A REVIEW OF BASIC KNOWLEDGE OF DATABASE SYSTEM

1. Relational model

I still remember it is my first time to read such an academic article in detail. I have to say that it’s really a hard time for me at the beginning. Edgar F. Codd originated the relational approach to database management in a series of research papers published commencing in 1970. After read the seminal paper "A Relational Model of Data for Large Shared Data Banks" which Prof. koo recommends us to read, I got some basic knowledge about relational model. The relational model for database management is a database model based on first-order predicate logic.

The basic thing of relational model is table. The fundamental assumption of the relational model is that all data is represented as mathematical n-ary relations, an n-ary relation being a subset of the Cartesian product of n domains.

Data are operated upon by means of a relational calculus or relational algebra, these being equivalent in expressive power. The relational model of data permits the database designer to create a consistent, logical representation of information. Consistency is achieved by including declared constraints in the database design, which is usually referred to as the logical schema. The theory includes a process of
database normalization whereby a design with certain desirable properties can be selected from a set of logically equivalent alternatives.

2. Entity-relationship model

Keep moving on the adventure of database, then we meet another interesting model--- Entity-Relationship Model. Although relational model deals with data at data structure level, it’s not as simple and clear as E-R model. But many operations can be directly carried on this model because it has the mathematical foundation of relational algebra and relational calculus. However, E-R model deals with data at the level of information concerning entities and relationships. So it’s more general and natural. It doesn’t depend on any kind of computer realization.

The building blocks of E-R model are entities, relationships, and attributes. Entity-relationship modeling is a database modeling method, used to produce a type of conceptual schema or semantic data model of a system, often a relational database, and its requirements in a top-down fashion.

3. Database design and normalization

Database design is the process of producing a detailed data model of a database. This logical data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a Data Definition
Language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

Normalization consists of normal forms that are 1NF, 2NF, 3NF, BOYCE-CODD NF (3.5NF), 4NF and 5NF. There is a tradeoff between space and efficiency.

4. Application design

While in some related Database material, design process can be divided into the following 6 steps: Requirement analysis, Conceptual database design, Logical database design, Schema refinement, Physical database design, Application and security design.

In the design process, the requirement is most important because it determines what we should do next. We need to understand what the data to be stored is and what application must be built on top of it. Then we should analyze our data source and make it feasible for our application. Conceptual database design and Logical database design are very important next. What we have to do is to refine the schema by examine the ICs much more closely. After this, we need get a balance between space and efficiency, so how to normalization is very important. Last but not least, operation and maintenance is what we should not forget to do.
5. Data Serialization

Serialization is the process of converting a data structure or object into a sequence of bits so that it can be stored in file or memory file. XML is one of important serialization formats. It is used to produce a human readable text-based encoding. Such an encoding can be useful for persistent objects that may be read and understood by humans, or communicated to other systems regardless of programming language. It has the disadvantage of losing the more compact, byte-stream-based encoding, but by this point larger storage and transmission capacities made file size less of a concern than in the early days of computing.

6. Transaction

Transaction is a collection of operations that forms a logic unit of work. The properties of database transactions are summed up with the acronym ACID:

1. Atomicity - all or nothing

   All of the tasks (usually SQL requests) of a database transaction must be completed; If incomplete due to any possible reasons, the database transaction must be aborted.

1. Consistency - serializability and integrity
The database must be in a consistent or legal state before and after the database transaction. It means that a database transaction must not break the database integrity constraints.

1 Isolation

Data used during the execution of a database transaction must not be used by another database transaction until the execution is completed. Therefore, the partial results of an incomplete transaction must not be usable for other transactions until the transaction is successfully committed. It also means that the execution of a transaction is not affected by the database operations of other concurrent transactions.

1 Durability

All the database modifications of a transaction will be made permanent even if a system failure occurs after the transaction has been completed.

The purpose of obeying these rules is to provide reliable units of work that allow correct recovery from failures and keep a database consistent even in case of system failure, and provide atomic and isolating process so that our data and database is correct when the operation accidently erupts.
7. System Architecture

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system.

An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structure of the system which comprises system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them, and provides a plan from which products can be procured, and systems developed, that will work together to implement the overall system.