

**BCS LEVEL 5 DIPLOMA IN IT**

# **BIG DATA MANAGEMENT**

**SYLLABUS**



**September 2021 v3.0**

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

- 3. Introduction
- 4. Qualification Suitability and Overview
- 4. SFIA Levels
- 5. Learning Outcomes
- 6. Syllabus
- 13. Examination Format
- 13. Question Weighting
- 14. Recommended Reading
- 17. Using BCS Books



# Introduction

The second stage within the BCS three-stage Higher Education Qualification programme, the Level 5 Diploma enables candidates who have already achieved the Level 4 Certificate in IT to progress to higher levels of knowledge and competency.

This internationally-recognised qualification introduces you to the business-related aspects of the IT industry, developing your technological expertise while also considering the potential challenges of the day-to-day running of an organisation, such as legal obligations and intellectual property.

Our modules have been created in-line with the latest developments in the industry, giving you a competitive edge in the IT job market. You will have the opportunity to learn about object-oriented programming, user experience, systems analysis and design, as well as to build upon knowledge and skills developed during the Level 4 Certificate.

To successfully achieve the qualification, candidates need to complete:

- One core module
- Three optional modules
- One Professional Project in IT

Candidates who wish to progress onto the next stage will need to complete the Project at end of the Level 6 Professional Graduate Diploma in IT.

## Big Data Management Optional Module

The Big Data Management module is an optional module that forms part of the Level 5 Diploma in IT – the second stage within the BCS three-stage Higher Education Qualification programme.

Candidates will be introduced to Big Data architectures, infrastructures, as well as tools and techniques for data analysis. They will explore issues relating to Big Data management and develop specialist knowledge and understanding relating to the storage, data and programming interface models used for Big Data. Candidates will have the opportunity to learn to construct simple R language scripts.

# Qualification Suitability and Overview

Candidates must have achieved the Certificate in IT or have an appropriate exemption to be entered for the Diploma in IT. Candidates can study for this diploma by attending a training course provided by a BCS accredited Training Provider or through self-study, although it is strongly recommended that all candidates register with an approved centre. Studying with an approved centre will deliver significant benefits.

Candidates are required to become a member of BCS, The Chartered Institute for IT, to sit and be awarded the qualifications. Candidates may apply for a four-year student membership that will support them throughout their studies.

The Level 5 Diploma is suitable for professionals wishing to gain a formal IT qualification, and this module may be particularly relevant for candidates interested in career opportunities such as machine learning engineering, data architecture, data analysis, or business/analytics translation.

---

<b>Total Qualification Time (Certificate)</b>	<b>Guided Learning Hours (Module)</b>	<b>Assessment Time (Exam)</b>
<b>1086 hours</b>	<b>225 hours</b>	<b>Two hours</b>

---

## 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.

<b>Level</b>	<b>Levels of Knowledge</b>	<b>Levels of Skill and Responsibility (SFIA)</b>
K7		Set strategy, inspire and mobilise
K6	Evaluate	Initiate and influence
K5	Synthesise	Ensure and advise
K4	Analyse	Enable
K3	<b>Apply</b>	<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 5.

### INAN3:

Undertakes analytical activities and delivers analysis outputs, in accordance with customer needs and conforming to agreed standards.

### DBDS3

Develops appropriate physical database or data warehouse design elements, within set policies, to meet business change or development project data requirements. Interprets installation standards to meet project needs and produces database or data warehouse component specifications.

## SCTY3

Communicates information security risks and issues to business managers and others. Performs basic risk assessments for small information systems. Contributes to vulnerability assessments. Applies and maintains specific security controls as required by organisational policy and local risk assessments. Takes action to respond to security breaches in line with security policy and records the incidents and action taken.

### NTDS3

Produces outline system designs and specifications, and overall architectures, topologies, configuration databases and design documentation of networks and networking technology within the organisation. Specifies user/system interfaces, including validation and error correction procedures, processing rules, access, security and audit controls. Assesses associated risks, and specifies recovery routines and contingency procedures. Translates logical designs into physical designs.

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

---

# Learning Outcomes

Upon completion of this module, candidates will be able to:

- Comply with applicable government regulations and maintain the security of Big Data.
- Manage appropriate architectures and storage for Big Data applications.
- Analyse data held within typical Big Data systems.

# Syllabus

## 1. Big Data Fundamentals

### Learners will be able to:

**1.1** Explain and describe defining characteristics.

#### Indicative content

- a. Laney's Vs (volume, velocity, variety, veracity, valence and value) and later extensions
- b. Typical measures of magnitude

#### Guidance

Candidates will need an awareness of the recognised attributes of big data as shown in the indicative content, and the implications that they will have on data storage and big data processing.

---

**1.2** Explain categories and examples of big data.

#### Indicative content

- a. Machines
- b. People
- c. Organisations

#### Guidance

Candidates will need to draw on real-world examples of systems based on big data, such as Facebook and Twitter.

---

**1.3** Describe the challenges of big data.

#### Indicative content

- a. Storage
- b. Infrastructure
- c. Technologies
- d. Processing

#### Guidance

Candidates should have an understanding of organisations' obligations to plan for storage, employ staff with appropriate skills, keep up-to-date with new technology, and efficiently process and analyse data.

---

## 2. Management Issues for Big Data

### Learners will be able to:

**2.1** Explain the key management issues for big data.

#### Indicative content

- a. Staffing
- b. Skillsets
- c. Strategies
- d. Total cost of ownership

#### Guidance

Big data is often associated with new technologies – it implies the use of large amounts of storage and specialised kit. Candidates should be able to explain the potential challenges (and expenses) for organisations in terms of finding experienced staff and, as data tends to grow exponentially, forward planning in terms of staff and their skillsets is essential.

---

**2.2** Describe the service provisions for big data management.

#### Indicative content

- a. Sourcing (ingestion)
- b. Storage
- c. Quality
- d. Operations
- e. Scalability

#### Guidance

Candidates should be able to discuss issues around data gathering, efficient storage and assurance of data quality, and resilience to data growth.

---

**2.3** Explain security, ethical and legal considerations.

#### Indicative content

- a. Personal data
- b. Internal data
- c. Considerations of data privacy in national laws, e.g. EU GDPR, US data regulations and UK data privacy regulations

#### Guidance

It is important for candidates to have an understanding of how and why governmental controls and data privacy are important, especially in the context of unstructured, personal data.

---

## 3. Storage, data and programming interface models used for Big Data

### Learners will be able to:

**3.1** Describe the key storage data, data types and documentation.

#### Indicative content

- a. Key-value
- b. Relational
- c. Extensible
- d. Document

#### Guidance

Candidates should be familiar with new storage organisations and the potential implications of using them. It would also be helpful for candidates to have knowledge of some commercial systems such as Mongo DB and NoSQL products.

---

**3.2** Explain data storage.

#### Indicative content

- a. Block
- b. File
- c. Object-based

#### Guidance

The physical characteristics of data storage, its positioning, processing and transfer to memory are key concepts for Big Data practitioners, and candidates should be able to discuss these.

---

**3.3** Describe data modelling.

#### Indicative content

- a. Structured data
- b. Unstructured data
- c. Streamed data

#### Guidance

Candidates will need to compare and contrast structured and unstructured data, as well as the characteristics of streamed data.

---

**3.4** Demonstrate data modelling in real time.

#### Indicative content

- a. Pipelines with Kafka and similar technologies

#### Guidance

Increased flexibility in data design approaches is becoming more and more important. Candidates will need to demonstrate the versatility of Kafka and other similar technologies.

---

**3.5** Compare programming interface models.

### Indicative content

- a. Hadoop
- b. MapReduce
- c. Spark
- d. Yarn

### Guidance

Candidates will be expected to be familiar with big data systems' languages and interfaces when handling such systems, as well as comparing these systems and considering their advantages and disadvantages.

## 4. Architecture for Big Data

### Learners will be able to:

**4.1** Describe techniques for storing data in secondary storage.

### Indicative content

- a. MapReduce
- b. HDFS
- c. Yarn

### Guidance

Candidates are expected to have knowledge of underlying concepts and techniques used in software, such as MapReduce.

---

**4.2** Explain and compare noSQL and RDBMS.

### Indicative content

- a. Definition, benefits, and comparison to RDBMS
- b. ACID properties

### Guidance

Candidates should be able to assess the use of BASE versus ACID approaches when dealing with transaction processing and concurrency control.

---

**4.3** Describe Brewers theorem.

### Indicative content

- a. Use of CAP 9 Consistency Availability Partition Tolerance Correct in NoSQL frameworks

### Guidance

Candidates should be aware of how to process transactions and be able to describe this process.

---

**4.4** Demonstrate and analyse database frameworks.

### **Indicative content**

- a. Examples that provide CA, CP, and AP, e.g. MongoDB, MySQL, and Cassandra

### **Guidance**

Candidates should be able to evaluate how effective common database products are.

## **5. Big Data infrastructures**

### **Learners will be able to:**

**5.1** Explain the use of data with cloud-based networks.

### **Indicative content**

- a. In-house
- b. Outsourced

### **Guidance**

Candidates should have a solid understanding of potential implications of using Big Data in cloud systems versus on-premises.

---

**5.2** Describe distributed file systems.

### **Indicative content**

- a. Specialised applications, e.g. GFS

### **Guidance**

Candidates should be able to show understanding of the impact of distributed file systems, such as Hadoop, on typical big data processing.

---

**5.3** Describe ownership and financial implications of cloud use.

### **Guidance**

Candidates should be able to contrast security and financial risk of cloud-based technology with the expense and management of on-premises systems.

---

## 6. Tools and techniques for analysis

### Learners will be able to:

**6.1** Explain Gartner's model.

#### Indicative content

- a. Descriptive
- b. Diagnostic
- c. Predictive
- d. Prescriptive

#### Guidance

Candidates are expected to be familiar with Gartner's model, mapping each stage across seven dimensions: goal, data, skill sets, organisational structure, use cases (applications), analytics techniques, and supporting technology. Candidates should also understand the connections between these dimensions.

---

**6.2** Analyse data using basic algorithms of machine learning.

#### Guidance

Candidates will be required to explain fundamentals of data analysis. Basic machine learning algorithms may include linear regression, logistic regression, decision trees, SVM, dimensionality reduction algorithms, and gradient boosting algorithms, but this is not designed to be an exhaustive list.

---

**6.3** Demonstrate and explain the use of AI in big data applications.

#### Guidance

Candidates should be able to describe examples of using AI when analysing big data. For example, AI may be used to identify data types, find possible connections among datasets, and recognise knowledge using natural language processing. However, this is not an exhaustive list.

---

**6.4** Describe types of neural networks.

#### Indicative content

- a. Classifiers
- b. Feedback
- c. Feedforward
- d. Deep learning

#### Guidance

Candidates will be expected to have a general understanding of this topic and be able to describe these terms in relation to big data, but not to use or demonstrate them in the exam.

---

**6.5** Explain and demonstrate the use of the hash function.

### **Indicative content**

- a. Methods and uses e.g. MapReduce.
- b. Bloom filters and their application in data mining

### **Guidance**

Candidates should be familiar with optimisation techniques such as Bloom filters, both explaining them and actively demonstrating them, or using a case system to do so in the exam.

## **7. Introduction to R**

### **Learners will be able to:**

**7.1** Demonstrate statistical analysis.

### **Indicative content**

- a. Use of the R language
- b. Basic statistical analysis

### **Guidance**

Candidates will be expected to have knowledge of the R language and its practical use.

---

**7.2** Utilise and demonstrate R language.

### **Indicative content**

- a. Language and ecosystem basic syntax.

### **Guidance**

Candidates will also be required to demonstrate their understanding of R by constructing simple R scripts.

---

**7.3** Explain the basic statistics used in R.

### **Indicative content**

- a. Mode
- b. Mean
- c. Variance
- d. 5d
- e. Moments
- f. Clustering

### **Guidance**

Candidates should be able to understand the role and use of statistics in data analysis (e.g. mode, mean, variance).

---

**7.5** Describe CRAN packages.

## Indicative content

a. Cluster analysis

## Guidance

Candidates will be expected to describe and use cluster analysis in identifying patterns and trends.

---

# Examination Format

This module is assessed through completion of an invigilated written exam.

<b>Type</b>	Four written questions from a choice of six, each with equal marks
<b>Duration</b>	Two hours
<b>Supervised</b>	Yes
<b>Open Book</b>	No (no materials can be taken into the examination room)
<b>Passmark</b>	10/25 (40%)
<b>Delivery</b>	Paper 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.

---

# Question Weighting

Candidates will choose four questions from a choice of six. All questions are equally weighted and worth 25 marks.

# Recommended Reading

## Primary texts

**Title:** Big Data Fundamentals: Concepts, Drivers, and Techniques.  
**Author:** T. Erl and W. Khattak  
**Publisher:** Prentice Hall  
**Publication date:** 2016  
**ISBN:** 978-0-134-29107-9

**Title:** Big Data Demystified: How to use big data, data science and AI to make better business decisions and gain competitive advantage  
**Author:** D. Stephenson  
**Publisher:** Pearson Education  
**Publication date:** 2018  
**ISBN:** 978-1-292-21810-6

**Title:** Big Data Management and Analysis  
**Author:** K. Adam, M. A. I. Fakharaldien, J. M. Zain and M. A. Majid  
**Publisher:** University Malaysia Pahang  
**Publication date:** 2014  
**Available at:** [Web link](#) (Accessed 5 July 2021)

**Title:** Benefits, Challenges and Tools of Big Data Management  
**Author:** F. Almeida  
**Publisher:** Journal of Systems Integration  
**Publication date:** 2017  
**Volume and page numbers:** 8, pp. 12-20

**Title:** NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence  
**Author:** P. J. Sadalage and M. Fowler  
**Publisher:** RR Donnelley  
**Publication date:** 2013  
**ISBN:** 978-0-321-82662-6

## Additional texts

**Title:** Principles of Database Management: Practical Guide to Storing, Managing and Analyzing Big and Small Data  
**Author:** W. Lemahieu, S. vanden Broucke and B. Baesens  
**Publisher:** Cambridge University Press  
**Publication date:** 2018  
**ISBN:** 978-1-107-18612-5

**Title:** Big Data Management and Processing  
**Author:** K.-C. Li, H. Jiang and A. Y. Zomaya  
**Publisher:** CRC Press  
**Publication date:** 2017  
**ISBN:** 978-1-498-76807-8

**Title:** Big Data Analytics. Tools and Technology for Effective Planning  
**Author:** A. K. Somani and G. C. Deka  
**Publisher:** CRC Press  
**Publication date:** 2018  
**ISBN:** 978-1-138-03239-2

## Online resources

**Title:** HDFS Architecture  
**Author:** Apache Hadoop  
**Publication date:** 2019  
**Available at:** [Web link](#) (Accessed 5 July 2021)

**Title:** Apache Hadoop YARN  
**Author:** Apache Hadoop  
**Publication date:** 2019  
**Available at:** [Web link](#) (Accessed 5 July 2021)

**Title:** Big Data Management  
**Author:** Datamation  
**Publication date:** 2017  
**Available at:** [Web link](#) (Accessed 5 July 2021)

**Title:** Hadoop Streaming  
**Author:** Apache Hadoop  
**Publication date:** 2019  
**Available at:** [Web link](#) (Accessed 5 July 2021)

**Title:** simpleR – Using R for Introductory Statistics  
**Author:** J. Verzani  
**Publication date:** 2002  
**Available at:** [Web link](#) (Accessed 5 July 2021)

# 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 its intended use.

## 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.

<b>Version Number</b>	<b>Changes Made</b>
<b>Version 1.0</b> July 2021	Document Creation

# CONTACT

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)

© 2021 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.

