Higher National Unit specification

General information

Unit title: Big Data (SCQF level 7)

Unit code: H8W8 34

Superclass: CB

Publication date: February 2015

Source: Scottish Qualifications Authority

Version: 01

Unit purpose

The purpose of this Unit is to provide 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 it differs from traditional analytical techniques, the applications of big data, and the societal implications of its use. The practical content relates to how big data can be used in real life situations.

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

At the completion of this Unit learners may progress to *Data Science* at SCQF level 8.

Outcomes

On successful completion of the Unit the learner will be able to:

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.

Higher National Unit specification: General information (cont)

Unit title: Big Data (SCQF level 7)

Credit points and level

1 Higher National Unit credit at SCQF level 7: (8 SCQF credit points at SCQF level 7)

Recommended entry to the Unit

No previous knowledge or experience is required. However, it would be beneficial if learners possessed *Numeracy Skills*. This may be evidenced by possession of the Core Skills Unit in *Numeracy* at SCQF level 6. Some previous knowledge of descriptive statistics is desirable but not essential.

Core Skills

Opportunities to develop aspects of Core Skills are highlighted in the Support Notes for this Unit specification.

There is no automatic certification of Core Skills or Core Skill components in this Unit.

Context for delivery

If this Unit is delivered as part of a Group Award, it is recommended that it should be taught and assessed within the subject area of the Group Award to which it contributes.

The Assessment Support Pack (ASP) for this Unit provides assessment and marking guidelines that exemplify the national standard for achievement. It is a valid, reliable and practicable assessment. Centres wishing to develop their own assessments should refer to the ASP to ensure a comparable standard. A list of existing ASPs is available to download from SQA's website **(http://www.sqa.org.uk/sqa/46233.2769.html)**.

Equality and inclusion

This Unit specification has been designed to ensure that there are no unnecessary barriers to learning or assessment. The individual needs of learners should be taken into account when planning learning experiences, selecting assessment methods or considering alternative evidence.

Further advice can be found on our website **www.sqa.org.uk/assessmentarrangements**.

Higher National Unit specification: Statement of standards

Unit title: Big Data (SCQF level 7)

Acceptable performance in this Unit will be the satisfactory achievement of the standards set out in this part of the Unit specification. All sections of the statement of standards are mandatory and cannot be altered without reference to SQA.

Where evidence for Outcomes is assessed on a sample basis, the whole of the content listed in the Knowledge and/or Skills section must be taught and available for assessment. Learners should not know in advance the items on which they will be assessed and different items should be sampled on each assessment occasion.

Outcome 1

Explain the concepts behind big data.

Knowledge and/or Skills

- 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)

Outcome 2

Describe the applications of big data in business, science and society.

Knowledge and/or Skills

- 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

Outcome 3

Explain the implications of big data for individuals and society.

Knowledge and/or Skills

- 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

Higher National Unit specification: Statement of standards (cont)

Unit title: Big Data (SCQF level 7)

Outcome 4

Apply big data techniques to a problem.

Knowledge and/or Skills

- Types of problem suited to big data analysis
- Data mining methods
- Types of visualisation
- Application of big data techniques to a problem

Evidence Requirements for this Unit

Learners will need to provide evidence to demonstrate their Knowledge and/or Skills across all Outcomes.

The Evidence Requirements for this Unit will take two forms:

1 Evidence of cognitive competence (for Outcomes 1, 2 and 3).
2 Evidence of practical competence (for Outcome 4).

The evidence of cognitive competence will be the definitions, descriptions and explanations required for Outcomes 1, 2 and 3. The evidence of practical competence will be 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, sampling may be used in a specific circumstance (see below).

The amount of evidence should be the **minimum** consistent with the defined knowledge and skills. For Outcome 2, it is sufficient for the learner 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 may be wholly or partly produced under controlled conditions. When evidence is produced in uncontrolled or loosely controlled conditions it must be authenticated. The *Guide to Assessment* provides further advice on methods of authentication. Evidence of authentication must be provided when any of the evidence is generated under loosely controlled conditions.

There are **no** time limitations on the production of evidence (but see exception below). The evidence may be produced at any time during the life of the Unit. Learners may use reference materials when undertaking assessment (but see exception below).

Sampling is permissible when the evidence for Outcomes 1, 2 and 3 is produced by a test of knowledge and understanding. The test may 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 approximately equal weighting for each Outcome. Access to reference material is not appropriate for this type of assessment.

The evidence of practical competence (Outcome 4) may relate to a real or fictitious problem. Learners may 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 Guidelines on Approaches to Assessment (see the Support Notes section of this specification) provides specific examples of instruments of assessment.

Higher National Unit Support Notes

Unit title: Big Data (SCQF level 7)

Unit Support Notes are offered as guidance and are not mandatory.

While the exact time allocated to this Unit is at the discretion of the centre, the notional design length is 40 hours.

Guidance on the 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, technological considerations are included.

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 may be offered as part of a wide range of programmes. No prior knowledge of computer science or statistics is required; however, numerical competency is presumed.

The level of treatment of each topic should reflect this. For example, the technological requirements of big data (Outcome 2) should be presented in a high-level manner, which focuses on the broad technologies that are 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.

The Unit should be delivered in an appropriate context for the learners and reflect their vocational and personal interests. For example, learners with an interest and/or background in business should be taught in that context. However, all learners must be exposed 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 under-pinning to the subject of big data. Outcome 4 provides an opportunity to apply this knowledge. Outcomes 2 and 3 may be considered as the pros-and-cons of big data respectively.

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.

Higher National Unit Support Notes (cont)

Unit title: Big Data (SCQF level 7)

The historical development of big data should emphasise its roots in computer science and artificial intelligence (AI). It should be framed as one of a growing array of applications of AI that are currently emerging. The treatment of the technologies and techniques involved in the evolution of big data should be light. 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 pedagogical perspective) to consider big data under three headings: technology, algorithms (software) and business (decision making) with data spanning all three.

Learners may be unaware of the huge growth in data in recent years, and the reasons for this growth. This may best be explained through examples such as the massive data sets generated by smartphones and online services (such as Facebook™). The standard Units of measurement of digital data (byte, kilobyte, megabyte, gigabyte, terabyte, etc.) should be introduced.

Learners may not appreciate the value of information, and how this can be monetised by online services. This may be explained by examples such as the (very) high market value of services such as Twitter™ in spite of its (relatively) low asset value. The concept of future value should also be introduced (the future value of Facebook's information is one of the main reasons for Facebook's™ exceptionally high market value).

Learners should be introduced 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). This should be contrasted with the huge data sets that big data employs, and the advantages that this confers.

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. Exhaustive descriptions are not required. 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.

Learners should be 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.

The technological requirements of big data should be explained 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 possible the retention of huge amounts of data on server farms; the large scale adoption of smart phones has created new sources of data relating to location; and improved computer power has made possible improved software to analyse massive data sets.

It may be worthwhile exploring the new skills and job opportunities (such as 'data scientist') that will emerge as a result of the growth of big data applications.

Higher National Unit Support Notes (cont)

Unit title: Big Data (SCQF level 7)

Outcome 3: This Outcome relates to the implications of big data for individuals and society. The limitations of predictive analytics should be explored. 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, his/her data is used. Privacy advocates have proposed stronger privacy rights for individuals with fewer exemptions for state authorities. The pros and cons of such solutions should be explored.

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 that will be made possible due to recent technological developments. This may be challenging for the learner and the teacher since this is an emerging area of research in computer science. It may be best understood through example. Using familiar online services, such as Facebook™, to illustrate how the data stored in such services may be mined for utility may help learners to appreciate the types of problem that big data could 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 the learner to actually use big data techniques to solve a problem. However, it is recommended that learners are given 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).

Guidance on approaches to delivery of this Unit

The Outcomes may be delivered in the order in which they are written. They have been written with a learning sequence in mind.

The actual distribution of time between Outcomes is at the discretion of the centre. However, one possible approach is to distribute the available time as follows:

Outcome 1 8 hours
Outcome 2 12 hours
Outcome 3 8 hours
Outcome 4 12 hours

Higher National Unit Support Notes (cont)

Unit title: Big Data (SCQF level 7)

It is anticipated that the required concepts will be introduced by the teacher and reinforced by 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. While the majority of time in this Unit will be spent on the theoretical aspects of the Unit, these can be explained through real-world applications of big data.

Throughout this Unit, learner 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.

Guidance on approaches to assessment of this Unit

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 practical abilities (Outcome 4).

The end of Unit test would **sample** from the knowledge and understanding contained in Outcomes 1, 2 and 3. The test could comprise a number of extended response questions (ERQs) and would be marked and assessed traditionally. For example, the test may comprise of five extended response questions, selected across all three Outcomes, each worth 10 marks, with the learner responses marked out of 50, with a pass mark of 25. If this approach is taken, it is recommended that some (or all) of the ERQs combine the knowledge and understanding within and across Outcomes. This test would be taken, sight-unseen, in controlled and timed conditions without reference to teaching materials. A suitable duration could be 90 minutes.

The practical assessment (Outcome 4) may comprise of an explanation of how big data techniques could 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/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 web log to record learning throughout the life of the Unit. If this approach is taken then sampling would not be 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). This could be done by the blog explaining 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 since it would be done on an on-going basis throughout the life of the Unit.

Centres are reminded that prior verification of centre-devised assessments would help to ensure that the national standard is being met. Where learners experience a range of assessment methods, this helps them to develop different skills that should be transferable to work or further and higher education.

Opportunities for e-assessment

E-assessment may be appropriate for some assessments in this Unit. By e-assessment we mean assessment which is supported by Information and Communication Technology (ICT), such as e-testing or the use of e-portfolios or social software. Centres which wish to use e-assessment must ensure that the national standard is applied to all learner evidence and that conditions of assessment as specified in the Evidence Requirements are met, regardless of the mode of gathering evidence. The most up-to-date guidance on the use of e-assessment to support SQA's qualifications is available at **www.sqa.org.uk/e-assessment**.

Opportunities for developing Core and other essential skills

This Unit provides opportunities to develop some of the following Core Skills:

- *Information and Communication Technology (ICT)* (SCQF level 6)
- *Numeracy* (SCQF level 6)

Several of the Core Skills in *Information and Communication Technology (ICT)* may be addressed in this Unit. There are opportunities to start software, enter and edit data, locate and extract information, apply a complex search strategy, evaluate information, and present information.

All of the Core Skills in *Numeracy* may be addressed in this Unit. There are opportunities to analyse situations to identify relevant data and relationships, decide which operations to carry out, use numerical and statistical theory, extract, analyse and interpret information, identify significant features in complex graphical information, and select an appropriate graphical form.

History of changes to Unit

Version	Description of change	Date

General information for learners

Unit title: Big Data (SCQF level 7)

This section will help you decide whether this is the Unit for you by explaining what the Unit is about, what you should know or be able to do before you start, what you will need to do during the Unit and opportunities for further learning and employment.

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:

- Definition of big data.
- Historical development of big data.
- 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.
- 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 will 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.

The assessment may take different forms. It will be straight-forward and not take much time away from your learning. It may involve a test of your knowledge and some practical tasks, or it 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.

The Unit will provide an opportunity to develop your *Information and Communication Technology (ICT) and Numeracy* Core Skills.

You may progress to the Higher National Unit entitled *Data Science* (SCQF level 8) on completion of this Unit if you wish to improve your knowledge and skills in this area.