

# Big Data

## BCS Foundation Award



September 2020 v1.0

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

## CONTENTS

- 3. Introduction
- 4. Qualification Suitability and Overview
- 4. Trainer Criteria
- 5. SFIA Levels
- 6. Learning Outcomes
- 6. Syllabus
- 14. Examination Format
- 15. Recommended Reading List
- 15. Using BCS Books

## 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
----------------	--------------

V1.0	Document Creation
------	-------------------



# Qualification Suitability and Overview

There are no specific entry requirements for this award. However, some professional experience in a business or IT environment may be advantageous.

The BCS Foundation Award in Big Data has been designed for those interested in furthering their understanding of Big Data within AI or to discover how their existing skillsets in Data Management can apply to a career in AI. This Foundation Award is ideal for candidates who are looking to move into analytical or research roles within AI.

This award has been created alongside a selection of other awards in the AI space which offer candidates a clear pathway of progression into other disciplines of IT along with a broader knowledge of AI in the workplace. This makes it ideally suited for those looking for a change in career, an upskilling workforce, sustainable employers and individuals with a background in science, engineering, knowledge engineering, finance, education or IT services.

This list is not exhaustive and many other roles may benefit.

**This award represents 4 credits that can count towards the credits required for a BCS Foundation Certificate or Diploma in a relevant discipline.**

Candidates can study for this award by attending a training course provided by a BCS accredited Training Provider or through self study.

Total Qualification Time	Guided Learning Hours	Independent Learning	Assessment Qualification Time	Credits
40 hours	16 hours	31.5 hours	0.5 hours	4

\*Examples of Independent Learning include reading of articles or books, watching videos, attendance of other types of training or work shadowing.

## Introduction

Big Data has the potential to enable organisations to gain a much greater insight into the market in which they operate and empower them to make better business decisions. As such there is an increasing demand for individuals with the knowledge and the skills to analyse and make use of Big Data. This award will enable candidates to understand how to manage Big Data and its use with Machine Learning.

Candidates will explore the terminology, general principles, concepts and approaches related to the management of Big Data. This includes the relationship between data and information, as well as an understanding of Big Data frameworks (including Hadoop) and applicable solutions.

## Trainer Criteria

It is recommended that to effectively deliver this award, trainers should possess:

- BCS Foundation Certificate in a relevant subject e.g. Data, Business Analysis, Artificial Intelligence.
- A minimum of 2 years' training experience or a recognised training qualification.

# SFIA 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	<b>Apply</b>
K2	<b>Understand</b>	<b>Assist</b>
K1	<b>Remember</b>	<b>Follow</b>

## SFIA Plus

This syllabus has been linked to the SFIA knowledge skills and behaviours required at level 3 for an individual working in an analytical role.

### KSC10

The application of automated (software) tools which enable selective access to information, or extraction of patterns and knowledge from large amount of data, typically held within some form of database or "data warehouse". Also referred to as "Data Mining".

### KSB17

Applying specific quality standards to all tasks undertaken to ensure that deliverables are accurate and complete.

### KSB04

Identifying gaps in the available information required to understand a problem or situation and devising a means of resolving them.

### KSCA5

The ability to harvest, clean, curate, manage, process and manipulate data in a variety of formats.

### KSC51

The discipline associated with data sets so large and/or complex that traditional data processing applications are inadequate. The data files may include structured, unstructured and/or semi-structured data, such as unstructured text, audio, video, etc. Challenges include analysis, capture, curation, search, sharing, storage, transfer, manipulation, analysis, visualization and information privacy.

Further detail around the SFIA Levels can be found at [www.bcs.org/levels](http://www.bcs.org/levels).

# Learning Outcomes

Candidates will be able to demonstrate knowledge and understanding of the following areas:

1. What is Big Data?
2. Working with a Big Data framework
3. Big data analytics
4. Big data and machine learning

# Syllabus

## 1. What is Big Data? (20%) (K1,K2)

### Candidates will be able to:

#### 1.1 Define Big Data.

#### Indicative content

- a. Dictionary definition

#### Guidance

Big Data; "extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions. Candidates should explore examples.

### Candidates will be able to:

#### 1.2 Describe the difference between data and Big Data.

#### Indicative content

- a. Big Data characteristics
- b. The 6 Vs, Volume, Variety, Velocity, Veracity, Value, Variability.
- c. Data processing methods.

#### Guidance

Understanding how Big Data differs to other forms of data will enable the candidates understand how to treat them and how to use them effectively.

# Syllabus

## 1. What is Big Data? (20%) (K1,K2)

### Candidates will be able to:

**1.3** Explain the different types of data structures and what can be included within each type.

#### Indicative content

- a. Unstructured
- b. Structured
- c. Semi-structured

#### Guidance

Candidates should be able to understand the different structures as well as the differences between Structured, Unstructured, and Semi-Structured data. It is useful to consider the pros and cons of data structures.

### Candidates will be able to:

**1.4** Explain the need for a Big Data strategy.

#### Indicative content

- a. Data is money
- b. Understanding the customer
- c. Improving process
- d. Improving experience

#### Guidance

Candidates should understand that using Big Data is not a local decision made within a team; there is often an organisational strategy that defines how it is used. Having this strategy provides clarity to all those involved in using data in the organisation. Through the process of analysing and using Big Data it should enable an organisation to improve eventual outputs and make better predictions.

## 2. Working with a Big Data framework (20%) (K1/2)

### Candidates will be able to:

**2.1** Explain the need for a Big Data framework/network.

#### Indicative content

- a. Functionality
- b. Software
- c. High-performance computing and cloud computing (AWS, IBM etc.)
- d. Assess, manage, control, improve

#### Guidance

Candidates need to consider that Big Data is not just produced in one locality, but it can cross borders and providing a framework to follow gives consistency as to how it is managed and controlled.

### Candidates will be able to:

**2.2** Identify Big Data frameworks and the benefits of using them.

#### Indicative content

- a. Hadoop
- b. Spark
- c. Scalability
- d. Minimise the risk of system failure
- e. Faster data processing

#### Guidance

Although there are many Big Data frameworks we focus on Hadoop and Spark (as the most common, currently). It is useful for candidates to understand the purpose of the frameworks and the benefits which these frameworks offer, as well as the drawbacks. It is equally important for candidates to appreciate how quickly this area is changing and the future technologies that may be used.

# Syllabus

## 2. Working with a Big Data framework (20%) (K1/2)

### Candidates will be able to:

**2.3** Explain the need for data architecture.

### Indicative content

- a. Data sources
- b. Data storage
- c. Data governance – data protection, ISO27001

### Guidance

Candidates need to understand the link between Big Data, data frameworks, data architecture and Artificial Intelligence. They should understand that without the correct architecture, Big Data analysis will not be able to be undertaken effectively.

## 3. Big data analytics (30%) (K1/2)

### Candidates will be able to:

**3.1** Identify Big Data analytics.

### Indicative content

- a. Text analytics
- b. Predictive analytics
- c. Data mining
- d. Machine learning
- e. Linear regression
- f. Bayesian classifiers
- g. Time series analysis
- h. Decision trees

### Guidance

It is important that candidates understand the types of analytics that can be applied to Big Data and their uses. Candidates are not expected to complete time series analysis (for example) but they do need to understand what it is and when it would be used.

### Candidates will be able to:

**3.2** Explain why data analytics is important.

### Indicative content

- a. Real time
- b. Business intelligence
- c. Decision making

### Guidance

Candidates need to understand that without Big Data analytics, the information an organisation needs to make decisions would not be available. Candidates should appreciate that decisions often need to be made quickly to respond to changes and to have an impact.

## 4. Big Data and Machine Learning (30%) (K1/2)

### Candidates will be able to:

**4.1** Describe why Machine Learning needs Big Data.

#### Indicative content

- a. The more data the better
- b. Data reliance
- c. Reduce bias
- d. Learning from experience

#### Guidance

The key message is that Machine Learning models need to be trained using Big Data. For example, they need to see lots of pictures of cats to recognise a cat. It is however equally important to address data bias and that lots of the same or very similar data will not support the training process. It is key to consider what could reduce data bias. It is equally important to understand that applications of Machine Learning have the potential to learn from experience to improve.

### Candidates will be able to:

**4.2** Explain how Machine Learning uses Big Data

#### Indicative content

- a. Data input
- b. Data analysis
- c. Training and test data
- d. Batch or online modelling

#### Guidance

Candidates should understand the principles of how Machine Learning uses Big Data, and how it is dependent on data. It is useful to consider the Machine Learning process. It is pertinent to note that other AI techniques can exist without the data and Big Data. A basic understanding of how a machine learning model uses data is required at this stage.

### Candidates will be able to:

**4.3** Illustrate the challenges of Big Data

#### Indicative content

- a. Quality vs quantity
- b. Privacy

#### Guidance

As per LO 4.1 it is important to discuss that the answer to the problem is not necessarily more data. There is a need to pursue quality data. It is essential that the laws around privacy are followed as well as the ethical principles an organisation has in place. It is important to address the need to consider this at the stage of design so that privacy, morals, and ethical behaviour are built into the design. Consider the vast amounts of data that an organisation gathers and the challenge of managing it and ensuring the right level of permissions are in place.

# Examination Format

This award 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.

<b>Type</b>	16 Multiple Choice questions, 2 Scenario Based Questions
<b>Duration</b>	60 minutes
<b>Supervised</b>	Yes
<b>Open Book</b>	No (no materials can be taken into the examination room)
<b>Passmark</b>	13/20 (65%)
<b>Delivery</b>	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:

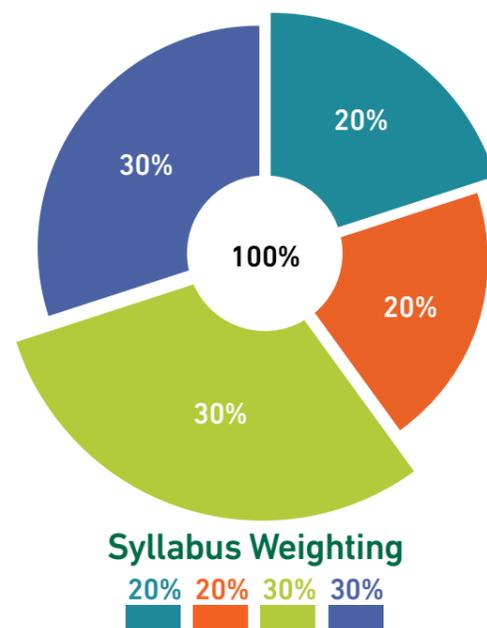
1. Guidance on the proportion of content allocated to each topic area of an accredited course.
2. Guidance on the proportion of questions in the exam.

## Syllabus Area

- 1. What is Big Data?
- 2. Working With a Big Data Framework
- 3. Big Data Analytics
- 4. Big Data and Machine Learning

## Question type

- Multiple Choice **20%**
- Multiple Choice **20%**
- Scenario Based Multiple Choice **30%**
- Scenario Based Multiple Choice **30%**



# Recommended Reading

The following titles are suggested reading for anyone undertaking this award. Candidates should be encouraged to explore other available sources.

<b>Title:</b>	Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems
<b>Author:</b>	Aurélien Géron
<b>Publisher:</b>	O'Reilly
<b>Publication Date:</b>	2017
<b>ISBN:</b>	1491962291

# 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 license from BCS. To request a license, please contact the Head of Publishing at BCS outlining the material you wish to copy and the use to which it will be put.

# CONTACT

For any queries relating to this document or the delivery of this award, please contact;

**T:** 01793 417445

**E:** bcssales@bcs.uk

If you have any queries relating to the online assessments, please contact;  
Service Delivery - eprofessional@bcs.uk

For further information please contact:

## **BCS**

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

**T** +44 (0)1793 417 445

[www.bcs.org](http://www.bcs.org)

© 2020 Reserved. BCS, The Chartered Institute for IT

All rights reserved. No part of this material protected by this copyright may be reproduced or utilised in any form, or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system without prior authorisation and credit to BCS, The Chartered Institute for IT.

Although BCS, The Chartered Institute for IT has used reasonable endeavours in compiling the document it does not guarantee nor shall it be responsible for reliance upon the contents of the document and shall not be liable for any false, inaccurate or incomplete information. Any reliance placed upon the contents by the reader is at the reader's sole risk and BCS, The Chartered Institute for IT shall not be liable for any consequences of such reliance.

