Lecture 1&2

A database system is a collection of interrelated data and a set of programs that allow users to access and modify these data.
A major purpose of a database system is to provide users with an abstract view of the data

Data models A collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints.

Data abstraction: Hide the complexity of data structures to represent data in the database from users through several levels of data abstraction

DDL (Data Definition Language) Specification notation for defining the database schema, compiler generates a set of table templates stored in a data dictionary

Data dictionary contains metadata (i.e., data about data)
• Database schema
• Integrity constraints (Primary key (ID uniquely identifies instructors))
• Authorization

Data Manipulation Language(DML) Language for accessing and updating the data organized by the appropriate data model, also known as query language

There are basically two types of data-manipulation language
• Procedural DML -- require a user to specify what data are needed and how to get those data.
• Declarative DML(Non procedural) -- require a user to specify what data are needed without specifying how to get those data, (easier).

SQL query language is nonprocedural. A query takes as input several tables (possibly only one) and always returns a single table, Application program interface (e.g., ODBC/JDBC) which allow SQL queries to be sent to a database

Non-procedural query languages such as SQL are not as powerful as a universal Turing machine.
♣ SQL does not support actions such as input from users, output to displays, or communication over the network.
♣ Such computations and actions must be written in a **host language**, such as C/C++, Java or Python, with embedded SQL queries that access the data in the database.

♣ **Application programs** -- are programs that are used to interact with the database in this fashion.

Database design requires that we find a "good" collection of relation schemas.

Logical Design – Deciding on the database schema

Physical Design – Deciding on the physical layout of the database

A database system is divided into modules that deal with each of the responsibilities of the overall system.
The functional components of a database system can be divided into
• The storage manager, • The query processor component, • The transaction management component.

The storage manager implements several data structures as part of the physical system implementation:
• **Data files** -- store the database itself
• **Data dictionary** -- stores metadata about the structure of the database, in particular the schema of the database.
• **Indices** -- can provide fast access to data items. A database index provides pointers to those data items that hold a particular value.

DDL interpreter -- interprets DDL statements and records the definitions in the data dictionary.
• **DML compiler** -- translates DML statements in a query language into an evaluation plan consisting of low-level instructions that the query evaluation engine understands
Query evaluation engine -- executes low-level instructions generated by the DML compiler.

A transaction is a collection of operations that performs a single logical function in a database application
♣ **Transaction-management component** ensures that the database remains in a consistent (correct) state despite system failures (e.g., power failures and operating system crashes) and transaction failures.
♣ **Concurrency-control manager** controls the interaction among the concurrent transactions, to ensure the consistency of the database.

Two-tier architecture -- the application resides at the client machine, where it refer to database system functionality at the server machine

Three-tier architecture -- the client machine acts as a front end and does not contain any direct database calls

Database administrator (DBA):A person who has central control over the system

The set of allowed values for each attribute is called **the domain** of the attribute

♣ Attribute values are (normally) required to be **atomic**; that is, indivisible

♣ The special value null is a member of every domain. Indicated that the value is "unknown", The null value causes complications in the definition of many operations

Database schema -- is the logical structure of the database.
Database instance -- is a snapshot of the data in the database at a given instant in time.

Relational Algebra: A procedural language consisting of a set of operations that take one or two relations as input and produce a new relation as their result.
Six basic operators: select(σ), project($\prod$), union($\cup$), set difference($-$) , Cartesian product(x) , rename(ρ).