

Monash University
Semester Examination Period
Faculty of Information Technology

EXAM CODES: Sample-1

TITLE OF PAPER: Sample-1

EXAM DURATION:

READING TIME:

Sample Exam

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

- | | | | | |
|---|---|------------------------------------|---|--|
| <input type="checkbox"/> Berwick | <input checked="" type="checkbox"/> Clayton | <input type="checkbox"/> Malaysia | <input checked="" type="checkbox"/> Off Campus Learning | <input type="checkbox"/> Open Learning |
| <input checked="" type="checkbox"/> Caulfield | <input type="checkbox"/> Gippsland | <input type="checkbox"/> Peninsula | <input type="checkbox"/> Enhancement Studies | <input type="checkbox"/> Sth Africa |
| <input type="checkbox"/> Pharmacy | <input type="checkbox"/> Other (specify) | | | |

During an exam, you must not have in your possession, a book, notes, paper, calculator, pencil case, mobile phone or other material/item which has not been authorised for the exam or specifically permitted as noted below. Any material or item on your desk, chair or person will be deemed to be in your possession. You are reminded that possession of unauthorised materials in an exam is a discipline offence under Monash Statute 4.1.

No examination papers are to be removed from the room.

Attempt all questions. All answers must be printed neatly on this paper. Answer questions with concisely expressed factual information. The backs of pages in this paper may be used for any rough work. **Any material written on the backs of pages will not normally be corrected.** If an answer needs to overflow from its designated answer space to a blank page, clearly indicate that this is the case and that the material on the blank page is for correction. This paper must be handed up at the end of the examination, even if no questions are attempted. **There are 8 questions, each of which is worth 10 marks. The total is 80 marks.** This exam counts as 60% of the final assessment for the unit.

AUTHORISED MATERIALS

CALCULATORS

☐ YES

☒ NO

OPEN BOOK

☐ YES

☒ NO

SPECIFICALLY PERMITTED ITEMS

☐ YES

☒ NO

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

STUDENT ID

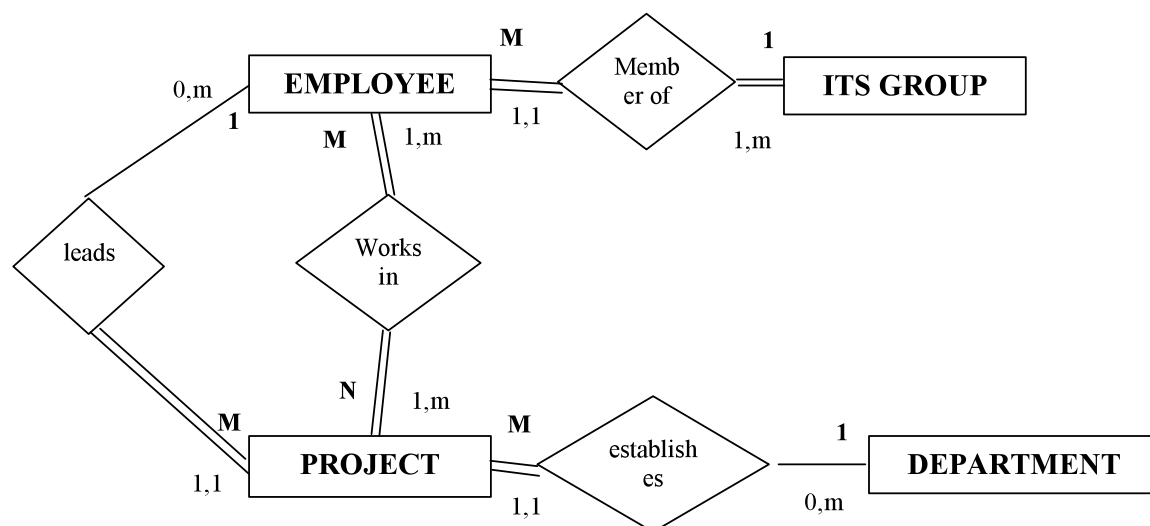
DESK NUMBER

Question 1:

The Information Technology Services (ITS) department at Monash University needs a database to keep track of ITS staff members and the projects they are working on. The system must satisfy the following requirements:

- For each ITS employee, list the employee number, employee name, job title (position), and the number and name of the ITS group in which he or she works. In addition, for each project to which the employee is assigned, list the project number and name, the percent of the employee's time assignment to the project, and the total number of hours the employee has worked on the project so far.
- For each ITS project, list the project number and name (description), the name of the department that requested the project, the name of the contact person in that requesting department, the project type (maintenance, database, etc.), project status, start date, end date, total budgeted person-hours, total budgeted dollars, and the name of the ITS employee serving as project leader.

The E/R diagram for the above process is given as follows:



The system is now operating with the following tables:

EMPLOYEE (emp#, empname, position, ITSgroup#)

ITS-GROUP (itsgroup#, itsgroupname)

DEPARTMENT (dept#, deptname)

PROJECT (proj#, projname, contactperson, projtype, projstatus, start_date, end_date, budgperhours, budgtotals\$, emp#, dept#)

EMP-PROJ (emp#, proj#, %timeassignment, projemptotalhrs)

Monash University management would like to analyse the performance of the ITS department, and in order to assist this process, you are asked to develop a data warehouse for analysis purposes. The analysis is needed for identifying at least the following questions:

- What is the total number of projects that are long duration?
- What is the total budget for a certain project type?
- What is an average budget cost per hour of a certain department?

Based on the above requirements, the fact measures that the management is interested in are total budgets, average budget per hours, and total projects; and the dimensions are project type, department, and project duration. Assuming that the management classifies projects into short term (less than 10 days), medium term (between 10 and 30 days), and long term (more than 30 days).

Questions:

- a. Draw a star schema containing dimensions and fact, together with their attributes.
- b. Define the SQL statements for the implementation of the star schema.
- c. Write the SQL statements to answer the above three query requirements.

Write your answer here:

Continue your answer here:

Continue your answer here:

Question 2:

Explain the differences between **RANK()**, **DENSE_RANK()**, and **ROW_NUMBER()** in OLAP SQL.

Write your answer here:

Continue your answers here:

Question 3:

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
...						
...						

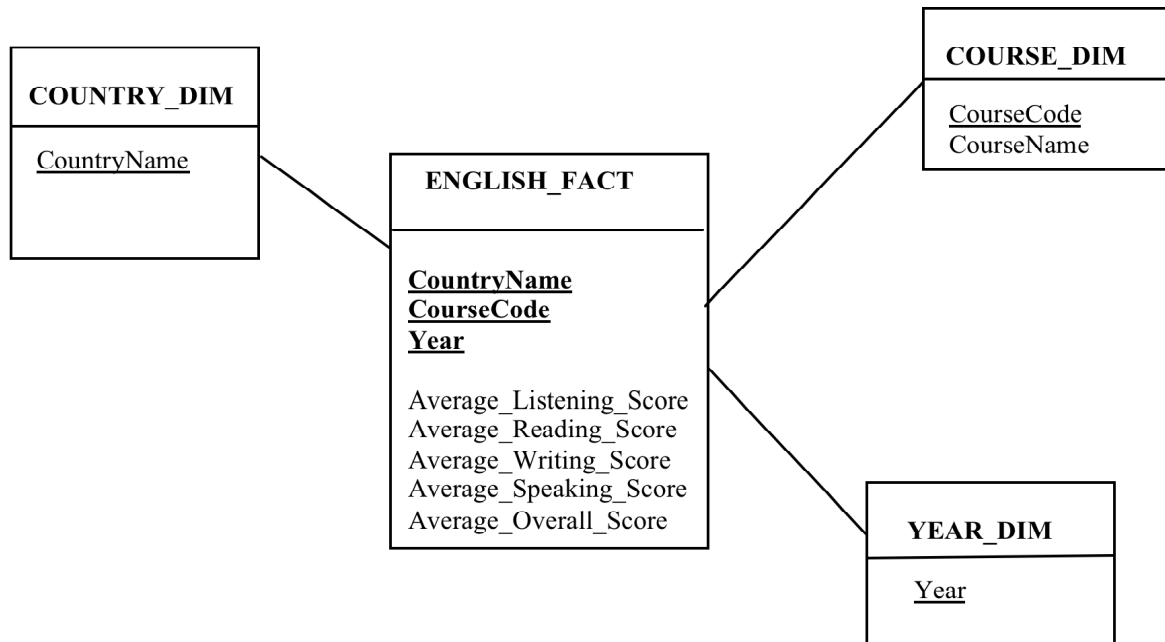
Table: Student_Course

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.
- 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.

Write 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
- (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.
- (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.

Write your answer here:

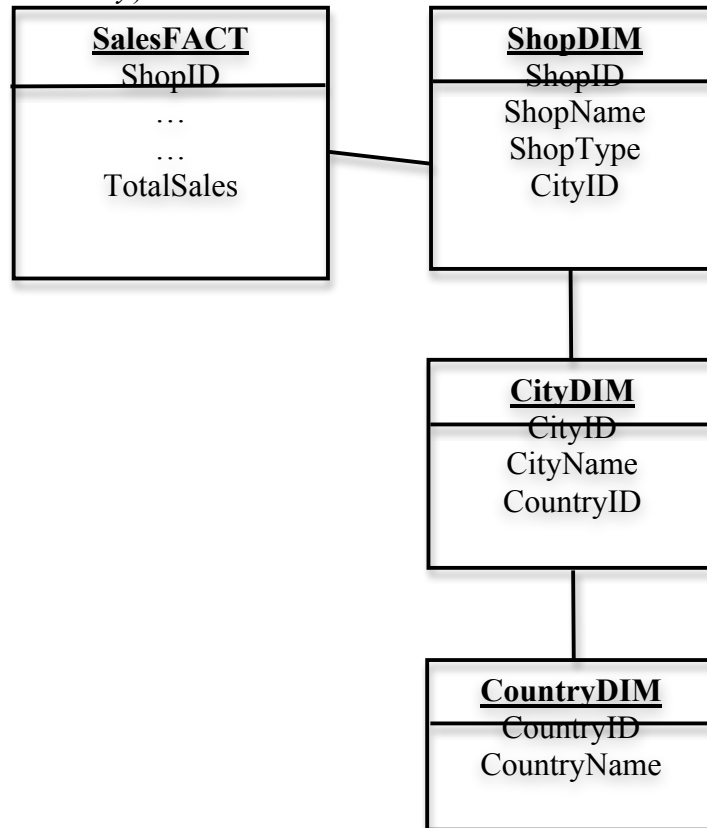
Continue your answer here:

Continue your answer here:

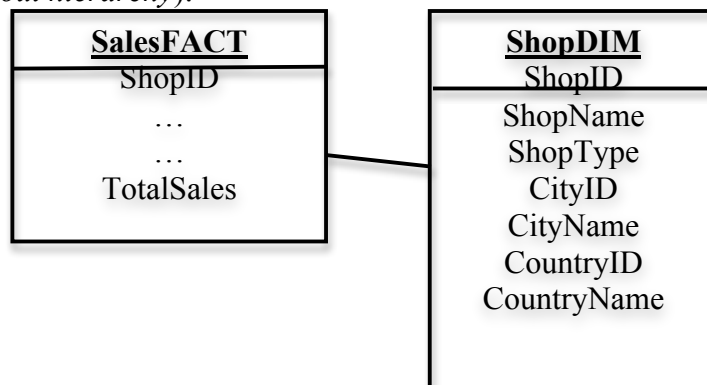
Question 5:

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.
- Compare and contrast the two star schemas using the sample tables in question (a) above. Explain the pros and cons of each star schema.

Write your answers here:

Continue your answers here:

Question 6:

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
- 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

Write your answers here:

Question 7:

An established real estate agent in Melbourne has started their business many years ago and has implemented a very simple database system. The simple database system consists of one large table listed below.

Table Name: PROPERTY	
Field Name	Description
Key	Unique key
Date_offered	Date property offered to the public
Summary	Short description of the property
Adtext	Longer description of the property
Url	The URL of the advertisement
Address	Property address
Suburb	Property suburb name
Postcode	Property postcode
Longitude	Longitude of address
Latitude	Latitude of address
Category	'Residential' or 'Commercial'
Zoning	Commercial Zoning Type
Property_type	Residential Property Type: 'House', 'apartment', or 'lot'
Houseprice	Price of property
Num_bedrooms	Number of bedrooms
Lot_size	Size of the lot
Heating	'ducted', 'gas', 'open fireplace' or 'wood'
Garage	Type of garage
Ensuite	'yes' or 'no'
Balcony	'yes' or 'no'
Pool	'yes', 'no'
Tennis_court	'yes', 'no'
Spa	'yes', 'no'
Aspect_facing	'north', 'south', 'east', or 'west'
School_distance	Distance to nearest school – in km
Shops_distance	Distance to nearest shops – in km
Train_distance	Distance to nearest train station – in km
Bus_distance	Distance to nearest bus stop – in km
Hospital_distance	Distance to nearest hospital – in km
Major_road_distance	Distance to nearest major road – in km

Based on this information, the manager of the real estate agent requires a data warehouse for analysis purposes. The final star schema may include a junk dimension.

Questions:

- What is a junk dimension? Explain!!! Also, use a sample data to illustrate a junk dimension
- Design two versions of star schema for the above case study; one without a junk dimension, and the other with a junk dimension. Compare and contrast these two schemas; focusing on the junk dimension only

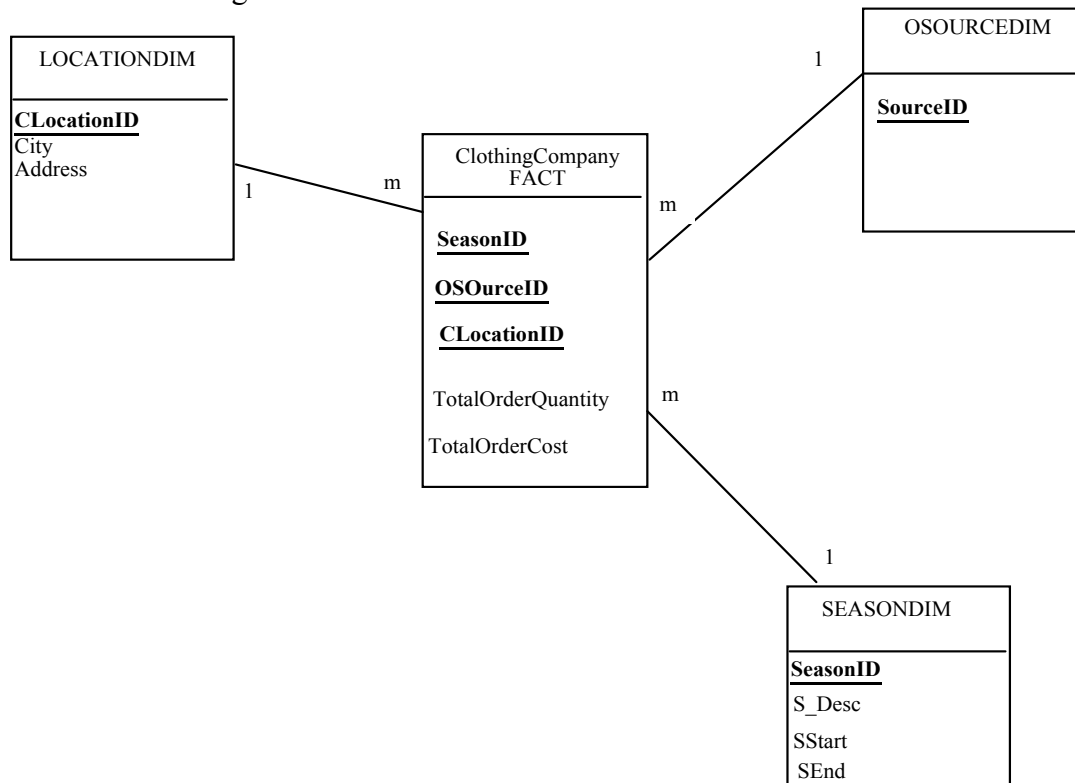
Write your answer here:

Continue your answer here:

Continue your answer here:

Question 8:

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.
- Like question (a) above, but now perform a **ROLLUP** operation.
- Perform a **CUMMULATIVE SUM** of the TotalOrderCost of all WEBSITE orders (use all dimensions).
- 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.
- Show the total order costs of each source order, and **RANK** them.
- Display the source order that generates the highest total order cost.

Write your answer here:

Continue your answer here:

Continue your answer here:

Continue your answer here:

THE END