1 Data Modeling

1.1 Date and Text Representations

- *Date:* Epoch times since 1970 in seconds. Dates are complex so it's best practice to use language libraries.
- *ASCII:* 8 bits (or 1 byte) per a character. Can represent 127 characters. [1]
- *UTF-8:* The most common text encoding. Can represent up to 1 million characters. Has variable-length encoding; characters are represented by one, two, three, or four bytes. Backward-compatible with ASCII.

1.2 Data Modeling

What data types are appropriate for the following situations?

- Quantity of goods stocked in a warehouse
- A Northwestern student ID number
- A student's name
- Energy readings produced by a particle accelerator
- Temperature in Celsius, recorded to the tenth of degree
- A transaction in dollars and cents in a banking application
- A ten-digit phone number
- A plain-text password

Fundamental data types for integers and floating-point constants [2]

Data Type	32-bit size	64-bit size
char	1 byte	1 byte
short	2 bytes	2 bytes
int	4 bytes	4 bytes
float	4 bytes	4 bytes
long	4 bytes	4/8 bytes
double	8 bytes	8 bytes
long long	8 bytes	8 bytes

[1] ASCII Uses 7 bits per character, but in practice each character is stored in 8 bits and the top bit is zero. Hence $2^7 - 1 = 127$

[2] Floating-point data types are `float, double, long double.` Integer data types are `char, short, int, long, long long.` Integer types may be prefixed with the `signed` or `unsigned` qualifier. If no sign qualifier is present, the type is assumed to be signed.

2 Relational Databases

2.1 Relational Databases

Database schemas define:

- Tables
- Columns (column names and data types)
- Primary keys
- Foreign keys

Normalization is the process of breaking one redudant table into multiple tables.

2.2 Primary Keys and Foreign Keys

A table's **primary key** uniquely identifies each row. No two rows in a table can have the same primary key. A table's **foreign key(s)** refer to keys in other tables.

2.3 Data Normalization Example

An office supply store is digitizing its customer database. All of their customer data is currently stored on paper order forms.

Order Form

Order number: 1234 Date: 11/04/98
Customer number: 9876
Customer name: Billy
Customer address: 456 HighTower Street
City-Country: Hong Kong, China

ProductNo	Desscription	Quantity	Unit Price
A123	Pencil	100	$3.00
B234	Eraser	200	$1.50
C345	Sharpener	5	$8.00

- What is the widest schema possible that can be used to represent an order?
- What are some problems with using such a wide schema?
- Normalize the database

3 Data Modeling

You are designing a database to manage patient data for a hospital.

- A patient has a name, an age, and a phone number.
- Patients also have a diagnosis, which is a disease.
- Patients can be prescribed to taking drugs with a dosage to treat their illness.
- Each patient has physicians that attend to them. Each physician has a name and a specialty.
- Each physician has a head of department that they need to report to, and a list of residents and nurses that report to them. Residents and nurses also have patients.
- Combinations of drugs can produce side-effects in patients, which are also diseases

Design an ER diagram for this database and be explicit about the relationships in the diagram. For simplilcity, we can use `Employee` to represent physicians, nurses and department heads. See the answer on the next page when finished.