Week 2 - Data Structures and Databases

Data Fun Facts

- 90 % of all data in existence today was created **in the last two years**.

- In 2018, nearly $\frac{3}{4}$ of companies reported that unstructured data was growing "somewhat faster" or "much faster" than other business data, leading to the estimate that by 2025, 80 % of data will be unstructured. *Source: IDC International Data Corporation*

But what is structured and unstructured data?

Structured Data

- Pre-defined model for storing the data

- Data usually comes in tabular form with relationships between different rows and columns
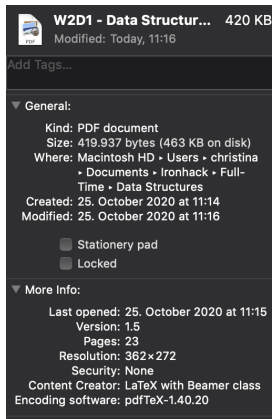
- Examples: Excel files, SQL databases

Unstructured Data

- Not organized in a pre-defined manner.

- Often text-heavy, but contains different types of content.
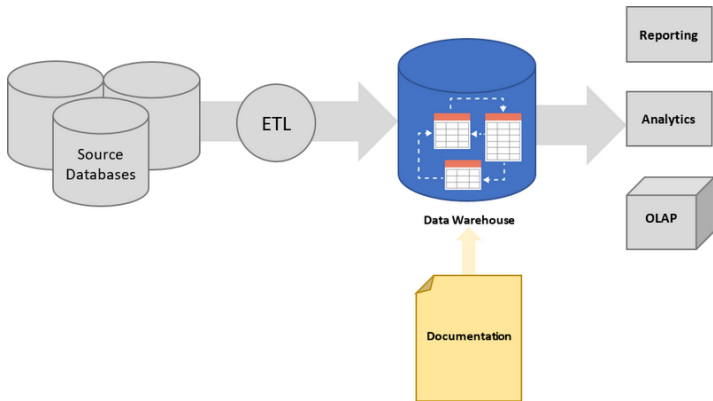
- Storage: e.g. MongoDB.

Semi-Structured Data

- Data does not come in table form, but contains e.g. tags for identification.

- Storage in formats such as XML and json.
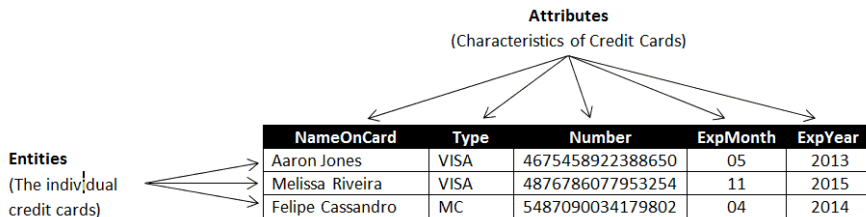
- You'll see this in web scraping!

Metadata

Database vs Data Warehouse

Structured Data

Entities vs Attributes

Attributes
(Characteristics of Credit Cards)

Entities
(The individual credit cards)

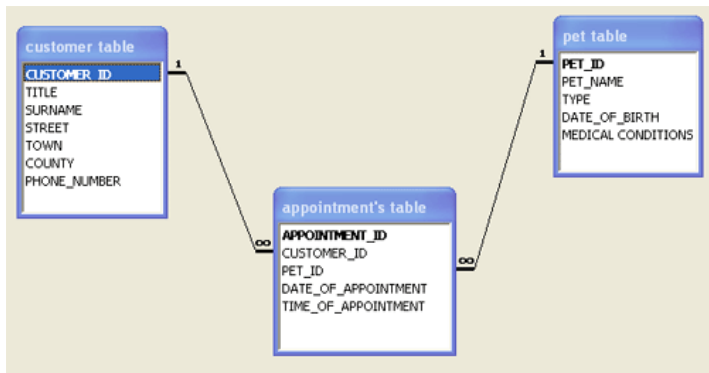| NameOnCard	Type	Number	ExpMonth	ExpYear
Aaron Jones	VISA	4675458922388650	05	2013
Melissa Riveira	VISA	4876786077953254	11	2015
Felipe Cassandro	MC	5487090034179802	04	2014

What is a Relational Database?

Flat File Structure

Date	Time	Pet Name	Type	Medical Conditions	Owner Title	Owner Surname	Owner Street	Owner Town	Owner Phone
20.09.2020	15:00	Meowy	Cat	Heartworm	Mr	Smith	Beethovenstr. 20	Berlin	01766262678
20.09.2020	16:00	Woof	Dog	Broken Leg	Ms	Jane	Goethe Str. 54	Berlin	01756272909
21.09.2020	12:00	Meowy	Cat	Heartworm	Mr	Smith	Beethovenstr. 20	Berlin	01766262678

Relational Structure

Relational Structure