UV Distributed Information Management
Summer semester 2024

Supplementary Material – Database Systems Terminology

This document provides supplementary material regarding terms that are commonly used in the context of database systems.

Data integrity (Datenintegrität, in german) Data integrity [1] has different meanings depending on the domain. In the context of information security, data integrity refers to the *prevention of unauthorized modification of information.* In other words, data integrity defines the *correctness* or *validity* of the data upon modification (done by humans or machines). In our context, *consistency (or integrity) constraints* are used to describe conditions that must be satisfied for the data to be correct. By definition, the data are *consistent* (or *of integrity*) if all the constraints are satisfied.

Example: If an attribute A stores the balance of a bank account, we may want to disallow values that are smaller than EUR -5,000 (i.e., we allow a credit of at most EUR 5,000 per bank account). If a customer then tries to lend more than EUR 5,000 from our bank, this constraint is violated and the system may prevent it.

Key (Schlüssel, in german) In the relational model [2] (record-based tables), we need a way to distinguish the rows. Let K be a subset of the attributes (columns) of a relation (table). K is called *superkey* [3] if the attributes of K suffice to uniquely identify a tuple (row) in the relation. K is a *candidate key* [4] if K is a super key and K cannot be further reduced (i.e., no attributes can be removed) without losing the super key property. In other words, a candidate key is a *minimal super key.* Any candidate key qualifies as *primary key* [5] of a relation (hence the name: it is a candidate to be chosen as primary key). Typically, a primary key is a candidate key that is composed of attributes that are rarely subject to updates.

Example: The social security number, the bank account identifier, or the registration number at universities are prototypical examples of primary keys. However, also a combination of multiple attributes can serve as primary key, e.g., the combination of firstname, lastname, and birthyear.

Schema In our lecture, the schema refers to the overall design of a database, i.e., it defines the structure of the data (similar to a variable declaration in a programming language) and

[1] Data integrity: `https://en.wikipedia.org/wiki/Data_integrity`
[2] The relational model: `https://en.wikipedia.org/wiki/Relational_model`
[3] Superkey: `https://en.wikipedia.org/wiki/Superkey`
[4] Candidate key: `https://en.wikipedia.org/wiki/Candidate_key`
[5] Primary key: `https://en.wikipedia.org/wiki/Primary_key`

the relationships between the data. We distinguish between the schema of a relation and of a database. The *relation schema* [6] refers to the schema of a single relation (table; in the relational model), whereas the *database schema* [7] refers to the collection of all relation schemata in the database.

Caveat: The term *schema* may have a different meaning in some database systems, e.g., a schema may subsume multiple tables. Therefore, we recommend to read the manual of the database system at hand. Nonetheless, we use the term *schema* as described above.

Pointers to Additional Material In this section, we provide pointers (references) to other material/videos that mostly provide more in-depth knowledge on certain topics related to database systems:

Database Courses at the University of Salzburg: The Database Research Group at the University of Salzburg teaches many topics in the area of database systems in more depth. Feel free to check the respective websites for more information on the specific topics [8] [9] [10] [11] [12] [13].

Big Data Engineering on Youtube: Jens Dittrich, a database systems and data science professor at the Saarland University, streams his lecture on Big Data Engineering live on Youtube (old videos are also available) [14].

Youtube Channel of the CMU Database Group: The Carnegie Mellon University is one of the top universities in Computer Science. The Database Group has many videos in their Youtube channel [15] that cover most of the topics we covered in class in more detail [16] [17].

[6] Relation: https://en.wikipedia.org/wiki/Relation_(database)
[7] Database schema: https://en.wikipedia.org/wiki/Database_schema
[8] Databases 1: https://dbresearch.uni-salzburg.at/teaching/2024ss/db1/
[9] Databases 2: https://dbresearch.uni-salzburg.at/teaching/2023ws/db2/
[10] Databases Tuning: https://dbresearch.uni-salzburg.at/teaching/2024ss/dbt/
[11] Advanced Databases: https://dbresearch.uni-salzburg.at/teaching/2023ws/adb/
[12] Non-Standard Database Systems: https://dbresearch.uni-salzburg.at/teaching/2024ss/nsdb/
[13] Similarity Search: https://dbresearch.uni-salzburg.at/teaching/2023ws/ssdb/
[14] Jens Dittrich's Youtube channel: https://www.youtube.com/user/jensdit/videos
[15] CMU-DB Youtube channel: https://www.youtube.com/c/CMUDatabaseGroup/videos
[16] Intro database systems: https://www.youtube.com/playlist?list=PLSE8ODhjZXjbohkNBWQs_otTrBTrjyohi
[17] Adv. database systems: https://www.youtube.com/playlist?list=PLSE8ODhjZXjasmrEd2_Yi1deeE360zv5O