8/28/2021

Sasha Stepanov [als0885]

BCDE103

Assessment 1

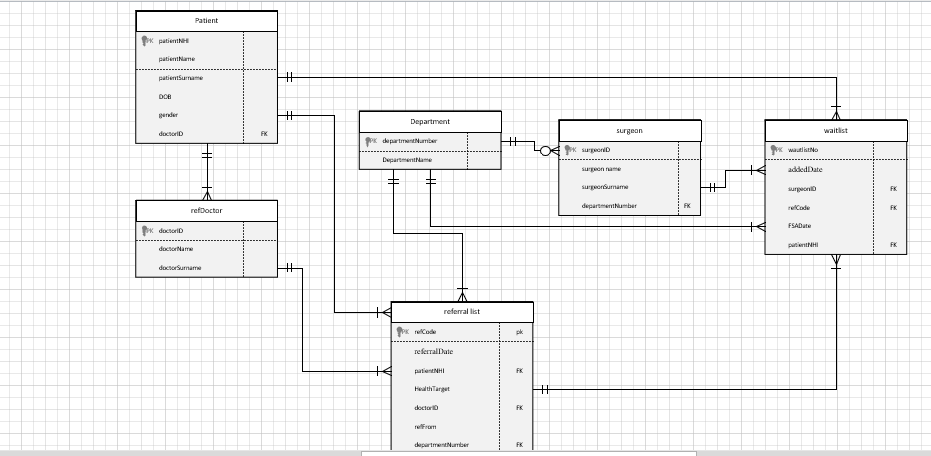
Database design

# DESIGN PROCESS

One of the first steps to accomplish any task is planning. ERD diagram was created before any work begun. After analyzing a task 6 tables (which is a mistake) where considered which are :

* Department
* Doctor
* Surgeon
* Patient
* Referral
* Waiting List

## First ERD



After the database was created and all information about the hospital information was loaded into SQL, it became obvious that there is no need for a waiting list table and it can be merged to the referral. This step was very important to make a query easier and data less bulky. Diagram

Description automatically generated

## Cardinality

* Department can have zero or many surgeons, surgeon can be attached to only one department
* Department can have zero to many Referrals, referral must have only one department.
* Doctor (referral doctor) can have zero to many patients, but patient should have a one doctor(probably)
* Doctor can have zero or many different referrals, but every referral should come from only one doctor
* Surgeon can have zero to many referrals, but referral should have only one surgeon
* Patient can have zero to many referrals, and referral should be only for one patient

# Entity, attributes, keys.

* Primary and Foreign keys were created to make associations or interactions between entities in database. For Patient Primary key was given as a NHI number which is unique. This number was chosen as a foreign key for Referral.
* Referral table consist of all 4 foreign keys, it seemed appropriate to make relationship through one table. For the rest of tables such as Department, Doctor (referral doctor), surgeon special unique ID were created for use as a Primary Key
* Almost for all attributes constraints a not a null, apart from FSA. I assumed that this could be left blanked, as FSA was not arranged confirmed yet.
* Seemed unnecessary to create any composite or multi-valued primary keys
* Composite attributes are names for Patient, Surgeon, Doctor. Were split into simple attributes. For example. Full name of a patient was divided into patient name and patient surname.

## Difficulties.

* Difficulties arose along the way of the entire project. For example, creating a bridge table, I was trying to place it somewhere as a very necessary attribute but turns out task can be accomplished without it. It was also difficult to determine or decide with cardinalities- what kind of relationship between the tables. For example, I'm still not sure about the relationship between the patient and the referral. Can a patient have zero referral, because if he does not have a referral, is he still a patient? With a surgeon, a department, a doctor, I did not have this problem, since even without a referral the surgeon will remain a surgeon.
* What could be altered is Referral table. I am not sure if it is good practice to have so many columns in one table. Need more experience in DATABASE to find out.

# Normalization

To organize original data into structured data normalization was done.

* For 1NF primary keys are NHI and Referral date. To determine a whole row, we need to know NHI and a referral date, because patient can come back, but will not have more than one visit per day.
* Partial dependencies are:

1. between NHI and a patient name, because if patient NHI is known it means his name can be found and identified.
2. Between NHI +Referral Date and Surgeon ID. Because NHI+Referral Date are known then Surgeon ID can be found.
3. Between NHI + Referral Date and Doctor ID. Because NHI+Referral Date are known then Doctor ID can be found.

* Transitive dependencies are:

1. Doctor ID and Doctor Type.
2. Surgeon ID and Surgeon Name.
3. Department ID and Department Name.

# DATA CLEANSING AND PREPARATION

A lot of work has been done to clean up the data. Below are all the steps to achieve the desired result

Step1: Remove all tracing and leading spaces (TRIM function).

Step2: Remove duplicates where it is necessary. For example, we cannot remove duplicates on dates, but it is necessary to remove duplicates for NHI, because it is supposed to be unique. (Format Cells)

Step3: all dates must be in a specific format that SQL accepts which is yyyy-mm-dd (Can be accomplished through some basic operation in Excel, such as “Data-Text – to Column” and “Formatting cells”.

Step4: Create a unique key (ID) for a Department, Doctor, Surgeon, Referral.

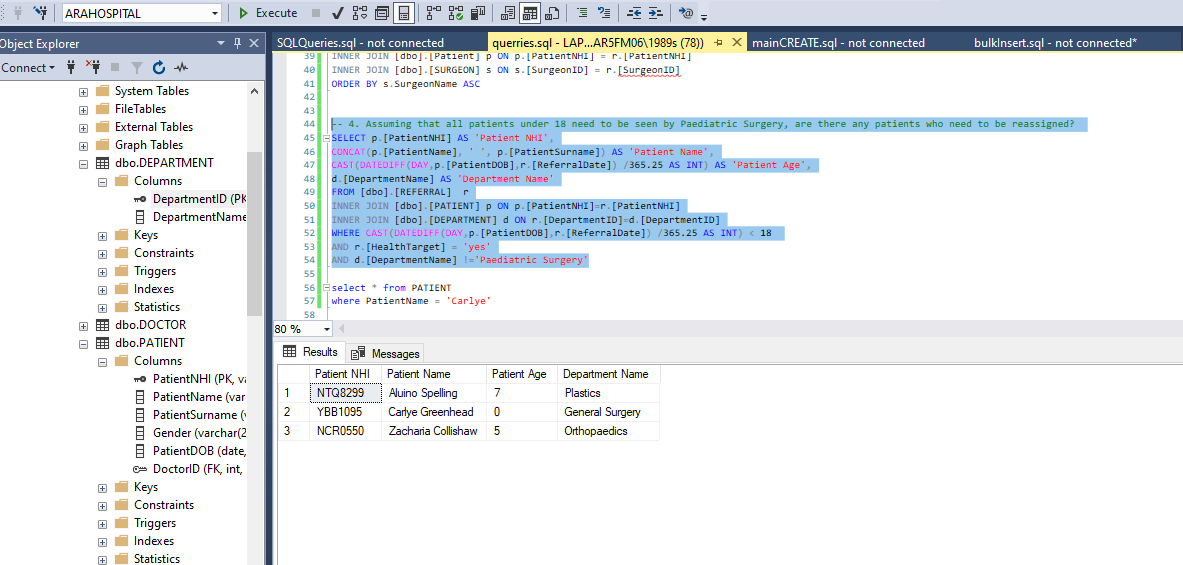
Step5: Remove all duplicates or repeating departments, surgeons, doctors. There are only 5 departments, 15 surgeons and 10 doctors with a unique ID. There is no point to store unnecessary data. (CTR + H > Replace All)

Step:6 One of the final steps to make spreadsheet workable is – all tables should be saved as appropriate file (CSV) for a SQL database.

In conclusion of the above, I would like to note that several operations have been done to achieve “clean data”.

* Column year-month seems unnecessary - was removed
* The last column, since it has no name, was ignored for further investigation.
* Persons who do not have a special NHI cannot be admitted to the appointment.
* BAK4481 Wandis Clipson Female 2027-12-10 couldn't be born this year as It has not come yet(unless time travel machine was involved). Must be mistake or type error, unfortunately it is not clear it could be 2017 or 2007.
* Simple misspelling for Mr Graeme row 81 was forgiven and ignored (instead of 1700 year it has been changed to a 1900,which is logical)

In the query number 4 “Assuming that all patients under 18 need to be seen by Pediatric Surgery, are there any patients who need to be reassigned?” one patient appeared with an age of 0 . After further investigations an error was found in spreadsheet data. Date of birth of YBB1095 Carlye Greenhead is after referral date-which is impossible. Clarification is required.



# Queries

Additional query was created to display Patient Age and Days Waiting From Referral

Table

Description automatically generated with low confidence

First query: amount of people who have been referred for cardiothoracic.

Table

Description automatically generated

Second query: the average time taken (in days) to see a Surgeon by Department.

Table

Description automatically generated with medium confidence

Third query: patient with each Surgeon on their list and how long have they been waiting.

Graphical user interface

Description automatically generated with medium confidence

Forth query: patients who need to be reassigned

Graphical user interface, application, Word

Description automatically generated

Fifth query: percentage of patient were seen within the target of 75 days by department

Table

Description automatically generated

Version control was carried out through GitHub.

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated