



BCS Professional Certificate in Data Analysis

Version 2.1.4

May 2024

This professional certification is not regulated by the following United Kingdom Regulators – Ofqual, Qualification in Wales CCEA or SQA.

Contents

Document Change History	3
Introduction	4
Qualification Suitability and Overview	4
Trainer Criteria	4
SFIA Levels.....	5
Learning Outcomes.....	5
Syllabus	6
Examination Format.....	8
Question Weighting.....	9
Using BCS Books	10
Recommended Reading	10
Glossary.....	12

Document 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.

Version Number	Changes Made
V2.0	Document Creation
V2.1	Version number updated to V2.1. Date amended. Glossary terms expanded to include: ADDITIVITY, DATA NAVIGATION DIAGRAM, MEAN, MEDIAN, MODE and NULL VALUE. Reading list expanded to include: Business Analysis (4th Edition), Business Analysis Techniques: 123 Essential Tools for Success (3rd Edition), DAMA Data Management Knowledge System Guide (Original Book 2nd Edition), UML 2 and the Unified Process: Practical Object-Oriented Analysis and Design (2nd Edition).
V2.1.1	Exam duration corrected, from 60 to 90 minutes.
V2.1.2	GLH/TQT updated. CONCATENATED KEY and COMPOUND KEY edited in glossary. Calculator allowed included in exam details. Reading list updated.
V2.1.3	Glossary definition of 'hierarchic key' updated.
V2.1.4	Addition of bullet point 'Third Normal Form data model' to learning objective 3.2. New entry for CANDIDATE KEY to glossary. New entry for THIRD NORMAL FORM DATA MODEL to glossary. Minor tweak to the wording of the entry for COMPOUND KEY in the glossary.

Introduction

The BCS Professional Certificate in Data Analysis comprises fundamental principles, concepts and techniques used to identify, analyse and model data. The aim of this certification is to enable candidates to define data requirements with detailed understanding and rigour.

The certification examination assesses knowledge and understanding of a range of activities and techniques that may be used by business analysts to elicit and analyse data requirements and the business rules inherent in the data, and to define the structure of the data that will support the business requirements in an unambiguous fashion.

Qualification Suitability and Overview

Total Qualification Time	Guided Learning Hours	Independent Learning	Assessment Time
17.5 hours	12 hours	4 hours	1 hour 30 minutes

Trainer Criteria

Summary of criteria:

- Hold the BCS Professional Certificate in Data Analysis.
- Have 10 days training experience or have a 'Train the Trainer' qualification.
- Have a minimum of 3 years practical experience in the subject area.

SFIA Levels

This syllabus has been linked to the SFIA knowledge, skills and behaviours required at level 5 for an individual working in a data role.

Further detail around the SFIA Levels can be found at www.bcs.org/levels.

This award provides candidates with the level of knowledge highlighted within the table, enabling candidates to develop the skills to operate successfully at the levels of responsibility indicated.

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

Learning Outcomes

Upon completion of the certificate candidates will be able to demonstrate an understanding of:

- the basics of Data Analysis.
- how to model data using class diagrams.
- how to define data requirements.
- the ways in which data is obtained and recorded.
- how to analyse data for decision-making.
- how data is protected.

Syllabus

For each top-level area of the syllabus, a percentage and K level are identified. The percentage is the examination coverage for that area; the K level identifies the highest level (within Bloom's taxonomy of cognitive domains) that may be examined for that area.

1. Introduction to data (10%; K2)

- 1.1 Define the terms: data, data analysis, data model, information and business intelligence.
- 1.2 Distinguish between structured and unstructured data
- 1.3 Explain the following data concepts:
 - Conceptual, logical, physical data models
 - Static and dynamic views
- 1.4 Define the stages in the data lifecycle
 - Identifying data sources
 - Modelling data requirements
 - Obtaining data
 - Recording data
 - Using data for business decisions and operations
 - Removing data

2. Modelling data using Class Diagrams (35%; K4)

- 2.1 Define concepts and notations used in class diagrams
 - Classes and objects
 - The structure of a class: name, attributes, operations
 - Modelling classes
 - Associations
 - Labelling associations
 - Multiplicity
 - Composition and Aggregation
 - Attributes
- 2.2 Interpret a class diagram
- 2.3 Explain the use of generalisation in class diagrams

3. Defining data requirements (15%; K3)

- 3.1 Define data modelling concepts
 - Metadata
 - Domain definitions
- 3.2 Explain data normalisation
 - Rationale for data normalisation
 - Unnormalised form
 - First normal form, second normal form and third normal form relations
 - Simple, compound, hierarchic and foreign keys
 - Third Normal Form data model
- 3.3 Identify aspects of data quality

4. **Obtaining and recording data (10%: K3)**

4.1 Identify sources of data: surveys, sampling exercises, records

4.2 Validate data models using a CRUD matrix

4.3 Validate data models against requirements using Data Navigation Paths

5. **Analysing data for decision-making (25%: K4)**

5.1 Explain and apply data analytics concepts

- Obtaining the data set: context, source and lineage
- Validating the data set: confirmation bias, sample size, outliers, consistency
- Dataset calculations: counts, totals, averages, probabilities
- Data relationships: regression analysis; correlation and causation; time-series forecasting

5.2 Explain data cleansing: rationale and key steps

5.3 Interpret data using data analytics concepts

6. **Protecting data (5%: K2)**

6.1 Define data protection principles

6.2 Define aspects relating to online data and ethics

Examination Format

This certificate is assessed through completion of an invigilated online exam which candidates will only be able to access at the date and time they are registered to attend.

Type	40 Multiple Choice Questions
Duration	90 minutes
Supervised	Yes
Open Book	No (no materials can be taken into the examination room)
Calculator Allowed	Yes
Passmark	26/40 (65%)
Delivery	Digital format only

Adjustments and/or additional time can be requested in line with the [BCS Reasonable Adjustments Policy](#) for candidates with a disability or other special considerations, including English as a second language.

Question Weighting

Each major subject heading in this syllabus is assigned a percentage weighting. The purpose of this is:

- 1) Guidance on the proportion of content allocated to each topic area of an accredited course.
- 2) Guidance on the proportion of questions in the exam.

Syllabus Area	Syllabus Weighting	Question type
1. Introduction to data	10%	Multiple choice
2. Modelling data using Class Diagrams	35%	Multiple choice / Scenario
3. Defining data requirements	15%	Multiple choice / Scenario
4. Obtaining and recording data	10%	Multiple choice / Scenario
5. Analysing data for decision-making	25%	Multiple choice
6. Protecting data	5%	Multiple choice
Total	100%	

Using BCS Books

Accredited training organisations may include excerpts from BCS books in the course materials. If you wish to use excerpts from the books, you will need a licence from BCS to do this. If you are interested in taking out a licence to use BCS published material, you should contact the Head of Publishing at BCS, outlining the material you wish to copy and the use to which it will be put.

Recommended Reading

Title: Business Analysis (4th Edition)
Author: Debbie Paul and James Cadle
Publisher: BCS, Learning and Development Limited
Publication Date: July 2020
ISBN: 978 1780175102

Title: Business Analysis Techniques: 123 Essential Tools for Success (3rd Edition)
Author: James Cadle, Debbie Paul, Jonathan Hunsley, Adrian Reed, David Beckham and Paul Turner
Publisher: BCS, Learning and Development Limited
Publication Date: August 2021
ISBN: 978 1780175690

Title: DAMA Data Management Knowledge System Guide (Original Book 2nd Edition)
Author: DAMA International
Publisher: Technics Publications
Publication Date: July 2017
ISBN: 978 1634622349

Title: Data Modeling Essentials
Author: Graeme Simsion and Graham C Witt.
Publisher: Morgan Kaufmann Publishers Inc
Publication Date: November 2004
ISBN: 978 0126445510

Title: Data Modeling Made Simple: A Practical Guide for Business and IT Professionals
Author: Steve Hoberman.
Publisher: Take IT with You Series
Publication Date: August 2009
ISBN: 978 0977140060

Title: Developing Information Systems
Author: James Cadle.
Publisher: BCS, Learning and Development Limited
Publication Date: September 2014
ISBN: 978 1780172453

Title: Introducing Systems Development
Author: Steve Skidmore and Malcolm Eva.
Publisher: Palgrave Macmillan
Publication Date: August 2003
ISBN: 978 0333973691


Title: UML 2 and the Unified Process: Practical Object-Oriented Analysis and Design (2nd Edition)
Author: Jim Arlow and Ila Neustadt
Publisher: Addison Wesley
Publication Date: July 2005
ISBN: 978 0321321275

Glossary

The full syllabus and recommended reading list may be found on the [BCS website](#).

TERM	DEFINITION
ADDITIVITY	The extent to which the addition of values to other values results in a meaningful outcome. Values may be fully-additive, semi-additive or non-additive.
AGGREGATION	An association between two classes that specifies that one class is the 'whole' and the set of related classes are the component parts. In an aggregation association, the component parts can exist independently from the class that represents the 'whole'. This is similar to the COMPOSITION association.
ARTEFACT	A diagram or supporting description providing a representation of the system of interest.
ASSOCIATION	A business link that represents a meaningful relationship between two classes. The link is required in order to navigate from one class to another.
ATTRIBUTE	An individual item of data required to describe something of interest in the business system. An attribute may belong to an entity or a class depending upon the modelling approach in use.
BUSINESS INTELLIGENCE	Business intelligence concerns the use of data analysis and analytics techniques, tools and best practices in order to make data-enabled business decisions.
CANDIDATE KEY	Any possible key for a relation regardless of whether it has been nominated as the primary key or not.
CARDINALITY	The degree of occurrence indicated on a relationship between two entities. The cardinality reflects the business rules for a relationship. Also see OPTIONALITY.
CAUSATION	The cause-and-effect relationship between two variables whereby one variable causes a particular effect to arise regarding the other variable.
CLASS	A class provides the template definition used to describe a set of objects that share the same attributes, operations, methods, relationships and behaviour. Each object is an instance of a particular class. (Based on Rumbaugh, Booch, Jacobson, UML Reference Guide, 1999).
CLASS MODEL	A technique from the Unified Modeling Language (UML). A class model consists of a diagram that represents the classes in a system and their associations with each other plus the definitions of the classes and associations.
CONFIRMATION BIAS	The tendency to search for and interpret information in order to confirm or support a particular viewpoint or set of beliefs.

COMPOUND KEY	A concatenated key where each element of the key is a unique identifier and also acts as a foreign key.
COMPOSITION	An association between two classes that specifies that one class is the 'whole' and the set of related classes are the component parts. In a composition association, the component parts cannot exist independently from the class that represents the 'whole'. This is a stronger form of the AGGREGATION association.
CONCATENATED KEY	A unique identifier for a relation that is made up of more than one attribute.
CONCEPTUAL DATA MODEL	A data model that represents the high-level data requirements. It contains the key business entities relevant to a particular business area.
CORRELATION	The degree of association between two variables. A correlation between two variables does not imply CAUSATION.
CRUD MATRIX	A matrix that shows the cross-references between the event/use case view and the data view for a system. CRUD stands for Create, Read, Update, Delete.
DATA	A collection of related facts and values from which information may be derived (Data Administration and Business Information System analysis)
DATA ANALYSIS	The activity concerned with conceptually understanding, organising and defining the data items that are relevant and useful in a business system.
DATA ANALYTICS	The activities concerned with interrogating and interpreting data for the purpose of business decision-making.
DATA CLEANSING	Reviewing data to detect and remove errors in order that the data aligns with required standards and business rules.
DATA ETHICS	The principles that govern the sourcing, recording, processing and disposal of data. Three core concepts related to data ethics are: the impact on people; the potential for misuse; the economic value associated with the data.
DATA LINEAGE	The provenance of the data based on understanding the sources of the data, how it is used, where it is transferred, and where and how it is stored.
DATA MODEL	A diagram and definitions of the data required to support a business or IT system and the business rules that govern the creation, amendment and deletion of those elements. A data model may include an entity relationship diagram or a class diagram (from the UML).
DATA MODELLING	The activity concerned with building a model of the data required to support a business or IT system.

DATA NAVIGATION DIAGRAM	<p>A diagram representing the navigation path to access the data required to fulfil a particular enquiry requirement. The diagram represents the classes or entities (data tables) to be accessed and the sequence of access. The arrows identify the route of the navigation through the data model. A round-cornered box is used to represent a class and an arrow is used to show the original class accessed and the navigation path between two classes. A single arrowhead indicates the access of a single occurrence of the destination class and a double arrowhead indicates that one or more occurrences of the destination class are accessed.</p>  <pre> graph LR EPD[Entry point data] --> C1[Class 1] C1 --> C2[Class 2] C2 ==> C3[Class 3] </pre>
DATA PROTECTION	<p>The need to ensure that personal and sensitive data is collected, recorded, accessed, used and retained such that each individual's privacy rights are maintained and respected. The DAMA-DMBOK lists seven GDPR principles: Fairness, Lawfulness, Transparency; Purpose Limitation; Data Minimization; Accuracy; Storage Limitation; Integrity and Confidentiality; Accountability.</p>
DATA QUALITY	<p>The degree to which data meets the needs of the context within which it is to be used. The DAMA-DMBOK defines six core dimensions of data quality: completeness; uniqueness; timeliness; validity; accuracy; consistency.</p>
DOMAIN	<p>A defined format and set of values that is applied to a set of ATTRIBUTES.</p>
ENTITY	<p>An entity is a grouping of the attributes that describe something of interest within the business system.</p>
ENTITY RELATIONSHIP MODEL	<p>A diagram, and supporting descriptions, produced to represent the data to be held within a business system under investigation. The supporting descriptions provide information about attributes and relationships between entities. The data groupings are represented as entities and the logical business connections between them are shown as relationships.</p>
FOREIGN KEY	<p>The primary key for a relation that is incorporated as a non-key attribute in another relation in order to enable a relationship that exists between them.</p>
HIERARCHIC KEY	<p>A concatenated key that is made up of at least one attribute that is a foreign key and at least one additional identifying element such as a sequential number.</p>
KEY	<p>A unique identifier for a relation that comprises one or more data items (attributes).</p>

LOGICAL DATA MODEL	A data model that represents the detailed data requirements for a particular product or software application. The model is technology-independent and may be developed from a CONCEPTUAL DATA MODEL.
MEAN	The average value calculated by adding up the values within a set of values and dividing by the total number of values within that set.
MEDIAN	The middle value within a set of values.
MODE	The most frequently-occurring value within a set of values.
MULTIPLICITY	The degree of association between two classes. The multiplicity represents the business rules for an association.
NULL VALUE	A NULL value represents the absence of a value and is sometimes represented as 'Ø'. NULL is not the same as '0' zero, as zero is a value.
OBJECT	An object is something within a business system for which a set of attributes and functions can be specified. An object is an instance of a class.
OPERATION	A set of actions performed on the data within a class.
OPTIONALITY	The ability of an entity occurrence to exist without being linked to another entity with which there is a business relationship.
OUTLIER	A data value within a set that is significantly different from the other data values within that set.
NORMALISATION	The process used to ensure that all attributes within a relation are identified by the key to the relation and nothing else. The data is put into third normal form so that it is organised without duplication or redundancy.
PHYSICAL DATA MODEL	A data model that represents the detailed data requirements within the context of a particular technical solution. The model is typically developed from a LOGICAL DATA MODEL.
REGRESSION ANALYSIS	The analysis of the relationship between two (or more) variables: the dependent variable and the independent variable(s). It is used to predict future performance based on the data from previous performance.
RELATION	A data set formed during data NORMALISATION.
RELATIONSHIP	A business link that represents a meaningful association between two entities. The link is required in order to navigate from one entity to another. A similar concept to an association in UML class modelling.
SIMPLE KEY	A unique identifier for a relation comprised of just one attribute.

STRUCTURED DATA	Data that is stored with a high level of organisation where data types, formats and relationships are defined and the data may be queried and reported upon. Structured data is typically held in databases or spreadsheets and has a pre-defined schema (or structured framework).
THIRD NORMAL FORM DATA MODEL (TNF MODEL)	A data model derived from a set of third normal form relations that have been derived using data normalisation. TNF data models use entity-relationship diagram notation but are not ERDs as they do not show optionality, exclusivity or relationship names; these constructs are determined when conducting business discussions to derive a top-down logical data model.
TIME-SERIES FORECASTING	A method of analysing a set of data values that are plotted in a sequence of equally-spaced points in time. The purpose of this analysis is to identify trends or patterns that enable prediction about future data values.
UNSTRUCTURED DATA	Data that is stored without an underlying structure so may not be accessed for reporting or query purposes. Unstructured data typically takes the form of audio, video and unstructured text in documents.