

COMP518 Assignment 2 (of 3): 25% of the final grade

Due: 17:00 on Friday, November 4th

Please submit your solutions **electronically (a single file in PDF format only)** at the electronic submission system of the Computer Science Department which you can find at the following url. **Do not use red colour in your solutions.**

<https://sam.csc.liv.ac.uk/COMP/Submissions.pl?strModule=COMP518>.

Learning outcomes:

1. Critical understanding of the nature of relational databases.
2. Design and implement relational databases with multiple tables.

Assessment's purpose:

1. Create the logical data model from an Entity-Relationship diagram.
2. Decompose a large table using the process of normalisation.

Failure in the assessment may be compensated for by higher marks in other components of the module. Marking of subquestions will be based on the marking descriptors of the University's Code of Practice on Assessment. Standard UoL penalty applies for late submission. The last possible date of submission is 2 weeks after the original deadline, unless a feedback is given in an earlier date (students will be informed if the second case apply at least 5 days earlier). Please be aware of the University guidelines on plagiarism and collusion.

Total: 100 marks

Question 1

Background

Cloud computing is changing the world. It is Internet-based computing: hundreds of thousands of remote computers are interconnected to provide all kinds of resources, including data, applications, as well as computing power. For example, Google provides online services named *Google Docs*¹, which we may consider as an alternative to Microsoft Office. Google Docs differs from conventional software products in the sense that, when using Google Docs, the data and the software locate no longer in the our own machines but rather somewhere “in the cloud.” We need not store the data or install any software. What we have to do is to get a Google Docs account and then simply open an Internet browser. This is also the situation when we rent online movies on Amazon². We pay about 4 dollars to enjoy a movie online. We do not store the movie in our own machines, nor do we install any movie players. We need only get an Amazon account to do the business.

One of the key observations from the above two examples is that machines become virtual to users in the cloud era. When we are editing online documents or watching online movies, we are also using certain computing power provided by Google or Amazon.

Task

You are given as an input, the Enhanced Entity-Relationship diagram of Figure 1, that is meant to be one possible conceptual representation of the users’ requirements. You are required to create the **logical data model**.

1. **(45 marks)** Create the appropriate **relations** to represent the **entities** and **relationships** of the E-R diagram.
2. **(15 marks)** Designate the **primary key**, and any **alternate** or **foreign keys** of each relation together with their **references**.

For all the above, use the methodology covered in the lectures and described in Chapter 17 of [1], providing clear annotations and explanations of all the steps. In case you deviated from the general guidelines, explain in detail why you did so. Also discuss in detail any other possible alternatives you may have considered (if any), and why you decided not to choose them.

¹See <https://docs.google.com>.

²See <http://www.amazon.com>.

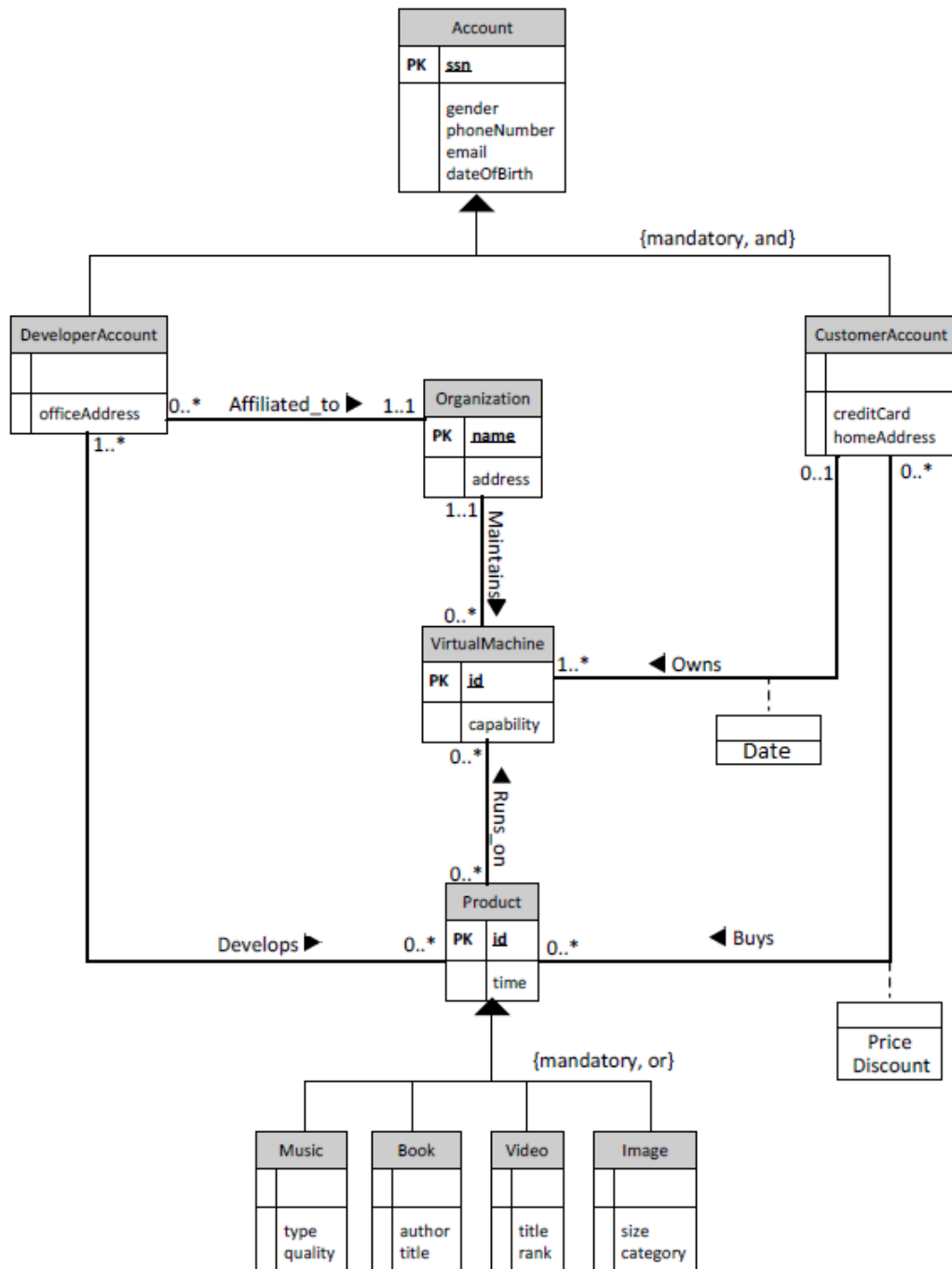


Figure 1: Entity-Relationship Diagram.

Question 2

Background

You are asked by the organisers of the Conference on Web and Internet Economics (WINE), to design a database that captures as many requirements as possible. The Conference on Web and Internet Economics (WINE, formerly Workshop on Internet and Network Economics) is an interdisciplinary forum for the exchange of ideas and results on incentives and computation arising from these various fields. Invited speakers of previous WINEs include not only many famous theorists from prestigious universities in the world, but also top researchers from many big businesses, such as Microsoft Research, Yahoo! Research, IBM Research, Google Inc., ebay Research Labs and so on. Participants of WINEs come from universities and industries as well.

The organisers anticipate several hundred papers to be submitted for the conference. All the submissions are to be reviewed by referees and, due to the length of the conference, no more than 100 of the submitted papers are to be accepted. Each paper that is submitted will be classified into one of the conference topics.

WINE19 will last for 4 days, during which 8 topics (e.g., “Mechanism Design”) will be discussed in two parallel streams. The conference sessions will be held in 2 large lecture theatres, with each session lasting a maximum of 2 hours and being dedicated to one of the topics; each topic takes 4 sessions (one per day). Therefore, each day there will be 4 sessions at each lecture theatre, summing up to 8 sessions per day (one for each topic). Each paper is presented in exactly one session during a 20-minute slot followed by 5 minutes of questions.

The chair of the Programme Committee appoints the Senior Programme Committee (SPC’s) members. Every SPC must be an authority in at least one of the conference topics and is solely responsible for reviewing all submissions within his or her topics of expertise. Normally an SPC handles several submitted papers and assigns each paper to three referees. It is the organisers intention that at most 16 papers per topic will be accepted.

Several hundreds participants are expected to attend the conference. Each participant typically plans to attend some of the sessions of their interest and may also present their own accepted papers, if they have any. When enrolling, every participant is required to provide some personal data (e.g., title, name, affiliation, address, etc.) and state which session(s) they wish to register for; this is required for calculating the individual fee for every single participant.

The standard single session fee is £30. The following discount scheme will apply: 20% reduction for all participants with papers accepted, 30% for participants who actually present their papers at the conference and 40% for the session chairperson. Every participant can register for as many sessions as they wish, provided the selected sessions do not clash.

The required database is to hold all the relevant data about the organization of the conference, including information about participants, the sessions they registered for etc. The necessary details about all the papers which are accepted for presentation (together with their session allocations) are also to be stored. The basic function of the database is to provide information on the conference programme and registration status of the participants.

Task

(40 marks) A database designer that possibly didn’t know how to create a proper Conceptual Data Model using Entity-Relationship modeling, created a large relation schema to store all the data associated with the registration information of participants to sessions, together with information about the accepted papers that are to be presented in a session (e.g. who was the SPC that handled a particular paper etc). The relation has the following schema:

ConferenceData (participantID, participantName, participantAddress, sessionLocation, sessionDate, sessionStartingTime, sessionDuration, topicID, topicName, paperID, paperTitle, SPCID, SPCName).

You are required to decompose, using the formal process of normalisation as is described in Chapters 14 and 15 of [1], the ConferenceData table into relations which are in 3rd normal form. In the absence of any data you may assume that the table is flattened. You are not supposed to add any further attribute, like authors or presenters of the papers. You should present your work logically providing details about the result of each stage of the normalisation process with clear annotations and explanations.

Note

- Note that the problem description, discusses a much more general scenario, so some of the information that is described there, might not be useful for your task. However, the scenario will be useful for identifying the functional dependencies.

References

- [1] Thomas M. Connolly, Carolyn E. Begg. Database systems: *A Practical Approach to Design, Implementation and Management. 5th Edition*, Addison-Wesley, 2010.