

ARTICLE

Structured vs. Unstructured Data: An Overview

Keith D. Foote

Published: April 11, 2023

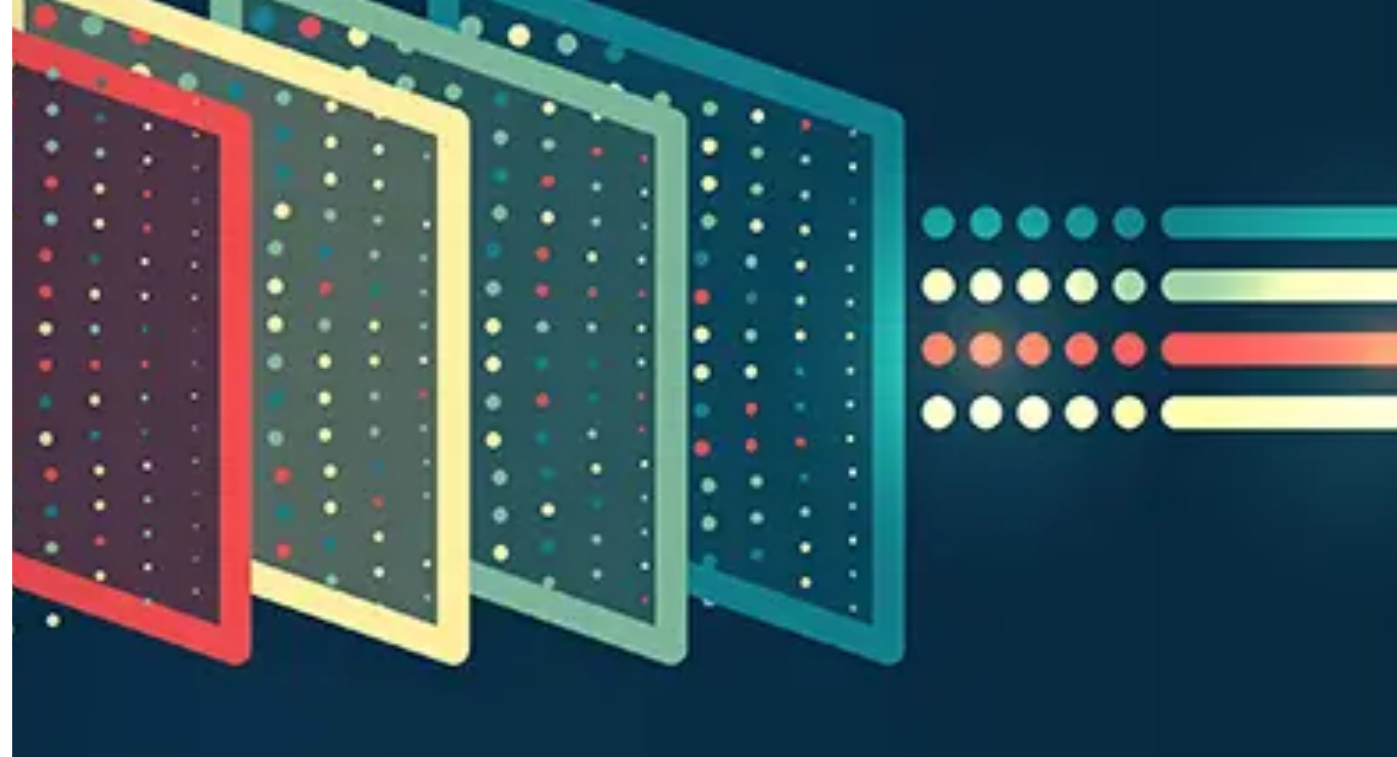


Table of Contents

→ The Expanded Use of the Internet and Unstructured Data

[The Benefits of Using Structured Data](#)[The Benefits of Using Unstructured Data](#)[The Future of Structured and Unstructured Data](#)

Structured data and unstructured data are both forms of data, but the first uses a single standardized format for storage, and the second does not. Structured data must be appropriately formatted (or reformatted) to provide a standardized data format before being stored, which is not a necessary step when storing unstructured data.

The relational database provides an excellent example of how structured data is used and stored. The data is normally formatted into specific fields (for example, credit card numbers or addresses), allowing the data to be easily found using SQL.

Non-relational databases, also called NoSQL, provide a way to work with unstructured data.

Edgar F. Codd invented relational databases (RDBMs) in 1970, and they became popular during the 1980s. Relational databases allow users to access data and write in SQL (Structured Query Language). RDBMs and SQL gave organizations the ability to analyze stored data on demand, providing a significant advantage against the competition of those times.

Relational databases are user-friendly, and very, very efficient at maintaining accurate records. Regrettably, they are also quite rigid and cannot work with other languages or data formats.

Unfortunately for [relational databases](#), during the mid-1990s, the internet gained significantly in popularity, and the rigidity of relational databases could not handle the variety of languages and formats that became accessible. This made research difficult, and NoSQL was developed as a solution between 2007 and 2009.

A NoSQL database translates data written in different languages and formats efficiently and quickly and avoids the rigidity of SQL. Structured data is often stored in relational databases and data warehouses, while unstructured data is often stored in [NoSQL](#) databases and data lakes.

For broad research, unstructured data used by NoSQL databases, compared to relational databases, are the better choice because of their speed and flexibility.

The Expanded Use of the Internet and Unstructured Data

During the late 1980s, the low prices of hard disks, combined with the development of data warehouses, resulted in remarkably inexpensive data storage. This, in turn, resulted in organizations and individuals embracing the habit of storing all data gathered from customers, and all the data collected from the internet for research purposes. A data warehouse allows analysts to [access research data](#) more quickly and efficiently.

Unlike a relational database, which is used for a variety of purposes, a data warehouse is specifically designed for a quick response to queries.

Data warehouses can be cloud-based, or part of a business's in-house mainframe server. They are compatible with SQL systems because by design, they rely on structured datasets. Generally speaking, [data warehouses](#) are not compatible with unstructured, or NoSQL, databases. Before the 2000s, businesses focused only on extracting and analyzing information from structured data.

The internet began to offer unique data analysis opportunities and data collections in the early 2000s. With the growth of web research and online shopping, businesses such as Amazon, Yahoo, and eBay began analyzing their customer's behavior by including such things as search logs, click-rates, and IP-specific location data. This abruptly opened up a whole new world of research possibilities. The profits resulting from their research prompted other organizations to begin their own expanded [business intelligence](#) research.

Data lakes came about as a way to deal with unstructured data in roughly 2015. Currently, [data lakes](#) can be set up both in-house and in the [cloud](#) (the cloud version eliminates in-house installation difficulties and costs). The advantages of moving a data lake from an in-house location to the cloud for analyzing unstructured data can include:

- **Cloud-based tools that are more efficient:** The tools available on the cloud can build data pipelines much more efficiently than in-house tools. Often, the data pipeline is pre-integrated, offering a working solution while saving hundreds of hours of in-house set up costs.
- **Scaling as needed:** A cloud provider can provide and manage scaling for stored data, as opposed to an in-house system, which would require adding machines or managing clusters.
- **A flexible infrastructure:** Cloud services provide a flexible, on-demand infrastructure that is charged for based on time used. Additional services can also be accessed. (However, confusion and inexperience will result in wasted time and money.)
- **Backup copies:** Cloud providers strive to prevent service interruptions, so they store redundant copies of the data, using physically different servers, just in case your data gets lost.

Data lakes, sadly, have not become the perfect solution for working with unstructured data. The data lake industry is about seven years old and is not yet mature – unlike structured/SQL data systems.

Cloud-based data lakes may be easy to deploy but can be difficult to manage, resulting in unexpected costs. Data reliability issues can develop when combining batch and streaming data and corrupted data. A lack of experienced data lake professionals is also a significant problem.

[Data lakehouses](#), which are still in the development stage, have the goal of storing and accessing unstructured data, while providing the benefits of structured data/SQL systems.

The Benefits of Using Structured Data

Basically, the primary benefit of structured data is its ease of use. This benefit is expressed in three ways:

- **A great selection of tools:** Because this popular way of organizing data has been around for a while, a significant number of tools have been developed for structured/SQL databases.
- **Machine learning algorithms:** Structured data works remarkably well for training [machine learning](#) algorithms. The clearly defined nature of structured data provides a language machine learning can understand and work with.
- **Business transactions:** Structured data can be used for business purposes by the average person because it's easy to use. There is no need for an understanding of different types of data.

The Benefits of Using Unstructured Data

Examples of unstructured data include such things as social media posts, chats, email, presentations, photographs, music, and IoT sensor data. The primary strength of NoSQL and data lakes working with unstructured data is their flexibility in working with a variety of data formats. The benefits of working with NoSql databases or data lakes are:

- **Faster accumulation rates:** Because there is no need to transform different types of data into a standardized format, it can be gathered quickly and efficiently.
- **More efficient research:** A broader base of data taken from a variety of sources typically provides more accurate predictions of human behavior.

The Future of Structured and Unstructured Data

Over the next decade, the use of unstructured data will become much easier to work with, and much more commonplace. It will have no problems working with structured data. Tools for structured data will continue to be developed, and it will continue to be used for business purposes.

Although very much in the early stages of development, [artificial intelligence](#) algorithms have been developed that help find meaning automatically when searching unstructured data.

Currently, Microsoft's Azure AI is using a combination of optical character recognition, voice recognition, text analysis, and machine vision to scan and understand unstructured collections of data that may be made up of text or images.

Google offers a wide range of tools using AI algorithms that are ideal for working with unstructured data. For example, Vision AI can decode text, analyze images, and even recognize the emotions of people in photos.

In the next decade, we can predict that AI will play a significant role in processing unstructured data. There will be an urgent need for "recognition algorithms." (We currently seem to be limited to [image recognition](#), [pattern recognition](#), and [facial recognition](#).) As artificial intelligence evolves, it will be used to make working with unstructured data much easier.

Image used under license from Shutterstock.com

ABOUT THE AUTHOR

Keith D. Foote

Keith D. Foote is a researcher, writer, and consultant focusing on AI, machine learning, big data, and GDPR. His background is in physics and business with an emphasis on data science.

[Learn More](#)

SHARE THIS RESOURCE



Related Articles

ARTICLE

7 Cybersecurity Tips for 2026 No One Will Tell You About

[Read](#)

ARTICLE

How to Learn Data Governance in Six Months

[Read](#)

ARTICLE

Data Governance in the AI Era: Are We Solving the Wrong Problem?

[Read](#)[See More Articles](#)

Lead the Data Revolution from Your Inbox.

Never miss a beat with updates on our conferences, certifications, webinars, and newest resources.

EMAIL ADDRESS:

[Submit](#)