UNIVERSITI KEBANGSAAN MALAYSIA

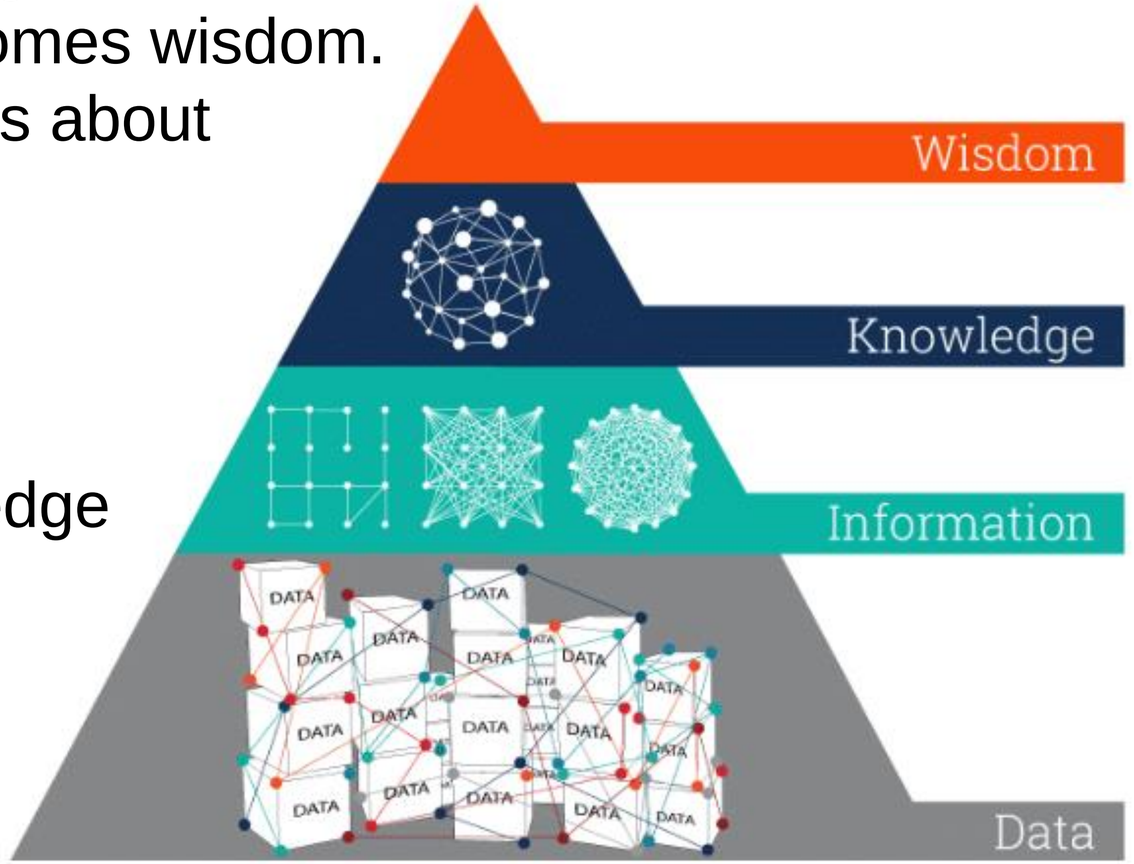
The National University of Malaysia

STQD6114
INTRODUCTION TO DATA STRUCTURES

Nor Hamizah Miswan

What is data?

- The Data, Information, Knowledge, Wisdom (DIKW) Pyramid represent the relationship between data, information, knowledge.
- Each building block is a step towards a higher level – first comes data, then is information, next is knowledge and finally comes wisdom.
- Each step up the pyramid answers questions about the initial data and adds value to it.
- The more questions we answer, the higher we move up the pyramid.
- In other words, the more we enrich our data with meaning and context, the more knowledge and insights we get out of it.
- At the top of the pyramid, we have turned the knowledge and insights into a learning experience that guides our actions.



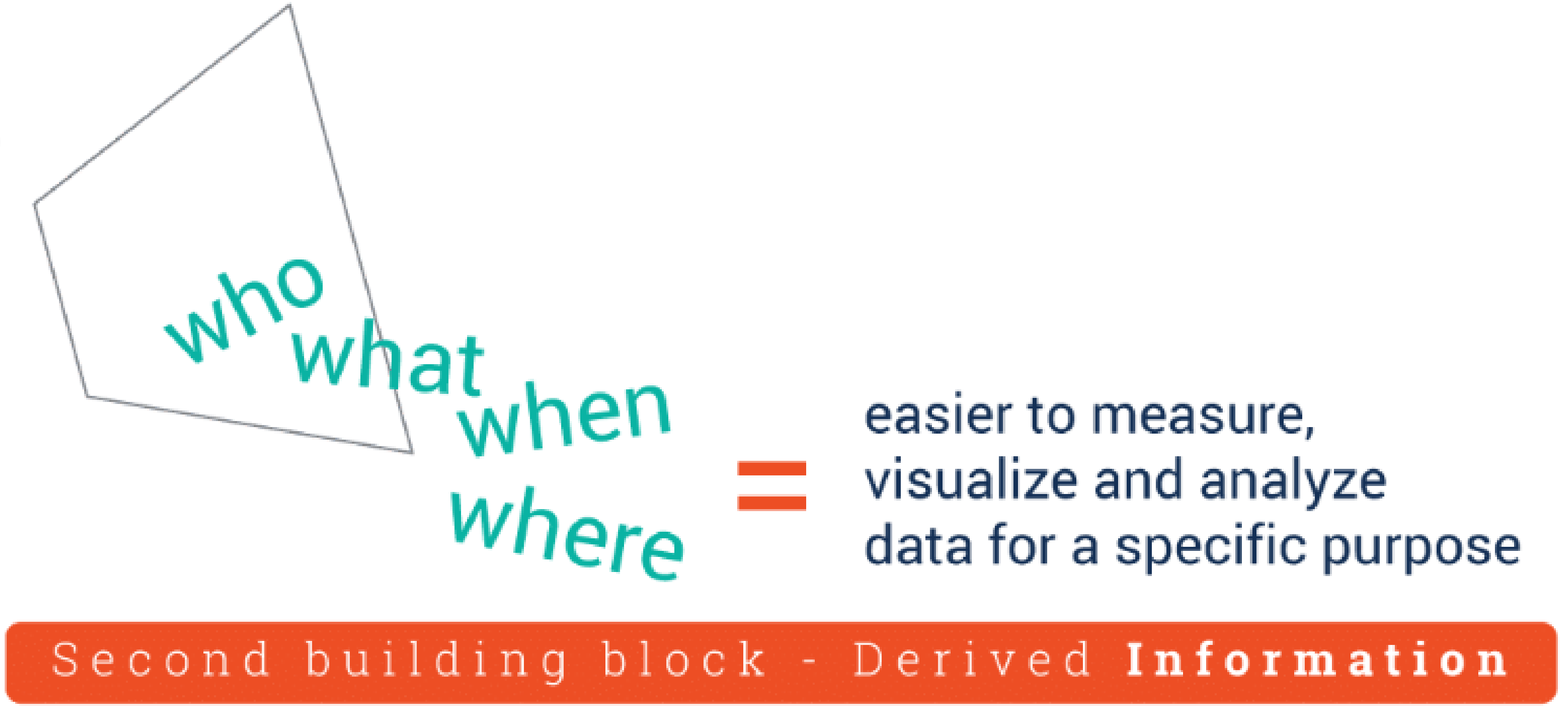
What is data?

- DATA is a collection of facts in a raw or unorganized form such as numbers or characters.
- However, without context, data can mean little. For example, *12012012* is just a sequence of numbers without apparent importance. But if we view it in the context of ‘this is a date’, we can easily recognize *12th of January, 2012*. By adding context and value to the numbers, they now have more meaning.
- In this way, we have transformed the raw sequence of numbers into INFORMATION.

= a collection of facts in a raw or unorganized form

Base building block - Raw **Data**

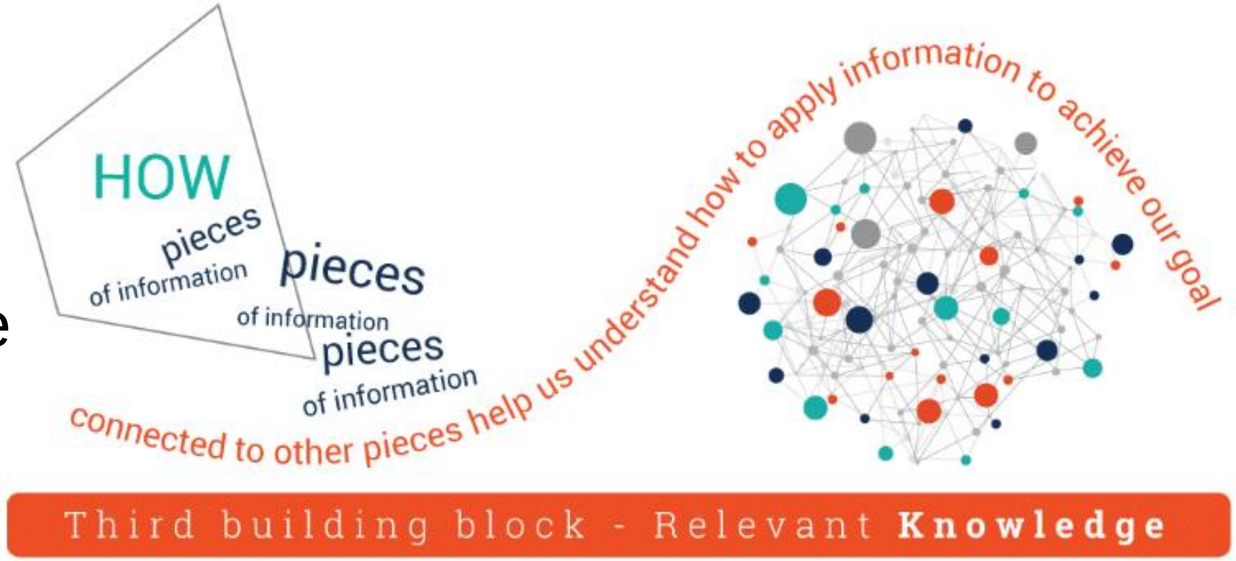
What is data?



- INFORMATION is the next building block of the DIKW Pyramid.
- This is data that has been "cleaned" of errors and further processed in a way that makes it easier to measure, visualize and analyze for a specific purpose.
- By asking relevant questions about 'who', 'what', 'when', 'where', etc., we can derive valuable information from the data and make it more useful for us.
- But when we get to the question of 'how', this is what makes the leap from information to KNOWLEDGE.

What is data?

- “How” is the information, derived from the collected data, relevant to our goals? “How” are the pieces of this information connected to other pieces to add more meaning and value?
- When we don’t just view information as a description of collected facts, but also understand how to apply it to achieve our goals, we turn it into knowledge.
- But only when we use the knowledge and insights gained from the information to take proactive decisions, we can say that we have reached the final – ‘wisdom’ – step of the Knowledge Pyramid.



What is data?

- In the heart of the computer, all data are represented in binary. One binary digit, or bit, is the smallest chunk of data that we can send from one place to another.
- Although all data are at heart binary, computers and software help to represent data in more convenient forms for people to see.
- Three important representations are: "character" for representing text, "integer" for representing numbers with no digits after the decimal point, and "floating point" for numbers that may have digits after the decimal point.

What is structured data?

- Structured data is a data whose elements are addressable for effective analysis
- It has been organized into a formatted repository that is typically a database
- Stored in tabular format
- Clearly defined
- Today, those data are most processed in development and simplest way to manage information.
- Example:

Excel files

SQL databases

What is structured data?

- The rows and columns are related to each other

ID	NAME	ADDRESS	PHONE NO

Proper view and understanding of data

What is unstructured data?

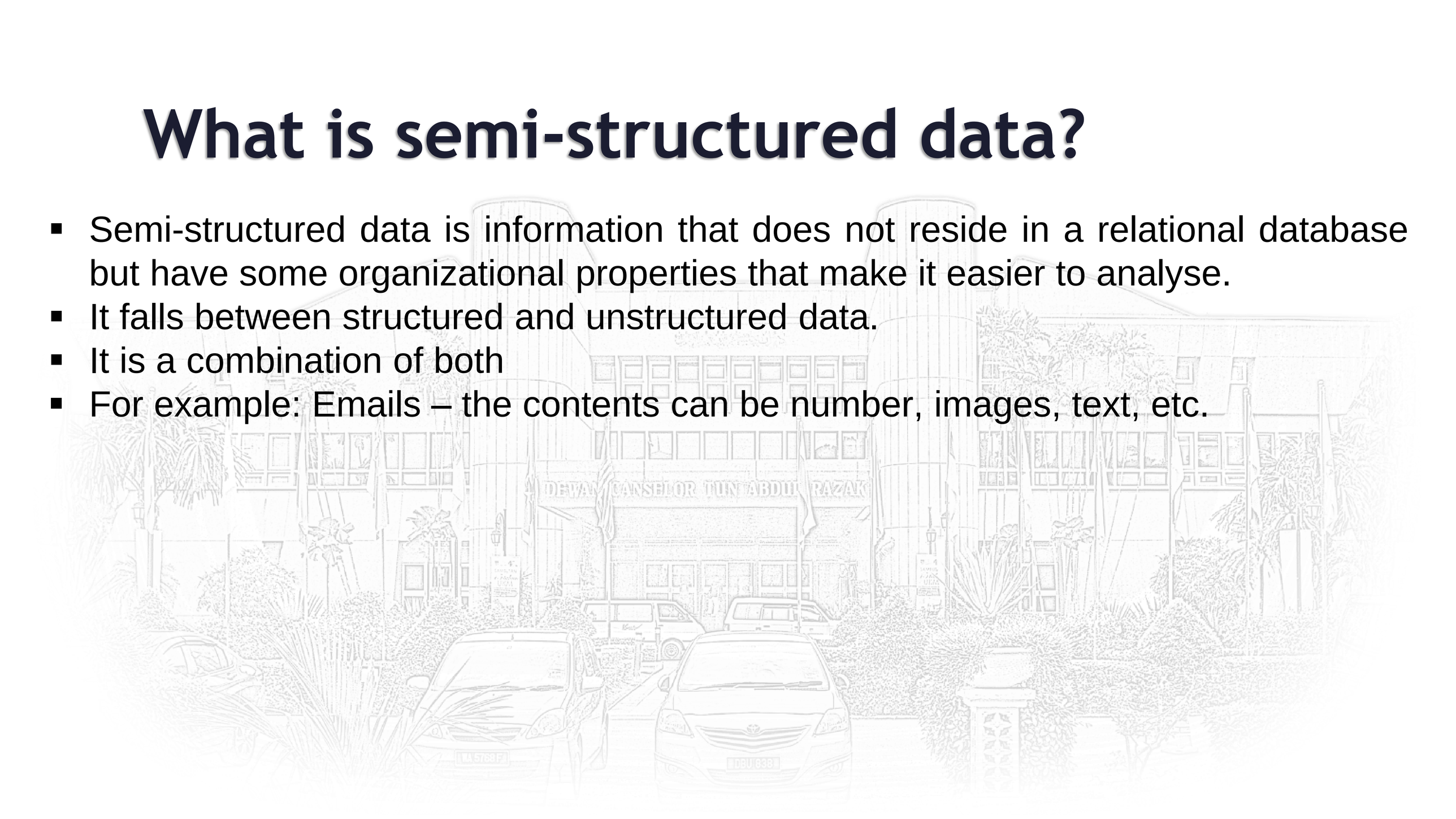
- No pre-defined structure
- No data model
- Data is irregular and ambiguous
- Easier to extract data
- Hence 80-90% of data is unstructured
- It is a combination of text, numbers, audio, video, images, messages, social media posts, etc.

What is unstructured data?

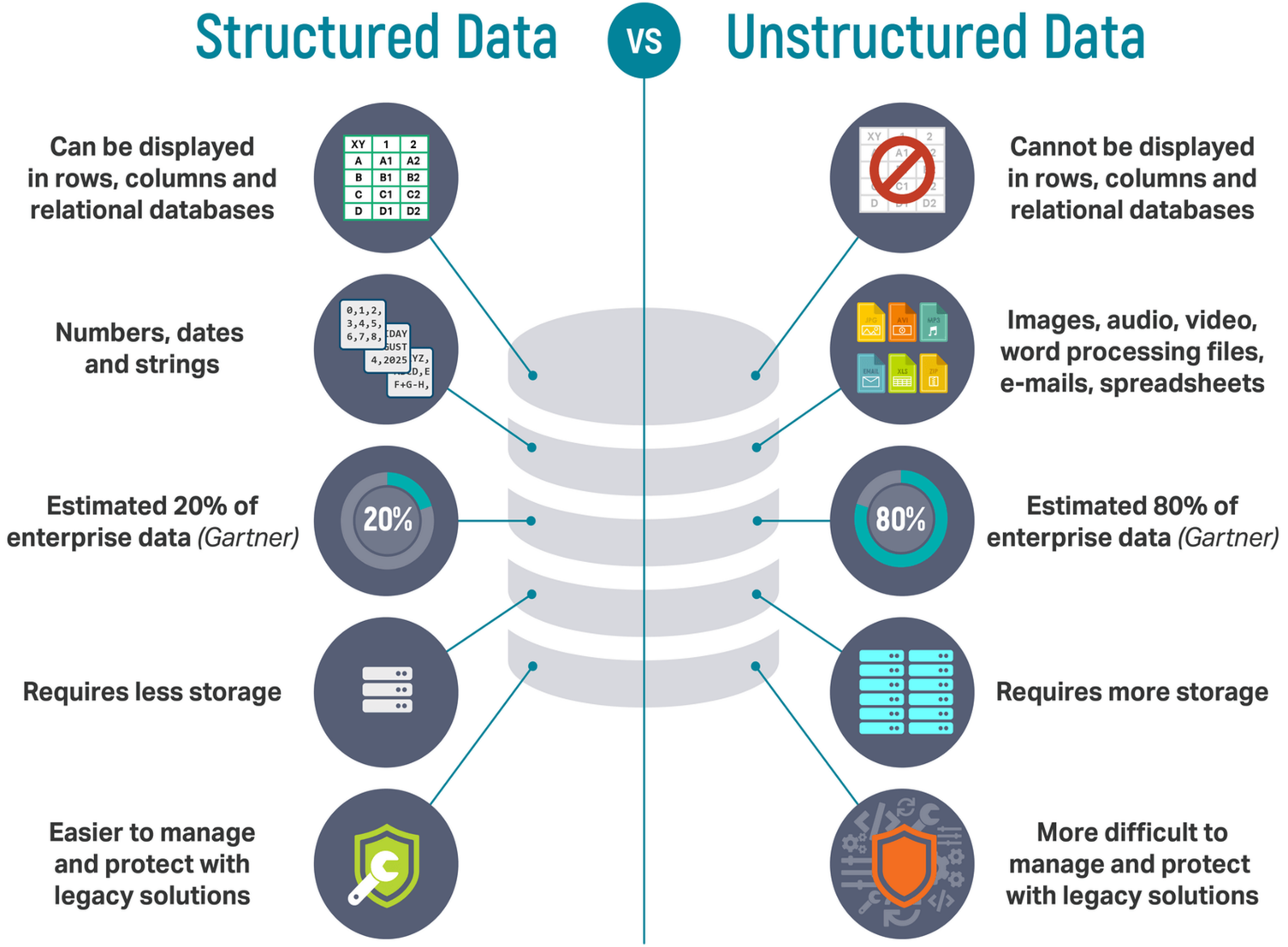
- Surveys are also unstructured data when the required answers are open-ended questions.
  Eg. What is your favourite kind of coffee?
  How does coffee make you feel? Please elaborate.
- Previously only structured data was used extensively.
- Unstructured data is the most useful kind of data.
- It provides a lot of information.

What is semi-structured data?

- Semi-structured data is information that does not reside in a relational database but have some organizational properties that make it easier to analyse.
- It falls between structured and unstructured data.
- It is a combination of both
- For example: Emails – the contents can be number, images, text, etc.



Unstructured vs structured data



Project 1 - Question 1 (3%)

Answer the following questions based on video https://youtu.be/dK4aGzeBPkk.

1. Two popular examples of big data technology are Netflix and credit-card. Explain how big data is used in these examples.
2. For EVERY application of big data found in different fields (e.g., banking telecommunication, healthcare, media, etc.) mentioned in the video, describe two real-life examples of each field.
3. Construct the table to differentiate between structured, unstructured and semi-structured. Give at least 3 points, including sources of the data.