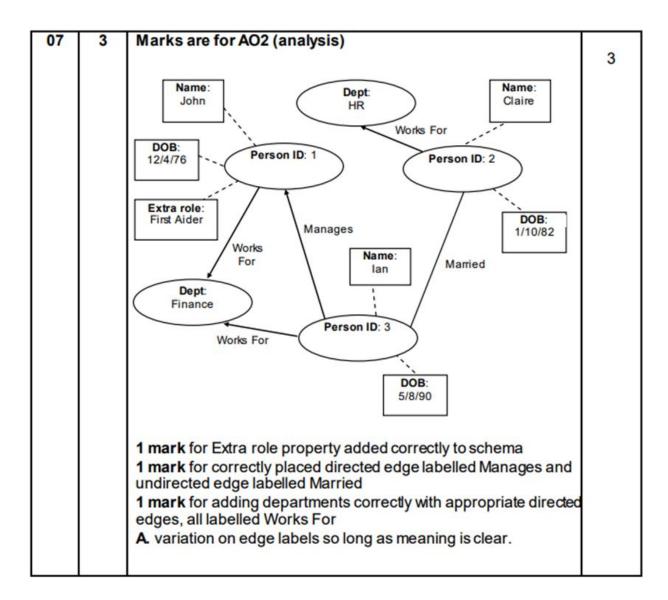# AQA Computer Science A-Level
4.11.1 Big Data
Past Paper Mark Schemes

Additional Specimen Paper 2

07	1	**Marks are for AO1 (knowledge)** The velocity/speed at which the data is generated/received; The variety/range of data types in the data set (**A.** examples) // unstructured nature of data; **N.E.** Just the words "velocity" or "variety"/"range"	2

07	2	**Marks are for AO1 (understanding)** Data in a functional language is immutable // there are no side-effects of computations // one part of a functional program cannot change data and thus affect another part; Map/reduce operations can be easily parallelised // functional programming operations are often collection oriented (so can be easily parallelised); Order of execution less rigidly defined in a functional language than for procedural/object-oriented/other paradigms; **MAX 2**	2

| 07 | 3 | **Marks are for AO2 (analysis)** | 3 |

1 mark for Extra role property added correctly to schema
1 mark for correctly placed directed edge labelled Manages and undirected edge labelled Married
1 mark for adding departments correctly with appropriate directed edges, all labelled Works For
A. variation on edge labels so long as meaning is clear.

Specimen Paper 2

04	1	**All marks AO2 (apply)**		12

Level	Description	Mark Range
4	A line of reasoning has been followed to produce a coherent, relevant, substantiated and logically structured response. The response covers all four areas indicated in the guidance below and in at least three of these areas there is sufficient detail to show that the student has a good level of understanding of the technologies required. A good level of understanding would be indicated by three substantiated points being made per area. To reach the top of this mark range, a good level of understanding must be shown of all four areas.	10-12
3	A line of reasoning has been followed to produce a coherent, relevant, substantiated and logically structured response but the response may only cover three of the areas indicated in the guidance below, with two or three substantiated points being made per area.	7-9
2	A limited attempt has been made to follow a line of reasoning by covering at least two of the topic areas in the guidance below. Overall, at least four valid points must have been made which can relate to any of the topic areas in the guidance.	4-6
1	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 four areas from the guidance or may be made in a superficial way with little substantiation.	1-3

Guidance – Indicative Response

1. Fridge capturing data from food

RFID well suited as completely automatic short-range wireless transmission so no user involvement

- tag does not contain a power source but is energised by reader in fridge
- this causes wireless transmission of data stored in memory on tag to reader

Alternatively, scan barcode/QR code as food put into fridge

Barcode less suitable than RFID as only identifies product not use by date and must be manually scanned

Problem of how to deal with untagged produce – possible use of voice recognition or touch screen interface

Can identify products and potentially track use by dates, but how to work out how much of the product is left – refrigerators redesigned with load cells to weigh items automatically?

2. Networking technologies

IPv4 does not have a big enough address space for the number of devices, hence introduction of Ipv6

Higher bandwidth Internet connections required for so many devices

- copper-based transmission systems replaced with fibre optic

Need for a standard (application layer) protocol for devices

Security issues with many devices connected to Internet that could be hacked

Would data be communicated to retailers directly from each device or through a server in the home?

Need to consider how to deal with interference between wireless devices, collisions etc with many more devices communicating

3. The data gathered and storage

Automatic collection of data from devices will produce vast amounts of data

This volume of data would be classified as big data

May also be classified as big data due to the velocity of data collection with so many devices

Storage could be cloud based for flexibility or close to processing cores for speed

Velocity at which data generated would make solid state storage appropriate as has fast access speeds but volume of data and lower cost per megabyte of hard disk storage may mean hard disks more likely to be used

Need to consider how long to keep data for in context of

- Storage capacity available
- Complying with relevant laws about privacy

4. Processing

		Volume of data means parallel processing or distributed processing architectures required	
		Volume of data collected makes it unsuitable for processing by traditional relational databases	
		Functional programming is one approach that could be used	
		Functional programming appropriate as works well on parallel processing systems as programs do not specify order of execution	
		Would software that managed contents of the fridge be run as embedded system in fridge or in the cloud / by the retailer?	
		Retailers may develop a standard API to interface with devices	