

RESPUESTAS ESPERADAS

Data Lakes

Data Lakes are consolidated, centralized storage areas for raw, unstructured, semi-structured, and structured data, taken from multiple sources and lacking a predefined schema. Data Lakes have been created to save data that "may have value." The value of data and the insights that can be gained from it are unknowns and can vary with the questions being asked and the research being done.

Data Lakes allow Data Scientists to mine and analyze large amounts of Big Data. Big Data, which was used for years without an official name, was labeled by Roger Magoulas in 2005. He was describing a large amount of data that seemed impossible to manage or research using the traditional SQL tools available at the time. Hadoop (2008) provided the search engine needed for locating and processing unstructured data on a massive scale, opening the door for Big Data research.

In October of 2010, James Dixon, founder of Pentaho, came up with the term "Data Lake."

Glosario: SQL: lenguaje de programación de dominio específico.

A. Traduzca las siguientes frases nominales.

1. consolidated, centralized storage areas: **áreas de almacenamiento centralizadas y consolidadas**
2. a large amount of data: **una gran cantidad de información/datos**
3. the traditional SQL tools available at the time: **las herramientas de SQL tradicionales disponibles en ese momento/en ese entonces**

B. Responda en castellano.

1. ¿Para qué se utilizan los lagos de datos y qué características presentan estos datos?

Se utilizan para guardar/almacenar gran cantidad de datos que pueden tener algún valor (o no). Se pueden almacenar datos sin procesar, estructurados, semiestructurados, o no estructurados que provienen de distintas fuentes y no tienen un esquema/patrón predefinido.

2. ¿Cómo se originó el término Big Data?

El término Big Data fue acuñado por Roger Magoulas en 2005 para referirse a una gran cantidad de información que parecía imposible de manejar o investigar utilizando las herramientas tradicionales de SQL disponibles en ese momento.

3. ¿Cuál fue el aporte del entorno Hadoop?

Hadoop proporcionó un motor de búsqueda para localizar y procesar datos no estructurados a gran escala y este hecho fue un primer paso/marcó un avance en la investigación de Big Data.

C. Traduzca el siguiente texto. (continuation)

In describing his concept of a Data Lake, he said:

"If you think of a Data Mart as a store of bottled water, the Data Lake is a large body of water in a more natural state. The contents of the Data Lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples."

Al describir su concepto de lago de datos, afirmó/dijo:

"Si piensa en el mercado de datos (Data Mart) como una tienda/un almacén de agua envasada/embotellada, el lago de datos es una gran masa de agua en un estado más natural. Los contenidos del lago de datos fluyen desde una fuente para llenar el lago, y varios usuarios de este lago pueden acercarse a analizar/examinar, tomar muestras o sumergirse en ellas."