# Homework 6 Big data Answers

1. Big Data can be represented using graph schema. Here is part of such a schema relating to a sensor network.



 Complete the graph schema to represent the following facts:

 (a) Any sensor with an ID starting “20-“ has the function “Temperature” [1]

 (b) The smoke alarm is located in the kitchen [1]

 (c) The towel rail is currently heated to 32 degrees Celsius [1]

 (d) Sensor 20-31 is connected to a door sensor in the same location. [2]

 (e) XYZ alarms are contracted to service all smoke alarms [1]



Connected to

Minor variations are possible: mark for the essential details and accept eg undirected edges

2. Big Data often involves a large volume of data. Explain two other distinguishing
features of Big Data and give examples. [6]

Velocity (1) where data is collected/captured/generated/streamed and needs to be processed in near real time (1) for example when managing traffic flow on a smart motorway/using facial recognition with CCTV to identify suspected criminals (1)

Variety (1) where data is in many forms such as structured, unstructured, text or multimedia (1) for example when analysing social media response to a film/seeking patterns of criminal behaviour (1)

Accept other well-argued ideas.

3. Describe an appropriate computer architecture for processing Big Data and
explain why it is appropriate. [4]

Parallel or distributed (1) architecture involves dividing processing tasks between multiple processors (1) and allows scaleable (1) and resilient (1) solutions. Massive amounts of data to be processed requires many processors working in parallel. (1)

4. Functional programming languages contain map and fold/reduce functions, and map-reduce algorithms are often used to process Big Data.

 (a) Explain the general principle of a map-reduce algorithm. [2]

 A processing task is subdivided by filtering and sorting (1) and then applying a map function to data elements in parallel (1)so that a reduce method can then be applied to calculating or summarising and returning a result (1) (max 2)

 (b) Explain why these algorithms are commonly used with Big Data [2]

 This approach allows (optimisation of) parallel processing (1) , which is an efficient/fault-tolerant/scaleable (1) way to deal with Big Datasets.

 [Total 20 Marks]