Next Generation Higher National Unit Specification

Big Data (SCQF level 7)

Unit code: J68W 47

SCQF level: (8 SCQF credit points)

Valid from: session 2022–23

Prototype unit specification for use in pilot delivery only (version 1.0) May 2022

This unit specification provides detailed information about the unit to ensure consistent and transparent assessment year on year.

This unit specification is for teachers and lecturers and contains all the mandatory information required to deliver and assess the unit.

This edition: May 2022 (version 1)

Unit purpose

This unit provides an introduction to the theory and practice of big data. The unit is for non-specialists who require a basic understanding of the concepts behind big data and its potential social, commercial and scientific uses. It is suitable for a wide range of learners.

The unit covers a mix of theory and practice. The theoretical content includes:

- the concepts behind big data
- how big data differs from traditional analytical techniques
- the applications of big data
- the societal implications of using big data

The practical content relates to how big data can be used in real-life situations.

The unit seeks to provide foundational knowledge of this emerging discipline so that learners can appreciate its actual and potential uses in a range of contexts.

On completion of this unit, learners can progress to the Data Science unit at SCQF level 8 or the Big Data unit at SCQF level 8.

Unit outcomes

Learners who complete this unit can:

1 explain the concepts behind big data
2 describe the applications of big data in business, science and society
3 explain the implications of big data for individuals and society
4 apply big data techniques to a problem

Evidence requirements

Learners need to provide evidence to demonstrate their knowledge and skills across all outcomes.

The evidence requirements for this unit take two forms:

- Knowledge evidence — the definitions, descriptions and explanations required for outcomes 1, 2 and 3
- Product evidence — the application of big data techniques to a specific problem required for outcome 4

Evidence is normally required for all of the knowledge and skills in every outcome. This means that every knowledge and skills statement must be evidenced. However, you can use sampling in a specific circumstance (see table below).

The amount of evidence should be the minimum consistent with the defined knowledge and skills. For outcome 2, it is sufficient for learners to describe one application of big data in each area and two future applications of big data (in any area). For outcome 4, it is sufficient to apply big data techniques to one problem.

Evidence can be produced at any time during the life of the unit. When carrying out assessment, learners can use reference materials and there are no time limitations (except under test conditions). When evidence is produced in uncontrolled or lightly controlled conditions it must be authenticated. The Guide to Assessment provides further advice on methods of authentication.

Sampling is permissible when the evidence for outcomes 1, 2 and 3 is produced by a test of knowledge and understanding. The test can take any form (including oral) but must be supervised, unseen and timed. The contents of the test must sample broadly and proportionately from the contents of outcomes 1, 2 and 3 with equal weighting for each outcome. Access to reference material is not appropriate for this type of assessment.

The product evidence for outcome 4 can relate to a real or fictitious problem. Learners can apply big data techniques to a real problem **or** explain how big data techniques could be applied to a real or fictitious problem. In the former case (application to a real problem), the evidence would consist of the results of the use of big data techniques; in the latter case (explanation of a real or fictitious application) all of the knowledge and skills statements must be evidenced in the explanation.

The ‘Additional guidance’ section in this unit provides specific examples of assessment.

Knowledge and skills

The following table shows the knowledge and skills covered by the unit outcomes:

Knowledge	Skills
Learners should understand the: ♦ definition of big data ♦ historical development of big data (technologies and techniques) ♦ growth of data (including measures of data) ♦ reasons for the growth of data ♦ value of data (including future value) ♦ traditional statistics (descriptive and inferential) ♦ limitations of traditional data analysis ♦ characteristics of big data analysis (including visualisations) ♦ contemporary applications of big data in business ♦ contemporary applications of big data in science ♦ contemporary applications of big data in society ♦ future applications of big data ♦ technological requirements of big data ♦ limitations of predictive analytics ♦ implications of big data for individuals ♦ implications of big data for society ♦ strategies for limiting the negative effects of big data	Learners can: ♦ identify types of problem suited to big data analysis ♦ use data mining methods ♦ create different types of visualisations ♦ apply big data techniques to a real-world problem

Meta-skills

Throughout the unit, learners develop meta-skills to enhance their employability in the computing sector.

Self-management

This meta-skill includes:

- integrity: ethics
- adapting: critical reflection, resilience
- initiative: independent thinking

Social intelligence

This meta-skill includes:

- communicating: receiving information, storytelling
- feeling: social conscience

Innovation

This meta-skill includes:

- curiosity: questioning, problem recognition
- creativity: idea generation and visualising
- sense-making: pattern recognition, holistic thinking, synthesis and analysis
- critical thinking: logical thinking, computational thinking

Literacies

Numeracy

Learners develop numeracy skills by learning and applying statistical methods for data analysis of large datasets.

Communication

Learners develop communication skills by selecting and applying visualisation to a supplied data set.

Digital

This unit contributes towards learners' digital skills by introducing them to the concepts, issues and practical application of big data and data analysis.

Delivery of unit

You can deliver the outcomes in the order in which they are numbered. They are numbered with a learning sequence in mind.

While the exact time allocated to this unit is at your your discretion, the notional design length is 40 hours. One possible approach is to distribute the available time as follows:

Outcome 1 — Explain the concepts behind big data
(8 hours)
Outcome 2 — Describe the applications of big data in business, science and society
(12 hours)
Outcome 3 — Explain the implications of big data for individuals and society
(8 hours)
Outcome 4 — Apply big data techniques to a problem
(12 hours)

We recommend you introduce the required concepts and reinforce them with appropriate examples.

There is significant scope in this unit to illustrate concepts and skills with case studies of actual or potential uses of big data. Most of your time in this unit is spent covering the theoretical aspects, and you can explain them through real-world applications of big data.

Throughout this unit, learners' activities should relate to their vocational interests. For example, the applications and implications of big data (outcomes 2 and 3) should relate to a relevant vocational area; the application of big data techniques (outcome 4) should relate to their vocational interests.

If you deliver this unit as part of a group award, we recommend that you teach and assess it within the subject area of the group award to which it contributes.

Where evidence for outcomes is assessed on a sample basis, the whole of the content listed in the 'Knowledge and skills' section must be taught and available for assessment. You should not give learners advance information about the items they are being assessed on, and you should sample different items on each assessment occasion.

Additional guidance

The guidance in this section is not mandatory.

Content and context for this unit

The contents of this unit provide a **basic** introduction to the principles and practice of big data. The focus is the applications and implications of big data, rather than the technological aspects of this subject. However, you should include technological considerations.

The general context for this unit is the rising public interest in this emerging technique, and its vocational relevance to an increasing number of occupations and professions. The unit is intended for non-specialists and can be offered as part of a wide range of programmes. Learners do not need prior knowledge of computer science or statistics; however, numerical competency is presumed. The level of treatment of each topic should reflect this. For example, you should present the technological requirements of big data (outcome 2) in a high-level manner that focuses on the broad technologies required to support big data analysis. While the treatment should not trivialise these complex requirements, a detailed understanding of relational database technologies, for example, is not appropriate.

You should deliver this unit in an appropriate context for learners and reflect their vocational and personal interests. For example, learners with an interest or background in business should be taught in that context. However, you must introduce all learners to a variety of applications of big data and not just those directly relevant to their vocational interests.

Throughout this unit it is vital to present the applications and implications of big data in a balanced way, neither overstating the opportunities nor understating the threats posed by this technology.

Outcomes 1, 2 and 3 provide a theoretical underpinning to the subject of big data. Outcome 4 provides learners with an opportunity to apply this knowledge. You can consider both outcomes 2 and 3 as the pros and cons of big data.

Explain the concepts behind big data (outcome 1)

This outcome covers the concepts behind the emerging discipline of big data. The key aspect of this outcome is that learners appreciate the difference between traditional (descriptive) data analysis and the new (predictive) data analysis made possible by massive data sets.

When teaching the historical development of big data, you should emphasise its roots in computer science and artificial intelligence (AI). You should frame it as one of a growing number of applications of AI that are currently emerging. You should lightly cover the technologies and techniques involved in the evolution of big data. It is sufficient for learners to appreciate the technological milestones (what and when) that contributed to the development of big data analysis (rather than the how and why).

It may be useful (from a teaching perspective) to consider big data under three headings:

- technology
- algorithms (software)
- business (decision making)

Data should span all three of these areas.

Learners may be unaware of the huge growth in data in recent years, and the reasons for this growth. You should explain this through examples, such as the massive data sets generated by smartphones and online services (like Facebook). You should introduce the standard units of measurement of digital data (such as byte, kilobyte, megabyte, gigabyte and terabyte).

Learners may not appreciate the value of information, and how it can be monetised by online services. You can explain this with examples such as the (very) high market value of services, such as Twitter, despite its (relatively) low asset value. You should introduce the concept of future value (the future value of Facebook's information is one of the main reasons for Facebook's exceptionally high market value).

You should introduce learners to traditional descriptive and inferential statistics (based on small samples) and the inherent limitations of traditional approaches (when small samples are used to approximate large populations). You should contrast this with the huge data sets that big data employs, and the advantages that it confers.

Describe the applications of big data in business, science and society (outcome 2)

This outcome covers the applications of big data in business, science and society. It is sufficient to cover a limited number of applications in each area. Your descriptions do not need to be exhaustive. Business applications include the use of big data techniques to improve products and services; scientific applications include those in health care and cosmology; and societal applications include improved public services and crime prevention.

You should make learners aware of the potential future applications of big data. At the time of writing, there is controversy about the potential use of big data to predict crime and take pre-emptive action to prevent crime.

You should explain the technological requirements of big data in simple terms. The critical aspect is that learners appreciate the basic technologies (both hardware and software) that have made big data possible. For example, the falling cost and rising capacity of digital storage have made it possible to retain huge amounts of data on server farms; the large-scale adoption of smartphones has created new sources of data relating to location; and improved computer power has made improved software to analyse massive data sets possible.

It may be worthwhile exploring the new skills and job opportunities (such as 'data scientist') that are emerging from the growth of big data applications.

Explain the implications of big data for individuals and society (outcome 3)

This outcome relates to the implications of big data for individuals and society. You should explore the limitations of predictive analytics. The main (current) criticism of big data relates to the over-confidence in the accuracy of large data sets when applied to individuals. While big data draws on huge data sets, these data sets are never (or rarely) entire populations and, therefore, some of the limitations of traditional data analysis still apply.

The implications for individuals relate to personal privacy and personal liberty. Numerous cases have already arisen about the abuse of data by state security services. The controversy of using big data to predict future crime has also been well documented. There are a number of societal issues too. The use of big data for crime prediction and prevention is one.

Solutions to these real or potential issues have been proposed. These include giving the data owner (the person who generates the data) the right to decide how, and by whom, their data is used. Privacy advocates have proposed stronger privacy rights for individuals with fewer exemptions for state authorities. You should explore the pros and cons of such solutions.

Apply big data techniques to a problem (outcome 4)

This outcome seeks to apply the knowledge and skills acquired in outcomes 1, 2 and 3.

Learners should appreciate the types of problem that are best suited to predictive analytics, including the new applications made possible by recent technological developments. This can be challenging for both you and learners because this is an emerging area of research in computer science. It can be understood through examples. Using familiar online services, such as Facebook, to illustrate how the data stored in such services can be mined for utility can help learners to appreciate the types of problem that big data can be used to solve.

There is significant scope in this outcome to use case studies to illustrate actual or potential ways in which big data techniques can be used to solve problems.

There is no requirement for learners to actually use big data techniques to solve a problem. However, we recommend that you give them some practical experience of applying big data techniques, even if this is only a small part of the overall process (such as selecting and applying a visualisation to a supplied data set).

Approaches to assessment

Evidence can be generated using different types of assessment. The following are suggestions only. There may be other methods that would be more suitable to learners.

A traditional approach to assessment would involve an end of unit test of the knowledge and understanding (outcomes 1, 2 and 3) and a practical assessment of skills (outcome 4).

The end-of-unit test would sample from the knowledge and understanding contained in outcomes 1, 2 and 3. For example, you could set a test comprising five extended-response questions, selected from across all three outcomes, and then mark and assess these traditionally. Each would be worth 10 marks, with learners' responses marked out of 50 marks, with a pass mark of 25 marks. If you take this approach, we recommend that some (or all) of the extended-response questions combine the knowledge and understanding within and across outcomes. This test would be taken, sight-unseen, in controlled and timed conditions without access to reference materials. A suitable duration could be 1 hour and 30 minutes.

The practical assessment (outcome 4) could be an explanation of how big data techniques can be applied to a real-world problem. This could be assessed holistically, without a marking scheme and not assigned a specific score, and given a simple 'pass' or 'fail' grade. All of the knowledge and skills would be evidenced in this assessment. There would be no time limitations (beyond the practicality of completing the unit within the scheduled timetable) for this assessment.

A more contemporary approach to assessment could use a blog to record learning throughout the life of the unit. If you take this approach, then sampling is not appropriate. The blog would contain evidence for all knowledge and skills statements. The blog would record, on a daily or weekly basis, the learning that has occurred. It would contain textual definitions, descriptions and explanations as required by the knowledge and skills statements in the outcomes (all outcomes), including hyperlinks and embedded multimedia (audio, graphic or video).

The blog could encompass all outcomes, including outcome 4 (which is practical). It should explain how big data techniques could be applied to a real (or fictitious problem) in a similar way to the theoretical description in the traditional approach (see above). Given that the blog would, most likely, be completed at various times and locations throughout the life of the unit, some form of authentication would be necessary. There would be no time limitation on the completion of the blog because it would be done on an on-going basis throughout the life of the unit.

Equality and inclusion

This unit is designed to be as fair and as accessible as possible with no unnecessary barriers to learning or assessment.

You should take into account the needs of individual learners when planning learning experiences, selecting assessment methods or considering alternative evidence.

Guidance on assessment arrangements for disabled learners and/or those with additional support needs is available on the assessment arrangements web page: www.sqa.org.uk/assessmentarrangements.

Information for learners

Big Data (SCQF level 7)

This section explains:

- what the unit is about
- what you should know or be able to do before you start
- what you need to do during the unit
- opportunities for further learning and employment

Unit information

This unit aims to introduce you to the theory and practice of big data. No previous knowledge of computing or statistics is required, but you should be comfortable with numbers before commencing this unit.

Big data is an emerging technique for analysing human behaviour and natural phenomena, and is becoming widely used in various fields, ranging from business to cosmology. This unit seeks to explain what it is, how it might affect you, and how you might use it in your job. It shows how organisations (such as Facebook) currently use big data techniques to 'mine' your personal data.

The unit covers:

- the definition of big data
- the historical development of big data
- the value of data
- how big data differs from traditional analytical methods
- how big data is used in business, science and society
- how big data might be used in the future
- the implications of big data for individuals and businesses
- how society can protect itself from the negative effects of big data
- how to apply big data techniques

You learn in a variety of ways, ranging from listening to your teacher explain the concepts behind big data to watching case studies of big data in the real world.

Assessments can take different forms. They should be straightforward and not take much time away from your learning. They may involve a test of your knowledge and some practical tasks, or may simply be a record of your activities during the unit. The focus of the unit is learning, not assessing.

The key goal of this unit is to introduce you to big data so that you understand what it is, know how you could use it, and appreciate the opportunities and threats posed by this emerging technology.

This unit also provides opportunities for you to enhance your meta-skills in self-management, social intelligence and innovation.

If you want to improve your knowledge and skills in this area, you can progress to:

- Data Science at SCQF level 8
- Big Data at SCQF level 8

Administrative information

Published: May 2022 (version 1.0)

Superclass: CB

History of changes

Version	Description of change	Date

Note: please check SQA's website to ensure you are using the most up-to-date version of this document.