

¿Qué es la ciencia de datos?

Cree una cuenta de AWS



Descubra ofertas de análisis gratuitas

Consulte las ofertas gratuitas de servicios de análisis en la nube



Eche un vistazo a los servicios de análisis

Innove más rápido con el conjunto más completo de servicios de análisis



Examine las capacitaciones de análisis

Empiece con la capacitación de análisis con contenido creado por expertos de AWS



Lea publicaciones de blog sobre análisis

Lea sobre las últimas novedades de los productos de análisis de AWS y las mejores prácticas

¿Qué es la ciencia de datos?

La ciencia de datos **es el estudio de datos con el fin de extraer información significativa para empresas**. Es un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación para analizar grandes cantidades de datos. Este análisis permite que los científicos de datos planteen y respondan a preguntas como “qué pasó”, “por qué pasó”, “qué pasará” y “qué se puede hacer con los resultados”.

¿Por qué es importante la ciencia de datos?

La ciencia de datos es importante **porque combina herramientas, métodos y tecnología para generar significado a partir de los datos**. Las organizaciones modernas están inundadas de datos; hay una proliferación de dispositivos que pueden recopilar y almacenar información de manera automática. Los sistemas en línea y los portales de pago capturan más datos en los campos del comercio electrónico, la medicina, las finanzas y cualquier otro aspecto de la vida humana. Disponemos de grandes cantidades de datos de texto, audio, video e imágenes.

Historia de la ciencia de datos

Aunque el término ciencia de datos no es nuevo, sus significados y connotaciones cambiaron con el tiempo. **La palabra apareció por primera vez en los años 60 como nombre alternativo de la estadística**. A finales de los 90, los profesionales de la computación formalizaron el término. Una propuesta de definición la consideraba un campo independiente con tres aspectos: diseño, recopilación y análisis de datos. Todavía tuvo que pasar otra década para que el término se utilizara fuera del ámbito académico.

Futuro de la ciencia de datos

La inteligencia artificial y las innovaciones del machine learning han hecho que el procesamiento de datos sea más rápido y eficiente. La demanda del sector ha creado un ecosistema de cursos, grados académicos y puestos de trabajo en el campo de la ciencia de datos.

Debido al conjunto de competencias multidisciplinares y a la experiencia necesaria, la ciencia de datos promete un fuerte crecimiento en las próximas décadas.

¿Para qué se utiliza la ciencia de datos?

La ciencia de datos se utiliza para estudiar los datos de cuatro maneras principales:

1. Análisis descriptivo

El análisis descriptivo **examina los datos para obtener información sobre lo que ha ocurrido u ocurre en el entorno de datos. Se caracteriza por las visualizaciones de datos, como los gráficos circulares, de barras o líneas, las tablas o las narraciones generadas.** Por ejemplo, un servicio de reserva de vuelos registra datos como el número de billetes reservados cada día. El análisis descriptivo revelará los picos y las caídas de las reservas, así como los meses de alto rendimiento del servicio.

2. Análisis de diagnóstico

El análisis de diagnóstico **es un examen profundo o detallado de datos para entender por qué ha ocurrido algo. Se caracteriza por técnicas como el análisis detallado, el descubrimiento y la minería de datos o las correlaciones. Se pueden llevar a cabo varias operaciones y transformaciones de datos en un determinado conjunto con el fin de descubrir patrones únicos en cada una de estas técnicas.** Por ejemplo, el servicio de vuelos podría hacer el análisis detallado de un mes con un rendimiento particularmente alto para entender mejor el pico de reservas. Esto puede revelar que muchos clientes visitan una determinada ciudad para asistir a un evento deportivo mensual.

3. Análisis predictivo

El análisis predictivo **utiliza los datos históricos para hacer previsiones precisas sobre los patrones de datos que pueden producirse en el futuro. Se caracteriza por técnicas como el machine learning, la previsión, la coincidencia de patrones y el modelado predictivo. En cada una de estas técnicas, se entrena a las computadoras para aplicar ingeniería inversa a las conexiones de causalidad en los datos.** Por ejemplo, el equipo de servicios de vuelo podría utilizar la ciencia de datos para predecir los patrones de reserva de vuelos del año siguiente al inicio de cada año. El programa o algoritmo de la computadora pueden examinar datos anteriores y predecir picos de reservas de determinados destinos en mayo. Al anticiparse a las futuras necesidades de viaje de los clientes, la empresa podría empezar desde febrero a hacer publicidad específica para esas ciudades.

4. Análisis prescriptivo

El análisis prescriptivo lleva los datos predictivos al siguiente nivel. **No solo predice lo que es probable que ocurra, sino que sugiere una respuesta óptima para ese resultado. Puede analizar las posibles implicaciones de las diferentes alternativas y recomendar el mejor curso de acción. Utiliza el análisis de gráficos, la simulación, el procesamiento de eventos complejos, las redes neuronales y los motores de recomendación del machine learning.**

De vuelta al ejemplo de la reserva de vuelos, el análisis prescriptivo podría examinar las campañas de marketing históricas para maximizar la ventaja del próximo pico de reservas. Un científico de datos podría proyectar los resultados de las reservas de diferentes niveles de gasto en varios canales de marketing. Estas previsiones de datos dan a la empresa de reserva de vuelos una mayor confianza en sus decisiones de marketing.

¿Cuáles son los beneficios de la ciencia de datos para las empresas?

La ciencia de datos revoluciona el funcionamiento de las empresas. Muchas empresas, independientemente de su tamaño, necesitan una sólida estrategia de ciencia de datos para impulsar el crecimiento y mantener una ventaja competitiva. Algunos beneficios clave son:

Descubrir patrones desconocidos de transformación

La ciencia de datos **permite a las empresas descubrir nuevos patrones y relaciones con el potencial de transformar la organización. Puede revelar cambios de bajo coste en la administración de recursos para obtener el máximo impacto en los márgenes de beneficio.** Por ejemplo, una empresa de comercio electrónico utiliza la ciencia de datos para descubrir que se generan demasiadas consultas de clientes fuera del horario comercial. Las investigaciones revelan que es más probable que los clientes compren si reciben una respuesta rápida en lugar de una respuesta al día siguiente. Al implementar un servicio de atención al cliente las 24 horas del día, los 7 días de la semana, la empresa aumenta sus ingresos en un 30 %.

Innovar con nuevos productos y soluciones

La ciencia de datos **puede revelar lagunas y problemas que de otro modo pasarían desapercibidos. Mejor información sobre las decisiones de compra, los comentarios de los clientes y los procesos empresariales puede impulsar la innovación en las operaciones internas y las soluciones externas.** Por ejemplo, una solución de pago en línea utiliza la ciencia de datos para cotejar y analizar los comentarios que hacen los clientes sobre la empresa en redes sociales. Los análisis revelan que los clientes olvidan las contraseñas durante los periodos de pico de compra y que no están satisfechos con el actual sistema de recuperación de contraseñas. **La empresa puede innovar para obtener una mejor solución y ver un aumento significativo en la satisfacción del cliente.**

Optimización en tiempo real

Para las empresas, en especial las grandes, es un gran reto responder en tiempo real a las condiciones cambiantes. Esto puede causar importantes pérdidas o interrupciones en la actividad empresarial. **La ciencia de datos puede ayudar a las empresas a predecir los cambios y reaccionar de forma óptima ante las distintas circunstancias.** Por ejemplo, una compañía de envíos que use camiones utiliza la ciencia de datos para reducir el tiempo de inactividad si los camiones se rompen. Identifican las rutas y los patrones de turnos que propician averías

más rápidas y ajustan los horarios de los camiones. Además, crean un inventario de piezas de repuesto comunes que se necesitan sustituir con frecuencia para que los camiones se puedan reparar con mayor rapidez.

¿Qué es el proceso de la ciencia de datos?

Un problema empresarial suele iniciar el proceso de la ciencia de datos. Un científico de datos trabajará con las partes interesadas del negocio para entender las necesidades del mismo. Una vez definido el problema, el científico de datos puede resolverlo con el proceso que consiste en obtener, depurar, explorar y modelar datos e interpretar los resultados (OSEMN):

Obtener datos

Los datos pueden ser preexistentes, recién adquiridos o un repositorio descargable de Internet. Los científicos de datos pueden extraerlos de las bases de datos internas o externas, del software CRM de la empresa, de los registros del servidor web, de las redes sociales o adquirirlos de terceros de confianza.

Depurar datos

La depuración o limpieza de datos consiste en el proceso de normalizarlos según un formato predeterminado. Incluye la gestión de los datos que faltan, la corrección de errores en estos y la eliminación de datos atípicos. Algunos ejemplos de la depuración de datos son:

- Cambiar todos los valores de fecha a un formato estándar común.
- Corregir las faltas de ortografía o los espacios adicionales.
- Corregir inexactitudes matemáticas o eliminar comas de números grandes.

Explorar datos

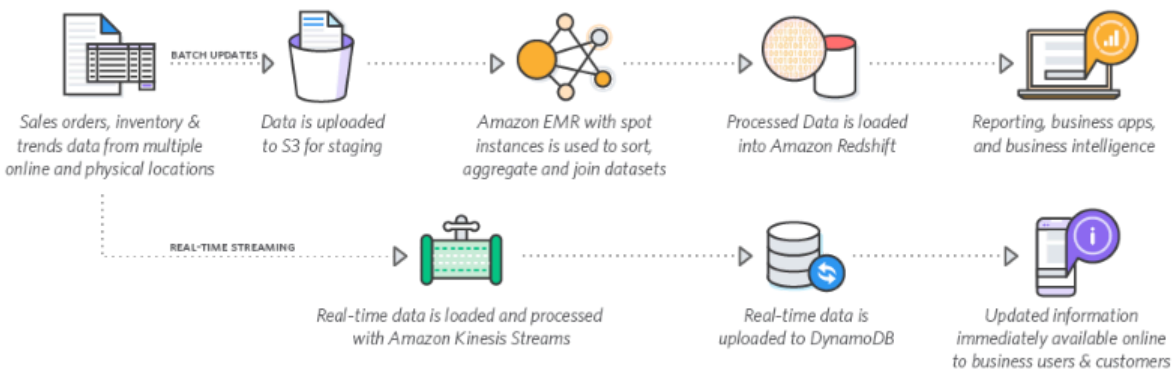
La exploración de datos es un análisis preliminar de estos que se utiliza para planificar otras estrategias para su modelado. Los científicos de datos obtienen una comprensión inicial de los datos mediante estadísticas descriptivas y herramientas de visualización de los mismos. A continuación, exploran los datos para identificar patrones interesantes que se puedan estudiar o utilizar.

Modelar datos

El software y los algoritmos de machine learning se utilizan para obtener información más profunda, predecir resultados y prescribir el mejor curso de acción. Las técnicas de machine learning, como la asociación, clasificación y agrupación, se aplican al conjunto de datos de entrenamiento. El modelo podría probarse con datos de prueba predeterminados para evaluar la precisión de los resultados. El modelo de datos se puede ajustar muchas veces para mejorar los resultados.

Interpretar los resultados

Los científicos de datos trabajan junto a los analistas y las empresas para convertir la información de datos en acción. Hacen diagramas, gráficos y tablas para representar tendencias y predicciones. La síntesis de datos ayuda a las partes interesadas a comprender y aplicar con eficacia los resultados.



¿Cuáles son las técnicas de la ciencia de datos?

Los profesionales de la ciencia de datos utilizan sistemas de computación para seguir el proceso de la ciencia de datos. Las principales técnicas que utilizan los científicos de datos son:

Clasificación

La clasificación consiste en ordenar los datos en grupos o categorías específicas. Las computadoras están entrenadas para identificar y ordenar datos. Los conjuntos de datos conocidos se utilizan para crear algoritmos de decisión en una computadora que procesa y categoriza rápidamente los datos. Por ejemplo:

- Ordenar los productos como populares o no populares.
- Ordenar las solicitudes de seguro como de alto o bajo riesgo
- Ordenar los comentarios en redes sociales como positivos, negativos o neutros.

Los profesionales de la ciencia de datos utilizan sistemas de computación para seguir el proceso de la ciencia de datos.

Regresión

La regresión es el método para encontrar una relación entre dos puntos de datos que aparentemente no se relacionan. La conexión se suele modelar en torno a una fórmula matemática y se representa en forma de gráfico o curvas. Cuando se conoce el valor de un punto de datos, se utiliza la regresión para predecir el otro punto de datos. Por ejemplo:

- La tasa de propagación de enfermedades transmitidas por aire.
- La relación entre la satisfacción del cliente y el número de empleados.
- La relación entre el número de estaciones de bomberos y el número de heridos por un incendio en un lugar determinado.

Agrupación

El método de clústeres consiste en agrupar datos estrechamente relacionados para buscar patrones y anomalías. El método de clústeres se diferencia del ordenamiento porque los datos no se pueden clasificar con precisión en categorías fijas. De ahí que los datos se agrupen en relaciones más probables. Con los clústeres se pueden descubrir nuevos patrones y relaciones. Por ejemplo:

- Agrupe clientes con un comportamiento de compra similar para mejorar el servicio.
- Agrupe el tráfico de red para identificar los patrones de uso diario o un ataque a la red de forma más rápida.
- Forme un clúster al agrupar los artículos en varias categorías de noticias diferentes y utilice esta información para encontrar contenidos de noticias falsas.

El principio básico de las técnicas de la ciencia de datos

Aunque los detalles varían, los principios subyacentes a estas técnicas son:

- Enseñar a una máquina a ordenar los datos a partir de un conjunto de datos conocido. Por ejemplo, las palabras clave de muestra se dan al ordenador con su valor de ordenación. “Feliz” es positivo, mientras que “odio” es negativo.
- Dar datos desconocidos a la máquina y permitir que el dispositivo ordene el conjunto de datos de forma independiente.
- Permitir inexactitudes en los resultados y manejar el factor de probabilidad de los mismos.

¿Cuáles son las diferentes tecnologías de la ciencia de datos?

Los profesionales de la ciencia de datos trabajan con tecnologías complejas como:

1. **Inteligencia artificial:** los modelos de machine learning y el software relacionado con este se utilizan para los análisis predictivos y prescriptivos.
2. **Computación en la nube:** las tecnologías en la nube ofrecen a los científicos de datos la flexibilidad y la capacidad de procesamiento necesarias para el análisis avanzado de datos.
3. **Internet de las cosas:** IoT se refiere a varios dispositivos que se pueden conectar de forma automática a Internet. Estos dispositivos recogen datos para iniciativas de la ciencia de datos. Generan datos masivos que se pueden utilizar para la minería y extracción de datos.
4. **Computación cuántica:** las computadoras cuánticas pueden realizar cálculos complejos a alta velocidad. Científicos de datos cualificados las utilizan para crear algoritmos cuantitativos complejos.

¿Cómo se compara la ciencia de datos con otros campos relacionados?

La ciencia de datos es un término que engloba otras funciones y campos relacionados con los datos. Veamos algunos ejemplos:

¿Cuál es la diferencia entre la ciencia de datos y el análisis de datos?

Aunque los términos se pueden usar de manera indistinta, el análisis de datos es un subconjunto de la ciencia de datos. La ciencia de datos es un término que engloba todos los aspectos del procesamiento de datos, desde la recopilación hasta el modelado y la información. Por otro lado, el análisis de datos se ocupa principalmente de la estadística, las matemáticas y el análisis estadístico. Solo se centra en el análisis de datos, mientras que la ciencia de datos está relacionada con el panorama general de los datos de la organización. En la mayoría de los lugares de trabajo, los científicos y analistas de datos trabajan juntos para alcanzar objetivos empresariales comunes. Un analista de datos puede dedicar más tiempo a los análisis rutinarios y proporcionar informes periódicos. Un científico de datos puede diseñar la forma de almacenar, manipular y analizar los datos. En pocas palabras, un analista da sentido a los datos existentes, mientras que un científico crea nuevos métodos y herramientas para procesarlos y que los usen los analistas.

¿Cuál es la diferencia entre ciencia de datos y el análisis empresarial?

Aunque ambos se superpongan entre sí, la diferencia clave consiste en el uso de la tecnología en cada campo. Los científicos de datos trabajan de manera más estrecha con la tecnología de datos que los analistas empresariales. Definen casos empresariales, recopilan información de las partes interesadas o validan soluciones. Por su parte, los científicos de datos usan la tecnología para trabajar con datos empresariales. Pueden escribir programas, aplicar técnicas de machine learning para crear modelos y desarrollar nuevos algoritmos. Los científicos de datos no solo entienden el problema, sino que también pueden crear una herramienta para solucionarlo. No es raro encontrar que los analistas empresariales y científicos de datos trabajan en el mismo equipo. Los analistas empresariales toman resultados de los científicos de datos y los utilizan para contar una historia que la empresa, en general, pueda entender.

¿Cuál es la diferencia entre la ciencia de datos y la ingeniería de datos?

Los ingenieros de datos crean y mantienen los sistemas que permiten a los científicos de datos acceder e interpretar los datos. Trabajan de manera más estrecha con la tecnología subyacente que un científico de datos. Por lo general, el rol implica la creación de modelos y canalizaciones de datos, así como la supervisión del servicio ETL (extracción, transformación y carga). Según la configuración y el tamaño de la organización, el ingeniero de datos también puede administrar la infraestructura relacionada, como el almacenamiento de macrodatos, el streaming y las plataformas de procesamiento como Simple Storage Service (Amazon S3). Los científicos utilizan los datos que los ingenieros han procesado para crear y entrenar modelos predictivos. A continuación, los científicos de datos pueden entregar los resultados a los analistas con vistas a la toma de decisiones.

¿Cuál es la diferencia entre la ciencia de datos y el machine learning?

El machine learning es la ciencia de entrenamiento de máquinas para que puedan analizar y aprender mediante datos, como lo hacen los humanos. Es uno de los métodos que se utilizan en los proyectos de ciencia de datos con el fin de obtener información automatizada de estos. Los ingenieros de machine learning se especializan en computación, algoritmos y habilidades de codificación específicas de los métodos de machine learning. Los científicos de datos pueden utilizar métodos de machine learning como herramientas o trabajar con otros ingenieros de machine learning para procesar los datos.

¿Cuál es la diferencia entre la ciencia de datos y la estadística?

La estadística es un campo con bases matemáticas que busca recopilar e interpretar datos cuantitativos. En cambio, la ciencia de datos es un campo multidisciplinario que utiliza métodos, procesos y sistemas científicos para extraer conocimientos a partir de los datos de maneras diversas. Los científicos de datos utilizan métodos de muchas disciplinas, incluida la estadística. Sin embargo, los campos difieren en sus procesos y los problemas que estudian.

¿Cuáles son las diferentes herramientas de la ciencia de datos?

AWS dispone de una serie de herramientas para apoyar a los científicos de datos de todo el mundo:

Almacenamiento de datos

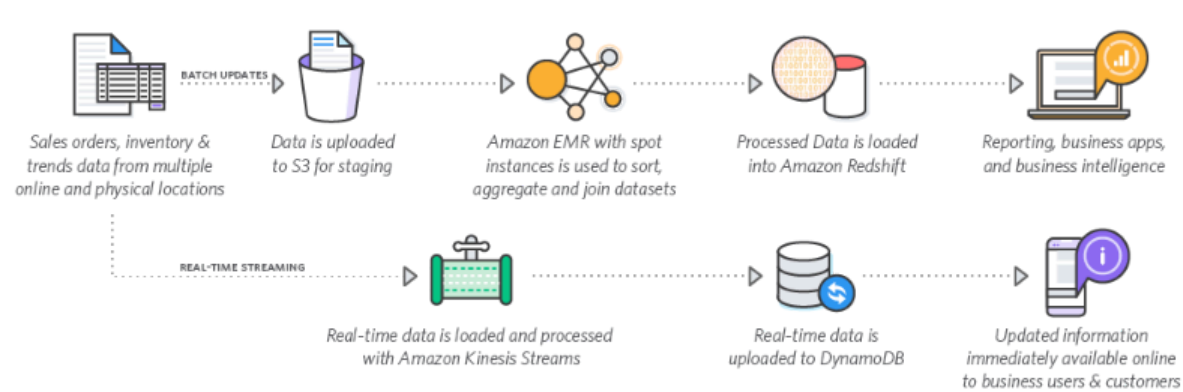
Para el almacenamiento de datos, Amazon Redshift puede ejecutar consultas complejas de datos estructurados o no estructurados. Los analistas y científicos de datos pueden utilizar AWS Glue para administrar y buscar datos. AWS Glue crea automáticamente un catálogo unificado de todos los datos del lago de datos, con metadatos asociados de modo que se puedan detectar.

Machine learning

Amazon SageMaker es un servicio de machine learning completamente administrado que se ejecuta en Amazon Elastic Compute Cloud (EC2). Permite a los usuarios organizar los datos, crear, entrenar y desplegar modelos de machine learning, así como escalar las operaciones.

Análisis

- Amazon Athena es un servicio de consultas interactivo que facilita el análisis de datos en Amazon S3 o Glacier. Es rápido, sin servidor y funciona con consultas SQL estándar.
- Amazon Elastic MapReduce (EMR) procesa macrodatos mediante servidores, como Spark y Hadoop.
- Amazon Kinesis permite la agregación y el procesamiento de datos de streaming en tiempo real. Utiliza secuencias de clics de sitios web, registros de aplicaciones y datos de telemetría de dispositivos del Internet de las cosas.
- Amazon OpenSearch permite buscar, analizar y visualizar petabytes de datos.



¿Qué hace un científico de datos?

Un científico de datos puede utilizar una serie de técnicas, herramientas y tecnologías diferentes como parte del proceso de la ciencia de datos. En función del problema, eligen las mejores combinaciones para obtener resultados más rápidos y precisos.

El rol y trabajo diario de un científico de datos varían en función del tamaño y las necesidades de la organización. Aunque suelen seguir el proceso de la ciencia de datos, los detalles pueden variar. En los equipos de ciencia de datos más grandes, un científico puede trabajar con otros analistas, ingenieros, expertos en machine learning y estadísticos para garantizar que el proceso de la ciencia de datos se siga de principio a fin y se alcancen los objetivos empresariales.

Sin embargo, en equipos más pequeños, un científico de datos puede cumplir varias funciones. En función de la experiencia, las aptitudes y la formación, pueden desempeñar varios roles que se superpongan entre sí. En este caso, sus responsabilidades diarias podrían incluir la ingeniería, el análisis y el machine learning, además de las metodologías básicas de la ciencia de datos.

¿A qué retos se enfrentan los científicos de datos?

Varios orígenes de datos

Los diferentes tipos de aplicaciones y herramientas generan datos en varios formatos. Los científicos tienen que limpiar y preparar los datos para que sean coherentes. Esto puede resultar tedioso y llevar mucho tiempo.

Entender el problema de la empresa

Los científicos de datos tienen que trabajar con varias partes interesadas y con administradores empresariales para definir el problema que se debe resolver. Esto puede suponer un reto, particularmente en empresas grandes que cuentan con múltiples equipos de trabajo con necesidades diferentes.

Eliminación del sesgo

Las herramientas de machine learning no son completamente precisas, por lo que puede existir cierta incertidumbre o sesgo. Los sesgos son desajustes en el comportamiento de las predicciones o los datos de entrenamiento del modelo entre diferentes grupos, como la edad o el nivel de ingresos. Por ejemplo, si una herramienta se entrena principalmente con datos de personas de mediana edad, puede ser menos preciso cuando se hagan predicciones que impliquen a personas más jóvenes o mayores. El ámbito del machine learning ofrece la oportunidad de abordar los sesgos detectándolos y midiéndolos en los datos y el modelo.

¿Cómo puedo convertirme en un científico de datos?

Por lo general, hay tres pasos para convertirse en científico de datos:

1. Obtener un título de grado en TI, ciencias de la computación, matemáticas, física u otro campo relacionado
2. Obtener una maestría en ciencia de datos o un campo relacionado
3. Obtener experiencia en un campo de interés

Próximos pasos de la ciencia de datos



Descubra otros recursos relacionados con el producto

[Más información sobre los lagos de datos y el análisis »](#)



Regístrese para obtener una cuenta gratuita

Obtenga acceso instantáneo a la capa gratuita de AWS.

[Registrarse »](#)



Comience a crear en la consola

Comience a crear con AWS en la consola de administración de AWS.

[Iniciar sesión »](#)

Información sobre AWS

[¿Qué es AWS?](#)

[¿Qué es la computación en la nube?](#)

Recursos para AWS

[Introducción](#)

[Formación técnica y certificación](#)

[Biblioteca de soluciones de AWS](#)

[Centro de arquitectura](#)

[Preguntas frecuentes sobre cuestiones técnicas y productos](#)

Desarrolladores en AWS

[Centro de desarrolladores](#)

[SDK y herramientas](#)

[.NET en AWS](#)

[Python en AWS](#)

[Java en AWS](#)

[Accesibilidad de AWS](#)

[Informes de analistas](#)

[PHP en AWS](#)

[¿Qué es DevOps?](#)

[Socios de AWS](#)

[JavaScript en AWS](#)

[¿Qué es un contenedor?](#)

[¿Qué es un lago de datos?](#)

[¿Qué es la inteligencia artificial \(IA\)?](#)

[¿Qué es la IA generativa?](#)

[¿Qué es el machine learning \(ML\)?](#)

[Seguridad en la nube de AWS](#)

[Novedades](#)

[Blogs](#)

[Notas de prensa](#)

Ayuda

[Contáctenos](#)

[Reciba ayuda de expertos](#)

[Abra un ticket de soporte técnico](#)

[AWS re:Post](#)

[Centro de conocimientos](#)

[Información general de AWS Support](#)

[Legal](#)

[Carreras en AWS](#)

[AWS en español](#)

[AWS en español](#)

[AWS en español](#)

[AWS en español](#)

[AWS en español](#)

[AWS en español](#)

[AWS en español](#)

Amazon es un empleador que ofrece igualdad de oportunidades: *minorías, mujeres, discapacitados, veteranos, identidad de género, orientación sexual y edad.*

Idioma

[عربي](#) |

[Bahasa Indonesia](#) |

[Deutsch](#) |

[English](#) |

[Español](#) |

[Français](#) |

[Italiano](#) |

[Português](#) |

[Tiếng Việt](#) |

[Türkçe](#) |

[Русский](#) |

[ไทย](#) |

[日本語](#) |

[한국어](#) |

[中文 \(简体\)](#) |

[中文 \(繁體\)](#)

Privacidad

|

Accesibilidad

|

Términos del sitio

|

Preferencias de cookies

|

© 2024, Amazon Web Services, Inc. o sus empresas afiliadas. Todos los derechos reservados.