



Big Data: Big what?

Age range: 16-19 and 19+

 **BARCLAYS** | LifeSkills



Session overview

Time	Key learning outcomes	Resources
120 mins	<ul style="list-style-type: none"> Explain what big data is and how it is processed and analysed Demonstrate an ability to interpret a visualisation of big data Explain examples of careers and industries that work with big data 	<ul style="list-style-type: none"> Big Data interactive tool Big Data: Big what? student worksheet

Always start the session by agreeing ground rules with the group. For advice on this and other ways to establish a safe learning environment, download the content guide [here](#).

Contents

Activities	Time	Page
Activity one: Big what?	30 mins	3
Activity two: How is big data processed, stored and analysed?	30 mins	5
Activity three: The big data sector	45 mins	10
Summary	5 mins	11

Activity one

Big what?

1. Discuss how big data is utilised

Big data is being utilised by businesses, governments and scientists to improve lives, and new ways of using the data are being explored all the time. To understand how and why big data is collected, and how it can be used, it is important that young people see the context in which this is happening.

Explain that by exploring big data, building their confidence in analysing data, spotting patterns and extracting useful information, they can start to develop some of the skills employers are increasingly seeking.

Ask the group if they have heard of the term 'big data', and ask some questions to lead a discussion. Can they explain what big data is?

How is it different from just data?

Have they heard it mentioned in the news, or in a job context?

Why is it becoming an increasingly discussed phenomenon?

To gauge students' starting points and enable them to demonstrate their progress later, ask them to rate their knowledge, skills and understanding of big data on a 'confidence continuum'. Or invite them to record their thoughts in the form of a mind map, drawings or a graffiti wall; either of these activities can be revisited at the end of the lesson to demonstrate students' progress.

Some key elements to draw out:

- Big data is a term for large amounts of layered data that can be analysed to reveal patterns and trends, which can be used to inform decisions
- Using data from multiple sources means that deeper and more accurate information could be gained
- Much of the activity in our daily lives leaves behind a large volume of data, such as searching online, using debit or credit cards, tracking a workout or even getting the bus

2. Explore how big data is used

Launch the [Big Data interactive tool](#). Click 'Start', and open the 'Big What?' section.

As a group, work through each of the tabs in this section to introduce the concept of big data, reading and discussing the information presented.

Pause at the tab titled 'Where does it come from?' After reading the text, ask the group to consider what data someone might have generated through their actions today, from waking up this morning to going to work or school until now. How much data might have been used or generated in this time?

For example, they may have viewed the news via an app, paid for something using a contactless bank card or smartphone, planned a route using a map app, used a fitness tracker, checked in or posted on social media.

Pause at the tab titled 'What do we do with it?' Considering the examples displayed, allow a few minutes for small groups to come up with other ways that they think big data is being analysed and used to improve experiences.

Once their ideas have been shared, return to the three examples of big data in action on the tab. Ask the groups to suggest the benefits and potential challenges that come with each.

Activity one

Big what? (cont'd)

3. How big is big?

Examples	Benefits	Challenges
Health data	Can speed up and improve patient diagnosis	Could be used to dispute a health insurance claim
Online shopping consumer data	Customers see relevant advertising and similar products	Companies have high volumes of information, customer preferences and personal details – if there was a data breach this could be exposed
Train ticket data	Can provide real time journey updates	Over reliance on quantitative data in place of direct feedback from customers which could lead to a misinterpretation of the data, such as making a causal link when in fact there is a coincidence

Activity two

How is big data processed, stored and analysed?

1. Storing big data



This section is to help students understand how big data is processed, stored and analysed. You can use the infographic in the Appendix of this discussion guide to explain the process, by either displaying it on a screen or printing copies to share.

Some key elements to draw out:

- When the amount of data becomes too large to store and process on one computer, multiple processors are utilised in parallel
- There are different tools available which deal with the storing and processing of big data, distributing it across many different locations (usually storage units inside a data centre)
- Once the data has been processed, visualisations help to reveal useful trends and patterns

2. Big data visualisations

There are plenty of examples of big data visualisations available across the internet. The options provided below all allow for personalisation so they can be used to reveal information about a particular area or category.

If the group has access to multiple PCs or tablets, ask them to explore the visualisations in pairs and answer the questions. Alternatively, select one of the options to explore as a whole group on a screen or whiteboard.

Allow ten minutes to explore the visualisation and answer the following questions. If they are working in pairs, ask them to feed back on what they have discovered to the whole group.

Can you ascertain the source of the data?

Activity two

How is big data processed, stored and analysed? (cont'd)

Is it updated in real-time or is it a snap shot of a particular point in time?

What insights can be gained from the visualisation?

How is it helpful and who might it be useful for?

How important is it that you can substantiate the information presented?

How might the data be misinterpreted?

The Internet of Things

Shows how data is being used to improve user experiences and increase business profitability across a range of categories

onemilliontweetmap

Maps the most recent one million tweets from across the world in real-time – around fifty new tweets are added every second. You can play with filters, adjust to a heat map, and drill down to geo-localised tweets. Ask students to zoom in on their local area and compare with somewhere further afield

Earth wind map

Discover world weather patterns using live satellite data

Commuting patterns

Select an area and see how far people commute to and from that area each day

Data breaches

View the size of data breaches and hacks of major companies over the last few years

3. Big data in action

Return to the online tool and launch the second section. Work through the examples of big data in action presented through Harry's story. At each stage of the journey allow time to discuss and use the prompts below to check for learning.

Getting a bus

This is a simple way of explaining and visualising how big data is generated and how multiple sources are used in tandem to provide a better experience for users.

For each of the data sources displayed, ask the group to suggest how each of them contribute to an accurate estimated arrival time.

- The bus's average speed measured by sensors or GPS
- Crowd-sourced traffic data GPS on smartphones
- The planned bus route and distance between stops
- Average passenger numbers from historical ticketing data
- Real-time weather data

Activity two

How is big data processed, stored and analysed? (cont'd)

Music streaming

After looking at the data sources used for targeted advertising, encourage the group to explore some of the questions below.

What are the benefits of data being used to target the advert for Harry?

E.g. Harry hears more relevant ads, easy for him to find out about local events and things he is interested in, the advertising is more likely to be successful as it reaches the target audience

What are the possible implications of using data in this way?

E.g. the app has access to lots of his personal information, and connected devices, which could put him at risk of identity theft if there was a security issue, Harry may only find out about related things – the concept of the 'echo chamber'. This describes a situation where one set of information or opinions is reinforced and repeated, whilst alternative views are underrepresented

- You could explore the implications further by searching online for recent articles relating to how big data is being used by businesses, and how data is collected on social media platforms

Going to a festival

After following Harry and his friend at the festival, break into small groups and set the scene for this activity.

Imagine that you are a small food and drink company and this is your first time hosting a stall as a vendor at the festival.

Allow groups two minutes to think about what products they think would be suitable to sell. Then ask each group to consider the following questions:

What data would you want to analyse before the event?

E.g. weather conditions, the number of festival tickets sold, demographics of attendees, which other vendors are booked

What data could be collected during the event?

E.g. sale trends, peak times, leftover products

How could you collect this data?

E.g. recording sales electronically, tracking contactless payments

How could you use the data that was collected to improve your productivity at future events? E.g. buying more of the most popular products, changing your advertising, honing your target audience, choosing which events to attend

Activity two

How is big data processed, stored and analysed? (cont'd)

Extension

Much of the big data discussed so far in this session is collected and analysed to improve customer experience. In the medical and health sector, it is being used to save lives and transform diagnostics.

One example is how scientists are using computers to manipulate big data in the form of DNA code when diagnosing the rarest conditions to treat patients more accurately

Can the group think of any other ways that big data could be used in the medicine and health sector? E.g.

- Some hospitals are providing real time big data to help patients get the fastest emergency treatment – sharing how many people are waiting at each hospital in a region, the number of staff on shift, and average wait times
- Combined with traffic updates, this information can also be shared with paramedics and ambulance drivers to inform their route and destination A&E.

4. Big data and the workplace

Discuss the idea that there is a predicted skills gap in the UK, particularly in the area of big data, including roles that involve data collection, processing and analysis, and young people need to be equipped to meet the needs of this future job market.

However, highlight that as more data is generated and increasingly relied on by businesses, analytical and digital skill sets are needed across many careers – even those you might not associate with data, computing or numeracy. Examples include:

- Geographers require skills to be able to gather, interpret and communicate big data, whether this be related to weather, changes to the physical environment, or population and migration trends
- Graphic designers will need to use their creative skills to communicate findings from analysis of big data with non-technical audiences
- Geographer James Cheshire and designer Oliver Uberti have combined data interpretation and design to produce a book of maps and graphics all about London. You can show some of their images [here](#)

There are plenty of examples of big data visualisations available across the internet. The options provided below all allow for personalisation so they can be used to reveal information about a particular area or category.

If the group has access to multiple PCs or tablets, ask them to explore the visualisations in pairs and answer the questions. Alternatively, select one of the options to explore as a whole group on a screen or whiteboard.

Allow ten minutes to explore the visualisation and answer the following questions. If they are working in pairs, ask them to feed back on what they have discovered to the whole group.

Activity two

How is big data processed, stored and analysed? (cont'd)

Can you ascertain the source of the data?

Is it updated in real-time or is it a snap shot of a particular point in time?

What insights can be gained from the visualisation?

How is it helpful and who might it be useful for?

How important is it that you can substantiate the information presented?

How might the data be misinterpreted?

5. Your turn with big data

Return to the online tool, and launch the third section. Real data from three areas of the UK has been selected for students to analyse in order to answer a series of questions. You could work through the questions as a whole class, or, if students have access to tablets/PCs, they can work individually or in small groups.

Encourage students to discuss the answers in pairs or groups, before sharing their decisions.

The questions and answers from the tool are listed below. You'll also find follow-on questions should you want to prompt further discussion points after running through this section of the tool

1. A leisure centre company is looking to open a new gym in a city where there is a low number of gyms compared to other types of businesses. Where do you think the company would choose based on this data alone?
Answer: Newport
2. A clothes retailer has asked you to identify whether consumers across the three locations prefer to buy their clothes online or in traditional physical stores. Based on the data provided which one would you recommend?
Answer: Physical store
3. From the data on cash purchases alone, which area would you say would be the least profitable to set up a shop or market stall that only accepts cash payments?
Answer: Oxford
4. A group of investors in Oxford wants to identify the two best business sectors to start investing in based on increased spending by consumers this year. Which two sectors would they be likely to choose?
Answer: Leisure and travel
5. A large supermarket chain is looking to open a new store in one of the locations. Where would be the best location to open the store based only on the frequency of visits?
Answer: Aberdeen
6. If councils in the three areas change business rates (taxes) in line with changes in average business turnover, which location would you expect to have a decrease in these business rates?
Answer: Oxford
7. A home insurance company is looking to target sales in locations with high mortgage spend – based on the data which of the three locations would be of most interest to the insurance company?
Answer: Aberdeen
8. Which city has seen the biggest reduction in spending across the <£15k income band?(Income bands are a way of grouping people by their annual income)
Answer: Aberdeen

Activity two

How is big data processed, stored and analysed? (cont'd)

9. The manager of a pharmacy in Oxford calculates that the average card transaction in their store is £20.45. Is this value higher or lower than the city average for their industry?
Answer: Below
10. An electronics business with stores in Oxford and Aberdeen wants to identify the age band that it should target new local marketing campaigns at to increase revenue. Which age band has shown a reduction in spending across both cities?
Answer: 60+

What data could be collected during the event?

E.g. sale trends, peak times, leftover products

How could you collect this data?

E.g. recording sales electronically, tracking contactless payments

How could you use the data that was collected to improve your productivity at future events?

E.g. buying more of the most popular products, changing your advertising, honing your target audience, choosing which events to attend

6. Explore the following questions

Once you have completed the tool, you could ask some of the following questions to encourage discussion around the uses of big data.

How can big data analytics tools be used to help businesses strategically plan and be more successful?

Do businesses need to 'look beyond the numbers'? E.g. would it be better for shops of the same type to cluster together, or look for audiences in new areas?

How can selection and simplification of certain stats help to prove/support a viewpoint? Could this lead to bias?

What is targeted advertising? Is it beneficial to the consumer to only see things that are relevant to them?

Activity three

The big data sector

1. Jobs in big data

Use this case study of a fictional company and ask the group to consider each of the job profiles within the big data sector, which can be found on the **Big data and the workplace: case study student worksheet**.

A large clothing and footwear retailer sells products both instore and online, with every transaction generating huge amounts of data. They want to increase sales and offer their customers a personalised experience and believe that spending time and money on data will help them do this.

Each of the roles on the student sheet demonstrates the skills required and gives an example of the tasks they might complete. For many jobs within this sector, young people will need to be able to gather and interpret information, as well have excellent organisational and interpersonal skills.

- Big data scientist
- Customer insight analyst
- App developer
- Big data visualiser
- Product designer

Students should choose one of the roles featured and create a job advert for the position, describing the skills they need. They can also create a CV or covering letter applying for that job. Less able students may want to focus on the job description activity and look at the skills required in more depth, which they could record as a leaflet, presentation or report.

More about how technology is influencing the world of work can be found in our guide to the top [Your guide to the jobs of the future](#)

Ask students to revisit their initial thoughts from the mind map, drawing or graffiti wall. They should now add anything new they have learnt about big data or make any changes to their initial ideas using a different coloured pen.

For students studying humanities subjects, you could encourage them to think of themselves as data scientists, by exploring the Zooniverse, the world's largest platform for people-powered research. Professional researchers from around the world publish projects and ask for volunteers to help classify and analyse large volumes of data.

Visit [zooniverse.org/projects](https://www.zooniverse.org/projects) and showcase some of the options for the subjects that might be relevant to your students.



Activity three

The big data sector (cont'd)

Try next

As a follow up, you can use the following related resources on the LifeSkills website. Suggest that young people spend some time exploring the tips and interactive content

Tips and interactive content

How the world of work is changing

(barclayslifeskills.com/futureworkplace)

Choosing your next steps

(barclayslifeskills.com/nextsteps)

Lesson plans

Understand and identify the core transferable skills employers are looking for

(barclayslifeskills.com/transferableskills)

If students have any anxieties or concerns about how their own data is used, signpost them to the following links about staying safe online, getting their social media profile employer ready and online fraud.

barclayslifeskills.com/staysafeonline

barclayslifeskills.com/socialmediaready

barclayslifeskills.com/phishingfromyoursmishing

Summary

Ask the group to share what they have discovered about big data, and recap the session objectives. As a summary activity, ask small groups to carry out some further online research about big data and create their own infographic, visualisation or presentation about what they have learnt.

For example, they could include:

- How many bytes of data does the world generate each day?
- How many jobs will involve data by 2030?
- By 2020, what percentage of cars are expected to have built-in connectivity?
- What is the estimated size of the big data skills gap?
- Choose a job website. How many jobs are currently advertised that include big data in the role title?
- Why is it important for us as individuals to keep some data private?
- How could knowledge of big data and its use help a consumer to make decisions?
- How could knowledge of big data and its use improve a person's employability?

Big data: Big what?

Big data and the workplace: Case study

A large clothing and footwear retailer sells products both instore and online, with every transaction generating huge amounts of data. They want to increase sales and offer their customers a personalised experience and believe that spending time and money on data will help them do this.

The following roles are examples of those that a company could employ to use their data more effectively.

Big data scientist

The retail company decides to use machine learning to explore which products were sold together and offer recommended items at checkout in the hope of boosting sales. The computers use algorithms and large volumes of historic data to predict and calculate trends, like spikes in sales of certain products, or when stock is likely to run out.

The big data scientist needs to design and implement ways of storing and processing large amounts of complex data to solve problems and improve service for users, such as:

- Detecting fraud
- Real-time pricing
- Product recommendations

Daily tasks may involve cleaning data, testing and training a model.

Skills needed:

- ✓ Digital
- ✓ Analytical
- ✓ Communication

Familiar with machine learning techniques and have a basic understanding of statistics.

Customer insight analyst

The company run several reports, which provide results such as top 10 products sold online vs instore. A customer insight analyst investigating these reports found that the most popular product instore was not in the top 10 sold online. To find out why this could be the case, the analyst looked at website traffic data and saw that the product was in the top 10 most viewed, but wasn't being purchased. They discovered that the product was priced incorrectly, listed for £500 instead of £50. The analyst was able to correct this error and improve sales.

Customer insight analysts also need to be able to interpret visualisations of big data from multiple sources, in order to understand the needs and behaviour of customers to make recommendations to improve a service or product.

Skills needed:

- ✓ Digital
- ✓ Analytical
- ✓ Problem solving
- ✓ Communication

Confident using databases, and ability to communicate insights to different audiences, including those who are non-technical.

App developer

The application developer took the knowledge gained by the customer insight analyst and results provided by the data scientist and used this to improve the online experience for customers. They developed a recommended products panel to be displayed under each product, and increased the number of servers needed during predicted website traffic peaks, e.g. when a sale is on, or after a marketing campaign.

Skills needed:

- ✓ Digital
- ✓ Problem solving
- ✓ Team work
- ✓ Creativity

Knowledge of programming languages, and ability to work closely with clients and users to ensure that their needs are met.

Big data: Big what?

Big data and the workplace: Case study

Big data visualiser

Visualisations and infographics are a useful way of showcasing data, and it's the visualiser's job to make the outcomes of the analysis that has been done easy for everyone to understand. The company decided to install a screen in the office displaying key statistics like number of customers instore per day, sales, most popular store, or most visited website page. This helped other employees make decisions, like the best location for the next retail store.

Skills needed:

- ✓ Problem solving
- ✓ Creativity
- ✓ Digital

Knowledge of computer-aided design, and the ability to communicate insights to different audiences in creative ways, including those who are non-technical.

Product designer

The product designer may not work with big data directly, but needs to understand which products are most popular and why, in order to design new products that meet the needs of the customer and keep sales for the company high. The analysts and developers will need to communicate this information to the designer.

Skills needed:

- ✓ Problem solving
- ✓ Creativity
- ✓ Project management
- ✓ Teamwork

Knowledge of computer-aided design, and the ability to explain creative ideas to others.



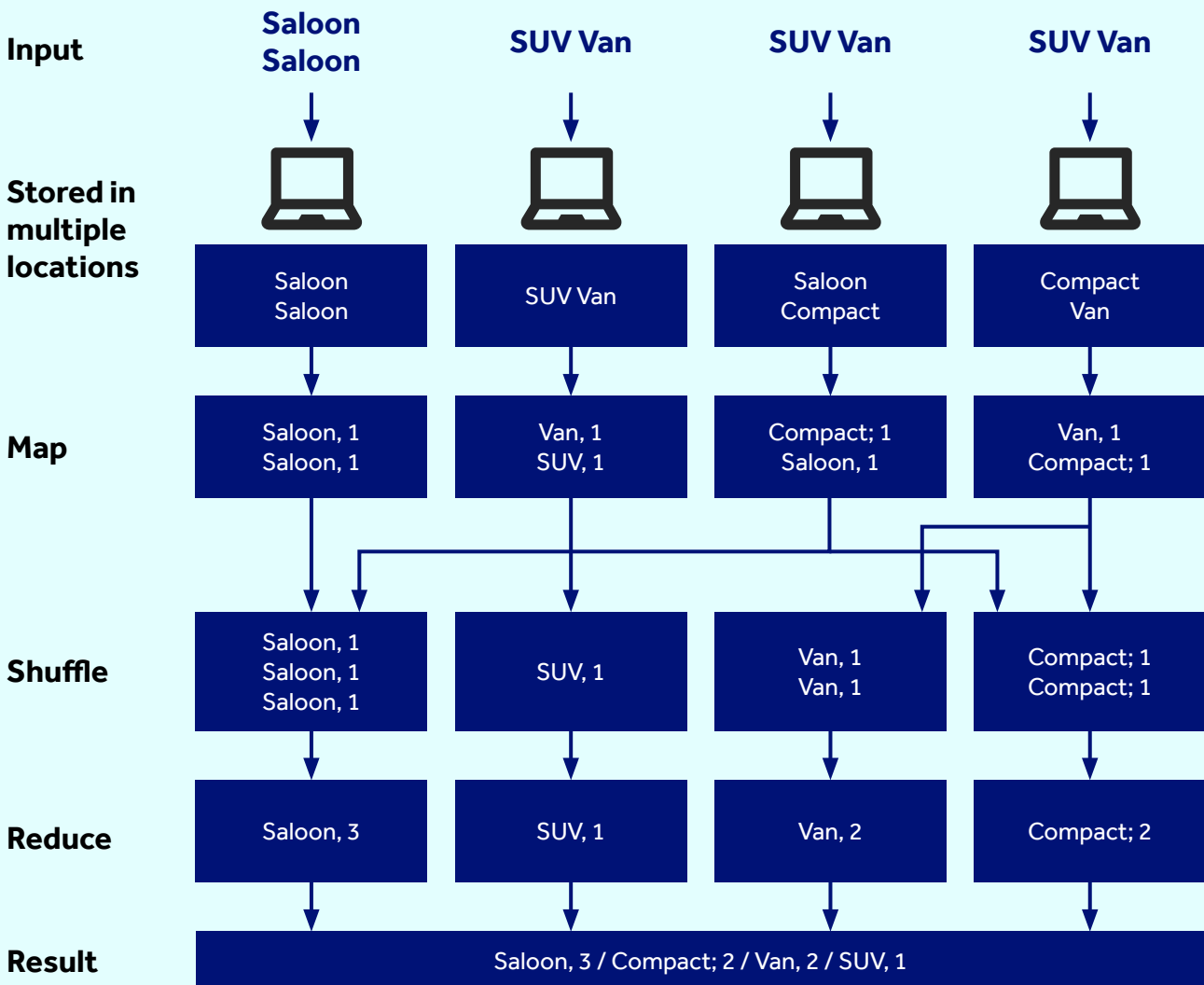
Big data: Big what?

Appendix: How Big Data is processed

Apache Hadoop is an open source framework which stores multiple data sources in one place; both structured data (e.g. a shop's product database) and unstructured data (e.g. transaction logs and website traffic). Hadoop stores the entire dataset in small fragments distributed across a collection of servers. The tool then analyses the data and reports results via a process called MapReduce.

How is big data processed using MapReduce?

Imagine you have a list of recently sold cars so big that there is too much information to store on one computer. You want to find out which is the best selling type.



Why use a big data tool

- ✓ Low cost: Hadoop is 'open source' which means it can be modified and enhanced by anyone
- ✓ Scalability: Hardware can be added automatically if you need to grow the amount of data you are working with
- ✓ Speed: Increased by processing in parallel
- ✓ Flexibility: There is no need to pre-process data before storing it
- ✓ Troubleshooting: If hardware goes down, the task is automatically redirected to working hardware by Hadoop