8. Database

8.1 Database Concepts

- **Relational Database**: a database where **internal pointers** link the data items
- **Limitations of file-based approaches**
    1. **Data redundancy**: data is repeated in more than one file
    2. **Data dependency**: changes to data means changes to programs accessing that data
    3. **Lack of data integrity**: entries that should be the same can be different in different places
    4. **Lack of data privacy**: all users have access to all data if a single flat file
- **Advantages of relational database**
    - **Reduced data redundancy** → because each item of data is only stored once and is referenced by a primary key, at different tables by primary and foreign keys, going through the normalization processes
    - **Improved data integrity**
        - changes in one table will automatically update in another
        - linked data cannot be entered differently in two tables
    - **Program-data independence** → changes to data do not require programs to be re-written
    - **Provision of different views** → so users can only see specific aspects of the database

Keywords

- **Entity**: anything that can have data stored about it, e.g. a person
- **Table**: a group of similar data in a database, with rows for each instance of an entity and columns for each attribute
- **Record**: a row in a table in a database
- **Field**: a column in a table in a database
- **Attribute**: an individual data item stored for an entity, e.g., name for a person
- **Candidate Key**: an attribute or smallest set of attributes in a table where no tuple has the same value
- **Primary Key**: a unique identifier for a table; a special case for the candidate key
- **Secondary Key**: a candidate key that is an alternative to the primary key
- **Foreign Key**: a set of attributes in one table that refer to the primary key in another table
- **Relationship**: situation in which one table has a foreign key that refers to a primary key in another table
- **Referential Integrity**: property of a database that does not contain any values of a foreign key that are not matched to the corresponding primary key
- **Indexing**: a data structure built from one or more columns in a database table to speed up searching for data

Database Normalization

1. **First Normal Form (1NF)**
    - All fields are **atomic**
    - **No repeating** groups
    - Each field should contain **a single value** for each row (no multiple values for a cell)
2. **Second Normal Form (2NF)**
    - It is already in 1NF
    - **No partial dependencies**: all non-key attributes are **fully functionally dependent** on the primary key

3. **Third Normal Form (3NF)**
   - It is already in 2NF
   - **No transitive dependencies**: all non-key attributes are not dependent on any other

8.2 Database Management Systems (DBMS)

- software systems used to store, retrieve, and run queries on data

Features of DBMS

- **Data Management**: DBMS maintains a **Data dictionary** to store the <mark>metadata about the data</mark>, such as its fields, data types, validations, and keys
- **Data Modelling**: a **relational model** which decides what can be present in the schema; It provides a database user with a conceptual framework in which we specify the database requirements of the database user and the structure of the database to fulfill these requirements
- **Logical Schema**: a blueprint of the database which specifies what fields will be present and their type constraints, using representations such as an ER diagram
- **Data Integrity**: DBMS enforces data integrity through constraints (e.g. primary keys and foreign keys) and transactions that maintain consistency
- **Data Security**
  - **Access rights** → unauthorized users cannot access the database
  - **Views** → restricts which parts of the database specific users can see
  - **Password** → prevents unauthorized access
  - **Automatic backup** → create regular copies of data in case of loss
  - **Encryption** → data is incomprehensible to unauthorized users

DBMS helps Data Redundancy

- Each record/piece of data stored once and referenced using primary key
- Since data stored in individual tables and tables linked using relationships
- By proper use of primary and foreign keys
- By enforcing referential integrity
- By going through normalization process

Tools of DBMS

1. **Developer Interface**
   - To create user-friendly features, e.g., forms to enter new bookings
   - To create outputs, e.g., report of bookings on a given data
   - To create interactive features, e.g., buttons and menus
2. **Query Processor** → takes a query written in SQL and processes it
   - To create SQL queries
   - To search for data that meets **set criteria**
   - To perform calculations on extracted data, e.g., summation

Referential Integrity

- Make sure tables don't try to reference data that don't exist
- Primary key cannot be deleted unless all dependent records have already deleted
- Cascading delete and update

- Primary key cannot be updated unless all dependent records are already updated
  Reason For Secondary Key

8.3 Data Definition Language (DLL) and Data Manipulation Language (DML)

- **Data Definition Language (DLL)**: carries out all creation/modification of the database structure
- **Data Manipulation Language (DML)**: carries out all queries and maintenance of data
- The industry standard for both DDL and DML is **Structured Query Language (SQL)**