



BCS Level 4 Diploma in Data Analysis Concepts QAN 603/0823/0

**Version 4
July 2020**

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

BCS Level 4 Diploma in Data Analysis Concepts

Contents

- Introduction5
- Objectives5
- Course Format and Duration5
- Eligibility for the Examination.....6
- Format and Duration of the Examination6
- Additional time for learners requiring Reasonable Adjustments due to a disability6
- Additional time for learners whose language is not the language of the examination.....6
- Guidelines for Training Providers.....6
- Syllabus8
- Levels of Knowledge / SFIA Levels 16
- Question Weighting 16
- Format of Examination 17
- Trainer Criteria 17
- Classroom Size 17
- Recommended Reading List 18
- Useful Websites 19

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
Version 1.0 December 2016	Syllabus Created
Version 1.1 December 2016	Compliance statement added to title.
Version 2.0 July 2017	Amendments following review
Version 3.0 June 2020	Update to DPA references in LO's 7.2 and 7.3.
Version 4 July 2020	Document amended to make it suitable for a range of different learners as well as apprentices

Introduction

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

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

Objectives

Learners should be able to demonstrate knowledge and understanding of Data Analysis and its underlying architecture, principles, and techniques. Key areas are:

1. Explore the different types of data, including open and public data, administrative data, and research data
2. Understand the data lifecycle
3. Illustrate the differences between structured and unstructured data
4. Understand the importance of clearly defining customer requirements for data analysis
5. Understand the quality issues that can arise with data and how to avoid and/or resolve these
6. Explore the steps involved in carrying out routine data analysis tasks
7. Understand the range of data protection and legal issues
8. Explore the fundamentals of data structures
9. Explore the database system design, implementation, and maintenance
10. Understands the organisation's data architecture
11. Understands the importance of the domain context for data analytics

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

Course Format and Duration

Candidates can study for this diploma by attending a training course provided by a BCS accredited Training Provider. The estimated total qualification time for this diploma is 600 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.

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

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.

Types of Data (10%, K2)

In this topic, the learner will explore the different types of data, including open and public data, administrative data, and research data. The successful learner should be able to:

- 1.1 Describe the differences between data (raw or unorganised facts), information (processed data to make it useful) and knowledge (understanding of information).
 - Typical formats and sources are: CSV, XML, RTF, TXT and File.
 - Benefits and limitations.
 - Database transformations needed of each type are organisation, structuring and processing (or Concept, mapping, matching).

- 1.2 Understand and explain the range of different types of data and the implications for allowable use, data quality, privacy concerns and availability.
 - Open and public vs. proprietary data.
 - Operational (data used in the day-to-day business operations) vs. administrative data (data used for the administration and management).
 - Research data.

- 1.3 Understand the importance of data classification and describe how to classify data which are:
 - Structured and unstructured data.
 - Quantitative data.
 - Categorical (Discrete)
 - Continuous (Time-series)
 - Qualitative data.
 - Binomial
 - Nominal
 - Ordinal

The Data Lifecycle (5%, K2)

In this topic area, the learner will explore the data lifecycle. The successful candidate should be able to:

- 2.1 Understand and describe how the flow of an information system's data and associated metadata follows a lifecycle.
- 2.2 Explain each of the stages of a data lifecycle, which are:
 - Creation;
 - Initial storage;
 - Archived;
 - Obsolete;
 - Deleted.

3. Structured and Unstructured Data (10%, K2)

In this topic area, the learner will explain the differences between structured and unstructured data. The successful candidate should be able to:

- 3.1 Describe that structured data is information which can be ordered and processed by data analysis tools.
- 3.2 Recognise common sources of structured data:
 - Data files organised sequentially or organised serially.
 - Tables stored within a database management system.
 - Extensible Markup Language.
- 3.3 Explain that unstructured data can take various formats:
 - Word processor, spreadsheet and PowerPoint files;
 - Audio;
 - Video;
 - Sensor and log data;
 - External data (such as social media feeds);
 - Paper-based documents.
- 3.4 Recognise how structured and unstructured data could complement each other to derive rich insight.
 - Enhance analysis of the other (Structured or Unstructured text data).
 - Combined into a common model.
 - Big data analytics.
- 3.5 Understand the importance of being able to rapidly analyse structured and unstructured data to maximise insight for the business.

4. Requirements for Data Analysis (15%, K2)

In this topic area, the learner will show the importance of clearly defining customer requirements for data. The successful candidate should be able to:

- 4.1 Recognise and understand why data does not provide the answers to business problems.
- 4.2 Understand the customer requirements and recognise the best way to obtain the relevant information through:
 - Classifying different types of requirements:
 - General requirements, such as business policies and standards
 - Technical requirements
 - Explain the difference between validation and verification.
- 4.3 Explain the requirements elicitation process.
 - Documentation included / used.
 - Explicit vs. tacit knowledge.
 - Different elicitation techniques. For example, apprentice, observe, recount, enact.
- 4.4 Recognise and interpret various data models used in the requirements gathering process
 - Recognise and interpret logical, physical, and conceptual data models.

5. Quality Issues for Data Analysis (10%, K2)

In this topic area, the learner will develop an understanding of the quality issues that can arise with data and how to avoid and/or resolve issues experienced. The successful candidate should be able to:

- 5.1 Understand the importance and necessity of good quality data in respect to:
 - Legal and regulatory compliance.
 - Commercial and intellectual property.
 - Confidentiality, integrity, and availability.
- 5.2 Identify the common sources of errors (such as completeness, uniqueness, timeliness, accuracy, and consistency) and how to avoid and/or resolve them through:
 - Entry / Transcription;
 - Process;
 - Identification;
 - Usage;
 - Validity;
 - Structure
- 5.3 Explain that minor data errors can cause major issues for data analysis:
Major issues are:
 - Cost;
 - Accuracy;
 - Inconsistency;
 - Cleanliness.

5.4 Understand that there will be a direct benefit to the value of data analytics through improving the data quality and having a defined organisational strategy for data creation and storage.

- Improved business decision making

6. Data Analysis Tasks (15%, K3)

In this topic area, the learner will explore the steps involved in carrying out routine data analysis tasks. The successful candidate should be able to:

6.1 List the typical routine steps of data analysis:

- Problem hypothesis;
- Identifying what to measure;
- Collect data;
- Cleanse data;
- Model data; Visualise data;
- Analyse data;
- Interpret results;
- Document and communicate results.

6.2 Understand and explain that routine data analysis includes creating a problem hypothesis and identifying what to measure.

- Creating a problem hypothesis:
 - Understanding the importance of null and alternative hypotheses
 - Understanding the subject area for analysis
 - Finding similar previous analysis and exploring existing definition, assumptions, and reconciliation requirements
- Identifying what to measure:
 - Selecting the data sources
 - Selecting aggregation and / or summarisation level

6.3 Understand and explain that routine data analysis includes clarification and confirmation of the requirement and identification of the right data and location through:

- Collecting data:
 - Understand the size, nature and content of the data
 - Identification of the data security and accessibility
 - Complete data extraction
 - Complete data transfer
- Data loading cleansing data through:
 - Filtering
 - Interpolation
 - Transformation
 - Masking
 - Blending

6.4 Understand and explain that routine data analysis includes modelling and visualising data.

- Modelling data:
 - Provide assumptions made to data
 - Train the model
 - Data preparation
 - Model definition
 - Validation / verification of predictive models
 - Troubleshooting
 - Validation testing
 - Strategy for improving model performance
 - Identifying and selecting an appropriate model
- Visualising data:
 - Understand which type of visual data is suitable for the customer:
 - Types of charts (such as line graph, column and bar charts, pie chart, scatter plot, histogram, radar / spider chart, waterfall chart)
 - Geospatial distributions such as heat maps, bubble maps
 - Time series such as time plot, Gantt chart
 - Unstructured data such as Word Cloud

6.5 Understand and explain that routine data analysis includes analysing data, interpreting, documenting, and communicating results.

- Analysing data:
 - Reconcile and compare with other sources
- Interpreting results:
 - Understand the relationship between variables
 - Show and compare the results in terms of real world objects
- Documenting and communicating results:
 - List the models and assumptions
 - Understand your customer and stakeholders needs and communication style

7. Compliance and Audit Considerations (5%, K2)

In this topic area, the learner will explore and gain knowledge on the range of data protection and legal issues. The successful candidate should be able to:

- 7.1 Describe the data protection and privacy issues that can occur during data analysis activities.
 - Discuss the types, formats and activities that are protected:
 - Personally Identifiable Information
 - Protected Health Information
- 7.2 Recall and describe the key principles of the Data Protection Act.
- 7.3 Explain the need to comply with the Data Protection Act.
 - Rights and obligations.
 - Enforcement agencies.
 - Regulatory and legal penalties.

8. Data Structures (10%, K3)

In this topic area, the learner will explore the fundamentals of data structures and database system design, implementation, and maintenance. The successful candidate should be able to:

- 8.1 Understand that data structure refers to different ways of describing different types of information.
 - Files;
 - Lists;
 - Arrays;
 - Records;
 - Trees;
 - Tables.
- 8.2 Identify that data structure refers to formalised ways of identifying, accessing, and manipulating data attributes by forming logical groupings of attributes into:
 - Files;
 - Lists;
 - Arrays;
 - Records;
 - Trees;
 - Tables.

9. Database Design, Implementation, and Maintenance (10%, K3)

In this topic area, the learner will explore database system design, implementation, and maintenance. The successful candidate should be able to:

- 9.1 Apply data modelling techniques within database design, producing data models from different perspectives including:
- Conceptual;
 - Logical;
 - Physical.
- 9.2 Recognise the most common forms of database, including:
- Relational;
 - Hierarchical;
 - Network;
 - Object-oriented;
 - Multi-dimensional (data cubes and hypercubes);
 - NoSQL.
- 9.3 Demonstrate how a logical data model can be transformed into a physical database design, including normalisation and de-normalisation:
- Normalisation:
 - Redundancy free
 - Unambiguous
 - Flexible / extensible
 - De-normalisation:
 - Introduction of derivable data (cumulative values, flags / status values)
 - Splitting logical data structures
 - Combining logical data structures
 - Introducing potentially redundant relationships
- 9.4 Recognise that database maintenance is an activity designed to keep a database running smoothly and that a database can become sluggish and otherwise lose functionality.
- Log file maintenance;
 - Data compaction;
 - Defragmentation;
 - Integrity Check;
 - Data Warehousing.
- 9.5 Explain the importance of maintaining a database by backing up the data securely.

10. Data Architecture (5%, K2)

In this topic area, the learner will understand the organisation's data architecture. The successful candidate should be able to:

10.1 Explain how an organisation's data architecture defines how data is stored, managed, used and integrated within an organisation and its database systems.

- Understand that it typically comprises of:
 - Rules
 - Policies
 - Standards
 - Models

10.2 Define the nature of the Data Architecture functions.

- Describe why a business needs to map its data to its infrastructure, applications, and business functions.
- Explain the need for definition and governance of data architectures.
- Data Architecture Functions:
 - Data migration
 - Data modelling
 - Data integration
 - Data warehousing
 - Database design

11. The Domain Context for Data Analytics (5%, K2)

In this topic area, the learner will explore the importance of the domain context for data analytics. The successful candidate should be able to:

11.1 Understand and explain the importance of having relevant domain (industry / organisation) knowledge to enable effective data analysis.

11.2 Describe the role of:

- Decision analytics;
- Descriptive analytics;
- Predictive analytics;
- Prescriptive analytics.

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. 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)
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

Question Weighting

Syllabus Area	Target number of questions
1. Types of data	4
2. The data lifecycle	2
3. Structured and unstructured data	4
4. Requirements for data analysis	6
5. Quality issues for data analysis	4
6. Data analysis tasks	6
7. Compliance and audit considerations	2
8. Data structures	4
9. Database design, implementation, and maintenance	4
10. Data architecture	2
11. The domain context for data analytics	2
Total	40 Questions

Format of Examination

Type	40 Question Multiple Choice.
Duration	1 Hour. An additional 15 minutes 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)	600 Hours 400 GLH recommended.
Delivery	Online

Trainer Criteria

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

Classroom Size

Trainer to Apprentice ratio	1:16
-----------------------------	------

Recommended Reading List

Title: [Business Analysis](#)
Author: Cadle, J et al.
Publisher: BCS, The Chartered Institute for IT; 3rd edition
Publication Date: 22 Sept. 2014
ISBN-10: 178017277X
ISBN-13: 978-1780172774

Title: [Data Mining: Concepts and Techniques \(The Morgan Kaufmann Series in Data Management Systems\)](#)
Author: Han, J & Kamber, M
Publisher: Morgan Kaufmann; 3 edition
Publication Date: 25 July 2011
ISBN-10: 9380931913
ISBN-13: 978-9380931913

Title: [Data Science for Business: What you need to know about data mining and data-analytic thinking](#)
Author: Provost, F. & Fawcett, T.
Publisher: O'Reilly Media; 1 edition
Publication Date: 19 August 2013
ISBN-10: 1449361323
ISBN-13: 978-1449361327

Title: [Data Warehouse Design: Modern Principles and Methodologies](#)
Author: Golfarelli, M & Rizzi, S
Publisher: McGraw-Hill Education
Publication Date: 1 July 2009
ISBN-10: 0071610391
ISBN-13: 978-0071610391

Title: [The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data](#)
Author: Kimball, R & Caserta, J
Publisher: John Wiley & Sons; 1 edition
Publication Date: 24 Sept. 2004
ISBN-10: 0764567578
ISBN-13: 978-0764567575

Title: [The Data Warehouse Lifecycle Toolkit](#)
Author: Kimball, R et al.
Publisher: John Wiley & Sons; 2nd Revised edition
Publication Date: 4 Jan. 2008
ISBN-10: 0470149779
ISBN-13: 978-0470149775

Title: [Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods \(Business in the Digital Economy\)](#)
Author: Finlay, S.
Publisher: Palgrave Macmillan; 2014 edition
Publication Date: 1 July 2014
ISBN-10: 1137379278
ISBN-13: 978-1137379276

Title: [Storytelling with Data: A Data Visualization Guide for Business Professionals](#)
Author: Cole Nussbaumer Knaflic
Publisher: John Wiley & Sons; 1 edition
Publication Date: 20 November 2015
ISBN-10: 1119002257
ISBN-13: 978-1119002253

Useful Websites

<http://www.kdnuggets.com/websites/index.html>

<https://www.sisense.com/blog/11-websites-every-data-analyst-should-follow/>

<http://www.datasciencecentral.com/profiles/blogs/my-top-websites-for-big-data>

<http://www.kimballgroup.com/category/articles-design-tips/>