TECHNISCHE UNIVERSITÄT DRESDEN

Dresden Database Systems Group

Data Preparation

Programming for Data Science

Warm Up

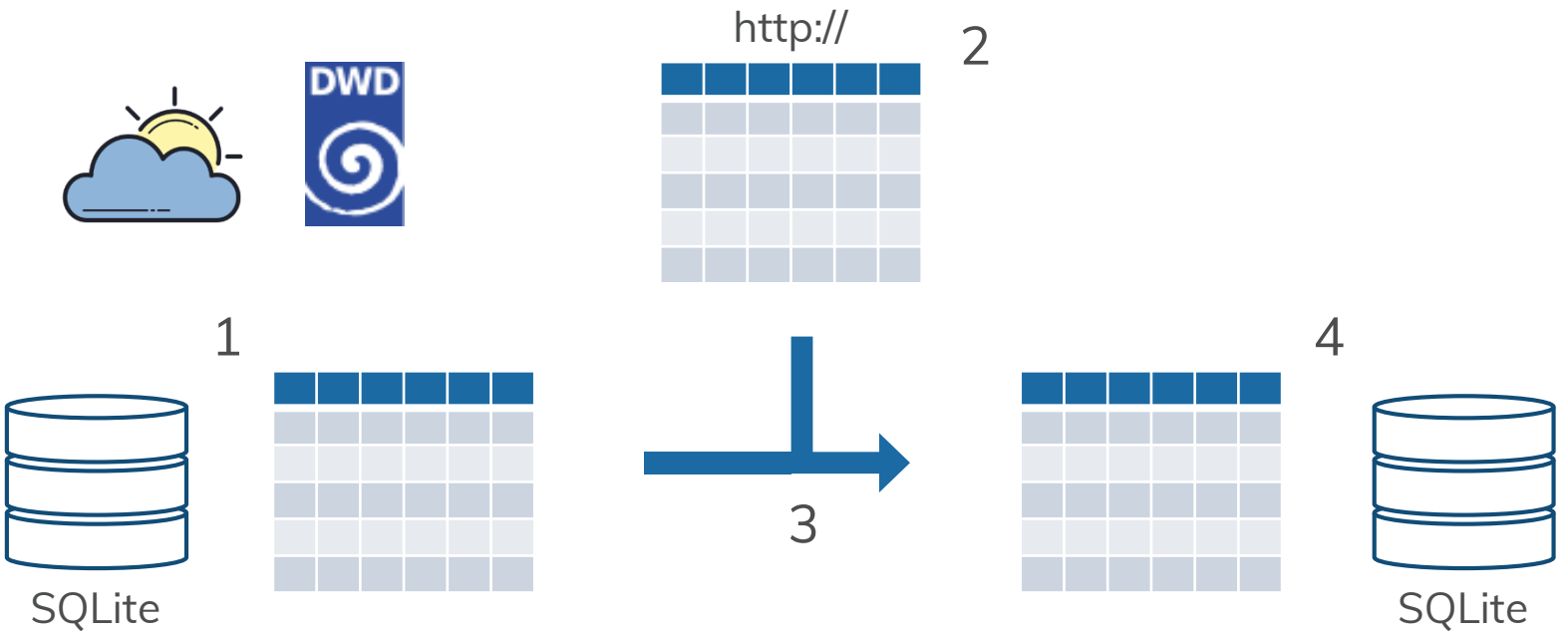
Dresden Database Systems Group

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?

Task

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



Task

Dresden Database Systems Group

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 temperature data from the last timestamp in the data to 15.10.2019 given the data in db
- https://opendata.dwd.de/
- Download the data for evaluation, but also implement a 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

Dresden Database Systems Group

R

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

python3

- sqlite3
- pandas
- requests
- zipfile

Exercise Appointment

Dresden Database Systems Group

We compare and discuss the results

- Tuesday, 22.10.2019,
- Consultation: Thursday, 17.10.2019, APB/E008,
- Please prepare your solutions! Send us your code!

If you have questions, please mail us:

claudio.hartmann@tu-dresden.de Orga + Code
lucas.woltmann@tu-dresden.de Tasks + Python
lars.kegel@tu-dresden.de Tasks + R