**Introductory Assignment for Semester ISE, course DMDD 29/8/2018**

The course DMDD (Data Modeling and Database Design) is all about the structure of data.

This introductory assignment introduces you intuitively to the most important data structure aspects. The tasks to be carried out are given after the explanation below, and a few concrete examples of the conference data are given on page 2.

The assignment concerns a conference on computer science, in which the participants can enroll in several workshops. A workshop is held in one of the rooms of the conference center, and can be held in more than one session. There are two or more workshops in each session. All participants receive a unique participant number, and their name is recorded and written on their badge.

A few important data structure aspects are:

* What are the types of facts that the data are made up of?

For instance (see page 2): the following three types of facts are in the data:

* + Fact type 1: Workshops have titles.
  + Fact types 2 and 3: Sessions start and end at a certain time.
* What are the business rules that the data should obey?

For instance (see page 2): the following two business rules apply:

* + Business rule 1: Every workshop has exactly one title
  + Business rule 2: For each session: the end time must be later than the start time.

The term data quality is used for how well the data meets various requirements, such as:

* Do we have all the facts we need to know (completeness of the data)?
* Is each fact stored only once (is there no redundancy in the data)?
* Are the facts correct, i.e: do the facts obey all the business rules (integrity of the data)?

In practice, it is very hard to make sure that the data quality is high. If data integrity is important, then a relational database is a good choice as implementation platform (as opposed to several NoSQL options like MongoDB). In this course we will focus on relational databases.

**Tasks**

The conference schedule and enrollment data are to be stored in a small relational database. See page 2 for a few concrete examples of all the data to be stored.

1. How many different types of facts are to be stored?

For example:

* the title of a workshop is one type of fact, with facts like:  
   Workshop A is titled: Model Driven Architecture.  
   Workshop E is titled: LINQ.
* the closing time of a session is another type of fact, with facts like:  
   Session 2 ends at 12:30.  
   Session 4 ends at 15:45.

1. What are the business rules that apply to the data? There are more than those given above.
2. Does the population satisfy the business rules? The examples on page 2 contain a few errors. Which ones?
3. Give a redundancy-free relational database schema (tables, columns, primary keys, foreign key references, NOT NULL integrity rules, etc) to store the conference data in.
4. Fill the tables with all the data shown on page 2, but without the errors.

Bring your answers to the first regular class meeting of the DMDD course, and be prepared to explain your database schema.

Here is the conference schedule, and a few enrollments of participants in the workshops:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Session** | **Time** | **Workshop** | **Title** | **Room** | **Participants** |
| 1 | 9.15 – 10.45 | A | Model Driven Architecture | Eagle | 99 Janny Mulder  12 Jan Verkerk  40 Janny Mulder  … ………………. |
| B | Scrum Project Management | Cockatoo | 33 Carel Doorman  62 Fred Overbeek  15 Geert van Straaten  … ………………. |
| 2 | 11.00 – 12.30 | C | Extreme Programming | Toucan | 33 Carel Doorman  99 Janny Mulder  62 Fred Overbeek  … ………………. |
| D | Data Mining | Cockatoo | 17 Pieter Carelse  58 Marten Goudsmid  12 Herman Geest  … ………………. |
| E | LINQ | Toucan | 86 John van de Berg  62 Fred Overbeek  15 Geert van Straaten  … ………………. |
| 3 | 13.30 – 14.00 | F | Enterprise Application Integration | Cockatoo | 86 John van de Berg  55 Maria Cornelisse  41 Joris Dekker  … ………………. |
| G |  | Toucan | 90 Vera van Aartsen  62 Fred Overbeek  15 Geert van Straaten  … ………………. |
| 4 | 14.15 – 15.45 | B | Scrum Project Management | Cockatoo | 86 John van de Berg  41 Joris Dekker  99 Janny Mulder  … ………………. |
| D | Data Mining | Toucan | 90 Vera van Aartsen  33 Carel Doorman  38 Jack Evans  … ………………. |
| E | LINQ | Eagle | 17 Pieter Carelse  12 Jan Verkerk  40 Janny Mulder  … ………………. |

1. Starting time of a session:  
   - Session 1 begins at 09:15.  
   - Session 2 begins at 11:00.  
   Letter of a workshop:  
   - Scrum project management has the letter B.  
   -Data mining has the letter D  
   Each room has a name:  
   - One room is called Eagle  
   - Another room is called Toucan  
     
     
   Each participant has a unique identifier:  
   - Pieter Carelse has the identifier 17  
   - Janny Mulder has the identifier 40
2. Identifiers are unique  
   Workshop letters and titles are a unique combination
3. There is a title missing at session 3 workshop G.   
   Geert van Staaten has no number in front of his full name.   
   Showing student ID is unnecessary as well.   
   Both rooms Toucan are used at the same time during session 2.   
   Fred Overbeek is in 2 rooms at the same time in session 2.   
   Janny Mulder in twice.