

Marcos, protocolos y herramientas de IA de las agencias sobre AWS

AWS Guía prescriptiva



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

AWS Guía prescriptiva: Marcos, protocolos y herramientas de IA de las agencias sobre AWS

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

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

Table of Contents

Introducción	1
Destinatarios previstos	2
Objetivos	2
Acerca de esta serie de contenido	2
Marcos de IA para agencias	3
Strands Agents	4
Características clave de Strands Agents	4
Cuándo se debe usar Strands Agents	5
Enfoque de implementación para Strands Agents	5
Ejemplo real de Strands Agents	6
LangChain y LangGraph	6
Características clave de y LangChainLangGraph	6
Cuándo usar LangChain y LangGraph	7
Enfoque de implementación para y LangChainLangGraph	8
Ejemplo real de y LangChainLangGraph	8
CrewAI	8
Características clave de CrewAl	9
Cuándo se debe usar CrewAl	9
Enfoque de implementación para CrewAl	10
Ejemplo real de CrewAl	10
Agentes de Amazon Bedrock	11
Características principales de Amazon Bedrock Agents	11
Cuándo usar Amazon Bedrock Agents	12
Enfoque de implementación para Amazon Bedrock Agents	12
Ejemplo real de Amazon Bedrock Agents	13
AutoGen	13
Características clave de AutoGen	13
Cuándo se debe usar AutoGen	14
Enfoque de implementación para AutoGen	15
Ejemplo del mundo real de AutoGen	15
Comparación de los marcos de IA de las agencias	16
Consideraciones a la hora de elegir un marco de IA para agencias	17
Protocolos de agencia	18
¿Por qué es importante la selección de protocolos	18

Ventajas de los protocolos abiertos	19
Agent-to-agent protocolos	19
Decidir entre las opciones de protocolo	20
Selección de los protocolos de los agentes	21
Consideraciones sobre el protocolo empresarial	22
Consideraciones sobre los protocolos para empresas emergentes y pequeñas y	
medianas	22
Consideraciones sobre los protocolos gubernamentales y de la industria regulada	22
Estrategia de implementación de protocolos de agencia	23
Cómo empezar con MCP	23
Herramientas	25
Categorías de herramientas	25
Herramientas basadas en protocolos	25
Herramientas nativas de Framework	26
Metaherramientas	26
Herramientas basadas en protocolos	26
Características de seguridad de las herramientas MCP	27
Cómo empezar con las herramientas de MCP	28
Herramientas nativas de Framework	28
Meta-herramientas	29
Metaherramientas de flujo de trabajo	29
Metaherramientas Agent Graph	30
Metaherramientas de memoria	30
Estrategia de integración de herramientas	30
Mejores prácticas de seguridad para la integración de herramientas	31
Autenticación y autorización	31
Protección de los datos	32
Supervisión y auditoría	32
Conclusión	33
Recursos	34
AWS Blogs	34
AWS Guía prescriptiva	34
AWS recursos	34
Otros recursos	35
Historial de documentos	36
Glosario	37

#	37
A	38
В	41
C	43
D	46
E	50
F	53
G	55
H	56
1	57
L	60
M	61
O	65
P	68
Q	71
R	71
S	74
T	78
U	80
V	81
W	81
Z	82
	xxxi\

Marcos, protocolos y herramientas de inteligencia artificial de las agencias en AWS

Aaron Sempf y Joshua Samuel, Amazon Web Services ()AWS

Julio de 2025 (historial del documento)

La IA de agencia es un poderoso paradigma en la intersección de la IA, los sistemas distribuidos y la ingeniería de software. Es una clase de sistemas inteligentes que se compone de agentes de software autónomos y asíncronos. Los agentes muestran capacidad de acción, son capaces de percibir el contexto, razonar por encima de los objetivos, tomar decisiones y emprender acciones intencionadas en nombre de los usuarios o los sistemas. Estos agentes funcionan de forma independiente, a menudo en colaboración, en entornos distribuidos y están diseñados para perseguir objetivos delegados con inteligencia, memoria e intención integradas.

Además AWS, las organizaciones pueden aprovechar la IA de los agentes para automatizar flujos de trabajo complejos, mejorar los procesos de toma de decisiones y crear sistemas con mayor capacidad de respuesta. Esta guía proporciona información sobre los componentes clave que son necesarios para crear soluciones de inteligencia artificial eficaces para los agentes:

- Los marcos de IA de las agencias describen los marcos de IA de las agencias actuales e incluyen revisiones de sus beneficios y casos de uso. Descubra cómo estos marcos reducen el trabajo pesado indiferenciado entre patrones, protocolos y herramientas. Comprenda los criterios de selección clave para elegir el marco adecuado para sus requisitos.
- Los protocolos de <u>agencia exploran los protocolos</u> de comunicación estandarizados esenciales para las interacciones entre los agentes. Agent-to-agentestán surgiendo protocolos, como el Model Context Protocol (MCP) y el Agent2Agent (A2A) de código abierto, junto con otras implementaciones patentadas. Descubra cómo los protocolos comunes permiten que diferentes protocolos interactúen sin problemas.
- <u>Las herramientas</u> proporcionan información sobre las herramientas basadas en protocolos (como
 el MCP), las herramientas nativas del marco y las metaherramientas. Las organizaciones pueden
 crear un conjunto de herramientas que se integre con los sistemas clave de sus flujos de trabajo, lo
 que permite flujos de trabajo de agentes basados en servidores y usuarios finales.

1

Destinatarios previstos

Esta guía está dirigida a arquitectos, desarrolladores y líderes tecnológicos que buscan aprovechar el poder de los agentes de software impulsados por la IA en aplicaciones modernas nativas de la nube.

Objetivos

Esta guía lo ayuda a hacer lo siguiente:

- Compare diferentes marcos de IA de agencias para seleccionar el más adecuado para su caso de uso.
- Comprenda las ventajas de los protocolos abiertos para crear arquitecturas de IA de agencias sostenibles.
- Cree una estrategia de integración de herramientas adecuada al crear sistemas de agentes.

Acerca de esta serie de contenido

Esta guía forma parte de un conjunto de publicaciones que proporcionan planos arquitectónicos y orientación técnica para crear agentes de software basados en la IA. AWS La serie de guías AWS prescriptivas incluye lo siguiente:

- · Operacionalización de la IA de los agentes en AWS
- Fundamentos de la IA agencial en AWS
- · Los patrones y flujos de trabajo de la IA de los agentes están activos AWS
- Los marcos, protocolos y herramientas de inteligencia artificial de las agencias están disponibles (esta guía) AWS
- Creación de arquitecturas sin servidor para la IA de los agentes en AWS
- Creación de arquitecturas multiusuario para la IA de los agentes en AWS

Para obtener más información sobre esta serie de contenido, consulte Agentic Al.

Destinatarios previstos 2

Marcos de IA para agencias

<u>Foundations of Agentic AI on AWS</u> examina los patrones y flujos de trabajo principales que permiten un comportamiento autónomo y orientado a objetivos. La clave de la implementación de estos patrones es la elección del marco. Un marco es la base del software que proporciona la estructura, las herramientas y las capacidades de orquestación necesarias para crear agentes de IA autónomos listos para la producción.

Los marcos de inteligencia artificial eficaces proporcionan varias capacidades esenciales que transforman las interacciones sin procesar con un modelo de lenguaje amplio (LLM) en agentes sólidos y autónomos capaces de funcionar de forma independiente:

- La orquestación de agentes coordina el flujo de información y la toma de decisiones entre uno o varios agentes para lograr objetivos complejos sin intervención humana.
- La integración de herramientas permite a los agentes interactuar con sistemas externos y fuentes de datos para ampliar sus capacidades más allá del procesamiento del lenguaje. APIs Para obtener más información, consulte la descripción general de las herramientas en la Strands Agents documentación.
- La administración de la memoria proporciona un estado persistente o basado en sesiones para mantener el contexto en todas las interacciones, algo esencial para las tareas autónomas de larga duración. Para obtener más información, consulte <u>Cómo pensar en los marcos de agentes en el</u> blog. LangChain
- La definición del flujo de trabajo admite patrones estructurados como las cadenas, el enrutamiento, la paralelización y los bucles de reflexión que permiten un razonamiento autónomo sofisticado.
- La implementación y el monitoreo facilitan la transición del desarrollo a la producción con la observabilidad de los sistemas autónomos. Para obtener más información, consulte el anuncio de <u>LangGraphPlatform GA</u>.

Estas capacidades se implementan con diferentes enfoques y énfasis en todo el panorama estructural, y cada una de ellas ofrece ventajas distintas para los diferentes casos de uso de agentes autónomos y contextos organizacionales.

En esta sección, se describen y comparan los principales marcos para crear soluciones de inteligencia artificial basadas en agentes, centrándose en sus puntos fuertes, limitaciones y casos de uso ideales para el funcionamiento autónomo:

- Agentes de Strands
- LangChain y LangGraph
- Crew Al
- Agentes de Amazon Bedrock
- AutoGen
- Comparación de los marcos de IA de las agencias



Note

Esta sección cubre los marcos que respaldan específicamente a la agencia de la IA y no cubre las interfaces frontend o la IA generativa sin agencia.

Strands Agents

Strands Agentses un SDK de código abierto que fue lanzado inicialmente por AWS, tal y como se describe en el blog de código AWS abierto. Strands Agentsestá diseñado para crear agentes de lA autónomos con un enfoque centrado en el modelo. Proporciona un marco flexible y ampliable diseñado para funcionar sin problemas y, al Servicios de AWS mismo tiempo, abierto a la integración con componentes de terceros. Strands Agents es ideal para crear soluciones totalmente autónomas.

Características clave de Strands Agents

Strands Agentsincluye las siguientes características clave:

- Diseño basado en el modelo: se basa en el concepto de que el modelo básico es el núcleo de la inteligencia de los agentes, lo que permite un razonamiento autónomo sofisticado. Para obtener más información, consulte Agent Loop en la Strands Agents documentación.
- Integración con el MCP: compatibilidad nativa con el Model Context Protocol (MCP), lo que permite el suministro de contexto estandarizado LLMs para un funcionamiento autónomo y uniforme.
- Servicio de AWS integración: conexión perfecta a Amazon Bedrock y otros Servicios de AWS para flujos de trabajo autónomos integrales. AWS Lambda AWS Step Functions Para obtener más información, consulte el resumen AWS semanal (AWS blog).
- Selección de modelos de base: admite varios modelos de base, incluidos Anthropic Claude, Amazon Nova (Premier, Pro, Lite y Micro) en Amazon Bedrock y otros para optimizarlos para

Strands Agents 4 diferentes capacidades de razonamiento autónomo. Para obtener más información, consulte Amazon Bedrock en la Strands Agents documentación.

- Integración de la API LLM: integración flexible con diferentes interfaces de servicios LLM, incluidas Amazon Bedrock, OpenAI y otras, para la implementación en producción. Para obtener más información, consulte Uso básico de Amazon Bedrock en la Strands Agents documentación.
- Capacidades multimodales: Support para múltiples modalidades, incluido el procesamiento de texto, voz e imágenes para interacciones integrales entre agentes autónomos. Para obtener más información, consulte Amazon Bedrock Multimodal Support en la Strands Agents documentación.
- Ecosistema de herramientas: amplio conjunto de herramientas para la Servicio de AWS
 interacción, con capacidad de ampliación para herramientas personalizadas que amplían las
 capacidades autónomas. Para obtener más información, consulte la descripción general de las
 herramientas en la Strands Agents documentación.

Cuándo se debe usar Strands Agents

Strands Agentses especialmente adecuado para escenarios de agentes autónomos, que incluyen:

- Organizaciones que se basan en una AWS infraestructura que desean una integración nativa con flujos Servicios de AWS de trabajo autónomos
- Equipos que requieren funciones de seguridad, escalabilidad y cumplimiento de nivel empresarial para los sistemas autónomos de producción
- Proyectos que necesitan flexibilidad a la hora de seleccionar modelos entre distintos proveedores para realizar tareas autónomas especializadas
- Casos de uso que requieren una estrecha integración con los AWS flujos de trabajo y los recursos existentes para lograr procesos autónomos de principio a fin

Enfoque de implementación para Strands Agents

Strands Agentsproporciona un enfoque de implementación sencillo para las partes interesadas de la empresa, tal como se describe en su Guía de inicio rápido. El marco permite a las organizaciones:

- Seleccione modelos de base como Amazon Nova (Premier, Pro, Lite o Micro) en Amazon Bedrock en función de los requisitos empresariales específicos.
- Defina herramientas personalizadas que se conecten a los sistemas y fuentes de datos empresariales.

- Procese múltiples modalidades, incluyendo texto, imágenes y voz.
- Implemente agentes que puedan responder de forma autónoma a las consultas empresariales y realizar tareas.

Este enfoque de implementación permite a los equipos empresariales desarrollar y desplegar rápidamente agentes autónomos sin necesidad de una amplia experiencia técnica en el desarrollo de modelos de IA.

Ejemplo real de Strands Agents

AWS Transform para Strands Agents que .NET potencie sus capacidades de modernización de aplicaciones, tal y como se describe en <u>AWS Transform forma.NET</u>, el primer servicio de <u>inteligencia artificial para modernizar las aplicaciones.NET a gran escala</u> (Blog).AWS Este servicio de producción emplea a varios agentes autónomos especializados. Los agentes trabajan juntos para analizar las aplicaciones de .NET heredadas, planificar estrategias de modernización y ejecutar transformaciones de código en arquitecturas nativas de la nube sin intervención humana. <u>AWS Transform para .NET</u> demuestra la preparación para la producción de los Strands Agents sistemas autónomos empresariales.

LangChain y LangGraph

LangChaines uno de los marcos más consolidados en el ecosistema de la IA de las agencias.

LangGraphamplía sus capacidades para admitir flujos de trabajo de agentes complejos y detallados, tal y como se describe en el blog. LangChain En conjunto, proporcionan una solución integral para crear agentes de IA autónomos sofisticados con amplias capacidades de orquestación para un funcionamiento independiente.

Características clave de y LangChainLangGraph

LangChaine LangGraph incluyen las siguientes características clave:

- Ecosistema de componentes: amplia biblioteca de componentes prediseñados para diversas capacidades de agentes autónomos, lo que permite el rápido desarrollo de agentes especializados. Para obtener más información, consulte la Documentación de LangChain.
- Selección de modelos de base: Support para diversos modelos de base, incluidos Anthropic
 Claude, modelos Amazon Nova (Premier, Pro, Lite y Micro) en Amazon Bedrock y otros para

diferentes capacidades de razonamiento. Para obtener más información, consulte Entradas y salidas en la LangChain documentación.

- Integración de la API LLM: interfaces estandarizadas para varios proveedores de servicios de modelos de lenguaje grandes (LLM), incluido Amazon BedrockOpenAI, y otros para una implementación flexible. Para obtener más información, consulte <u>LLMs</u> en la documentación del LangChain.
- Procesamiento multimodal: soporte integrado para el procesamiento de texto, imágenes y audio para permitir interacciones multimodales ricas entre agentes autónomos. Para obtener más información, consulte Multimodalidad en la documentación. LangChain
- Flujos de trabajo basados en gráficos: LangGraph permiten definir comportamientos complejos de agentes autónomos como máquinas de estados, lo que respalda una sofisticada lógica de decisiones. Para obtener más información, consulte el anuncio de LangGraphPlatform GA.
- Abstracciones de memoria: múltiples opciones para la gestión de la memoria a corto y largo plazo, algo esencial para los agentes autónomos que mantienen el contexto a lo largo del tiempo.
 Para obtener más información, consulte Cómo añadir memoria a los chatbots en la LangChain documentación.
- Integración de herramientas: amplio ecosistema de integraciones de herramientas en varios servicios y APIs que amplía las capacidades de los agentes autónomos. Para obtener más información, consulte las herramientas en la LangChain documentación.
- LangGraph plataforma: solución gestionada de implementación y supervisión para entornos de producción, que admite agentes autónomos de larga duración. Para obtener más información, consulte el anuncio de LangGraphPlatform GA.

Cuándo usar LangChain y LangGraph

LangChainy LangGraph son especialmente adecuados para escenarios de agentes autónomos, que incluyen:

- Flujos de trabajo complejos de razonamiento de varios pasos que requieren una orquestación sofisticada para una toma de decisiones autónoma
- Proyectos que necesitan acceso a un gran ecosistema de componentes e integraciones prediseñados para diversas capacidades autónomas
- Equipos con una infraestructura y experiencia Python en aprendizaje automático (ML) existentes que desean crear sistemas autónomos

 Casos de uso que requieren una gestión del estado compleja en sesiones de agentes autónomos de larga duración

Enfoque de implementación para y LangChainLangGraph

LangChainy LangGraph proporcionar un enfoque de implementación estructurado para las partes interesadas de la empresa, tal como se detalla en la <u>LangGraphdocumentación</u>. El marco permite a las organizaciones:

- Defina gráficos de flujo de trabajo sofisticados que representen los procesos empresariales.
- Cree patrones de razonamiento de varios pasos con puntos de decisión y lógica condicional.
- Integre las capacidades de procesamiento multimodal para gestionar diversos tipos de datos.
- Implemente el control de calidad mediante mecanismos integrados de revisión y validación.

Este enfoque basado en gráficos permite a los equipos empresariales modelar procesos de decisión complejos como flujos de trabajo autónomos. Los equipos tienen una visibilidad clara de cada paso del proceso de razonamiento y la capacidad de auditar las vías de toma de decisiones.

Ejemplo real de y LangChainLangGraph

Vodafoneha implementado agentes autónomos que utilizan LangChain (yLangGraph) para mejorar sus flujos de trabajo de operaciones e ingeniería de datos, como se detalla en su <u>estudio de caso LangChain empresarial</u>. Crearon asistentes de IA internos que supervisan de forma autónoma las métricas de rendimiento, recuperan información de los sistemas de documentación y presentan información útil, todo ello mediante interacciones en lenguaje natural.

La Vodafone implementación utiliza cargadores de documentos LangChain modulares, integración vectorial y soporte para múltiples LLMs (OpenAI, LLaMA 3 yGemini) para crear prototipos y comparar rápidamente estas canalizaciones. Luego, solían LangGraph estructurar la organización multiagente mediante el despliegue de subagentes modulares. Estos agentes realizan tareas de recopilación, procesamiento, resumen y razonamiento. LangGraphintegraron estos agentes APIs en sus sistemas en la nube.

CrewAl

CrewAles un marco de código abierto centrado específicamente en la orquestación autónoma de múltiples agentes, disponible en. GitHub Proporciona un enfoque estructurado para crear equipos de

agentes autónomos especializados que colaboran para resolver tareas complejas sin intervención humana. CrewAlhace hincapié en la coordinación basada en funciones y la delegación de tareas.

Características clave de CrewAl

CrewAlproporciona las siguientes características clave:

- Diseño de agentes basado en funciones: los agentes autónomos se definen con funciones, objetivos e historias de fondo específicos para disponer de conocimientos especializados. Para obtener más información, consulte Cómo crear agentes eficaces en la documentación. CrewAl
- Delegación de tareas: mecanismos integrados para asignar tareas de forma autónoma a los agentes apropiados en función de sus capacidades. Para obtener más información, consulte <u>las</u> tareas en la CrewAl documentación.
- Colaboración entre agentes: marco para la comunicación autónoma entre agentes y el intercambio de conocimientos sin mediación humana. Para obtener más información, consulte <u>la colaboración</u> en la CrewAl documentación.
- Gestión de procesos: flujos de trabajo estructurados para la ejecución secuencial y paralela de tareas autónomas. Para obtener más información, consulte Procesos en la CrewAl documentación.
- Selección de modelos de base: Support para varios modelos de base, incluidos los modelos Anthropic Claude, Amazon Nova (Premier, Pro, Lite y Micro) en Amazon Bedrock y otros para optimizarlos para diferentes tareas de razonamiento autónomo. Para obtener más información, consulte LLMs en la documentación del CrewAI.
- Integración de la API LLM: integración flexible con múltiples interfaces de servicios LLM, incluida Amazon BedrockOpenAI, e implementaciones de modelos locales. Para obtener más información, consulte los ejemplos de configuración de proveedores en la documentación. CrewAI
- Soporte multimodal: capacidades emergentes para gestionar texto, imágenes y otras modalidades para interacciones integrales entre agentes autónomos. Para obtener más información, consulte Uso de agentes multimodales en la CrewAl documentación.

Cuándo se debe usar CrewAl

CrewAles especialmente adecuado para escenarios de agentes autónomos, que incluyen:

- Problemas complejos que se benefician de una experiencia especializada y basada en funciones que funcione de forma autónoma
- Proyectos que requieren la colaboración explícita entre varios agentes autónomos

Características clave de CrewAI

- Casos de uso en los que la descomposición de problemas en equipo mejora la resolución autónoma de problemas
- Escenarios que requieren una separación clara de las preocupaciones entre las diferentes funciones de los agentes autónomos

Enfoque de implementación para CrewAl

CrewAlproporciona un enfoque de implementación basado en roles de equipos de agentes de IA para las partes interesadas de la empresa, tal como se detalla en la sección Introducción en la CrewAl documentación. El marco permite a las organizaciones:

- Defina agentes autónomos especializados con funciones, objetivos y áreas de experiencia específicos.
- Asigne tareas a los agentes en función de sus capacidades especializadas.
- Establezca dependencias claras entre las tareas para crear flujos de trabajo estructurados.
- Organice la colaboración entre varios agentes para resolver problemas complejos.

Este enfoque basado en roles refleja las estructuras de los equipos humanos, lo que hace que los líderes empresariales lo entiendan e implementen de forma intuitiva. Las organizaciones pueden crear equipos autónomos con áreas de experiencia especializadas que colaboren para alcanzar los objetivos empresariales, de forma similar a como funcionan los equipos humanos. Sin embargo, el equipo autónomo puede trabajar de forma continua sin intervención humana.

Ejemplo real de CrewAl

AWS <u>ha implementado sistemas autónomos multiagente mediante CrewAl integrado con Amazon</u>

<u>Bedrock, como se detalla en el CrewAl estudio de caso publicado.</u> AWS y CrewAl desarrolló un marco seguro e independiente de los proveedores. La arquitectura CrewAl de código abierto «flujos y personal» se integra perfectamente con los modelos básicos, los sistemas de memoria y las barreras de cumplimiento de Amazon Bedrock.

Los elementos clave de la implementación incluyen:

 Planos y código abierto, y <u>diseños de referencia CrewAl publicados</u> que mapean CrewAl los agentes con los modelos y las herramientas de observabilidad de Amazon Bedrock. AWS También presentaron sistemas ejemplares, como un equipo de auditoría de AWS seguridad compuesto por varios agentes, flujos de modernización del código y automatización administrativa de bienes de consumo envasados (CPG).

- Integración de la pila de observabilidad: la solución incorpora la supervisión con Amazon
 CloudWatch y permite la trazabilidad y LangFuse la depuración desde la prueba de concepto hasta la producción. AgentOps
- Retorno de la inversión (ROI) demostrado: los primeros proyectos piloto muestran mejoras importantes: una ejecución un 70 por ciento más rápida para un proyecto de modernización de código de gran tamaño y una reducción de aproximadamente un 90 por ciento en el tiempo de procesamiento para un flujo administrativo de CPG.

Agentes de Amazon Bedrock

Amazon Bedrock Agents es un servicio totalmente gestionado que le permite crear y configurar agentes autónomos en sus aplicaciones. Puede organizar las interacciones entre los modelos básicos, las fuentes de datos, las aplicaciones de software y las conversaciones de los usuarios. Su enfoque simplificado para crear agentes no requiere aprovisionar capacidad, administrar la infraestructura ni escribir código personalizado.

Características principales de Amazon Bedrock Agents

Amazon Bedrock Agents incluye las siguientes funciones clave:

- Servicio totalmente gestionado: gestión completa de la infraestructura sin necesidad de aprovisionar capacidad ni gestionar los sistemas subyacentes. Para obtener más información, consulte <u>Automatizar tareas en su aplicación mediante agentes de IA</u> en la documentación de Amazon Bedrock.
- Desarrollo basado en API: defina y ejecute agentes mediante sencillas llamadas a la API
 especificando modelos, instrucciones, herramientas y parámetros de configuración. Para obtener
 más información, consulte Crear y configurar el agente manualmente en la documentación de
 Amazon Bedrock.
- Grupos de acciones: defina las acciones específicas que su agente puede realizar mediante la creación de grupos de acciones con esquemas de API. Para obtener más información, consulte <u>Utilizar grupos de acciones para definir las acciones que debe realizar su agente</u> en la documentación de Amazon Bedrock.
- Integración de la base de conocimientos: conéctese sin problemas a las bases de conocimiento de Amazon Bedrock para aumentar las respuestas de los agentes con los datos de su organización.

Agentes de Amazon Bedrock 11

Para obtener más información, consulte <u>Aumente la generación de respuestas para su agente con</u> la base de conocimientos en la documentación de Amazon Bedrock.

- Plantillas de solicitudes avanzadas: personalice el comportamiento de los agentes mediante
 plantillas de solicitudes para el preprocesamiento, la organización, la generación de respuestas a
 la base de conocimientos y el posprocesamiento. Para obtener más información, consulte Mejorar
 la precisión de los agentes mediante plantillas de mensajes avanzadas en Amazon Bedrock en la
 documentación de Amazon Bedrock.
- Rastreo y observabilidad: realice un seguimiento del proceso de step-by-step razonamiento del agente mediante las funciones de rastreo integradas. Para obtener más información, consulte el proceso de step-by-step razonamiento de un agente mediante el rastreo en la documentación de Amazon Bedrock.
- Control de versiones y alias: cree varias versiones de su agente y despliéguelas mediante alias para una implementación controlada. Para obtener más información, consulte <u>Implementación y</u> uso de un agente de Amazon Bedrock en su aplicación en la documentación de Amazon Bedrock.

Cuándo usar Amazon Bedrock Agents

Amazon Bedrock Agents es especialmente adecuado para escenarios de agentes autónomos, que incluyen:

- Organizaciones que desean una experiencia totalmente gestionada para crear e implementar agentes sin administrar la infraestructura
- Proyectos que requieren un rápido desarrollo y despliegue de agentes mediante la configuración en lugar de mediante el código
- Casos de uso que se benefician de una estrecha integración con otras capacidades de Amazon Bedrock, como Knowledge Bases y Guardrails
- Los equipos no cuentan con los recursos internos necesarios para crear agentes partiendo de cero, pero necesitan capacidades autónomas listas para la producción

Enfoque de implementación para Amazon Bedrock Agents

Amazon Bedrock Agents ofrece un enfoque de implementación basado en la configuración para las partes interesadas de la empresa. El servicio permite a las organizaciones:

 Defina los agentes mediante llamadas a la API AWS Management Console o a la API sin escribir código complejo.

- Cree grupos de acciones que especifiquen APIs las operaciones que el agente puede realizar.
- Connect bases de conocimiento para proporcionar información específica del dominio al agente.
- Pruebe e itere el comportamiento del agente a través de una interfaz visual.

Este enfoque gestionado permite a los equipos empresariales desarrollar y desplegar rápidamente agentes autónomos sin necesidad de una amplia experiencia técnica en el desarrollo de modelos de IA o en la gestión de infraestructuras.

Ejemplo real de Amazon Bedrock Agents

Una solución de operaciones financieras (FinOps) descrita en esta entrada de <u>AWS blog</u> utiliza el marco multiagente de Amazon Bedrock para crear un asistente de gestión de costes en la nube impulsado por la IA. El rentable modelo básico de Amazon Nova potencia la solución, en la que un agente FinOps supervisor central delega tareas en agentes especializados. Estos agentes recopilan y analizan los datos de AWS gastos mediante el uso AWS Cost Explorer y generan recomendaciones de ahorro de costes mediante el uso. AWS Trusted Advisor

El sistema incluye el acceso seguro de los usuarios a través de Amazon Cognito, una interfaz alojada en AWS Amplify, y grupos de AWS Lambda acción para realizar análisis y pronósticos en tiempo real. Los equipos financieros pueden hacer preguntas en lenguaje natural, como «¿Cuáles fueron mis costes en febrero de 2025?» El sistema responde con desgloses detallados, sugerencias de optimización y previsiones, todo ello dentro de una arquitectura escalable y sin servidores que se implementa mediante el uso. AWS CloudFormation

AutoGen

AutoGenes un marco de código abierto que fue lanzado inicialmente por. Microsoft AutoGense centra en habilitar agentes de IA autónomos conversacionales y colaborativos. Proporciona una arquitectura flexible para crear sistemas multiagente, con énfasis en las interacciones asincrónicas y basadas en eventos entre los agentes para flujos de trabajo autónomos complejos.

Características clave de AutoGen

AutoGenproporciona las siguientes características clave:

• Agentes conversacionales: se basan en conversaciones en lenguaje natural entre agentes autónomos, lo que permite un razonamiento sofisticado a través del diálogo. Para obtener

más información, consulte el <u>marco de conversación entre múltiples agentes</u> en la AutoGen documentación.

- Arquitectura asíncrona: diseño basado en eventos para interacciones de agentes autónomos sin bloqueo, que admite flujos de trabajo paralelos complejos. Para obtener más información, consulte Resolución de varias tareas en una secuencia de chats asíncronos en la documentación. AutoGen
- H uman-in-the-loop Se apoya firmemente la participación humana opcional en flujos de trabajo de agentes que, de otro modo, serían autónomos cuando sea necesario. Para obtener más información, consulte <u>Permitir la retroalimentación humana en los agentes</u> en la AutoGen documentación.
- Generación y ejecución de código: capacidades especializadas para agentes autónomos centrados en el código que pueden escribir y ejecutar código. Para obtener más información, consulte la sección Ejecución de código en la AutoGen documentación.
- Comportamientos personalizables: configuración flexible de agentes autónomos y control de conversaciones para diversos casos de uso. Para obtener más información, consulte agentchat.conversable_agent en la documentación. AutoGen
- Selección de modelos de base: Support para varios modelos de base, incluidos los modelos Anthropic Claude, Amazon Nova (Premier, Pro, Lite y Micro) en Amazon Bedrock y otros para diferentes capacidades de razonamiento autónomo. Para obtener más información, consulte Configuración de LLM en la AutoGen documentación.
- Integración de la API LLM: configuración estandarizada para múltiples interfaces de servicios LLM, incluidas Amazon Bedrock, yOpenAI. Azure OpenAI Para obtener más información, consulte oai.openai_utils en la referencia de la API. AutoGen
- Procesamiento multimodal: Support para el procesamiento de texto e imágenes para permitir ricas interacciones multimodales entre agentes autónomos. Para obtener más información, consulte Uso de modelos multimodales: GPT-4V en la documentación. AutoGen AutoGen

Cuándo se debe usar AutoGen

AutoGenes especialmente adecuado para escenarios de agentes autónomos, que incluyen:

- Aplicaciones que requieren flujos de conversación naturales entre agentes autónomos para un razonamiento complejo
- Proyectos que requieren un funcionamiento totalmente autónomo y capacidades opcionales de supervisión humana

Cuándo se debe usar AutoGen 14

- Casos de uso que implican la generación, ejecución y depuración de código autónomas sin intervención humana
- Escenarios que requieren patrones de comunicación entre agentes autónomos, asíncronos y flexibles

Enfoque de implementación para AutoGen

AutoGenproporciona un enfoque de implementación conversacional para las partes interesadas de la empresa, tal como se detalla en <u>Primeros pasos</u> en la AutoGen documentación. El marco permite a las organizaciones:

- Cree agentes autónomos que se comuniquen a través de conversaciones en lenguaje natural.
- Implemente interacciones asincrónicas y basadas en eventos entre varios agentes.
- Combine un funcionamiento totalmente autónomo con la supervisión humana opcional cuando sea necesario.
- Desarrolle agentes especializados para diferentes funciones empresariales que colaboren a través del diálogo.

Este enfoque conversacional hace que el razonamiento del sistema autónomo sea transparente y accesible para los usuarios empresariales. Los responsables de la toma de decisiones pueden observar el diálogo entre los agentes para comprender cómo se llega a las conclusiones y, opcionalmente, participar en la conversación cuando se requiere el juicio humano.

Ejemplo del mundo real de AutoGen

Magentic-Onees un sistema multiagente generalista y de código abierto diseñado para resolver de forma autónoma tareas complejas y de varios pasos en diversos entornos, tal y como se describe en el blog Al Frontiers. Microsoft En esencia, está el agente Orchestrator, que descompone los objetivos de alto nivel y realiza un seguimiento del progreso mediante libros de contabilidad estructurados. Este agente delega las subtareas en agentes especializados (comoWebSurfer,, yComputerTerminal) y se adapta de forma dinámica FileSurfer Coder replanificándolas cuando es necesario.

El sistema se basa en la AutoGen estructura y es independiente del modelo; por defecto, es GPT-4o. Logra un rendimiento de última generación en puntos de referencia como, y todo ello sin necesidad de ajustes específicos para cada tarea. GAIA AssistantBench WebArena Además, apoya la extensibilidad modular y la evaluación rigurosa mediante sugerencias. AutoGenBench

Comparación de los marcos de IA de las agencias

Al seleccionar un marco de inteligencia artificial para el desarrollo de agentes autónomos, tenga en cuenta cómo se ajusta cada opción a sus requisitos específicos. Tenga en cuenta no solo sus capacidades técnicas, sino también su capacidad organizativa, incluida la experiencia del equipo, la infraestructura existente y los requisitos de mantenimiento a largo plazo. Muchas organizaciones podrían beneficiarse de un enfoque híbrido, que aproveche múltiples marcos para diferentes componentes de su ecosistema de IA autónomo.

En la siguiente tabla se comparan los niveles de madurez (más sólidos, adecuados o débiles) de cada marco en función de las dimensiones técnicas clave. Para cada marco, la tabla también incluye información sobre las opciones de implementación en producción y la complejidad de la curva de aprendizaje.

Plataform a	AWS integrati on	Soporte multiagen te autónomo	ad del flujo de	Capacidad es multimoda les	del	Integraci ón de la API LLM	Despliegu e de producció n	de
Amazon BedrockAg ents	Más gfuerte	Adecuada	Adecuada	Fuerte	Fuerte	Fuerte	Totalment e administr ado	Bajo
AutoGen	Débil	Fuerte	Fuerte	Adecuado	Adecuada	Fuerte	Hágalo usted mismo (DIY)	Empapado
CrewAl	Débil	Fuerte	Adecuado	Débil	Adecuado	Adecuada	BRICOLA	J E ∕loderado
LangChair / LangGrap h	n Adecuada	Fuerte	Más fuerte	Más fuerte	Más fuerte	Más fuerte	Plataform a o bricolaje	Acero
Strands Agents	Más fuerte	Fuerte	Más fuerte	Fuerte	Fuerte	Más fuerte	BRICOLA	J ⊠ oderado

Consideraciones a la hora de elegir un marco de IA para agencias

Al desarrollar agentes autónomos, tenga en cuenta los siguientes factores clave:

- AWS integración de la infraestructura: las organizaciones en las que se invierta mucho AWS se beneficiarán más de las integraciones nativas o de Strands Agents los flujos Servicios de AWS de trabajo autónomos. Para obtener más información, consulte el resumen AWS semanal (AWS blog).
- Selección del modelo de base: considere qué marco proporciona el mejor soporte para sus modelos de base preferidos (por ejemplo, los modelos Amazon Nova en Amazon Bedrock o Anthropic Claude), en función de los requisitos de razonamiento de su agente autónomo. Para obtener más información, consulte Cómo crear agentes eficaces en el sitio Anthropic web.
- Integración de la API LLM: evalúe los marcos en función de su integración con las interfaces de servicio del modelo de lenguaje grande (LLM) preferidas (por ejemplo, Amazon Bedrock oOpenAI) para la implementación en producción. Para obtener más información, consulte las <u>interfaces de</u> modelo en la documentación. Strands Agents
- Requisitos multimodales: para los agentes autónomos que necesitan procesar texto, imágenes y voz, tenga en cuenta las capacidades multimodales de cada marco. Para obtener más información, consulte Multimodalidad en la documentación. LangChain
- Complejidad del flujo de trabajo autónomo: los flujos de trabajo autónomos más complejos con una administración de estado sofisticada podrían favorecer las capacidades avanzadas de las máquinas de estados. LangGraph
- Colaboración autónoma en equipo: los proyectos que requieren una colaboración autónoma explícita y basada en roles entre agentes especializados pueden beneficiarse de la arquitectura orientada al equipo de. CrewAl
- Paradigma de desarrollo autónomo: los equipos que prefieran patrones conversacionales y asíncronos para los agentes autónomos podrían preferir la arquitectura basada en eventos de. AutoGen
- Enfoque gestionado o basado en código: las organizaciones que desean una experiencia totalmente gestionada con un mínimo de codificación deberían considerar Amazon Bedrock Agents. Las organizaciones que requieren una personalización más profunda pueden preferir Strands Agents otros marcos con capacidades especializadas que se ajusten mejor a los requisitos específicos de los agentes autónomos.
- Preparación para la producción de sistemas autónomos: considere las opciones de implementación, las capacidades de monitoreo y las funciones empresariales para los agentes autónomos de producción.

Protocolos de agencia

Los agentes de IA requieren protocolos de comunicación estandarizados para interactuar con otros agentes y servicios. Las organizaciones que implementan arquitecturas de agentes se enfrentan a importantes desafíos en torno a la interoperabilidad, la independencia de los proveedores y la preparación de sus inversiones para el futuro.

Esta sección le ayuda a navegar por el panorama de los agent-to-agent protocolos centrándose en los estándares abiertos que maximizan la flexibilidad y la interoperabilidad. (Para obtener información sobre agent-to-tool los protocolos, consulte <u>Estrategia de integración de herramientas</u> más adelante en esta guía).

En esta sección se destaca el Model Context Protocol (MCP), un estándar abierto desarrollado originalmente Anthropic en 2024. En la actualidad, apoya AWS activamente al MCP mediante contribuciones al desarrollo e implementación del protocolo. AWS colabora con los principales marcos de agentes de código abierto, incluidos LangGraph CrewAlLlamaIndex, y para dar forma al futuro de la comunicación entre agentes en el protocolo. Para obtener más información, consulte Protocolos abiertos para la interoperabilidad de los agentes, parte 1: Comunicación entre agentes en el MCP (blog).AWS

En esta sección

- Por qué es importante la selección de protocolos
- Agent-to-agent protocolos
- Selección de protocolos de agencia
- Estrategia de implementación de los protocolos de los agentes
- Cómo empezar con MCP

¿Por qué es importante la selección de protocolos

La selección de protocolos determina de manera fundamental la forma en que puede crear y evolucionar su arquitectura de agentes de IA. Al elegir protocolos que admitan la portabilidad entre marcos de agentes, obtiene la flexibilidad necesaria para combinar diferentes sistemas de agentes y flujos de trabajo para satisfacer sus necesidades específicas.

Los protocolos abiertos le permiten integrar agentes en varios marcos. Por ejemplo, utilícelos LangChain para la creación rápida de prototipos e implemente sistemas de producción que se

comuniquen a través de un protocolo común, como el MCP o el protocolo Agent2Agent (A2A). Strands Agents Esta flexibilidad reduce la dependencia de proveedores de IA específicos, simplifica la integración con los sistemas existentes y permite mejorar las capacidades de los agentes a lo largo del tiempo.

Los protocolos bien diseñados también establecen patrones de seguridad consistentes para la autenticación y la autorización en todo el ecosistema de agentes. Y lo que es más importante, la portabilidad de los protocolos preserva la libertad de adoptar nuevos marcos y capacidades de agentes a medida que vayan surgiendo. La elección de protocolos abiertos protege su inversión en el desarrollo de agentes y, al mismo tiempo, mantiene la interoperabilidad con sistemas de terceros.

Ventajas de los protocolos abiertos

Al implementar sus propias extensiones o crear sistemas de agentes personalizados, los protocolos abiertos ofrecen ventajas convincentes:

- Documentación y transparencia: normalmente proporcionan documentación completa e implementaciones transparentes
- Soporte comunitario: acceso a comunidades de desarrolladores más amplias para la solución de problemas y las mejores prácticas
- Garantías de interoperabilidad: mayor garantía de que sus extensiones funcionarán en diferentes implementaciones
- Compatibilidad futura: se reduce el riesgo de que se produzcan cambios importantes o de que queden obsoletos
- Influencia en el desarrollo: oportunidad de contribuir a la evolución del protocolo

Agent-to-agent protocolos

La siguiente tabla proporciona una descripción general de los protocolos de los agentes que permiten a varios agentes colaborar, delegar tareas y compartir información.

Protocolo	Ideal para	Consideraciones
Comunicación entre agentes MCP	Organizaciones que buscan patrones flexibles de colaboración entre agentes	 Una extensión del Protocolo de Contexto Modelo (MCP) propuesta por él AWS que se basa en su base de

		agent-to-agent comunicac ión existente • Permite una colaboración fluida entre los agentes con una seguridad OAuth basada en la seguridad
Protocolo A2A	Ecosistemas de agentes multiplataforma	 Respaldado por Google Estándar más nuevo con una adopción más limitada en comparación con el MCP
AutoGenmultiagente	Sistemas multiagente centrados en la investigación	 Respaldados por Microsoft Fuerte para interacciones complejas entre agentes
CrewAI	Equipos de agentes basados en roles	 Implementación independi ente Bueno para simular estructuras organizacionales

Decidir entre las opciones de protocolo

Al implementar la agent-to-agent comunicación, haga coincidir sus requisitos de comunicación específicos con las capacidades de protocolo adecuadas. Los diferentes patrones de interacción requieren diferentes características de protocolo. En la siguiente tabla se describen los patrones de comunicación más comunes y se recomiendan las opciones de protocolo más adecuadas para cada escenario.

Patrón	Descripción	La elección de protocolo ideal
Solicitud y respuesta sencillas	Interacciones puntuales entre agentes	MCP con flujos sin estado
Diálogos sensatos	Conversaciones continuas con el contexto	MCP con gestión de sesiones

Colaboración entre múltiples agentes	Interacciones complejas entre varios agentes	Interagente MCP o AutoGen
Flujos de trabajo basados en equipos	Equipos de agentes jerárquic os con funciones definidas	Interagente MCP, o CrewAl AutoGen

Más allá de los patrones de comunicación, varios factores técnicos y organizativos pueden influir en la selección del protocolo. En la siguiente tabla se describen las consideraciones clave que pueden ayudarle a evaluar qué protocolo se ajusta mejor a sus requisitos de implementación específicos.

Consideración	Descripción	Ejemplo
Modelo seguridad	Requisitos de autenticación y autorización	OAuth 2.0 en MCP
Entorno de despliegue	Donde los agentes correrán y se comunicarán	Máquina distribuida o única
Compatibilidad con los ecosistemas	Integración con los marcos de agentes existentes	LangChain o Strands Agents
Necesidades de escalabilidad	Crecimiento esperado de las interacciones entre los agentes	Capacidades de transmisión de MCP

Selección de los protocolos de los agentes

Para la mayoría de las organizaciones que crean sistemas de agentes de producción, el Model Context Protocol (MCP) ofrece la base de comunicación más completa y mejor respaldada. agent-to-agent MCP se beneficia de las contribuciones activas al desarrollo AWS y de la comunidad de código abierto.

Seleccionar los protocolos de agencia correctos es importante para las organizaciones que buscan implementar la IA de manera efectiva. Las consideraciones difieren según el contexto organizacional.

Consideraciones sobre el protocolo empresarial

Las empresas deberían considerar las siguientes medidas:

- Priorice los protocolos abiertos, como el MCP, para las implementaciones estratégicas de los agentes a largo plazo.
- Implemente capas de abstracción cuando utilice protocolos propietarios para facilitar las futuras migraciones.
- Participe en el desarrollo de estándares para influir en la evolución del protocolo.
- Considere los enfoques híbridos que utilizan protocolos abiertos para la infraestructura principal y protocolos propietarios para casos de uso específicos.

Consideraciones sobre los protocolos para empresas emergentes y pequeñas y medianas

Las empresas emergentes y small-to-medium (pequeñas y medianas) deberían considerar las siguientes medidas:

- Equilibre la velocidad y la flexibilidad comenzando con protocolos patentados bien compatibles para un desarrollo rápido.
- Planifique las rutas de migración para utilizar estándares más abiertos a medida que vayan surgiendo sus necesidades.
- Evalúe las tendencias de adopción de protocolos para evitar invertir en estándares en declive.
- Considere los servicios gestionados que resumen la complejidad del protocolo.

Consideraciones sobre los protocolos gubernamentales y de la industria regulada

El gobierno y las industrias reguladas deberían considerar las siguientes medidas:

- Haga hincapié en los estándares abiertos para garantizar el acceso a largo plazo y evitar la dependencia de los proveedores.
- Priorice los protocolos con modelos de seguridad y mecanismos de autenticación sólidos.
- Tenga en cuenta las implicaciones de los modelos de implementación remota y local para la soberanía de los datos.

• Documente las decisiones de protocolo para cumplir con los requisitos de cumplimiento y gobierno.

Estrategia de implementación de protocolos de agencia

Para implementar de manera efectiva los protocolos de los agentes en toda su organización, considere los siguientes pasos estratégicos:

- 1. Comience con la alineación de los estándares: adopte protocolos abiertos establecidos siempre que sea posible.
- 2. Cree capas de abstracción: implemente adaptadores entre sus sistemas y protocolos específicos.
- 3. Contribuya a los estándares abiertos: participe en las comunidades de desarrollo de protocolos.
- 4. Supervise la evolución de los protocolos: manténgase informado sobre las nuevas normas y actualizaciones.
- 5. Pruebe la interoperabilidad con regularidad: compruebe que sus implementaciones siguen siendo compatibles.

Cómo empezar con MCP

Para implementar el Model Context Protocol (MCP) en la arquitectura de su agente, lleve a cabo las siguientes acciones:

- 1. Explore las implementaciones del MCP en marcos como el SDK. Strands Agents
- 2. Revise la documentación técnica del Model Context Protocol.
- Lea los protocolos abiertos para la interoperabilidad de los agentes, parte 1: Comunicación entre agentes en el MCP (AWS blog), para obtener más información sobre la interoperabilidad de los agentes.
- 4. Únase a la comunidad de MCP para influir en la evolución del protocolo.

El MCP proporciona una capa de comunicación que permite a los agentes interactuar con datos y servicios externos y también se puede utilizar para permitir que los agentes interactúen con otros agentes. La implementación de <u>transporte HTTP Streamable</u> del protocolo ofrece a los desarrolladores un conjunto completo de patrones de interacción sin tener que reinventar la rueda. Estos patrones admiten tanto los request/response flujos sin estado como la gestión de sesiones con estado de forma persistente. IDs

Al adoptar protocolos abiertos como el MCP, usted posiciona a su organización para crear sistemas de agentes que sigan siendo flexibles, interoperables y adaptables a medida que la tecnología de IA evoluciona.

Cómo empezar con MCP 24

Herramientas

Los agentes de IA aportan valor al interactuar con herramientas y fuentes de datos externas para realizar tareas útiles. APIs La estrategia de integración de herramientas adecuada afecta directamente a las capacidades, la postura de seguridad y la flexibilidad a largo plazo del agente.

Esta sección le ayuda a navegar por el panorama de la integración de herramientas centrándose en los estándares abiertos que maximizan su libertad y flexibilidad. La sección destaca el <u>Protocolo de contexto modelo (MCP)</u> para la integración de herramientas y analiza las herramientas específicas del marco y las metaherramientas especializadas que mejoran los flujos de trabajo de los agentes.

En esta sección

- Categorías de herramientas
- Herramientas basadas en protocolos
- Herramientas nativas de Framework
- Metaherramientas
- Estrategia de integración de herramientas
- Mejores prácticas de seguridad para la integración de herramientas

Categorías de herramientas

La creación de sistemas de agentes implica tres categorías principales de herramientas.

Herramientas basadas en protocolos

<u>Las herramientas basadas en protocolos</u> utilizan protocolos estandarizados para la comunicación: agent-to-tool

- Herramientas MCP: herramientas estándar abiertas que funcionan en varios marcos con opciones de ejecución local y remota.
- OpenAlllamada a funciones: herramientas patentadas que son específicas de los OpenAl modelos.
- Anthropicherramientas: herramientas patentadas que son específicas de los modelos Anthropic Claude.

Categorías de herramientas 25

Herramientas nativas de Framework

Las herramientas nativas de Framework se integran directamente en marcos de agentes específicos:

- Strands Agents Pythonherramientas: herramientas ligeras quick-to-implement y específicas del marco. Strands Agents
- LangChainherramientas: herramientas Python basadas en herramientas que están estrechamente integradas con el LangChain ecosistema.
- LlamaIndexherramientas: herramientas que están optimizadas para la recuperación y el procesamiento internos LlamaIndex de datos.

Metaherramientas

<u>Las metaherramientas</u> mejoran los flujos de trabajo de los agentes sin tomar acciones externas directas:

- Herramientas de flujo de trabajo: administre el flujo de ejecución de los agentes, la lógica de ramificación y la administración del estado.
- Herramientas gráficas de agentes: coordine varios agentes en flujos de trabajo complejos.
- Herramientas de memoria: proporcionan almacenamiento y recuperación persistentes de la información en todas las sesiones de los agentes.
- Herramientas de reflexión: permiten a los agentes analizar y mejorar su propio rendimiento.

Herramientas basadas en protocolos

Al considerar las herramientas basadas en protocolos, el <u>Model Context Protocol (MCP)</u> proporciona la base más completa y flexible para la integración de herramientas. Como se indica en la entrada del <u>blog de código AWS abierto sobre la interoperabilidad de los agentes</u>, AWS ha adoptado el MCP como un protocolo estratégico y ha contribuido activamente a su desarrollo.

En la siguiente tabla se describen las opciones para el despliegue de la herramienta MCP.

Modelo de despliegue Descripción Ideal para Implementación

Basado en un estudio local	Las herramientas se ejecutan en el mismo proceso que el agente	Desarrollo, pruebas y herramientas sencillas	Rápida de implement ar sin sobrecarga de red
Basado en eventos enviados por el servidor local (SSE)	Las herramientas se ejecutan localmente pero se comunican a través de HTTP	Herramientas locales más complejas con separación de preocupaciones	Mejor aislamiento pero baja latencia
Basado en SSE remoto	Las herramientas se ejecutan en servidore s remotos	Entornos de producción y herramientas compartidas	Escalable y gestionad o de forma centraliz ada

El protocolo Model Context oficial SDKs está disponible para crear herramientas de MCP:

- PythonSDK: implementación integral con soporte completo de protocolos
- TypeScriptSDK:JavaScript/TypeScriptimplementación para aplicaciones web
- JavaSDK: implementación de Java para aplicaciones empresariales

Estos SDKs proporcionan los componentes básicos para crear herramientas compatibles con MCP en su idioma preferido, con implementaciones coherentes de la especificación del protocolo.

Además, AWS ha implementado el MCP en el SDK. Strands Agents El Strands Agents SDK proporciona una forma sencilla de crear y utilizar herramientas compatibles con el MCP. La documentación completa está disponible en el Strands Agents GitHub repositorio. Para casos de uso más sencillos o cuando se trabaja fuera del Strands Agents marco, el MCP oficial SDKs ofrece implementaciones directas del protocolo en varios idiomas.

Características de seguridad de las herramientas MCP

Las características de seguridad de las herramientas MCP incluyen las siguientes:

- OAuth Autenticación 2.0/2.1: autenticación estándar del sector
- Alcance de los permisos: control de acceso detallado para las herramientas
- Descubrimiento de la capacidad de la herramienta: descubrimiento dinámico de las herramientas disponibles

Gestión estructurada de errores: patrones de error consistentes

Cómo empezar con las herramientas de MCP

Para implementar el MCP para la integración de herramientas, lleve a cabo las siguientes acciones:

- 1. Explore el Strands AgentsSDK para obtener una implementación de MCP lista para la producción.
- 2. Revise la documentación técnica del MCP para comprender los conceptos básicos.
- 3. Utilice los ejemplos prácticos que se describen en esta entrada de blog de código AWS abierto.
- 4. Comience con herramientas locales sencillas antes de pasar a herramientas remotas.
- 5. Únase a la comunidad de MCP para influir en la evolución del protocolo.

Herramientas nativas de Framework

Si bien el <u>Model Context Protocol (MCP)</u> proporciona la base más flexible, las herramientas nativas del framework ofrecen ventajas para casos de uso específicos.

El <u>Strands AgentsSDK</u> ofrece herramientas Python basadas en herramientas que se caracterizan por su diseño liviano que requiere una sobrecarga mínima para operaciones sencillas. Permiten una implementación rápida y permiten a los desarrolladores crear herramientas con solo unas pocas líneas de código. Además, están estrechamente integrados para funcionar sin problemas dentro del Strands Agents marco.

El siguiente ejemplo demuestra cómo crear una herramienta meteorológica sencilla utilizandoStrands Agents. Los desarrolladores pueden transformar rápidamente Python las funciones en herramientas accesibles a los agentes con una sobrecarga de código mínima y generar automáticamente la documentación adecuada a partir de la cadena de documentos de la función.

```
#Example of a simple Strands native tool
@tool
def weather(location: str) -> str:
"""Get the current weather for a location""" #
Implementation here
```

return f"The weather in {location} is sunny."

Para la creación rápida de prototipos o para casos de uso sencillos, las herramientas nativas del framework pueden acelerar el desarrollo. Sin embargo, para los sistemas de producción, las herramientas MCP ofrecen una mejor interoperabilidad y flexibilidad en el futuro que las herramientas nativas del marco.

La siguiente tabla proporciona una descripción general de otras herramientas específicas del marco.

Plataforma	Tipo de herramienta	Ventajas	Consideraciones
AutoGen	Definiciones de funciones	Sólido soporte multiagente	Microsoftecosistema
LangChain	Pythonclases	Amplio ecosistem a de herramientas prediseñadas	Bloqueo de un marco
LlamaIndex	Funciones de Python	Optimizado para las operaciones de datos	Limitado a LlamaInde x

Meta-herramientas

Las metaherramientas no interactúan directamente con sistemas externos. En cambio, mejoran las capacidades de los agentes mediante la implementación de patrones de los agentes. En esta sección se analiza el flujo de trabajo, el gráfico de agentes y las metaherramientas de memoria.

Metaherramientas de flujo de trabajo

Las metaherramientas de flujo de trabajo gestionan el flujo de ejecución de los agentes:

- Gestión del estado: mantenga el contexto en las interacciones entre varios agentes
- Lógica de ramificación: habilite las rutas de ejecución condicionales
- · Mecanismos de reintento: gestione los errores con estrategias de reintento sofisticadas

Entre los ejemplos de marcos con metaherramientas de flujo de trabajo se incluyen LangGraphlas capacidades de flujo de trabajo. Strands Agents

Meta-herramientas 29

Metaherramientas Agent Graph

Las metaherramientas Agent Graph coordinan el trabajo conjunto de varios agentes:

- Delegación de tareas: asigne subtareas a agentes especializados
- Agregación de resultados: combine los resultados de varios agentes
- Resolución de conflictos: resuelva los desacuerdos entre los agentes

Los marcos <u>CrewAl</u>se especializan en la coordinación de gráficos de agentes <u>AutoGen</u>y se especializan en ella.

Metaherramientas de memoria

Las metaherramientas de memoria proporcionan almacenamiento y recuperación persistentes:

- Historial de conversaciones: mantenga el contexto en todas las sesiones
- Bases de conocimiento: almacene y recupere información específica del dominio
- Almacenes vectoriales: habilitan las capacidades de búsqueda semántica

El sistema de recursos de MCP proporciona una forma estandarizada de implementar metaherramientas de memoria que funcionan en diferentes marcos de agentes.

Estrategia de integración de herramientas

La estrategia de integración de herramientas que elijas repercute directamente en lo que tus agentes pueden conseguir y en la facilidad con la que puede evolucionar tu sistema. Priorice los protocolos abiertos, como el <u>Model Context Protocol (MCP)</u>, y utilice estratégicamente las metaherramientas y las herramientas nativas del marco. De esta forma, podrá crear un ecosistema de herramientas que siga siendo flexible y potente a medida que avance la tecnología de IA.

El siguiente enfoque estratégico para la integración de herramientas maximiza la flexibilidad y, al mismo tiempo, satisface las necesidades inmediatas de su organización:

1. Adopte el MCP como base: el MCP proporciona una forma estandarizada de conectar a los agentes con herramientas con sólidas funciones de seguridad. Comience con el MCP como su protocolo de herramientas principal para:

- Herramientas estratégicas que se utilizarán en múltiples implementaciones de agentes.
- Herramientas sensibles a la seguridad que requieren una autenticación y una autorización sólidas.
- Herramientas que necesitan ejecución remota en entornos de producción.
- 2. Utilice herramientas nativas del marco cuando sea apropiado: considere usar herramientas nativas del marco para:
 - Creación rápida de prototipos durante el desarrollo inicial.
 - Herramientas sencillas y no esenciales con requisitos de seguridad mínimos.
 - Funcionalidad específica del marco que aprovecha capacidades únicas.
- 3. Implemente metaherramientas para flujos de trabajo complejos: añada metaherramientas para mejorar la arquitectura de sus agentes:
 - Comience de forma sencilla con patrones de flujo de trabajo básicos.
 - Añada complejidad a medida que vayan madurando sus casos de uso.
 - Estandarice las interfaces entre los agentes y las metaherramientas.
- 4. Planifique para la evolución: construya pensando en la flexibilidad del futuro:
 - Documente las interfaces de las herramientas independientemente de las implementaciones.
 - Cree capas de abstracción entre los agentes y las herramientas.
 - Establezca rutas de migración de protocolos propietarios a protocolos abiertos.

Mejores prácticas de seguridad para la integración de herramientas

La integración de herramientas afecta directamente a su postura de seguridad. En esta sección se describen las prácticas recomendadas que debe tener en cuenta para su organización.

Autenticación y autorización

Utilice los siguientes controles de acceso robustos:

- Utilice la OAuth versión 2.0/2.1: implemente la autenticación estándar del sector para las herramientas remotas.
- Implemente los privilegios mínimos: conceda a las herramientas solo los permisos que necesitan.
- Cambie las credenciales: actualice periódicamente las claves de API y los tokens de acceso.

Protección de los datos

Para ayudar a proteger los datos, adopta las siguientes medidas:

- Valide las entradas y salidas: implemente la validación del esquema para todas las interacciones entre herramientas.
- Cifre los datos confidenciales: utilice TLS para todas las comunicaciones remotas con las herramientas.
- Implemente la minimización de datos: transfiera solo la información necesaria a las herramientas.

Supervisión y auditoría

Mantenga la visibilidad y el control mediante el uso de estos mecanismos:

- Registre todas las invocaciones de herramientas: mantenga registros de auditoría exhaustivos.
- Supervise en busca de anomalías: detecte patrones de uso inusuales de las herramientas.
- Implemente la limitación de velocidad: evite el abuso mediante el uso excesivo de herramientas.

El modelo de seguridad MCP aborda estas preocupaciones de manera integral. Para obtener más información, consulte Consideraciones de seguridad en la documentación del MCP.

Protección de los datos 32

Conclusión

El panorama de la IA de los agentes sigue evolucionando rápidamente y ofrece a las organizaciones nuevas y poderosas formas de crear sistemas inteligentes y autónomos. Esta guía ha explorado tres componentes esenciales para una implementación exitosa: los marcos que proporcionan la base, los protocolos que permiten la comunicación y las herramientas que amplían las capacidades.

A medida que los marcos vayan madurando, cabe esperar una mayor interoperabilidad, una estandarización en torno a protocolos como <u>el Model Context Protocol (MCP)</u> y capacidades de orquestación más sofisticadas para los agentes autónomos. Las organizaciones que adquieran experiencia en estos marcos en la actualidad estarán bien posicionadas para crear agentes cada vez más autónomos e inteligentes que ofrezcan un valor empresarial significativo.

La elección de los protocolos de los agentes representa una decisión estratégica que equilibra las necesidades de desarrollo inmediatas con la flexibilidad y la interoperabilidad a largo plazo. Al dar prioridad a los protocolos abiertos y crear las capas de abstracción adecuadas, las organizaciones pueden crear sistemas de agentes que se adapten a las tecnologías en evolución y, al mismo tiempo, cumplan con los requisitos empresariales actuales.

Para la mayoría de las organizaciones, el MCP representa una base sólida debido a su estándar abierto, su creciente ecosistema, su compatibilidad con los patrones de agent-to-agent comunicación y sus capacidades de integración de herramientas. AWS ha adoptado el MCP como un protocolo estratégico y ha contribuido activamente a su desarrollo e implementación en servicios como el Strands Agents SDK. Al utilizar MCP junto con las herramientas y metaherramientas nativas del marco adecuadas, puede crear sistemas de agentes que ofrezcan un valor inmediato y, al mismo tiempo, se adapten a las innovaciones futuras.

Recursos

Utilice los siguientes AWS y otros recursos relacionados con el desarrollo de agentes autónomos.

AWS Blogs

- Mejores prácticas para crear aplicaciones sólidas de IA generativa con Amazon Bedrock Agents —
 Parte 1
- Mejores prácticas para crear aplicaciones sólidas de IA generativa con Amazon Bedrock Agents —
 Parte 2
- Presentamos un Strands Agents SDK de código abierto para agentes de IA
- Protocolos abiertos para la interoperabilidad de los agentes, parte 1: Comunicación entre agentes en el MCP
- AWS Transform para .NET, el primer servicio de inteligencia artificial para modernizar las aplicaciones.NET a gran escala
- AWS Resumen semanal: Strands Agents

AWS Guía prescriptiva

- Operacionalización de la IA de los agentes en AWS
- Fundamentos de la IA agencial en AWS
- Los patrones y flujos de trabajo de la IA de los agentes están activos AWS
- Creación de arquitecturas sin servidor para la IA de los agentes en AWS
- · Creación de arquitecturas multiusuario para la IA de los agentes en AWS
- Recupere las opciones y arquitecturas de generación aumentada en AWS

AWS recursos

- Documentación de Amazon Bedrock
- Documentación de Amazon Nova
- AWS Servidores MCP () GitHub

AWS Blogs 34

Otros recursos

- AutoGendocumentación () Microsoft
- Creación de agentes eficaces (Anthropic)
- CrewAl GitHubrepositorio
- Documentación de LangChain
- LangGraphplataforma
- · Documentación sobre el protocolo Model Context
- Documentación de Strands Agents
- Strands AgentsDescripción general de las herramientas
- Strands AgentsGuía de inicio rápido

Otros recursos 35

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 <u>notificaciones RSS</u>.

Cambio	Descripción	Fecha
Publicación inicial	_	14 de julio 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.

Α

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 de agregación

Función SQL que opera en un grupo de filas y calcula un único valor de retorno para el grupo. Entre los ejemplos de funciones agregadas se incluyen SUM yMAX.

IΑ

Véase inteligencia artificial.

AIOps

Consulte las operaciones de inteligencia artificial.

Ā 38

anonimización

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

antipatrones

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

control de aplicaciones

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 AlOps se utiliza en la estrategia de AWS migración, consulte la <u>guía de integración de operaciones</u>.

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.

Ā 39

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 <u>Página web de AWS CAF</u> y el <u>Documento técnico de AWS CAF</u>.

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

 $\overline{\mathsf{A}}$

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.

В

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 <u>endianness</u>. 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.

B 41

bot

Una 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 <u>bots</u> que están infectadas por <u>malware</u> 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 <u>Implemente procedimientos de rotura de cristales en la guía Well-Architected</u> 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.

B 42

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 <u>Organizado en torno a las capacidades empresariales</u> del documento técnico <u>Ejecutar microservicios en contenedores en AWS</u>.

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 se tiene confianza, se despliega la nueva versión y se 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 <u>AWS Fault Injection Service (AWS FIS)</u> para realizar experimentos que estresen tus AWS cargas de trabajo y evalúen su respuesta.

C 43

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 <u>publicaciones de CCo E</u> 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 <u>Creación de su modelo operativo de nube</u>.

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 CCo E, establecer un modelo de operaciones)
- · Migración: migración de aplicaciones individuales
- Reinvención: optimización de productos y servicios e innovación en la nube

C 44

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

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 oBitbucket 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 <u>IA</u> que utiliza el aprendizaje automático para analizar y extraer información de formatos visuales, como imágenes y vídeos digitales. Por ejemplo, Amazon SageMaker Al 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

C 45

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 <u>paquetes de conformidad</u> 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 se describe comúnmente como una canalización. CI/CD puede ayudarlo a automatizar los procesos, mejorar la productividad, mejorar la calidad del código y entregar más rápido. Para obtener más información, consulte Beneficios de la entrega continua. CD también puede significar implementación continua. Para obtener más información, consulte Entrega continua frente a implementación continua.

CV

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.

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 <u>esquema en estrella</u>, 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 <u>desastre</u>. Para obtener más información, consulte <u>Recuperación</u> <u>ante desastres de cargas de trabajo en AWS: Recovery in the Cloud in the AWS Well-Architected</u> 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 desviaciones

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

DVSM

Consulte el mapeo del flujo de valor del desarrollo.

Ε

EDA

Consulte el análisis exploratorio de datos.

EDI

Véase intercambio electrónico de datos.

E 5

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 <u>la computación en nube</u>, <u>la computación</u> 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 bigendianos 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.

E 51

cifrado de sobre

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

entorno

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

- entorno de desarrollo: instancia de una aplicación en ejecución que solo se encuentra disponible para el equipo principal responsable del mantenimiento de la aplicación. Los entornos de desarrollo se utilizan para probar los cambios antes de promocionarlos a los entornos superiores. Este tipo de entorno a veces se denomina entorno de prueba.
- entornos inferiores: todos los entornos de desarrollo de una aplicación, como los que se utilizan para las compilaciones y pruebas iniciales.
- entorno de producción: instancia de una aplicación en ejecución a la que pueden acceder los usuarios finales. En un CI/CD proceso, el entorno de producción es el último entorno de implementación.
- entornos superiores: todos los entornos a los que pueden acceder usuarios que no sean del equipo de desarrollo principal. Esto puede incluir un entorno de producción, entornos de preproducción y entornos para las pruebas de aceptación por parte de los usuarios.

epopeya

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

PERP

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

E 52

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 <u>esquema en forma de estrella</u>. 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 <u>Límites de AWS aislamiento</u> 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 <u>Interpretabilidad del modelo de aprendizaje automático con AWS</u>.

F 53

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 <u>LLM</u> 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. <u>Consulte también el apartado de mensajes sin intervención.</u>

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 <u>captura de datos modificados</u> 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.

FΜ

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 <u>IA</u> 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 <u>Qué es la IA generativa</u>.

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 <u>la sección</u> 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 <u>flujo de trabajo basado en enlaces troncales</u> 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 <u>implementación sobre infraestructura existente</u>. Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de implementación desde cero.

G 55

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.

Н

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 <u>automático</u>. Puede utilizar los datos de reserva para evaluar el rendimiento del modelo comparando las predicciones del modelo con los datos de reserva.

H 56

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 identidades

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.

57

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, parchear 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 <u>arquitectura AWS de referencia de seguridad</u> 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.

Industria 4.0

Un término que <u>Klaus Schwab</u> 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 laC 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.

58

Internet de las cosas industrial (T) llo

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 <u>arquitectura AWS de referencia de seguridad</u> recomienda configurar su cuenta de red con entrada, salida e inspección VPCs para proteger la interfaz bidireccional entre la aplicación e Internet en general.

Internet de las cosas (IoT)

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

interpretabilidad

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

IoT

Consulte Internet de las cosas.

biblioteca de información de TI (ITIL)

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

administración de servicios de TI (ITSM)

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

ITIL

Consulte la biblioteca de información de TI.

I 59

ITSM

Consulte Administración de servicios de TI.

ı

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 <u>IA</u> 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. <u>Para obtener más información, consulte</u> Qué son. LLMs

migración grande

Migración de 300 servidores o más.

LBAC

Consulte el 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 <u>Aplicar permisos de privilegio mínimo</u> en la documentación de IAM.

migrar mediante lift-and-shift

Ver 7 Rs.

L 6

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 los ajustes necesarios. Un mecanismo es un ciclo que se refuerza y mejora a sí mismo a medida que funciona. Para obtener más información, consulte <u>Creación de mecanismos</u> 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 <u>Guía de preparación para la migración</u>. La MRA es la primera fase de la <u>estrategia de migración de AWS</u>.

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 <u>7 R</u> de este glosario y consulte <u>Movilice a su organización</u> 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 <u>Estrategia para modernizar</u> 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.

O 65

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.

O 66

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 <u>Guía</u> 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 del 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.

O 67

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 <u>arquitectura AWS de referencia de seguridad</u> 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.

P 68

policy

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

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 ofalse, 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 <u>Controles preventivos</u> en Implementación de controles de seguridad en AWS.

P 69

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 <u>control de seguridad</u> diseñado para evitar el despliegue de recursos que no cumplan con las normas. 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 <u>guía de referencia de controles</u> en la AWS Control Tower documentación y consulte <u>Controles</u> 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 un mensaje de <u>LLM</u> como entrada para el siguiente mensaje para generar mejores respuestas. Esta técnica se utiliza para dividir una tarea compleja en subtareas o para

P 70

refinar o ampliar de forma iterativa una respuesta preliminar. Ayuda a mejorar la precisión y la relevancia de las respuestas de un modelo y permite obtener resultados más detallados y personalizados.

seudonimización

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

publish/subscribe (pub/sub)

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

Q

plan de consulta

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

regresión del plan de consulta

El optimizador de servicios de la base de datos elige un plan menos óptimo que antes de un cambio determinado en el entorno de la base de datos. Los cambios en estadísticas, restricciones, configuración del entorno, enlaces de parámetros de consultas y actualizaciones del motor de base de datos PostgreSQL pueden provocar una regresión del plan.

R

Matriz RACI

Véase responsable, responsable, consultado, informado (RACI).

RAG

Consulte Retrieval Augmented Generation.

Q 71

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 y es 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).

R 72

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. <u>La alta disponibilidad</u> y la <u>recuperación ante desastres</u> son consideraciones comunes a la hora de planificar la resiliencia en el. Nube de AWS Para obtener más información, consulte <u>Nube de AWS Resiliencia</u>.

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.

R 73

retain

Consulte 7 Rs.

jubilarse

Ver 7 Rs.

Generación aumentada de recuperación (RAG)

Tecnología de <u>inteligencia artificial generativa</u> en la que un máster <u>hace referencia</u> 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 <u>secreto</u> 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

S 74

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 <u>Acerca de la federación basada en SAML 2.0</u> 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: <u>preventivos</u>, <u>de detección</u>, con <u>capacidad</u> de <u>respuesta</u> y <u>proactivos</u>.

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

S 75

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 <u>las políticas de control de servicios</u> 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.

S 76

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 datos grande para almacenar datos transaccionales o medidos 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 desmantelar el sistema heredado. Este patrón utiliza la analogía de una higuera que crece hasta convertirse en un árbol estable y, finalmente, se apodera y reemplaza a su host. El patrón fue presentado por Martin Fowler como una forma de gestionar el riesgo al reescribir sistemas monolíticos. Para ver un ejemplo 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 <u>Amazon CloudWatch Synthetics</u> para crear estas pruebas.

indicador del sistema

Una técnica para proporcionar contexto, instrucciones o pautas a un <u>LLM</u> para dirigir su comportamiento. Las indicaciones del sistema ayudan a establecer el contexto y las reglas para las interacciones con los usuarios.

Т

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.

T 78

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

T 79

por usted. Para obtener más información, consulte <u>AWS Organizations Utilización con otros AWS</u> 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 <u>Cuantificación de la incertidumbre en los sistemas de aprendizaje profundo</u>.

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.

U 80

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.

Interconexión con 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.

 $\overline{\mathsf{V}}$

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 <u>día cero</u>.

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 <u>LLM</u> instrucciones para realizar una tarea, pero sin ejemplos (imágenes) que puedan ayudar a guiarla. El LLM debe utilizar sus conocimientos previamente entrenados para

Z 82

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.

Z 83

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