

STRUCTURED VS UNSTRUCTURED DATA

A look into structured and unstructured data, their key differences and which form best meets your business needs.

All data is not created equal. Some data is structured, but most of it is unstructured. Structured and unstructured data is sourced, collected and scaled in different ways, and each one resides in a different type of database.

In this article, we'll take a deep dive into both types so that you can get the most out of your data.

What is structured data?

Structured data — typically categorized as quantitative data — is highly organized and easily decipherable by [machine learning algorithms](#). [Developed by IBM in 1974](#), structured query language (SQL) is the programming language used to manage structured data. By using a [relational \(SQL\) database](#), business users can quickly input, search and manipulate structured data.

Pros and cons of structured data

Examples of structured data include dates, names, addresses, credit card numbers, etc. Their benefits are tied to ease of use and access, while liabilities revolve around data inflexibility:

Pros

- **Easily used by machine learning (ML) algorithms:** The specific and organized architecture of structured data eases manipulation and querying of ML data.
- **Easily used by business users:** Structured data does not require an in-depth understanding of different types of data and how they function. With a basic understanding of the topic relative to the data, users can easily access and interpret the data.
- **Accessible by more tools:** Since structured data predates unstructured data, there are more tools available for using and analyzing structured data.

Cons

- **Limited usage:** Data with a predefined structure can only be used for its intended purpose, which limits its flexibility and usability.

STRUCTURED VS UNSTRUCTURED DATA

- **Limited storage options:** Structured data is generally stored in data storage systems with rigid schemas (e.g., “[data warehouses](#)”). Therefore, changes in data requirements necessitate an update of all structured data, which leads to a massive expenditure of time and resources.

Structured data tools

- **OLAP:** Performs high-speed, multidimensional data analysis from unified, centralized data stores.
- **SQLite:** Implements a self-contained, [serverless](#), zero-configuration, transactional relational database engine.
- **MySQL:** Embeds data into mass-deployed software, particularly mission-critical, heavy-load production system.
- **PostgreSQL:** Supports SQL and JSON querying as well as high-tier programming languages (C/C+, Java, [Python](#), etc.).

Use cases for structured data

- **Customer relationship management (CRM):** CRM software runs structured data through analytical tools to create datasets that reveal customer behavior patterns and trends.
- **Online booking:** Hotel and ticket reservation data (e.g., dates, prices, destinations, etc.) fits the “rows and columns” format indicative of the pre-defined data model.
- **Accounting:** Accounting firms or departments use structured data to process and record financial transactions.

What is unstructured data?

Unstructured data, typically categorized as qualitative data, cannot be processed and analyzed via conventional data tools and methods. Since unstructured data does not have a predefined data model, it is best managed in [non-relational \(NoSQL\) databases](#). Another way to manage unstructured data is to use [data lakes](#) to preserve it in raw form.

The importance of unstructured data is rapidly increasing. [Recent projections](#) indicate that unstructured data is over 80% of all enterprise data, while 95% of businesses prioritize unstructured data management.

Pros and cons of unstructured data

Examples of unstructured data include text, mobile activity, social media posts, Internet of Things (IoT) sensor data, etc. Their benefits involve advantages in

STRUCTURED VS UNSTRUCTURED DATA

format, speed and storage, while liabilities revolve around expertise and available resources:

Pros

- **Native format:** Unstructured data, stored in its native format, remains undefined until needed. Its adaptability increases file formats in the database, which widens the data pool and enables data scientists to prepare and analyze only the data they need.
- **Fast accumulation rates:** Since there is no need to predefine the data, it can be collected quickly and easily.
- **Data lake storage:** Allows for massive storage and pay-as-you-use pricing, which cuts costs and eases scalability.

Cons

- **Requires expertise:** Due to its undefined/non-formatted nature, [data science](#) expertise is required to prepare and analyze unstructured data. This is beneficial to data analysts but alienates unspecialized business users who may not fully understand specialized data topics or how to utilize their data.
- **Specialized tools:** Specialized tools are required to manipulate unstructured data, which limits product choices for data managers.

Unstructured data tools

- **[MongoDB](#):** Uses flexible documents to process data for cross-platform applications and services.
- **[DynamoDB](#):** Delivers single-digit millisecond performance at any scale via built-in security, in-memory caching and backup and restore.
- **[Hadoop](#):** Provides distributed processing of large data sets using simple programming models and no formatting requirements.
- **[Azure](#):** Enables agile cloud computing for creating and managing apps through Microsoft's data centers.

Use cases for unstructured data

- **[Data mining](#):** Enables businesses to use unstructured data to identify consumer behavior, product sentiment, and purchasing patterns to better accommodate their customer base.
- **[Predictive data analytics](#):** Alert businesses of important activity ahead of time so they can properly plan and accordingly adjust to significant market shifts.
- **[Chatbots](#):** Perform text analysis to route customer questions to the appropriate answer sources.

STRUCTURED VS UNSTRUCTURED DATA

What are the key differences between structured and unstructured data?

While structured (quantitative) data gives a “birds-eye view” of customers, unstructured (qualitative) data provides a deeper understanding of customer behavior and intent. Let’s explore some of the key areas of difference and their implications:

- **Sources:** Structured data is sourced from GPS sensors, online forms, network logs, web server logs, [OLTP systems](#), etc., whereas unstructured data sources include email messages, word-processing documents, PDF files, etc.
- **Forms:** Structured data consists of numbers and values, whereas unstructured data consists of sensors, text files, audio and video files, etc.
- **Models:** Structured data has a predefined data model and is formatted to a set data structure before being placed in data storage (e.g., schema-on-write), whereas unstructured data is stored in its native format and not processed until it is used (e.g., schema-on-read).
- **Storage:** Structured data is stored in tabular formats (e.g., excel sheets or SQL databases) that require less storage space. It can be stored in data warehouses, which makes it highly scalable. Unstructured data, on the other hand, is stored as media files or NoSQL databases, which require more space. It can be stored in data lakes which makes it difficult to scale.
- **Uses:** Structured data is used in machine learning (ML) and drives its algorithms, whereas unstructured data is used in [natural language processing](#) (NLP) and text mining.

What is semi-structured data?

Semi-structured data (e.g., JSON, CSV, XML) is the “bridge” between structured and unstructured data. It does not have a predefined data model and is more complex than structured data, yet easier to store than unstructured data.

Semi-structured data uses “metadata” (e.g., tags and semantic markers) to identify specific data characteristics and scale data into records and preset fields. Metadata ultimately enables semi-structured data to be better cataloged, searched and analyzed than unstructured data.

- **Example of metadata usage:** An online article displays a headline, a snippet, a featured image, image alt-text, slug, etc., which helps differentiate one piece of web content from similar pieces.
- **Example of semi-structured data vs. structured data:** A tab-delimited file containing customer data versus a database containing CRM tables.

STRUCTURED VS UNSTRUCTURED DATA

- **Example of semi-structured data vs. unstructured data:** A tab-delimited file versus a list of comments from a customer's Instagram.

The future of data

Recent developments in [artificial intelligence](#) (AI) and machine learning (ML) are driving the future wave of data, which is enhancing business intelligence and advancing industrial innovation. In particular, the data formats and models covered in this article are helping business users to do the following:

- **Analyze digital communications for compliance:** Pattern recognition and email threading analysis software that can search email and chat data for potential noncompliance.
- **Track high-volume customer conversations in social media:** Text analytics and sentiment analysis that enables monitoring of marketing campaign results and identifying online threats.
- **Gain new marketing intelligence:** ML analytics tools that can quickly cover massive amounts of data to help businesses analyze customer behavior.

Furthermore, smart and efficient usage of data formats and models can help you with the following:

- Understand customer needs at a deeper level to better serve them
- Create more focused and targeted marketing campaigns
- Track current metrics and create new ones
- Create better product opportunities and offerings
- Reduce operational costs

Structured and unstructured data and IBM

Whether you are a seasoned data expert or a novice business owner, being able to handle all forms of data is conducive to your success. By leveraging structured, semi-structured and unstructured data options, you can perform optimal data management that will ultimately benefit your mission.