TEESSIDE UNIVERSITY
SCHOOL OF COMPUTING

ICA SPECIFICATION

Module Title: Advanced Database Systems	**Module Leader: G.Capper**
	Module Code: COM3018-N-YEAR-2013
Assignment Title: ICA2013-14	**Deadline Date:** **25th April 2014**
	Deadline Time: 16:00
	Submission Method: Blackboard ☑
	Feedback Method: Word document via Blackboard

ONLINE Submission Notes:

- Please follow carefully the instructions given on the Assignment Specification
- When an extension has been granted, a fully completed and signed Extension form must be submitted to the module leader as soon as it had been signed.

FULL DETAILS OF PHASE 1 OF THE ASSIGNMENT ARE ATTACHED, INCLUDING MARKING & GRADING CRITERIA. PHASE 2 WILL BE RELEASED LATER IN THE ACADEMIC YEAR

Phase 1: You should refer to the supplied documents to complete the following tasks.

Task 1

Correct the initial data model (Figure 1) as shown in Document 1. You should redraw the diagram using a named notation and explain the corrections you have made. At this stage you should **not** add any new objects to the diagram.

Marks are awarded for identifying an error and for providing a correct solution.

Task 2

Update your solution to Task 1 by completing the data model. You should add new objects you identify from the supplied documents. Show as much implementation detail as possible. Name your notation.

Marks are awarded for modelling the remaining data from the attached documents.

Task 3

Implement the model above as a SQLServer 2012 database schema & associated tables. Document your solution by reproducing the SQLServer Database Diagram. Select **two associated** tables and show how you derived the Data Definition Language code (T-SQL) from the model. If you built the tables using the Visual Designer you must *script* the tables and relate your model to the generated code.

Marks are awarded for implementing SQLServer database tables that correspond to your model in Task 2 (as shown in your diagrams). Marks are awarded for Data Definition Language code and an explanation on how it was derived.

You must now insert some sample data into the tables, the supplied documents can be used as a source.

Task 4

Create a stored procedure that returns for a supplied candidate name all the offers made to that candidate, showing points, the course name and degree type. Reproduce the Data Manipulation Language code (T-SQL), a script showing its execution and the text outcome of that execution. Draw a graphical representation (parse tree) of what might be an optimised execution of this query.

Marks are awarded for Data Manipulation Language code and evidence of its successful execution. Marks are awarded for the parse tree of optimised execution.

Task 5

Create an application using Visual Studio 2012 that imports part of your database schema created in Task 3 and meets the requirements as outlined in Document 8. This must be implemented using C# and utilise the Entity Framework classes. Identify a potential problem existing in your current database design that this set of requirements exposes. Outline a possible solution that could be implemented.

Marks are awarded for C# code and evidence of its successful execution. Marks are awarded for identification of a problem and a proposed solution.

Each of these 5 tasks will be marked out of **10** making a total of **50** for phase 1 of the ICA. Phase 2 will be available later to complete the ICA.

Note that the use of any other technology than those specified will be marked as 0 for that task.

Introduction:

SomeCity University holds interview days for candidates applying to its degree courses. The purpose is to appraise the students' ability and then make them an offer regarding acceptance onto a degree course; an offer is the total number of points they need to gain from their A-level or other qualifications. SomeCity would like to develop a database and associated applications to manage this activity. The following documents highlight the data in the system. Read through the documents carefully. If you have queries you can contact your local tutor who will act as a client from SomeCity University.

Document 1: Draft data model created by A.Analyst 27.10.13

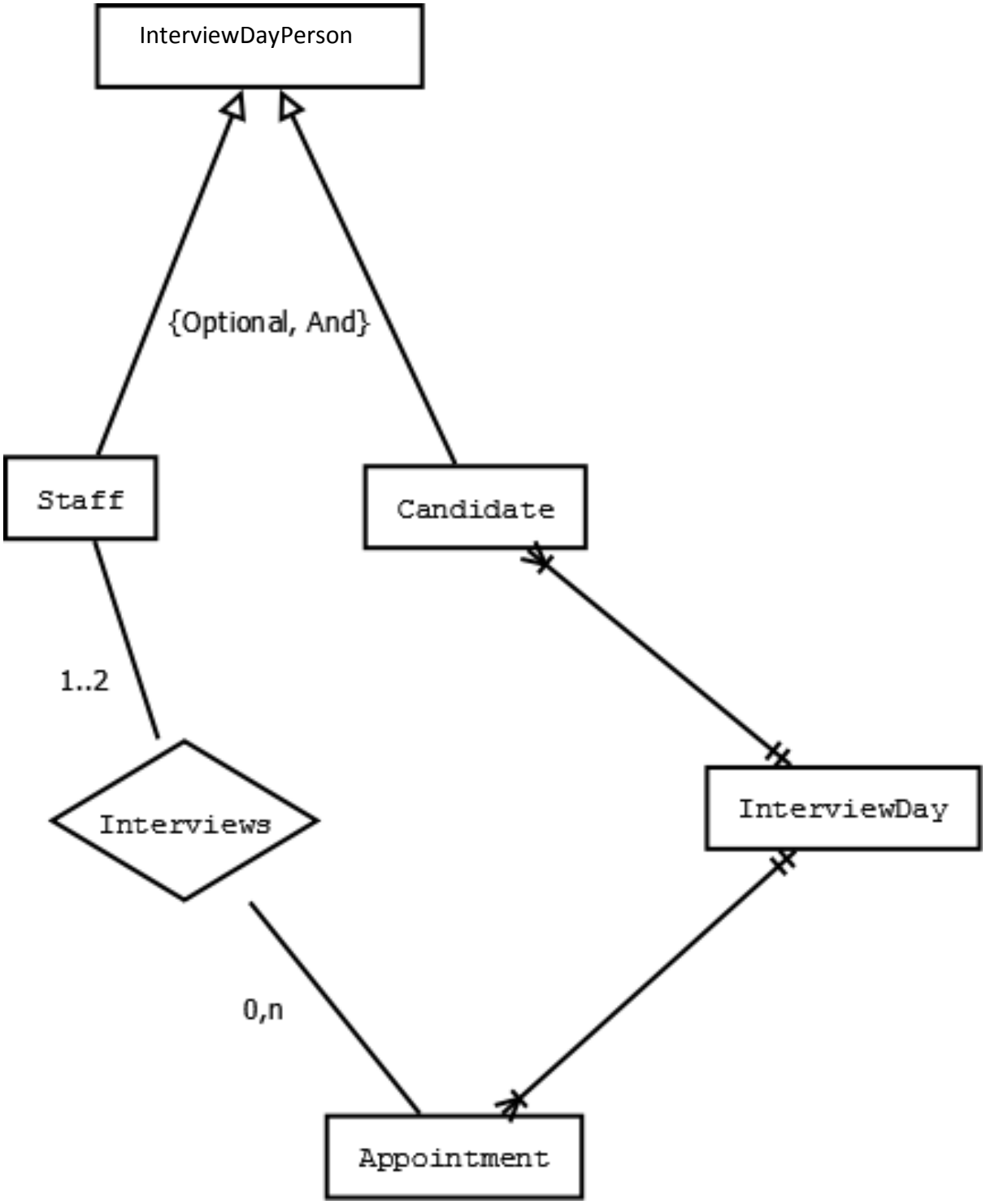

Figure 1: Initial Data Model

Document 2

SomeCity University Interview Days 2014:
Staff Interview Rota

	jan		feb			mar		
	15	22	5	12	26	5	19	26
A Smith	**√√**			**√√**	**√√**			
B Jones	√		√		√		√	
C White		**√√**		√		**√√**		√
D Brown			√		√			√
E Black		√				√	√	
F Smith	√					√		**√√**
G Green		√	**√√**	√			**√√**	

Double Ticks indicate the member of staff in-charge of that interview day.

Document 3

Interview Day Date: 15 Jan'14 Interviews

Location: it05

	Interviewer: **A.Smith**	Interviewer: **B.Jones**
Time: 13.30	Candidate: T.Till	Candidate: J.Jem
Time: 13.50	Candidate: F.Froom	Candidate: P.Pugh
Time: 14.10	Candidate: G.Gates	Candidate: L.Low

Location: g05

	Interviewer: **F.Smith**	
Time: 13.30	Candidate: C.Childs	
Time: 13.50	Candidate: V.Singh	

Document 4

Sample Staff Information

Dr A.Smith

a.smith@someCity.ac.uk

office it01

Document 5

Sample Candidate Information

UCAS number: 123456

Miss J.Jem

jemmy@hotmail.com

12 high street, Hartlepool

Document 6

Sample Degree Information

BSc Computer Science

Offer: type 1

FnD Computer Science

Offer: type 3

Document 7

Offer Type Information

Type	Points
1	280
2	240
3	160

Document 8

Prototype Application Requirements

1. Staff Management
    a. Add new staff
    b. Edit existing staff details
    c. Delete existing staff

Document 9

Answers to frequently asked questions. This list will be maintained on blackboard, you should check on a regular basis for updates.

Must the offer match the type associated with the degree?

- No, an offer can be made with any number of points – the type level is the default guide.

Can a candidate be made multiple offers?

- Yes, a candidate could be made one or several offers.

You should be working on the ICA in tutorial time from 3rd – 12th December 2013.

You should NOT be working on the ICA in tutorial time from 7th January 2014 onwards. These slots are for working on the tutorial activities set by Dr J.J. Longstaff that will lead towards phase 2 of the ICA.

Phase 2:

Preparatory work. Run, explore and understand the Demo4 and Demo7 Mendix projects. Explore the AdventureWorks2012 and AdventureWorksDW2012 SQL Server databases, particularly the tables in the Internet Sales database diagram of AdventureWorksDW2012. Complete the Analysis Services Tutorial and the Basic Data Mining Tutorial on the Microsoft MSDN site, as follows:

Analysis Services Tutorial
http://msdn.microsoft.com/en-us/library/ms170208(v=sql.110).aspx Lessons 1, 2 and 3.

Basic Data Mining Tutorial
http://msdn.microsoft.com/en-us/library/ms167167.aspx Lessons 1,2,3 and 4,
http://msdn.microsoft.com/en-us/library/ms174907 (Singleton Queries).

(You may need to create a Login/User for Analysis Services for AdventueWorksDW2012 if these do not already exist – see the module Blackboard site for directions.)

Task 6

1. Develop a Mendix domain model which directly corresponds to the tables and relationships in the Internet Sales database diagram of AdventureWorksDW2012. Include a Mendix entity for every table in the Internet Sales diagram, and an attribute for each table column except Foreign Key columns. (Use UML associations in your Mendix domain model, and not Primary Key – Foreign Key relational data model relationships.) Take screenshot(s) of your Mendix domain model, and briefly explain what they show.

   (5 marks)

2. For each entity in your Mendix domain model, state whether that entity is strictly in BCNF, giving a brief justification.

   (4 marks)

3. Develop a Mendix form which will display the data displayed by the IS_for_logged_in_Customer form from Demo7. This form will display exactly the same columns and data as does the form in Demo7, but will retrieve that data from the Mendix entities in your domain model developed in part 1 above. (You will need to build extra forms to enter test data into your new Mendix entities.) Take screenshots of your form design, and of your working forms displaying test data. Briefly explain the contents of your screenshots.

   (6 marks)

(15 marks)

Task 7

1. Using SQL Server 2012 Data Tools, and the AdventureWorksDW2012 database, develop an OLAP cube which is based on the following tables.

    - DimCustomer
    - DimDate
    - DimGeography
    - DimProduct
    - DimProductSubCategory
    - DimProductCatgory
    - FactInternetSales
    - FactInternetSalesReason
    - DimSalesReason

    Develop appropriate dimensions and measures which will allow at least the functionality in the Analysis Services Tutorial (Lessons 1, 2 and 3) to be applied, plus additional queries for ProductCategory and ProductSubcategory data. Take screenshots of your dimension and cube structures, and briefly explain what the screenshots show.

    (5 marks)

2. Explain the benefits of user-defined attribute hierarchies for dimensions. Define TWO hierarchies for the Product dimension which involve ProductCategory and ProductSubCategory, and perform several queries on your cube using the DataTools Cube Browser which use your hierarchies to their best advantage. Take screenshots of your dimension structure, and Cube Browser queries, and briefly describe what they show.

    (5 marks)

3. Discuss the purpose of the FactInternetSalesReason and the DimSalesReason tables in the AdventureWorksDW2012 database. Comment on the storage and indexing structures for these tables. How you expect these tables to be used in a Data Tools multidimensional project?

    (5 marks)

(15 marks)

Task 8

Using Data Tools data mining facility, create the **Decision Tree** and **Clustering** mining models by following the Basic Data Mining Tutorial lessons listed above.

1. Explore the two mining models to investigate the likelihood of a person aged about 50, with 3 children at home, and owning 2 cars, buying a bike. Take two or three screenshots for each model which support your conclusions, and briefly describe the contents of your screenshots.

   (6 marks)

2. Use the mining model prediction facility (Singleton Query) to predict the probability of a person aged exactly 50 with 3 children at home and owning 2 cars buying a bike, for each of the two models. Take screenshots of your probability predictions, giving brief descriptions of what they show.

   (2 marks)

3. BRIEFLY discuss which of these two models was more suited to the explorations and predictions you carried out in parts 1 and 2, stating which model is likely to have given the more accurate results.

   (4 marks)

(12 marks)

Task 9

BRIEFLY discuss (500 words minimum) how the tools and techniques of Big Data analytics might generally be used to extend the OLAP and Data Mining analyses for the AdventureWorks application which is the subject of Tasks 7 and 8 above?

(8 marks)

You should be working on tasks 6-9 in from March onwards.

Deliverable:

You must submit only a **single pdf** (portable document format) file that documents all your answers.

No other document format will be opened or marked. **Do not submit** your database file. **Do not submit** your Visual Studio solution. **Do not submit** Mendix files. **Do not submit** AdventureWorks database files. **Do not submit** multiple pdf files. **Do not submit** a word document (word can be used to create your document but you must save the final version for submission as a pdf file).

Use the guide on blackboard to help you format your submission document. If you have any queries contact your local tutor.

Your attention is drawn to the university's plagiarism regulations; do not copy or allow anyone to copy your work. This is an individual ICA; do not work in partnership with other students.