BCS Level 4 Diploma in Data Analysis Concepts
Answer Key and Rationale – QAN 603/0823/0

| Question | Answer | Explanation / Rationale | Syllabus Sections |
|---|---|---|---|
| 1 | A | Proprietary data is internally generated data or documents that contain technical or other types of information controlled by a firm to safeguard its competitive edge. It may be protected under copyright or trade secret laws. | 1.2 |
| 2 | A | Operational data includes internal control and operational environment information such as data on the company's workforce, direct competitors, creditors, suppliers and information on customers. | 1.2 |
| 3 | B | Open data is information that can be freely used, reused and redistributed by anyone with no existing local, national or international legal restrictions on access or usage. | 1.2 |
| 4 | B | This is a representation of the chronological order of the ETL (Extract, Transform and Load) process. | 10.2 |
| 5 | A | Predictive tools do not provide the actual outcome and other business knowledge will not provide key relevance and therefore accuracy. | 11.1 |
| 6 | A | Migration does not feature in the data lifecycle. | 2.2 |
| 7 | A | Extensible Markup Language (XML) is a common source of structured data. Recognising the way XML looks as it structures data allows a data analyst to understand how that data can be queried, transported, exported and more. | 3.2 |
| 8 | D | Structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spreadsheets. Unstructured data formats are often layered within other formats. For example, plain text could be stored inside a document contained within a zip file. | 3.2 |
| 9 | D | Structured and unstructured data are both used extensively in data analysis. | 3.5 |
| 10 | D | Requirements analysis takes account of the possibly conflicting requirements of the various stakeholders, analysing, documenting, validating and managing software or system requirements. | 4.2 |
| 11 | A | Requirements elicitation is the practice of collecting the requirements of a system from users, customers and other stakeholders | 4.3 |
| 12 | C | Data validation is the process of ensuring that a program operates on clean, correct and useful data. | 4.2 |
| 13 | D | The quality of data is not subjective to the location in which it is stored. | 5.1 |

| Question | Answer | Explanation / Rationale | Syllabus Sections |
|---|---|---|---|
| 14 | C | Inaccurate data will prevent reliable analysis and insight. Correcting this will then take additional time to remedy. Typically, poor data quality will not have a positive effect on a business. | 5.2 |
| 15 | A | Obscuring or removing personal data will protect the individual by preventing accurate, genuine data from being available. | 5.1 |
| 16 | D | A high level of data quality will ensure compliance with legislative direction as well as providing insight into key organisational statistics, allowing informed business planning. | 5.4 |
| 17 | A | A hypothesis test is a statistical test that is used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population. | 6.1 |
| 18 | A | The (null) hypothesis is that there is no significant difference between specified populations, any observed difference being due to sampling or experimental error. | 6.2 |
| 19 | A | Interpolation is a method of constructing new data points within the range of a discrete set of known data points. | 6.3 |
| 20 | D | A pie chart is an accurate graphical representation of the categorical distribution of data. | 6.4 |
| 21 | A | The correct answer is A because an organisation may face a fine from the ICO following a data breach if they haven't complied with requirements of the Act. Answer B isn't correct because the cost of doing business is likely to be lower if they don't try and comply. Answer C isn't correct because while some customers may insist on compliance many won't care. Answer D isn't correct because every organisation handles some personal data even if it's only about their own staff. | 7.3 |
| 22 | C | General data structure types include the array, the file, the record, the table and the tree. A folder is a named collection of related files. | 8.2 |
| 23 | B | A table space is a storage location where the actual data underlying database objects can be kept. | 8.2 |
| 24 | D | A conceptual data model maps concepts and their relationships, describing an organisations semantics and demonstrating assertions about its nature. | 9.1 |
| 25 | B | Redundant relationships are those that are already indicated by other relationships, although not directly. | 9.3 |
| 26 | D | Database normalisation is the process of organising the columns (attributes) and tables (relations) of a relational database to reduce data redundancy and improve data integrity. | 9.3 |
| 27 | C | Descriptive analytics is a preliminary stage of data processing that creates a summary of historical data. | 11.2 |
| 28 | C | The Primary Key is the set of mandatory attributes in a relation that enforces uniqueness of tuples. | 8.2 |

| Question | Answer | Explanation / Rationale | Syllabus Sections |
|---|---|---|---|
| 29 | D | Only the index and the name of a data attribute within a class represent data that describes data (metadata). The remaining options refer to physical data artefacts. | 10.2 |
| 30 | A | Digital invoices are structured data. | 3.3 |
| 31 | A | Interviews are good for capturing tacit knowledge. | 4.2 |
| 32 | A | A Gantt chart is a series of horizontal lines that show the amount of work done or production completed in certain periods of time, in relation to the amount planned for those periods.<br>A line graph is a chart representing values of a variable (e.g. time) for suitable values of an independent variable are connected by a broken line. | 6.4 |
| 33 | B | The data protection principles are:<br>• Lawfulness, fairness and transparency<br>• Purpose limitation<br>• Data minimisation<br>• Accuracy<br>• Storage limitation<br>• Integrity and confidentiality (security)<br>Whilst data subjects have the right to access their information under the Data Protection Act 2018 this must be done through a subject access request. | 7.2 |
| 34 | D | Linear is not a recognised database technology. | 9.2 |
| 35 | B | Verification may cause a problem in testing but not in requirements analysis. The agile principles of requirements will not have make analysis more difficult. | 4.2 |
| 36 | C | Interface and document analysis do not require user contact. | 4.3 |
| 37 | D | A null hypothesis is a hypothesis that says there is no statistical significance between the two variables. It is usually the hypothesis a researcher or experimenter will try to disprove or discredit. | 6.2 |
| 38 | D | Creating is always the first step in the data lifecycle. | 2.2 |
| 39 | D | Relational Databases are not restricted to containing only a single table. The relationships of relational databases are generally logical connections between multiple tables. | 8.2 |
| 40 | C | Storing and backing up data does not transform it into information. | 1.1 |