BCS Professional Certificate in Data Analysis

Version 3.0

February 2026

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

Copyright © Assist Knowledge Development 2026

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 11
Question Weighting 11
Using BCS Books 12
Recommended Reading 12
Glossary 14

Copyright © Assist Knowledge Development 2026

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, NULL VALUE, DATA CONTROLLER, DATA PROCESSOR, DATA SUBJECT, DESCRIPTIVE ANALYTICS, PREDICTIVE ANALYTICS, REAL-TIME ANALYTICS. 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	Syllabus 3.2 added Third Normal Form data model. Glossary definitions added – Candidate key, Third Normal Form data model (TNF Model) and updated Compound key.
V3.0	• Existing topics have been clarified for better understanding. • Topics have been restructured and weightings rebalanced across the syllabus. • New content has been added on how organisations use and source data, and the use of AI in data analytics. • The syllabus now includes key concepts such as **big data**, **semi-structured data**, **veracity**, and **confidence levels**. • Topics on data protection and the ethical use of data have been further expanded. • Revised glossary definitions for: **Aggregation**, **Composition**, **Median**, **Object** and **Operation**. • Added the following definitions to the glossary: **Big data**, **Confidence level**, **Consistency**, **Dynamic data**, **Generalisation**, **Hypothesis**, **Information**, **Metadata**, **Population**, **Primary key**, **Sampling**, **Sample size**, **Semi-structured data**, **Standard deviation**, **Static data** and **Veracity.** • The recommended reading list has been refreshed.

Introduction

The BCS Professional Certificate in Data Analysis comprises fundamental principles, concepts and techniques used to identify, model, obtain and analyse 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.

Copyright © Assist Knowledge Development 2026

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 basic concepts and terminology underlying 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
- the principles and key concepts associated with data protection and the ethical use of data
- the potential benefits of using artificial intelligence when interpreting the results of data analytics

Copyright © Assist Knowledge Development 2026

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 following terms:

- Data
- Information
- Big data
- Data analysis
- Data analytics
- Data model
- Business intelligence

1.2 Distinguish between the following types of data:

- Structured, unstructured and semi-structured data
- Static data and dynamic data
- Qualitative and quantitative data

1.3 Explain the following ways of representing data:

- Conceptual, logical and physical data models
- Static and dynamic views

1.4 Define the stages in the data lifecycle:

- Investigate data sources and requirements
- Define and model data requirements
- Obtain and record data (including validation and cleansing)
- Use data for operations and business decision making
- Dispose of data

2. Modelling data using Class Diagrams (30%: K4)

2.1 Define the following concepts used in class diagrams, and describe the associated notation used to represent them:

- Classes and objects
- The structure of a class: name, attributes, operations
- Associations
- Labelling associations
- Association class

Copyright © Assist Knowledge Development 2026

- Multiplicity
- Composition and aggregation
- Generalisation

2.2 Analyse a class diagram and describe the business rules represented.

2.3 Validate data models using a CRUD matrix.

2.4 Validate data models against requirements using data navigation paths (Data Navigation Diagrams).

3. Defining data requirements (10%: K3)

3.1 Define the concept of Metadata and recognise examples of the following types:

- Structural metadata
- Descriptive metadata
- Statistical metadata

3.2 Explain the purpose and contents of a Domain definition.

3.3 Explain the rationale for data normalisation and describe common problems associated with unnormalised data:

- Redundant duplication of data
- Data inconsistencies
- Update anomalies

3.4 Define the following normalisation concepts:

- Primary keys (to uniquely identify data)
- Foreign keys (to represent relationships between different data groupings (classes/relations))
- Simple, compound and hierarchic keys
- Unnormalised form, first normal form, second normal form and third normal form relations
- Third Normal Form data model

3.5 Identify the following core dimensions of data quality:

- Completeness
- Uniqueness
- Timeliness
- Validity
- Accuracy
- Consistency

4. Sourcing data (10%: K3)

4.1 Identify the following best practices when sourcing data:

- Define the purpose
- Ensure data quality
- Respect privacy and consent
- Use secure and ethical sources
- Document data provenance
- Consider bias and representation
- Review legal and licencing terms

4.2 Identify a variety of different sources of data:

- Research
- Operational business data from applications, documents, spreadsheets and databases
- Sensor and internet of things (IoT) data
- Web and social media data
- Public and open source
- Pay-per-use and subscription data services

4.3 Describe the purpose of validating data and typical activities involved in cleansing data:

- Determining data lineage and provenance
- Reformatting data
- Handling duplicated and missing data
- Correcting erroneous or inconsistent data

5. Analysing data for decision-making (30%: K4)

5.1 Describe the following uses of data within organisations:

- Informed decision making: operational, tactical, strategic
- Performance of operational processes and tasks
- Improving operational effectiveness
- Understanding the customer
- Identifying opportunities for competitive advantage

5.2 Analyse a scenario and identify the following potential problems and their impact on analysing data:

- Confirmation bias
- Sample size
- Outliers
- Consistency

5.3 Apply the following dataset calculations:

- Counting
- Totalling
- Averages: mean, media, mode and standard deviation
- Probability

5.4 Explain the characteristics and applications of the following data relationships in analytics:

- Regression analysis
- Correlation and causation
- Time-series forecasting

5.5 Describe the role of visualisation when interpreting the results of data analytics:

- Translate complex data into intuitive, graphical formats (charts, graphs, dashboards)
- Help identify patterns, trends and outliers
- Makes data accessible to non-technical audiences

5.6 Describe the role of storytelling when interpreting the results of data analytics:

- Provides context and narrative around the data
- Connects insights to real-world implications
- Helps the audience answer questions like ‘why?’ and ‘so what?’
- Enhances retention and actionability of data-driven insights

5.7 Identify where Artificial Intelligence (AI) can add value when interpreting the results of data analytics:

- Descriptive, predictive and real-time analytics
- Natural Language Processing (NLP) for querying, sentiment analysis and text analysis
- Pattern recognition and clustering

5.8 Explain the following challenges when interpreting the results of data analytics:

- Misleading presentation of data
- Data veracity
- Bias
- Confidence levels

6. Protecting data (10%: K2)

6.1 Explain the CIA triad (confidentiality, integrity, availability) and its relevance to data protection.

6.2 Define Personally Identifiable Information (PII).

6.3 Describe the following roles in data protection.

- Data subject
- Data controller
- Data processor

6.4 Explain the following data protection principles:

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality (security)
- Accountability

6.5 Define data ethics.

6.6 Identify a range of issues associated with the ethical use of data:

- The impact on people (privacy, bias, fairness and consent)
- The potential for misuse (surveillance, manipulation and data breaches)
- The economic value associated with the data (monetisation of personal data and the need to balance profit with responsible data stewardship)

Copyright © Assist Knowledge Development 2026

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	30%	Multiple choice / Scenario
3. Defining data requirements	10%	Multiple choice / Scenario
4. Sourcing data	10%	Multiple choice / Scenario
5. Analysing data for decision-making	30%	Multiple choice / Scenario
6. Protecting data	10%	Multiple choice
Total	**100%**	

Copyright © Assist Knowledge Development 2026

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: Defining Digital Solutions
Author: Peter Thompson.
Publisher: BCS, Learning and Development Limited
Publication Date: April 2025
ISBN: 9781780177052

Title: Designing Digital Solutions
Author: Peter Thompson.
Publisher: BCS, Learning and Development Limited
Publication Date: May 2025
ISBN: 9781780177083

Title: Artificial Intelligence and Machine Learning Foundations
Author: Andrew Lowe and Steve Lawless.
Publisher: BCS, Learning and Development Limited
Publication Date: Feb 2021
ISBN: 9781780176734

Copyright © Assist Knowledge Development 2026

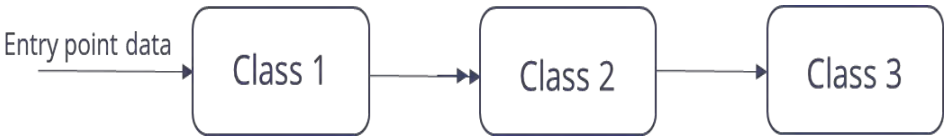
https://ico.org.uk/for-organisations/

Glossary

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	A type of association that represents a 'whole–part' relationship between two classes. It is used to show that one class (the whole) is made up of one or more instances of another class (the parts). Aggregation indicates a weaker relationship compared to COMPOSITION, another type of whole–part relationship.
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.
BIG DATA	Extremely large or complex datasets that exceed the capabilities of traditional data processing tools and techniques. Big data is characterised by the '3 Vs': volume (massive scale of data), velocity (rapid ingestion, generation and processing) and variety (different types and formats). Advanced analytics and specialised technologies are often used to extract insights from big data and inform strategic decision-making.
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).

Copyright © Assist Knowledge Development 2026

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.
CONFIDENCE LEVEL	The probability that a statistical estimate (e.g. a sample mean) accurately reflects the true population parameter, within a given CONFIDENCE INTERVAL.
CONFIDENCE INTERVAL	A range within which the true population parameter (e.g. mean) is expected to fall.
CONFIRMATION BIAS	The tendency to search for and interpret information in order to confirm or support a particular viewpoint or set of beliefs.
CONSISTENCY	A data quality that is measured by comparing data values against their meta-data definitions to assure that the values share the same underlying meaning and formatting.
COMPOUND KEY	A concatenated key where each element of the key is a unique identifier and also acts as a foreign key.
COMPOSITION	A strong form of the AGGREGATION association that defines a whole-part relationship between two classes, where the parts are entirely dependent on the whole. In this association, part objects cannot exist independently—if the whole is destroyed, the parts are also destroyed. It is typically used to model scenarios where the lifecycle of the parts is tightly bound to that of the whole.
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 CONTROLLER	A person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.

DATA ETHICS	The principles that govern the sourcing, recording, processing and disposal of data. These include: • Ownership • Transparency • Privacy • Intention • Outcomes
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	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 though 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. 
DATA PROCESSOR	A person, public authority, agency or other body which processes personal data on behalf of the controller.
DATA PROTECTION	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.
DATA SUBJECT	The identified or identifiable living individual to whom personal data relates.
DATA QUALITY	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.

Copyright © Assist Knowledge Development 2026

DESCRIPTIVE ANALYTICS	The process of using current and historical data to identify trends and relationships.
DOMAIN	A defined format and set of values that is applied to a set of ATTRIBUTES.
DYNAMIC DATA	Data that changes frequently or in real time, often reflecting ongoing processes or updates.
ENTITY	An entity is a grouping of the attributes that describe something of interest within the business system.
ENTITY RELATIONSHIP MODEL	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.
FOREIGN KEY	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.
GENERALISATION	A relationship that abstracts common features from two or more entities into a single, more general entity or superclass. Properties (attributes, operations and associations) are inherited by the subclasses from the superclass.
HIERARCHIC KEY	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.
HYPOTHESIS	A proposed explanation or assumption that can be tested through analysis or experimentation to determine its validity.
INFORMATION	Facts, concepts or instructions that inform in order to aid communication, interpretation and processing.
KEY	A unique identifier for a relation that comprises one or more data items (attributes).
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 an ordered set of values.
METADATA	Data that provides information about other data. It serves as a descriptive summary or contextual information about a file, record, dataset or digital resource, making it easier to understand, organise, search and use. There are three common types of metadata: • Structural metadata - describes the structural elements related to data containers (such as classes or entities) and the

	relationships between them. • Descriptive metadata – describes the content, context, and characteristics of a data resource, making it easier to identify, discover, and understand. • Statistical metadata – describes quantitative information about a data resource such as predicted volumes and sizes of data that can be compared with actual volumes and sizes. Metadata is often used in disciplines such as content management, where the content itself is the data and descriptive metadata is used to describe that content.
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.
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.
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	Something within a business system for which a set of attributes (properties) and operations (actions) can be specified. An object is an instance of a class.
OPERATION	A specific action that can be performed on an object.
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.
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.
POPULATION	The entire set of individuals, items, or data points that share common characteristics and from which a sample may be drawn for analysis.
PREDICTIVE ANALYTICS	The use of data to predict future trends and events.
PRIMARY KEY	An attribute or combination of multiple attributes within a relation or table, the value of which is used to identify a single row.
REAL-TIME ANALYTICS	The process of analysing data as it becomes available.
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.
SAMPLE SIZE	The number of individual observations or data points selected from a population to form a sample for analysis.
SAMPLING	The process of selecting a subset of individuals, items, or data points from a larger POPULATION to analyse and draw conclusions about the whole population.
SEMI-STRUCTURED DATA	Data that does not follow a strict tabular format but contains some organisational structure, such as tags or markers, making it easier to interpret than completely unstructured data. Common in formats like JSON, XML, and CSV.
SIMPLE KEY	A unique identifier for a relation comprised of just one attribute.
STANDARD DEVIATION	A statistical measure that quantifies the amount of variation or dispersion in a set of values.
STATIC DATA	Data that remains fixed and does not change after it is created or recorded.
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.
VERACITY	The accuracy, reliability, and truthfulness of data, indicating how trustworthy and credible the information derived from it is.

Copyright © Assist Knowledge Development 2026