1		**6 marks for AO1 (understanding)** **Area 1: How it could work:** • members could specify their interests / views and stories could be matched to these; • consider basic facts about member eg age, gender, location; • consider what stories have been read by friends of the member; • analyse the type of stories that the user has read before; • analyse the information that a member shares about themselves to identify characteristics/interests etc; • track how popular a story is to display the most popular ones; • look at member's search history; • look at member's reaction to other similar stories eg likes; • show stories viewed by others with a similar profile to this user; • display articles that have been more popular // had more hits // received more positive feedback; • compare keywords in articles with keywords in articles previously viewed by the member; • how can the algorithm avoid displaying click-bait?; **Area 2: Legal** • who owns the copyright in the story?; • is it legal for the company to reproduce a news story that someone else has written?; • is the company legally responsible for the content/accuracy of stories?; • do contracts need to be signed between the company and the organisations/ individuals that stories will be displayed from?; • do laws in some countries prevent some types of stories being displayed? // need to ensure laws in different countries are followed; • need to ensure that stores are age-appropriate; • need to notify members about how their information is being used to select stories; **Area 3: Ethical / Moral** • by choosing what news stories to display, will the service influence the views of members?; • how should the company deal with governments/organisations who might want to influence/control which stories are displayed?; • should the company accept payments to promote stories?; • how should the company deal with complaints / issues raised by members (in a timely fashion)?; • will the reproduction of news stories adversely (or positively) affect the number of people who go to read the original stories from their authors?; • how can / should reliability of stories be checked / shown (fake news)?; • how can / should the company assess bias / prevent spread of propaganda; • does the company have a duty to try to provide balance?; • should a method be provided so members can request their data is not analysed for this purpose? // importance of consent; • should the company let them know that the news they are seeing is being tailored to them / not everyone sees the same news?;	**6**

		Area 2 or 3: Legal OR Ethical / Moral • should the company have people who read/check each story?; Is it practical to do this?; • how should the company select which organisations/individuals it will display stories from?; **Max 4 if all points are from one area**	

<table>
<tr><th>Qu</th><th>Pt</th><th>Marking guidance</th><th>Total marks</th></tr>
<tr><td>2</td><td></td><td>

All marks AO1 (understanding)

<table>
<tr><th>Level</th><th>Description</th><th>Mark Range</th></tr>
<tr><td>4</td><td>A line of reasoning has been followed to produce a coherent, relevant, substantiated and logically structured response. The response covers all three areas indicated in the guidance below and in at least two of these areas there is sufficient detail to show that the student has a good level of understanding. To reach the top of this mark range, a good level of understanding must be shown of all three areas.</td><td>10–12</td></tr>
<tr><td>3</td><td>A line of reasoning has been followed to produce a coherent, relevant, substantiated and logically structured response which shows a good level of understanding of two areas indicated in the guidance below or a good level of understanding of one area and a reasonable level of understanding of the other two areas. To reach the top of this mark range, a good level of understanding must be shown of two areas.</td><td>7–9</td></tr>
<tr><td>2</td><td>A limited attempt has been made to follow a line of reasoning and the response has a mostly logical structure. A good level of understanding has been shown of at least one area or some understanding has been shown of all three areas.</td><td>4–6</td></tr>
<tr><td>1</td><td>A few relevant points have been made but there is no evidence that a line of reasoning has been followed. The points may only relate to one or two of the areas from the guidance. There is insufficient evidence of a good understanding of any of the three areas.</td><td>1–3</td></tr>
</table>

<u>Guidance – Indicative Content</u>

Area 1: What Big Data is

Overarching description: Data that can't be processed or analysed using traditional processes or tools.

<table>
<tr><th>Characteristic</th><th>Expansions / Examples</th></tr>
<tr><td>Variety of different forms of information // data may lack structure</td><td>Cannot be represented in a table // by a relational database

Email messages
Videos
Images
Web site contents
Facial recognition</td></tr>
</table>

</td><td>12</td></tr>
</table>

There is a lot / high volume of data (to process as one dataset) // data will not fit on one server	Hundreds of terabytes Large medical datasets for diagnosis Gene sequencing Predicting disease outbreaks Results of large-scale scientific experiments
The data is generated / received / must be processed at high velocity / very quickly	Thousands of items to process per second. Data must be processed as it is received – it cannot be batched and processed later Card payment fraud detection Recommendations systems

Good level of understanding = Either all three characteristics covered or two characteristics and the overarching description. Some examples or expansions covered.

Area 2: Challenges and How Overcome

Challenges:

- Data cannot be stored on one server / computer.
- Not possible to process data quickly enough with one computer.
- Data cannot be represented in a table // by a relational database.
- Some forms of data / unstructured data are difficult to analyse.

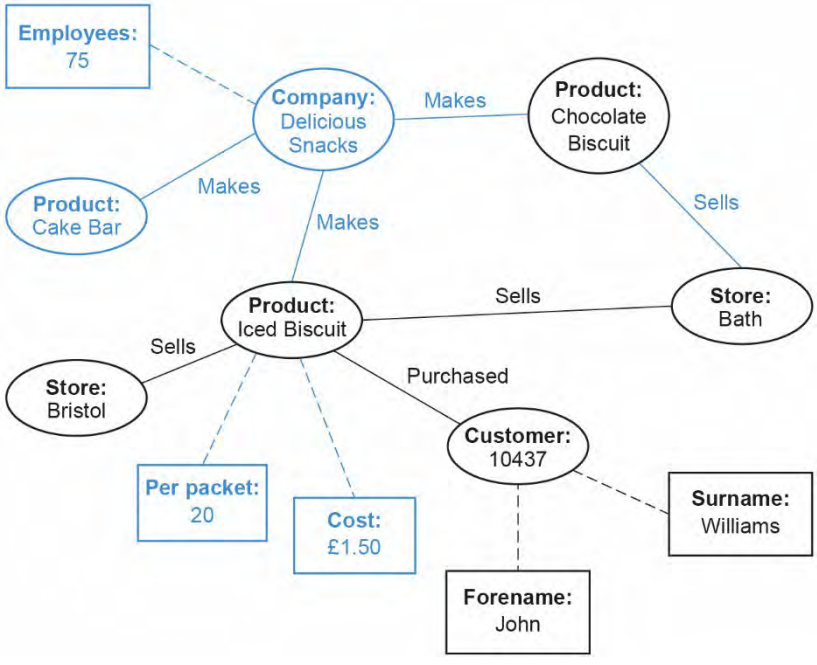
How overcome**:**

- Distributed database systems // distributed file systems // blocks of individual files distributed across multiple servers.
- Use of functional programming.
- (Massively) parallelising the execution of programs.
- MapReduce // input split into parts then mapper executed on each part then all results combined by reducer(s) // function-to-data model.
- Functional programming makes it easier to write distributable code // determine which parts of code can be run independently.
- Functional programming makes it easier to write correct code // example features of functional programming that facilitate writing correct code
- Use of many thousands of commodity servers.
- Use of servers with multiple CPUs / cores / drives.
- Machine learning can identify patterns / the value in the data // use of predictive data models.
- Use of languages such as XML or JSON to describe semi-structured data.

		• Use of fact-based model can manage bigger data sets better than relational model. *Good level of understanding = A range of challenges and how to overcome them are discussed.* **Area 3: Ethical and Legal Issues** • How can data be kept securely? • Who should have access to what data? • Will people know what data is being stored about them? • Where should / will the data be stored // concerns relating to data being stored in other countries. • What rights do people have in relation to data stored about them? • Example laws (allow two examples): Computer Misuse Act, General Data Protection Regulations / GDPR / Data Protection Act, Regulation of Investigatory Powers Act / RIPA. • Who owns data about individuals? *Good level of understanding = A range of issues described*	

Qu	Pt	Marking guidance	Total marks
3	1	**Mark is AO1 (knowledge)** The data is generated/received/must be processed/responded to at high velocity/very quickly; **NE.** "velocity" on its own. **NE.** High velocity of data **NE.** Speed data transmitted/sent at **A.** "changed", "modified" or similar instead of "processed" but **NE.** accessed **A.** An example of what very quickly would be eg milliseconds, but not a long time period eg seconds **A.** Instantly/immediately for very quickly as **BOD**	1

Qu	Pt	Marking guidance	Total marks
3	2	**All marks AO2 (analysis)** **1 mark** for representing "The Bath store sells chocolate biscuits" with a solid line joining the store and product labelled "Sells". **A.** Alternative labels which clearly have the same meaning eg "Stocks" **A.** Use of directed arrow **1 mark** for representing "There are 20 individual biscuits in a packet of iced biscuits and each packet costs £1.50" by drawing rectangular boxes containing this data, connected to the Iced Biscuits with dashed lines. **R.** Both pieces of data in in one box **1 mark** for representing "Both chocolate biscuits and iced biscuits are made by the company Delicious Snacks. The company has 75 employees and also makes cake bars." by: • adding an oval for the company Delicious Snacks • connecting a rectangle to the company oval by dashed lines indicating the number of employees • adding an oval for the new product Cake Bar • linking the oval for the company to all three products using solid lines and giving these an appropriate label eg "makes". **A.** Alternative labels which clearly have the same meaning eg "Manufactures" **A.** Use of directed arrows **I.** Incorrect spelling as long as meaning is clear. **DPT.** Use of incorrect line styles (solid/dashed) **DPT.** Wrong shapes used for rectangles and ovals **DPT.** Missing labels on lines **DPT.** Labels on dashed lines	3

		An example fully correct diagram is shown below:  	

Qu	Pt	Marking guidance	Total marks
3	3	**All marks AO1 (understanding)** Immutable data structures // the state of a data structure cannot be changed (after creation); Statelessness // functions do not have side-effects // all functions are pure; Functions can be distributed to servers and executed on data sets then the results can be combined // map-reduce; Higher-order functions can compose the results of processing on multiple processors/cores // functions are first-class objects; The order of execution can be determined at run-time // the order of execution can be determined by the translator/compiler/interpreter (**A.** language) // the order of execution is not defined by the program code // programs are not a sequence of instructions that must be followed in a specific order; **NE.** Suitable for parallel processing **Max 2**	2

Qu	Pt	Marking guidance	Total marks

4	1	**All marks are AO1 (understanding)**
		
		(Data structures/variables are immutable which means that) the state/values stored in data structures/variables cannot be changed (after they are created) // functional programming languages do not have variables;
		
		(Functions / programs are stateless which means that) functions do not have side-effects // the output of a function depends only on its inputs // functions are pure // the output of a function is not influenced by a stored state;
		
		Higher-order functions can compose the results of processing on multiple processors/cores // higher order functions can take a function as an argument and apply it to every element in a list // map-reduce can be used // functions are first-class objects and so can be passed to other functions as an argument;
		
		The order of execution can be determined at run-time // the order of execution can be determined by the translator/compiler/interpreter (**A.** language) // the order of execution is not defined by the program code // programs are not a sequence of instructions that must be followed in a specific order;
		
		Max 2
			2