**1. Data structures**

Awesome job on Chapter 1! Let's continue our exploration of the world of data engineering. This second chapter will focus on storage. In this lesson, we're going to learn more about data structure.

**2. Structured data**

Structured data is easy to search and organize. Data is entered following a rigid structure, like a spreadsheet where there are set columns. Each column takes values of a certain type, like text, data, or decimal. It makes it easy to form relations, hence it's organized in what is called a relational database. About 20% of the data is structured. SQL, which stands for Structured Query Language, is used to query such data.

**3. Employee table**

Here is an example of structured data. This is an extract of Spotflix's employee table. It's easy to read the table and well-organized. You can see it follows a model: each row expects an employee and each column a specific information about that employee (team, role). Each column needs to be of a certain type. The index is a number, and acts as a unique ID, because two employees may have the same name, last name, or both. The penultimate column holds logical values: values can only be true or false. For example, Rick Sanchez is part-time. The rest of the columns are text.

**4. Relational database**

Because it's structured we can easily relate this table to other structured data. For example, if there's another table holding information about offices,

**5. Relational database**

we can connect on the office column. Tables that can be connected that way form a relational database.

**6. Semi-structured data**

Semi-structured data resembles structured data, but allows more freedom. It's therefore relatively easy to organize, and pretty structured, but allows more flexibility. It also has different types and can be grouped to form relations, although this is not as straightforwards as with structured data - you have to pay for that flexibility at some point. Semi-structured data is stored in NoSQL databases (as opposed to SQL) and usually leverages the JSON, XML or YAML file formats.

**7. Favorite artists JSON file**

Here is an example of a JSON file storing the favorite artists of each Spotflix user. As you can see, the model is consistent: each user id contains the user's last and first name, and their favorite artists. However, the number of favorite artists may differ: I have four, Sara has two and Lis has three favorite artists. Relational databases don't allow that kind of flexibility, but semi-structured formats let you do it.

**8. Unstructured data**

Unstructured data is data that does not follow a model and can't be contained in a rows and columns format. This makes it difficult to search and organize. It's usually text, sound, pictures or videos. It's usually stored in data lakes, although it can also appear in data warehouses or databases - don't worry, we will cover the differences between these at the end of this chapter. Most of the data around us is unstructured. Unstructured data can be extremely valuable, but because it's hard to search and organize, this value could not be extracted until recently, with the advent of machine learning and artificial intelligence.

**9. Lyrics**

At Spotflix, unstructured data consists in lyrics,

**10. Songs**

songs,

**11. Pictures**

albums pictures and artists profile pictures,

**12. Videos**

and music videos.

**13. Adding some structure**

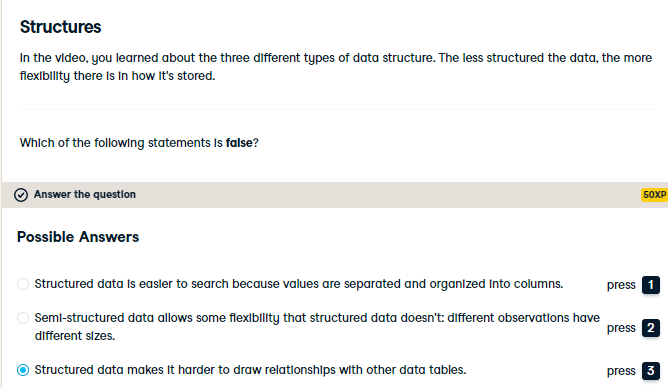
At Spotflix, we could use machine learning algorithms to parse song spectrums, analyze beats per minute, chord progressions, genres to help categorize songs. Or, we could have artists give additional information when they upload their songs. Having them add the genre, and some tags, would make it semi-structured data, and would make searching and organizing easier.

**14. Summary**

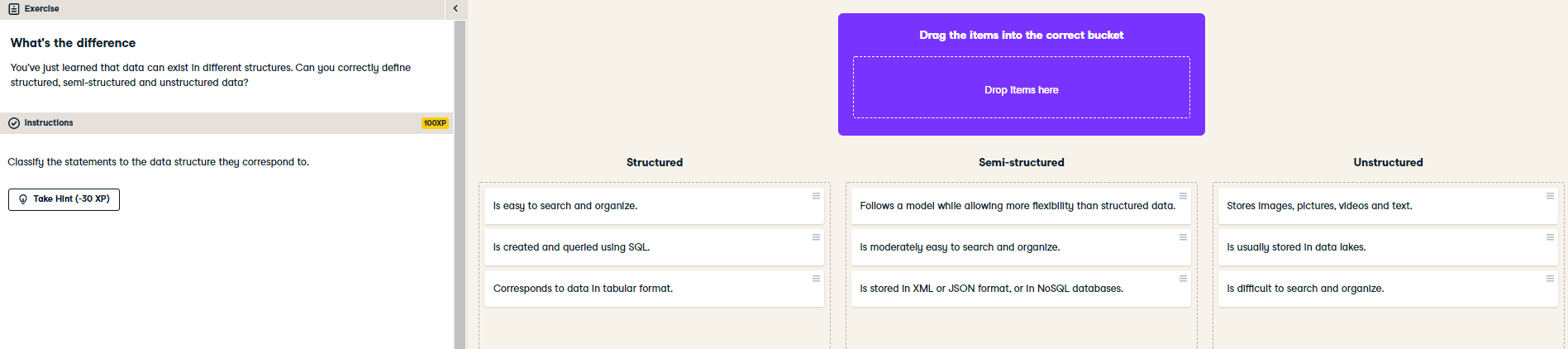
All right, now you know what is characteristic of structured data, semi-structured data and unstructured data, the differences between the three, and you're able to give examples for each of them.

**15. Let's practice!**

Let's consolidate this knowledge with some exercises!



Good job! This statement is completely false: structured data make it super easy to draw relationships between tables.



Excellent job on this one! Structured, semi-structured and unstructured data are fundamental concepts in data engineering.