

AMAZON WEB SERVICES PROVEEDOR

Líder en servicios en la nube que tiene una amplia gama de servicios que están relacionados con ciencia de datos

INGESTA O RECOPIACIÓN DE DATOS

1



Amazon OpenSearch

- Capacidad de recopilar datos de diversas fuentes
- Variedad de herramientas incorporadas para transformar datos
- No requiere servidores
- Se ajusta automáticamente para manejar cargas de trabajo intensas
- Simplifica la gestión de flujos de datos complejos
- Enfoque en objetivos comerciales

Facilita la ingesta y procesamiento de datos desde diversas fuentes, lo que es esencial en Big Data y Data Science

Amazon Kinesis Data Firehose

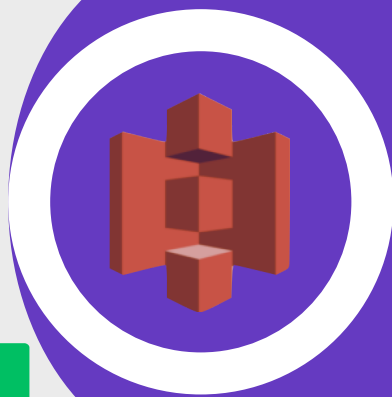
- Simplifica la conversión de datos en streaming
- Utiliza AWS Lambda
- Automatiza el procesamiento y carga de datos
- Proporciona plantillas Lambda listas para usarse
- Opciones para la recuperación de errores
- Copias de seguridad de datos sin procesar

Amazon Kinesis Data Firehose

Simplifica la conversión y carga de datos en tiempo real en Amazon OpenSearch, útil para el procesamiento de flujos de datos en tiempo real en aplicaciones de análisis.

ALMACENAR DATOS, BASES DE DATOS, DATA WAREHOUSE Y DATA LAKES

3



Amazon S3 (Simple Storage Service).

- Servicio de almacenamiento de objetos que ofrece escalabilidad, disponibilidad de datos, seguridad y rendimiento.
- Almacenar y salvaguardar cualquier cantidad de datos: data lakes, aplicaciones nativas en la nube o móviles.
- Almacenamiento rentables y herramientas de gestión intuitivas.
- Optimizar gastos, organizar datos y establecer controles de acceso para requisitos específicos.

Sirve como un componente para almacenar grandes volúmenes de datos, como en Data Lakes, lo que es esencial en Big Data y Data Science.

4



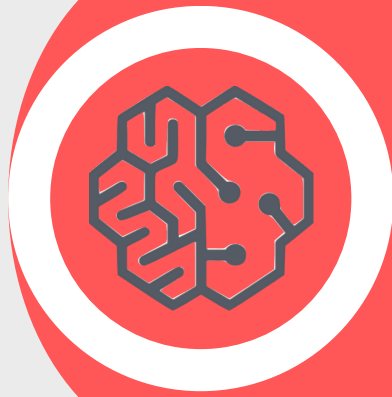
Amazon Redshift

- Analiza datos estructurados y semiestructurados en almacenamiento de datos, bases de datos operativas y lagos de datos.

Proporciona capacidades de análisis en data warehouses, vital para el almacenamiento y análisis de datos estructurados en Business Intelligence y Data Science.

HACER TRATAMIENTO DE DATOS COMO LIMPIEZA Y MINERÍA DE DATOS

5



Amazon SageMaker Data Wrangler.

- Prepara y limpia datos para aprendizaje automático.
- Interfaz visual para transformaciones sin codificación.
- Admite múltiples fuentes de datos como CSV y S3.
- Genera código Python a partir de transformaciones.
- Integración directa con flujos de trabajo de SageMaker.

Limpieza y preprocesamiento de datos, pasos en Data Science antes de crear modelos.

Jade Ullola Castro | Alexis Giovanni Valencia Medina
Donovan Gared Orozco García | Gustavo De Jesús Barriga
Mathieu | Jesús Alejandro Valencia Valadez

AMAZON WEB SERVICES PROVEEDOR

6

Amazon Redshift Spectrum

- Consulta datos en Amazon S3 desde Amazon Redshift.
- Herramienta para minería de datos almacenados en Amazon S3 mediante SQL
- Predicción de particiones para consultas eficientes.
- Ahorro de costos al usar almacenamiento en S3.
- Cambios en datos en tiempo real.
- Escalamiento y administración separados.
- Simplifica la administración de datos

Permite la minería de datos almacenados en Amazon S3, esencial para analizar datos en Data Lakes.

HACER MODELOS DE ANÁLISIS DE DATOS, LENGUAJES DE PROGRAMACIÓN, ACCESO A MODELOS EXISTENTES PARA USARLOS, FRAMEWORKS DE APOYO PARA DESARROLLAR PROYECTOS DE CIENCIA DE DATOS.

Entorno completo para la construcción y despliegue de modelos de aprendizaje automático, esencial en Data Science.



Amazon SageMaker

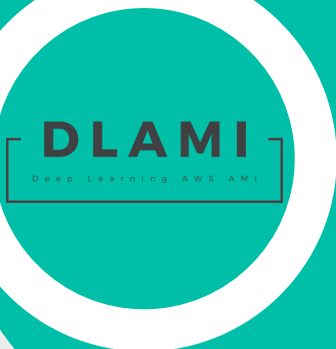
- Plataforma de ciencia de datos
- Asignando un entorno integrado con herramientas para selección de algoritmos
- Ajuste y evaluación de modelos
- Permite el despliegue sencillo de modelos entrenados para su uso en aplicaciones en tiempo real.

8

AWS Deep Learning AMIs.

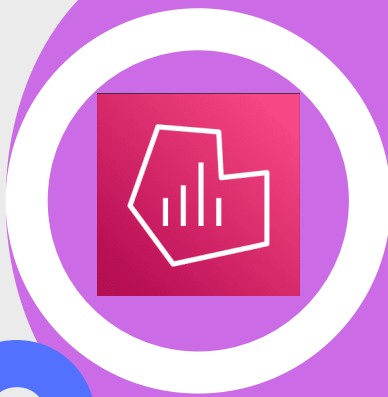
- Es una biblioteca de aprendizaje automático
- Da una colección de imágenes preconfiguradas con bibliotecas y frameworks para la creación y entrenamiento de modelos de aprendizaje profundo
- Ofrecer herramientas que ayudan a desarrollar proyectos de ciencias de datos y análisis avanzados de datos.

Herramientas y bibliotecas preconfiguradas fundamental en Data Science y Machine Learning.



HACER VISUALIZACIONES

Creación de visualizaciones para el análisis de métricas y registros, esencial en Business Intelligence y Data Science.



Amazon Managed Grafana

- Solución gestionada para Grafana, plataforma de análisis de código abierto.
- Grafana se usa para consultar y visualizar métricas de almacenamiento diverso.
- Amazon Managed Grafana se integra fácilmente con fuentes de datos de AWS.
- Ofrece visualizaciones preconfiguradas para un inicio rápido en análisis.
- Evita la necesidad de diseñar paneles desde cero.

10

Amazon QuickSight

- Solución de análisis empresarial en la nube.
- No requiere servidores.
- Proporciona representaciones visuales y paneles interactivos.
- Utiliza tecnología de aprendizaje automático.
- Revela conocimientos en los datos y hace pronósticos precisos.
- QuickSight responde a consultas en lenguaje natural usando aprendizaje automático.

Representaciones visuales de datos, lo que es clave en Business Intelligence.



Jade Ullola Castro | Alexis Giovanni Valencia Medina
Donovan Gared Orozco García | Gustavo De Jesús Barriga
Mathieu | Jesús Alejandro Valencia Valadez

AMAZON WEB SERVICES PROVEEDOR

PROCESAR DATOS NO ESTRUCTURADOS

Permite el análisis de imágenes y videos, fundamental en aplicaciones de visión artificial en Data Science.



Amazon Rekognition

- Agilidad en la incorporación de API de visión artificial a aplicaciones.
- Utiliza modelos preentrenados o personalizables según preferencias.
- No es necesario crear modelos de machine learning desde cero.
- Realiza análisis rápido de imágenes, flujos de datos en tiempo real y videos almacenados.
- Escalabilidad según demandas comerciales, gestionada por capacidades de IA.

12

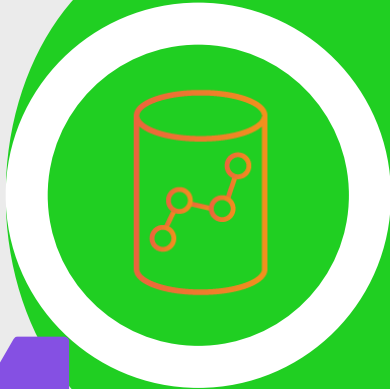
Amazon Comprehend

- Analiza contenido textual en diversos tipos de documentos.
- Extrae de forma automatizada texto, frases clave, temas y sentimientos.
- Simplifica el manejo de documentos.
- Crea un modelo capaz de categorizar documentos y reconocer términos específicos.
- No se requiere experiencia en machine learning.
- Garantiza seguridad y control de acceso, ocultando datos personales identificables.

Fácil extracción de información valiosa de texto no estructurado, como opiniones de clientes, en Data Science y Business Intelligence.

PROCESAMIENTO DE DATOS EN TIEMPO REAL (STREAMING)

Análisis en tiempo real de datos de streaming, útil en escenarios de análisis en Data Science y Business Intelligence.



Amazon Redshift Streaming Ingestion

- Analiza datos de streaming en tiempo real.
- Usa SQL familiar para hacer análisis valiosos y crear y gestionar flujos de ELT
- Procesa grandes cantidades de datos de streaming con baja latencia y alto rendimiento en segundos.

14

Kinesis Data Streams

- Permite capturar, procesar y almacenar de forma continua flujos de datos en tiempo real.
- Uso del framework de streaming que prefieras: librería de clientes de Amazon Kinesis (KCL), Apache Storm o Apache Spark Streaming, para trabajar con los datos.

Captura y procesa flujos de datos en tiempo real, esencial en aplicaciones de streaming en Data Science y Business Intelligence.

BIBLIOGRAFÍA

- Amazon Web Services. (2023). Análisis en AWS. Recuperado de: <https://aws.amazon.com/es/big-data/datalakes-and-analytics/>
- Amazon Web Services. (2023). Integraciones. Recuperado de: <https://aws.amazon.com/es/opensearch-service/features/integration/>
- Amazon Web Services. (2023). Amazon S3. Recuperado de: <https://aws.amazon.com/es/s3/?c=a&sec=srvn>
- Amazon Web Services. (2023). Amazon Redshift. Recuperado de: <https://aws.amazon.com/es/redshift/>
- Amazon Web Services. (2023). ¿Qué son los datos de streaming? Recuperado de: <https://aws.amazon.com/es/what-is/streaming-data/>
- Amazon Web Services. (2023). ¿Qué es la visualización de datos? Recuperado de: <https://aws.amazon.com/es/what-is/data-visualization/>
- Amazon Web Services. (2023). Automatización del procesamiento de datos no estructurados con Amazon SageMaker Recuperado de: <https://aws.amazon.com/startups/learn/automating-unstructured-data-processing-with-amazon-sagemaker?lang=es>
- Amazon Web Services. (2023). Amazon Rekognition Recuperado de: <https://aws.amazon.com/es/rekognition/?nc=sn&loc=0>
- Amazon Web Services. (2023). Amazon Comprehend. Recuperado de: <https://aws.amazon.com/es/comprehend/>
- Amazon Web Services. (2023). Amazon SageMaker Data Wrangler Recuperado de: <https://aws.amazon.com/sagemaker/data-wrangler/>
- Amazon Web Services. (2023). Amazon Redshift Spectrum Recuperado de: <https://aws.amazon.com/es/redshift/>
- Amazon Web Services. (2023). Deep Learning AMIs Recuperado de: <https://aws.amazon.com/es/machine-learning/amis/>