Summary Database Knowledge

1. Database is a collection of data for retrieval, insertion, deletion, revision.
2. DBMS stands for database management system. It is a software to manage database such as MySQL, Oracle, SQL Server… We do not retrieval data from database directly. We retrieval data via operation of DBMS.
3. SQL is language for view and changing the data in database. Notice that SQL might has slight difference between different DBMS.
4. File System manage and organize the files in storage medium while DBMS manage the database; File System has redundant data while DBMS has no redundant data. File System is not efficient on query processing while DBMS is efficient on query processing. At last, File System is less security and less expensive than DBMS.
5. There are 4 steps to build a database. First, we need to have ideas about what kind of database should we need. Second is hign level design. Third is schema, this part decides what kind of relationship between each table, what kind of data should us use. The last step is relational DBMS.
6. ER Model is entity relationship diagram, the picture we draw before creating the database.
7. Components of an ER Diagram are entity, attribute and relationship. For example, a student can choose many courses, student and courses are entities, student information is attribute and the one to many is relationship.
8. The entity in ER Model is rectangle.

9. Attribute is the property of an entity. The unique attribute such as ID is key attribute. Combination of other attributes like address is composite attribute. One person might have multiple phone no., the phone no. attribute is multivalued attribute. The age attribute is dynamic and derived from DPB also can be changed as time. That is derived attribute.
10. Unary relationship means only one entity participate. Like the friendship between people. Binary relationship is two entities participate. N-ary…
11. One to one relationship, one to many, many to many…
12. Participation Constraint has Total Participation and Partial participation. The first one requires each entity must participate. We use double line in ER Model. Like a passport must belongs to someone. The latter one does not require participation of entities. We use single line in ER Model. Like a person might or might not have a passport.
13. Domain Constraints means the content of each attribute must be atomic values, only one valid value or none. Key Constraints require primary key must be uniquely. Referential Integrity Constraints require the foreign key must refer to a exist item in another table.
14. Case1: person and passport. We can optimize the three tables (Person Passport JoinTable) to one PersonAndPassport table by allowing the passport attribute allows null value.
15. Case2: male and female. We can optimize the three tables(Male, Female, JoinTable) to Female and Male with Female as foreign key.
16. Case3: one student and many courses. We can optimize the three tables (Student, Course, JoinTable) to Course and Student with course as Foreign key.

17. Case4: Many students to many courses relationships. The three tables (Students, Course, JoinTable) cannot be optimized because many to many relation table need a join table.
18. Non-relational database does not use tabular schema. The term NoSQL refers to data stores that do not use SQL for queries. There are Document data stores which store json data type, Columnar data stores store items, Key-value pair data stores store hash code, graph data stores store graphs…
19. MangoDB is a popular Non-Relational Database. It belongs to document store. It is easy to setup and has lots of professional supports but does not allow joins which is a normal problem to Non-relational database.
20. Redis is another popular Non-relational database belongs to key value pair stores. It is also easy to install and supports lots of data types but does not support join and requires language of "LUA".
21. Summary difference between relational and non-relation database: SQL is relational database, it needs predefined schema which means it needs design the relationship before creating while NoSQL is non-relational database, its schema is dynamic. SQL is vertically scale which means if we need to enhance the database, we can enhance the computer while NoSQL is horizontally scale which means if we need to enhance the database, we can add more computers.
22. ACID (atomicity, consistency, isolation, durability) is a set of properties of database transactions intended to guarantee data validity despite errors, power failures, and other mishaps. In the context of databases, a sequence of database operations that satisfies the ACID properties (which can be perceived as a single logical operation on the data) is called a transaction. For example, a transfer of funds from one bank account to another,

even involving multiple changes such as debiting one account and crediting another, is a single transaction.

The theorem is a set of basic requirements that describe any distributed system. If you imagine a distributed database system with multiple servers, here's how the CAP theorem applies:

Consistency - All the servers in the system will have the same data so users will get the same copy regardless of which server answers their request.

Availability - The system will always respond to a request (even if it's not the latest data or consistent across the system or just a message saying the system isn't working).

Partition Tolerance - The system continues to operate as a whole even if individual servers fail or can't be reached.

It's theoretically impossible to have all 3 requirements met, so a combination of 2 must be chosen and this is usually the deciding factor in what technology is used.