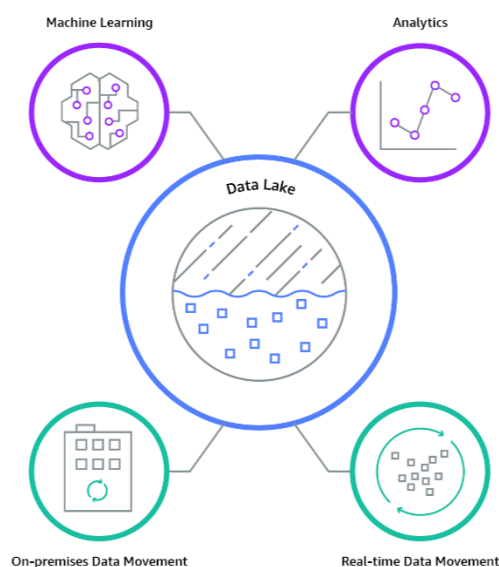


Lagos de datos análisis de big data AWS (Amazon Web Services) Lake Formation

¿Qué es un lago de datos?

Un lago de datos es un repositorio centralizado que le permite almacenar todos sus datos estructurados y no estructurados a cualquier escala. Puede almacenar sus datos sin modificarlos y sin tener que estructurarlos primero. También puede ejecutar diferentes tipos de análisis: desde paneles y visualizaciones hasta procesamiento de macrodatos, análisis en tiempo real y *machine learning* para tomar mejores decisiones.



Beneficios

Analice todos sus datos

Debido a que Amazon S3 escala de manera rentable, prácticamente sin límite, puede almacenar todos sus datos, desde cualquier fuente y desbloquear su valor.

Aumente la innovación

Con todos sus datos disponibles para el análisis, las organizaciones pueden acelerar la innovación, como descubrir nuevas oportunidades de ahorro o personalización. Se puede acceder a un continuo de datos más amplio para ML y análisis predictivo.

Utilice la mejor herramienta para el trabajo

Con los servicios de análisis de AWS especialmente diseñados, puede extraer rápidamente información de datos utilizando la herramienta más adecuada para el trabajo, optimizada para brindarle el mejor rendimiento, escala y costo para sus necesidades.

Elimine la administración del servidor

Con la mayoría de las opciones sin servidor para el análisis de datos en la nube, los servicios de análisis de AWS son fáciles de usar.

Lagos de datos en comparación con almacenamientos de datos: dos enfoques diferentes

Un almacenamiento de datos es una base de datos optimizada para analizar datos relacionales que provienen de sistemas transaccionales y aplicaciones de línea de negocio. La estructura y el esquema de los datos se definen con anticipación para optimizar y agilizar las consultas SQL, donde los resultados suelen usarse para informes y análisis operativos. Los datos se limpian, enriquecen y transforman para que puedan funcionar como el “único origen de información” en el que los usuarios puedan confiar.

Un lago de datos es diferente, ya que almacena datos relacionales de aplicaciones de línea de negocio y datos no relacionales de aplicaciones móviles, dispositivos de IoT y redes sociales. La estructura de los datos o el esquema no se definen cuando se capturan los datos. Esto significa que puede almacenar todos sus datos sin un diseño cuidadoso y sin saber qué preguntas tal vez tenga que responder en el futuro. Para descubrir información, se pueden utilizar diferentes tipos

de análisis sobre sus datos, como consultas SQL, análisis de macrodatos, búsqueda de texto completo, análisis en tiempo real y *machine learning*.

Características	Almacenamiento de datos	Lago de datos
Datos	Relacionales provenientes de sistemas transaccionales, bases de datos operativas y aplicaciones de línea de negocio	No relacionales y relacionales provenientes de dispositivos de IoT, sitios web, aplicaciones móviles, redes sociales y aplicaciones corporativas
Esquema	Diseñado con anterioridad a la implementación del almacenamiento de datos (esquema en escritura)	Escrito al momento del análisis (esquema en lectura)
Precio/desempeño	Resultados de búsqueda más rápidos con almacenamiento de mayor costo	Resultados de consultas que se tornan más rápidos con almacenamiento de bajo costo
Calidad de los datos	Datos seleccionados cuidadosamente que funcionan como la versión central de la verdad	Cualquier dato seleccionado o no (es decir, datos sin procesar)
Usuarios	Analistas de negocios	Científicos de datos, desarrolladores de datos y analistas de negocios (con datos seleccionados)
Análisis	Generación de informes en lotes, inteligencia empresarial y visualizaciones	<i>Machine learning</i> , análisis predictivo, detección de datos y creación de perfiles

Los elementos esenciales de un lago de datos

Migración de datos

Los lagos de datos le permiten importar cualquier cantidad de datos que pueda ingresar en tiempo real. Los datos se recopilan de múltiples orígenes y se transfieren al lago de datos en su formato original. Este proceso le permite escalar hasta alcanzar cualquier volumen de datos, a la vez que ahorra tiempo en la definición de estructuras de datos, esquemas y transformaciones.

Almacenamiento seguro y catalogación de los datos

Los lagos de datos le permiten almacenar datos relacionales, como los que surgen de bases de datos operativas y datos de aplicaciones de línea de negocio; y datos no relacionales, como los provenientes de aplicaciones móviles, dispositivos de IoT y redes sociales. También le brindan la capacidad de comprender qué datos hay en el lago a través del rastreo, la catalogación y la indexación de datos. Finalmente, los datos deben estar seguros para garantizar que sus activos de datos estén protegidos.

Análisis

Los lagos de datos permiten a distintos roles en su organización, como científicos de datos, desarrolladores de datos y analistas de negocios, acceder a los datos con las herramientas y los marcos de análisis que prefieran. Esto incluye marcos de código abierto, como Apache Hadoop, Presto y Apache Spark, además de ofertas comerciales de proveedores de almacenamiento de datos e inteligencia empresarial. Los lagos de datos le permiten ejecutar análisis sin tener que transferir sus datos a un sistema de análisis independiente.

Machine learning

Los lagos de datos permitirán a las organizaciones generar diferentes tipos de información, lo que incluye generar informes de datos históricos, aplicar *machine learning* donde se crean los modelos para predecir posibles resultados y sugerir un rango de acciones recomendadas para lograr el mejor resultado posible.

¿Por qué necesita un lago de datos?

Las organizaciones que generan valor empresarial de forma exitosa a partir de sus datos superarán a sus colegas. Una encuesta de Aberdeen notó que las organizaciones que implementaron lagos de datos superaron el nivel de rendimiento de empresas similares en un 9 % en cuanto al crecimiento orgánico de los ingresos. Estos líderes pudieron realizar nuevos tipos de análisis, como *machine learning* sobre nuevos orígenes, tales como archivos de registro, datos de secuencias de clics, redes sociales y dispositivos conectados a Internet almacenados en lagos de datos. Esto los ayudó a identificar oportunidades de crecimiento empresarial en menos tiempo y a actuar para aprovecharlas al atraer y retener clientes, potenciar la productividad, mantener dispositivos de manera proactiva y tomar decisiones informadas.