



MONASH University

Office Use Only

--	--	--

**Semester Two 2022
Examination Period**

Faculty of Information Technology

Sample Exam

EXAM CODES:

FIT3003

TITLE OF PAPER:

Business Intelligence and Data Warehousing - SAMPLE 1

EXAM DURATION:

2 hours 10 minutes or 130 minutes

THIS PAPER IS FOR STUDENTS STUDYING AT: (tick where applicable)

- ☐ Caulfield ☒ Clayton ☐ Parkville ☐ Peninsula
☐ Monash Extension ☐ Off Campus Learning ☒ Malaysia ☐ Sth Africa
☐ Other (specify)

During an exam, you must not have in your possession any item/material that has not been authorised for your exam. This includes books, notes, paper, electronic device/s, mobile phone, smart watch/device, calculator, pencil case, or writing on any part of your body. Any authorised items are listed below. Items/materials on your desk, chair, in your clothing or otherwise on your person will be deemed to be in your possession.

No examination materials are to be removed from the room. This includes retaining, copying, memorising or noting down content of exam material for personal use or to share with any other person by any means following your exam.

Failure to comply with the above instructions, or attempting to cheat or cheating in an exam is a discipline offence under Part 7 of the Monash University (Council) Regulations, or a breach of instructions under Part 3 of the Monash University (Academic Board) Regulations.

AUTHORISED MATERIALS

OPEN BOOK

☐ YES

☒ NO

CALCULATORS

☐ YES

☒ NO

SPECIFICALLY PERMITTED ITEMS

☐ YES

☒ NO

if yes, items permitted are:

Candidates must complete this section if required to write answers within this paper

STUDENT ID: _ _ _ _ _

DESK NUMBER: _ _ _ _ _

Question 1:

Monash International would like to analyse their policy in regard to English requirement for admission into a course. Monash International has the following data:

Table: Student_IELTS

Student ID	Student Name	Listening	Reading	Writing	Speaking	Overall
228493	Sooying Tan	6.5	6.5	6.0	7.0	6.5
229094	Xuebing Lu	5.5	5.5	5.5	5.5	5.5
231289	Amandh Kumar	6.0	7.0	6.0	7.0	6.5
234354	Agus Hidayat	5.5	6.0	6.0	6.5	6.0
234355	Budi Rahayu	7.0	7.0	7.0	7.0	7.0
...						
...						

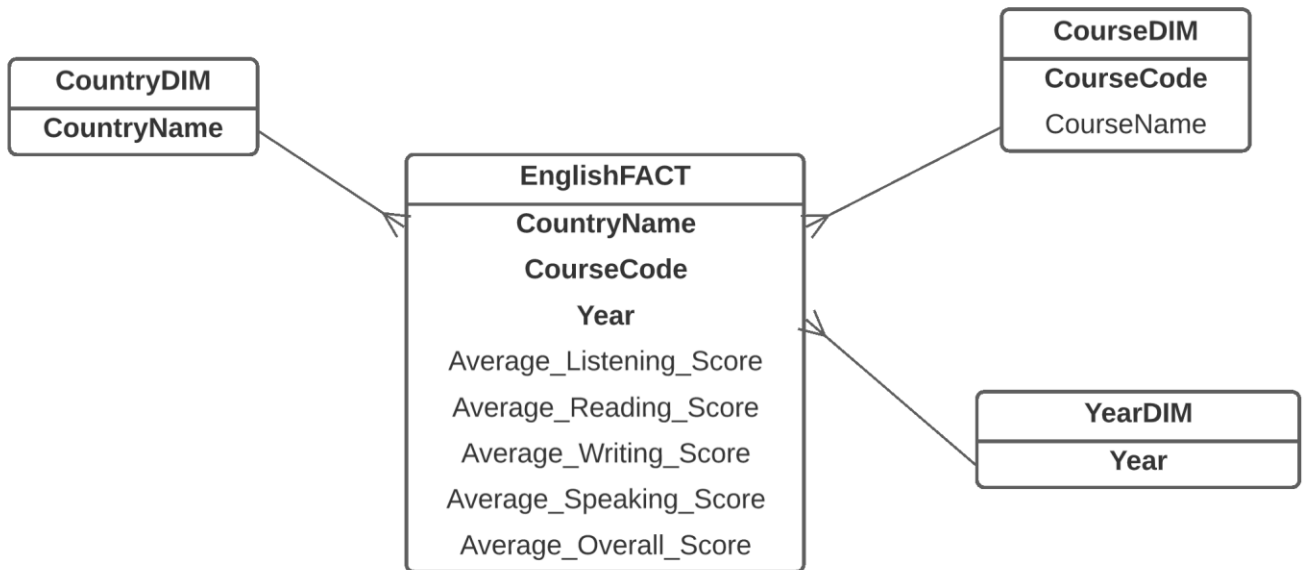
Table: Student_Course

Student ID	Student Name	Course	StarYear
228493	Sooying Tan	MBIS	2013
229094	Xuebing Lu	MBIS	2013
231289	Amandh Kumar	MIT	2013
234354	Agus Hidayat	MIT	2013
234355	Budi Rahayu	MIT	2013
...			
...			

Table: Student

Student ID	Student Name	Address	Suburb	Phone Number	Country
228493	Sooying Tan				Singapore
229094	Xuebing Lu				China
231289	Amandh Kumar				India
234354	Agus Hidayat				Indonesia
234355	Budi Rahayu				Indonesia
...					
...					

A data warehouse based on the above data has been created. A star schema is shown as follows:



Questions:

- The above star schema will not produce a correct analysis of the fact measures. Explain why. Explain your answer using more concrete examples or data. **(5 marks)**
- How do you correct this problem by changing the fact measures of the above star schema? Explain your solution using more concrete examples or data. **(5 marks)**

Write your answer here:

Continue your answer here:

Question 2:

Data cleaning is an important part in building a clean and correct data warehouse. Data cleaning is often needed, because there are mistakes and inconsistencies in the operational database. Before data cleaning is done, we need to do data exploration on the operational database in order to find out if there are any mistakes and inconsistencies in the operational database.

The following are the four tables in the operational database:

```
SQL> desc dw.uselog;
```

Name	Null?	Type
LOG_DATE	NOT NULL	DATE
LOG_TIME	NOT NULL	DATE
STUDENT_ID	NOT NULL	CHAR(11)
ACT		CHAR(1)

```
SQL> desc dw.student;
```

Name	Null?	Type
SEX		VARCHAR2(2)
FULL_PART		VARCHAR2(2)
TYPE		VARCHAR2(4)
CLASS_ID		VARCHAR2(6)
MAJOR_CODE		VARCHAR2(8)
STUDENT_ID	NOT NULL	CHAR(11)

```
SQL> desc dw.class;
```

Name	Null?	Type
CLASS_DESCRIPTION		CHAR(50)
CLASS_ID		VARCHAR2(6)

```
SQL> desc dw.major;
```

Name	Null?	Type
MAJOR_NAME		CHAR(35)
MAJOR_CODE		VARCHAR2(8)

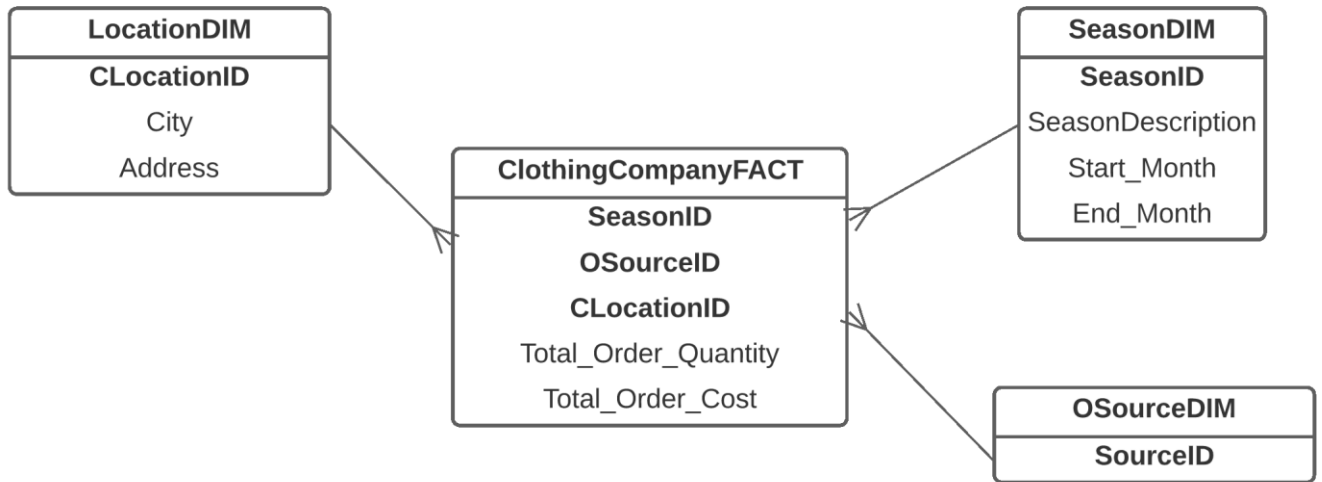
Questions:

- Write the SQL command to find out if there are duplicate student records (**5 marks**)
- Write the SQL command to find out if there are records in dw.uselog whereby the Student_ID exists in dw.uselog actually do not exist in dw.student (**5 marks**)

Write your answers here:

Question 3:

Given the following schema:



The tables (e.g. ClothingCompany fact and the three dimensions) have been created and populated with an adequate number of records.

The table names and attributes are shown in the above star schema. In the Fact table, the total order quantity and total order cost attributes are included.

Write the SQL for the following advanced OLAP queries:

- Perform a **CUBE** operation (use all dimensions). Display each TotalOrderCost and the subtotals. **(3 marks)**
- Like question (a) above, but now perform a **ROLLUP** operation. **(3 marks)**
- Perform a **CUMMULATIVE SUM** of the TotalOrderCost of all WEBSITE orders (use all dimensions). **(4 marks)**
- Like question (c) above, perform a CUMMULATIVE SUM of the TotalOrderCost but **PARTITIONED** based on the OSourceID, that is one partition for Phone orders, one partition for Fax orders, and one partition for Website orders. **(4 marks)**
- Show the total order costs of each source order, and **RANK** them. **(3 marks)**
- Display the source order that generates the highest total order cost. **(3 marks)**

Write your answer here:

Continue your answers here:

Continue your answer here:

Continue your answer here:

Question 4:

There is a toll way (or toll road) in a metropolitan city (such as CityLink or EastLink in Melbourne, or any similar toll roads in other major cities in the world). This toll way has a number of gates, where the motorist needs to pay. Every time a motorist passes through this toll gate, the registration number of the vehicle, vehicle type (e.g. car, bus, truck, etc), amount paid, and time, are recorded in the operational database.

A data warehouse needs to be built, for analysing the *revenue* from the toll payments. The management would like to drill down this revenue based on the *tollgate* (there is a number of toll gates along the toll way), *day of week* (e.g. weekdays, weekends), and *time period of a day* (e.g. peak hours, non-peak hours, late nights).

You are required to draw **three levels of star schemas** showing three different levels of aggregation for the above data warehouse. You also need to explain each of the three star schemas, by contrasting the level of aggregation. Level-0 star schema contains the most detailed data, whereas level-2 star schema is the highly aggregated (e.g. containing highly aggregated data).

Questions:

- (a) Draw a level-2 star schema and explain why it is a level-2 schema **(6 marks)**
- (b) Draw a level-1 star schema and explain why it is a level-1 schema. You may want to add a new dimension, called *vehicle* (e.g. cars, trucks, busses, etc). You need to also explain the difference between level-1 and level-2 schemas. **(6 marks)**
- (c) Draw a level-0 star schema and explain why it is a level-0 schema. You also need to explain the difference between level-1 and level 0 schemas. **(8 marks)**

Write your answer here:

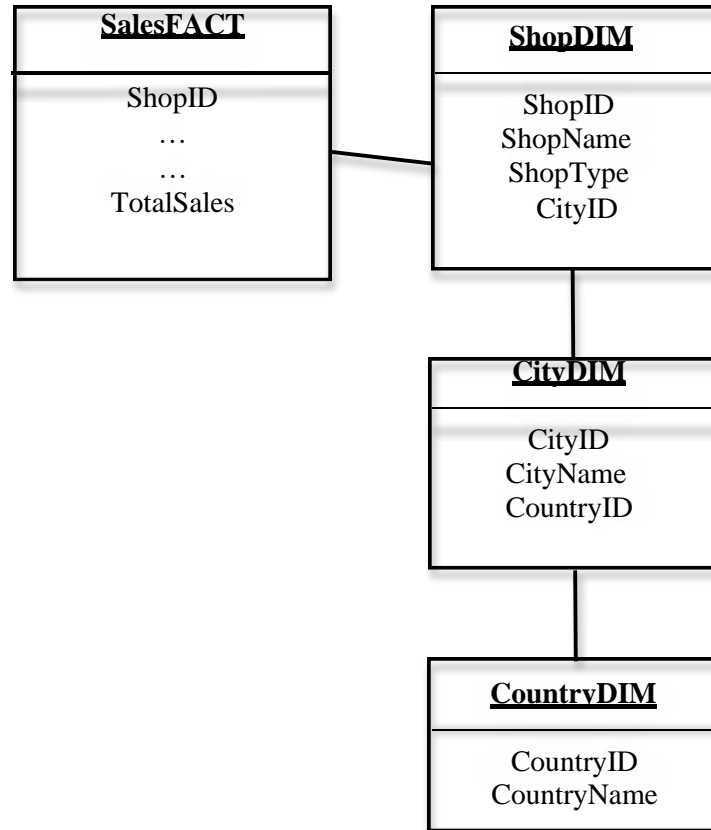
Continue your answer here:

Continue your answer here:

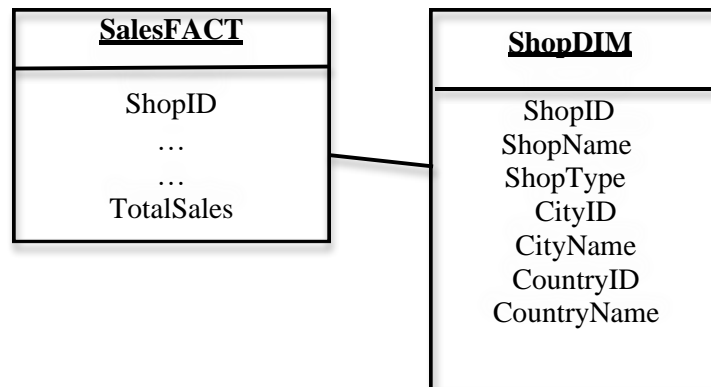
Question 5:

Consider the following star schemas. Star Schema-1 contains a hierarchy in the dimension, whereas Star Schema-2 collapses the hierarchy into one dimension.

Star Schema-1 (with hierarchy):



Star Schema-2 (without hierarchy):



Questions:

- Draw sample table contents of the fact and dimension tables of the two star schemas. **(6 marks)**
- Compare and contrast the two star schemas using the sample tables in question (a) above. Explain the pros and cons of each star schema. **(4 marks)**

Write your answers here:

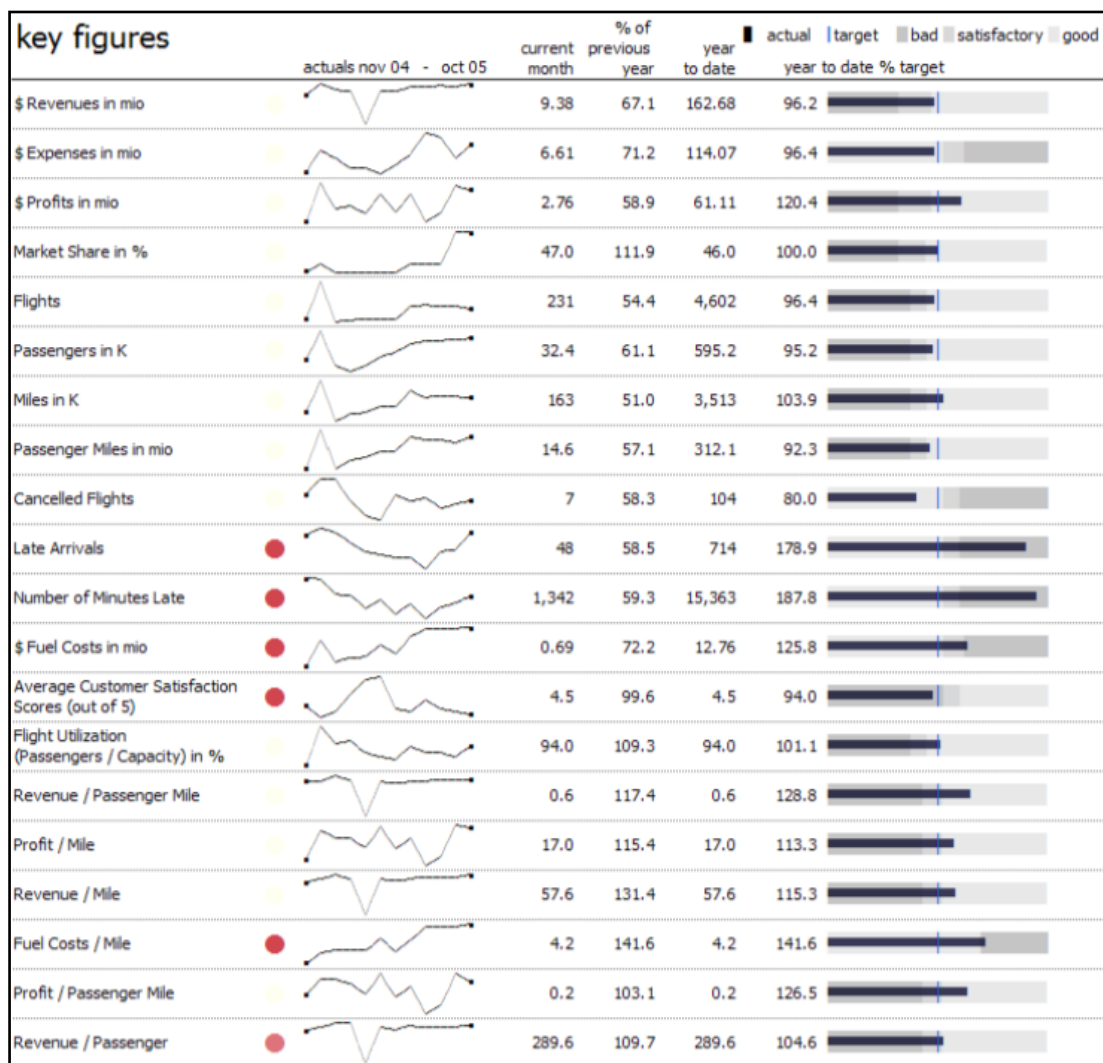
Continue your answers here:

Question 6:

- a) Dashboards provide overview of a system in a visual and often interactive display, with the most important information about the system. It can be customised for specific groups of users, or for different purposes. There are basically three types of dashboard, namely strategic, analytic and operational.

The following figure shows a part of the dashboard in a BI system for an airline company. What type of dashboard is this? What type of users would use this dashboard? Explain your answer.

[5 marks]



- b) Compared to dashboard, reports provide minimal interactivity and analysis. A good report should require minimal cognitive effort in order to understand the contents. There are a few types of navigational methods when viewing reports in BI systems.

Explain drill up and drill-through navigational methods using examples.

[5 marks]

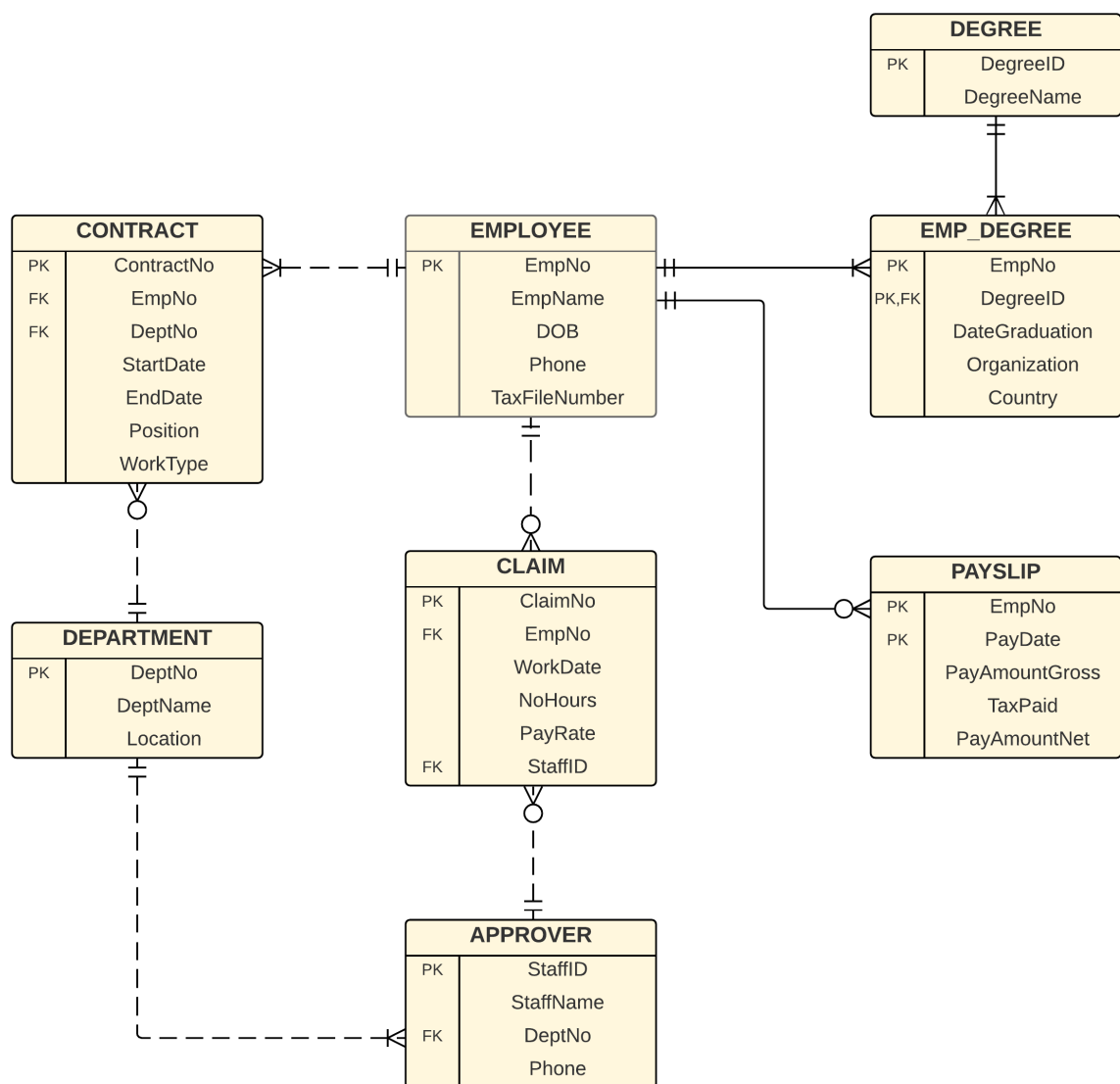
Write your answers here:

Question 7

Monash University employs its students to do various jobs, such as tutoring, programming, etc. These jobs are called sessional jobs. For each sessional job, students need to sign a contract. For example, to do tutoring (one or more units), the student will sign a contract with Monash for one semester.

These sessional workers (e.g. sessional tutors) need to claim their work hours every week. This claim will need to be approved by a designated person in the faculty. Every fortnight, the sessional workers will get their pay.

The operational database, which keeps track of this system, is shown by the following E/R diagram.



You are required to build a data warehouse to analyse the following:

- The total number of contracts made every year.
- The total payment made to each employee in 2020.
- The yearly number of contracts made in each department.

Question:

(a) Draw the star schema for this system. **(10 marks)**

(b) Write the SQL queries to create the fact and dimension tables. **(10 marks)**

Write your answers here:

THE END