Dylan Phoutthavong
February 7th, 2025
CSCU 3287

HW3

Consider a STUDENT relation in a UNIVERSITY database with the following attributes (Name, Ssn, Local_phone, Address, Cell_phone, Age, Gpa). Note that the cell phone may be from a different city and state (or province) from the local phone. A possible tuple of the relation is shown below:

Name	Ssn	Local_phone	Address	Cell_phone	Age	Gpa
George Shaw William Edwards	123-45-6789	555-1234	123 Main St., Anytown, CA 94539	555-4321	19	3.75

1. Identify the critical missing information from the Local_phone and Cell_phone attributes. (Hint: How do you call someone who lives in a dif-ferent state or province?)
    a. Missing information:
        i. The area code is missing for both Local_phone and Cell_phone. Without the area code, the numbers are too vague to verify and cannot be dialed, especially if they belong to different state or province.
2. Would you store this additional information in the Local_phone and Cell_phone attributes or add new attributes to the schema for STUDENT?
    a. I would choose option 2: Add separate attributes for Local_phone_area_code and Cell_phone_area_code.
        i. This will provide better clarity and allow easier querying of the area code but increases schema complexity.
3. Consider the Name attribute. What are the advantages and disadvantages of splitting this field from one attribute into three attributes (first name, middle name, and last name)?
    a. Advantages:
        i. Easier querying: You can search or sort specifically by first, middle, or last name.
        ii. Flexibility: Useful for formatting names differently such as displaying "Last, First".
        iii. Localization: Supports different naming conventions globally such as family names that comes first.
    b. Disadvantages:
        i. Increased complexity: The schema becomes more complex with additional attributes.

ii. Data entry errors: Users might leave fields blank or make mistakes such as entering the middle name in the first name field.
iii. Storage overhead: Slightly more storage is required for the additional attributes.

4. What general guideline would you recommend for deciding when to store information in a single attribute and when to split the information?
    a. Guidelines:
        i. Split when:
            1. The components of the attribute have independent meaning or are frequently queried/filtered separately such as splitting Name into First_name and Last_name.
            2. The attribute supports different formats or localizations such as phone numbers with area codes.
        ii. Store as a single attribute when:
            1. The information is typically used as a whole and does not need to be split for querying or processing like Gpa.
5. Suppose the student can have between 0 and 5 phones. Suggest two different designs that allow this type of information.
    a. Design 1: Add multiple phone number attributes to the schema:
        i. Attributes: Phone1, Phone2, Phone3, Phone4, Phone5.
        ii. Pros:
            1. Simple to implement.
            2. No need for an additional table.
        iii. Cons:
            1. Schema is less flexible and wastes space if most students have fewer than 5 numbers.
    b. Design 2: Use a separate table for phone numbers:
        i. Create a Phone table with attributes: Ssn, Phone_number, and Type such as Local or Cell.
        ii. Each student's Ssn in the STUDENT table can have multiple rows in the Phone table.
        iii. Pros:
            1. More flexible and normalized design.
            2. Avoids wasting storage for students with fewer phone numbers.
        iv. Cons:
            1. Slightly more complex queries due to the additional table.