

BCS Level 4 Certificate in Data Analysis Tools  
Answer Key and Rationale – QAN 603/0824/2

| Question | Answer | Explanation / Rationale   | Syllabus Sections |
|----------|--------|---|-------------------|
| 1        | D      | Completeness analysis is used to determine how attributes are used to populate blank or null records.   | 1.4               |
| 2        | B      | A subquery is a query within a query.   | 1.3               |
| 3        | D      | The mean is typically greater than the median, which is always greater than the mode in positively skewed records.  | 2.2               |
| 4        | D      | In excel, data is organised in rows and columns.  | 2.2               |
| 5        | D      | NA is used to represent missing value in R.   | 2.1               |
| 6        | D      | ISNULL is used to set specific values in SQL.   | 1.3               |
| 7        | B      | R is only extendable as packages.   | 2.1               |
| 8        | B      | They are all unstructured data types. Unstructured data is information that either does not have a pre-defined data model or is not organised in a pre-defined manner.  | 1.5               |
| 9        | B      | Data synchronisation is the process of establishing consistency amongst data from a source to a target data storage and vice versa and the continuous harmonisation of the data over time.                                    | 1.5               |
| 10       | B      | The incorrect options listed are non-functional requirements.   | 1.1               |
| 11       | B      | The median of a set of numbers is the number in the middle.   | 2.2               |
| 12       | C      | The mean is the value obtained by dividing the sum of several quantities by the number of quantities there are. If one of the quantities changes, the mean will ultimately change.  | 2.2               |
| 13       | C      | Only a Union join is used to perform this function.   | 1.3               |
| 14       | C      | Data integration reduces complexity by providing a unified view of the data taken from multiple sources.  | 1.1               |
| 15       | A      | ETL is Extract, Transform, Load.  | 1.2               |
| 16       | D      | Ignore or create null value is not used when an error is found, whereas accept and correct would be.  | 1.4               |
| 17       | B      | In business, Master Data Management (MDM) comprises the processes, governance, policies, standards and tools that consistently define and manage the critical data of an organisation to provide a single point of reference. | 1.5               |
| 18       | C      | SELECT DISTINCT picks single records, avoiding duplicates.  | 2.1               |
| 19       | C      | Data cleanliness is not a data quality dimension.   | 1.4               |
| 20       | D      | Data transparency is not a user interface.  | 1.2               |
| 21       | A      | User Acceptance Testing (UAT) - also called end user testing - is a phase of software development in which the software is tested in the 'real world' by the intended audience.   | 1.6               |
| 22       | A      | Dashboard is a combination of data and charts.  | 1.2               |
| 23       | C      | Structured data refers to data that is organised and defined to the extent that it can easily be included and identified in e.g. a database.  | 1.5               |

| Question | Answer | Explanation / Rationale   | Syllabus Sections |
|----------|--------|---|-------------------|
| 24       | A      | Class is a construct that lets data be structured in a particular way. Lists, Dictionary and Tuples all feature in Python.  | 2.1               |
| 25       | C      | Legality is not a reason for testing.   | 1.6               |
| 26       | D      | 'Insert into' is the only option that is not a recognised type of join.   | 1.3               |
| 27       | C      | Data migration is done to change application, system or database - never to remove without a replacement.   | 1.5               |
| 28       | C      | Only Group by is used to perform this function in SQL.  | 1.3               |
| 29       | B      | Data profiling is the process of examining the data available from an existing information source (e.g. a database or a file) and collecting statistics or informative summaries about that data.   | 1.4               |
| 30       | D      | A stress test is focused on determining or validating performance characteristics of the product such as scalability and reliability.   | 1.6               |
| 31       | A      | One reason for integrating data for the quick analysis of separate data sources that together allow analysis that is not possible when they are separate. Integration of data has nothing to do with GDPR. Storage is more likely to go up than down because original data sources and integrated data are all stored. It is very unlikely that integrating data sources will reduce storage. Data integration has no relationship to database normalisation. | 1.1               |
| 32       | B      | Only "Select" retrieves/extracts data.  | 1.2               |
| 33       | A      | A data integration dashboard typically displays the progress of processing the data.  | 1.2               |
| 34       | D      | Union appends the output of one select statement to another and hence integrates data into one tabular output. The other statements imply using joins to provide a subset of the data.  | 1.3               |
| 35       | A      | Validity is checking the correct format, type and range of data.  | 1.4               |
| 36       | B      | Social media is unstructured data and large volumes would imply big data. Whilst this could be processed using conventional tools it is likely this would be processed using big data tools such as Hadoop.   | 1.5               |
| 37       | D      | Current, complete and accurate data are collected to make sure that the systems are tested on data that reflects as closely as possible real-world data. Test data not in the real-world may crash the system so 'A' and 'B' cannot be correct. There is no relationship between test data and legislations so 'C' cannot be correct.   | 1.6               |
| 38       | A      | R is a programming language specifically designed with built in statistical functions.  | 2.1               |
| 39       | C      | Whilst Python and R could be used to sort large volumes of structured data, SQL is a tool that is specifically designed and optimised to do this. UIMA is a java framework for unstructured data  | 2.1               |
| 40       | A      | Significance is the measure of certainty that a result produced from statistical analysis is not because of a random chance.  | 2.2               |