



# Data Preparation

Programming for Data Science

# Warm Up

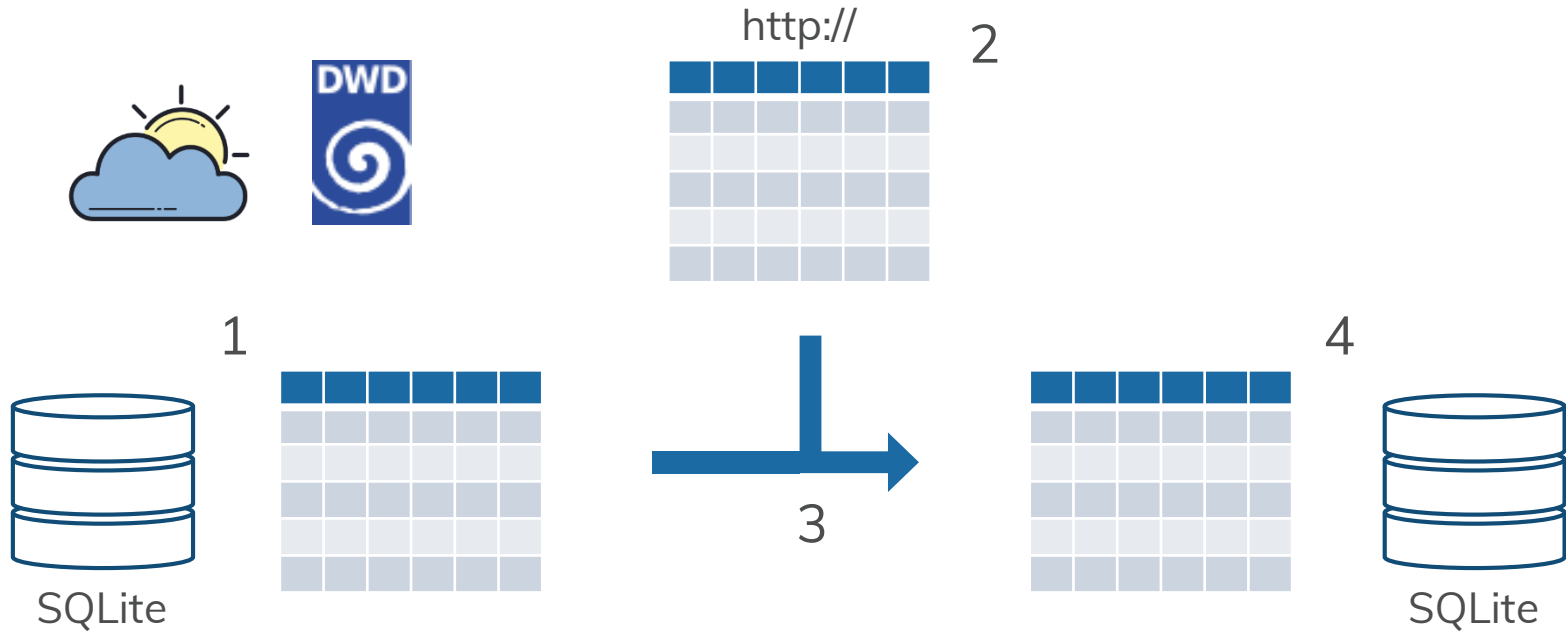
*Answer the following statements! Give reason for your answers.*

1. What kind of attribute heterogeneities are there?
2. How do you read data from a database?
3. How do you read the table/column names from a database?
4. What is a csv file?
5. How do you download files over http in code?
6. What are typical data types in databases?
7. What do you need to consider when writing data to a database?

<https://amcs.website>

# Task

*Expand the given SQL database with new content from an online source.*



# Task

## Step 1

- You will get a SQLite .db file from us. This already contains data.
- Connect to the db in your language/environment.
- Explore the data in it (schema).
- Load the data from the db into your environment.

## Step 2

- Find the weather data from the last timestamp in the database to 23.10.2020.
- <https://opendata.dwd.de/>
- Download the data for evaluation, but also implement the download in your code.

## Step 3

- Adapt the downloaded data to the db data representation (data types, columns)

## Step 4

- Append the new data to the db and export the complete table to json
- Generate the average temperature per hour in a new table

# Package suggestions

## R

- data.table
- RSQLite
- (DBI)
- jsonlite

## python3

- sqlite3
- pandas
- requests
- zipfile

# Exercise Appointment

## *We compare and discuss the results*

- Tuesday, 03.11.2020,
- Consultation: Please use the forum in Opal.
- Please prepare your solutions! Send us your code!

*If you have questions, please mail us:*

[claudio.hartmann@tu-dresden.de](mailto:claudio.hartmann@tu-dresden.de) Orga + Code + R

[lucas.woltmann@tu-dresden.de](mailto:lucas.woltmann@tu-dresden.de) Tasks + Python

