FIT3003 – Business Intelligence and Data Warehousing

Semester 2, 2019

SAMPLE Exam Paper

Question 1 [20 marks]

a) What are the two types of information systems in enterprise information technology? Given two examples for each type.

[6 marks]

Answer:
1. Transaction processing systems - Systems for collecting and recording information about business activities [2 marks]
   - E.g. ERP systems, payroll systems, point-of-sale systems, inventory management systems, etc [1 mark for each correct example]
2. Systems for *collating and analysing* information (typically collected from transaction processing systems) to support organisational decision-making. [2 marks]
   - E.g. Decision support system, business intelligence system [1 mark for each correct example]

b) There are three types of dashboard for business intelligence (BI) systems, namely strategic, analytic and operational. Describe these three types of dashboards.

[6 marks]

Answer: [2 marks for each correct description]
- Strategic – Used by managers to monitor business health and progress towards meeting strategic objectives – High-level measures of past and forecast performance
- Analytic – Used by analysts as a way to gain access to a broad range of more detailed information
- Operational – Used to monitor the operation of a factory or service – Display of information is real-time or near-real-time – Not only display measures but also displays warnings and alerts

c) There are a few main navigational methods that should be supported by BI systems. Explain drill-up and drill through. Give example for each of these two navigational methods.

[8 marks]

Answer:

Drill-up [2 marks correct explanation and 2 marks for correct example]

- Navigating to a report with less detail following the levels of a pre-defined dimension hierarchy.
- Example: Data in the following sequence: Daily sales ➜ Monthly sales ➜ Quarterly sales ➜ Yearly sales

Drill-through [2 marks correct explanation and 2 marks for correct example]

- Navigating from a summary report to report that shows the detailed transactions that made up the summary.
- Example: From a report that shows total sales by region to another report that shows the sales by invoice in that particular region.

Question 2 [20 marks]

a) The table below shows a report generated from a business intelligence system for XYZ University.

<table>
<tr><th colspan="7">Student Intake for Undergraduate (UG) and Postgraduate (PG) Courses
Semester 1 2019</th></tr>
<tr><th rowspan="3">Country of Origin</th><th colspan="2">South Campus</th><th colspan="3">North Campus</th><th rowspan="3">Total</th></tr>
<tr><th>UG Program</th><th>PG by Research</th><th>UG Program</th><th>PG by Coursework</th><th>PG by Research</th></tr>
<tr><td colspan="5">Asia</td></tr>
<tr><td>China</td><td>100</td><td>20</td><td>60</td><td>130</td><td>20</td><td>330</td></tr>
<tr><td>Malaysia</td><td>50</td><td>5</td><td>30</td><td>15</td><td>10</td><td>110</td></tr>
<tr><td>Indonesia</td><td>150</td><td>10</td><td>70</td><td>50</td><td>20</td><td>300</td></tr>
<tr><td>Subtotal</td><td>300</td><td>35</td><td>160</td><td>195</td><td>50</td><td>740</td></tr>
<tr><td colspan="7"></td></tr>
<tr><td colspan="5">Europe</td></tr>
<tr><td>France</td><td>70</td><td>10</td><td>100</td><td>30</td><td>10</td><td>220</td></tr>
<tr><td>Italy</td><td>100</td><td>2</td><td>120</td><td>40</td><td>5</td><td>267</td></tr>
<tr><td>Subtotal</td><td>170</td><td>12</td><td>220</td><td>70</td><td>15</td><td>487</td></tr>
<tr><td>Grant Total</td><td>470</td><td>47</td><td>380</td><td>265</td><td>65</td><td>1227</td></tr>
</table>

Draw the Thomsen Diagram that represents the dimension(s) and fact(s) of the data in the report above.

[8 marks]

Answers: [2 marks for each correct dimension and fact, with correct members/hierarchy] – STUDENTS TO DRAW THE DIAGRAM

Dimensions
- Time: Semester -> Year
- Region: Country -> Region
- Course: Course -> Campus

Fact
- Intake: Total Intake

b) Customer Relationship Management (CRM) system is an example that shows 3NF can be combined with star schema to develop a data warehouse.

Assume that you are asked to design the data warehouse for premium clothing store. Its customer has a set of basic attributes, changing circumstances and behaviors to be stored in the data warehouse.

Identify at least two suitable changing circumstances, two behaviors as well as two derived segments for the data warehouse.

Draw the logical schema for the resulting data warehouse.

[12 marks]

Expected answer:
Customer dimension in the DW with a set of basic attributes [1 mark]

Changing circumstances [2 marks, 1 for each correct example]
    e.g. address, marital status, employment type

Behaviors to be modeled as fact tables [4 marks, 2 marks for each correct example with correct PK/FK of fact table]
    e.g. purchase transaction, lay-by transaction, subscription to loyalty program etc

Derived segments [2 marks, 1 for each correct example]
    e.g. loyalty program type, age category, frequent buyer etc

0.5 mark for every correct relationship shown in DW schema, total 3 marks

Question 3 [20 marks]

a) Assume that you have a table with 100 rows of records with the following attributes:

```
TableA (keyAttribute, attribute1, attribute2)
```

Explain how linear search and binary search take place on *TableA* for the following query:

```
SELECT *
FROM TableA
WHERE keyAttribute = 125;
```

[4 marks]

Answer:

In linear search, every record in the table is retrieved, and tested whether its attribute values satisfy the selection condition. [2 marks]

If the selection condition involves an equality comparison on a key attribute on which the file is ordered, binary search can be used. It is more efficient than linear search. [2 marks]

b) The table below shows a fact table named *Booking_Fact* for a hotel to analyze the booking of its rooms across different seasons and different branches. Answer question (i) to (iii) based on the table below:

BRANCH	ROOM_TYPE	SEASON	TOTAL_ROOMS_BOOKED
100	Single	Spring	100
100	Deluxe	Summer	200
100	Suite	Spring	80
200	Single	Summer	500
200	Single	Autumn	400
300	Deluxe	Autumn	200
300	Suite	Winter	300
100	Deluxe	Winter	120
200	Single	Spring	350

(i) Write a SQL command to generate the following result based on the table above.

[6 marks]

Branch	Room_Type	SUM(Total_Rooms_booked)
100	Suite	80
100	Deluxe	320
100	Single	100
100	(null)	500
200	Single	1250
200	(null)	1250
300	Suite	300
300	Deluxe	200
300	(null)	500
(null)	(null)	2250

Answer:

SELECT Branch, Room_Type, SUM(Total_Rooms_booked) [3 marks]
FROM Room_Fact [1 mark]
GROUP BY ROLLUP (Branch, Room_Type); [2 marks]

(ii) Write a SQL command to show the ranking of room types in every season. The SQL command should return the following result.

[6 marks]

Season	Room_Type	Total_Booked	Ranked_Room_Type
Autumn	Single	400	1
Autumn	Deluxe	200	2
Spring	Single	450	1
Spring	Suite	80	2
Summer	Single	500	1
Summer	Deluxe	200	2
Winter	Suite	300	1
Winter	Deluxe	120	2

Answer:

SELECT Season, Room_Type, SUM(Total_Rooms_Booked) AS Total_Booked, [2 marks]

RANK() OVER (PARTITION BY Season ORDER BY SUM(Total_Rooms_Booked) DESC) [2 marks]
as Ranked_Room_Type [2 marks]
FROM Room_Fact [1 mark]
GROUP BY Season, Room_Type; [1 mark]

(iii) Modify your SQL command above to show the top room type booked in each season. The SQL command should return the following result.

[4 marks]

Season	Room_Type	Total_Booked	Room_Rank
Autumn	Single	400	1
Spring	Single	450	1
Summer	Single	500	1
Winter	Suite	300	1

Answer:

select * from ([1 mark]
select season, room_type, SUM(Total_Rooms_Booked) as Total_Booked, [1 mark]
rank() over (partition by season order by SUM(Total_Rooms_Booked) desc)
Room_Rank [2 marks]
from room_fact
group by season, room_type)

where Room_Rank < = 1; [1 mark]

Question 4 [10 marks]

List and describe all Data Quality dimensions and give four examples of Data Quality Management tasks.

[10 marks]

Answer:
[0.5 mark for each identified dimension; 3 marks in total for dimensions]
[0.5 mark for each correct description; 3 marks in total for descriptions]

a. Correctness → when the data item captured in the data warehouse reflects the real-world state it is intended to represent
b. Completeness → When all real-world states that one wishes to model are captured in the data warehouse *OR* When all values for a certain data element are recorded (either def is correct)
c. Precision → when a single state of the data warehouse maps onto a single state in the real world
d. Timeliness → When the recorded values are not out of date
e. Usability → The usefulness of data for the particular application at hand
f. Comprehensibility → The ease with which users of the data can comprehend its meaning

Data Quality Management involves:
[1 mark for each example; 4 marks in total for DQM examples]

- Data Information gathering
- Analysis of the impact of data deficiencies
- Ascertain the nature of the data deficiencies
- Determine the appropriate level of data quality
- Choose procedures to ensure data quality

Question 5 [10 marks]

a) When designing charts for business intelligence system, one should avoid chart junk. One of the relevant standards is called data-ink ratio, which was defined by Tufte. What is data-ink ratio? Should it be maximised or minimised? Why?

[5 marks]

Answer:

According to Tufte, data-ink ratio is defined as ink used to display data / total ink to print chart. [1 mark]

Data-ink ratio should be maximised within reason [2 mark].

The reason is because every bit of ink on a graphic requires a reason. And nearly always that reason should be that the ink presents new information. [2 marks]

b) Discuss the concept and the consequences of Pre-attentive processing for Business Intelligence Interface Design.

[5 marks]

Answer:
Performed automatically on the entire visual field detecting basic features of objects in the display. This occurs in "Stage 1" of visual processing. [2 marks]

It is recommended to:
- use features 'seen' during pre-attentive processing: shapes, curves, etc. to communicate the message of the display [1 mark]
- use white-space to group, sort data to easily identify outliers [1 mark]
- avoid 'wasted' use of these features – 3D effects, irrelevant colours [1 mark]

Question 6 [20 marks]

a) List and define the five key phases of a technology's life cycle, also known as hype cycles.

[10 marks]

Answer: [1 mark for each correct phase; 1 mark for each correct description]

The 5 key phases of a technology lifecycles are:

Technology Trigger: A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger significant publicity. Often no usable products exist, and commercial viability is unproven.

Peak of Inflated Expectations: Early publicity produces several success stories — often accompanied by scores of failures. Some companies act; many do not.

Trough of Disillusionment: Interest decreases as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters.

Slope of Enlightenment: More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious.

Plateau of Productivity: Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

c) Discuss the purpose of a dashboard and identify at least 3 (three) things to avoid and 3 (three) things to consider when designing a dashboard.

[10 marks]

Answer:
The purpose of a dashboard is to quickly present relevant data that informs the user about the current status of something. [2 marks]

Things that should be avoided: [1.5 mark for each correct; 4 marks in total]
- In-depth analysis
- Long historical comparisons
- Navigation (tabs)
- Non-critical information

Things that need to be considered: [1.5 mark for each correct; 4 marks in total]
- *Form – Form follows function*
  Paper "one-pager", paper presentation, Excel spreadsheet, Online application, Email or text message, large screen
- *Structure – How will you layout the objects?*
  Flow, relationships, grouping, top-left to bottom-right
- *Design Principles – What will guide your design?*
  Information density, gradual reveal, guide attention, usability
- *Functionality – What can the user do?*
  Drill-down, filters, comparison, alerts, export and print