



Qualification Specification for the Knowledge Modules that can form part of the BCS Level 4 Data Analyst Apprenticeship

BCS Level 4 Certificate in Data Analysis Tools

QAN: 603/0824/2

BCS Level 4 Diploma in Data Analysis Concepts

QAN: 603/0823/0

Version 4.1

December 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.

Contents

1. About BCS	3
2. Equal Opportunities	3
3. Introduction to the Qualification	3
3.1 Qualification summary	3
3.2 Purpose of the qualifications	4
3.3 Structure of the qualifications	4
3.4 Prior learning	5
3.5 Learner progression	5

4. Units	6
4.1 Guidance on the qualifications' content	6
4.2 Learning Outcomes and Assessment Criteria	7

5. Assessment	26
5.1 Summary of assessment methods	26
5.2 Availability of assessments	26
5.3 Grading	26
5.4 Externally assessed units	26
5.5 Specimen assessment materials	26
5.6 Support materials	26
5.7 Access to Assessment	27

6. Contact Points	27
--------------------------	-----------

1. About BCS

Our mission as BCS, The Chartered Institute for IT, is to enable the information society. We promote wider social and economic progress through the advancement of information technology, science and practice. We bring together industry, academics, practitioners and government to share knowledge, promote new thinking, design new curricula, shape public policy and inform the public.

Our vision is to be a world class organisation for IT. Our 70,000 strong membership includes practitioners, businesses, academics and students in the UK and internationally. We deliver a range of professional development tools for practitioners and employees. As a leading IT qualification body, we offer a range of widely recognised qualifications.

2. Equal Opportunities

BCS wishes to ensure good practice in the area of Equal Opportunity. Equality of opportunity extends to all aspects for the provision of BCS qualifications.

3. Introduction to the Qualification

3.1 Qualification summary

Qualification Title	QAN	Accreditation Start
1. BCS Level 4 Certificate in Data Analysis Tools	603/0824/2	9/12/16
2. BCS Level 4 Diploma in Data Analysis Concepts	603/0823/0	9/12/16

The two knowledge module qualifications listed above have been developed based on the requirements set out in the Standard issued and approved by the Institute for Apprenticeships & Technical Education, details of which can be located in the Assessment Plan (Click [here](#)) and Occupational Brief (Click [here](#)) documents.

Apprenticeship learners:

Apprentices must achieve the vendor or professional qualification, from the right hand column in the table below (currently this is only the EMC: Data Science Associate). This exempts the knowledge module in the left hand column. Apprentices must also achieve the Data Analysis Concepts Knowledge Module.

The knowledge modules are summarised below, and further details are available in the occupational brief available from NSAR at:

<https://www.nsar.co.uk/wp-content/uploads/2019/02/occupational-brief-data-analyst.pdf>

Knowledge Modules Vendor or Professional Qualifications.

BCS qualification	Vendor certification alternative chosen
BCS Level 4 Certificate in Data Analysis Tools	EMC: Data Science associate
BCS Level 4 Diploma in Data Analysis Concepts	

All BCS qualifications are subject to our quality assurance and validation process. This ensures that new and revised qualifications are fit for purpose. Qualifications are reviewed to ensure the alignment of the qualification with agreed design principles, regulatory requirements and to ensure accuracy and consistency across units and qualifications. Through our quality assurance and validation process, we ensure the qualification, its units and assessments are fit for purpose and can be delivered efficiently and reasonably by Training Providers.

3.2 Purpose of the qualifications

The qualifications are designed for apprentices enrolled on the Level 4 Data Analyst Digital IT Apprenticeship, to provide them with the technical knowledge and understanding they require for their role detailed below. 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.

The primary role of a Data Analyst is to collect, organise and study data to provide business insight. Data analysts are typically involved with managing, cleansing, abstracting and aggregating data, and conducting a range of analytical studies on that data. They work across a variety of projects, providing technical data solutions to a range of stakeholders/customers issues. They document and report the results of data analysis activities making recommendations to improve business performance. They have a good understanding of data structures, database systems and procedures and the range of analytical tools used to undertake a range of different types of analyses.

3.3 Structure of the qualifications

This document covers the following qualifications which are used towards the Level 4 Data Analyst Apprenticeship, and which may also be completed as standalone units. When completed as part of the Level 4 Data Analyst Apprenticeship, the qualifications can be taken in any order however it is recommended that they be completed in the following sequence:

1. BCS Level 4 Certificate in Data Analysis Tools
2. BCS Level 4 Diploma in Data Analysis Concepts

Qualification Level 4 Descriptor	
Knowledge descriptor (the holder...)	Learners will develop knowledge and understanding in all aspects of data analysis tools and concepts including; the range of data protection and legal issues; the lifecycle of data; the different types of data; the differences between structured and unstructured data; the fundamentals of data structures and designing, implementing and maintenance of database systems; the importance of the domain context for data analytics; quality issues and how to avoid/resolve them; the importance of clearly defining customer requirements; processes and tools used for data integration; the steps involved in routine data analysis tasks; how to use and apply industry standard tools and methods.
Skills descriptor (the holder can...)	Learners will develop skills and be able to demonstrate; logical and creative thinking skills; analytical and problem solving skills, ability to work independently and take responsibility; can use own initiative; a thorough and organized approach, ability to work with a range of people; ability to communicate effectively in a variety of situations; maintain productive, professional and secure working environment.

3.4 Prior learning

Apprenticeship learners:

Individual employers will set the selection criteria for enrolment onto the Apprenticeship, but this is likely to include five GCSEs, (especially English, Mathematics and a Science or Technology subject); a relevant Level 3 Apprenticeship; other relevant qualifications and experience; or an aptitude test with a focus on IT skills.

Other learners:

It is recommended that learners have completed five 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.

3.5 Learner progression

Apprenticeship learners:

This document covers the qualifications that are part of the Level 4 Data Analyst Apprenticeship. The qualifications must be completed to allow the apprentice to progress onto the end-point assessment, detailed below.

The final end-point assessment is completed in the last few months of the apprenticeship. It is based on:

- a portfolio – produced towards the end of the apprenticeship, containing evidence from real work projects which have been completed during the apprenticeship, usually towards the end and which, taken together, cover the totality of the standard and which is assessed as part of the end-point assessment.
- a project - giving the apprentice the opportunity to undertake a business-related project over a one-week period away from the day to day workplace.
- an employer reference.
- a structured interview with an assessor - exploring what has been produced in the portfolio and the project, as well as looking at how it has been produced.

An independent assessor will evaluate each element of the end-point assessment and will then decide whether to award successful apprentices with a pass, a merit or a distinction.

Other learners:

For learners taking this module as a standalone unit, there is not a mandatory progression route. However, learners will benefit from further study at level 4 onwards, provided through the BCS qualification portfolio.

4. Units

4.1 Guidance on the qualifications' content

The content for each qualification has been developed based on the criteria set out in the Occupational Brief.

Qualification Title	TQT (Guided Learning + Direct Study + Assessment)
1. BCS Level 4 Certificate in Data Analysis Tools	145h (80h + 64h + 1h)
2. BCS Level 4 Diploma in Data Analysis Concepts	600h (400h + 199h + 1h)

4.2 Learning Outcomes and Assessment Criteria

Qualification Name	Learning Outcomes The learner will....	Assessment Criteria The learner can...
BCS Level 4 Certificate in Data Analysis Tools	1. Describe how data integration is achieved through the manipulation of data from different sources. They will also learn about how this data is manipulated using programming languages and how it is prepared for analysis.	1.1 Explain the purpose and outputs of data integration activities. <ul style="list-style-type: none"> • Function requirements • Non-functional requirements (such as speed and time available) • Information structure and rules relevant to the business (policies and practices) • Rationale for using and integrating data from multiple sources • Importance of data in a business context

		<p>1.2 Explain how data from multiple sources can be integrated to provide a unified view of the data.</p> <ul style="list-style-type: none">• 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<ul style="list-style-type: none">○ Common user interface<ul style="list-style-type: none">- Dashboard- Scorecard- Dynamic○ Virtual integration<ul style="list-style-type: none">- Communication channels- Data transfer○ Physical data integration<ul style="list-style-type: none">- ETL (Extract - Transform - Load)
--	--	--

--	--	--

		<p>1.3 Describe how programming languages for statistical computing (SQL) can be applied to data integration activities, improving speed and data quality for analysis.</p> <ul style="list-style-type: none">• Programming constructs<ul style="list-style-type: none">○ Sequence, selection and iteration• Single queries;• Multiple queries (UNION);• Expressions;<ul style="list-style-type: none">- CASE- DATETIME- Function- Compound• Functions;<ul style="list-style-type: none">- Avg- Count- Max- Min- Group by- Round- Cast- Convert- ISNULL• Querying multiple tables in different information<ul style="list-style-type: none">○ Joins<ul style="list-style-type: none">- Inner and outer- Right and left- Full
--	--	--

		<ul style="list-style-type: none">- Union- Select into- subqueries • Joins with duplicate values;• Joining on multiple fields;• Select and select* statements;• From;• Where;<ul style="list-style-type: none">- AND- OR- Use of wildcards and ordering• Selecting the first/last of occurrences;• Implicit data conversion
--	--	--

		<p>1.4 Explain how to take account of data quality in preparing data for analysis to improve accuracy, quality and usefulness.</p> <ul style="list-style-type: none">• Data profiling;• What happens when an error or issue is found;<ul style="list-style-type: none">- Accept- Reject- Correct error- Create default value• Data quality dimensions.<ul style="list-style-type: none">- 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
 - Data quality
 - Security
 - Performance
 - Maintenance

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

- Business testing & technical testing
 - Technical acceptance testing (TAT)
 - User acceptance testing (UAT)
 - Performance stress tests (PST)

	<p>2. Describe and use a range of tools, techniques and methods to prepare and analyse data.</p>	<p>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.</p> <ul style="list-style-type: none"> • Capabilities and functions of statistical programming language; <ul style="list-style-type: none"> - R • Programming language; <ul style="list-style-type: none"> - Python • Relational databases; <ul style="list-style-type: none"> - SQL • Non-relational databases; <ul style="list-style-type: none"> - Graph - Document - Column - Family • Software tools. <ul style="list-style-type: none"> - Excel
--	--	---

		<p>2.2 Explain how statistical programming languages are used in preparing data for analysis and within analysis projects.</p> <ul style="list-style-type: none">• Preparation techniques (such as, but not limited to; searching and sorting, grouping, filtering, modelling)• Data cleaning to remove a range of data issues (such as, but not limited to; errors, missing data, inconsistencies, redundancy)• Processing and analysing:<ul style="list-style-type: none">○ Mean, Median, Mode and Range○ Probability○ Bias o Statistical significance○ Linear Regression (simple & multiple)○ Scatter plots and correlation○ And/Or probability○ Stem and leaf plots (frequency and distribution)○ Factorials○ Box and whisker plots• Methods for presenting results (such as, but not limited to; tables, charts and graphs)
--	--	---

Qualification Name	Learning Outcomes The learner will....	Assessment Criteria The learner can...
BCS Level 4 Diploma in Data Analysis Concepts	1. Explore the different types of data, including open and public data, administrative data, and research data.	<p>1.1 Describe the differences between data (raw or unorganised facts), information (processed data to make it useful) and knowledge (understanding of information).</p> <ul style="list-style-type: none"> • 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). <hr/> <p>1.2 Understand and explain the range of different types of data and the implications for allowable use, data quality, privacy concerns and availability.</p> <ul style="list-style-type: none"> • 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.

		<p>1.3 Understand the importance of data classification and describe how to classify data which are:</p> <ul style="list-style-type: none"> • Structured and unstructured data. • Quantitative data. <ul style="list-style-type: none"> ○ Categorical (Discrete) ○ Continuous (Time-series) • Qualitative data. <ul style="list-style-type: none"> ○ Binomial ○ Nominal ○ Ordinal
	<p>2. Explore the data lifecycle.</p>	<p>2.1 Understand and describe how the flow of an information system's data and associated metadata follows a lifecycle.</p> <p>2.2 Explain each of the stages of a data lifecycle, which are:</p> <ul style="list-style-type: none"> • Creation; • Initial storage; • Archived; • Obsolete; • Deleted.
	<p>3. Explain the differences between structured and unstructured data.</p>	<p>3.1 Describe that structured data is information which can be ordered and processed by data analysis tools.</p> <p>3.2 Recognise common sources of structured data:</p> <ul style="list-style-type: none"> • Data files organised sequentially or organised serially. • Tables stored within a database management system. • Extensible Markup Language.

		<p>3.3 Explain that unstructured data can take various formats:</p> <ul style="list-style-type: none"> • Word processor, spreadsheet and PowerPoint files; • Audio; • Video; • Sensor and log data; • External data (such as social media feeds); • Paper-based documents.
		<p>3.4 Recognise how structured and unstructured data could complement each other to derive rich insight.</p> <ul style="list-style-type: none"> • Enhance analysis of the other (Structured or Unstructured text data). • Combined into a common model. • Big data analytics.
		<p>3.5 Understand the importance of being able to rapidly analyse structured and unstructured data to maximise insight for the business.</p>
	<p>4. Show the importance of clearly defining customer requirements for data.</p>	<p>4.1 Recognise and understand why data does not provide the answers to business problems.</p>
		<p>4.2 Understand the customer requirements and recognise the best way to obtain the relevant information through:</p> <ul style="list-style-type: none"> • Classifying different types of requirements: <ul style="list-style-type: none"> ○ General requirements, such as business policies and standards ○ Technical requirements • Explain the difference between validation and verification.

		<p>4.3 Explain the requirements elicitation process.</p> <ul style="list-style-type: none"> • Documentation included / used. • Explicit vs. tacit knowledge. • Different elicitation techniques. For example, apprentice, observe, recount, enact.
		<p>4.4 Recognise and interpret various data models used in the requirements gathering process</p> <ul style="list-style-type: none"> • Recognise and interpret logical, physical, and conceptual data models.
	<p>5. Develop an understanding of the quality issues that can arise with data and how to avoid and/or resolve issues experienced.</p>	<p>5.1 Understand the importance and necessity of good quality data in respect to:</p> <ul style="list-style-type: none"> • Legal and regulatory compliance. • Commercial and intellectual property. • Confidentiality, integrity, and availability. <p>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:</p> <ul style="list-style-type: none"> • Entry / Transcription; • Process; • Identification; • Usage; • Validity; • Structure.

		<p>5.3 Explain that minor data errors can cause major issues for data analysis: Major issues are:</p> <ul style="list-style-type: none"> • Cost; • Accuracy; • Inconsistency; • Cleanliness. <p>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.</p> <ul style="list-style-type: none"> • Improved business decision making.
	<p>6. Explore the steps involved in carrying out routine data analysis tasks.</p>	<p>6.1 List the typical routine steps of data analysis:</p> <ul style="list-style-type: none"> • Problem hypothesis; • Identifying what to measure; • Collect data; • Cleanse data; • Model data; • Visualise data; • Analyse data; • Interpret results; • Document and communicate results.

		<p>6.2 Understand and explain that routine data analysis includes creating a problem hypothesis and identifying what to measure.</p> <ul style="list-style-type: none">• Creating a problem hypothesis:<ul style="list-style-type: none">○ 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:<ul style="list-style-type: none">○ Selecting the data sources○ Selecting aggregation and / or summarisation level
--	--	--

		<p>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:</p> <ul style="list-style-type: none">• Collecting data:<ul style="list-style-type: none">○ 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:<ul style="list-style-type: none">○ Filtering○ Interpolation○ Transformation○ Masking○ Blending
--	--	--

		<p>6.4 Understand and explain that routine data analysis includes modelling and visualising data.</p> <ul style="list-style-type: none">• Modelling data:<ul style="list-style-type: none">○ Provide assumptions made to data○ Train the model<ul style="list-style-type: none">▪ Data preparation▪ Model definition○ Validation / verification of predictive models<ul style="list-style-type: none">▪ Troubleshooting▪ Validation testing▪ Strategy for improving model performance○ Identifying and selecting an appropriate model• Visualising data:<ul style="list-style-type: none">○ Understand which type of visual data is suitable for the customer:<ul style="list-style-type: none">▪ 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
--	--	---

		<p>6.5 Understand and explain that routine data analysis includes analysing data, interpreting, documenting, and communicating results.</p> <ul style="list-style-type: none"> • Analysing data: <ul style="list-style-type: none"> ○ Reconcile and compare with other sources • Interpreting results: <ul style="list-style-type: none"> ○ Understand the relationship between variables ○ Show and compare the results in terms of real world objects • Documenting and communicating results: <ul style="list-style-type: none"> ○ List the models and assumptions ○ Understand your customer and stakeholders needs and communication style
	<p>7. Explore and gain knowledge on the range of data protection and legal issues.</p>	<p>7.1 Describe the data protection and privacy issues that can occur during data analysis activities.</p> <ul style="list-style-type: none"> • Discuss the types, formats and activities that are protected: <ul style="list-style-type: none"> ○ Personally Identifiable Information ○ Protected Health Information <p>7.2 Recall and describe the 8 principles of the Data Protection Act.</p> <p>7.3 Explain the need to comply with the Data Protection Act 1998 UK.</p> <ul style="list-style-type: none"> • Rights and obligations. • Enforcement agencies. • Regulatory and legal penalties.

	<p>8. Explore the fundamentals of data structures and database system design, implementation, and maintenance.</p>	<p>8.1 Understand that data structure refers to different ways of describing different types of information.</p> <ul style="list-style-type: none"> • Files; • Lists; • Arrays; • Records; • Trees; • Tables.
	<p>8.2 Identify that data structure refers to formalised ways of identifying, accessing, and manipulating data attributes by forming logical groupings of attributes into:</p> <ul style="list-style-type: none"> • Files; • Lists; • Arrays; • Records; • Trees; • Tables. 	
	<p>9. Explore database system design, implementation, and maintenance.</p>	<p>9.1 Apply data modelling techniques within database design, producing data models from different perspectives including:</p> <ul style="list-style-type: none"> • Conceptual; • Logical; • Physical.

		<p>9.2 Recognise the most common forms of database, including:</p> <ul style="list-style-type: none"> • Relational; • Hierarchical; • Network; • Object-oriented; • Multi-dimensional (data cubes and hypercubes); • NoSQL. <hr/> <p>9.3 Demonstrate how a logical data model can be transformed into a physical database design, including normalisation and de-normalisation:</p> <ul style="list-style-type: none"> • Normalisation: <ul style="list-style-type: none"> ○ Redundancy free ○ Unambiguous ○ Flexible / extensible • De-normalisation: <ul style="list-style-type: none"> ○ Introduction of derivable data (cumulative values, flags / status values) ○ Splitting logical data structures ○ Combining logical data structures ○ Introducing potentially redundant relationships
--	--	--

		<p>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.</p> <ul style="list-style-type: none"> • Log file maintenance; • Data compaction; • Defragmentation; • Integrity Check; • Data Warehousing.
		<p>9.5 Explain the importance of maintaining a database by backing up the data securely.</p>
	<p>10. Understand the organisation's data architecture.</p>	<p>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.</p> <ul style="list-style-type: none"> • Understand that it typically comprises of: <ul style="list-style-type: none"> ○ Rules ○ Policies ○ Standards ○ Models

		<p>10.2 Define the nature of the Data Architecture functions.</p> <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> ○ Data migration ○ Data modelling ○ Data integration ○ Data warehousing ○ Database design
	<p>11. Explore the importance of the domain context for data analytics.</p>	<p>11.1 Understand and explain the importance of having relevant domain (industry / organisation) knowledge to enable effective data analysis.</p> <p>11.2 Describe the role of:</p> <ul style="list-style-type: none"> ○ Decision analytics; ○ Descriptive analytics; ○ Predictive analytics; ○ Prescriptive analytics.

11 Assessment

11.2 Summary of assessment methods

The qualification is assessed in controlled exam conditions by a one-hour multiple-choice examination, consisting of 40 questions.

The exams are externally marked.

11.3 Availability of assessments

To be able to offer BCS Qualifications, you need to become a BCS Approved Training Provider.

All staff members who are involved in the management, invigilation and training must be registered with BCS. Suitably qualified individuals may be registered for more than one role. At least two members of staff must be registered with BCS in one of the roles in order for the Training Provider to retain Training Provider approval.

11.4 Grading

The exam has a pass mark of 65%.

Please note: Whilst BCS would not normally want to make changes to either grade thresholds or grading algorithms there is potential for them to change in order to maintain standards.

11.5 Externally assessed units

External tests from BCS come in the form of automated tests. The tests offer instant results to the learner.

11.6 Specimen assessment materials

A sample test is available on the BCS Website.

11.7 Support materials

BCS provides the following resources specifically for these qualifications:

Description	How to access
Syllabus	Available on website
Sample tests	Available on website

11.8 Access to Assessment

BCS seeks to provide equal Access to Assessment for all learners, ensuring that there are no unnecessary barriers to assessment and that any reasonable adjustments for learners preserve the validity, reliability and integrity of the qualification.

We will consider requests from BCS Approved Training Providers for reasonable adjustments and special considerations to be approved for a learner. The decision will be based on the individual needs of the learner as assessed by suitably qualified professionals. In promoting this policy, BCS aims to ensure that a learner is not disadvantaged in relation to other learners and their certificate accurately reflects their attainment.

12 Contact Points

BCS Qualifications Client Services is committed to providing you with a professional service and support at all times through a single, dedicated point of contact. With a flexible and proactive approach, our team will work together with you to ensure we deliver quality solutions that are right for you.

BCS, The Chartered Institute for IT
3 Newbridge Square
Swindon
Wiltshire
SN1 1BY

T: +44 (0) 1793 417 424

W: www.bcs.org/qualifications

If you require this document in an accessible format, please call +44 (0) 1793 417 424

© BCS, The Chartered Institute for IT, is the business name of The British Computer Society (registered charity no. 292786).