

The University of Fiji

School of Science and Technology
Department of Information Technology and Computer Science

ITC251 Data Warehousing

Semester 1, 2013

Final Examination

Question Paper

Time Allowed: 3 Hours

Marks 100

Weighting 50%

Instructions:

1. All questions in the Question Paper are compulsory
2. Write all your answers in the Answer Book provided.
3. Any extra sheets used should be labeled with your id-number and attached inside the Answer Book.

Section A – Multiple Choice (15 marks)

Circle the best answer in the Multiple Choice Grid, and attach it inside the Answer Book.

1. According to Bill Inmon, a data warehouse is

- a. A system that retrieves and consolidates data periodically from the source systems into a dimensional or normalized data store.
- b. A subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management's decisions.
- c. A system that extracts, cleans, conforms and delivers source data into a dimensional data store and then supports and implements querying and analysis for the purpose of decision making.
- d. All of the above.

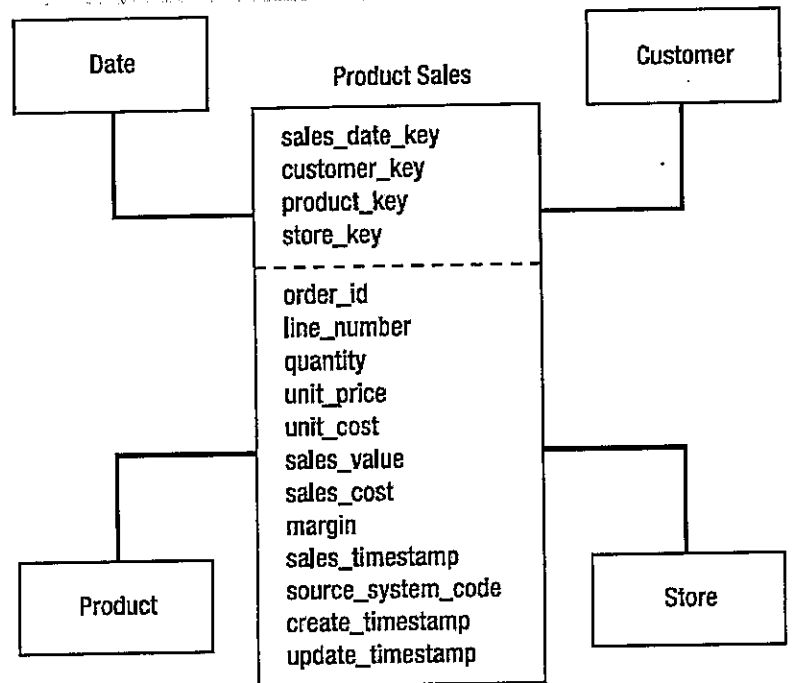
2. This is a user-facing data store, in the form of one or more relational databases, where the data is arranged in dimensional format for the purpose of supporting analytical queries.

- a. Stage
- b. ODS
- c. NDS
- d. DDS

3. Which of the following is the correct order of steps in waterfall methodology?

- a. Feasibility study, requirements, design, architecture, development, testing, deployment, operation
- b. Feasibility study, requirements, architecture, design, development, operation, deployment, testing
- c. Feasibility study, requirements, architecture, design, development, testing, deployment, operation
- d. Feasibility study, requirements, design, architecture, development, operation, deployment, testing

Study the data mart below and answer question 4.



4. Product_key in Product Sales Table is a _____ key.

- a. Primary
- b. Foreign
- c. Degenerate
- d. Natural

5. A _____ is a condition of an object at a particular point in time.

- a. Event
- b. Status
- c. Level
- d. Role

6. Which of the following is an example of a functional requirement?

- a. The data warehouse will store the data up to five years online and five years offline.
- b. Certain sensitive data is viewable by certain people only.
- c. The downloads from Jupiter can occur only from 4 am to 5 am.
- d. Who accessed what and when must be auditable.

7. In a data warehouse, Views are used for three purposes. Which of the following statements is not one of those purposes?

- a. To create conform dimensions in DDS.
- b. To shield users from physical tables, making it simpler for the users as well as to restrict access.
- c. To increase the availability, particularly to make the data warehouse up and running when we are populating it.
- d. To improve the query and loading performance of data warehousing.

8. A star schema has what type of relationship between a dimension and fact table?

- a. May-to-many
- b. One-to-one
- c. One-to-many
- d. All of the above

9. There are several basic principles to understand when extracting data from a source system for the purpose of populating a data warehouse.

What is the most important fundamental principle in data extraction?

- a. Be careful not to slow the source system down too much.
- b. Be careful not to disturb the source system too much.
- c. The extraction should be as small as possible.
- d. The extraction should be as infrequent as possible.

10. Which of the following defines Data scrubbing?

- a. A process to reject data from the data warehouse and to create the necessary indexes
- b. A process to load the data in the data warehouse and to create the necessary indexes
- c. A process to upgrade the quality of data after it is moved into a data warehouse
- d. A process to upgrade the quality of data before it is moved into a data warehouse

11. The illustration shown below is of a

ID	Date	Store	Product	Quantity
2893	2/1/08	32	A160	150
2897	2/4/08	29	B120	240

- a. Fixed position file
- b. Unstructured
- c. Delimited file
- d. Semi-structured file

12. What technique will you use to overcome the following potential problem in data extraction?

Scenario	Potential problem
<i>On the first day of every month, OLTP runs a month-end batch for the whole day.</i>	<i>The OLTP database is not available on certain days.</i>

- a. Skip extraction on that day. Tell users the reasons for this limitation.
- b. Ask the DBA to give "read-only" access. Ask the OLTP DBA to export data for you.
- c. Negotiate a time slot with the OLTP DBA. Break the extraction into smaller chunks.
- d. Program the ETL to rerun automatically later.

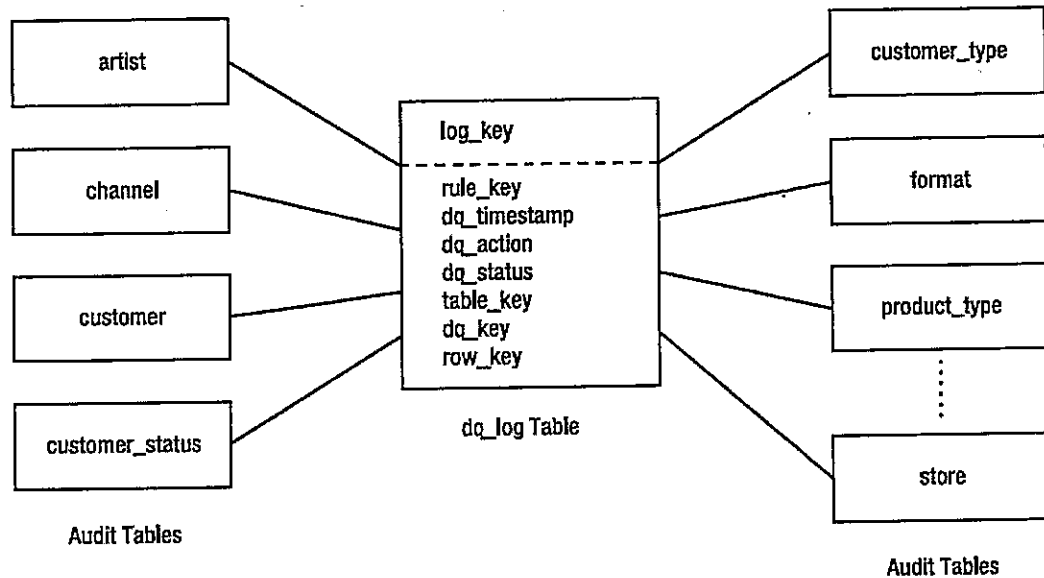
13. Which of the following is not one of the methods to extract data from relation database tables?

- a. Whole table every time
- b. Incremental extract
- c. Pull approach
- d. Fixed range

14. Which of the following is not an aspect of the *data quality process*?

- a. Loading
- b. Checking
- c. Reporting
- d. Correcting

Use the dq_log table below to answer questions 19.



15. Which of the following is a piece of information related to the violation event?

- a. log_key
- b. dq_action
- c. row_key
- d. All of the above

Section B - Short Answers/Descriptive [85 marks]

16. Draw a simple diagram of a data warehouse system and briefly explain it.
[1½+1½= 3 marks]

17. What is the purpose of ETL package?
[1 mark]

18. What is a stage and state **two** reasons why it is necessary.
[1+2=3 marks]

19. State one advantage and disadvantage of Single DDS architecture.
[1+1=2 marks]

20. Draw and discuss the iterative cycle.
[2+2= 4 marks]

21. What is the purpose of doing data feasibility study?
[1 mark]

22. Explain the difference between SCD type 1, SCD type 2, and SCD type 3.
[3 marks]

23. Explain **failover cluster**
[2 marks]

24. Explain the difference between SQL Server licensing models: **Per Processor License** and **Server + CAL License**.
[2+2= 4 marks]

25. Explain two approaches of implementing ETL.
[2 marks]

26. In terms of **who** moves the data out of the source system, we can categorize ETL methods into **four** approaches. Draw diagrams and discuss these approaches.

[2+2+2+2= 8 marks]

27. In terms of **where** the processes that move the data out are executed, we can categorize ETL into three approaches. Draw diagrams and discuss these approaches.

[2+2+2=6 marks]

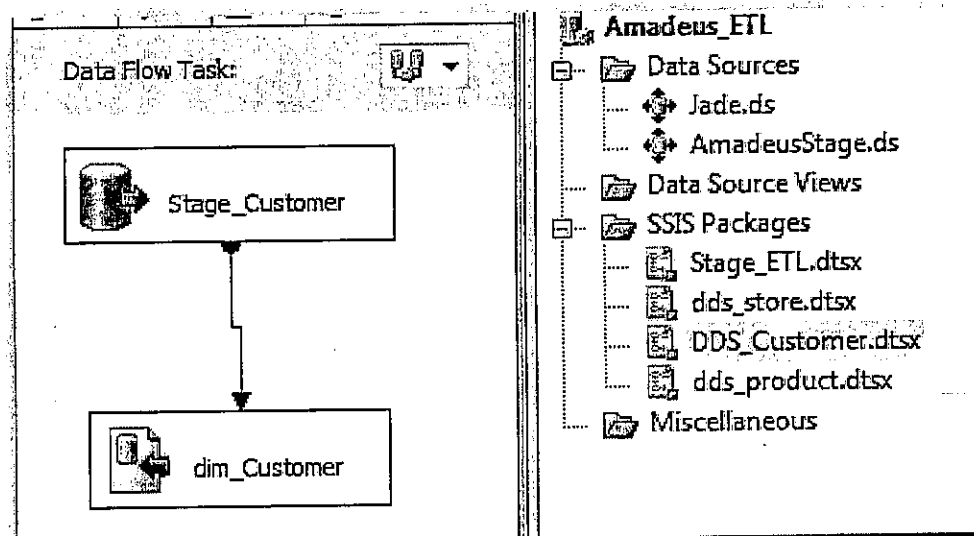
28. What is a data leak and why is it essential to test for data leaks?

[2 marks]

29. Briefly explain three approaches of how stage tables are structured.

[1+1+1=3 marks]

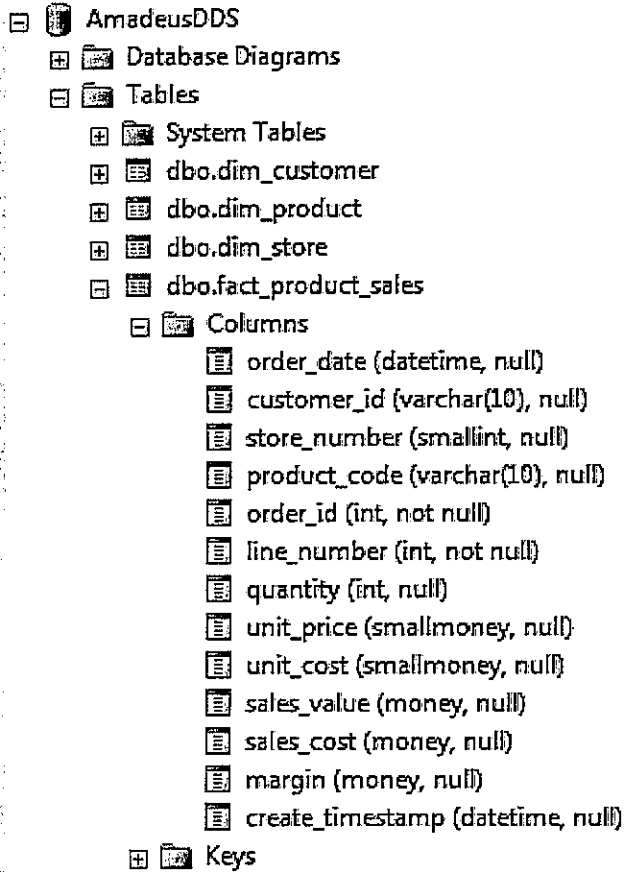
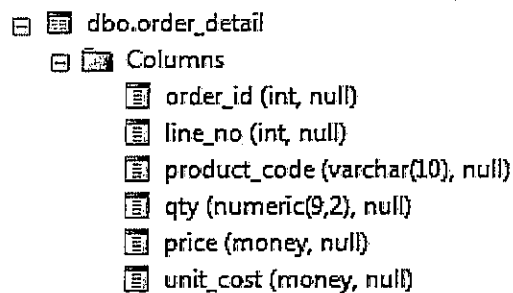
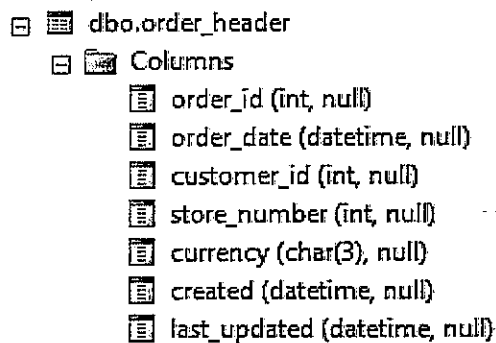
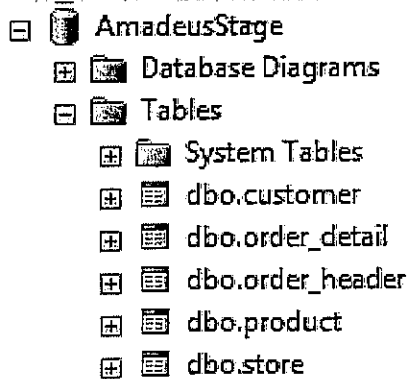
30. Briefly describe the changes in the *Data Flow Task* upon successful execution of the SSIS package *DDS_Customer.dtsx* as shown in the picture below.



[2 marks]

Study the pictorials below and answer question 31-33.

The Populated AmadeusStage and AmadeusDDS databases:



31. Build the SQL command to populate the AmadeusDDS *fact_product_sales* table.

[10 marks]

32. According to Kimball, it is important to declare the **grain** of the fact table. What is the grain of the *fact_product_sales* table?

[1 mark]

33. *order_id* and *line-number* are degenerate dimensions.

a. Explain *degenerate dimension*.

[1 mark]

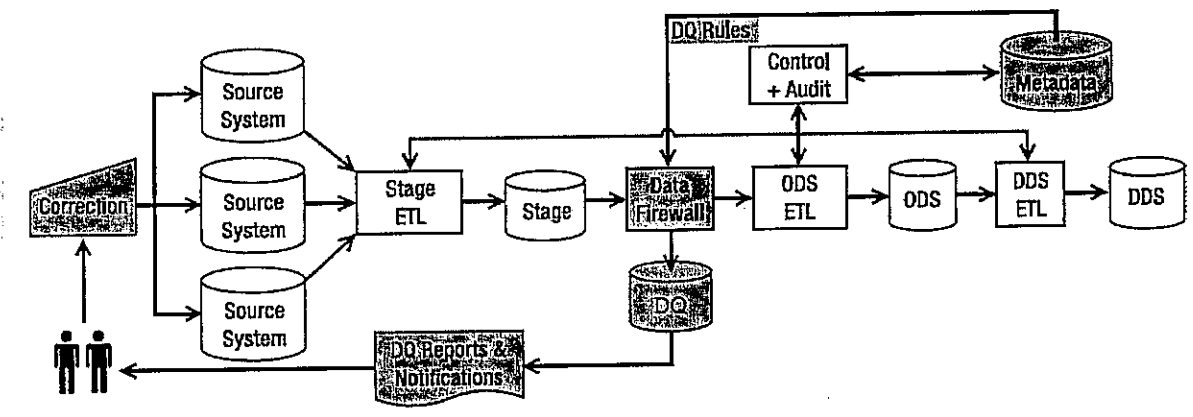
b. Can we use *order_id* and *line_number* as *primary key*? Give reason(s) to support your answer.

[2 marks]

34. Why is data firewall important in data warehousing?

[2 marks]

35. Discuss the data quality components in the data warehouse architecture given below.



[5 marks]

36. Discuss with examples the three types of data matching logic: **Exact**, **Fuzzy**, and **Rule-based**.

[1½+1½+1½=4½ marks]

37. When a data quality rule is violated, we have three choices of what to do with the data. What are the choices?

[1½marks]

38. There are seven kinds of metadata. List at least **three** and briefly state what each describes.

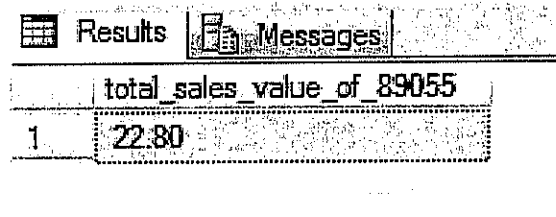
[6 marks]

39. Discuss the three panes of the **Query Builder** window in Business Intelligence Development Server *Report Wizard*.

[3 marks]

40. In reference to the fact_product_sales table on page 9, write down the SQL query that finds the total value of sales made by Customer 89055, as represented in the *Results* below.

[3 marks]



The screenshot shows a software window with two tabs: 'Results' and 'Messages'. The 'Results' tab is active and displays a table. The table has a header row with the text 'total_sales value of 89055'. Below the header is a single data row with the value '22.80'.

total_sales value of 89055	
1	22.80

THE END

