Relational Database

In this note, we learn about the data storage methods. In the modern world, advanced languages like Python, SQL, R operate on relational databases, therefore we need to understand this first.

1. *Introduction*:
   Let's start with the database. A database is a record of an organized collection of structured data that can be stored, accessed, and managed electronically. In simpler terms, a database is a system that allows you to store and retrieve large amounts of data easily and efficiently.

2. *Limitation of systems preceding databases:* Before databases came into existence, people used to store records in something called a file system. Following is an example of file system:

Name	Major	Enrollment
John Smith	Stat	STAT100A, MATH010, CS010A
Mary Jane	Math	MATH105, STAT130, CS135
Erin Key	CS	CS105, STAT100A, CS181

   There are several problems associated with this type of database. For example:
   What if we add/remove courses, or update course numbers? we would need to update in all relevant places. This may end up deleting some information that we may not want to drop.

   What if we want to answer the query "How many students are registered for STAT 100A?" The programmer has to write specific code to form the query. Writing specific code every time for these types of queries is not so efficient.

   What if we want to enforce the constraint that each student is registered for at least 1 course? It's hard to enforce as students add/drop, and courses are canceled and added continuously.

   These limitations can be summarized in the followings points:
   - Difficulty in accessing data,
   - Data redundancy and inconsistency,
   - Data Isolation, Integrity problems,
   - Atomicity of updates,
   - Concurrent access by multiple users ,
   - Security problems

3. *What are the properties of a good data recording system?*

   Before jumping to the databases, it will be helpful for us to discuss the requirements or properties we want to look for in a good data system. The following are the desired properties:

   - Atomicity: system should ensure that updates of a partially executed transaction are not reflected in the database.
   - Consistency: system should ensure that any changes to values in an instance are consistent with changes to other values in the same instance.
   - Durability: system should ensure updates of committed transactions is critical.
   - Isolation: system should ensure that transactions that occur in parallel will have the same effect if they were run sequentially.

   These properties are called ACID properties of a system.

4. *Finally our Relational Database:*

   Our database system satisfied the ACID properties and is a constantly flourishing industry. We might wonder what this 'relational' word means here: in a relational database, "relational" refers to the way data is organized and related within the database.

   A relational database stores data in one or more tables, which consist of rows and columns. Each table represents a specific entity or concept, and each row represents an instance or record of that entity. For example, a database for a university might have tables for students, courses, instructors, and enrollments. The relationship between tables is defined by common columns, which are used to establish links or connections between the tables. These links are typically based on the values of one or more columns that are common to both tables. For example, a "students" table might have a "student_id" column that links to the "enrollments" table, which has corresponding "student_id" and "course_id" columns.

5. *The terminologies of Relational Database:*

   Since we are familiar with R dataframes, a map of jargon will help. The following table describes what we know in R is called what in Relational databases:

R's dataframes/tibbles are actually tables

R jargon	Database jargon
column	field/attribute
row	record/tuple
dataframe	table/relation
types of the columns	table **schema**
bunch of related dataframes	database

So a dataframe is called a 'table' in a database system and a bunch of such tables are called databases. Columns are called fields or attributes