

BCS Level 4 Diploma in Data Analysis Concepts QAN 603/0823/0

Sample Paper A

Record your surname/last/family name and initials on the Answer Sheet.

Sample paper only 40 multiple-choice questions – 1 mark awarded to each question. Mark only one answer to each question. There are no trick questions.

A number of possible answers are given for each question, indicated by either **A. B. C. or D.** Your answers should be clearly indicated on the Answer Sheet.

Pass mark is 26/40.

Copying of this paper is expressly forbidden without the direct approval of BCS, The Chartered Institute for IT.

1 Which of the following include proprietary data?

- a) Internally generated documents relating to an organisation's research and development activities.
- b) Documents published on an organisation's website describing its product and services.
- c) Internally generated documents that are made freely available to the public.
- d) Documents an organisation wishes to keep confidential.

- A** a and d only.
- B** b and c only.
- C** a, b and c only.
- D** b, c and d only.

2 An analyst is working with data provided by their organisation's purchasing department, which contains workforce expenditure and is confidential. It has been provided so that workforce costs can be analysed. What sort of data has the analyst been provided with?

- A** Operational.
- B** Obsolete.
- C** Sensory.
- D** Public.

3 Which of the following CORRECTLY describes open data?

- A** Any data that is held in the public domain.
- B** Data that can be freely used, reused and redistributed.
- C** Data that is recorded and retained by the scientific community.
- D** Data that is obtained from collaboration between organisations.

4 Put the following steps involved in the ETL process in the CORRECT order.

- a) Create the mapping.
- b) Define the source.
- c) Create the session.
- d) Define the target.
- e) Create the work flow.

- A** c, e, a, b, d.
- B** b, d, a, c, e.
- C** a, c, d, e, b.
- D** d, b, a, c, e.

5 An analyst with little or no knowledge of the industry they are working in could inadvertently lead their organisation astray. Which of the following would help the analyst avoid this?

- a) Questioning the source and meaning of every data item before including it.
- b) Employing predictive analysis tools at random stages of the analysis.
- c) Assuming analysis techniques used in other industries are always applicable.

- A a only.
- B b and c only.
- C a and c only.
- D b only.

6 Which of the following is **NOT** a stage of the data lifecycle?

- A Migrated.
- B Created.
- C Archived.
- D Deleted.

7 What is the following an example of?

```
<?xml version="1.0"?>
<payment>
<from>Acme Inc</from>
<to>XYZ Ltd</to>
<amount>1234.56</amount>
<currency>GBP</currency>
</payment>
```

- A Extensible Markup Language.
- B Data Definition Language.
- C Entity-Relationship Model.
- D Unified Modelling Language.

8 Which of the following is **NOT** a common format for structured data?

- A Sequential data files, e.g. .CSV
- B Extensible markup records, e.g. .XML
- C Relational database tables, e.g. .Accdb
- D Data compression archives, e.g. .ZIP

- 9** Which of the following statements is TRUE about structured and unstructured data analysis for businesses?
- A** Statistical analysis is more difficult on structured data than on unstructured data.
 - B** Unstructured data analysis is easier if a business's requirements are not very clear.
 - C** Modern statistical analysis tools always struggle with processing well-structured data.
 - D** Structured and unstructured data can both be used during statistical data analysis.
- 10** What is requirements analysis designed to resolve?
- A** Functional and non-functional requirements.
 - B** Technical requirements.
 - C** Requirements that are changed due to system limitations.
 - D** Ambiguous, conflicting and duplicated requirements.
- 11** What is a requirements elicitation process?
- A** Gathering requirements for an analysis from stakeholders.
 - B** Filtering the initial requirements for an analysis.
 - C** Refining the formal requirements of an analysis.
 - D** Gathering the data that will be used in an analysis.
- 12** Which of the following processes are aspects of validation?
- a) Checking data for accuracy and consistency.
 - b) Checking to ensure that required data values are present.
 - c) Checking data for correctness and meaningfulness in a specific context.
 - d) Checking data for completeness.
- A** a and b only.
 - B** a and d only.
 - C** b and c only.
 - D** c and d only.
- 13** Which of the following statements about the importance of data quality is FALSE?
- A** Enterprise data is a valuable asset and its quality should be protected.
 - B** Poor quality data can cause legal and compliance issues.
 - C** Poor data quality can be very costly to correct.
 - D** Only good quality data can be stored in the Cloud.

14 What results can poor data quality have on a business?

- a) It causes loss of trust in the analytical system.
- b) It makes business continuity planning easier.
- c) It simplifies data modelling.
- d) It increases the time required for reconciliation.

- A** a, b and c only.
- B** b, c and d only.
- C** a and d only.
- D** c and d only.

15 Some business systems will deliberately reduce the quality of the data by omitting or obscuring personal information. What would be a good reason for doing this?

- A** It preserves customer confidentiality.
- B** It makes the conceptual model simpler.
- C** It increases the data accuracy.
- D** It avoids costly data backups.

16 Which of the following are reasons for defining an organisational strategy for data quality?

- a) To increase the amount of system downtime.
- b) To increase the licensing cost of analytical systems.
- c) To increase revenue by improving analytics.
- d) To improve compliance with relevant legislation.

- A** a, b and c only.
- B** a, c and d only.
- C** b and d only.
- D** c and d only.

- 17 What is the MAJOR purpose of hypothesis testing?
- A To choose between two competing hypotheses about the value of a population parameter.
 - B To prove that, with two competing hypotheses, the null hypothesis population parameter is true.
 - C To prove that, with two competing hypotheses, the alternative hypothesis population parameter is true.
 - D To show that there is no significant difference between competing hypotheses about the value of a population parameter.
- 18 Which of the following statements defines a null hypothesis?
- A The opposite of the alternative hypothesis.
 - B The suggestion that data analysed has been corrupted.
 - C The belief that more unstructured data is needed to draw a valid conclusion.
 - D The assumption that the data comes from a valid source.
- 19 Which of the following statements describes interpolation?
- A To create new values within the range of the existing data.
 - B To explain the meaning or result of the information gathered.
 - C To create new estimated values that are outside the range of the original data.
 - D To manipulate or change the values of the original data.
- 20 Which of the following chart types is **MOST** suitable for categorical data?
- A Scatter plot.
 - B Radar chart.
 - C Line graph.
 - D Pie chart.
- 21 What is the **MOST LIKELY** reason for an organisation to comply with the Data Protection Act?
- A They may be fined for non-compliance in the event of a data breach
 - B The cost of doing business will be higher if they don't comply
 - C Their customers will all insist on compliance
 - D They don't handle any personal data but its best to be safe than sorry.

- 22** Which of the following is **NOT** a data structure type?
- A** Arrays.
 - B** Files.
 - C** Folders.
 - D** Trees.
- 23** A layer of construct between physical and logical data, which serves to allocate storage for all DBMS managed segments, is known as what?
- A** List space.
 - B** Table space.
 - C** Records space.
 - D** Slack space.
- 24** Conceptual data modelling is one of the various perspectives of producing data models within database design. What differentiates it from the rest?
- a) It is a map of concepts and their relationships used for databases.
 - b) It describes the semantics of an organisation and represents a series of assertions about its nature.
 - c) It specifies all tables and columns.
 - d) It shows a detailed representation of an organisation's data, independent of any particular technology.
- A** a and c only.
 - B** b and c only.
 - C** b and d only.
 - D** a and b only.
- 25** Sometimes there are relationships in a database model that are indicated indirectly by other relationships. What are these relationships known as?
- A** Flexible.
 - B** Redundant.
 - C** Logical.
 - D** Associative.

- 26 Which of the following describes normalisation?
- A It produces more data for analysis.
 - B It increases data inaccuracy.
 - C It limits the ability to migrate data.
 - D It removes data redundancy.
- 27 What type of analytics summarises what has happened in the past?
- A Prescriptive.
 - B Decision.
 - C Descriptive.
 - D Predictive.
- 28 Which of the following is the method by which the uniqueness of tuples is enforced?
- A Foreign key.
 - B Part key dependency.
 - C Primary key.
 - D Candidate key.
- 29 Which of the following can be classified as metadata?
- a) An index to provide access to a data table.
 - b) The name of an attribute in an object.
 - c) A data object.
 - d) The value of a given data attribute at a point in time.
- A a and c only.
 - B b, c and d only.
 - C a, b and c only.
 - D a and b only.
- 30 Which of the following is **NOT** a source of unstructured data?
- A Digital invoice.
 - B Video file.
 - C Audio file.
 - D Handwritten document.

31 Which of the following requirements techniques extracts tacit knowledge?

- A Interviews.
- B Document analysis.
- C Project charter.
- D Non-functional requirements.

32 Which of the following charts would be used to show time-series data?

- a) Gantt chart.
- b) Line graph.
- c) Null hypothesis.
- d) Word cloud.

- A a and b only.
- B a and c only.
- C a, b and d only.
- D b, c and d only.

33 Which of the following is **NOT** a **KEY** data protection principle?

- A Data minimisation.
- B Freely available.
- C Integrity and confidentiality.
- D Purpose limitation.

34 Which of the following is **NOT** a recognised database technology?

- A Relational.
- B Hierarchical.
- C Object-oriented.
- D Linear.

- 35** Which of the following make requirements analysis difficult?
- a) Duplication and overlap across requirements.
 - b) The correct verification of requirements.
 - c) Lack of clarity in the wording.
 - d) The Agile principles of requirements.
- A** a and b only.
B a and c only.
C b and c only.
D c and d only.
- 36** Which of the following elicitation techniques can be used without the need for user contact?
- a) Interface Analysis.
 - b) Focus group.
 - c) Document Analysis.
 - d) Interviews.
- A** a, b and c only.
B a, c and d only.
C a and c only.
D b and d only.
- 37** Which of the following is a TRUE statement of a null hypothesis?
- A** A null hypothesis can never be wrong.
 - B** You must have a null hypothesis to do data analysis.
 - C** The null hypothesis is the same as the alternative hypothesis.
 - D** The null hypothesis is the outcome that we wish to disprove.
- 38** Which of the following is TRUE in relation to the data lifecycle?
- A** Data can never go back to an earlier stage.
 - B** Data must be archived before its deleted.
 - C** Data must be archived 5 years after it has been created.
 - D** Data is always created before it is archived.

39 Which of the following is **NOT** a TRUE statement relating to a relational database?

- A** They can be accessed with SQL.
- B** They often use foreign / primary keys.
- C** Their structure can be normalised.
- D** They cannot contain more than one single table.

40 Which of the following can transform data into information?

- a) Organisation.
- b) Structuring.
- c) Storage.
- d) Back up.
- e) Processing.

- A** a and b only.
- B** c and d only.
- C** a, b and e only.
- D** a, b and d only.

-End of Paper-