



14 Dec, 2020

Structured vs Unstructured Data: Compared and Explained

Reading time: 14 minutes

According to IBM, [the global volume of data](#) was predicted to reach 35 zettabytes in 2020. Since it increases daily, data scientists expect that the number will hit 175 zettabytes in 2025. Picture this: 35ZB holds approximately 1 trillion hours' worth of movies. It will take 115 million years to watch all those movies. Those are some impressive figures, aren't they? Well, there's something even more impressive about the global data sphere. The prevailing part of data, which is 80 percent or so, is unstructured. This means structured data only has about 20 percent of all generated information.

In this article, you'll get a closer look at structured vs unstructured data. Let's see what the difference between the two is and why you should know it in the first place. Also, we will help you understand how to handle each data type and what software tools are available for each purpose.

Structured vs unstructured data in a nutshell

Data exists in a plethora of different forms and sizes, but most of it can be presented as structured data and unstructured data.



Structured vs Unstructured Data

Structured data

Structured data stands for information that is highly organized, factual, and to-the-point.

Quantitative

**Data warehouses
Relational databases**

Several predetermined formats

Unstructured data

Unstructured data doesn't have any predefined structure to it and comes in all its diversity of forms.

Qualitative

**Data lakes
Non-relational databases**

A huge array of formats



altexsoft
software r&d engineering

The key differences between unstructured data and structured data.

Structured data stands for information that is highly organized, factual, and to-the-point. It usually comes in the form of letters and numbers that fit nicely into the rows and columns of tables. Structured data commonly exists in tables similar to Excel files and Google Docs spreadsheets.

Unstructured data doesn't have any pre-defined structure to it and comes in all its diversity of forms. The examples of unstructured data vary from imagery and text files like PDF documents to video and audio files, to name a few.

Structured data is often spoken of as quantitative data, meaning its objective and pre-defined nature allows us to easily count, measure, and express data in numbers. Unstructured data, alternately, is called qualitative data in the sense that it has a subjective and interpretive nature. This data can be categorized depending on its characteristics and traits.

With that summary, let's move on to more descriptive explanations of the differences.

What is structured data?



So, structured data is the type of data that is well-organized and accurately formatted. This data exists in a format of relational databases ([RDBMSs](#)), meaning the information is stored in tables with rows and columns that are connected. In this way, structured data is arranged and recorded neatly, so it can be easily found and processed. As long as data fits within the structure of RDBMSs, we can easily search for specific information and single out the relationships between its pieces. Such data can only be used for its intended purpose. On top of that, structured data doesn't normally require much storage space.

For analytical purposes, you can use [data warehouses](#). DWs are central data storages used by companies for data analysis and reporting.

There is a special programming language used for handling relational databases and warehouses called SQL, which stands for Structured Query Language and was developed back in the 1970s by IBM.

Structured data examples. Structured data is familiar to most of us. Google Sheets and Microsoft Office Excel files are the first things that spring to mind concerning structured data examples. This data can comprise both text and numbers, such as employee names, contacts, ZIP codes, addresses, credit card numbers, etc.

	A	B	C	D	E	F	G
1	Purchase ID	Last name	First name	Birthday	Country	Date of purchase	Amount of purchase
2	1	Davidson	Michael	04/03/1986	United States	10/12/2016	37
3	2	Vito	Jim	09/01/1994	United Kingdom	02/02/2016	85
4	3	Johnson	Tom	23/08/1972	France	02/11/2016	83
5	4	Lewis	Peter	18/10/1979	Germany	22/11/2016	27
6	5	Koenig	Edward	13/05/1983	Argentina	26/03/2015	43
7	6	Preston	Jack	16/06/1991	United States	06/11/2016	77
8	7	Smith	David	11/03/1965	Canada	15/11/2016	23
9	8	Brown	Luis	03/09/1997	Australia	03/07/2015	74
10	9	Miller	Thomas	07/01/1980	Germany	07/11/2016	13
11	10	Williams	Bill	26/07/1960	United States	20/11/2015	80
12	11	Gemini	Alexia	12/09/1995	Canada	11/03/2017	35
13	12	Bond	James	25/02/1975	United Kingdom	12/08/2017	40
14	13	Burgle	Patricia	01/12/1990	United States	18/01/2015	55
15	14	Reding	Michelle	07/04/1985	Canada	23/02/2017	28
16	15	Harvey	Billy	14/07/1971	United Kingdom	12/01/2016	41
17							

The typical structured data example: Excel spreadsheet that contains information about customers and purchases.

Pretty much everyone has dealt with booking a ticket via one of the [airline reservation systems](#) or withdrawing cash using an ATM. During these operations, we don't normally think of what kind of applications we deal with and what types of data they process. However, these are the systems that typically use structured data and relational databases as well.



What is unstructured data?

It makes sense that if the definition of structured data implies a neat organization of components in a predetermined manner, the definition of unstructured data will be the opposite. The pieces of such data aren't structured in a pre-defined way, meaning data is stored in its native formats.

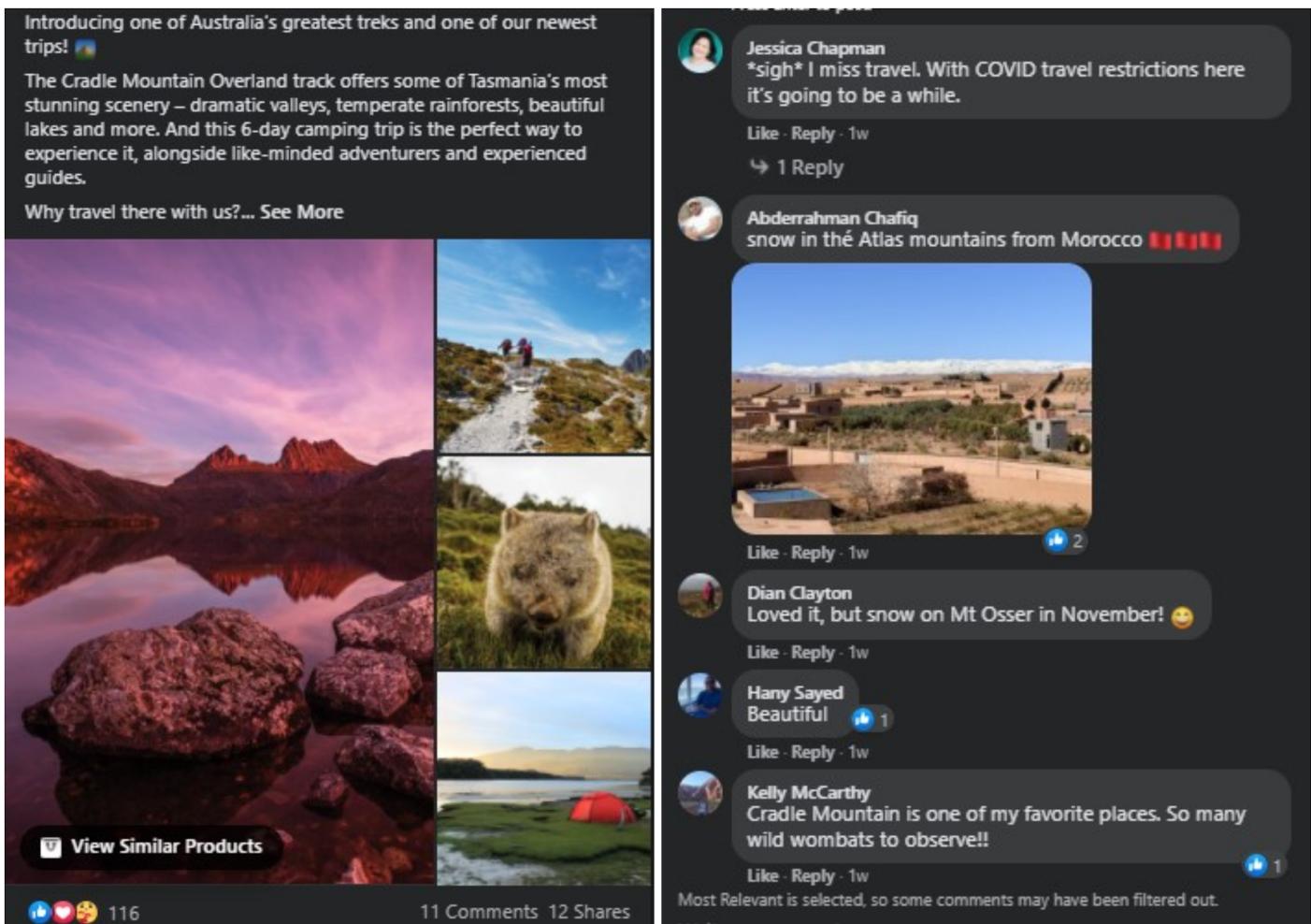
The thing with unstructured data is that traditional methods and tools can't be used to analyze and process it. One of the ways to manage unstructured data is to opt for non-relational databases, also known as [NoSQL](#).

If there's a need to keep data in its raw native formats for further analysis, storage repositories called [data lakes](#) will be the way to go. A data lake is a storage repository or system meant to store huge volumes of data in its natural/raw formats.

Taking into account the whole variety of file formats of unstructured data, it comes as no surprise that it makes up more than 80 percent of all data. Given this, companies ignoring unstructured data are left far behind as they don't get enough valuable information.

Unstructured data examples. There is a wide array of forms that make up unstructured data such as email, text files, social media posts, video, images, audio, sensor data, and so on.





The travel agency Facebook post: an example of unstructured data.

As an example, we can take social media posts of a travel agency or all posts for that matter. Each post contains some metrics like shares or hashtags that can be quantified and structured. However, the posts themselves belong to the category of unstructured data. What we're trying to say here is, it will take some time, effort, knowledge, and special software tools to analyze the posts and collect useful insights. If an agency posts new travel tours and wants to know the audience's reactions (comments), they will need to examine the post in its native format (view the post via social media app or use advanced techniques like [sentiment analysis](#)).

The key differences between structured and unstructured data



Structured vs Unstructured Data

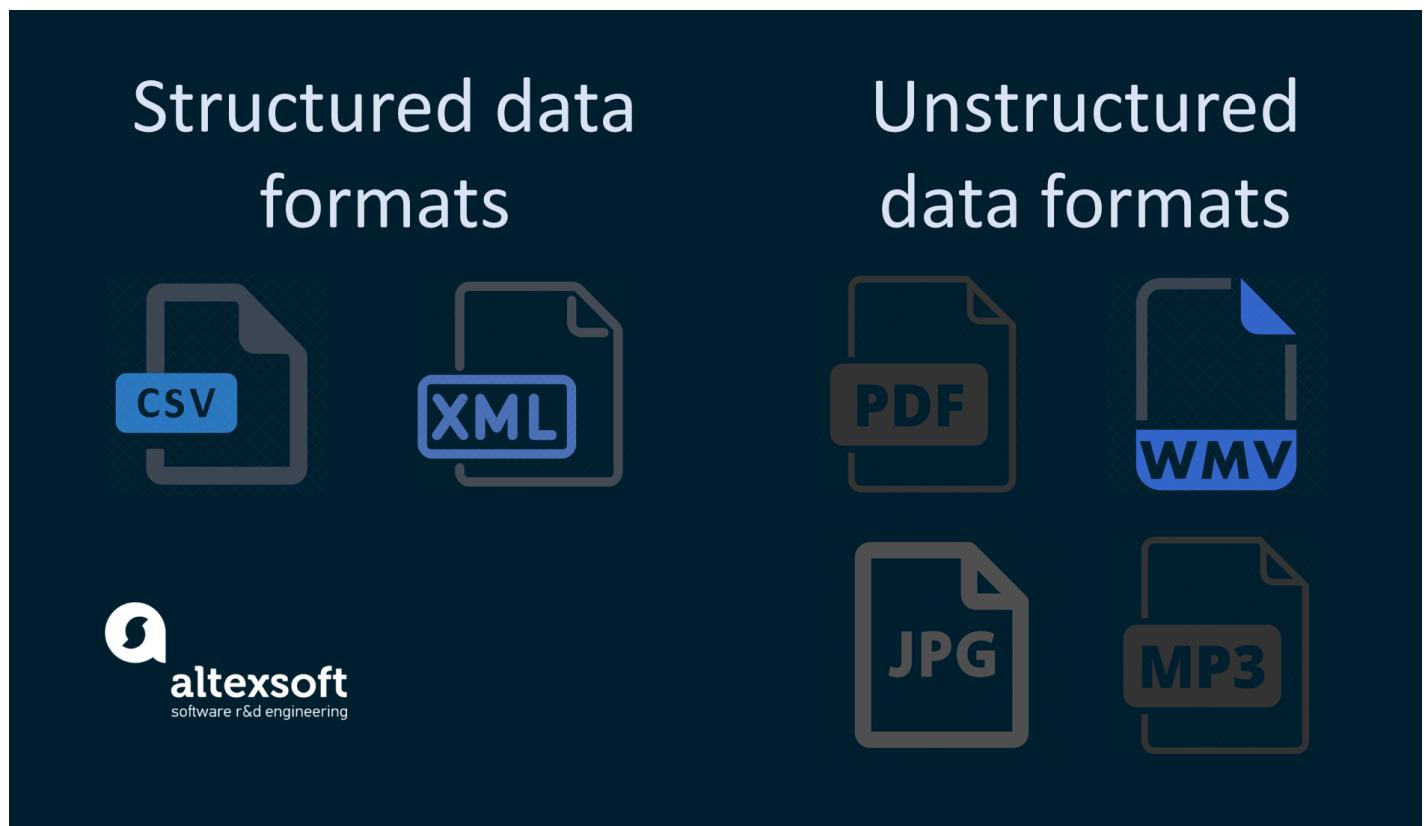
Properties	Structured data	Unstructured data
Formats	Several formats	A huge variety of formats
Data model	Pre-defined/not flexible	Not pre-defined/flexible
Storages	Data warehouses	Data lakes
Databases	SQL Relational databases	NoSQL Non-relational databases
Ease of search	Easy to search	Difficult to search
Data nature	Quantitative	Qualitative
Analysis methods	<ul style="list-style-type: none"> Classification Regression Data clustering 	<ul style="list-style-type: none"> Data stacking Data mining
Tools and technologies	<ul style="list-style-type: none"> RDBMS CRM OLAP OLTP 	<ul style="list-style-type: none"> NoSQL DBMS AI-driven tools Data storage architectures Data visualization tools
Specialists to handle data	Business analysts Software engineers Marketing analysts	Data scientists, engineers, and analysts with deep expertise

Differences between structured and unstructured data in detail.

Now let's discuss a few more important differences between structured and unstructured data:

Data formats: few formats vs plethora of formats

Structured data is usually presented in the form of text and numbers. Its formats are standardized and user-readable. The most common ones are CSV and XML. In a data model, the data format has been determined in advance.



Data formats.

Unlike structured data, unstructured data formats are presented in a surfeit of different shapes and sizes. Unstructured data doesn't have any pre-defined data model and it is stored in its *native formats* (aka "original" formats). Those can be audio (WAV, MP3, OGG, etc.) or video files (MP4, WMV, etc.), PDF documents, images (JPEG, PNG, etc.), emails, social media posts, sensor data, etc.

Data models: pre-defined vs flexible

Structured data is less flexible as it relies on a strict organization of a data model. Such data is schema dependent. The schema of the database stands for the configuration of columns (also called fields) and the types of data meant to be held in these columns. Such dependency is both an advantage and a disadvantage.



disadvantage. While the information here can be easily searchable and processed, all records have to follow the very strict requirements of the schema.

Unstructured data, on the other hand, offers more flexibility and scalability. The absence of the pre-defined purpose of unstructured data makes it super flexible as the information can be stored in various file formats. Yet, this data is subjective and more difficult to work with.

Storages for analytical use: data lakes vs data warehouses

If we apply data for analytical processing and use so-called [data pipelines](#), the final destination of the structured data's journey will be special **data warehouses**. These are space-saving storages or repositories with a defined structure that is difficult to change. Even minor changes to the schema may result in the need to reconstruct huge volumes of data, which might entail spending time and resources.

The bigger the data volume is, the more space it requires for storage. A picture with high resolution weighs a lot more than a textual file. Therefore, unstructured data requires more storage space and is usually kept in **data lakes**, storage repositories that allow for storing almost limitless amounts of data in its raw formats. Apart from data lakes, unstructured data resides in native applications.

There is the potential for cloud-use in both cases.

Databases: SQL vs NoSQL

As we have already mentioned, structured data lives in **relational databases**, also known as RDBMSs. The data here is set up in tables that have a lot of rows (also called records) and columns with labels, denoting specific data types they are supposed to keep. The configuration of data types and columns makes up the schema of the database table.

Relational databases use SQL, or Structured Query Language, to reach the stored data and manipulate it. SQL syntax is similar to that of the English language, providing the simplicity of writing, reading, and interpreting it.



SQL

```
SELECT column1, column2 FROM table1='customers';
```



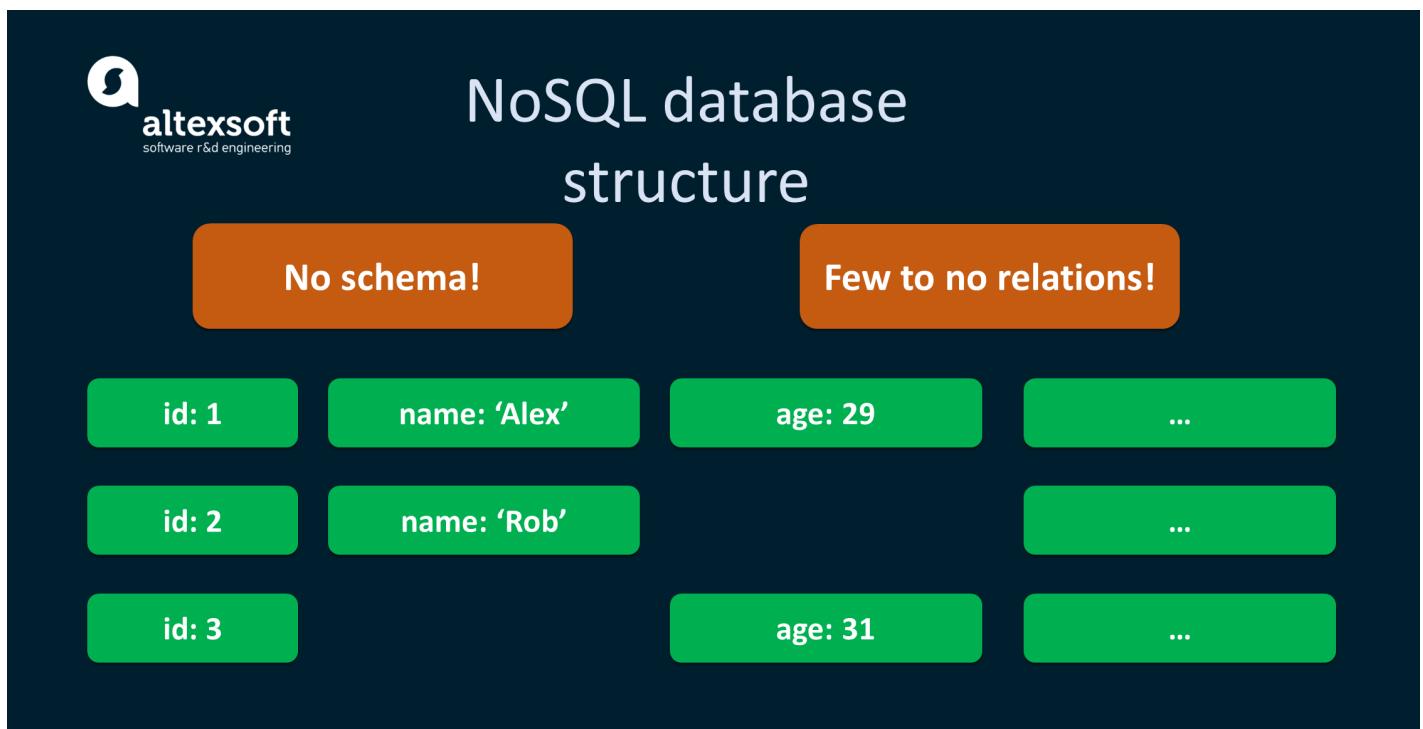
altexsoft
software r&d engineering

This is how SQL helps make queries.

Speaking of databases for unstructured data, the most suitable option for this type of data will be non-relational databases, also known as **NoSQL databases**.

NoSQL stands for “not only SQL.” These databases have various data models and they store data in a non-tabular way. The most common types of NoSQL databases are key-value, document, graph, and wide-column. Such databases can process huge volumes of data and deal with high user loads as they are quite flexible and scalable. In the NoSQL world, there are collections of data rather than tables. In these collections, there are so-called documents. While the documents may look like rows in tables, they don’t use the same schema. It’s possible to have multiple documents in one collection that have different fields. On top of that, there are few to no relations between items of data. The idea here is to have less relation merging going on and instead to have super-fast and efficient queries. Although, there will be some data duplicates.





The example of a NoSQL data structure.

Ease of search, analysis, and processing

One of the main differences between structured and unstructured data is how easily it can be subjected to analysis. Structured data is overall easy to search and process whether it is a human who processes data or program algorithms. Unstructured data, by contrast, is a lot more difficult to search and analyze. Once found, such data has to be processed attentively to understand its worth and applicability. The process is challenging as unstructured data can't fit within the fixed fields of relational databases until it is stacked and handled.

From a historical point of view, since structured data has been here longer, it's logical that there is a great choice of mature analytics tools for it. At the same time, those who work with unstructured data may face a poorer choice of analytics tools as most of them are still being developed. The usage of traditional data mining tools usually crashes into the rocks of the disorganized internal structure of this data type.

Data nature: quantitative vs qualitative

Structured data is often referred to as **quantitative data**. It means that such data commonly contains precise numbers or textual elements that can be counted. The analysis methods are clear and easy-to-apply. Among them there are:

- **classification** or arranging stored items of data into similar classes based on common features,
- **regression** or investigation of the relationships and dependencies between variables, and



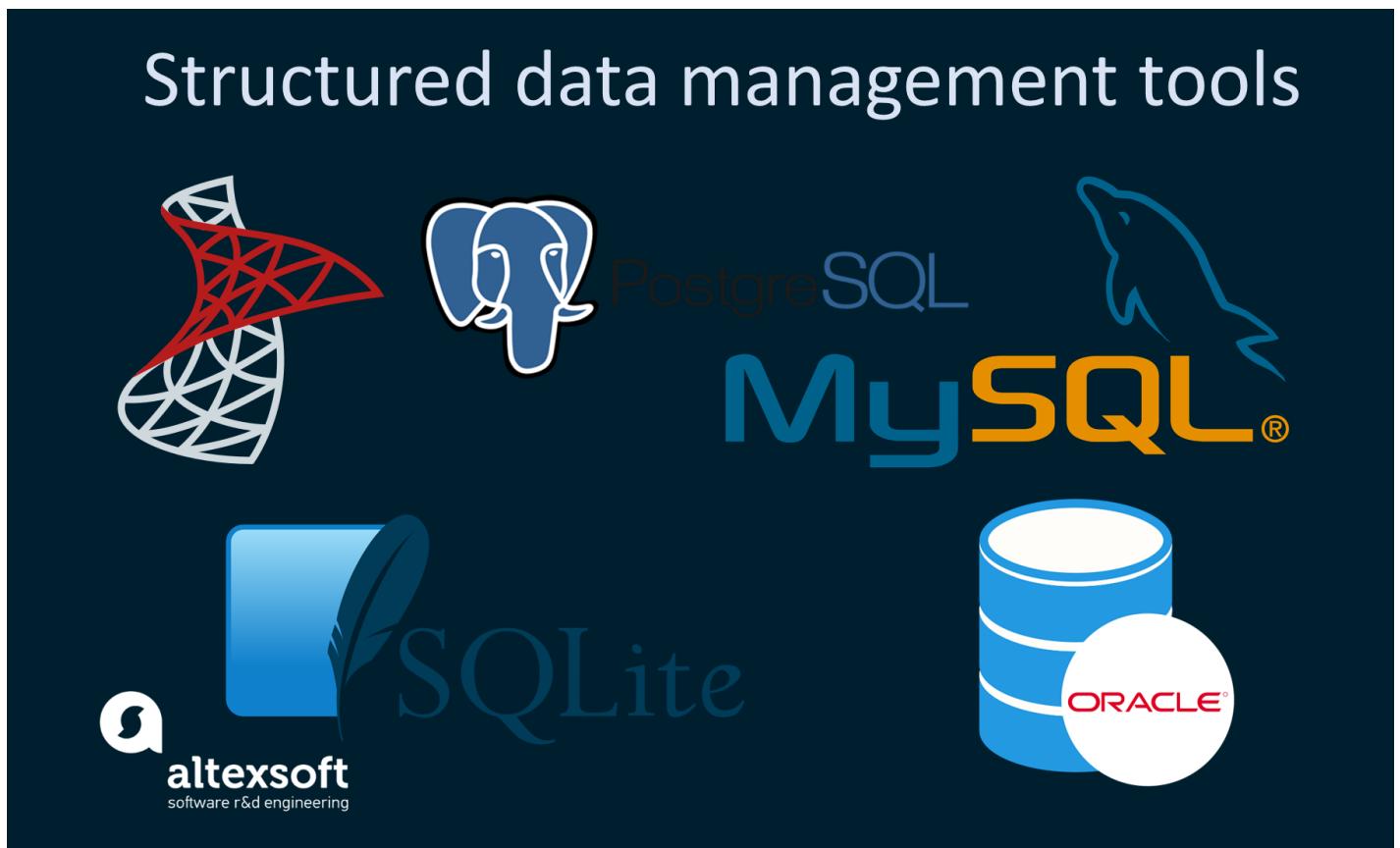
- **data clustering** or organizing the data points into specific groups based on various attributes.

Unstructured data, in turn, is often classified as **qualitative data** containing subjective information that can't be handled using traditional methods and software analytics tools. For instance, qualitative data can flow from customer surveys or social media feedback in a text form. To process and analyze qualitative data, more cutting-edge analytics techniques are required such as:

- **data stacking** or investigation of large volumes of data, splitting them into smaller items and stacking the variables with similar values into a single group, and
- **data mining** or the process of detecting certain patterns, oddities, and interactions in large data sets to express possible outcomes in advance.

Tools and technologies

Structured data tools. The clear-cut and highly organized essence of structured data contributes to a wide array of data management and analytics tools. This opens opportunities for **data teams** in terms of picking up the most fitting software product when working with structured data.



Structured data management⁺



Among the most commonly used relational database management systems, data tools, and technologies there are the following:

- *PostgreSQL*. It's a free, open-source RDBMS that supports both SQL and JSON querying as well as the most widely used programming languages such as Java, Python, C/C+, etc.
- *SQLite*. It's another popular choice of an SQL database engine contained in a C library. It's a lightweight and transactional system that doesn't rely on a separate server process as it is rather inserted into the end-program.
- *MySQL*. One of the most popular open-source RDBMSs that is fast and reliable. It runs on a server and allows for creating both small and large apps.
- *Oracle Database*. This is an advanced database management system with a multi-model structure. It can be used for data warehousing, online transaction processing, and mixed database workloads.
- *Microsoft SQL Server*. Developed by Microsoft, SQL Server is a reliable and functional relational database management system that makes it possible to store and retrieve data as per requests of other software applications.
- *OLAP applications*. A unit of business intelligence (BI), online analytical processing (OLAP) stands for an advanced computing approach that answers multi-dimensional queries effectively and swiftly. OLAP tools allow users to work with data from different perspectives, because they combine data mining, a relational database, and reporting features. *Apache Kylin* is one of the most popular open-source OLAP systems. It supports large data sets as it is synced with Hadoop.

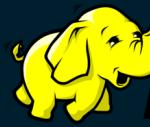
Unstructured data tools. As unstructured data comes in various shapes and sizes, it requires specially designed tools to be properly analyzed and manipulated. Also, there's a necessity of finding a qualified data science team. Not only is it useful to understand the topic of data, but it is also crucial to figure out the relations of that data.



Unstructured data management tools



DynamoDB



hadoop



mongoDB

Unstructured data management tools.

Below you find a few examples of tools and technologies to manage unstructured data effectively:

- *MongoDB*. This is a document-oriented database management system that does not require any rigid schema or structure of tables. It is thought of as one of the classic NoSQL examples. MongoDB uses JSON-like documents.
- *Amazon DynamoDB*. Offered by Amazon as a part of their AWS package, DynamoDB is an advanced NoSQL database service for complete data management. It supports document and key-value data structures and is a good fit for working with unstructured data.
- *Apache Hadoop*. This is an efficient, open-source framework used for processing large amounts of data and storing it on inexpensive commodity servers. Apart from being a powerful tool, Hadoop is also flexible as it does not require having a schema or a structure for the stored data. It helps with structuring unstructured data and then exporting this data to relational databases.
- *Microsoft Azure*. Presented by Microsoft, Azure is a comprehensive cloud service for building and managing applications and services via data centers. Azure Cosmos DB is a fast and scalable NoSQL database that helps with storing and analyzing masses of unstructured data.

Back in the day, unstructured data analysis was typically manual, and a time-consuming process. Nowadays there are quite a few advanced AI-driven tools that help sort out unstructured data, find relevant items, and store the results. The technologies and tools for unstructured data incorporate



natural language processing and machine learning algorithms. As such, it is possible to adjust software products to the needs of specific industries.

Data teams to handle data

Owing to relational databases having been here for longer, they are more familiar to a user. Data specialists with different levels of skills can work with any RDB quite easily and quickly as a data model is pre-defined. Any inputs, searches, queries, and manipulations are made within a highly-organized environment, resulting in opening self-service access to **different specialists from business analysts to software engineers.**

Unlike structured data tools, those designed for unstructured data are more complex to work with. Therefore, they require a certain level of expertise in data science and machine learning to conduct deep data analysis. Besides that, specialists who deal with unstructured data have to have a good understanding of a data topic and how the data is related. Given the above, to handle unstructured data, a company will need qualified help from **data scientists, engineers, and analysts.**

Structured and unstructured data examples and use cases

As we've partially touched on the subject matter of structured and unstructured data examples above, it would be useful to point out particular use cases.

So, when you think of dates, names, product IDs, transaction information, and so forth, you know that you have structured data in mind. At the same time, unstructured data has many faces like text files, PDF documents, social media posts, comments, images, audio/video files, and emails, to name a few.

More often than not industries need to leverage both data types to improve the efficiencies of their services.



Structured vs Unstructured Data

Industry	Structured data	Unstructured data
eCommerce	<ul style="list-style-type: none"> • Product IDs • Pricing data • Customer account data 	<ul style="list-style-type: none"> • Customer behavior and spending patterns • Customer service satisfaction (reviews, social media mentions)
Healthcare	<ul style="list-style-type: none"> • Patient forms • Medical insurance data • Medical billing data 	<ul style="list-style-type: none"> • X-Ray and MRI scans • Doctor notes • Treatment recommendations
Banking	<ul style="list-style-type: none"> • Financial transactions • Customer account data 	<ul style="list-style-type: none"> • Call logs and weblogs • Audio and video communication



How structured and unstructured data is used in different industries.

Structured data use case examples

Online booking. Different hotel booking and ticket reservation services leverage the advantages of the pre-defined data model as all booking data such as dates, prices, destinations, etc. fit into a standard data structure with rows and columns.

ATMs. Any ATM is a great example of how relational databases and structured data work. All the actions a user can do follow a pre-defined model.



Inventory control systems. There are lots of variants of inventory control systems companies use, but they all rely on a highly organized environment of relational databases.

Banking and accounting. Different companies and banks must process and record huge amounts of financial transactions. Consequently, they make use of traditional database management systems to keep structured data in place.

Unstructured data use case examples

Sound recognition. Call centers use speech recognition to identify customers and collect information about their queries and emotions.

Image recognition. Online retailers take advantage of image recognition so that customers can shop from their phones by posting a photo of the desired item.

Text analytics. Manufacturers make use of advanced text analytics to examine warranty claims from customers and dealers and elicit specific items of important information for further clustering and processing.

Chatbots. Using natural language processing (NLP) for text analysis, chatbots help different companies boost customer satisfaction from their services. Depending on the question input, customers are routed to the corresponding representatives that would provide comprehensive answers.

What is semi-structured data?

As the name suggests, semi-structured data is partially structured, meaning that it incorporates certain markers that can split semantic elements and implement data hierarchies, but it is still different from the tabular data models presented in relational databases. Such a structure is called self-describing. Markup languages such as XML are the forms of semi-structured data. JSON is also a semi-structured data model that is used by new-generation databases such as MongoDB and Couchbase. There are a bunch of other Big Data tools and solutions that use this category of data because it is significantly easier to process than, say, unstructured data.





How data is organized in JSON.

Source: [techEplanet](#)

While semi-structured data may seem like a happy medium, it is not like that. In today's highly competitive environment, businesses need to use all data sources to receive information and use it correctly to reap the benefits.

The blurred line between structured and unstructured data

Wrapping things up, it is worth saying that there is no real struggle between unstructured data and structured data. Both types of data carry great value for businesses of diverse fields and scale. Picking a data source may depend on the structure of data. But more often than not, we don't choose one type over the other and rather look for the software opportunities to handle all data.

In the past, companies had no real way of analyzing unstructured data, so it was discarded while the focus was put on the data that could be easily counted. Nowadays, companies can use artificial intelligence, machine learning opportunities, and advanced analytics to do the tricky unstructured data analysis for them. For example, corporations like Google have made huge advances in image recognition technology by creating AI algorithms that can automatically detect what or who is on a photograph.

Truth be told, those lines between structured and unstructured data are a little bit blurred because most datasets are semi-structured these days. Even if we take unstructured data like a photograph, it st



components of structured data such as image size, resolution, the date the image was taken, etc. This information can be organized in a tabular format of relational databases.

Now that you know the characteristics and differences between unstructured and structured data, you can make an informed decision on whether or not you should invest in technologies to grasp unstructured data benefits. The best-case scenario for corporations is to adopt both data types, improving the effectiveness of business intelligence.

