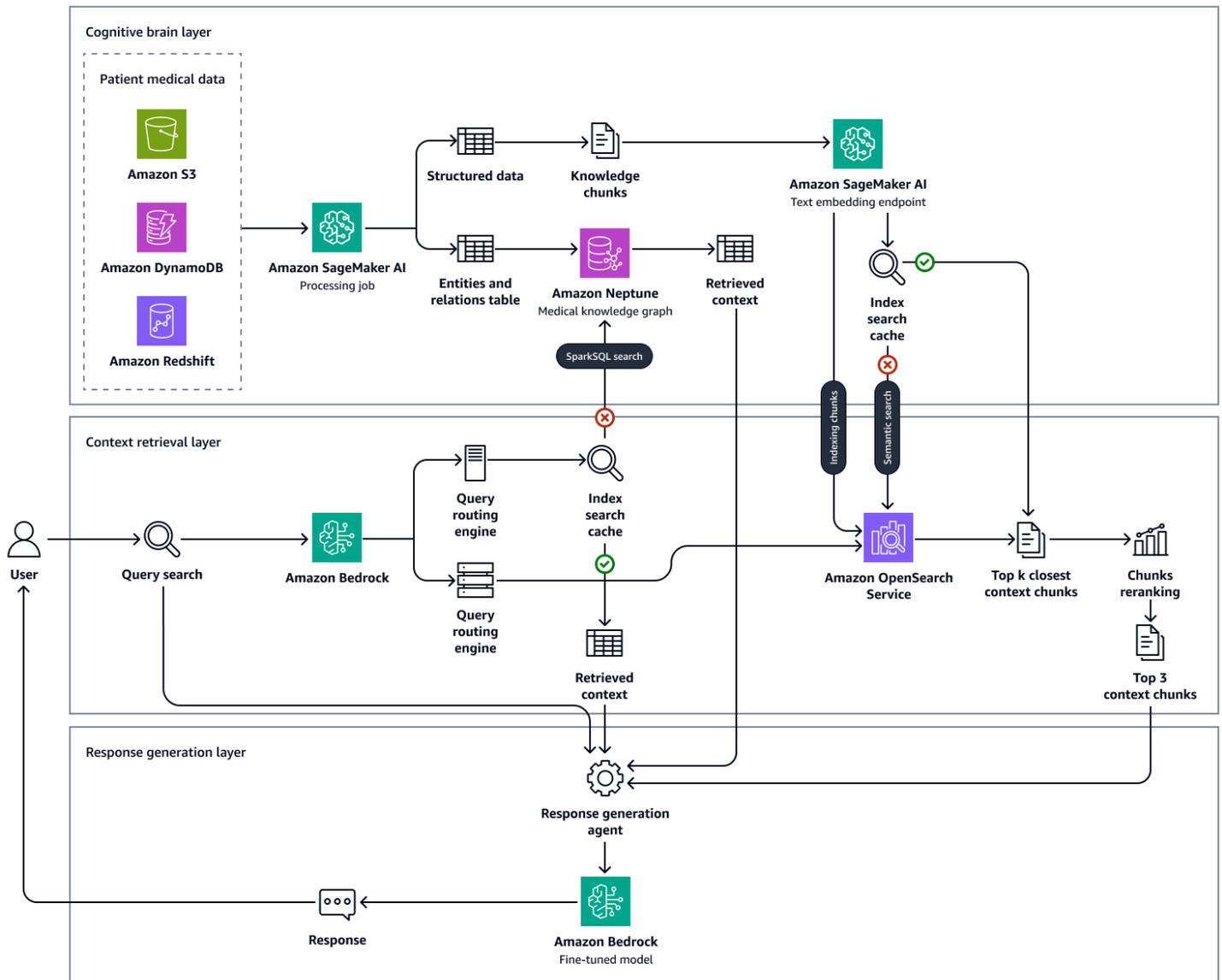
Al combinar modelos de IA generativos con gráficos de conocimiento, puede automatizar la organización cronológica de los registros electrónicos de los pacientes. Esto le ayuda a integrar datos en tiempo real de las interacciones entre el médico y el paciente, los síntomas, los diagnósticos, los resultados de laboratorio y el análisis de imágenes. Esto proporciona al médico datos completos del paciente. Estos datos ayudan al médico a tomar decisiones médicas más precisas y oportunas, lo que mejora tanto los resultados de los pacientes como la productividad de los proveedores de atención médica.

## Información general de la solución

La IA puede ayudar a los médicos y médicos al sintetizar los datos de los pacientes y el conocimiento médico para proporcionar información valiosa. Esta solución de generación aumentada de recuperación (RAG) es un motor de inteligencia médica que consume un conjunto integral de datos y conocimientos de los pacientes procedentes de millones de interacciones clínicas. Aprovecha el poder de la IA generativa para crear información basada en la evidencia para mejorar la atención a los pacientes. Está diseñado para mejorar los flujos de trabajo clínicos, reducir los errores y mejorar los resultados de los pacientes.

La solución incluye una capacidad de procesamiento de imágenes automatizada que funciona con LLMs. Esta capacidad reduce la cantidad de tiempo que el personal médico debe dedicar manualmente a buscar imágenes de diagnóstico similares y a analizar los resultados del diagnóstico.

En la siguiente imagen se muestra end-to-end-workflow la solución correspondiente. Utiliza Amazon Neptune, Amazon SageMaker AI, Amazon OpenSearch Service y un modelo básico en Amazon Bedrock. Para el agente de recuperación de contexto que interactúa con el gráfico de conocimientos médicos de Neptune, puede elegir entre un agente de Amazon Bedrock y un LangChain agente.



En nuestros experimentos con ejemplos de preguntas médicas, observamos que las respuestas finales generadas por nuestro enfoque, que utilizó un gráfico de conocimiento mantenido en Neptune, una base de datos OpenSearch vectorial que alberga la base de conocimientos clínicos y Amazon Bedrock, LLMs se basaban en la realidad y son mucho más precisas, ya que reducen los falsos positivos y aumentan los verdaderos positivos. Esta solución puede generar información basada en evidencias sobre el estado de salud del paciente y tiene como objetivo mejorar los flujos de trabajo clínicos, reducir los errores y mejorar los resultados de los pacientes.

La creación de esta solución consta de los siguientes pasos:

- [Paso 1: Descubrimiento de datos](#)

- [Paso 2: Crear un gráfico de conocimientos médicos](#)
- [Paso 3: Crear agentes de recuperación de contexto para consultar el gráfico del conocimiento médico](#)
- [Paso 4: Crear una base de conocimientos de datos descriptivos en tiempo real](#)
- [Paso 5: LLMs Utilízalo para responder a preguntas médicas](#)

## Paso 1: Descubrimiento de datos

Existen muchos conjuntos de datos médicos de código abierto que puede utilizar para respaldar el desarrollo de una solución sanitaria basada en la IA. Uno de estos conjuntos de datos es el conjunto de datos [MIMIC-IV, que es un conjunto](#) de datos de historias clínicas electrónicas (EHR) disponible públicamente y que se utiliza ampliamente en la comunidad de investigación sanitaria. El MIMIC-IV contiene información clínica detallada, incluidas notas de alta en texto libre extraídas de las historias clínicas de los pacientes. Puede utilizar estos registros para experimentar con técnicas de suma de texto y extracción de entidades. Estas técnicas le ayudan a extraer información médica (como los síntomas del paciente, los medicamentos administrados y los tratamientos recetados) a partir de un texto no estructurado.

También puede utilizar un conjunto de datos que proporcione resúmenes de alta de pacientes anotados y anónimos, seleccionados específicamente para fines de investigación. Un conjunto de datos resumidos sobre el alta médica puede ayudarte a experimentar con la extracción de entidades, ya que te permite identificar entidades médicas clave (como afecciones, procedimientos y medicamentos) a partir del texto. [Paso 2: Crear un gráfico de conocimientos médicos](#) en esta guía, se describe cómo puede utilizar los datos estructurados extraídos del conjunto de datos MIMIC-IV y del conjunto de datos resumidos del alta médica para crear un gráfico de conocimientos médicos. Este gráfico de conocimientos médicos es la columna vertebral de los sistemas avanzados de consulta y apoyo a la toma de decisiones para los profesionales de la salud.

Además de los conjuntos de datos basados en texto, puede utilizar conjuntos de datos de imágenes. Por ejemplo, el [conjunto de datos de radiografías musculoesqueléticas \(MURA\)](#), que es una base de datos completa de imágenes radiográficas de huesos con múltiples vistas. Utilice estos conjuntos de datos de imágenes para experimentar con la evaluación diagnóstica mediante técnicas de decodificación de imágenes médicas. Estas técnicas de decodificación son cruciales para el diagnóstico precoz de enfermedades, como las enfermedades musculoesqueléticas, las enfermedades cardiovasculares y la osteoporosis. Al ajustar con precisión los modelos básicos de visión y lenguaje del conjunto de datos de imágenes médicas, se pueden detectar anomalías

en las imágenes de diagnóstico. Esto ayuda al sistema a proporcionar a los médicos información diagnóstica temprana y precisa. Al utilizar conjuntos de datos de imágenes y texto, puede crear una aplicación sanitaria basada en la IA que sea capaz de procesar datos de texto e imágenes para mejorar la atención a los pacientes.

## Paso 2: Crear un gráfico de conocimientos médicos

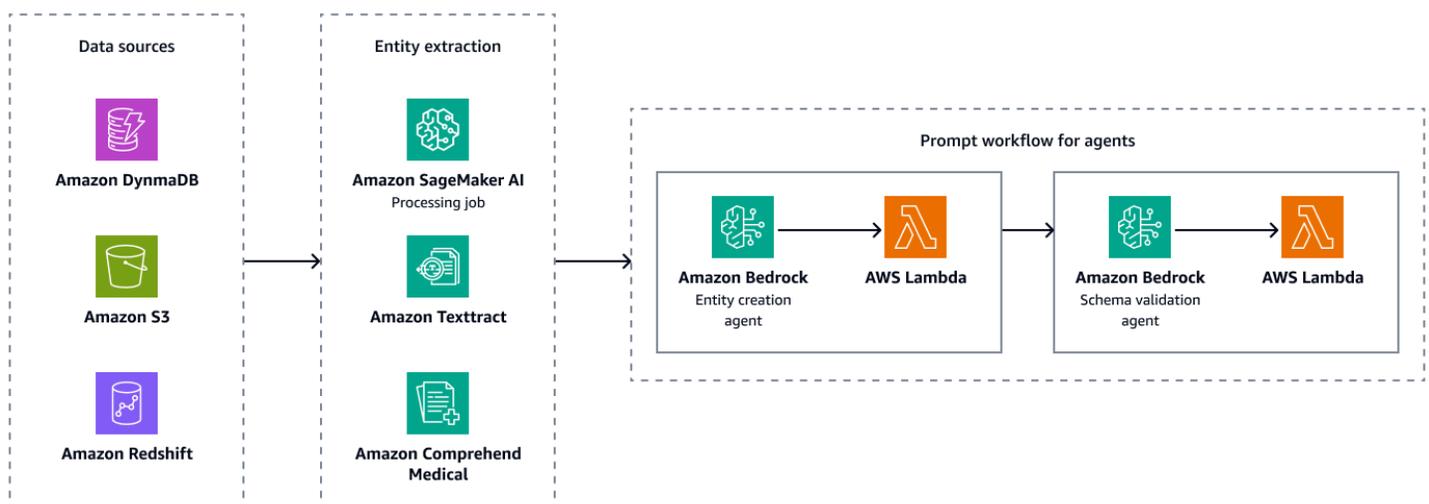
Para cualquier organización de salud que desee crear un sistema de apoyo a la toma de decisiones basado en una base de conocimientos masiva, un desafío clave es localizar y extraer las entidades médicas que están presentes en las notas clínicas, las revistas médicas, los resúmenes de alta y otras fuentes de datos. También es necesario recopilar las relaciones temporales, los temas y las evaluaciones de certeza de estos registros médicos para poder utilizar de forma eficaz las entidades, los atributos y las relaciones extraídas.

El primer paso consiste en extraer los conceptos médicos del texto médico no estructurado utilizando un mensaje de pocas tomas para un modelo básico, como Llama 3 en Amazon Bedrock. Las solicitudes de pocas tomas se dan cuando se le proporciona a un LLM un número reducido de ejemplos que muestran la tarea y el resultado deseado antes de pedirle que realice una tarea similar. Con un extractor de entidades médicas basado en el LLM, puede analizar el texto médico no estructurado y, a continuación, generar una representación de datos estructurados de las entidades del conocimiento médico. También puede almacenar los atributos del paciente para su posterior análisis y automatización. El proceso de extracción de entidades incluye las siguientes acciones:

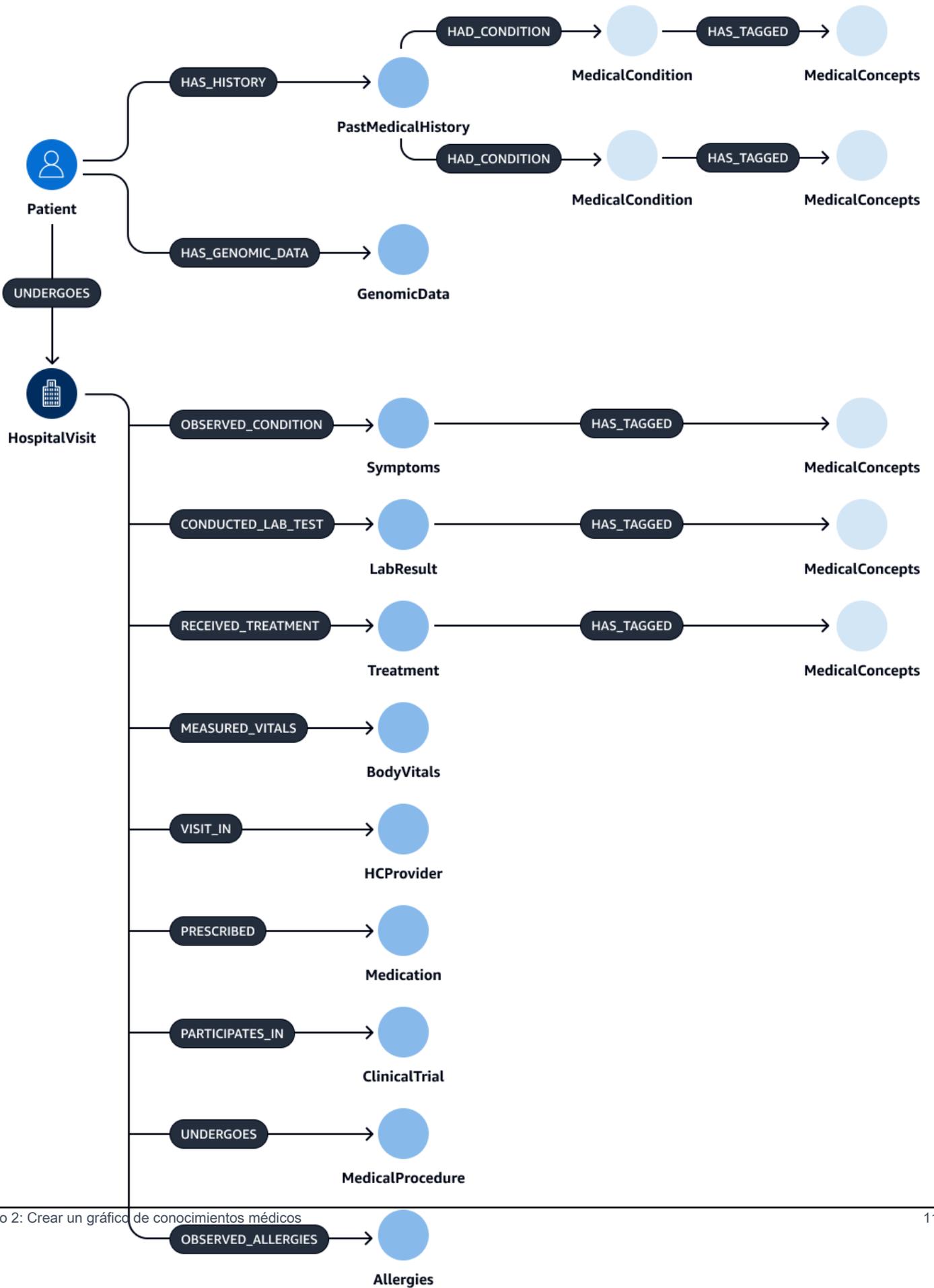
- Extraiga información sobre conceptos médicos, como enfermedades, medicamentos, dispositivos médicos, dosis, frecuencia y duración de los medicamentos, síntomas, procedimientos médicos y sus atributos clínicamente relevantes.
- Capture las características funcionales, como las relaciones temporales entre las entidades extraídas, los sujetos y las evaluaciones de certeza.
- Amplíe los vocabularios médicos estándar, como los siguientes:
  - [Identificadores conceptuales \(RxCUI\) de la base de datos RxNorm](#)
  - Códigos de la [Clasificación Internacional de Enfermedades, décima revisión, modificación clínica \(ICD-10-CM\)](#)
  - Términos de los [encabezados de materias médicas \(MeSH\)](#)
  - Conceptos de la [nomenclatura sistematizada de la medicina, términos clínicos \(SNOMED CT\)](#)
  - Códigos del Sistema [Unificado de Lenguaje Médico \(UMLS\)](#)

- Resuma las notas de alta y obtenga información médica a partir de las transcripciones.

La siguiente figura muestra los pasos de extracción de entidades y validación del esquema para crear combinaciones pareadas válidas de entidades, atributos y relaciones. Puede almacenar datos no estructurados, como resúmenes de alta o notas de pacientes, en Amazon Simple Storage Service (Amazon S3). Puede almacenar datos estructurados, como datos de planificación de recursos empresariales (ERP), historiales electrónicos de pacientes y sistemas de información de laboratorio, en Amazon Redshift y Amazon DynamoDB. Puede crear un agente de creación de entidades en Amazon Bedrock. Este agente puede integrar servicios, como los canales de extracción de datos de Amazon SageMaker AI, Amazon Textract y Amazon Comprehend Medical, para extraer entidades, relaciones y atributos de las fuentes de datos estructuradas y no estructuradas. Por último, utiliza un agente de validación de esquemas de Amazon Bedrock para asegurarse de que las entidades y relaciones extraídas se ajustan al esquema gráfico predefinido y mantienen la integridad de las conexiones entre los extremos del nodo y las propiedades asociadas.



Tras extraer y validar las entidades, las relaciones y los atributos, puede vincularlos para crear un subject-object-predicate triplete. Estos datos se ingieren en una base de datos de gráficos de Amazon Neptune, como se muestra en la siguiente figura. Las [bases de datos de gráficos](#) están optimizadas para almacenar y consultar las relaciones entre los elementos de datos.



Puede crear un gráfico de conocimiento completo con estos datos. Un [gráfico de conocimiento](#) le ayuda a organizar y consultar todo tipo de información conectada. Por ejemplo, puede crear un gráfico de conocimiento que tenga los siguientes nodos principales: HospitalVisit, PastMedicalHistory, Symptoms, Medication, MedicalProcedures, y Treatment.

En las tablas siguientes se enumeran las entidades y sus atributos que se pueden extraer de las notas de aprobación.

Entidad	Atributos
Patient	PatientID , Name, Age, Gender, Address, ContactInformation
HospitalVisit	VisitDate , Reason, Notes
HealthcareProvider	ProviderID , Name, Specialty , ContactInformation , Address, AffiliatedInstitution
Symptoms	Description , RiskFactors
Allergies	AllergyType , Duration
Medication	MedicationID , Name, Description , Dosage, SideEffects , Manufacturer
PastMedicalHistory	ContinuingMedicines
MedicalCondition	ConditionName , Severity, Treatment Received , DoctorinCharge , HospitalName , MedicinesFollowed
BodyVitals	HeartRate , BloodPressure , RespiratoryRate , BodyTemperature , BMI
LabResult	LabResultID , PatientID , TestName, Result, Date

Entidad	Atributos
ClinicalTrial	TrialID, Name, Description , Phase, Status, StartDate , EndDate
GenomicData	GenomicDataID , PatientID , Sequencedata , VariantInformation
Treatment	TreatmentID , Name, Description , Type, SideEffects
MedicalProcedure	ProcedureID , Name, Description , Risks, Outcomes
MedicalConcepts	UMLSCodes , MedicalVocabularies

En la siguiente tabla se enumeran las relaciones que pueden tener las entidades y sus atributos correspondientes. Por ejemplo, la Patient entidad podría conectarse a la HospitalVisit entidad con la [UNDERGOES] relación. El atributo de esta relación es VisitDate.

Entidad objeto	Relación	Entidad objeto	Atributos
Patient	[UNDERGOES]	HospitalVisit	VisitDate
HospitalVisit	[VISIT_IN]	HealthcareProvider	ProviderName , Location, ProviderID , VisitDate
HospitalVisit	[OBSERVED_CONDITION]	Symptoms	Severity, CurrentStatus , VisitDate
HospitalVisit	[RECEIVED_TREATMENT]	Treatment	Duration, Dosage, VisitDate

Entidad objeto	Relación	Entidad objeto	Atributos
HospitalVisit	[PRESCRIBED]	Medication	Duration, Dosage, Adherence , VisitDate
Patient	[HAS_HISTORY]	PastMedicalHistory	Ninguno
PastMedicalHistory	[HAD_CONDITION]	MedicalCondition	DiagnosisDate , CurrentStatus
HospitalVisit	[PARTICIPATES_IN]	ClinicalTrial	VisitDate , Status, Outcomes
Patient	[HAS_GENOMIC_DATA]	GenomicData	CollectionDate
HospitalVisit	[OBSERVED_ALLERGIES]	Allergies	VisitDate
HospitalVisit	[CONDUCTED_LAB_TEST]	LabResult	VisitDate , AnalysisDate , Interpretation
HospitalVisit	[UNDERGOES]	MedicalProcedure	VisitDate , Outcome
MedicalCondition	[HAS_TAGGED]	MedicalConcepts	Ninguna
LabResult	[HAS_TAGGED]	MedicalConcepts	Ninguna
Treatment	[HAS_TAGGED]	MedicalConcepts	Ninguna
Symptoms	[HAS_TAGGED]	MedicalConcepts	Ninguno

## Paso 3: Crear agentes de recuperación de contexto para consultar el gráfico del conocimiento médico

Tras crear la base de datos de gráficos médicos, el siguiente paso es crear agentes para la interacción con los gráficos. Los agentes recuperan el contexto correcto y necesario para la consulta que introduce un médico o un clínico. Existen varias opciones para configurar estos agentes que recuperan el contexto del gráfico de conocimiento:

- [Agentes de Amazon Bedrock](#)
- [LangChain agentes](#)

### Agentes de Amazon Bedrock para la interacción con gráficos

[Los agentes](#) de Amazon Bedrock trabajan a la perfección con las bases de datos de gráficos de Amazon Neptune. Puede realizar interacciones avanzadas a través de los [grupos de acción](#) de Amazon Bedrock. El grupo de acciones inicia el proceso llamando a una AWS Lambda función que ejecuta consultas OpenCypher de Neptune.

Para consultar un gráfico de conocimiento, puede utilizar dos enfoques distintos: la ejecución directa de consultas o la consulta con incrustación de contexto. Estos enfoques se pueden aplicar de forma independiente o combinados, según el caso de uso específico y los criterios de clasificación. Al combinar ambos enfoques, puede proporcionar un contexto más completo al LLM, lo que puede mejorar los resultados. Los dos enfoques de ejecución de consultas son los siguientes:

- Ejecución directa de consultas de Cypher sin incrustaciones: la función Lambda ejecuta consultas directamente en Neptune sin ninguna búsqueda basada en incrustaciones. A continuación se muestra un ejemplo de este enfoque:

```
MATCH (p:Patient)-[u:UNDERGOES]->(h:HospitalVisit) WHERE h.Reason = 'Acute Diabetes'
AND date(u.VisitDate) > date('2024-01-01')
RETURN p.PatientID, p.Name, p.Age, p.Gender, p.Address, p.ContactInformation
```

- Ejecución directa de consultas de Cypher mediante búsqueda incrustada: la función Lambda utiliza la búsqueda incrustada para mejorar los resultados de la consulta. Este enfoque mejora la ejecución de las consultas al incorporar incrustaciones, que son representaciones vectoriales densas de datos. Las incrustaciones son particularmente útiles cuando la consulta requiere una similitud semántica o una comprensión más amplia que vaya más allá de las coincidencias

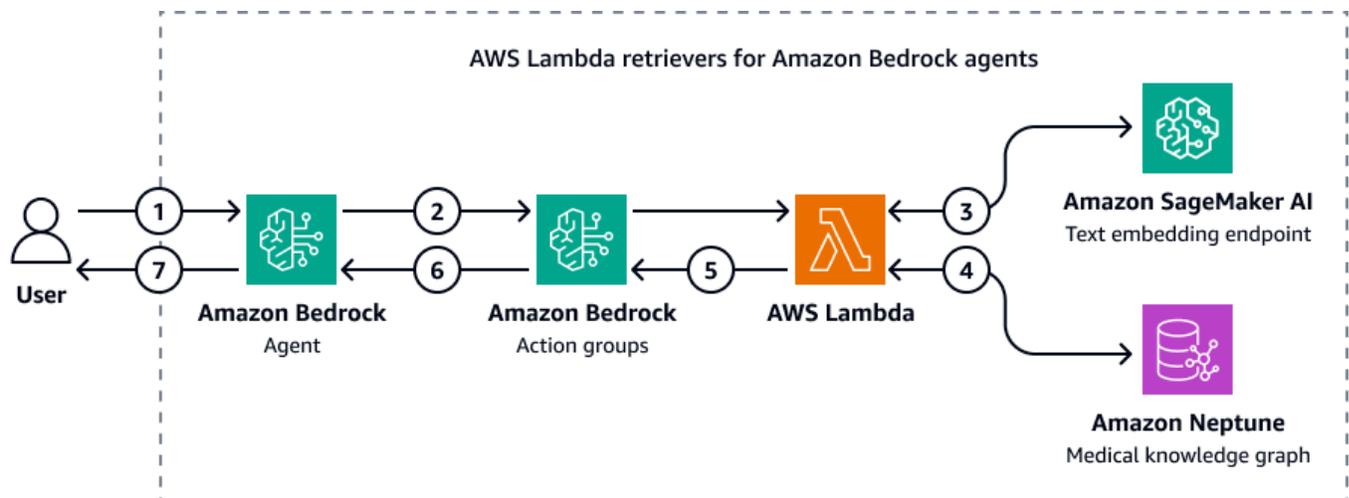
exactas. Puede utilizar modelos previamente entrenados o personalizados para generar integraciones para cada afección médica. El siguiente es un ejemplo de este enfoque:

```
CALL { WITH "Acute Diabetes" AS query_term RETURN search_embedding(query_term) AS
similar_reasons }

MATCH (p:Patient)-[u:UNDERGOES]->(h:HospitalVisit) WHERE h.Reason IN similar reasons
AND date(u.VisitDate) > date('2024-01-01')
RETURN p.PatientID, p.Name, p.Age, p.Gender, p.Address, p.ContactInformation
```

En este ejemplo, la `search_embedding("Acute Diabetes")` función recupera afecciones semánticamente cercanas a la «diabetes aguda». Esto ayuda a la consulta a encontrar también pacientes con afecciones como la prediabetes o el síndrome metabólico.

La siguiente imagen muestra cómo los agentes de Amazon Bedrock interactúan con Amazon Neptune para realizar una consulta cifrada de un gráfico de conocimientos médicos.



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

1. El usuario envía una pregunta al agente de Amazon Bedrock.
2. El agente de Amazon Bedrock pasa la pregunta y las variables del filtro de entrada a los grupos de acción de Amazon Bedrock. Estos grupos de acción contienen una función AWS Lambda que interactúa con el punto final de incrustación de texto de Amazon SageMaker AI y el gráfico de conocimientos médicos de Amazon Neptune.
3. La función Lambda se integra con el punto final de incrustación de texto de SageMaker IA para realizar una búsqueda semántica en la consulta de OpenCypher. Convierte la consulta

en lenguaje natural en una consulta de OpenCypher mediante el uso de la función subyacente LangChain agentes.

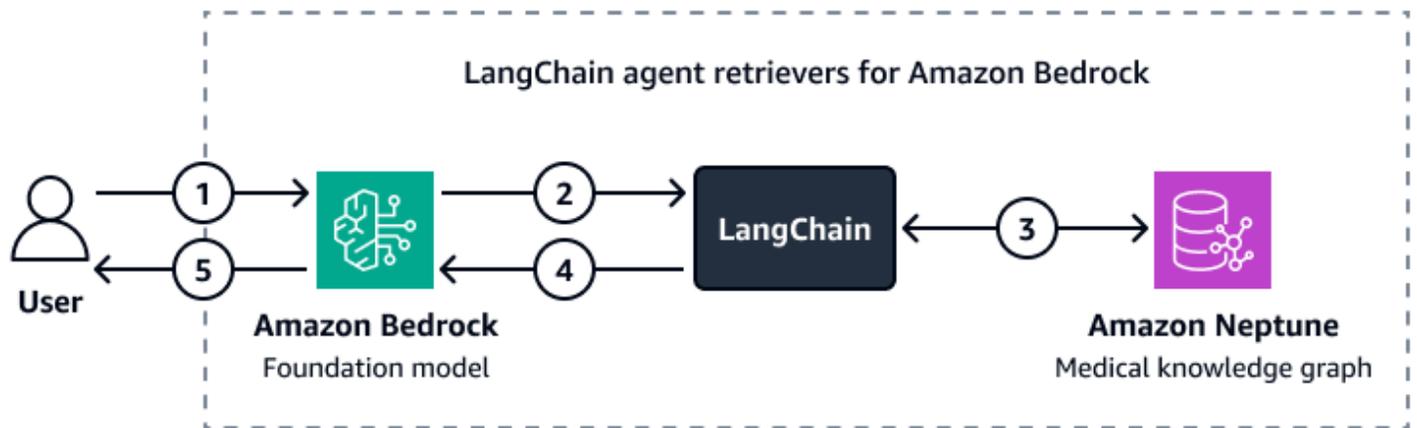
4. La función Lambda consulta el conjunto de datos correcto en el gráfico de conocimiento médico de Neptune y recibe el resultado del gráfico de conocimiento médico de Neptune.
5. La función Lambda devuelve los resultados de Neptune a los grupos de acción de Amazon Bedrock.
6. Los grupos de acción de Amazon Bedrock envían el contexto recuperado al agente de Amazon Bedrock.
7. El agente de Amazon Bedrock genera la respuesta mediante la consulta original del usuario y el contexto recuperado del gráfico de conocimiento.

## LangChain agentes para la interacción de gráficos

Puedes integrar LangChain con Neptune para permitir consultas y recuperaciones basadas en gráficos. Este enfoque puede mejorar los flujos de trabajo impulsados por la IA mediante el uso de las capacidades de la base de datos de gráficos de Neptune. La personalizada LangChain El retriever actúa como intermediario. El modelo fundamental de Amazon Bedrock puede interactuar con Neptune mediante consultas Cypher directas y algoritmos de gráficos más complejos.

Puede utilizar el recuperador personalizado para refinar la forma en que LangChain el agente interactúa con los algoritmos gráficos de Neptune. Por ejemplo, puede utilizar indicaciones de pocas tomas, lo que le ayuda a adaptar las respuestas del modelo base en función de patrones o ejemplos específicos. También puede aplicar filtros identificados por el LLM para afinar el contexto y mejorar la precisión de las respuestas. Esto puede mejorar la eficiencia y la precisión del proceso general de recuperación al interactuar con datos gráficos complejos.

La siguiente imagen muestra cómo se personaliza LangChain El agente organiza la interacción entre un modelo básico de Amazon Bedrock y un gráfico de conocimiento médico de Amazon Neptune.



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

1. Un usuario envía una pregunta a Amazon Bedrock y al LangChain agente.
2. El modelo de base de Amazon Bedrock utiliza el esquema de Neptune, que proporciona el LangChain agente, para generar una consulta para la pregunta del usuario.
3. La LangChain El agente compara la consulta con el gráfico de conocimiento médico de Amazon Neptune.
4. La LangChain el agente envía el contexto recuperado al modelo de base de Amazon Bedrock.
5. El modelo básico de Amazon Bedrock utiliza el contexto recuperado para generar una respuesta a la pregunta del usuario.

## Paso 4: Crear una base de conocimientos de datos descriptivos en tiempo real

A continuación, debe crear una base de conocimientos con notas descriptivas de la interacción médico-paciente en tiempo real, evaluaciones de imágenes de diagnóstico e informes de análisis de laboratorio. Esta base de conocimientos es una base de datos [vectorial](#). Al utilizar una base de datos vectorial, que puede almacenar el conocimiento médico descriptivo de forma indexada y vectorizada, los proveedores de atención médica pueden consultar y acceder de manera eficiente a la información relevante de un amplio repositorio. Estas representaciones vectorizadas ayudan a recuperar datos semánticamente similares. Los proveedores de atención médica pueden consultar rápidamente las notas clínicas, las imágenes médicas y los resultados de laboratorio. Esto acelera la toma de decisiones informadas al ofrecer acceso instantáneo a información relevante desde el punto

de vista del contexto, lo que mejora la precisión y la velocidad de los diagnósticos y los planes de tratamiento.

## Uso de una base de conocimientos médicos OpenSearch del Servicio

[Amazon OpenSearch Service](#) puede gestionar grandes volúmenes de datos médicos de alta dimensión. Es un servicio gestionado que facilita la búsqueda de alto rendimiento y el análisis en tiempo real. Es ideal como base de datos vectorial para aplicaciones RAG. OpenSearch El servicio actúa como una herramienta de back-end para administrar grandes cantidades de datos no estructurados o semiestructurados, como registros médicos, artículos de investigación y notas clínicas. Sus capacidades de búsqueda semántica avanzada le ayudan a recuperar información relevante desde el punto de vista contextual. Esto lo hace particularmente útil en aplicaciones como los sistemas de apoyo a la toma de decisiones clínicas, las herramientas de resolución de consultas de los pacientes y los sistemas de gestión del conocimiento sanitario. Por ejemplo, un médico puede encontrar rápidamente datos relevantes de pacientes o estudios de investigación que coincidan con síntomas o protocolos de tratamiento específicos. Esto ayuda a los médicos a tomar decisiones basadas en la información más relevante up-to-date.

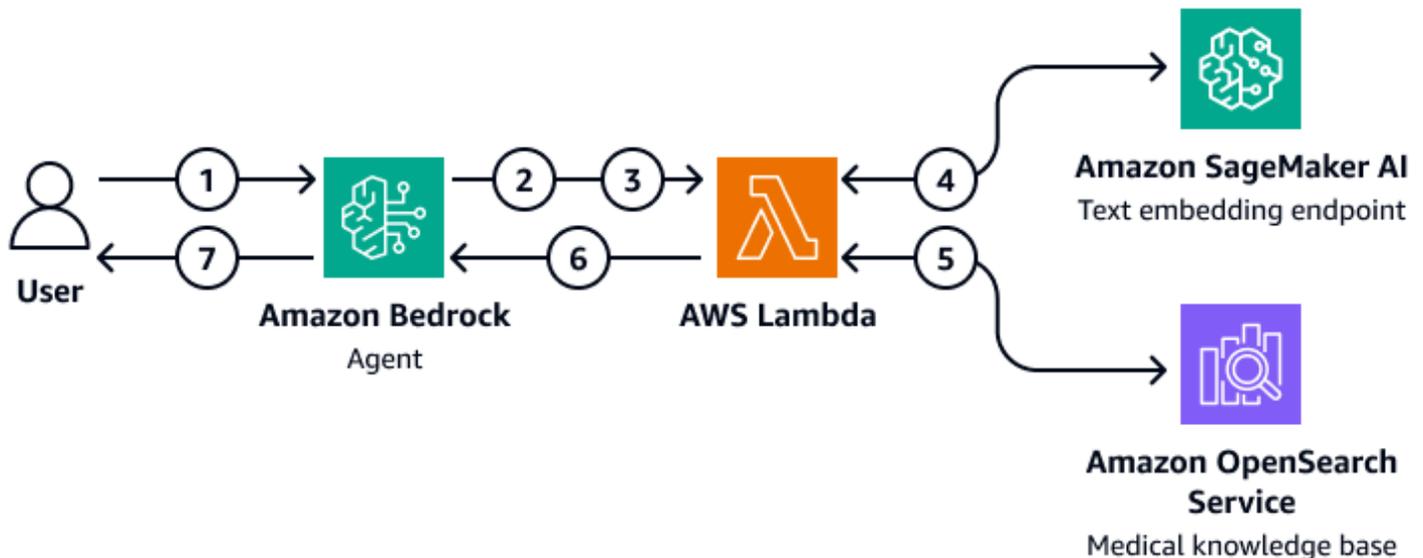
OpenSearch El servicio puede escalar y gestionar la indexación y consulta de datos en tiempo real. Esto lo hace ideal para entornos de atención médica dinámicos donde el acceso oportuno a información precisa es fundamental. Además, cuenta con capacidades de búsqueda multimodales que son óptimas para búsquedas que requieren múltiples entradas, como imágenes médicas y notas médicas. Al implementar las aplicaciones de OpenSearch Service for Health, es fundamental definir campos y mapeos precisos para optimizar la indexación y la recuperación de datos. Los campos representan los datos individuales, como los registros de los pacientes, los historiales médicos y los códigos de diagnóstico. Los mapeos definen cómo se almacenan estos campos (en forma incrustada o en forma original) y cómo se consultan. Para las aplicaciones sanitarias, es esencial establecer mapeos que se adapten a varios tipos de datos, incluidos los datos estructurados (como los resultados de las pruebas numéricas), los datos semiestructurados (como las notas de los pacientes) y los datos no estructurados (como las imágenes médicas)

En OpenSearch Service, puede realizar [búsquedas neuronales](#) de texto completo mediante instrucciones seleccionadas para buscar en historias clínicas, notas clínicas o trabajos de investigación y encontrar rápidamente información relevante sobre síntomas, tratamientos o historias clínicas específicas de pacientes. Las consultas de búsqueda neuronal gestionan automáticamente la incrustación del mensaje de entrada y las imágenes mediante modelos de redes neuronales integrados. Esto le ayuda a comprender y captar las relaciones semánticas más profundas de los datos multimodales, lo que ofrece resultados de búsqueda más precisos y sensibles al contexto en

comparación con otros algoritmos de consulta de búsqueda, como la búsqueda K-near Neighbor (k-NN).

## Crear una arquitectura RAG

Puede implementar una solución RAG personalizada que utilice los agentes de Amazon Bedrock para consultar una base de conocimientos médicos en OpenSearch Service. Para ello, debe crear una AWS Lambda función que pueda interactuar con el Servicio y OpenSearch consultarlo. La función Lambda incrusta la pregunta introducida por el usuario accediendo a un punto final de incrustación de texto de SageMaker IA. El agente de Amazon Bedrock transfiere parámetros de consulta adicionales como entradas a la función Lambda. La función consulta la base de conocimientos médicos de OpenSearch Service, que devuelve el contenido médico relevante. Tras configurar la función Lambda, agréguela como grupo de acciones en el agente de Amazon Bedrock. El agente de Amazon Bedrock toma la entrada del usuario, identifica las variables necesarias, pasa las variables y la pregunta a la función Lambda y, a continuación, inicia la función. La función devuelve un contexto que ayuda al modelo básico a proporcionar una respuesta más precisa a la pregunta del usuario.

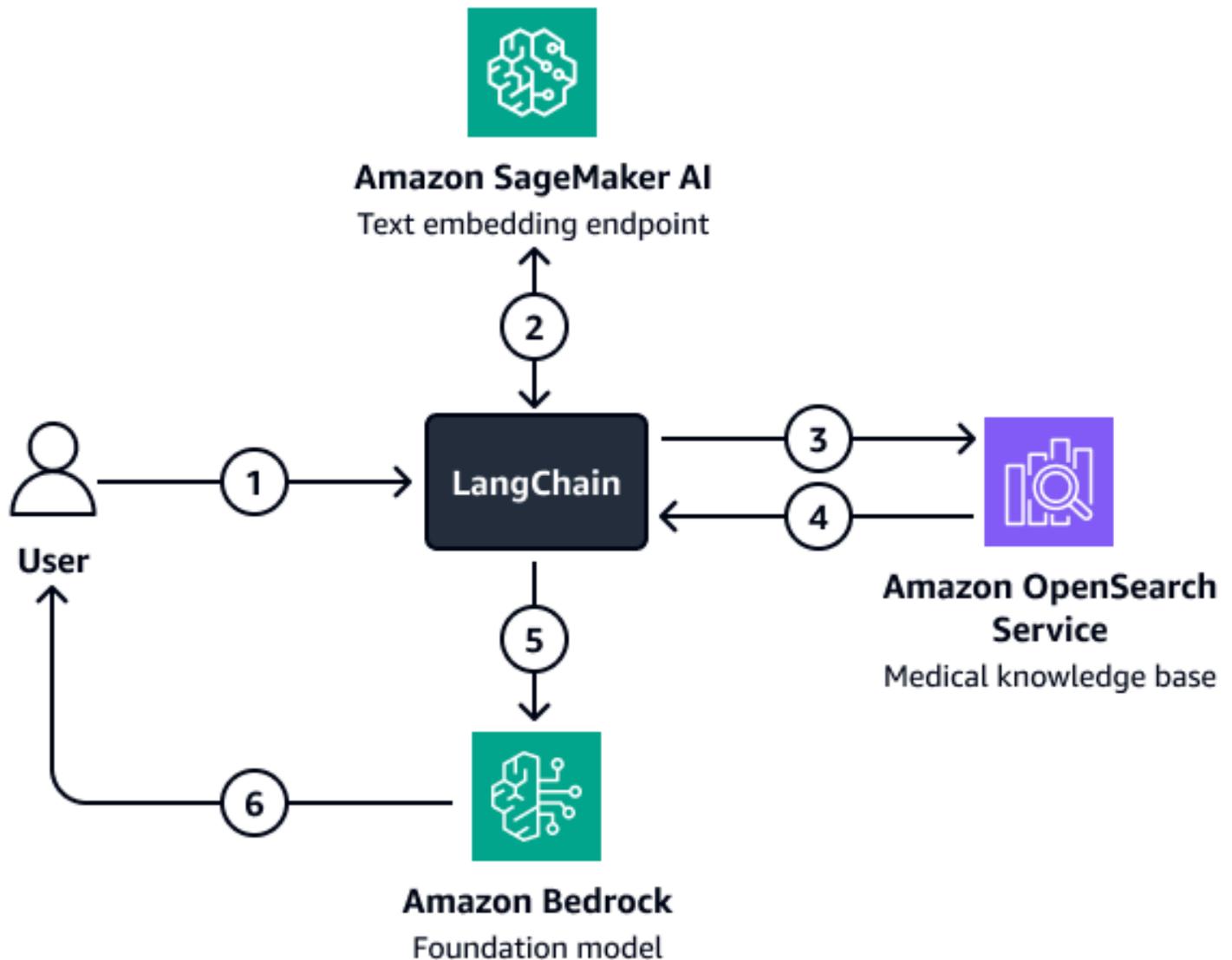


En el diagrama, se muestra el siguiente flujo de trabajo:

1. Un usuario envía una pregunta al agente de Amazon Bedrock.
2. El agente de Amazon Bedrock selecciona qué grupo de acción iniciar.
3. El agente de Amazon Bedrock inicia una AWS Lambda función y le pasa parámetros.

4. La función Lambda inicia el modelo de incrustación de texto de Amazon SageMaker AI para incrustar la pregunta del usuario.
5. La función Lambda transfiere el texto incrustado y los parámetros y filtros adicionales a Amazon OpenSearch Service. Amazon OpenSearch Service consulta la base de conocimientos médicos y devuelve los resultados a la función Lambda.
6. La función Lambda devuelve los resultados al agente de Amazon Bedrock.
7. El modelo básico del agente Amazon Bedrock genera una respuesta basada en los resultados y la devuelve al usuario.

Para situaciones en las que se trate de un filtrado más complejo, puede utilizar un filtro personalizado LangChain recuperador. Cree este recuperador configurando un cliente de búsqueda vectorial de OpenSearch Service que se cargue directamente en LangChain. Esta arquitectura le permite pasar más variables para crear los parámetros del filtro. Una vez configurado el retriever, utilice el modelo Amazon Bedrock y el retriever para configurar una cadena de preguntas y respuestas de recuperación. Esta cadena organiza la interacción entre el modelo y el recuperador al pasar la entrada del usuario y los posibles filtros al recuperador. El recuperador devuelve el contexto relevante que ayuda al modelo básico a responder a la pregunta del usuario.



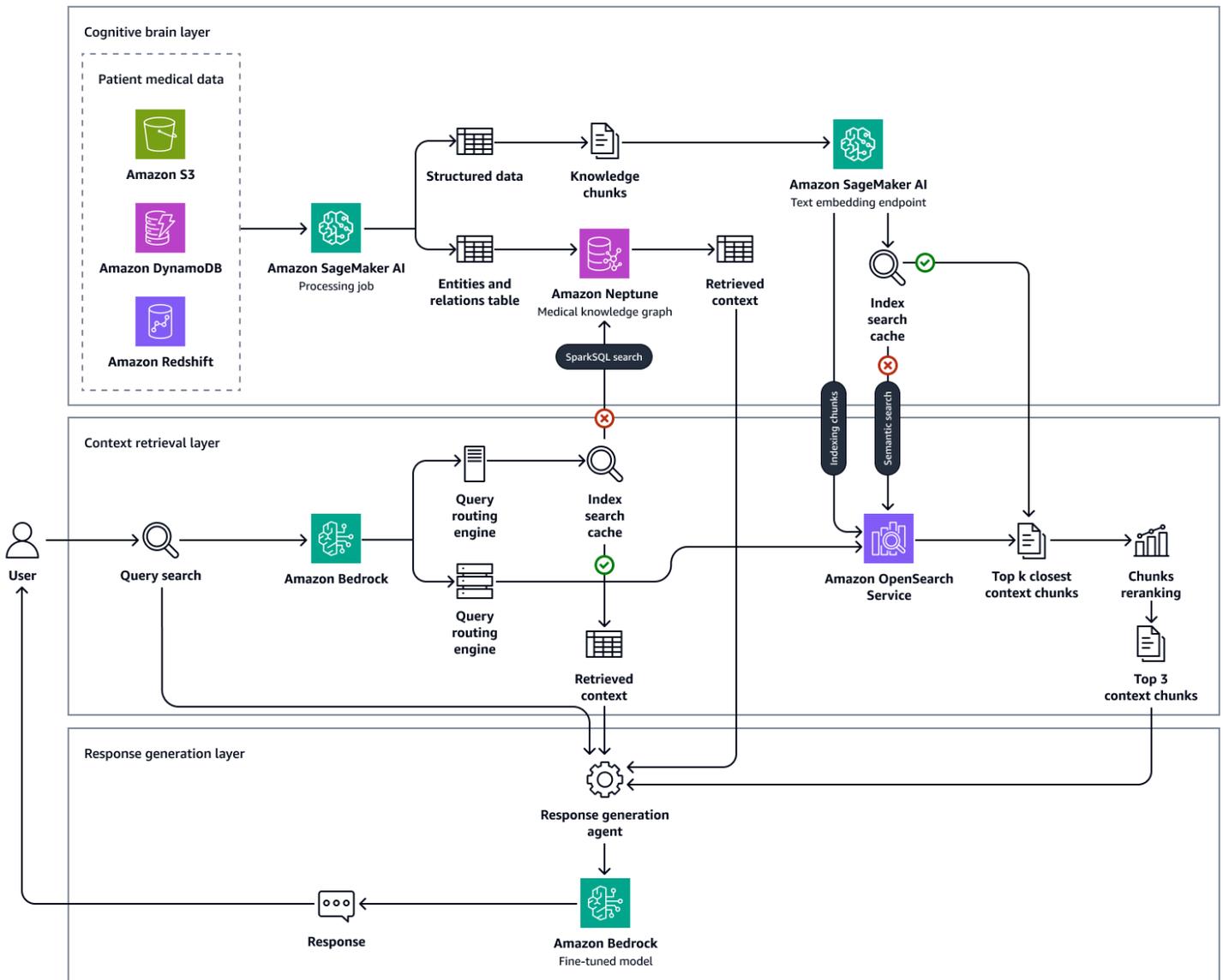
En el diagrama, se muestra el siguiente flujo de trabajo:

1. Un usuario envía una pregunta al LangChain agente recuperador.
2. La LangChain El agente de recuperación envía la pregunta al punto final de incrustación de texto de Amazon SageMaker AI para incrustarla.
3. La LangChain El agente de recuperación pasa el texto incrustado a Amazon OpenSearch Service.
4. Amazon OpenSearch Service devuelve los documentos recuperados al LangChain agente recuperador.
5. La LangChain El agente retriever pasa la pregunta del usuario y el contexto recuperado al modelo básico de Amazon Bedrock.
6. El modelo básico genera una respuesta y la envía al usuario.

## Paso 5: LLMs Utilízalo para responder a preguntas médicas

Los pasos anteriores le ayudan a crear una aplicación de inteligencia médica que pueda recuperar el historial médico del paciente y resumir los medicamentos relevantes y los posibles diagnósticos. Ahora, crea la capa de generación. Esta capa utiliza las capacidades generativas de un LLM en Amazon Bedrock, como Llama 3, para aumentar la producción de la aplicación.

Cuando un médico introduce una consulta, la capa de recuperación de contexto de la aplicación lleva a cabo el proceso de recuperación a partir del gráfico de conocimiento y devuelve los registros principales relacionados con el historial, la demografía, los síntomas, el diagnóstico y los resultados del paciente. De la base de datos vectorial, también recupera notas descriptivas y en tiempo real sobre las interacciones entre el médico y el paciente, información sobre la evaluación de las imágenes diagnósticas, resúmenes de los informes de análisis de laboratorio e información procedente de una enorme cantidad de libros académicos y de investigación médica. Estos resultados más recuperados, la consulta del médico y las indicaciones (que están diseñadas para seleccionar las respuestas en función de la naturaleza de la consulta) se pasan luego al modelo básico de Amazon Bedrock. Esta es la capa de generación de respuestas. El LLM utiliza el contexto recuperado para generar una respuesta a la consulta del médico. La siguiente figura muestra el end-to-end flujo de trabajo de los pasos de esta solución.



Puede utilizar un modelo básico previamente entrenado en Amazon Bedrock, como Llama 3, para una variedad de casos de uso que debe gestionar la aplicación de inteligencia médica. El LLM más eficaz para una tarea determinada varía según el caso de uso. Por ejemplo, un modelo previamente entrenado podría ser suficiente para resumir las conversaciones entre el paciente y el médico, buscar entre los medicamentos y las historias clínicas de los pacientes y obtener información a partir de conjuntos de datos médicos internos y conjuntos de conocimientos científicos. Sin embargo, podría ser necesario un LLM perfeccionado para otros casos de uso complejos, como las evaluaciones de laboratorio en tiempo real, las recomendaciones de procedimientos médicos y las predicciones de los resultados de los pacientes. Puede afinar un LLM entrenándolo en conjuntos de datos del dominio médico. Los requisitos específicos o complejos de la salud y las ciencias de la vida impulsan el desarrollo de estos modelos ajustados.

Para obtener más información sobre cómo ajustar un LLM o elegir un LLM existente que se haya formado con datos del dominio médico, consulte [Uso de modelos lingüísticos de gran tamaño para casos de uso de la salud](#) y las ciencias de la vida.

## Alineación con el marco de AWS Well-Architected

La solución se alinea con los seis pilares del [AWS Well-Architected](#) Framework de la siguiente manera:

- **Excelencia operativa:** la arquitectura está desacoplada para una supervisión y actualizaciones eficientes. Amazon Bedrock tiene agentes y le AWS Lambda ayuda a implementar y deshacer herramientas rápidamente.
- **Seguridad:** esta solución está diseñada para cumplir con las normas sanitarias, como la HIPAA. También puede implementar el cifrado, un control de acceso detallado y las barandillas Amazon Bedrock para ayudar a proteger los datos de los pacientes.
- **Fiabilidad:** los servicios AWS gestionados, como Amazon OpenSearch Service y Amazon Bedrock, proporcionan la infraestructura necesaria para la interacción continua con los modelos.
- **Eficiencia de rendimiento:** la solución RAG recupera los datos relevantes rápidamente mediante búsquedas semánticas optimizadas y consultas cifradas, mientras que un agente enrutador identifica las rutas óptimas para las consultas de los usuarios.
- **Optimización de costes:** el pay-per-token modelo de Amazon Bedrock y la arquitectura RAG reducen los costes de inferencia y formación previa.
- **Sostenibilidad:** el uso de infraestructura y pay-per-token computación sin servidores minimiza el uso de recursos y mejora la sostenibilidad.

# Caso de uso: predicción de los resultados de los pacientes y las tasas de reingreso

Los análisis predictivos basados en la IA ofrecen beneficios adicionales al pronosticar los resultados de los pacientes y permitir planes de tratamiento personalizados. Esto puede mejorar la satisfacción de los pacientes y los resultados de salud. Al integrar estas capacidades de IA con Amazon Bedrock y otras tecnologías, los proveedores de atención médica pueden lograr importantes aumentos de productividad, reducir los costos y mejorar la calidad general de la atención al paciente.

Puede almacenar datos médicos, como los historiales de los pacientes, las notas clínicas, los medicamentos y los tratamientos, en un gráfico de [conocimiento](#). Al combinar la comprensión profunda del contexto LLMs con los datos temporales estructurados de un gráfico de conocimiento médico, los proveedores de servicios de salud pueden obtener información adicional sobre los patrones individuales de los pacientes. Mediante el análisis predictivo, puede identificar de forma temprana la posible falta de adherencia o las complicaciones del tratamiento y generar puntuaciones personalizadas de propensión al reingreso.

Esta solución le ayuda a predecir la probabilidad de volver a ingresar. Estas predicciones pueden mejorar los resultados de los pacientes y reducir los costos de atención médica. Esta solución también puede ayudar a los médicos y administradores de los hospitales a centrar su atención en los pacientes con un mayor riesgo de reingreso. También les ayuda a iniciar intervenciones proactivas con esos pacientes mediante alertas, autoservicio y acciones basadas en datos.

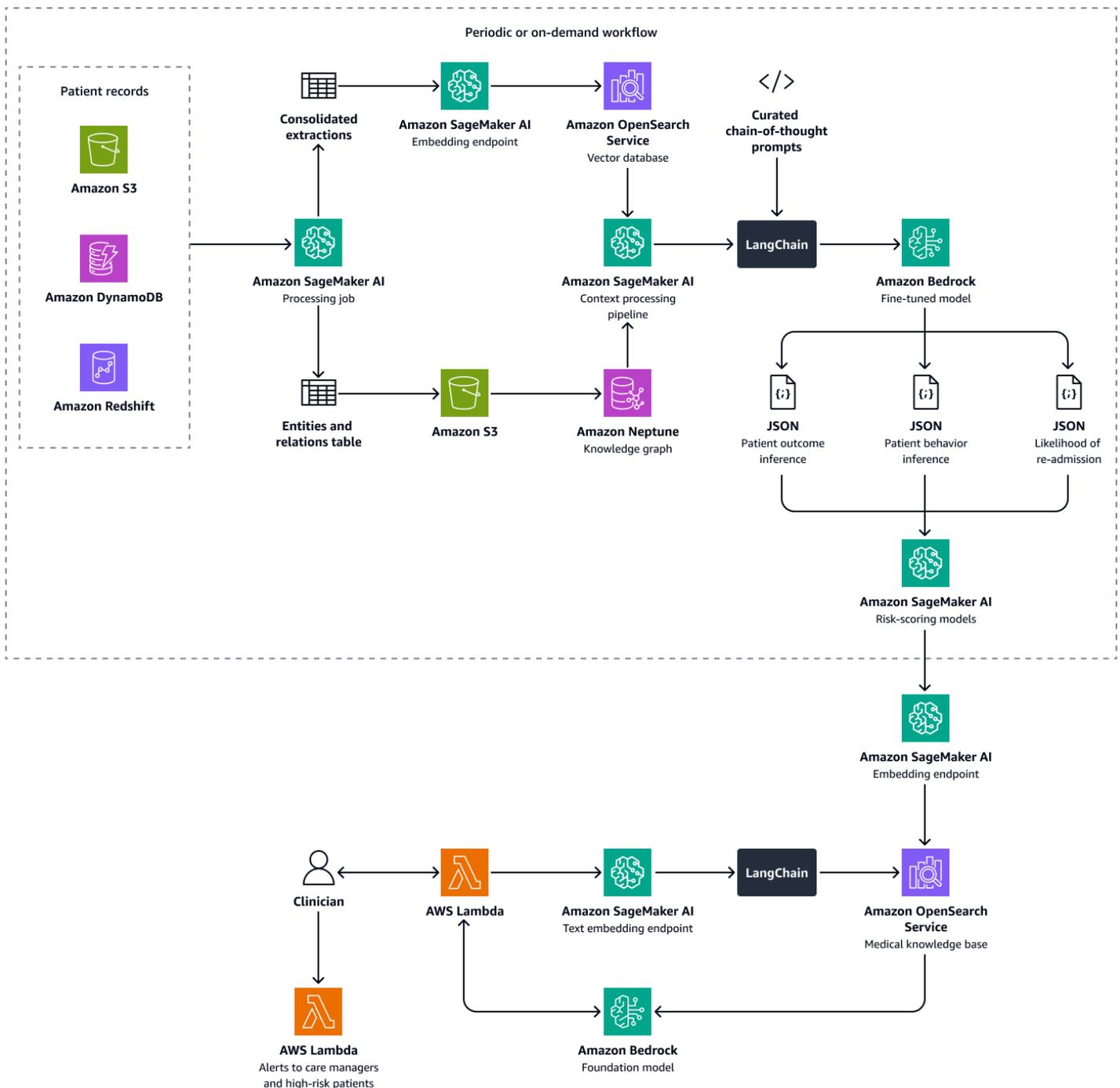
## Información general de la solución

Esta solución utiliza un marco de generación aumentada (RAG) de recuperación múltiple para analizar los datos de los pacientes. Predice la probabilidad de reingreso hospitalario de pacientes individuales y le ayuda a calcular una puntuación de propensión al reingreso a nivel hospitalario. Esta solución integra las siguientes funciones:

- Gráfico de conocimiento: almacena datos estructurados y cronológicos de los pacientes, como las visitas al hospital, los reingresos anteriores, los síntomas, los resultados de laboratorio, los tratamientos prescritos y el historial de adherencia a la medicación
- Base de datos vectorial: almacena datos clínicos no estructurados, como resúmenes de alta, notas del médico y registros de consultas no atendidas o de efectos secundarios notificados por los medicamentos

- LLM perfeccionado: utiliza tanto datos estructurados del gráfico de conocimiento como datos no estructurados de la base de datos vectorial para generar inferencias sobre el comportamiento del paciente, la adherencia al tratamiento y la probabilidad de reingreso

Los modelos de puntuación de riesgo cuantifican las inferencias del LLM en puntuaciones numéricas. Puede sumar las puntuaciones en una puntuación de propensión a la readmisión a nivel hospitalario. Esta puntuación define la exposición al riesgo de cada paciente y puede calcularla periódicamente o según sea necesario. Todas las inferencias y puntuaciones de riesgo se indexan y almacenan en Amazon OpenSearch Service para que los administradores de atención y los médicos puedan recuperarlas. Al integrar un agente de IA conversacional en esta base de datos vectorial, los médicos y los administradores de atención pueden extraer información sin problemas a nivel de paciente individual, de todo el centro o por especialidad médica. También puede configurar alertas automatizadas basadas en las puntuaciones de riesgo, lo que fomenta las intervenciones proactivas.



La creación de esta solución consta de los siguientes pasos:

- [Paso 1: Predecir los resultados de los pacientes mediante un gráfico de conocimientos médicos](#)
- [Paso 2: Predecir el comportamiento del paciente con respecto a los medicamentos o tratamientos recetados](#)
- [Paso 3: Predecir la probabilidad de reingreso del paciente](#)

- [Paso 4: Calcular la puntuación de propensión al reingreso hospitalario](#)

## Paso 1: Predecir los resultados de los pacientes mediante un gráfico de conocimientos médicos

En [Amazon Neptune](#), puede usar un gráfico de conocimiento para almacenar información temporal sobre las visitas de los pacientes y los resultados a lo largo del tiempo. La forma más eficaz de crear y almacenar un gráfico de conocimiento es utilizar un modelo gráfico y una base de datos de gráficos. Las bases de datos de gráficos están diseñadas específicamente para almacenar y navegar por las relaciones. Las bases de datos de gráficos facilitan el modelado y la administración de datos altamente conectados y cuentan con esquemas flexibles.

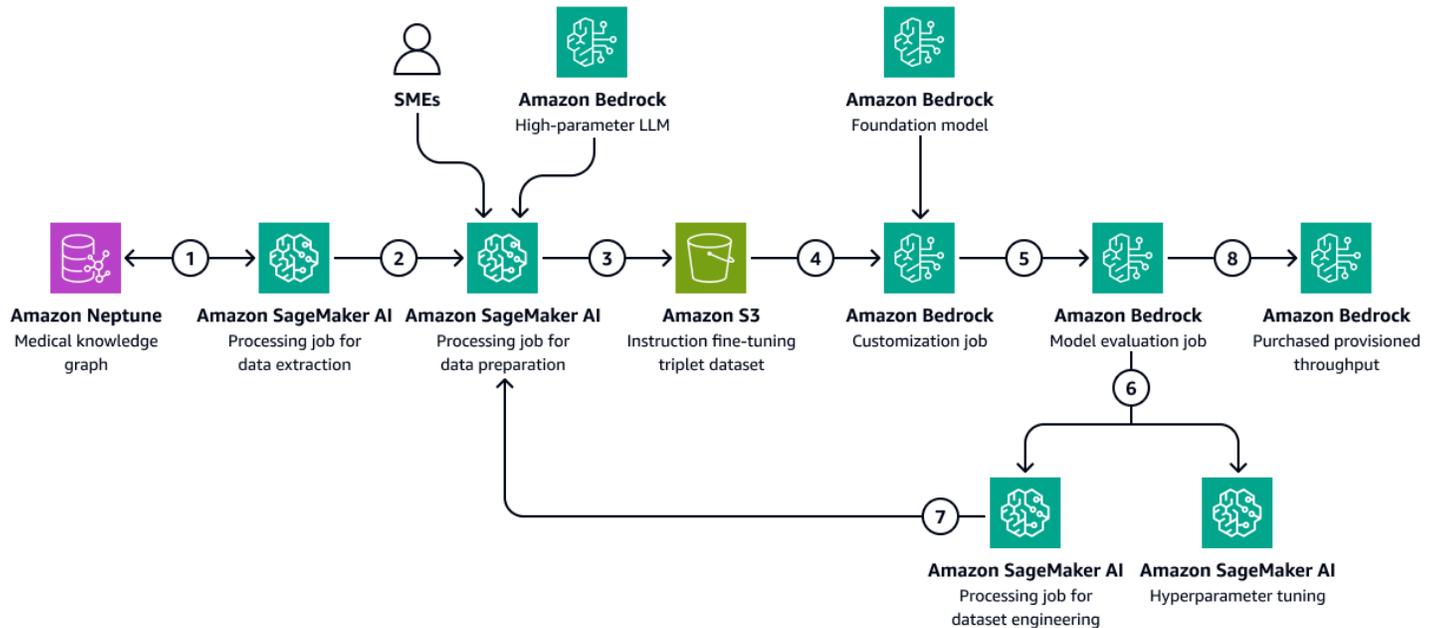
El gráfico de conocimiento le ayuda a realizar análisis de series temporales. Los siguientes son los elementos clave de la base de datos de gráficos que se utilizan para la predicción temporal de los resultados de los pacientes:

- Datos históricos: diagnósticos previos, medicación continuada, medicamentos utilizados anteriormente y resultados de laboratorio del paciente
- Visitas de los pacientes (cronológicas): fechas de las visitas, síntomas, alergias observadas, notas clínicas, diagnósticos, procedimientos, tratamientos, medicamentos recetados y resultados de laboratorio
- Síntomas y parámetros clínicos: información clínica y basada en los síntomas, incluida la gravedad, los patrones de progresión y la respuesta del paciente a los medicamentos

Puede utilizar la información del gráfico de conocimientos médicos para afinar un máster en Amazon Bedrock, como Llama 3. El LLM se ajusta con precisión con datos secuenciales del paciente sobre la respuesta del paciente a un conjunto de medicamentos o tratamientos a lo largo del tiempo. Utilice un conjunto de datos etiquetado que clasifique un conjunto de medicamentos o tratamientos y los datos de interacción entre el paciente y la clínica en categorías predefinidas que indiquen el estado de salud del paciente. Algunos ejemplos de estas categorías son el deterioro de la salud, la mejora o el progreso estable. Cuando el médico introduce un nuevo contexto sobre el paciente y sus síntomas, el LLM, bien ajustado, puede utilizar los patrones del conjunto de datos de entrenamiento para predecir el posible resultado del paciente.

La siguiente imagen muestra los pasos secuenciales necesarios para afinar un LLM en Amazon Bedrock mediante un conjunto de datos de formación específico para el sector sanitario. Estos datos

pueden incluir las afecciones médicas de los pacientes y las respuestas a los tratamientos a lo largo del tiempo. Este conjunto de datos de entrenamiento ayudaría al modelo a realizar predicciones generalizadas sobre los resultados de los pacientes.



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

1. El trabajo de extracción de datos de Amazon SageMaker AI consulta el gráfico de conocimiento para recuperar datos cronológicos sobre las respuestas de los diferentes pacientes a un conjunto de medicamentos o tratamientos a lo largo del tiempo.
2. El trabajo de preparación de datos de SageMaker IA integra un LLM de Amazon Bedrock y aportaciones de expertos en la materia (SMEs). El trabajo clasifica los datos recuperados del gráfico de conocimiento en categorías predefinidas (como el deterioro de la salud, la mejora o el progreso estable) que indican el estado de salud de cada paciente.
3. El trabajo crea un conjunto de datos preciso que incluye la información extraída del gráfico de conocimiento, las chain-of-thought indicaciones y la categoría de resultados de los pacientes. Carga este conjunto de datos de entrenamiento en un bucket de Amazon S3.
4. Un trabajo de personalización de Amazon Bedrock utiliza este conjunto de datos de entrenamiento para ajustar un LLM.
5. El trabajo de personalización de Amazon Bedrock integra el modelo fundamental de Amazon Bedrock preferido en el entorno de formación. Comienza el trabajo de ajuste y utiliza el conjunto de datos de entrenamiento y los hiperparámetros de entrenamiento que usted configure.

6. Un trabajo de evaluación de Amazon Bedrock evalúa el modelo ajustado mediante un marco de evaluación de modelos prediseñado.
7. Si es necesario mejorar el modelo, el trabajo de formación se vuelve a ejecutar con más datos tras una cuidadosa consideración del conjunto de datos de formación. Si el modelo no demuestra una mejora gradual del rendimiento, considere también la posibilidad de modificar los hiperparámetros del entrenamiento.
8. Una vez que la evaluación del modelo cumpla con los estándares definidos por las partes interesadas de la empresa, debe alojar el modelo ajustado según el rendimiento aprovisionado de Amazon Bedrock.

## Paso 2: Predecir el comportamiento del paciente con respecto a los medicamentos o tratamientos recetados

Tune-Tuned LLMs puede procesar notas clínicas, resúmenes de alta y otros documentos específicos del paciente a partir del gráfico temporal de conocimientos médicos. Pueden evaluar si es probable que el paciente siga los medicamentos o tratamientos recetados.

Este paso utiliza el gráfico de conocimiento creado en [Paso 1: Predecir los resultados de los pacientes mediante un gráfico de conocimientos médicos](#). El gráfico de conocimiento contiene datos del perfil del paciente, incluido el historial de adherencia del paciente como nodo. También incluye los casos de falta de adherencia a los medicamentos o tratamientos, los efectos secundarios de los medicamentos, la falta de acceso a los medicamentos o los obstáculos a su costo, o los regímenes de dosificación complejos, como atributos de dichos nodos.

Fine-Tuned LLMs puede consumir datos de entrega de recetas anteriores del gráfico de conocimientos médicos y resúmenes descriptivos de las notas clínicas de una base de datos vectorial de Amazon OpenSearch Service. Estas notas clínicas pueden mencionar la falta frecuente de citas o el incumplimiento de los tratamientos. El LLM puede usar estas notas para predecir la probabilidad de una futura falta de adhesión.

1. Prepare los datos de entrada de la siguiente manera:
  - Datos estructurados: extraiga los datos recientes de los pacientes, como las tres últimas visitas y los resultados del laboratorio, del gráfico de conocimientos médicos.
  - Datos no estructurados: recupera las notas clínicas recientes de la base de datos vectorial OpenSearch de Amazon Service.

2. Cree un mensaje de entrada que incluya el historial del paciente y el contexto actual. El siguiente es un ejemplo de mensaje:

```
You are a highly specialized AI model trained in healthcare predictive analytics.
Your task is to analyze a patient's historical medical records, adherence patterns,
and clinical context to predict the likelihood of future non-adherence to
prescribed medications or treatments.
```

```
### Patient Details
```

- **Patient ID:** {patient\_id}
- **Age:** {age}
- **Gender:** {gender}
- **Medical Conditions:** {medical\_conditions}
- **Current Medications:** {current\_medications}
- **Prescribed Treatments:** {prescribed\_treatments}

```
### Chronological Medical History
```

- **Visit Dates & Symptoms:** {visit\_dates\_symptoms}
- **Diagnoses & Procedures:** {diagnoses\_procedures}
- **Prescribed Medications & Treatments:** {medications\_treatments}
- **Past Adherence Patterns:** {historical\_adherence}
- **Instances of Non-Adherence:** {past\_non\_adherence}
- **Side Effects Experienced:** {side\_effects}
- **Barriers to Adherence (e.g., Cost, Access, Dosing Complexity):** {barriers}

```
### Patient-Specific Insights
```

- **Clinical Notes & Discharge Summaries:** {clinical\_notes}
- **Missed Appointments & Non-Compliance Patterns:** {missed\_appointments}

```
### Let's think Step-by-Step to predict the patient behaviour
```

1. You should first analyze past adherence trends and patterns of non-adherence.
2. Identify potential barriers, such as financial constraints, medication side effects, or complex dosing regimens.
3. Thoroughly examine clinical notes and documented patient behaviors that may hint at non-adherence.
4. Correlate adherence history with prescribed treatments and patient conditions.
5. Finally predict the likelihood of non-adherence based on these contextual insights.

```
### Output Format (JSON)
```

```
Return the prediction in the following structured format:
```

```
```json
{
```

```
"patient_id": "{patient_id}",  
"likelihood_of_non_adherence": "{low | moderate | high}",  
"reasoning": "{detailed_explanation_based_on_patient_history}"  
}
```

3. Pase el mensaje al LLM ajustado con precisión. El LLM procesa el mensaje y predice el resultado. El siguiente es un ejemplo de respuesta del LLM:

```
{  
  "patient_id": "P12345",  
  "likelihood_of_non_adherence": "high",  
  "reasoning": "The patient has a history of missed appointments, has reported side effects to previous medications. Additionally, clinical notes indicate difficulty following complex dosing schedules."  
}
```

4. Analice la respuesta del modelo para extraer la categoría de resultado prevista. Por ejemplo, la categoría de la respuesta de ejemplo del paso anterior podría ser una alta probabilidad de falta de adherencia.
5. (Opcional) Utilice registros modelo o métodos adicionales para asignar puntuaciones de confianza. Los logits son las probabilidades no normalizadas de que el elemento pertenezca a una determinada clase o categoría.

## Paso 3: Predecir la probabilidad de reingreso del paciente

Los reingresos hospitalarios son motivo de gran preocupación debido al alto coste de la administración de la asistencia sanitaria y a su impacto en el bienestar del paciente. Calcular las tasas de reingreso hospitalario es una forma de medir la calidad de la atención al paciente y el desempeño de un proveedor de atención médica.

Para calcular la tasa de reingresos, ha definido un indicador, como una tasa de reingresos a 7 días. Este indicador es el porcentaje de pacientes ingresados que regresan al hospital para una visita no planificada dentro de los siete días posteriores al alta. Para predecir la probabilidad de reingreso de un paciente, un LLM ajustado puede consumir datos temporales del gráfico de conocimientos médicos en el que se creó. [Paso 1: Predecir los resultados de los pacientes mediante un gráfico de conocimientos médicos](#) Este gráfico de conocimiento mantiene registros cronológicos de los encuentros con los pacientes, los procedimientos, los medicamentos y los síntomas. Estos registros de datos contienen lo siguiente:

- Tiempo transcurrido desde la última vez que el paciente fue dado de alta
- La respuesta del paciente a los tratamientos y medicamentos anteriores
- La progresión de los síntomas o afecciones a lo largo del tiempo

Puede procesar estas series temporales de eventos para predecir la probabilidad de reingreso de un paciente mediante un sistema seleccionado. El indicador imparte la lógica de predicción al LLM ajustado con precisión.

1. Prepare los datos de entrada de la siguiente manera:

- Historial de adherencia: extraiga las fechas de recogida de los medicamentos, las frecuencias de reabastecimiento de los medicamentos, los detalles del diagnóstico y del medicamento, el historial médico cronológico y otra información del gráfico de conocimientos médicos.
- Indicadores de comportamiento: recupere e incluya notas clínicas sobre las consultas faltantes y los efectos secundarios informados por los pacientes.

2. Cree un mensaje de entrada que incluya el historial de adherencia y los indicadores de comportamiento. El siguiente es un ejemplo de mensaje:

```
You are a highly specialized AI model trained in healthcare predictive analytics.
Your task is to analyze a patient's historical medical records, clinical events, and
adherence patterns to predict the likelihood of hospital readmission within the
next few days.
```

```
### Patient Details
```

- ```
- Patient ID: {patient_id}
- Age: {age}
- Gender: {gender}
- Primary Diagnoses: {diagnoses}
- Current Medications: {current_medications}
- Prescribed Treatments: {prescribed_treatments}
```

```
### Chronological Medical History
```

- ```
- Recent Hospital Encounters: {encounters}
- Time Since Last Discharge: {time_since_last_discharge}
- Previous Readmissions: {past_readmissions}
- Recent Lab Results & Vital Signs: {recent_lab_results}
- Procedures Performed: {procedures_performed}
- Prescribed Medications & Treatments: {medications_treatments}
- Past Adherence Patterns: {historical_adherence}
- Instances of Non-Adherence: {past_non_adherence}
```

```

### **Patient-Specific Insights**
- **Clinical Notes & Discharge Summaries:** {clinical_notes}
- **Missed Appointments & Non-Compliance Patterns:** {missed_appointments}
- **Patient-Reported Side Effects & Complications:** {side_effects}

### **Reasoning Process - You have to analyze this use case step-by-step.**
1. First assess **time since last discharge** and whether recent hospital encounters suggest a pattern of frequent readmissions.
2. Second examine **recent lab results, vital signs, and procedures performed** to identify clinical deterioration.
3. Third analyze **adherence history**, checking if past non-adherence to medications or treatments correlates with readmissions.
4. Then identify **missed appointments, self-reported side effects, or symptoms worsening** from clinical notes.
5. Finally predict the **likelihood of readmission** based on these contextual insights.

### **Output Format (JSON)**
Return the prediction in the following structured format:
```json
{
  "patient_id": "{patient_id}",
  "likelihood_of_readmission": "{low | moderate | high}",
  "reasoning": "{detailed_explanation_based_on_patient_history}"
}

```

3. Pase el mensaje al LLM ajustado con precisión. El LLM procesa el aviso y predice la probabilidad y los motivos de la readmisión. El siguiente es un ejemplo de respuesta del LLM:

```

{
  "patient_id": "P67890",
  "likelihood_of_readmission": "high",
  "reasoning": "The patient was discharged only 5 days ago, has a history of more than two readmissions to hospitals where the patient received treatment. Recent lab results indicate abnormal kidney function and high liver enzymes. These factors suggest a medium risk of readmission."
}

```

4. Clasifique la predicción en una escala estandarizada, como baja, media o alta.
5. Revise el razonamiento proporcionado por el LLM e identifique los factores clave que contribuyen a la predicción.

6. Asigne los resultados cualitativos a las puntuaciones cuantitativas. Por ejemplo, una probabilidad muy alta podría ser igual a 0,9.
7. Utilice conjuntos de datos de validación para calibrar los resultados del modelo con respecto a las tasas reales de readmisión.

## Paso 4: Calcular la puntuación de propensión al reingreso hospitalario

A continuación, se calcula la puntuación de propensión al reingreso hospitalario por paciente. Esta puntuación refleja el impacto neto de los tres análisis realizados en los pasos anteriores: los posibles resultados de los pacientes, el comportamiento de los pacientes con respecto a los medicamentos y los tratamientos y la probabilidad de reingreso del paciente. Al sumar la puntuación de propensión al reingreso a nivel del paciente al nivel de la especialidad y, luego, al nivel hospitalario, puede obtener información para los médicos, los directores de atención y los administradores. La puntuación de propensión al reingreso hospitalario le ayuda a evaluar el desempeño general por centro, especialidad o afección. Luego, puede usar esta puntuación para implementar intervenciones proactivas.

1. Asigne ponderaciones a cada uno de los diferentes factores (predicción del resultado, probabilidad de adherencia, readmisión). Los siguientes son ejemplos de pesos:
  - Peso de predicción del resultado: 0,4
  - Peso de predicción de adherencia: 0,3
  - Peso de la probabilidad de readmisión: 0,3
2. Utilice el siguiente cálculo para calcular la puntuación compuesta:

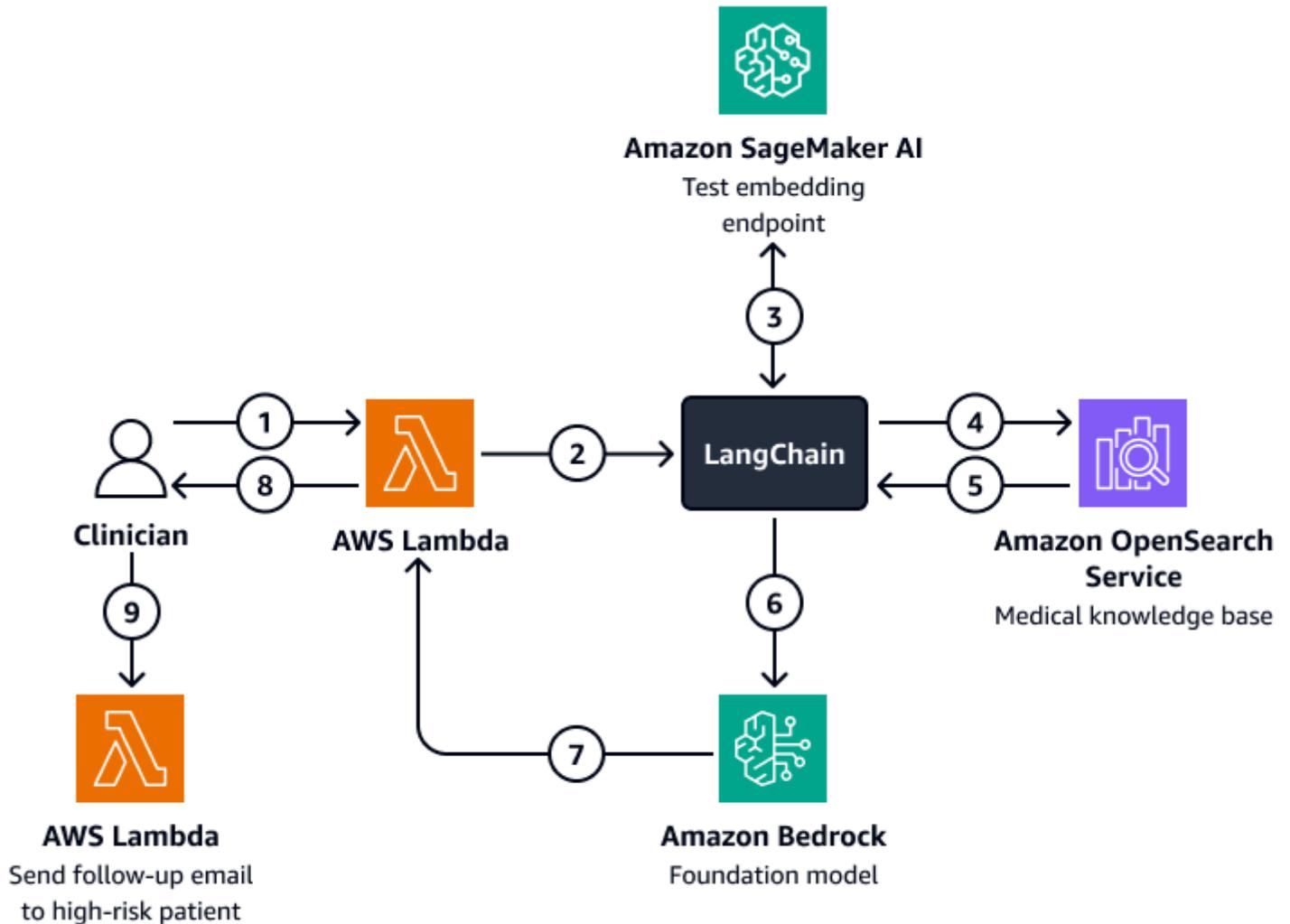
$$\text{ReadmissionPropensityScore} = (\text{OutcomeScore} \times \text{OutcomeWeight}) + (\text{AdherenceScore} \times \text{AdherenceWeight}) + (\text{ReadmissionLikelihoodScore} \times \text{ReadmissionLikelihoodWeight})$$

3. Asegúrese de que todas las puntuaciones individuales estén en la misma escala, por ejemplo, de 0 a 1.
4. Defina los umbrales de acción. Por ejemplo, las puntuaciones superiores a 0,7 inician las alertas.

Basándose en los análisis anteriores y en la puntuación de propensión al reingreso de un paciente, los médicos o los administradores de atención pueden configurar alertas para monitorear a sus

pacientes individuales en función de la puntuación calculada. Si está por encima de un umbral predefinido, se les notifica cuando se alcanza ese umbral. Esto ayuda a los administradores de atención a ser proactivos en lugar de reactivos a la hora de crear planes de atención de alta para sus pacientes. Guarda las puntuaciones de los resultados, el comportamiento y la propensión al reingreso de los pacientes de forma indexada en una base de datos vectorial de Amazon OpenSearch Service para que los administradores de atención puedan recuperarlas sin problemas mediante un agente de IA conversacional.

El siguiente diagrama muestra el flujo de trabajo de un agente de inteligencia artificial conversacional que un médico o un administrador de atención puede utilizar para obtener información sobre los resultados de los pacientes, el comportamiento esperado y la propensión a volver a ser ingresados. Los usuarios pueden obtener información a nivel de paciente, departamento o hospital. El agente de IA recupera estos datos, que se almacenan de forma indexada en una base de datos vectorial de Amazon OpenSearch Service. El agente utiliza la consulta para recuperar los datos relevantes y proporciona respuestas personalizadas, incluidas las acciones sugeridas para los pacientes que corren un alto riesgo de volver a ser admitidos. Según el nivel de riesgo, el agente también puede configurar recordatorios para los pacientes y los cuidadores.



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

1. El médico le hace una pregunta a un agente de IA conversacional, que alberga una función. AWS Lambda
2. La función Lambda inicia un LangChain agente.
3. La LangChain el agente envía la pregunta del usuario a un punto final de incrustación de texto de Amazon SageMaker AI. El punto final incorpora la pregunta.
4. La LangChain el agente pasa la pregunta incrustada a una base de conocimientos médicos en Amazon OpenSearch Service.
5. Amazon OpenSearch Service devuelve la información específica que es más relevante para la consulta del usuario al LangChain agente.
6. La LangChain los agentes envían la consulta y el contexto recuperado de la base de conocimientos a un modelo básico de Amazon Bedrock.

7. El modelo básico de Amazon Bedrock genera una respuesta y la envía a la función Lambda.
8. La función Lambda devuelve la respuesta al médico.
9. El médico inicia una función Lambda que envía un correo electrónico de seguimiento a un paciente que tiene un alto riesgo de reingreso.

## Alineación con el marco de AWS Well-Architected

[La arquitectura para rastrear el comportamiento de los pacientes y predecir las tasas de reingreso hospitalario integra gráficos del conocimiento médico y mejora los resultados de la atención médica](#)  
[Servicios de AWS, LLMs al tiempo que se alinea con los seis pilares del Well-Architected AWS Framework:](#)

- **Excelencia operativa:** la solución es un sistema automatizado y desacoplado que utiliza Amazon Bedrock y proporciona alertas en AWS Lambda tiempo real.
- **Seguridad:** esta solución está diseñada para cumplir con las normas sanitarias, como la HIPAA. También puede implementar el cifrado, un control de acceso detallado y las barandillas Amazon Bedrock para ayudar a proteger los datos de los pacientes.
- **Fiabilidad:** la arquitectura utiliza sistemas tolerantes a fallos y sin servidores. Servicios de AWS
- **Eficiencia del rendimiento:** Amazon OpenSearch Service y los productos afinados LLMs pueden proporcionar predicciones rápidas y precisas.
- **Optimización de costos:** las tecnologías y los pay-per-inference modelos sin servidor ayudan a minimizar los costos. Si bien el uso de un LLM ajustado puede generar cargos adicionales, el modelo utiliza un enfoque RAG que reduce los datos y el tiempo computacional necesarios para el proceso de ajuste.
- **Sostenibilidad:** la arquitectura minimiza el consumo de recursos mediante el uso de una infraestructura sin servidores. También es compatible con operaciones de atención médica eficientes y escalables.

# Caso de uso: gestionar y mejorar las competencias de su personal sanitario

La implementación de estrategias de transformación del talento y mejora de las habilidades ayuda a la fuerza laboral a seguir siendo experta en el uso de las nuevas tecnologías y prácticas en los servicios médicos y de salud. Las iniciativas proactivas de mejora de las competencias garantizan que los profesionales de la salud puedan brindar una atención de alta calidad a los pacientes, optimizar la eficiencia operativa y cumplir con las normas reglamentarias. Además, la transformación del talento fomenta una cultura de aprendizaje continuo. Esto es fundamental para adaptarse al cambiante panorama sanitario y abordar los nuevos desafíos de salud pública. Los enfoques de formación tradicionales, como la formación presencial y los módulos de aprendizaje estáticos, ofrecen un contenido uniforme a un público amplio. A menudo carecen de rutas de aprendizaje personalizadas, que son fundamentales para abordar las necesidades específicas y los niveles de competencia de los profesionales individuales. Esta one-size-fits-all estrategia puede provocar una desvinculación y una retención de conocimientos subóptima.

En consecuencia, las organizaciones de atención médica deben adoptar soluciones innovadoras, escalables e impulsadas por la tecnología que puedan determinar la brecha de cada uno de sus empleados en su estado actual y potencial futuro. Estas soluciones deberían recomendar itinerarios de aprendizaje hiperpersonalizados y el conjunto adecuado de contenidos de aprendizaje. Esto prepara eficazmente a la fuerza laboral para el futuro de la atención médica.

En la industria de la salud, puede aplicar la IA generativa para comprender y mejorar las habilidades de su fuerza laboral. Mediante la conexión de modelos de lenguaje grandes (LLMs) y recuperadores avanzados, las organizaciones pueden comprender qué habilidades tienen actualmente e identificar las habilidades clave que podrían ser necesarias en el futuro. Esta información le ayuda a cerrar la brecha mediante la contratación de nuevos trabajadores y la mejora de las habilidades de la fuerza laboral actual. Con Amazon Bedrock y los gráficos de conocimiento, las organizaciones de atención médica pueden desarrollar aplicaciones de dominio específico que faciliten el aprendizaje continuo y el desarrollo de habilidades.

Los conocimientos que ofrece esta solución le ayudan a gestionar eficazmente el talento, optimizar el rendimiento de la fuerza laboral, impulsar el éxito de la organización, identificar las habilidades existentes y elaborar una estrategia de talento. Esta solución puede ayudarlo a realizar estas tareas en semanas en lugar de meses.

## Información general de la solución

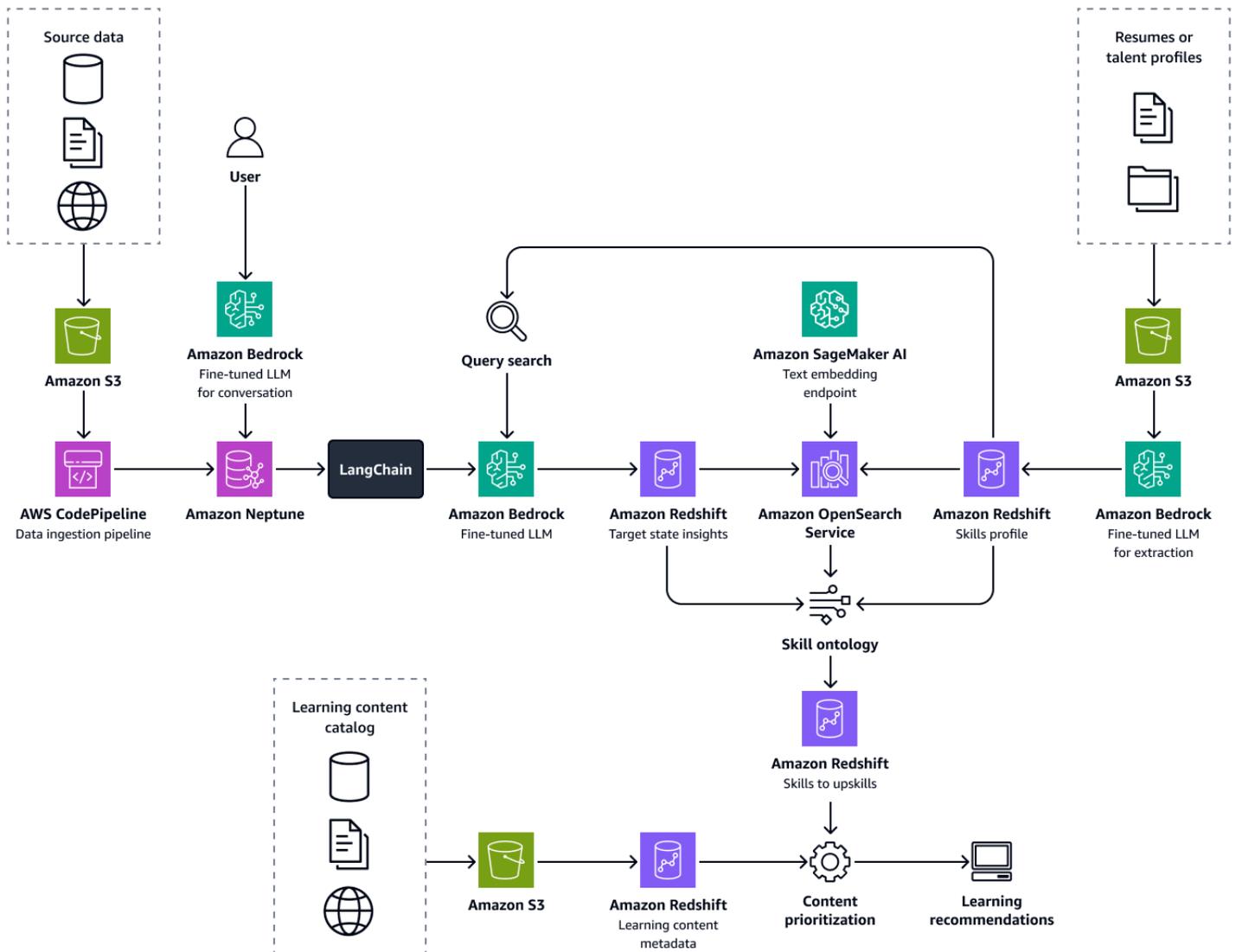
Esta solución es un marco de transformación del talento sanitario que consta de los siguientes componentes:

- **Analizador de currículum inteligente:** este componente puede leer el currículum de un candidato y extraer con precisión la información del candidato, incluidas sus habilidades. Solución de extracción de información inteligente creada con el modelo Llama 2 ajustado de Amazon Bedrock en un conjunto de datos de formación patentado que incluye currículos y perfiles de talentos de más de 19 industrias. Este proceso basado en la LLM ahorra cientos de horas al automatizar el proceso de revisión manual de los currículos y asignar a los mejores candidatos a los puestos vacantes.
- **Gráfico de conocimiento:** gráfico de conocimiento creado en Amazon Neptune, un repositorio unificado de información sobre talentos que incluye la taxonomía de funciones y habilidades de la organización y del sector, que captura la semántica del talento sanitario mediante definiciones de habilidades, funciones y sus propiedades, relaciones y restricciones lógicas.
- **Ontología de habilidades:** el descubrimiento de las proximidades de habilidades entre las habilidades de los candidatos y las habilidades ideales del estado actual o futuro (recuperadas mediante un gráfico de conocimiento) se logra mediante algoritmos de ontología que miden la similitud semántica entre las habilidades de los candidatos y las habilidades del estado objetivo.
- **Itinerario y contenido de aprendizaje:** este componente es un motor de recomendaciones de aprendizaje que puede recomendar el contenido de aprendizaje adecuado de un catálogo de materiales de aprendizaje de cualquier proveedor en función de las brechas de habilidades identificadas. Identificar las vías de mejora de competencias más óptimas para cada candidato mediante el análisis de las carencias de competencias y la recomendación de contenidos de aprendizaje priorizados, a fin de permitir un desarrollo profesional continuo y fluido para cada candidato durante la transición a un nuevo puesto.

Esta solución automatizada y basada en la nube se basa en servicios de aprendizaje automático LLMs, gráficos de conocimiento y recuperación aumentada (RAG). Puede ampliarse para procesar decenas o miles de currículums en un tiempo mínimo, crear perfiles de candidatos instantáneos, identificar brechas en su estado futuro actual o potencial y, luego, recomendar de manera eficiente el contenido de aprendizaje adecuado para cerrar estas brechas.

La siguiente imagen muestra el end-to-end flujo del marco. La solución se basa en LLMs Amazon Bedrock, optimizada con precisión. Estos LLMs recuperan datos de la base de conocimientos sobre

talentos de la salud en Amazon Neptune. Los algoritmos basados en datos hacen recomendaciones sobre vías de aprendizaje óptimas para cada candidato.



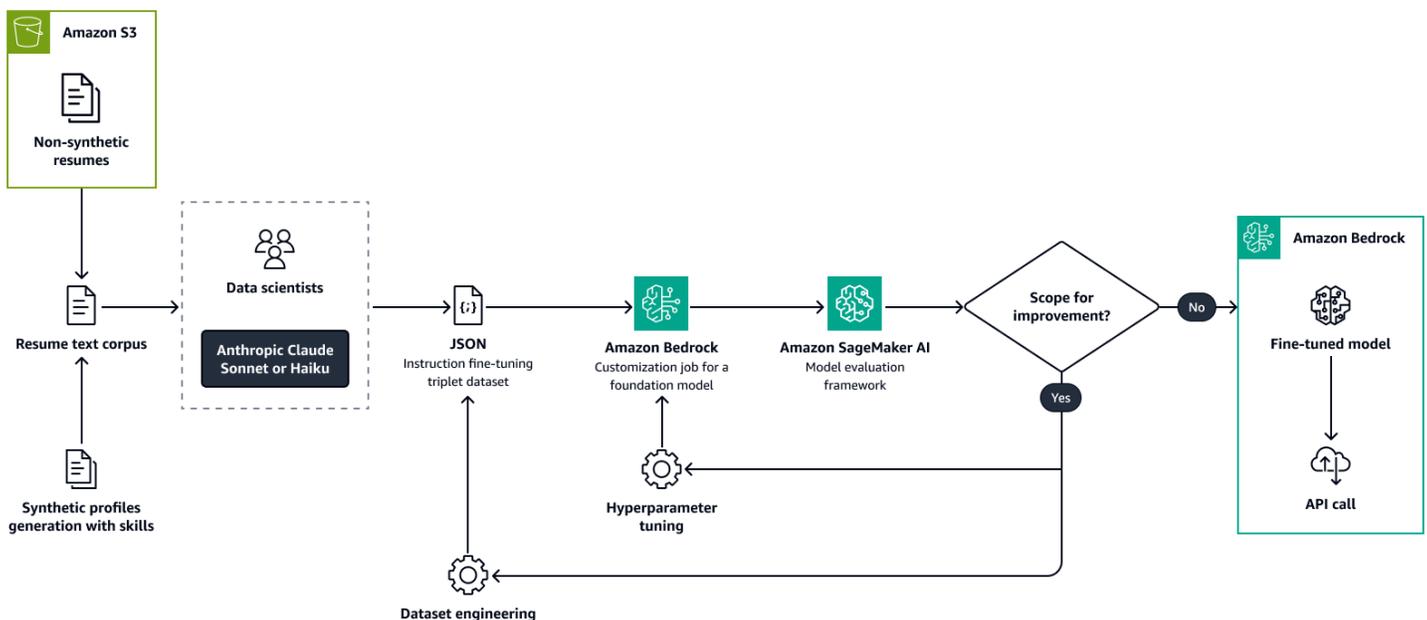
La creación de esta solución consta de los siguientes pasos:

- [Paso 1: Extraer información sobre el talento y crear un perfil de habilidades](#)
- [Paso 2: Descubrir la role-to-skill relevancia a partir de un gráfico de conocimiento](#)
- [Paso 3: Identificar las carencias de habilidades y recomendar la formación](#)

## Paso 1: Extraer información sobre el talento y crear un perfil de habilidades

En primer lugar, debe ajustar un modelo de lenguaje grande, como Llama 2, en Amazon Bedrock con un conjunto de datos personalizado. Esto adapta el LLM al caso de uso. Durante la formación, extraes de forma precisa y coherente los atributos clave del talento de los currículums de los candidatos o de perfiles de talentos similares. Estos atributos de talento incluyen las habilidades, el cargo actual, los títulos de experiencia con períodos de tiempo, la educación y las certificaciones. Para obtener más información, consulte [Personalice su modelo para mejorar su rendimiento para su caso de uso](#) en la documentación de Amazon Bedrock.

La siguiente imagen muestra el proceso para ajustar un modelo de análisis de currículum mediante Amazon Bedrock. Tanto los currículos reales como los creados sintéticamente se envían a un LLM para extraer la información clave. Un grupo de científicos de datos valida la información extraída comparándola con el texto original sin procesar. Luego, la información extraída se concatena utilizando las [chain-of-thought](#) indicaciones y el texto original para obtener un conjunto de datos de entrenamiento para su ajuste. Luego, este conjunto de datos se pasa a un trabajo de personalización de Amazon Bedrock, que ajusta el modelo. Un trabajo por lotes de Amazon SageMaker AI ejecuta un marco de evaluación de modelos que evalúa el modelo ajustado con precisión. Si es necesario mejorar el modelo, el trabajo se vuelve a ejecutar con más datos o con distintos hiperparámetros. Una vez que la evaluación cumpla con los estándares, usted hospeda el modelo personalizado a través del rendimiento aprovisionado por Amazon Bedrock.



## Paso 2: Descubrir la role-to-skill relevancia a partir de un gráfico de conocimiento

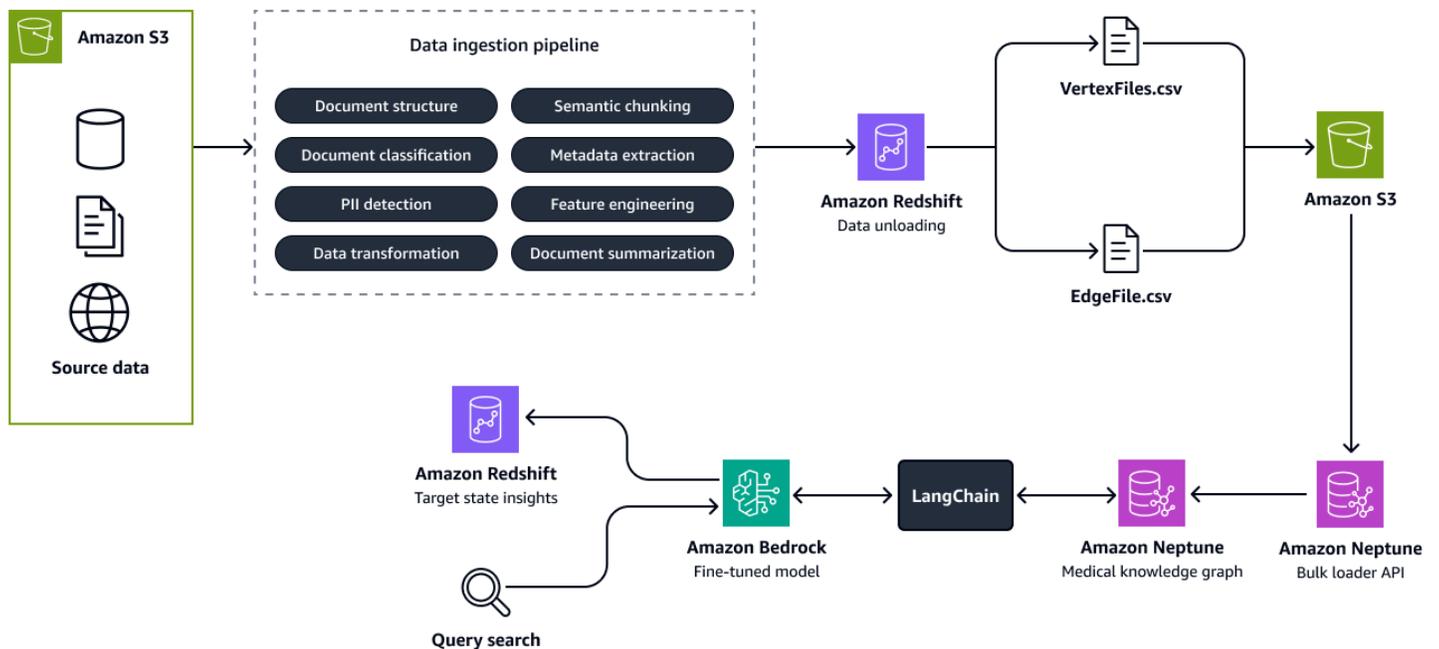
A continuación, debe crear un gráfico de conocimiento que resuma la taxonomía de las habilidades y funciones de su organización y de otras organizaciones del sector de la salud. Esta base de conocimientos enriquecida proviene de datos agregados de talento y organización en [Amazon Redshift](#). Puede recopilar datos sobre el talento de una variedad de proveedores de datos del mercado laboral y de fuentes de datos estructuradas y no estructuradas específicas de la organización, como los sistemas de planificación de recursos empresariales (ERP), un sistema de información de recursos humanos (HRIS), los currículos de los empleados, las descripciones de los puestos y los documentos sobre la arquitectura del talento.

Cree el gráfico de conocimiento sobre [Amazon Neptune](#). Los nodos representan habilidades y funciones, y los bordes representan las relaciones entre ellos. Amplíe este gráfico con metadatos para incluir detalles como el nombre de la organización, el sector, la familia laboral, el tipo de aptitud, el tipo de función y las etiquetas del sector.

A continuación, desarrollará una aplicación Graph Retrieval Augmented Generation (Graph RAG). Graph RAG es un enfoque de RAG que recupera datos de una base de datos de gráficos. Los siguientes son los componentes de la aplicación Graph RAG:

- Integración con un LLM en Amazon Bedrock: la aplicación utiliza un LLM en Amazon Bedrock para comprender el lenguaje natural y generar consultas. Los usuarios pueden interactuar con el sistema mediante el uso de un lenguaje natural. Esto lo hace accesible a las partes interesadas no técnicas.
- Orquestación y recuperación de información: uso o [LlamaIndexLangChain](#) orquestadores para facilitar la integración entre el LLM y el gráfico de conocimiento de Neptune. [Gestionan el proceso de conversión de consultas en lenguaje natural en consultas de OpenCypher](#). A continuación, ejecutan las consultas en el gráfico de conocimiento. Utilice la ingeniería rápida para instruir al LLM sobre las mejores prácticas para crear consultas de OpenCypher. Esto ayuda a optimizar las consultas para recuperar el subgráfico correspondiente, que contiene todas las entidades y relaciones pertinentes sobre las funciones y habilidades consultadas.
- Generación de información: el LLM de Amazon Bedrock procesa los datos gráficos recuperados. Genera información detallada sobre el estado actual y proyecta estados futuros para el rol consultado y las habilidades asociadas.

La siguiente imagen muestra los pasos para crear un gráfico de conocimiento a partir de los datos de origen. Los datos de origen estructurados y no estructurados se transfieren a la canalización de ingesta de datos. La canalización extrae y transforma la información en una formación de carga masiva CSV compatible con Amazon Neptune. La API de carga masiva carga los archivos CSV que están almacenados en un bucket de Amazon S3 al gráfico de conocimiento de Neptune. Para las consultas de los usuarios relacionadas con el estado futuro del talento, las funciones relevantes o las habilidades, el perfeccionado LLM de Amazon Bedrock interactúa con el gráfico de conocimiento a través de un LangChain orquestador. El orquestador recupera el contexto relevante del gráfico de conocimiento y envía las respuestas a la tabla de información de Amazon Redshift. La LangChain Orchestrator, al igual que [Graph QChain](#), convierte la consulta del usuario en lenguaje natural en una consulta de OpenCypher para consultar el gráfico de conocimiento. El modelo ajustado de Amazon Bedrock genera una respuesta basada en el contexto recuperado.



## Paso 3: Identificar las carencias de habilidades y recomendar la formación

En este paso, calcula con precisión la proximidad entre el estado actual de un profesional de la salud y las posibles funciones estatales futuras. Para ello, se realiza un análisis de afinidad entre las habilidades y se comparan los conjuntos de habilidades de la persona con el puesto de trabajo. En una base de datos vectorial de [Amazon OpenSearch Service](#), se almacena información de taxonomía de habilidades y metadatos de habilidades, como la descripción de la habilidad, el tipo de

habilidad y los grupos de habilidades. Utilice un modelo de incrustación de Amazon Bedrock, como los [modelos Amazon Titan Text Embeddings](#), para incrustar la habilidad clave identificada en los vectores. Mediante una búsqueda vectorial, puede recuperar las descripciones de las habilidades del estado actual y las habilidades del estado objetivo y realizar un análisis ontológico. El análisis proporciona puntuaciones de proximidad entre los pares de habilidades actuales y del estado objetivo. Para cada par, se utilizan las puntuaciones ontológicas calculadas para identificar las brechas en las afinidades de las habilidades. Luego, recomiendas la ruta óptima para mejorar las habilidades, que el candidato puede tener en cuenta durante las transiciones de roles.

Para cada puesto, recomendar el contenido de aprendizaje correcto para mejorar o volver a capacitarse implica un enfoque sistemático que comienza con la creación de un catálogo completo de contenido de aprendizaje. Este catálogo, que se almacena en una base de datos de Amazon Redshift, agrega contenido de varios proveedores e incluye metadatos, como la duración del contenido, el nivel de dificultad y el modo de aprendizaje. El siguiente paso consiste en extraer las habilidades clave que ofrece cada contenido y, a continuación, asignarlas a las habilidades individuales requeridas para el puesto objetivo. Para lograr este mapeo, se analiza la cobertura que proporciona el contenido mediante un análisis de proximidad de habilidades. Este análisis evalúa en qué medida las habilidades que se enseñan en el contenido se alinean con las habilidades deseadas para el puesto. Los metadatos desempeñan un papel fundamental a la hora de seleccionar el contenido más apropiado para cada habilidad, ya que garantizan que los alumnos reciban recomendaciones personalizadas que se adapten a sus necesidades de aprendizaje. Úselo LLMs en Amazon Bedrock para extraer conocimientos de los metadatos del contenido, realizar ingeniería de características y validar las recomendaciones de contenido. Esto mejora la precisión y la relevancia del proceso de mejora o recalcificación.

## Alineación con el marco de AWS Well-Architected

La solución se alinea con los seis pilares del [AWS Well-Architected](#) Framework:

- **Excelencia operativa:** una canalización modular y automatizada mejora la excelencia operativa. Los componentes clave de la canalización están desacoplados y automatizados, lo que permite actualizar los modelos más rápidamente y facilitar la supervisión. Además, los procesos de formación automatizados permiten lanzar modelos ajustados con mayor rapidez.
- **Seguridad:** esta solución procesa información confidencial y de identificación personal (PII), como los datos de los currículums y los perfiles de talentos. En [AWS Identity and Access Management \(IAM\)](#), implemente políticas de control de acceso detalladas y asegúrese de que solo el personal autorizado tenga acceso a estos datos.

- **Fiabilidad:** la solución utiliza Servicios de AWS, como Neptune, Amazon Bedrock y OpenSearch Service, que proporcionan tolerancia a errores, alta disponibilidad y acceso ininterrumpido a la información, incluso cuando hay una gran demanda.
- **Eficiencia del rendimiento:** las bases de datos vectoriales ajustadas en LLMs Amazon Bedrock y OpenSearch Service están diseñadas para procesar grandes conjuntos de datos de forma rápida y precisa a fin de ofrecer recomendaciones de aprendizaje personalizadas y oportunas.
- **Optimización de costos:** esta solución utiliza un enfoque RAG, lo que reduce la necesidad de una formación previa continua de los modelos. En lugar de ajustar todo el modelo de forma repetida, el sistema solo ajusta procesos específicos, como la extracción de información de los currículums y la estructuración de los resultados. Esto se traduce en un importante ahorro de costes. Al minimizar la frecuencia y la escala de los modelos de formación con un uso intensivo de recursos y al utilizar los servicios pay-per-use en la nube, las organizaciones sanitarias pueden optimizar sus costes operativos y, al mismo tiempo, mantener un alto rendimiento.
- **Sostenibilidad:** esta solución utiliza servicios escalables nativos de la nube que asignan los recursos informáticos de forma dinámica. Esto reduce el consumo de energía y el impacto ambiental y, al mismo tiempo, respalda las iniciativas de transformación del talento a gran escala y con uso intensivo de datos.

# Desarrollo y organización de soluciones de IA generativa para el cuidado de la salud

Para desarrollar las soluciones de esta guía, debe crear una arquitectura RAG optimizada para ofrecer a los proveedores de servicios de salud datos aumentados sobre LLMs los pacientes, información clínica y diagnóstica y pronósticos de los resultados de los pacientes. Esto requiere la integración de múltiples herramientas Servicios de AWS y herramientas para crear un flujo de trabajo coherente y eficiente. En esta sección se analiza lo siguiente:

- [Amazon Q Developer](#)— Utilice Amazon Q Developer para abordar cuestiones de ingeniería y errores de código durante el proceso de desarrollo.
- [Diseño RAG con múltiples recuperadores](#)— Diseñe e implemente soluciones RAG que utilicen varios recuperadores para encontrar el contexto médico correcto para la pregunta del usuario.
- [ReAct agentes](#)— Implementar agentes que combinen el razonamiento con la acción dinámica.

## Amazon Q Developer

Al crear una solución de IA generativa, puede resultar difícil crear agentes de IA y conectar los servicios clave. Sin embargo, [Amazon Q Developer](#) ayuda a los científicos de datos y a los ingenieros de IA al proporcionarles acceso a un asistente de IA generativo avanzado. Amazon Q puede abordar de forma rápida y precisa las preguntas de los usuarios y los errores de código, lo que puede ayudarle a optimizar el proceso de desarrollo de LLM. Amazon Q ofrece ventajas significativas para los desarrolladores que crean aplicaciones que utilizan los modelos básicos de Amazon Bedrock. Puede agilizar los flujos de trabajo y mejorar la calidad del código. Automatiza la generación de scripts y configuraciones de infraestructura como código (IaC) de Python, lo que reduce significativamente el tiempo y el esfuerzo de desarrollo. Mediante capacidades de refactorización avanzadas, Amazon Q puede mejorar el rendimiento del código, identificar las vulnerabilidades de seguridad y asegurarse de que los desarrolladores cumplan con las mejores prácticas. Además, facilita el aprendizaje y la adopción por parte de los principiantes al proporcionar sugerencias y explicaciones contextuales, lo que hace que las tareas de codificación complejas sean más accesibles y eficientes.

## Diseño RAG con múltiples recuperadores

En una aplicación de IA generativa, una canalización de RAG con múltiples recuperadores puede recuperar de manera eficiente información de múltiples fuentes de datos para ayudar a los proveedores de atención médica y a los médicos a responder a las preguntas médicas. Esta canalización utiliza diferentes tipos de recuperadores para extraer datos relevantes de distintas bases de conocimiento. Cada recuperador está especializado en obtener un tipo concreto de información, como historias clínicas de pacientes, información diagnóstica, notas clínicas o contenido de investigaciones médicas y textos académicos.

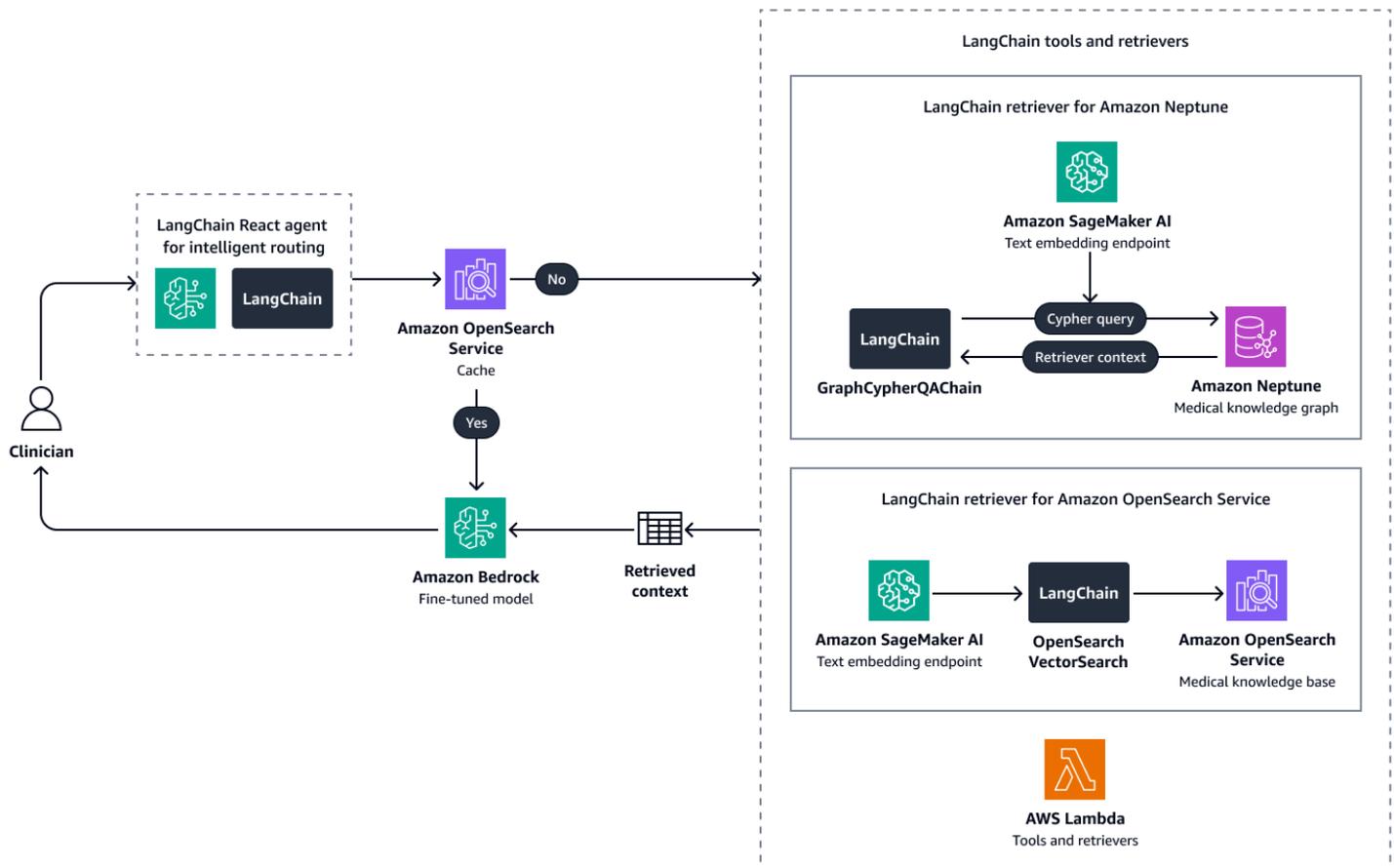
Utilice la naturaleza de los datos y los requisitos específicos de la aplicación para determinar cuál es la base de conocimientos de backend correcta para su caso de uso. Una base de datos vectorial de Amazon OpenSearch Service es ideal para grandes volúmenes de datos de salud no estructurados o semiestructurados, incluidos resúmenes de evaluaciones de diagnósticos por imágenes, resúmenes de alta, informes clínicos, investigaciones médicas y contenido de textos académicos. Por otro lado, un servicio de base de datos de gráficos, como Amazon Neptune, puede ser ideal para los casos de uso de la atención médica que requieren una exploración profunda de las relaciones temporales entre entidades, como el paciente, el historial del paciente, el proveedor de atención médica, los medicamentos, los síntomas y los tratamientos.

Un componente fundamental de este proceso es la predicción de la intención de consulta de los usuarios. Esto garantiza que el sistema dirija la consulta a la cadena de recuperadores correcta. Por ejemplo, si un médico pregunta sobre el historial de tratamiento del paciente, los síntomas, la interacción con el hospital, la probabilidad de que vuelva a ingresar en el hospital o los posibles resultados del paciente, el módulo de predicción de la intención de la consulta identifica esta intención. Dirige la solicitud a la cadena de recuperadores, que puede obtener los registros de los pacientes o los datos cronológicos del tratamiento a partir del gráfico de conocimientos médicos. Como alternativa, si la pregunta se refiere al descubrimiento de enfermedades, a evaluaciones diagnósticas específicas o a detalles de procedimientos clínicos específicos extraídos de libros de texto académicos, la consulta se dirige a la cadena de recuperadores, que puede obtener esta información de la base de datos vectorial del Servicio. OpenSearch [Puede utilizar la función de llamada a herramientas de](#) LangChain para vincular una herramienta personalizada al LLM de Amazon Bedrock que pueda clasificar una pregunta de usuario en intenciones predefinidas.

Este sistema RAG de recuperación múltiple incluye LangChain agentes diseñados para gestionar el acceso a la base de conocimientos específica. Puede usar... LangChain para organizar la interacción entre el Amazon Bedrock LLM, los diferentes recuperadores y las herramientas.

LangChain incluye una clase de uso de herramientas que le ayuda a crear herramientas personalizadas, como un clasificador de intenciones, un recuperador para Neptune, un recuperador para Service o cualquier otra herramienta que se pueda desarrollar OpenSearch para clasificar la intención del usuario y acceder a los datos de una base de conocimientos específica en un formato estructurado. A continuación, incorpore estas herramientas a la clase para crear un agente de Reasoning and Acting (). ReAct El ReAct agente procesa la pregunta del usuario, planifica los pasos secuenciales para responder a la pregunta y, a continuación, ejecuta de forma iterativa las herramientas disponibles y procesa las respuestas de la herramienta para finalmente responder a la consulta del usuario.

La siguiente imagen muestra cómo funciona un sistema RAG de recuperación múltiple diseñado para una recuperación eficiente del conocimiento y una resolución inteligente de consultas. A LangChain ReAct El agente analiza la intención del usuario, formula un plan estructurado de ejecución y selecciona las herramientas de recuperación más relevantes. El sistema consulta una caché de preguntas anteriores y comprueba si hay consultas similares en función de atributos clave, como la identificación del paciente, el estado de salud y la fecha de la visita. Si se encuentra una pregunta muy similar, la respuesta correspondiente se recupera directamente. De lo contrario, el agente ejecuta el recuperador correspondiente. Para recuperar información centrada en el paciente, como el historial de tratamiento, los síntomas, las interacciones en el hospital o la probabilidad de reingreso, el sistema utiliza un recuperador de gráficos. Para las evaluaciones diagnósticas, los procedimientos clínicos y los hallazgos médicos estructurados, el agente emplea un recuperador de bases de datos vectoriales. En los escenarios que requieren una combinación del conocimiento contextual de ambos almacenes de datos para generar una respuesta integral, el sistema utiliza una estrategia de recuperación híbrida que integra los resultados del gráfico de conocimiento y de la base de datos vectorial.



## ReAct agentes

Los agentes Reasoning and Acting (ReAct) están diseñados para aplicaciones RAG multifacéticas. Estos agentes proporcionan una poderosa combinación de razonamiento y acción dinámica, especialmente para aplicaciones complejas que implican step-by-step flujos de trabajo lógicos de recuperación de información. Para obtener más información, consulte [ReAct: Sinergizar el razonamiento y la actuación en modelos de lenguaje](#).

En el contexto médico y sanitario, las consultas de un clínico o un médico suelen tener múltiples facetas. Por ejemplo, un médico podría preguntar: «¿Qué tratamientos se administraron a pacientes similares con hipertensión y diabetes tipo 2?» Tras identificar la intención del usuario, que es buscar los tratamientos para la hipertensión y la diabetes tipo 2, el agente de IA debe dividir esta consulta en subtareas y, a continuación, elegir la estrategia de recuperación más eficaz. En este caso, el agente de IA debe identificar los nodos más relevantes (como la edad, el sexo, las afecciones, los tratamientos y los medicamentos del paciente) y, a continuación, consultar el gráfico para ver estas entidades y sus atributos y relaciones. ReAct los agentes son muy útiles porque combinan

la capacidad de razonamiento (inferencia lógica) de un LLM con una acción (consultar recursos o bases de conocimiento externos o interactuar con ellos).

Para responder a la pregunta del usuario «¿Qué tratamientos se administraron a pacientes similares con hipertensión y diabetes tipo 2?» , el siguiente ejemplo ilustra cómo funciona un ReAct agente:

1. Razonamiento con el ReAct agente: el agente deduce que la pregunta implica recuperar información sobre afecciones (diabetes e hipertensión). Considera la edad del paciente, los tratamientos, los medicamentos y el período a analizar.
2. Acción del agente: el agente utiliza OpenCypher para consultar el gráfico de conocimientos sobre los tratamientos específicos para la diabetes de tipo 2 y la hipertensión. También muestra los medicamentos administrados, las fechas de las visitas al hospital, los efectos secundarios de los medicamentos, los resultados conocidos de los pacientes y los datos de referencias cruzadas de pacientes similares (por ejemplo, pacientes del mismo sexo y edad).
3. Observación del agente: a partir del gráfico de conocimiento, el agente recupera los datos tabulares de los seis meses más recientes sobre los tratamientos administrados a pacientes con hipertensión y diabetes de tipo 2.
4. Razonamiento de los agentes: para clasificar los resultados de los registros recuperados, el agente identifica los atributos importantes, como la antigüedad, los efectos secundarios de los medicamentos o los resultados conocidos de los pacientes.
5. Acción del agente: el agente cambia la clasificación de los registros en función de los atributos identificados y de la lógica predefinida que se imparte a través de la línea de comandos del sistema.
6. Generación de respuestas: el LLM de Amazon Bedrock genera una respuesta en función del contexto que preparó el ReAct agente.

# Evaluación de soluciones de IA generativa para el cuidado de la salud

Evaluar las soluciones de IA para el sector sanitario que cree es fundamental para garantizar que sean eficaces, fiables y escalables en los entornos médicos del mundo real. Utilice un enfoque sistemático para evaluar el rendimiento de cada componente de la solución. El siguiente es un resumen de las metodologías y métricas que puede utilizar para evaluar su solución.

## Temas

- [Evaluación de la extracción de información](#)
- [Evaluación de soluciones de RAG con varios recuperadores](#)
- [Evaluar una solución mediante un LLM](#)

## Evaluación de la extracción de información

Evalúe el rendimiento de las soluciones de extracción de información, como el [analizador inteligente de currículums](#) y el [extractor de entidades personalizado](#). Puede medir la alineación de las respuestas de estas soluciones mediante un conjunto de datos de prueba. Si no tiene un conjunto de datos que abarque perfiles versátiles de talentos del sector de la salud y los historiales médicos de los pacientes, puede crear un conjunto de datos de pruebas personalizado utilizando la capacidad de razonamiento de un LLM. Por ejemplo, puedes usar un modelo de parámetros grande, como Anthropic Claude modelos, para generar un conjunto de datos de prueba.

Las siguientes son tres métricas clave que puede utilizar para evaluar los modelos de extracción de información:

- **Precisión e integridad:** estas métricas evalúan hasta qué punto el resultado capturó la información correcta y completa presente en los datos básicos. Esto implica comprobar tanto la exactitud de la información extraída como la presencia de todos los detalles relevantes en la información extraída.
- **Semejanza y relevancia:** estas métricas evalúan las similitudes semánticas, estructurales y contextuales entre los datos de salida y los datos basados en la verdad básica (la similitud) y el grado en que el resultado se alinea con el contenido, el contexto y la intención de los datos basados en la verdad básica (la relevancia) y los aborda.
- **Tasa de recuperación o captura ajustada:** estas tasas determinan empíricamente cuántos de los valores actuales de los datos basados en la verdad básica fueron identificados correctamente

por el modelo. La tasa debe incluir una penalización para todos los valores falsos que extraiga el modelo.

- Puntuación de precisión: la puntuación de precisión le ayuda a determinar cuántos falsos positivos están presentes en las predicciones, en comparación con los positivos verdaderos. Por ejemplo, puedes usar métricas de precisión para medir la exactitud de la habilidad extraída.

## Evaluación de soluciones de RAG con varios recuperadores

Para evaluar en qué medida el sistema recupera la información relevante y con qué eficacia la utiliza para generar respuestas precisas y adecuadas al contexto, puede utilizar las siguientes métricas:

- Relevancia de la respuesta: mida la relevancia de la respuesta generada, que utiliza el contexto recuperado, para la consulta original.
- Precisión del contexto: del total de resultados recuperados, evalúe la proporción de documentos o fragmentos recuperados que son relevantes para la consulta. Una mayor precisión del contexto indica que el mecanismo de recuperación es eficaz a la hora de seleccionar la información relevante.
- Fidelidad: evalúa la precisión con la que la respuesta generada refleja la información en el contexto recuperado. En otras palabras, mide si la respuesta se mantiene fiel a la información de origen.

## Evaluar una solución mediante un LLM

Puedes usar una técnica llamada LLM-as-a-judge para evaluar las respuestas de texto de tu solución de IA generativa. Implica utilizarla LLMs para evaluar y valorar el rendimiento de los resultados del modelo. Esta técnica utiliza las capacidades de Amazon Bedrock para emitir juicios sobre varios atributos, como la calidad de la respuesta, la coherencia, el cumplimiento, la precisión y la integridad de los datos según las preferencias humanas o la veracidad de los datos fundamentales. Para realizar una evaluación exhaustiva, se utilizan técnicas [chain-of-thought \(CoT\)](#) y [de indicaciones de pocos](#) pasos. El mensaje indica al LLM que evalúe la respuesta generada con una rúbrica de puntuación, y las pocas muestras del mensaje muestran el proceso de evaluación real. El mensaje también incluye pautas que debe seguir el evaluador del LLM. Por ejemplo, podría considerar usar una o más de las siguientes técnicas de evaluación que utilizan un LLM para evaluar las respuestas generadas:

- **Comparación por pares:** entregue al evaluador del LLM una pregunta médica y varias respuestas generadas por diferentes versiones iterativas de los sistemas RAG que creó. Pide al evaluador del LLM que determine la mejor respuesta en función de la calidad de la respuesta, la coherencia y el cumplimiento de la pregunta original.
- **Calificación con una sola respuesta:** esta técnica es adecuada para los casos de uso en los que es necesario evaluar la precisión de la categorización, como la clasificación de los resultados de los pacientes, la categorización del comportamiento de los pacientes, la probabilidad de reingreso del paciente y la categorización del riesgo. Utilice el evaluador LLM para analizar la categorización o clasificación individual de forma aislada y evalúe el razonamiento que ha proporcionado comparándolo con datos basados en datos básicos.
- **Calificación guiada por referencia:** proporcione al evaluador del LLM una serie de preguntas médicas que requieran respuestas descriptivas. Cree ejemplos de respuestas a estas preguntas, como respuestas de referencia o respuestas ideales. Pida al evaluador del LLM que compare la respuesta generada por el LLM con las respuestas de referencia o las respuestas ideales, y pídale al evaluador del LLM que califique la respuesta generada en función de su precisión, integridad, similitud, relevancia u otros atributos. Esta técnica le ayuda a evaluar si las respuestas generadas se alinean con una respuesta estándar o ejemplar bien definida.

# Recursos

## AWS documentación

- [Documentación de Amazon Bedrock](#)
- [Documentación de Amazon Neptune](#)
- [Documentación OpenSearch de Amazon Service](#)
- [Aplicación del marco AWS Well-Architected para Amazon Neptune](#) (guía prescriptiva)AWS
- [Mejores prácticas operativas para Amazon OpenSearch Service](#) (documentación OpenSearch del servicio)
- [Uso de Amazon Comprehend Medical LLMs y para la salud y las ciencias de la vida](#)AWS (orientación prescriptiva)

## AWS publicaciones de blog

- [Cree aplicaciones de IA generativa basadas en agentes y RAG con el nuevo modelo Amazon Titan Text Premier, disponible en Amazon Bedrock](#)
- [Complemente la inteligencia comercial creando un gráfico de conocimiento a partir de un almacén de datos con Amazon Neptune](#)
- [Uso de gráficos de conocimiento para crear aplicaciones GraphRag con Amazon Bedrock y Amazon Neptune](#)

## Otros recursos

- [Integración de la generación de recuperación aumentada con modelos lingüísticos de gran tamaño en nefrología: avances en las aplicaciones prácticas](#) (Central, Biblioteca Nacional de Medicina) PubMed
- [Introducción a LangChain](#) (LangChain documentación)

# Colaboradores

## Creación

- Nitu Nivedita, director general de inteligencia artificial, datos e inteligencia artificial, Accenture
- Manoj Appully, fundador y director de tecnología de Cadiem
- Conor Folan, consultor de datos e IA de Accenture
- Deepak Krishna AR, consultor de datos e IA, Accenture
- Almore Cato, director de datos e IA de Accenture
- Soonam Kurian, arquitecto principal de soluciones, AWS

## Revisando

- Sally Lin, directora sénior de ciencia de datos e inteligencia artificial de Accenture
- Terry Huang, director de ciencia de datos para datos e IA de Accenture
- William Lorenz, arquitecto de soluciones de Partners, AWS

## Redacción técnica

- Lilly AbouHarb, redactora técnica sénior, AWS

## Historial de documentos

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

Cambio	Descripción	Fecha
<a href="#">Publicación inicial</a>	—	14 de marzo 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

IaC

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 (IIoT)

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 IIoT?](#).

### 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. Laseudonimización puede ayudar a proteger la privacidad personal. Los datosseudonimizados 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.