4.11. Big Data – Test 1

1. It has been calculated that 90% of the world's data has been gathered and created in the last two years. This explosion in the amount of data we gather and need to store has led to new possibilities and problems in the modern world. This phenomenon is known as Big Data.

   a) The 'big' in Big Data represents the scale of the data in terms of the demands it puts on our systems. Other than the large volume of data (and the necessary storage capacity required to host it), describe two other demands Big Data may have. [4]

      1..

      ....

      2..

      ....

   b) The main challenges that Big Data brings are due to its lack of structure. Why does this lack of structure pose a problem? [2]

      ....

      ....

      ....

   c) The huge size of data being stored often means it cannot be stored on one server. Why is this a problem and what method can be used to help use data that is distributed across multiple servers? [2]

      ....

      ....

      ....

   d) Describe three developments in technology in recent years that have led to the creation of the Big Data phenomenon. [6]

      1..

      ....

      ....

      2..

      ....

      ....

      3..

      ....

      ....

e) George Lee (CIO at Goldman Sachs) referenced the fact that for years society has used 'experiential wisdom' (learning from and gaining wisdom from one's experiences) to make decisions in many areas of our lives. How does Big Data contribute to this way of thinking? Explain how Big Data can potentially be used in Medicine, Agriculture and Manufacturing (you may choose any area of manufacturing to talk about).   [14]

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Total marks =   /28

4.11. Big Data – Test 2

1. Functional programming languages are widely used to process Big Data.

   a) `Map` is a common function in Big Data applications. `Map` is an example of a higher-order function. What is meant by a higher-order function? [1]

   ....

   ....

   b) Big Data frameworks such as Apache Spark can distribute higher-order functions across multiple physical servers. Explain why it is possible for this to be done safely when a purely functional programming language is used. [2]

   ....

   ....

   ....

   ....

2. Stock exchanges update their prices every few microseconds. The following table gives some example data that might be produced by a stock exchange.

Timestamp	Stock Name	Min Price	Average Price	Max Price
612090	MGX	102	105	108
612094	MGX	102	104	108
612097	APP	80	83	85
612099	MGX	100	102	108
612102	GTP	3	3	3
612108	APP	80	84	86

   a) What is meant by a fact-based model? [2]

   ....

   ....

   ....

   b) Give an example of a fact from the table above. [1]

   ....

   c) Describe the elements that make up a graph schema. [3]

   ....

   ....

   ....

   ....

d) Draw a graph that represents the stock exchange data above. [3]
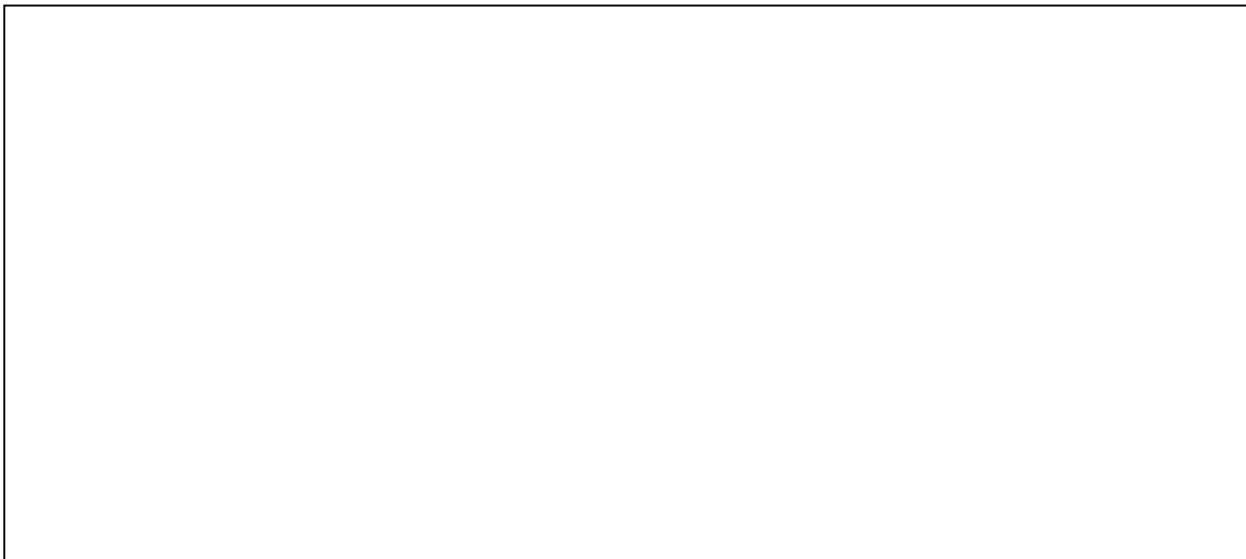
e) Stocks may be listed on more than one stock exchange at a time. Describe how you would need to modify your graph schema to represent the relationships that stocks have with the multiple stock exchanges. [3]

....

....

....

....

3. Smartphones, networked sensors and video surveillance are three modern sources of Big Data. Describe what kind of data these devices generate and describe why this data can be so difficult to store. [5]

....

....

....

....

....

....

....

....

....

4. Search-engine companies such as Google, Bing and Yahoo! generate a huge amount of data. Google alone responded to 1.2 trillion search requests in 2012. Discuss the types of data a search engine might collect and how the search engine could use this information to improve its accuracy or the effectiveness of its advertising. [12]

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Total marks =   /32