Linnéuniversitetet
Kalmar Växjö

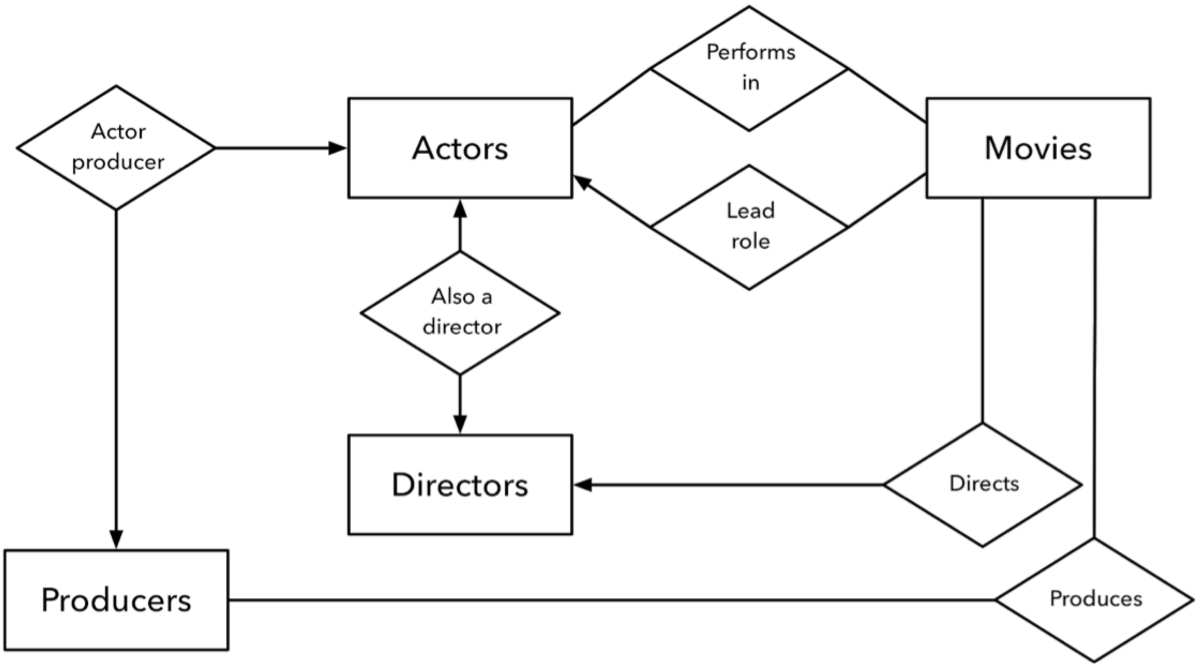
Linnaeus University

1DV503 – Database technology
Assignment 1

Group Members: <Fabian Dacic>, <Yuyao Duan>

Table of Contents

1. MoviesDB



1. There are no actors in this database that have been in no movies.

Analysis: Maybe. From the E/R diagram, we can see that actors "performs in" movies. According to the concept "Multiplicity", this relationship can represent "Many-Many relationships" which may lead to a certain situation that no actors in this database have been in no movies. However, without exact data, we cannot give a confirmed answer whether it is 100% existed or not.

2. There are some actors who have acted in more than ten movies.

Analysis: Maybe. Again, following the above understanding, the "Many-Many" relationships can lead to some actors performed in several movies. Regarding the task question, acting in more than ten movies theoretically exist in this situation. Still, without exact data, there is nothing that can either confirm or deny that an actor has acted in 10 movies. we cannot give a confirmed answer whether it is 100% existed or not.

3. Some actors have done a lead role in multiple movies.

Analysis: Maybe. According to the E/R diagram, we can clearly see the arrow from "Movies" via the "Lead role" relationship to "Actors". This shows a "Many-One" relationship in which the arrow enters the "One" side. This situation is the same as the "drinkers have favorite beers". So the statement that some actors have done a lead role in multiple movies is correct! However, to 100% confirm this statement is true, the exact data is need to confirm from this relation, so the final result for this statement based on the current information is maybe.

4. A movie can have only a maximum of two lead actors.

Analysis: False. According to the E/R diagram, “Movies”, “Lead role”, and “Actors” have the “Many-One” relationship. This means that an actor can lead in several different movies but each of these different movies should point to the typical actor as the lead one. If a movie can have two lead actors then the relationship will turn to be a “Many-Many” relationship.

5. Every director has been an actor in some movie.

Analysis: False. According to the E/R diagram, the relationship between “Actors” and “Directors” is a “One-One” relationship. However, according to the diagram, both sides of the “Also a director” relationship are described with the normal arrows which mean that some actors may not be directors and some directors may not be the actors. So, the statement “every director has been an actor in some movie” is incorrect. To make this statement true, the diagram needs to apply “rounded arrows” to describe.

6. No producer has ever been an actor.

Analysis: Maybe. According to the E/R diagram, the relationship between “Actors” and “Producer” is a “One-One” relationship. Following the same understanding from question 5, both sides of the relationship “Actor producer” are described with the normal arrows which mean that existing some producers are also actors, therefore the statement “no producer has ever been an actor” is incorrect. However, to confirm this statement is incorrect which demands exact which leads to the conclusion for this statement based on current situation is maybe.

7. A producer cannot be an actor in some other movie.

Analysis: False. According to the E/R diagram, the relationship between “Actors” and “Producers” is a “One-One” relationship and the relation between “Actors” and “Movies” (“Performs in”) is a “Many-Many” relationship. Due to the “Many-Many” relationship, a producer can be an actor in some other movies also, therefore, the statement “a producer cannot be an actor in some other movie” is incorrect.

8. There are movies with more than a dozen actors.

Analysis: Maybe. According to the E/R diagram, the relationship between “Actors” and “Movies” is “Many-Many” which means that there exist possibilities that some movies with more than a dozen actors. Considering that there is no count specified, movies that have a dozen actors are present however is not a must when it comes to casting.

9. Some producers have been a director as well.

Analysis: Maybe. This statement can be inferred from the relationships “Actor producer” and “Also a director”. According to the E/R diagram, we know that there is the possibility that existing some producers are actors who could also be directors. From the two “One-One” relationships we can deduce the statement can be true, however, due to the relationships being described with normal arrows (not the rounded arrows), we cannot 100% confirm that this statement exists.

10. Most movies have one director and one producer.

Analysis: Maybe. According to the E/R diagram, the relationship between “Movies” and “Directors” is “Many-One” while the relationship between “Movies” and “Producers” is “Many-Many”, therefore the analysis based on E/R diagram theoretically inferring the statement “most movies have one director and one producer” is incorrect. However, to 100% confirm this statement is true, the exact data is needed to confirm from this relation, so the final result for this statement based on the current information is maybe.

11. Some movies have one director but several producers.

Analysis: Maybe. According to the E/R diagram, the relationship between “Movies” and “Directors” is “Many-One” while the relationship between “Movies” and “Producers” is “Many-Many”. Most movies will fulfill these relationships, therefore, the statement “some movies have one director but several producers” is true theoretically. However, to 100% confirm this statement is true, the exact data is needed to confirm from this relation, so the final conclusion for this statement based on existing information is maybe.

12. There are some actors who have done a lead role, directed a movie, and produced a movie.

Analysis: Maybe. According to the E/R diagram, there exists such a possibility that some actors fulfill “Lead role” who are “Also a director” and “Actor producer”. The latter two “One-One” relationships may exist but not 100% confirmed due to the “normal arrow”, therefore, we say the statement “there are some actors who have done a lead role, directed a movie, and produced a movie” maybe true.

13. No movie has a director who also acted in that movie.

Analysis: Maybe. According to the E/R diagram, the relationship between “Movies” and “Directors” is “Many-One” and the relationship between “Directors” and “Actors” is “One-One”. Theoretically speaking, there exists such a possibility that some movies are directed by certain directors who are also the actors in them, therefore, the statement “no movie has a director who also acted in that movie” is incorrect. However, to 100% confirm this statement is true, the exact data is needed to confirm from this relation, so the final conclusion for this statement based on existing information is maybe.

2. Births

1. Every baby is the result of a unique birth, and every birth is of a unique baby.

Analysis: According to the description, the binary relationship between Births and Babies is a "One-One" relationship, therefore the arrows should be applied in this case. Moreover, due to the description stating that "every baby is a unique birth, and every birth is a unique baby", therefore this relationship should use rounded arrows to enter both sides to fulfill the "exactly one" relationship. The diagram should be described as following:

Design flaws: For babies it is quite common that different babies have the same birth. In reality, it is difficult to fulfill the requirements that "every baby has a unique birth". This design ignores that one birth may have more than one baby, i.e. the possibility of twins and multiple births. This design does not include other aspects of birth such as stillbirth, cesarean section or other ways of birth-occurring during the process which could act as attributes to births, and in that note, it does not include aspects of babies being born such as state, weight, gender which, once again, could be considered as attributes and so forth.

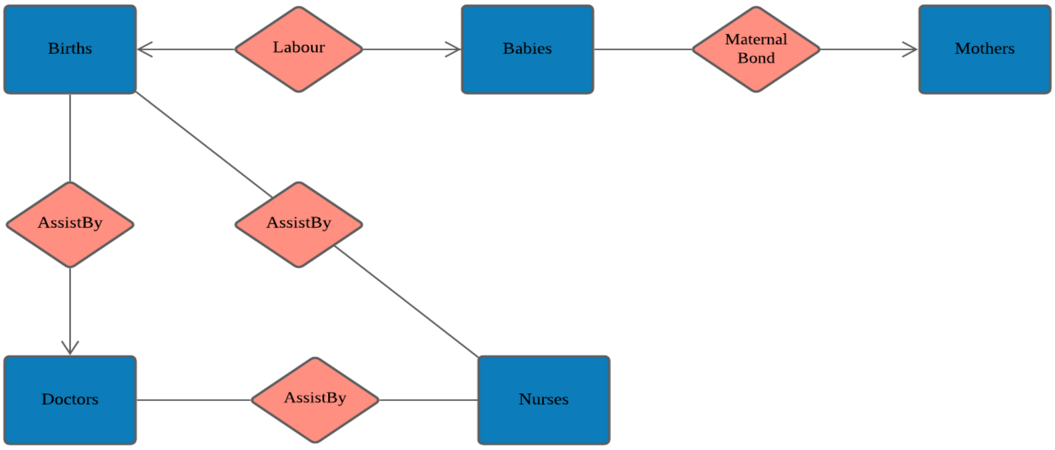
2. In addition to (1), every baby has a unique mother.

Analysis: According to the description, the relationship between Babies and Mothers can be described as the right side of the diagram.

Design flaws: This design ignores the situation that each birth may involve more than one baby and relating to the exact one mother. For example, multiple birth scenarios may involve a certain number of offsprings to be born at various intervals however they will have the same mother, which in this case, this design is not fitted for such scenarios. In addition, the Mother entity may face one situation that "mothers in labor" who still wait for birth, therefore, this E/R diagram system could not properly handle certain cases. Further design may need to differ different categorized mothers.

3. In addition to (1) and (2), for every birth there is a unique doctor.

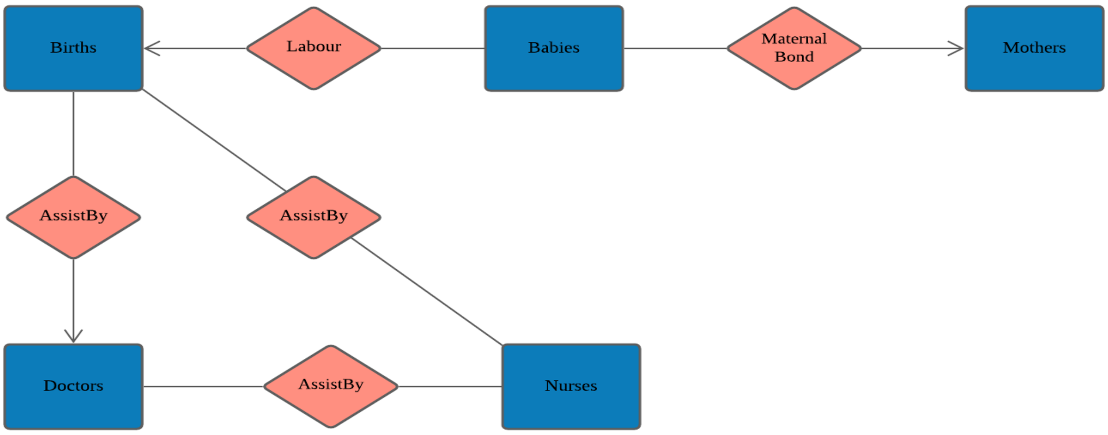
Analysis: According to the description, the relationship between Babies and Mothers can be described as a "Many-One" relationship, however, we should consider that even one Doctor relates to many Births, and every birth has a unique doctor. Same as “Beer-Bestseller” example, there are many births, however, each birth has to have a doctor and a rounded arrow should apply here. Furthermore, doctors are assisted by nurses and the relationship between them is “Many-Many”. Births are also assisted by nurses and the relationship between them is “Many-Many”.



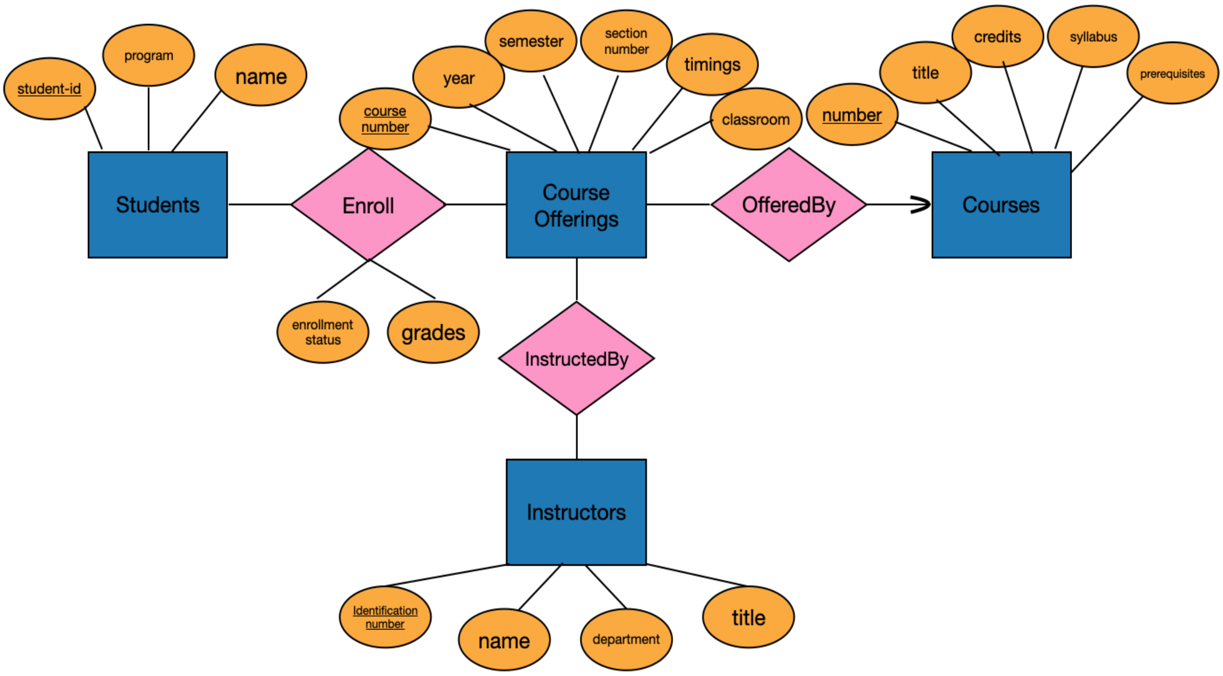
Design flaws: When it comes to entity “Doctors”, the main design flaw of this diagram is that not any doctor can perform delivering of the babies during birth however it has to be exactly an obstetrician therefore not all doctors all qualified to do so. In order to have a more precise E/R diagram, an extra relationship or an attribute should be added that specifies the qualification of said obstetricians to separate them from the general doctors would be ideal.

4. Suppose we change our viewpoint to allow a birth to involve more than one baby born to one mother. How would you represent the fact that every baby still has a unique mother?

Analysis: By applying the new viewpoint, the “Births” and “Babies” should be updated as the “Many-One” relationship, every baby should be able to find exactly one birth, but each birth may include more than one baby. On the other side, each baby should relate to a mother and each mother may have more than one baby.



3. The registrar's office

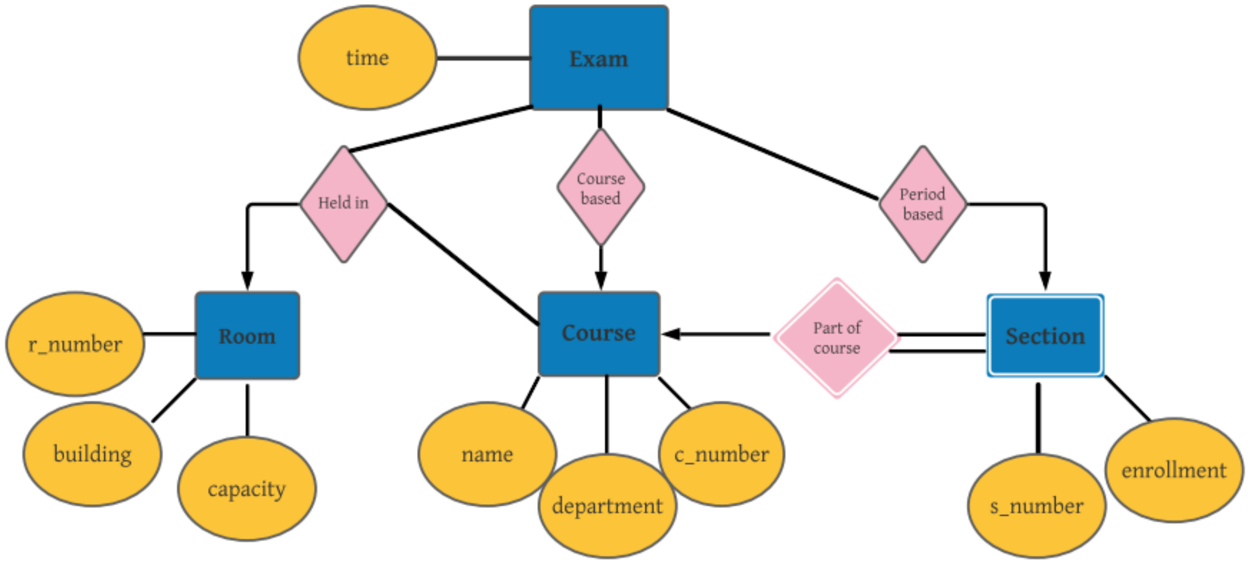


The E/R diagram of the registrar's office is as above: Students entity set connects Course Offerings entity set by the "Many-Many" relationship Enroll which owns attributes "grades" and "enrollment status" that are functions of both Students and Course Offerings. The students' enrollment status and grades of the enrolled courses can be checked. The key of Students entity set is "student-id" attribute and the key of Course Offerings entity set is "course number".

Course Offerings entity set connects Courses entity set by the relationship OfferedBy which is a "Many-One" relationship. Based on the understanding that Course entity is more like an overall structure/skeleton of a course, e.g. "1DV503" while Course Offerings more like detailed courses that students can register, e.g. "1DV503-20VT", "1DV503-21VT" etc. Therefore, based on the one course the university can provide various similar course offerings but inherent with differences. The key of Course Offerings is the attribute "course number" and the key of Courses is "number". Since a course offering has to have a course to support, therefore, the relationship is connected by a rounded arrow, e.g., no matter what versions of "1DV503-XXVT" have to have a course "DV503" to correspond.

In order to avoid redundancy and potential inconsistency, the Course Offerings entity set connects Instructors entity set with the "Many-Many" relationship InstructedBy. Therefore, the Course Offerings entity set removes the instructor(s) attribute which the function will be realized by the connected relationship. Through Instructors entity set, Course Offerings entity set can have related instructors' information. The key of the Instructors entity set is the "identification number" attribute in this case.

4. Classroom scheduling



Why all attributes have been included is to ease the identification process of an entity in the database therefore for most identity sets, there are exactly three attributes. However there are cases where there are no sections of a course or at the very least, one, and in such case it would better to just remove the entity sets "Course" and "Section" and instead add an attribute to the "Exam" entity in which the primary key and only key would be the ID of the course or section for that matter which in the case of the E/R diagram constructed, "c_number" and "s_number" are also the key attributes for the entities "Course" and "Section".

It is also to be noted that considering that entity "Section" is a weak entity, it has full participation through the relationship "Part of course" to "Course". That is due to the reason that if a section does indeed exist, it is indubitably part of a course (period-based course) however the course has partial participation due to the fact that not every course has sections. Now, for the only and primary key being the ID of the course or the section, the reason for that is depending on the location of the university (e.g. in the UK, some universities have courses which do not have any exams), having numerous attributes to identify the course in the "Exam" entity would just create problems for the identification process of it. The entity "Room" can also be removed, and instead be an attribute for the "Exam" entity instead however in the case of the E/R diagram constructed, it has a multi-attribute key which are "r_number" and "building" which help identify easier the entity "Room" and avoid confusion.

5. Relational algebra

student (id, name)

enrolledIn (id, code)

subject (code, lecturer)

1. What are the names of students enrolled in 2dv513?

$R_1 = \sigma_{(code = '2dv513')}$ (enrolledIn)

R_2 = student

$R_3 = R_1 ⟗ R_2$

$R_4 = \sigma_{(code\ != NULL)} (R_3)$

The names of students enrolled in 2dv513: $\pi_{(name)} (R_4)$

2. What are the names of students in both 1dv513 and 2dv513?

$R_5 = \sigma_{(code = '1dv513')}$ (enrolledIn)

R_2 = student

$R_6 = R_5 ⟗ R_2$

$R_7 = \sigma_{(code\ != NULL)} (R_6)$

The names of students enrolled in 1dv513: $\pi_{(name)} (R_7)$

The names of students in both 1dv513 and 2dv513: $\pi_{(name)} (R_4) \cap \pi_{(name)} (R_7)$

3. Who teaches 2dv610?

$R_8 = \sigma_{(code = '2dv610')}$ (subject)

The lecturer(s) teach(es) 2dv610: $\pi_{(lecturer)} (R_8)$

4. Who teaches 1dv513 and 2dv513?

$R_9 = \sigma_{(code = '1dv513')}$ (subject)

The lecturer(s) teach(es) 1dv513: $\pi_{(lecturer)} (R_9)$

$R_{10} = \sigma_{(code = '2dv513')}$ (subject)

The lecturer(s) teach(es) 2dv513: $\pi_{(lecturer)} (R_{10})$

The lecturer(s) teach(es) both 1dv513 and 2dv513: $\pi_{(lecturer)} (R_9) \cap \pi_{(lecturer)} (R_{10})$

5. What are the names of students who are taking a subject not taught by Ilir?

$R_{11} = \sigma_{(lecturer\ != 'Ilir')}$ (subject)

R_{12} = enrolledIn

$R_{13} = R_{12} ⋉ R_{11}$

$R_{14} = \boldsymbol{\pi}_{(id,\ code)} (R_{13})$

$R_{15} = R_{14} \bowtie R_{12}$

The names of students who are taking a subject not taught by Ilir: $\boldsymbol{\pi}_{(name)} (R_{15})$

6. FDs and Normalization

The system uses a single relation for all data, with the following schema:

Interviews (manager, applicant, day, time, room).

1. Find functional dependencies.

The functional dependencies are:

manager day → room

applicant day → manager time room

room day time → manager applicant

manager day time → applicant room

applicant day time → manager room

applicant day room → manager time

manager applicant day time → room

manager applicant day room → time

2. Find the keys of the relation.

L	M	R
day	manager, applicant, time, room	

This table helps to identify keys of the relation:

L column includes the attributes which only occur on the left side of FDs, it must be a part of the keys; M column includes the attributes which occur both on left and right sides of FDs, it might be a part of keys; R column only includes the attributes which occur only on the right sides of FDs. From above FDs, the different attributes have been allocated as above.

Analysis:

Step1: calculate $(day)^+$ → not a key

Step2: calculate $(day, manager)^+$ which can only infer day, manager, room; → not a key.

Step3: calculate $(day, applicant)^+$ which can infer day, manager, applicant, time, room; → <u>a key</u>.

Step4: calculate $(day, time)^+$ which can only infer day time; → not a key.

Step5: calculate $(day, room)^+$ which can only infer day, room; → not a key.

Step6: calculate $(day, manager, applicant)^+$ which can infer day, manager, applicant; → not a key.

Step7: calculate $(day, manager, time)^+$ which can infer day, manager, applicant, time, room; → <u>a key</u>.

Step8: calculate $(day, manager, room)^+$ which can infer day, manager, room; → not a key.

Step9: calculate $(day, applicant, time)^+$ which can infer day, manager, applicant, time, room; → <u>a superkey</u>.

Step10: calculate $(day, applicant, room)^+$ which can infer day, manager, applicant, time, room; → <u>a superkey</u>.

Step11: calculate (day, time, room)$^+$ which can infer day, manager, applicant, time, room; →a key.

In all, we get keys: 1) {day applicant};

2) {day manager time};

3) {day time room};

Furthermore, we get superkeys: 1) {day applicant time};

2) {day applicant room};

3) {day manager applicant time room};

3. Show that the relation is in 3NF but not in BCNF.

According to definition: A relation R is in BCNF ⇔ whenever X → A is a nontrivial FD that holds in R, X is a superkey. As we can see, (day manager → room), which is a FD and it is also nontrivial, however, left side of this FD is not a superkey that violates BCNF.

By comparison, the definition for 3NF: X → A violates 3NF if and only if X is not a superkey, and also A is not prime i.e. whenever X → A is a nontrivial FD, either X is a superkey, or A is a member of some key. Thus, we can see from above all the FDs, either the right sides of the FDs are part of the keys or the left sides are/contain superkeys, therefore it fulfills 3NF.

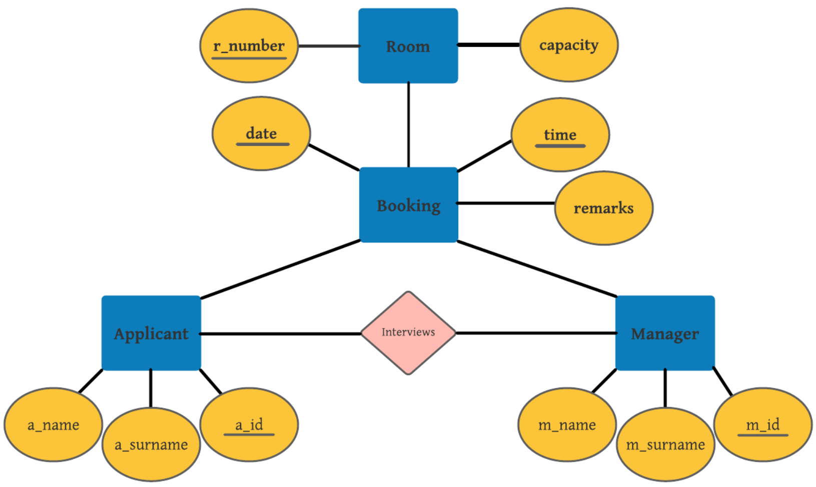
4. Decompose the relation in relations that are in BCNF.

Due to the FD (manager day → room) violates BCNF, therefore we can apply the formula of decomposition into BCNF: $R_1 = X^+ =$ {manager day}$^+$ = {manager day room};

$R_2 = R -$ {manager day}$^+$ – {manager day} = {manager applicant day time};

So, the two relations are: Interviews_1 (manager, day, room) and Interviews_2 (manager, applicant, day, time).

5. Draw an E/R diagram that describes the system. Try to incorporate all dependencies.



In this E/R diagram system, we identified four entities: Applicant, Manager, Booking and Room. Applicant and Manager are connected by the relationship Interviews. Manager and Applicant's relationship as described (managers make appointments with applicants) will together influence the booking system (Booking) in terms of date (day), time and remarks regarding the planned meetings. The booking system will eventually determine which rooms are available for the focal day's interviews.