



# **BCS Level 4 Certificate in Data Analysis Tools**

## **QAN 603/0824/2**

**Version 2.1**  
**July 2020**

This is a United Kingdom government regulated qualification which is administered and approved by one or more of the following: Ofqual, Qualifications Wales, CCEA Regulation or SQA.

# BCS Level 4 Certificate in Data Analysis Tools

## Contents

- Introduction ..... 4
- Objectives ..... 4
- Course Format and Duration ..... 5
- Eligibility for the Examination..... 5
- Format and Duration of the Examination ..... 5
- Additional time for learners requiring Reasonable Adjustments due to a disability ..... 5
- Additional time for learners whose language is not the language of the examination..... 5
- Syllabus ..... 7
- Levels of Knowledge / SFIA Levels ..... 11
- Question Weighting ..... 11
- Format of Examination ..... 12
- Trainer Criteria ..... 12
- Classroom Size ..... 12

## Change History

Any changes made to the syllabus shall be clearly documented with a change history log. This shall include the latest version number, date of the amendment and changes made. The purpose is to identify quickly what changes have been made.

| <b>Version Number</b>        | <b>Changes Made</b>   |
|------------------------------|---|
| Version 1.0<br>December 2016 | Syllabus Created  |
| Version 1.1<br>December 2016 | Compliance statement added to title page.   |
| Version 2.0<br>March 2018    | Revised syllabus created.   |
| Version 2.1 July<br>2020     | Document amended to make it suitable for a range of different learners and as well as apprentices |

## Introduction

This Certificate is the first module of the two knowledge modules required for the Level 4 Data Analyst apprenticeship programme. It covers the range of concepts, approaches, tools and techniques that are applicable to Data Analysts, for which learners are required to demonstrate their knowledge and understanding.

This Certificate can also be delivered as a standalone programme for learners working to develop their knowledge and understanding of the processes and tools used for data analysis.

## Objectives

Learners should be able to demonstrate knowledge and understanding of Data Analysis Tools and the underlying principles and processes of data integration. Key areas are:

- Explain the purpose and outputs of data integration activities.
- Explain how data from multiple sources can be integrated to provide a unified view of the data.
- Describe how programming languages for statistical computing (SQL) can be applied to data integration activities, improving speed and data quality for analysis.
- Explain how to take account of data quality when preparing data for analysis, improving quality, accuracy and usefulness.
- Explain the nature and challenges of data volumes being processed through integration activities and how a programming approach can improve this.
- Understand testing requirements to ensure that unified data sets are correct, complete and up to date.
- Explain the capabilities (speed, cost, function) of statistical programming languages and software tools, when manipulating, processing and cleaning data and the tools required to solve analysis issues.
- Explain how statistical programming languages are used in preparing data for analysis and within analysis projects.

Learners who are completing this as part of the apprenticeship programme should collate evidence of lessons learnt in these key areas and these should be reflected upon when the learner is compiling the Summative Portfolio. This will provide the learner with the opportunity to identify how the task might be done better/differently with knowledge subsequently gained.

## Target Audience

The Certificate is relevant to learners who are either enrolled on a Level 4 Data Analyst apprenticeship programme or want to gain more understanding of data analysis to a level 4 standard.

## Course Format and Duration

Candidates can study for this Certificate by attending a training course provided by a BCS accredited Training Provider. The estimated total qualification time for this Certificate is 145 hours.

## Eligibility for the Examination

Apprenticeship learners:

Individual employers will set the selection criteria, but this is likely to include 5 GCSEs (especially English, mathematics and a science or technology subject); other relevant qualifications and experience; or an aptitude test with a focus on IT skills. Learners should have a strong quantitative background and a solid understanding of statistics. They should have experience in statistical programming language, database and software tool types e.g. SQL, Python, R.

Level 2 English and Maths will need to be achieved, if not already, prior to taking the end point assessment.

Other learners:

It is recommended that learners have completed 5 GCSEs (especially English, mathematics and a science or technology subject); other relevant qualifications and experience; or an aptitude test with a focus on IT skills. Training providers may have selection criteria based on the above. Learners should have a strong quantitative background and a solid understanding of statistics. They should have experience in statistical programming language, database and software tool types e.g. SQL, Python, R.

## Format and Duration of the Examination

The format for the examination is a one-hour multiple-choice examination consisting of 40 questions. The examination is closed book (no materials can be taken into the examination room). The pass mark is 26/40 (65%).

## Additional time for learners requiring Reasonable Adjustments due to a disability

Learners may request additional time if they require reasonable adjustments. Please refer to the [reasonable adjustments policy](#) for detailed information on how and when to apply.

## Additional time for learners whose language is not the language of the examination

If the examination is taken in a language that is not the learners native/official language, then they are entitled to 25% extra time.

If the examination is taken in a language that is not the learners native/official language, then they are entitled to use their own **paper** language dictionary (whose purpose is translation between the examination language and another national language) during the examination. Electronic versions of dictionaries will **not** be allowed into the examination room.

## Guidelines for Training Providers

Each major subject heading in this syllabus is assigned an allocated time. The purpose of this is two-fold: first, to give both guidance on the relative proportion of time to be allocated to each section of an accredited course and an approximate minimum time for the teaching of each section; second, to guide the proportion of questions in the exam. Training Providers may spend more time than is indicated and learners may spend more time again in reading and research. Courses do not have to follow the same order as the syllabus. Courses may be run as a single module or broken down into two or three smaller modules.

This syllabus is structured into sections relating to major subject headings and numbered with a single digit section number. Each section is allocated a minimum contact time for presentation. Learners enrolled on an apprenticeship programme should be encouraged to consider their Summative Portfolio throughout the modules.

# Syllabus

For each top-level area of the syllabus a percentage and K level is identified. The percentage is the exam coverage of that area, and the K level identifies the maximum level of knowledge that may be examined for that area.

## 1. Processes and Tools Used for Data Integration (70%, K3)

In this topic, the learner will describe how data integration is achieved through the manipulation of data from different sources through the use of programming languages and how it is prepared for analysis. The successful learner should be able to:

1.1. Explain the purpose and outputs of data integration activities.

- functional requirements;
- non-functional requirements:
  - speed
  - time available
- information structure and rules;
  - policies
  - practices
- rationale for using and integrating data from multiple sources;
- importance of data in a business context.

1.2. Explain how data from multiple sources and systems, can be integrated to provide a unified view of the data.

- business need for analysis;
- reasons for using data from multiple sources;
- importance of data source quality to improve the quality of results;
- filtering data to ensure only relevant data is combined to underpin business objectives;
- data integration techniques.
  - common user interface
    - dashboard
    - scorecard
    - dynamic
  - virtual integration
    - communication channels
    - data transfer
  - physical data integration
    - ETL (extract - transform - load)

1.3. Describe how programming languages for statistical computing (SQL) can be applied to data integration activities, improving speed and data quality for analysis.

- programming constructs;
  - sequence, selection and iteration
- single queries;

- multiple queries (UNION);
- expressions;
  - CASE
  - DATETIME
  - Function
  - Compound
- Functions;
  - Avg
  - Count
  - Max
  - Min
  - Group by
  - Round
  - Cast
  - Convert
  - ISNULL
- Querying multiple tables in different information;
  - joins
    - inner and outer
    - right and left
    - full
    - union
    - select into
    - subqueries
  - joins with duplicate values;
  - joining on multiple fields;
  - select and select\* statements;
  - from;
  - where;
    - AND
    - OR
    - use of wildcards and ordering
  - selecting the first / last of occurrences;
  - implicit data conversion.

1.4. Explain how to take account of data quality in preparing data for analysis to improve accuracy, quality and usefulness.

- data profiling;
- what happens when an error or issue is found;
  - accept
  - reject
  - correct error
  - create default value
- data quality dimensions.



- completeness
- uniqueness
- timeliness
- validity
- accuracy
- consistency

1.5. Explain the nature and challenges of data volumes being processed through integration activities and how a programming approach can improve this.

- big data;
  - unstructured data
  - structured data
- technical requirements for managing large data set;
  - the location of data and challenge of restrictions due to the computer architecture (software and the system);
- data migration;
- master data management;
- integration design;
  - rules and requirements
  - objectives and deliverables
  - support models and SLAs
- data integration tools (SQL);
  - future scalability
  - implementation
  - support costs
- data synchronisation.
  - data ownership
  - frequency of updates
  - format
  - security
  - data quality
  - performance
  - maintenance

1.6. Understand testing requirements to ensure that unified data sets are correct, complete and up to date.

- business testing and technical testing.
  - technical acceptance testing (TAT)
  - user acceptance testing (UAT)
  - performance stress tests (PST)

## **2. Industry Standard Tools and Methods for Data Analysis (30%, K3)**

In this topic, the learner will describe and use a range of tools, techniques and methods to prepare and analyse data. The successful candidate should be able to:

2.1. Explain the capabilities (speed, cost, function) of statistical programming languages and software tools, when manipulating, processing and cleaning data and the tools required to solve analysis issues.

- capabilities and functions of statistical programming language;
  - R
- programming language;
  - Python
- relational databases;
  - SQL
- non-relational databases;
  - graph
  - document
  - column
  - family
- software tools.
  - Excel

2.2. Explain how statistical programming languages are used in preparing data for analysis and within analysis projects.

- preparation techniques;
  - searching and sorting
  - grouping
  - filtering
  - modelling
- data cleaning to remove a range of data issues;
  - types of errors
    - missing data
    - inconsistencies
    - redundancy
  - invalid values
  - data that is out of range
  - outliers
- processing and analysing:
  - mean, median, mode and range
  - probability
  - bias
  - statistical significance
  - linear regression (simple and multiple)
  - scatter plots and correlation
  - and / or probability
  - stem and leaf plots (frequency and distribution)
  - factorials
  - box and whisker plots
- methods for presenting results.
  - tables
  - charts
  - graphs

## Levels of Knowledge / SFIA Levels

This syllabus will provide learners with the levels of difficulty / knowledge skill highlighted within the following table, enabling them to develop the skills to operate at the levels of responsibility indicated. The levels of knowledge and SFIA levels are explained on the website [www.bcs.org/levels](http://www.bcs.org/levels). The levels of knowledge above will enable learners to develop the following levels of skill to be able to operate at the following levels of responsibility (as defined within the SFIA framework) within their workplace:

| Level     | Levels of Knowledge | Levels of Skill and Responsibility (SFIA) |
|-----------|---------------------|---|
| <b>K7</b> |                     | Set strategy, inspire and mobilise        |
| <b>K6</b> | Evaluate            | Initiate and influence                    |
| <b>K5</b> | Synthesise          | Ensure and advise                         |
| <b>K4</b> | Analyse             | Enable                                    |
| <b>K3</b> | Apply               | Apply                                     |
| <b>K2</b> | Understand          | Assist                                    |
| <b>K1</b> | Remember            | Follow                                    |

## Question Weighting

| Syllabus Area  | Target number of questions |
|--|----------------------------|
| 1. Processes and tools used for data integration         | 29                         |
| 2. Industry standard tools and methods for data analysis | 11                         |
| <b>Total</b>   | <b>40 Questions</b>        |

## Format of Examination

|                                |   |
|--------------------------------|---|
| Type                           | 40 Question Multiple Choice   |
| Duration                       | 1 hour. An additional 25% will be allowed for learners sitting the examination in a language that is not their native /mother tongue. |
| Pre-requisites                 | Training from a BCS accredited Training Provider is strongly recommended but is not a pre-requisite.                                  |
| Supervised                     | Yes   |
| Open Book                      | No  |
| Pass Mark                      | 26/40 (65%)   |
| Calculators                    | Calculators cannot be used during this examination.   |
| Total Qualification Time (TQT) | 145 Hours   |
| Delivery                       | Online  |

## Trainer Criteria

|          |   |
|----------|---|
| Criteria | <ul style="list-style-type: none"><li>▪ Have 10 days training experience or have a train the trainer qualification</li><li>▪ Have a minimum of 3 years practical experience in the subject area</li></ul> |
|----------|---|

## Classroom Size

|                          |      |
|--------------------------|------|
| Trainer to learner ratio | 1:16 |
|--------------------------|------|