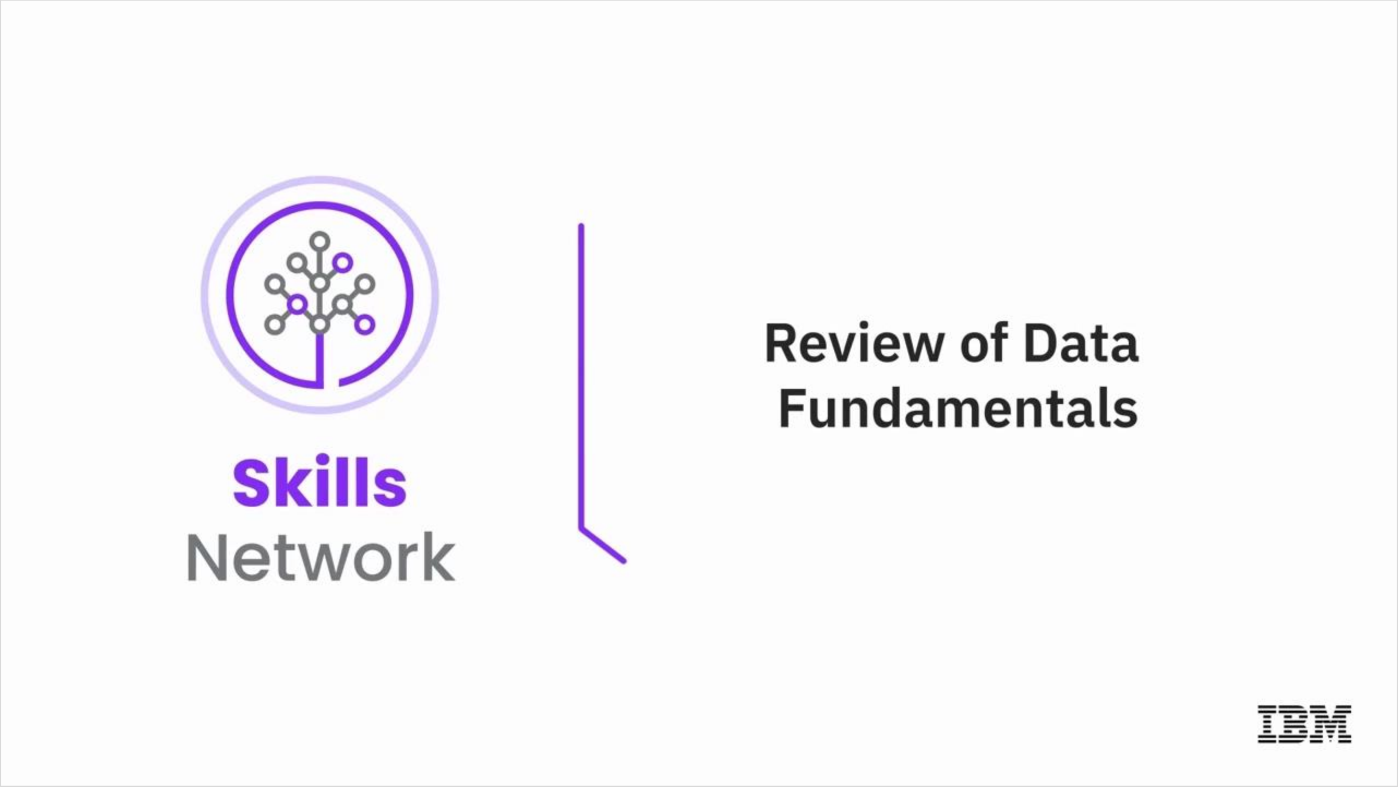
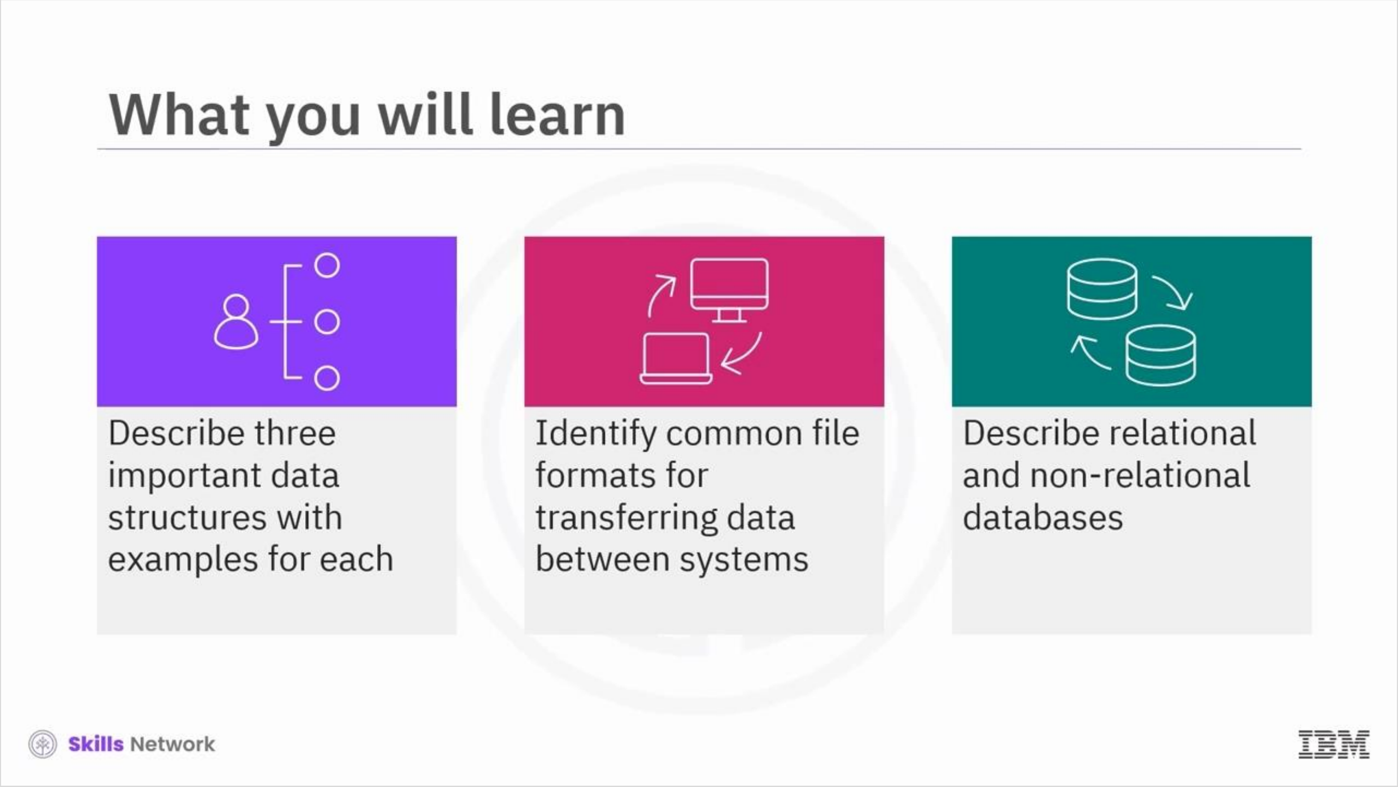
- **Fundamental Relational Database Concepts**



Skills Network

Review of Data Fundamentals

IBM



What you will learn

Describe three important data structures with examples for each

Identify common file formats for transferring data between systems

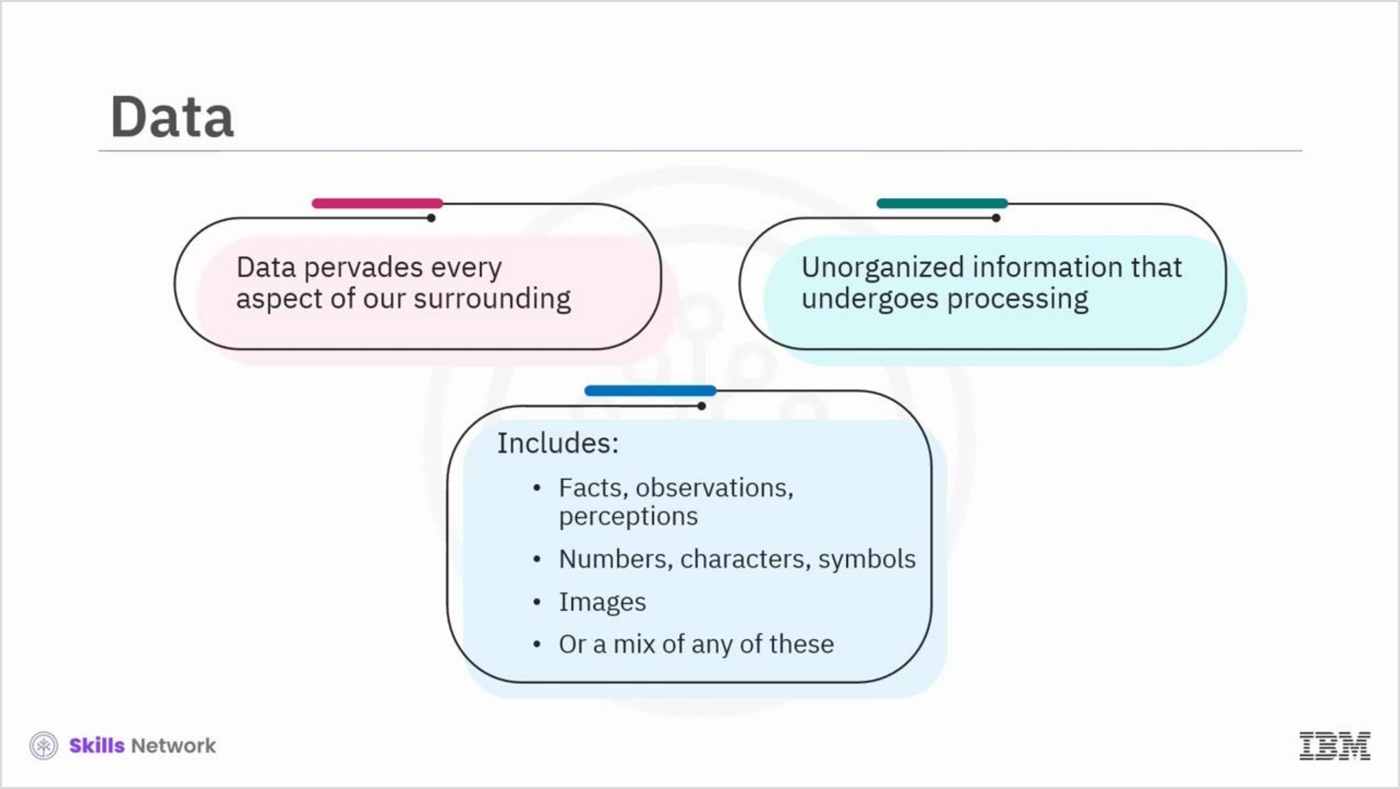
Describe relational and non-relational databases

Skills Network

IBM

Welcome to Review of Data Fundamentals. After watching this video, you will be able to describe three important data structures with examples for each, identify common file formats for transferring data between systems, describe relational and non-relational databases.

Data pervades every aspect of our surroundings in the rapidly evolving world of information and technology but how do we define data? Data refers to unorganized information that undergoes processing to make it meaningful. It includes facts, observations, perceptions, numbers, characters, symbols, images, or a combination of these elements.



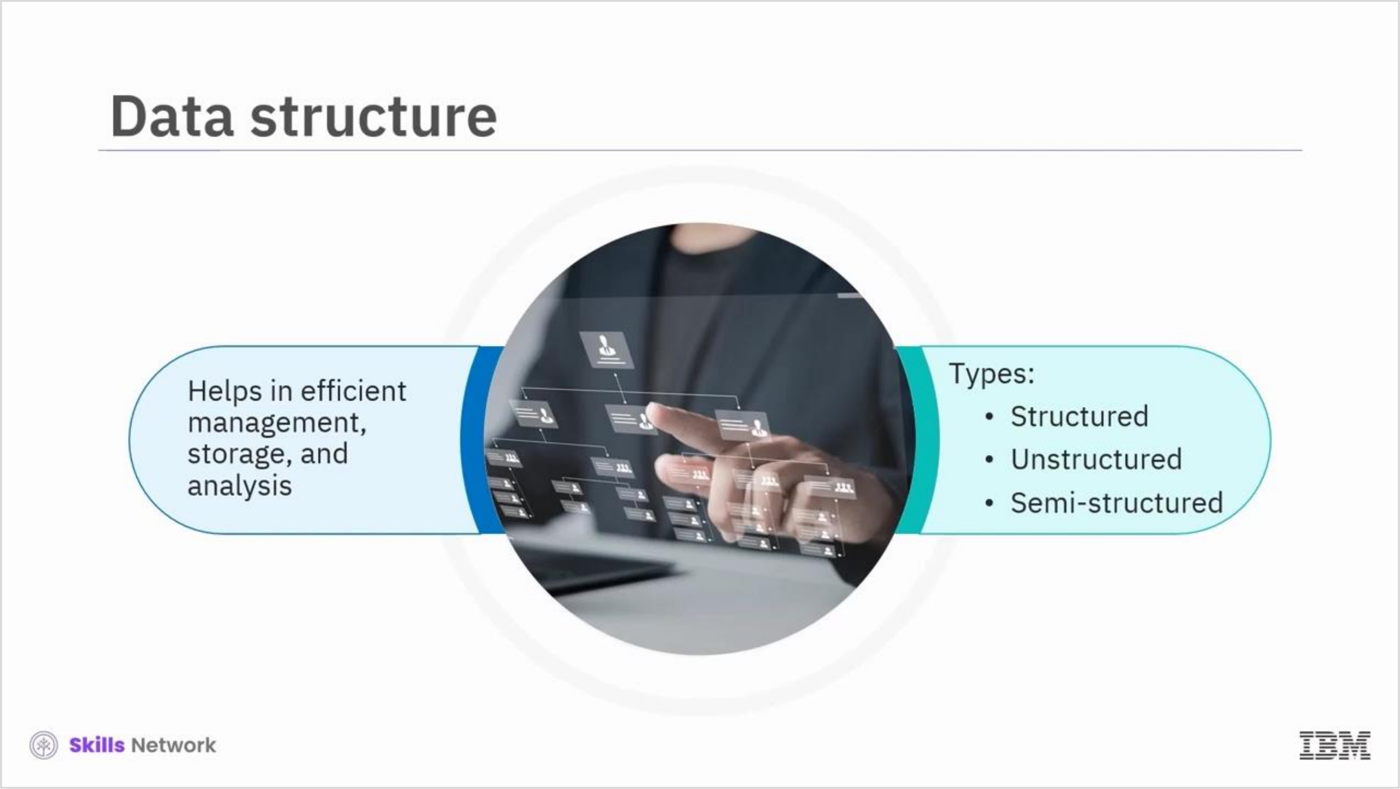
Data

Data pervades every aspect of our surrounding

Unorganized information that undergoes processing

Includes:

- Facts, observations, perceptions
- Numbers, characters, symbols
- Images
- Or a mix of any of these



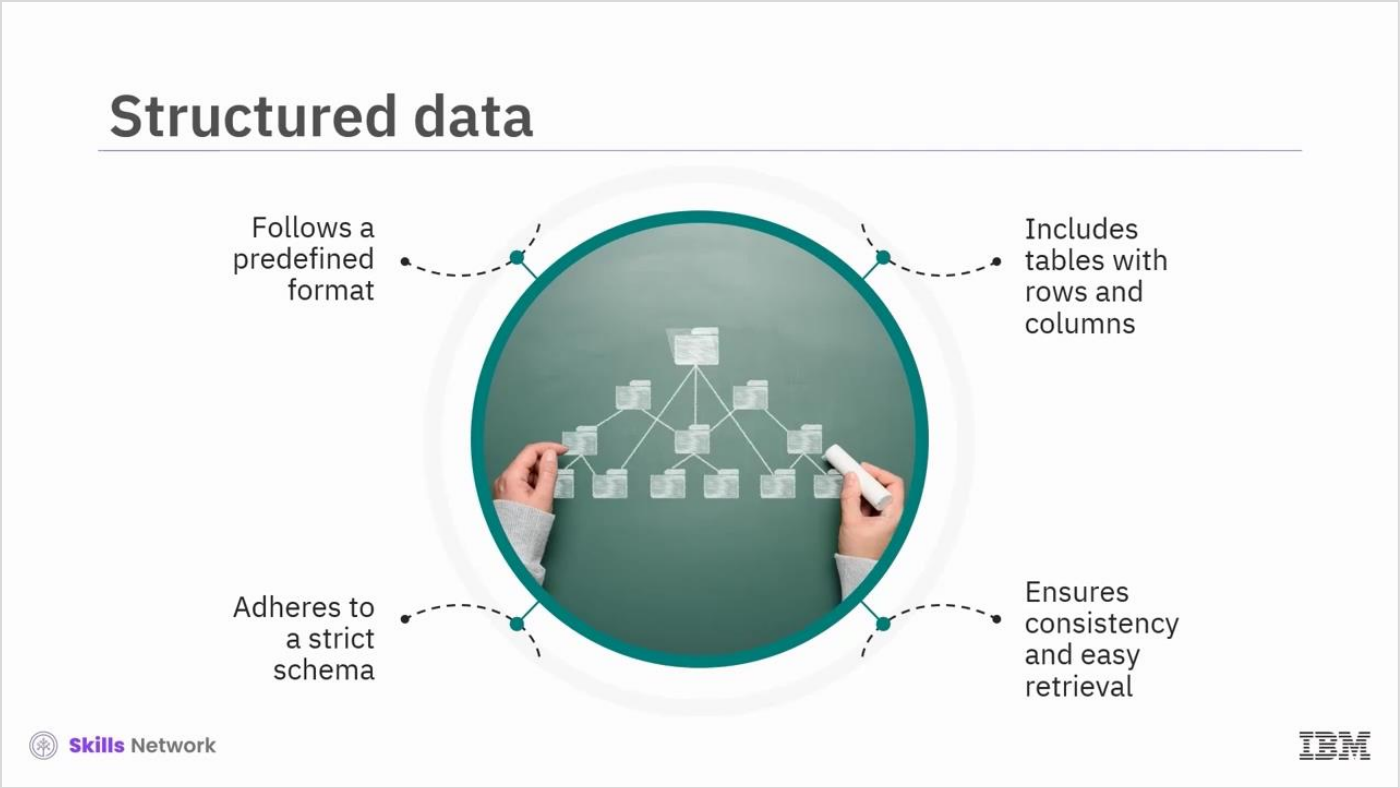
Data structure

Helps in efficient management, storage, and analysis

Types:

- Structured
- Unstructured
- Semi-structured

The structure of data plays an important role in determining its efficient management, storage, and analysis. Broadly, we can categorize data into three main types, structured, unstructured, and semi-structured. Let us learn about each of these data structures.



Structured data

Follows a predefined format

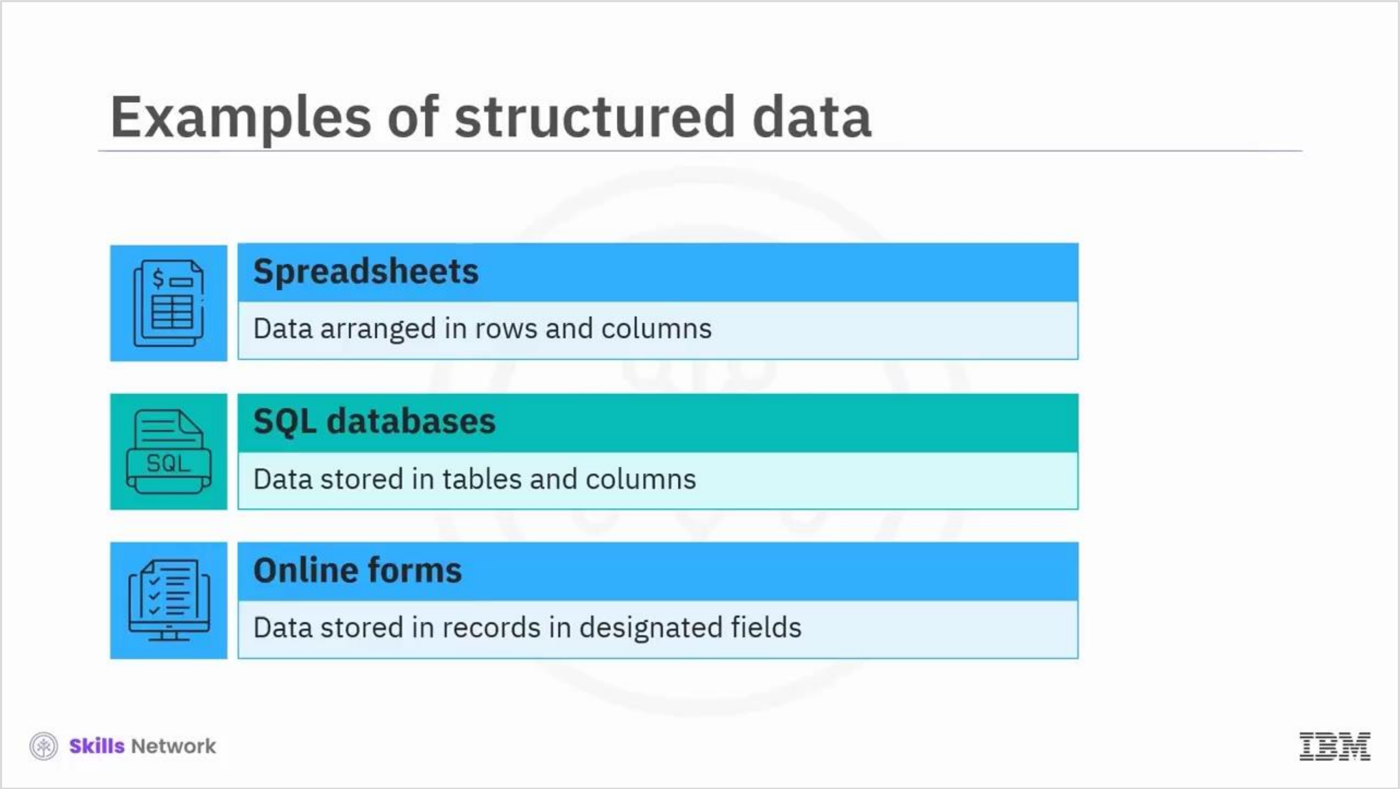
Includes tables with rows and columns

Adheres to a strict schema

Ensures consistency and easy retrieval

Skills Network
IBM

Structured data is highly organized and follows a predefined format, typically arranged in tables with rows and columns. It adheres to a strict schema and a rigid structure ensuring consistency and easy retrieval.



Examples of structured data

Spreadsheets
Data arranged in rows and columns

SQL databases
Data stored in tables and columns

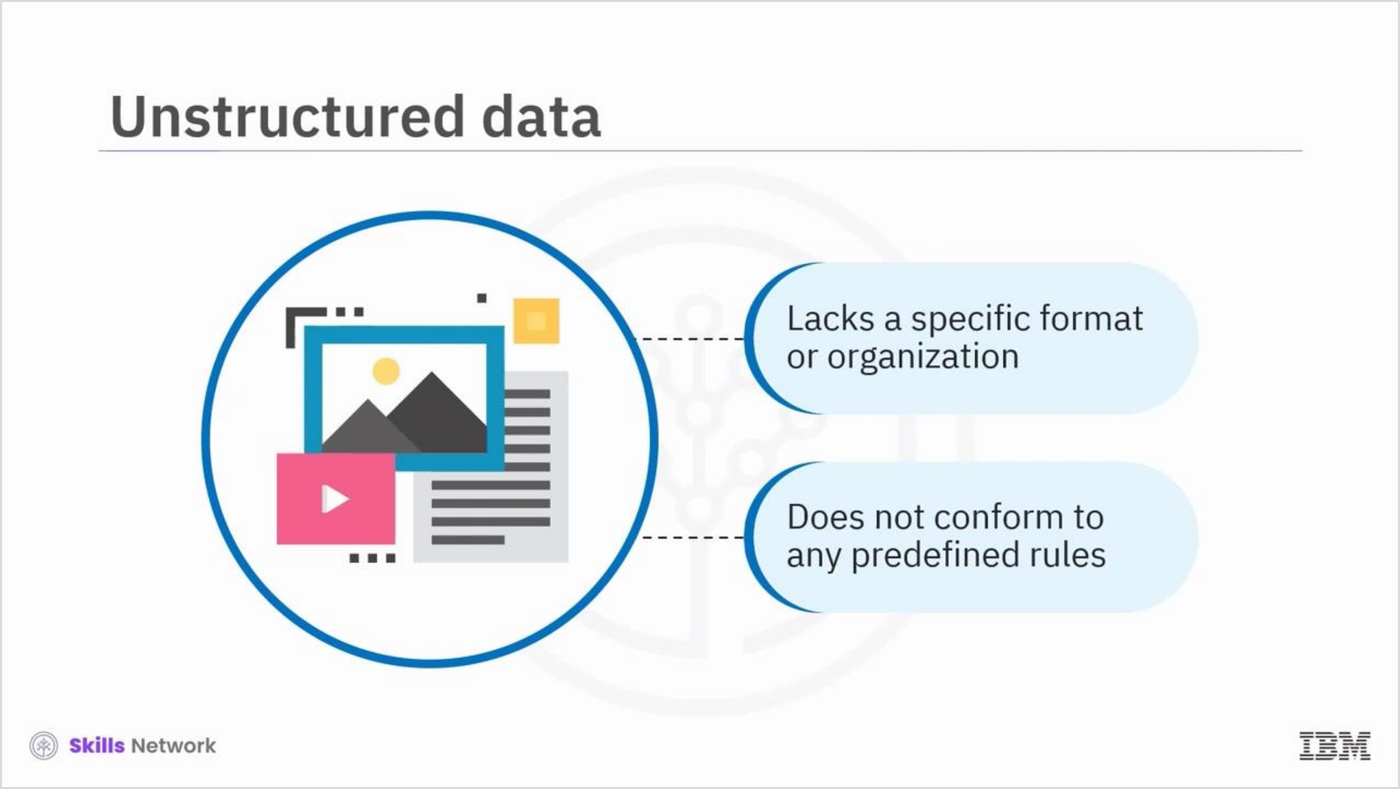
Online forms
Data stored in records in designated fields

SQL
Skills Network
IBM

Examples of structured data include Excel spreadsheets, arrange data into rows and columns, assigning each piece of data a specific cell address.

SQL databases that allow data to be stored in predefined tables and columns.

Online forms collect customer information by storing each piece of data; name, address, credit card number in designated fields.

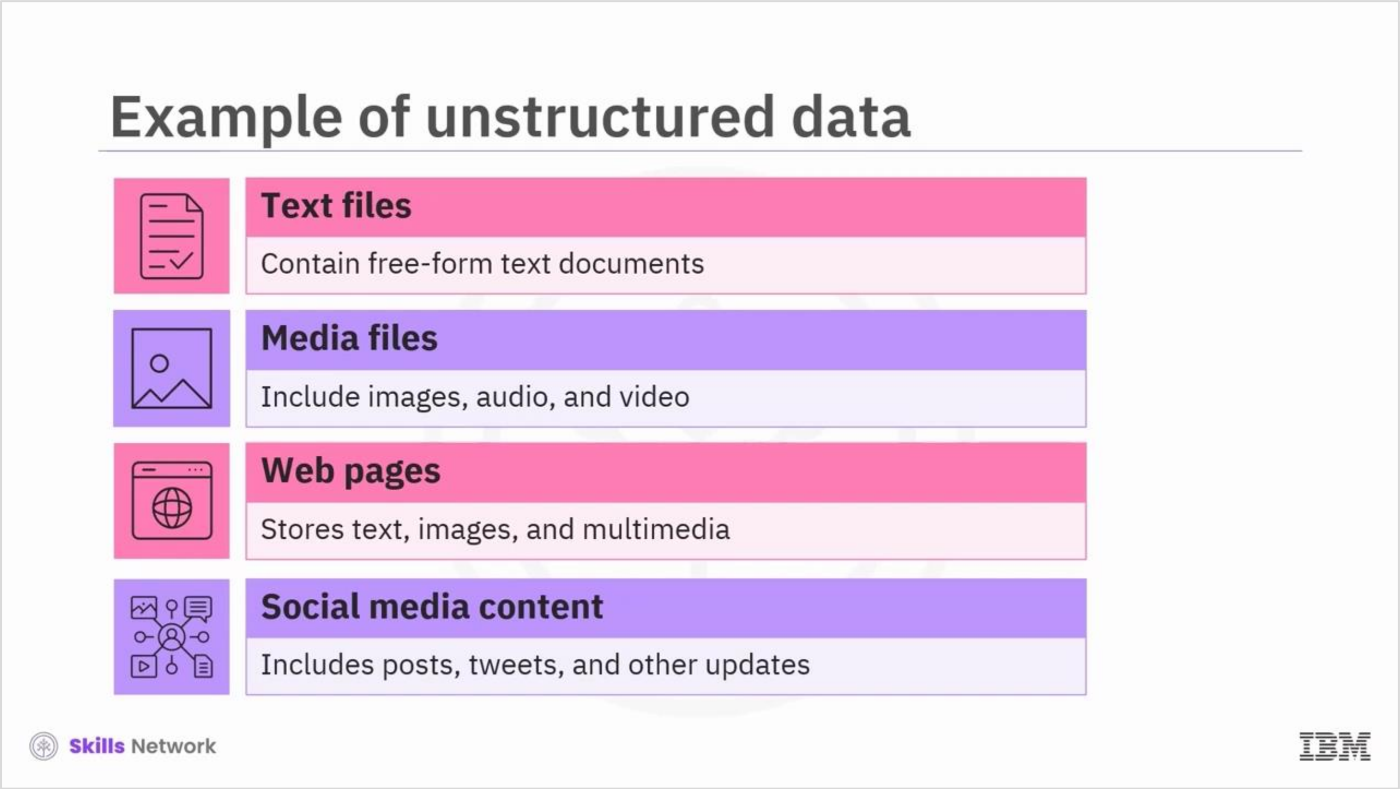


Unstructured data

Lacks a specific format or organization

Does not conform to any predefined rules

In contrast, unstructured data lacks a specific format or organization. It doesn't conform to any predefined rules or sequences, making it challenging to process and analyse using traditional methods.



Example of unstructured data

Text files
Contain free-form text documents

Media files
Include images, audio, and video

Web pages
Stores text, images, and multimedia

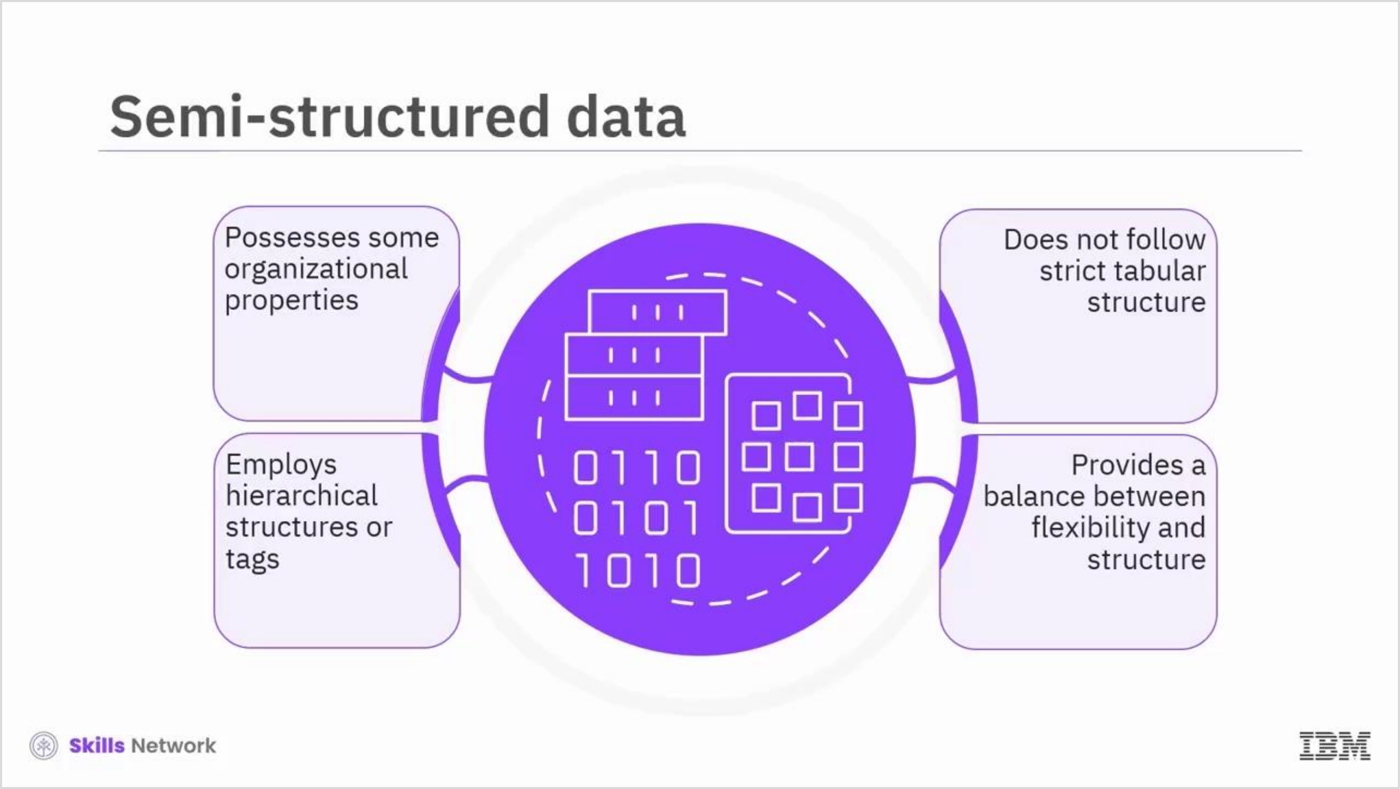
Social media content
Includes posts, tweets, and other updates

Examples of unstructured data include text files contain free-form text documents without a predefined structure.

Media files include images, audio, and video.

Web pages, main content, such as text, images, and multimedia is often unstructured despite potentially structured elements like HTML tags.

Social media content encompasses posts, tweets, and other updates with mixed text, images, and links.



Semi-structured data

Possesses some organizational properties

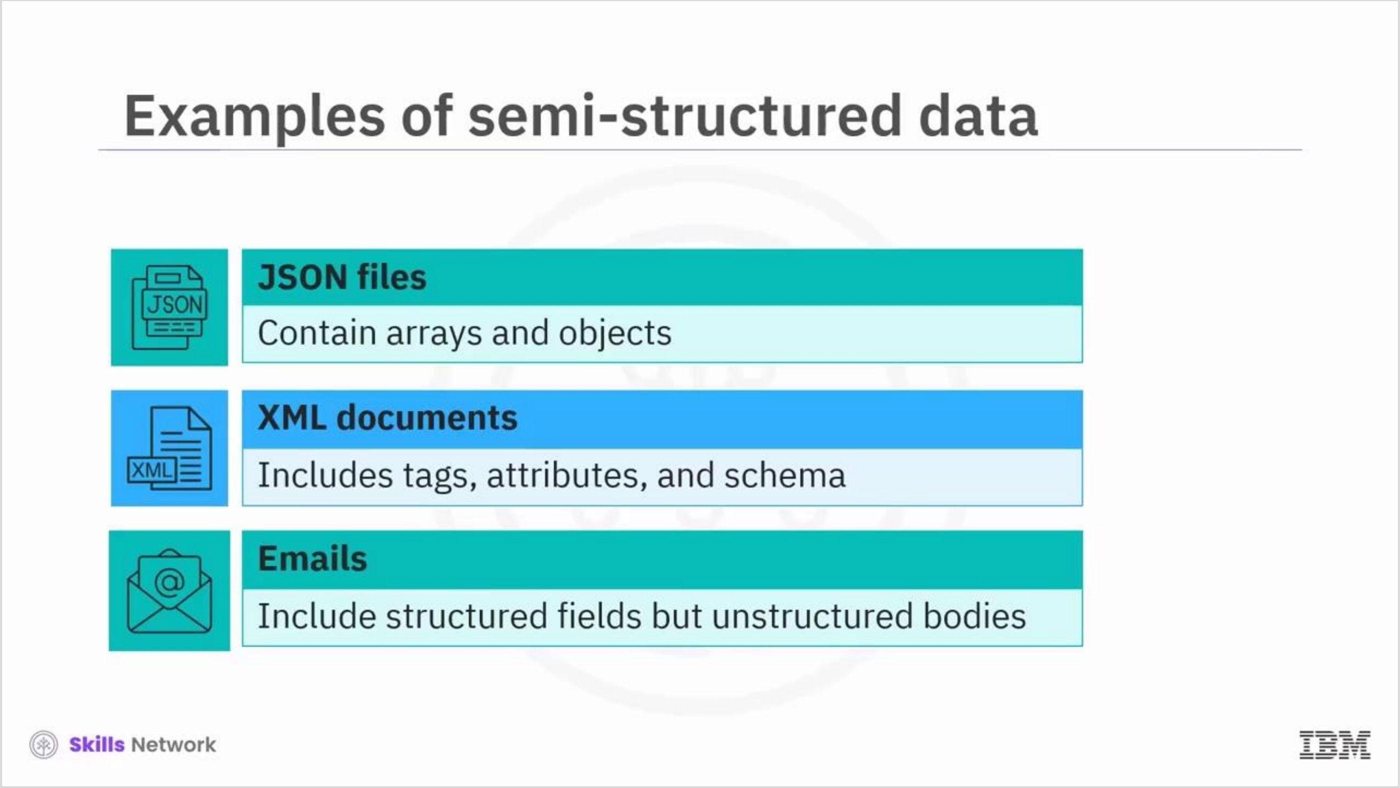
Employs hierarchical structures or tags

Does not follow strict tabular structure

Provides a balance between flexibility and structure

The last category we will discuss is semi-structured data.

This type of data possesses some organizational properties but does not adhere to a strict tabular structure. It often employs hierarchical structure or tags to organize information, providing a balance between flexibility and structure.



Examples of semi-structured data

JSON files

Contain arrays and objects

XML documents

Includes tags, attributes, and schema

Emails

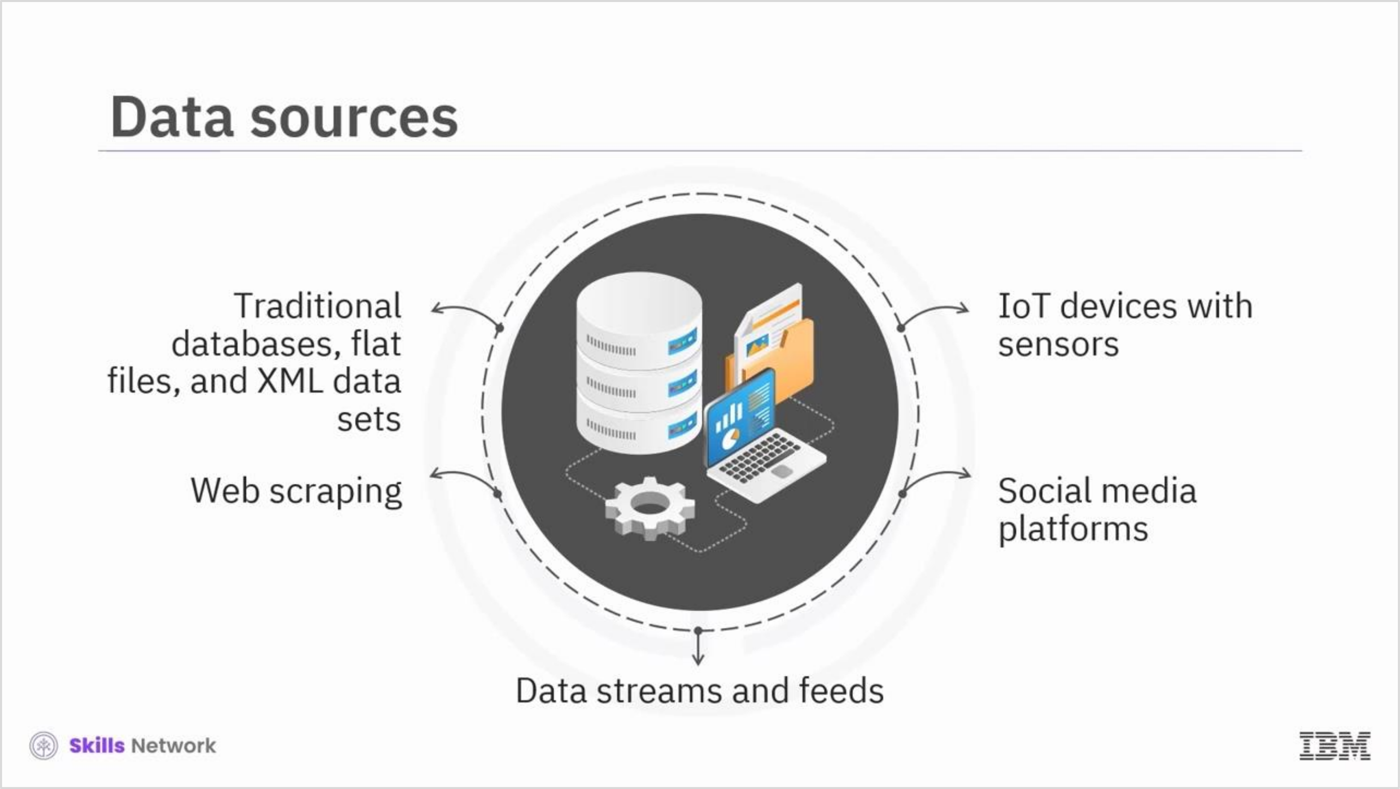
Include structured fields but unstructured bodies

Examples of semi-structured data include-

JSON files contain arrays and objects using specific tags or keys to mark different data elements.

XML documents define data structure using tags, attributes, and schema.

Emails have structured fields; to, from, subject but unstructured message bodies.



Data sources

Traditional databases, flat files, and XML data sets

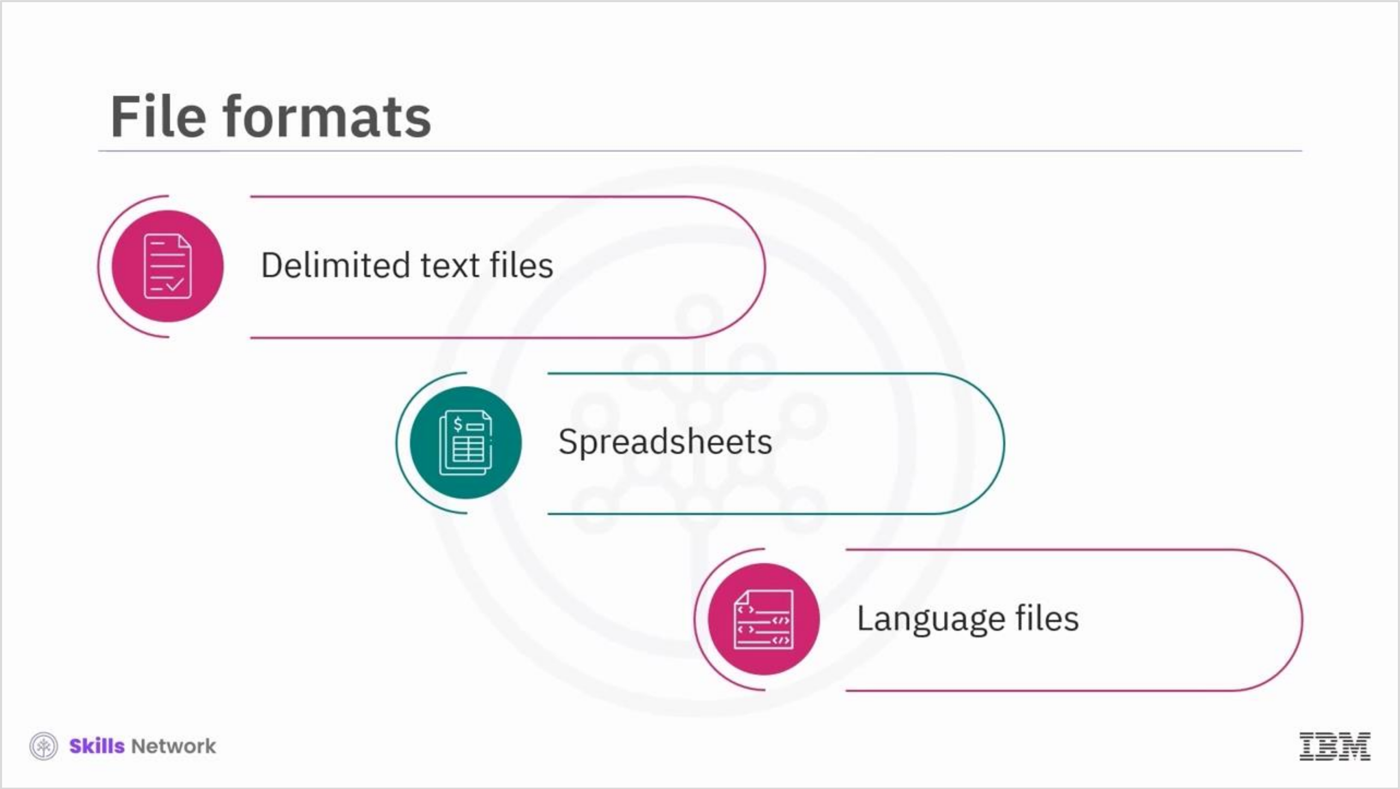
Web scraping

IoT devices with sensors

Social media platforms

Data streams and feeds

Businesses today access a wealth of data from various sources, including traditional databases, flat files, and XML data sets, web scraping, data streams, and feeds, social media platforms, IoT devices with sensors.



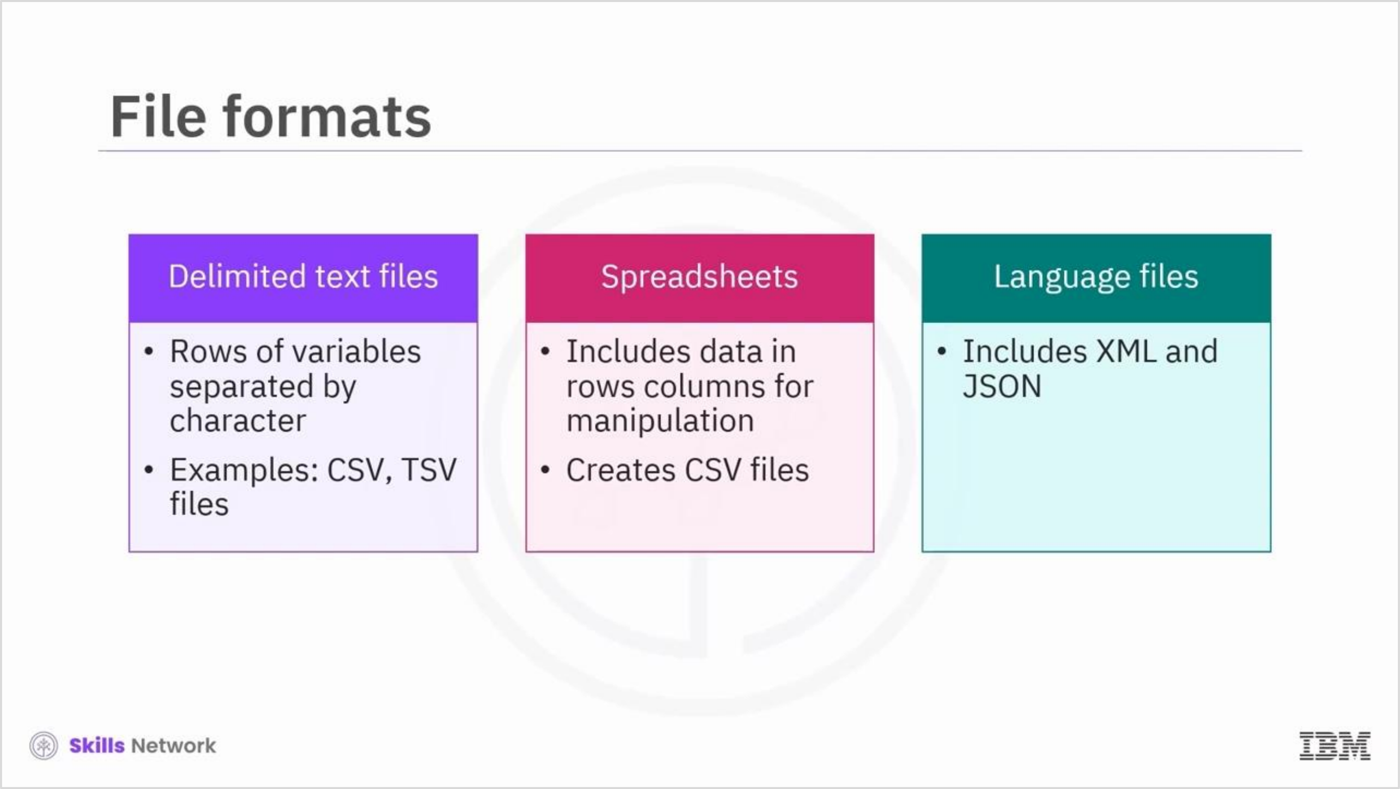
File formats

Delimited text files

Spreadsheets

Language files

We can hold or transfer data between systems in many different file formats. Common file formats include delimited text files, spreadsheets, language files.



File formats

Delimited text files

- Rows of variables separated by character
- Examples: CSV, TSV files

Spreadsheets

- Includes data in rows columns for manipulation
- Creates CSV files

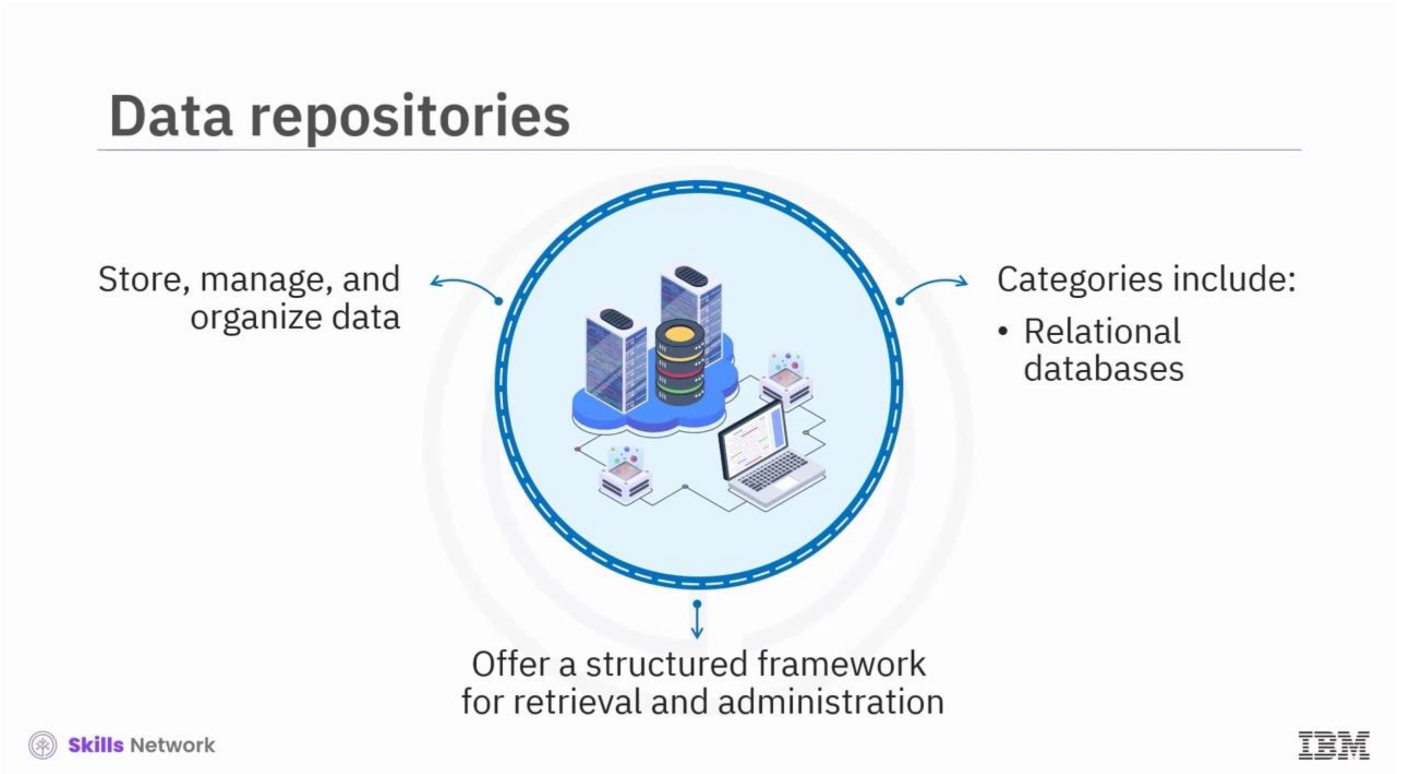
Language files

- Includes XML and JSON

Delimited text files. In these files, data resides in rows with each variable separated by a specific character, like a comma or a tab. Delimited files comprise comma-separated variable, CSV, and tab-separated variable, TSV, files.

Spreadsheets. In these files, data exists in rows and columns, the same as a table, facilitating easy access and manipulation. A spreadsheet helps to create CSV files.

Language files. Language files like extensible markup language (XML) and JavaScript Object Notation, (JSON), have set rules and structures for encoding data to send over the internet.



Data repositories

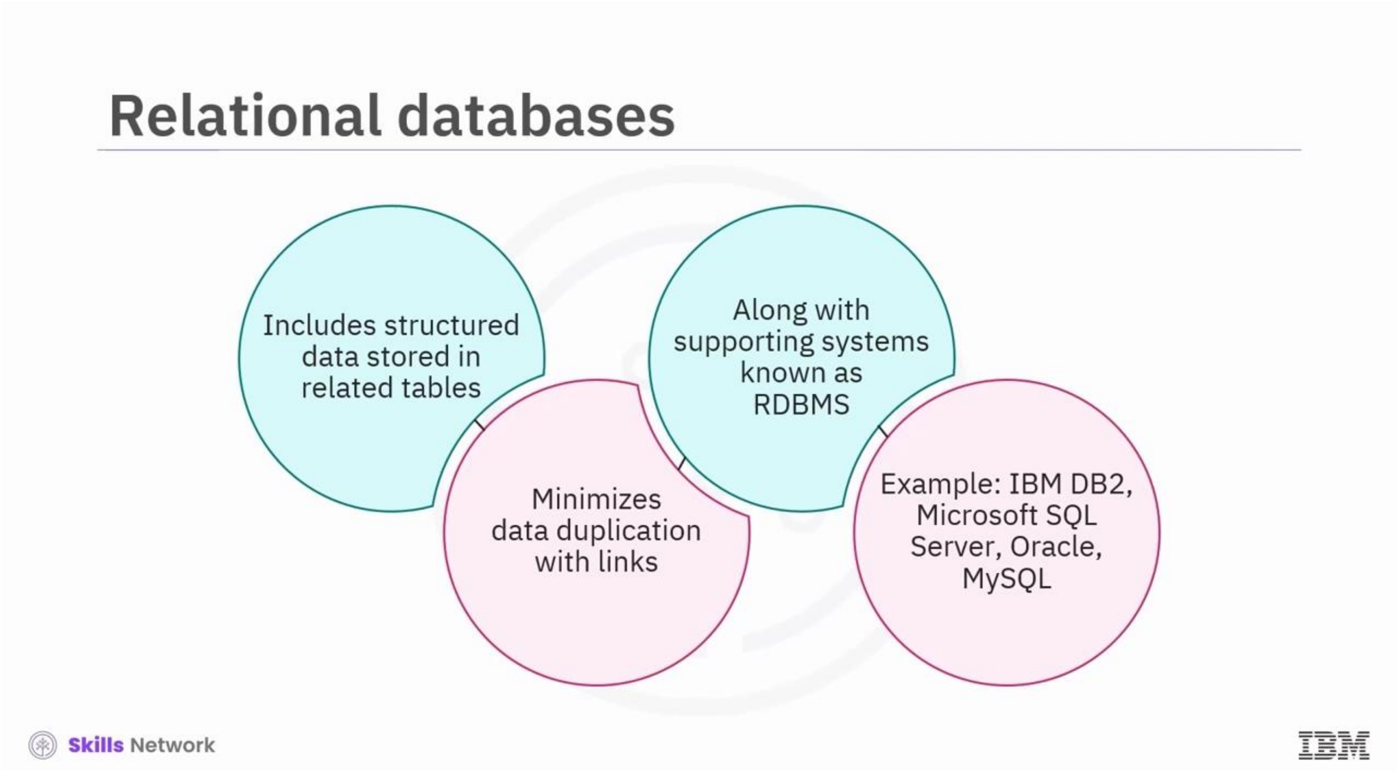
Store, manage, and organize data

Categories include:

- Relational databases

Offer a structured framework for retrieval and administration

After collecting your data, where's the best place to store it? Data repositories actively store, manage, and organize data in a centralized manner, offering a structured framework for efficient retrieval and administration. Two major categories of data repositories are relational databases and non-relational databases.



Relational databases

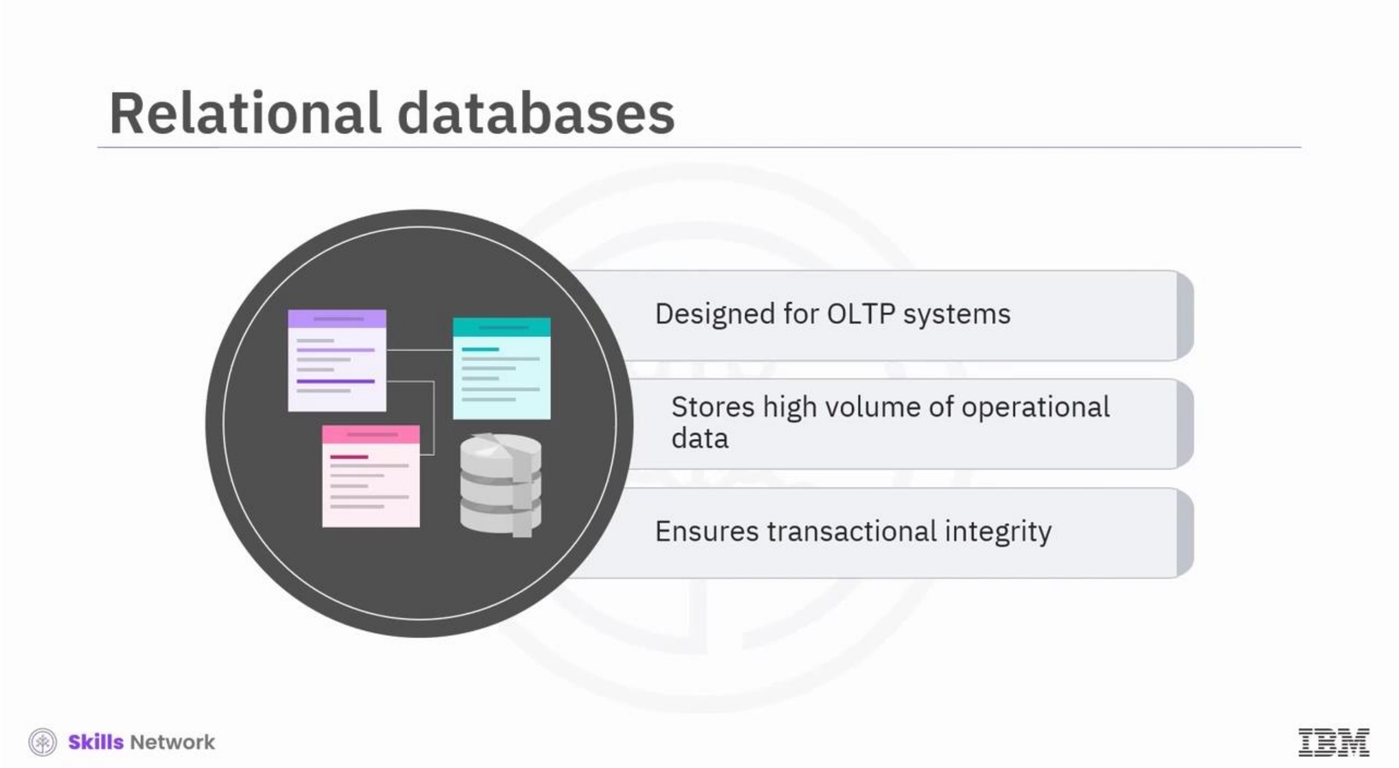
Includes structured data stored in related tables

Minimizes data duplication with links

Along with supporting systems known as RDBMS

Example: IBM DB2, Microsoft SQL Server, Oracle, MySQL

Relational databases consist of structured data stored in related tables. The links between the tables minimize data duplication while preserving intricate relationships. These databases and their supporting systems are collectively known as Relational Database Management Systems, RDBMS. Prominent examples include IBM DB2, Microsoft SQL Server, Oracle, and MySQL.



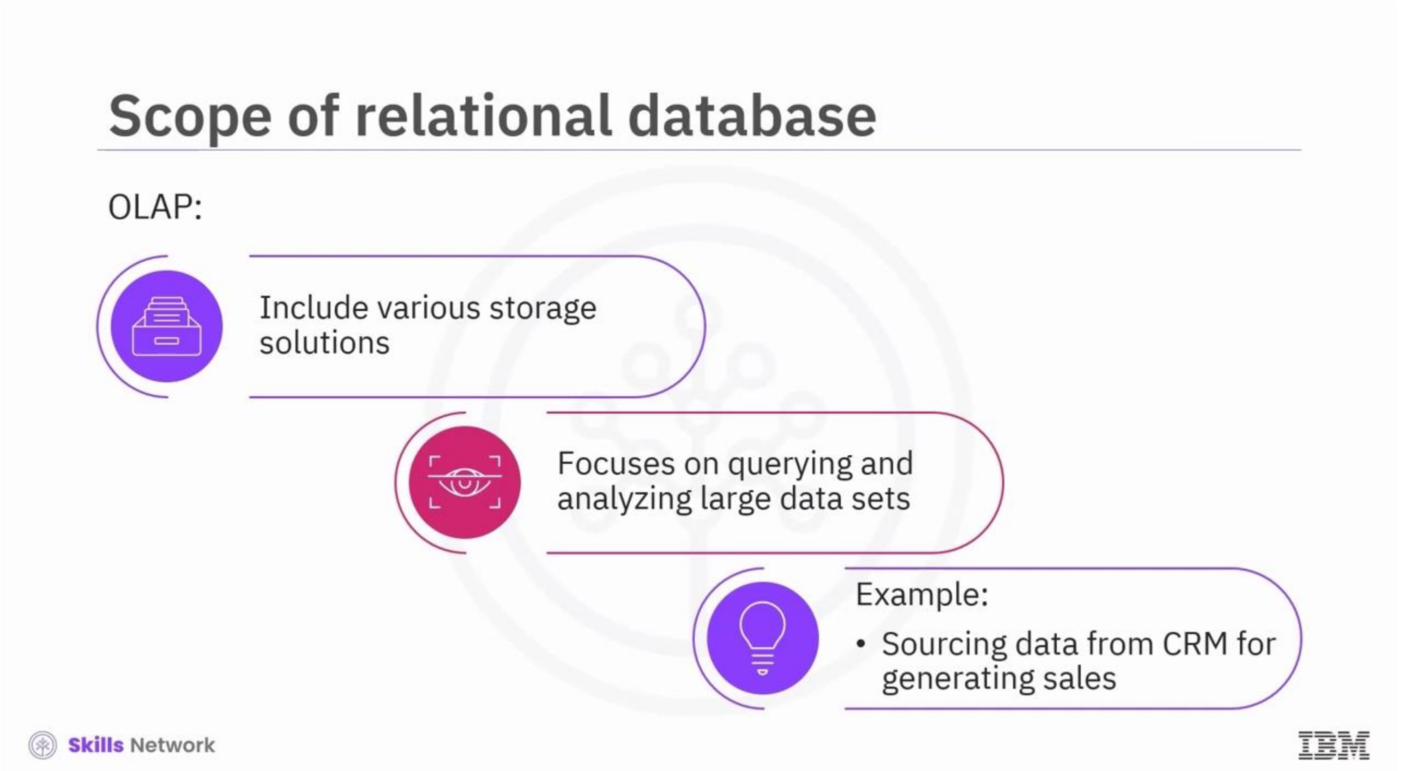
Relational databases

Designed for OLTP systems

Stores high volume of operational data

Ensures transactional integrity

Let us look at the scope of relational databases. Relational databases are primarily OLTP systems used to support day-to-day business activities such as customer transactions, human resource activities, and workflows. They help in storing a high volume of day-to-day operational data that many businesses rely on. Their normalized structure ensures transactional integrity and supports concurrent access for routine operations.



Scope of relational database

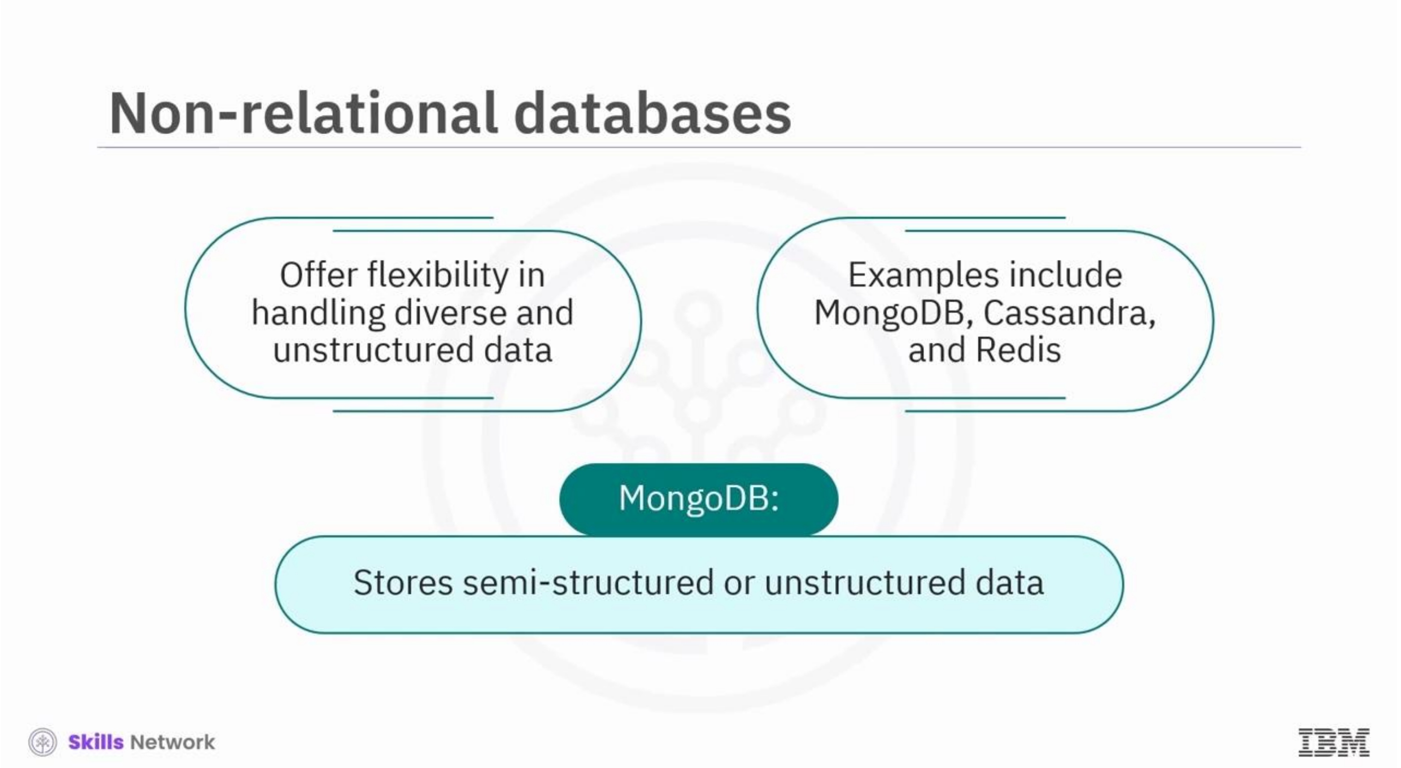
OLAP:

- Include various storage solutions
- Focuses on querying and analyzing large data sets
- Example:
  - Sourcing data from CRM for generating sales

Skills Network

IBM

OLAP systems include various storage solutions, like relational and non-relational databases, data warehouses, data lakes, and big data stores. The system focuses on querying and analysing large datasets to extract meaningful insights. For example, organizations can leverage data sourced from a customer relationship management CRM system for insightful analytics, such as generating sales projections.



Non-relational databases

Offer flexibility in handling diverse and unstructured data

Examples include MongoDB, Cassandra, and Redis

MongoDB:

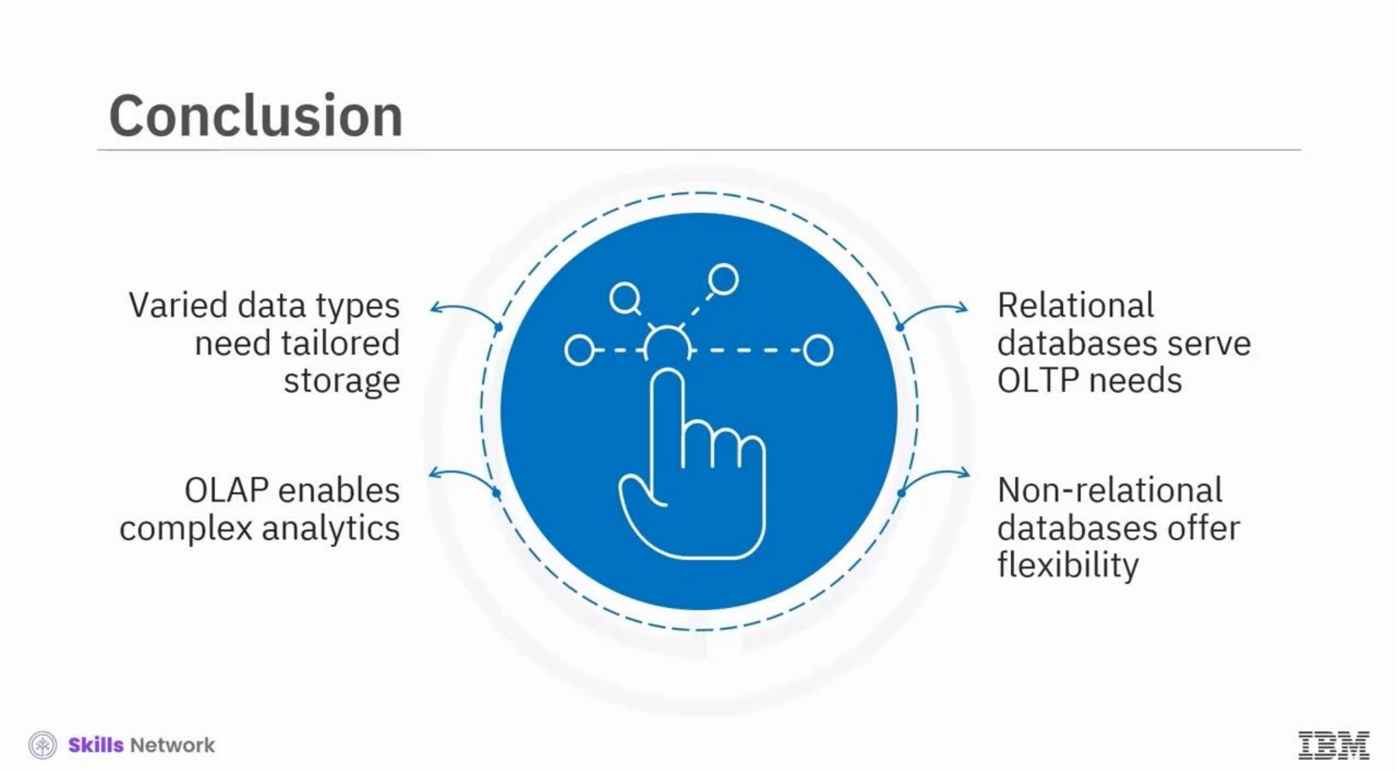
Stores semi-structured or unstructured data

Skills Network

IBM

Now that you know about relational databases, let us explore non-relational databases. They offer flexibility in handling diverse and unstructured data. MongoDB, Cassandra, and Redis represent this category.

MongoDB is a document-oriented database suitable for storing and managing semi-structured or unstructured data.



Conclusion

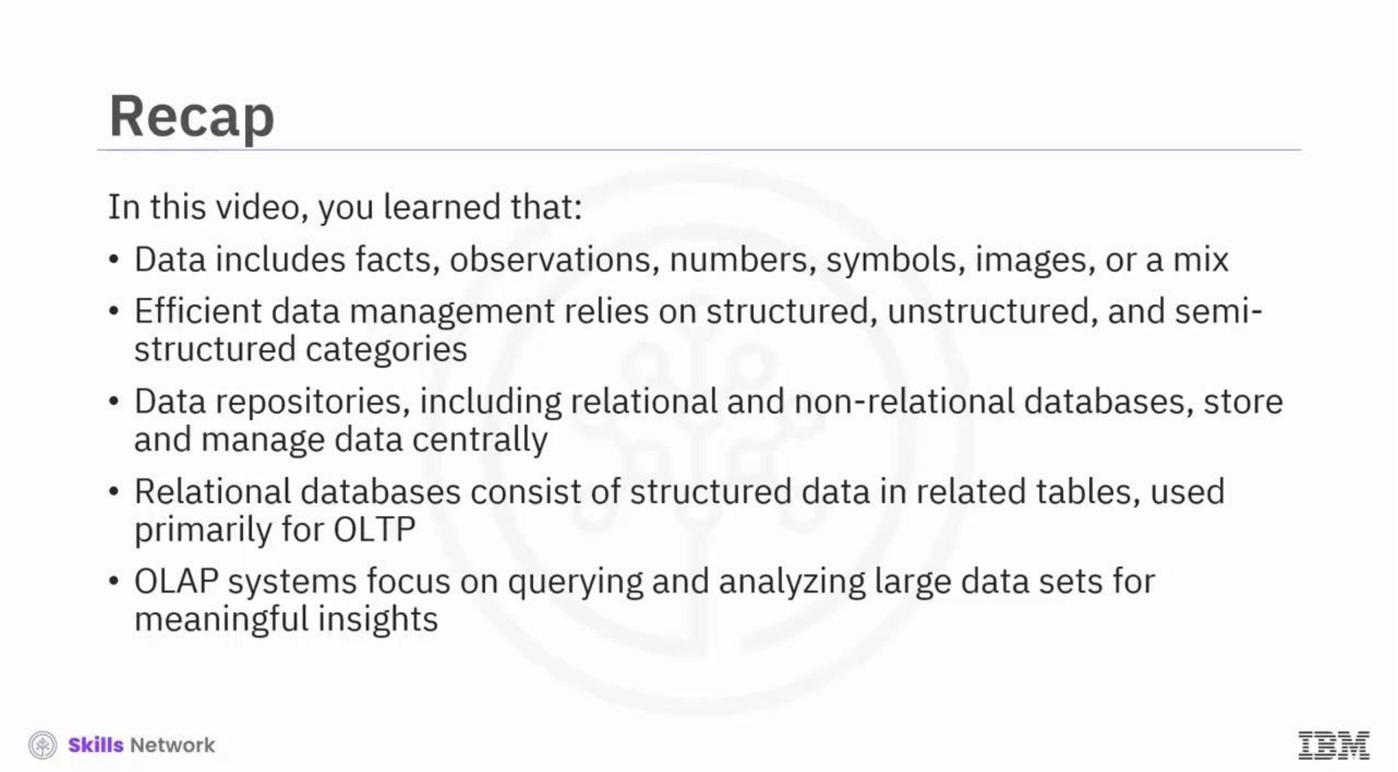
Varied data types need tailored storage

OLAP enables complex analytics

Relational databases serve OLTP needs

Non-relational databases offer flexibility

In conclusion, data's varied types and structures demand appropriate storage solutions with relational databases serving OLTP needs, OLAP systems enabling complex analytics, and non-relational databases providing flexibility for diverse data.



Recap

In this video, you learned that:

- Data includes facts, observations, numbers, symbols, images, or a mix
- Efficient data management relies on structured, unstructured, and semi-structured categories
- Data repositories, including relational and non-relational databases, store and manage data centrally
- Relational databases consist of structured data in related tables, used primarily for OLTP
- OLAP systems focus on querying and analyzing large data sets for meaningful insights

In this video, you learned that data includes facts, observations, numbers, symbols, images, or a mix. Efficient data management relies on structured, unstructured, and semi-structured categories. Data repositories, including relational and non-relational databases store and manage data centrally. Relational databases consist of structured data in related tables used primarily for OLTP. OLAP systems focus on querying and analysing large datasets for meaningful insights.