THE UNIVERSITY OF MELBOURNE

School of Computing and Information Systems
End of Semester 2 2017 SUPPLEMENTARY Examination

INFO90002 DATABASE SYSTEMS & INFORMATION MODELLING

Reading time: 15 minutes

Writing time: 180 minutes

This paper has 6 pages, including this page.

Authorised Materials:	No materials are authorised.

Instructions to Invigilators:

The examination paper IS TO REMAIN in the examination room.

Students are to be provided with standard script books.

Instructions to Students:

The total mark for this paper is 70 marks, representing 70% of your final assessment.

Ensure your student number is written on all script books during writing time.

This exam paper has 9 questions, some with multiple parts.

Attempt all questions.

Answer all questions on the right-hand lined pages of the script book.

Start the answer to each question on a new page in the script book.

The left-hand unlined pages of script books are for draft working and notes and will not be marked.

Write legibly in blue or black pen.

Textual answers can be in point form.

All electronic devices (including mobile phones and phone alarms) must be switched off and remain under your desk until you leave the examination venue. No items may be taken to the toilet.

Q1 – Database Design (16 Marks)

Scenario:

You are a data modeller working for a startup who are creating a new on-demand streaming music service, "Classic Play". The service allows customers to listen to recordings of classical music on their phone. We need a database to store searchable information about the music we offer and record customer activity.

We will allow customers to search for their favourite composers, performers, pieces, or performances. For example, the composer "Ludwig van Beethoven" composed many pieces, one of which is called "Moonlight Sonata". This piece has been recorded multiple times by different performers. A customer can choose which of these performance they wish to listen to, for example: "Moonlight Sonata, played by Glen Gould, recorded live at Carnegie Hall on 1st January 1960". We want customers to be able to search by composer, piece, performer, venue or date.

A piece may have multiple composers, and a performance may have multiple performers. We keep composers and performers in the same list, since some artists perform both roles, either on the same or on different recordings.

Several different performances may be collected into an album. Each album has a title, such as "Great Piano Pieces by Beethoven". Some popular performances will be put on several albums. For each of the performances on an album, we need to record their sequence number within the album, so they are played in the right order. All performances are classified as belonging to a genre. We have about 20 genres - this number is not likely to grow.

When someone registers to become a customer, we record their username, credit card number, email address, and favourite composer. Whenever a customer streams a performance, we keep track of who streamed what, and when. This information is used to generate business analytics, to make recommendations to customers based on music popularity, and to work out the correct amounts to pay the artists.

Question 1 instructions:

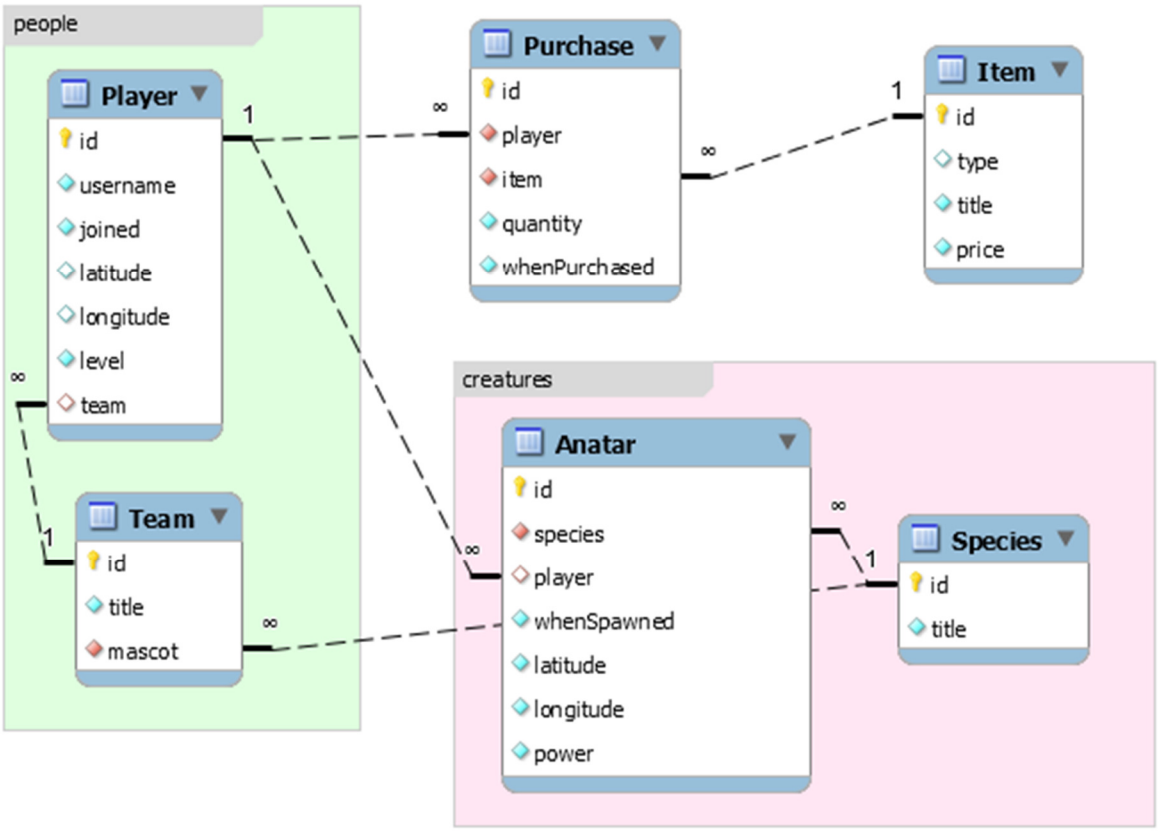
For the above scenario, draw a ***physical*** data model. (You must show the data-types of columns.)

Use crows-foot notation for relationships, and join the lines to the related columns. Show the cardinalities of relationships and whether they are optional or mandatory.

You do not need to add names to relationships. You do not need to write a data dictionary. If you wish, you may explain the reasoning behind any design decisions or assumptions you made, but this is not needed. You only need to submit the ER diagram.

Q2 – SQL (17 Marks)

Consider the following data model and *sample* data (not all data is shown) for a phone-based augmented-reality game. About 10 million human players move around the (physical) world, catching virtual animals called “Anatars”. Each Anatar is of a particular Species (there are currently 100 species). Player and Anatar locations are recorded as latitudes and longitudes. Each player belongs to one of 4 teams. Players can purchase virtual items, some of which are food or medicine items (marked with type ‘F’ or ‘M’).



Anatar

id	species	player	whenSpawned	latitude	longitude	power
1	69	10	2016-08-01 01:01:01	-37.784186	144.859335	10
2	10	9	2016-08-01 02:02:02	-37.785251	144.860921	20
3	50	8	2016-08-01 03:03:03	-37.786316	144.862506	30
4	84	7	2016-08-01 04:04:04	-37.787381	144.864092	40
5	96	6	2016-08-01 05:05:05	-37.788446	144.865677	50
6	23	5	2016-08-01 06:06:06	-37.789511	144.867263	60
7	133	4	2016-08-01 07:07:07	-37.790576	144.868848	70
8	92	3	2016-08-01 08:08:08	-37.791641	144.870434	80
9	69	2	2016-08-01 09:09:09	-37.792706	144.872019	90
10	10	1	2016-08-01 10:10:10	-37.793771	144.873604	100

Species

id	title
1	Elizabeth Springs Goby
2	Murray Hardyhead
3	Golden Galaxias
4	Swan Galaxias
5	Barred Galaxias
6	Clarence Galaxias
7	Northern River Shark
8	Clarence River Cod
9	Trout Cod
10	Mary River Cod

Player

id	username	joined	latitude	longitude	level	team
1	Barton	2016-07-01 01:01:01	-37.799040	144.962968	1	U
2	Deakin	2016-07-02 02:02:02	-37.805405	144.971206	2	U
3	Watson	2016-07-03 03:03:03	-37.809897	144.963579	3	K
4	Reid	2016-07-04 04:04:04	-37.811821	144.980993	4	D
5	Fisher	2016-07-05 05:05:05	-37.826703	144.976608	5	U
6	Cook	2016-07-06 06:06:06	-37.864674	144.971305	6	K
7	Hughes	2016-07-07 07:07:07	-37.830160	144.893845	7	U
8	Bruce	2016-07-08 08:08:08	-37.802347	145.002998	8	U
9	Scullin	2016-07-09 09:09:09	-37.781793	144.912453	9	K
10	Lyons	2016-07-10 10:10:10	-37.783369	144.951774	10	D

Team

id	title	mascot
D	Daintree	146
K	Kakadu	145
U	Uluru	144
NULL	NULL	NULL

Purchase

id	player	item	quantity	whenPurchased
1	1	2	2	2016-08-01 01:01:01
2	2	5	3	2016-08-02 02:02:02
3	3	2	1	2016-08-03 03:03:03
4	4	1	2	2016-08-04 04:04:04
5	5	2	3	2016-08-05 05:05:05
6	6	5	1	2016-08-06 06:06:06
7	7	2	2	2016-08-07 07:07:07
8	8	1	3	2016-08-08 08:08:08
9	9	2	1	2016-08-09 09:09:09
10	1	3	2	2016-08-10 01:01:01

Item

id	type	title	price
1	NULL	Powerup	1.11
2	NULL	Coin	2.22
3	F	Apple	3.33
4	F	Pear	4.44
5	NULL	Gem	5.55
6	M	Health pack	6.66
7	M	Bandage	7.77
8	NULL	Booster	8.88
NULL	NULL	NULL	NULL

Tasks:

Questions 2A-D require you to write one single SQL statement per question. Do not use Views or Temporary Tables. Use reader-friendly code layout. Ensure user friendly output by sorting and renaming fields where appropriate.

A) How many species names begin with each letter of the alphabet?
List the top 5 letters in order, each with its species count. Don't worry about ties. (3 Marks)

B) Find the items which have been purchased by at least 2 different players who are located south of latitude -37.5 degrees. Show each item's name and title. (5 Marks)

C) Which player(s) have spent the most money on Food items?
List the usernames and money spent for the top 3 spenders. Don't worry about ties. (5 Marks)

D) List any players who have captured at least one Anatar which is of the same species as that player's team's mascot. (4 Marks)

Q3 – Normalisation (4 Marks)

We are modelling a database to store soccer (football) players and the goals they score in games.

Each player has an id and name and belongs to a team.

Each goal kicked gets a unique ordinal number within the game: that is, the first goal in a particular game is goal 1, the next goal is goal 2. We track the time at which each goal was scored, the date on which the game took place, and the home and away teams that took part in the game.

Our modeller has arrived at the following relation (primary key is underlined).
But this relation is not in third normal form.

PLAYERGOALS (playerid, givenName, surname, teamid, teamname,
(goalId, gameId, gameDate, homeTeam, awayTeam, timeScored)
)

Your job is to convert the relation to 3rd normal form.
Mark your primary keys with a solid underline, and your foreign keys with a dotted underline.
(Any attributes that are both primary and foreign keys should get both underlines.)

You don't need to show intermediate normal forms – just the 3rd normal form you end up with. (4 marks)

Q4 – Physical Design (4 Marks)

For each of the following columns in the "Anatar" database shown in **question 2**,
write the column's correct MySQL data type, include the width of the column where appropriate.
Each is worth 1/2 mark.

i) Player.id
ii) Player.latitude
iii) Player.team

iv) Anatar.whenSpawned

v) Species.id

vi) Purchase.id
vii) Purchase.quantity

viii) Item.price

Q5 – Applications (8 Marks)

A) Explain two (2) major reasons why databases usually require application software. (4 Marks)

B) Draw a diagram of a 4-tier web-app architecture. Label each component. (4 Marks)

Q6 – Distributed Databases (6 Marks)

A) State the difference between replicated and partitioned databases,
and describe one real-world example of each. (4 Mark)

B) Explain one advantage and one cost of the synchronous updating of distributed databases. (2 Marks)

Q7 – Transactions (6 Marks)

A) Explain, using a timeline diagram and text, how 2 processes simultaneously accessing the same data can create the ***Uncommitted Data*** problem, using the example of booking seats on an aeroplane. (4 Marks)

B) List the 3 main SQL commands used in transaction control, and briefly describe each one. (2 Marks)

Q8 – Database Administration (5 Marks)

A) Explain the purpose and operation of the database buffer cache. (3 Marks)

B) What is the main advantage of having offsite backups? (2 Mark)

Q9 – NoSQL (4 Marks)

Explain the difference between a key-value database and a document database.

END OF EXAM