**Home Work – 05**

1. Please start by answering the following questions:

1. What is a database?

Ans: Database is a collection of data which can be also be manipulated as per the needs. It is organized in such a way that it can easily accessed, managed and updated.

E.g: An online telephone directory maintains the phone numbers of all the people.

2. What is a DBMS?

Ans: A Database Management System (DBMS) is a software that helps in creating and managing databases. Using DBMS, an end user can easily insert, delete and update the data in the database.

3. What are the problems of file systems?

Ans: Problems of file system:

* Data redundancy – The same information can be duplicated in multiple files.
* Data inconsistency – As there is a lot of redundant data, changes in one file can lead to inconsistency.
* Difficulty in accessing data – It is very difficult to access a data in file system.
* Concurrency issues – In file system, multiple users cannot access the same file.
* Data Sharing issues – Data is scattered in the different location and so it is difficult to share the data among different application.
* Integrity problems – Integrity means the data in the database is correct and consistent. But due to the data redundancy and inconsistency, there are integrity problems.
* Security problems – Data security is not maintained properly in file system.

4. What are the pros and cons of hierarchical databases?

Ans: Hierarchical Database – It is way of organizing the database in the form of multiple one-to-many relationship.

Pros of the hierarchical database:

* Addition and deletion of the new information is easy.
* The information which is on the top of the tree can retrieved easily.
* It works very well with anything that relates to one to many relationships.

Cons of hierarchical databases:

* The repeated data is stored in many different entities.
* It works very slow when the data to be searched is at the bottom of the tree.
* Searching of any data requires the DBMS to run from top to bottom making the query very slow.

5. Give an example of a network database.

Ans: A network database is a type of the database model wherein the child file or record is linked to the one or multiple file/records and vice-versa.

E.g. The Student and the Degree class has the Subject as their child class.

Student

Degree

Subject

6. What are the pros and cons of relational databases?

Ans: Pros of relational databases:

* Easy to use and understand as the data is stored in tables.
* It avoids data duplication.
* Data from the multiple tables can be retrieved using the join query.
* It avoids data inconsistency.
* The data can be made accessible to the users who needs it.

Cons of relational databases:

* Limit the length of the data fields. If the data to be stored is more than the field which can accommodate it. In that case, the data is lost.
* It is limited while accessing complex data.
* It does not provide support for the complex base types like drawing, etc.
* Knowledge of the database structure is required before creating the Ad hoc queries.

7. What are the major types of NoSQL databases?

Ans: There are 4 basic types of NoSQL databases:

1. Key-Value Store: In this type, the data is stored in the form of big hash tables which contains key and values.
2. Document-based store: It pairs each key with a complex data structure known as document. Each document contains multiple key-value pairs or key-array pairs or nested documents.
3. Graph stores: In uses edges and nodes to represent and store data.
4. Column based database – It stores columns of data together instead of rows.

8. What are the pros and cons of NoSQL databases?

Ans: Pros of NoSQL databases:

* It can handle huge data.
* As it is schema less and table free, it offers high level of flexibility with the data models.
* NoSQL is open source and so it is very affordable to the small enterprises.
* To upscale the database, you don’t need to increase the hardware instead you need to add more servers into the pool.
* Detailed modeling of the data is not required. So, it saves much time and efforts.

Cons of NoSQL Database:

* NoSQL only maintains consistency but does not maintain the ACID properties.
* It has less community support.
* It lacks standardization which causes issues during migration.

**Now consider a database Student-Course with the following three tables:**

S (SNO, SNAME, AGE, SEX, SDEPT)

SC(SNO, CNO , GRADE)

C (CNO, CNAME, CDEPT, TNAME)

Table S shows the

id number (SNO)

name (SNAME)

age (AGE)

gender (SEX) and

the department (SDEPT) of the students.

Table SC shows the

id number of the students (SNO) and

the courses they enrolled (CNO) as well as

the grade they get in the courses (GRADE).

Table C shows the

id number (CNO)

name (CNAME)

department (CDEPT), and

the name of the teacher (TNAME) of the courses.

B. Using this database, please write the following eight relational algebra queries:

1. the id number and the name of the courses that John Smith teaches.

Ans: πCNO, CNAME(σTNAME = ‘John Smith’(C))

1. the id number and the name of the male students who are 23 years old or more.

Ans: πSNO, SNAME(σAGE >= 23 AND SEX = ‘Male’(S))

1. the names of the courses and their teachers that student S3 enrolled. Note that S3 is the id number of the student.

Ans: πC.CNAME, C.TNAME(σS.SNO = S3(SC ⨝CNO = CNO C))

1. the names of the female students who at least enrolled one course taught by John Smith.

Ans: πS.SNAME(σS.SEX = ‘FEMALE’ AND C.TNAME = ‘John Smith’(S ⨝S.SNO = SC.SNO SC) ⨝CNO = CNO C )

1. the id number of the courses that Linda Wild did not enroll.

Ans: πSC.CNO(σS.SNAME ≠ ‘Linda Will’(SC ⨝SNO = SNO S))

1. the id number of the students who have at least enrolled two courses.

Ans: πSNO(σCOUNT(CNO) > 2(SC))

1. the id number and the name of the courses enrolled by all of the students in our database.

Ans: πC.CNO, C.CNAME(σCOUNT(SC.SNO) ƳC.CNO (SC ⨝ C))

1. the id number of the students who have enrolled the course taught by John Smith.

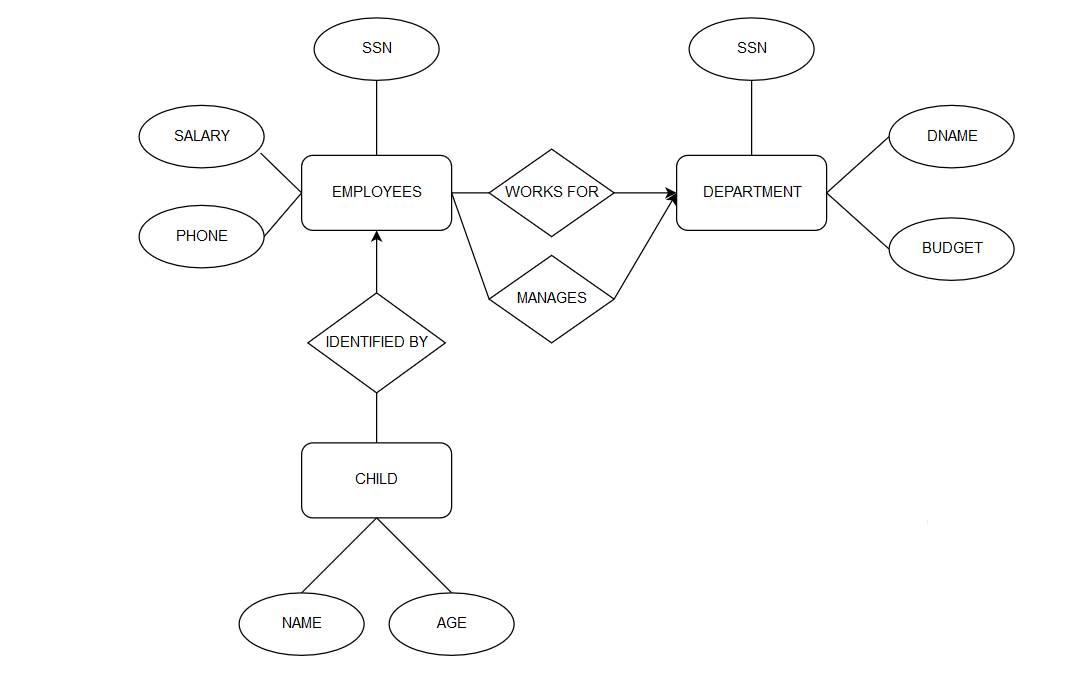
Ans: πSC.SNO(σC.TNAME = ‘John Smith’(SC ⨝SC.CNO = C.CNO C))

Hint: consider this and this entry points in the material we posted earlier this semester.

**C. For the next problem you need to do some ER modeling:**

A company database needs to store information about employees (identified by ssn, with salary and phone as attributes), departments (identified by dno, with dname and budget as attributes), and children of employees (with name and age as attributes). Employees work in departments; each department is managed by an employee; a child must be identified uniquely by name when the parent (who is an employee; assume that only one parent works for the company) is known. We are not interested in information about a child once the parent leaves the company. Draw an ER diagram that captures this information then upload it to your github repository.

**Solution:**

****

D. Normalization

Below is an invoice, and the 2NF of this invoice is provided.

Please provide 1NF and 3NF of this invoice.

HILLTOP ANIMAL HOSPITAL DATE: JAN 13/2017

INVOICE # 987

RICHARD COOK

123 THIS STREET

MY CITY, ONTARIO

Z5Z 6G6

PET PROCEDURE AMOUNT

-------------------------------------------------------------------

ROVER RABIES VACCINATION 30.00

MORRIS RABIES VACCINATION 24.00

-------------------------------------------------------------------

TOTAL 54.00

TAX (8%) 4.32

---------------------------------------------------

AMOUNT OWING 58.32

1NF:

Invoice [ invoice\_no, invoice\_date, cust\_name, cust\_addr]

invoice\_pet [ invoice\_no, pet\_id, procedure, amount, pet\_name]

2NF:

Invoice [ invoice\_no, invoice\_date, cust\_name, cust\_addr]

invoice\_pet [ invoice\_no, pet\_id, procedure, amount]

pet [ pet\_id, pet\_name]

Note: The underlined fields are the keys of the table.

3NF:

Invoice [ invoice\_no, invoice\_date, cust\_no]

invoice\_pet [ invoice\_no, pet\_id, procedure\_no]

pet [ pet\_id, pet\_name ]

customer [cust\_no, cust\_name, cust\_addr]

procedure [procedure\_no, procedure, amount]