Diploma Programme
Programme du diplôme
Programa del Diploma

Markscheme

May 2017

Computer science

Higher level

Paper 2

53 pages

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

This markscheme is **confidential** and for the exclusive use of examiners in this examination session.

It is the property of the International Baccalaureate and must **not** be reproduced or distributed to any other person without the authorization of the IB Global Centre, Cardiff.

The following are the annotations available to use when marking responses.

Annotation	Explanation	Associated shortcut
✔	Good Response/Good Point	(Alt+) 0
✘	Incorrect Point	(Alt+) 1
?	Unclear	
BOD	Benefit of the doubt	
NBOD	No benefit of doubt	
SEEN	Seen	
TV	Too vague	
REP	Repetition	
FT	Follow through	(Alt+) 2
L	(Comp Sci) Language	(Alt+) 3
D	Description	(Alt+) 4
A+	Analysis	(Alt+) 5
REF	Reference	(Alt+) 6
DEV	Development	(Alt+) 7
B+	Balanced argument	
OC	Off course	(Alt+) 8
EVAL	Evaluation	
O	Opinion	
═	Dynamic, horizontal line that can be expanded	(Alt+) 9
∿	Dynamic, horizontal wavy line that can be expanded	
∫	Dynamic, vertical wavy line that can be expanded	
T	Text box	

You **must** make sure you have looked at all pages. Please put the SEEN annotation on any blank page, to indicate that you have seen it.

General marking instructions

1. Follow the markscheme provided, award only whole marks and mark only in **RED**.

2. Make sure that the question you are about to mark is highlighted in the mark panel on the right-hand side of the screen.

3. Where a mark is awarded, a tick/check (✓) **must** be placed in the text at the **precise point** where it becomes clear that the candidate deserves the mark. **One tick to be shown for each mark awarded**.

4. Sometimes, careful consideration is required to decide whether or not to award a mark. In these cases use RM™ Assessor annotations to support your decision. You are encouraged to write comments where it helps clarity, especially for re-marking purposes. Use a text box for these additional comments. It should be remembered that the script may be returned to the candidate.

5. Personal codes/notations are unacceptable.

6. Where an answer to a part question is worth no marks but the candidate has attempted the part question, enter a zero in the mark panel on the right-hand side of the screen. Where an answer to a part question is worth no marks because the candidate has not attempted the part question, enter an "NR" in the mark panel on the right-hand side of the screen.

7. Please ensure you check all scanned pages. The candidate may have answered more than one option.

8. **If a candidate has attempted more than one Option** within a paper, mark all the candidate's work. RM™ Assessor will only award the marks for the higher scoring Option. Once all the work the candidate has attempted has been marked, please click "COMPLETE"; all the other questions from the other Options will auto complete to "NR" for "no response".

9. Ensure that you have viewed **every** page including any additional sheets. Please ensure that you stamp "SEEN" on any page that contains no other annotation.

10. A mark should not be awarded where there is contradiction within an answer. Make a comment to this effect using a text box or the "CON" stamp.

Subject details: **Computer science HL paper 2 markscheme**

Mark allocation

Candidates are required to answer **all** questions in **one** Option. Total 65 marks.

General

A markscheme often has more specific points worthy of a mark than the total allows. This is intentional. Do not award more than the maximum marks allowed for that part of a question.

When deciding upon alternative answers by candidates to those given in the markscheme, consider the following points:

- Each statement worth one point has a separate line and the end is signified by means of a semi-colon (;).

- An alternative answer or wording is indicated in the markscheme by a "/"; either wording can be accepted.

- Words in (…) in the markscheme are not necessary to gain the mark.

- If the candidate's answer has the same meaning or can be clearly interpreted as being the same as that in the markscheme then award the mark.

- Mark positively. Give candidates credit for what they have achieved and for what they have got correct, rather than penalizing them for what they have not achieved or what they have got wrong.

- Remember that many candidates are writing in a second language; be forgiving of minor linguistic slips. In this subject effective communication is more important than grammatical accuracy.

- Occasionally, a part of a question may require a calculation whose answer is required for subsequent parts. If an error is made in the first part then it should be penalized. However, if the incorrect answer is used correctly in subsequent parts then **follow through** marks should be awarded. Indicate this with "**FT**".

General guidance

Issue	Guidance
Answering more than the quantity of responses prescribed in the questions	- In the case of an "identify" question, read all answers and mark positively up to the maximum marks. Disregard incorrect answers. - In the case of a "describe" question, which asks for a certain number of facts *eg* "describe two kinds", mark the first two correct answers. This could include two descriptions, one description and one identification, or two identifications. - In the case of an "explain" question, which asks for a specified number of explanations *eg* "explain two reasons …", mark the first two correct answers. This could include two full explanations, one explanation, one partial explanation *etc*.

Option A — Databases

1. (a) *Award up to [2 max].*
Note: Reasons must be primarily focused on data, on processing (querying), on communications or system design, not on storage.

DBMS allows organisation of data / parents' details in a way that supports efficient and secure access (to parents) / (hence) chances of data loss are relatively low;
Enables specific operations/processing/querying of data (search/sort/analyse);
Permits the production/printing of reports on searched information in specific formats;
Supports the modular extension of system by adding new fields without need of redesigning the system;
Supports the possibility of re-using the resources/tools/scheme for comparable activities (e.g. when the school open a secondary campus);
Relies on widespread use of standards that ensures a smooth integration of systems / so that the school does not struggle with the upgrades of technology;
Automatically supports the data validation/data verification when filling the database;

Award [1 max] if the response only makes a direct comparison with a spreadsheet.
Provides the ability to store a greater number of records than a spreadsheet;
Provides more ability to customise the nature of the information output in records compared to a spreadsheet;
Provides the ability to manipulate the data better than in a spreadsheet, especially on screen;
It supports the operation of data mining, something that would not be possible in a spreadsheet;   **[2]**

(b) *Award up to [1 max].*
Name (*eg* Chris Smith) / Address / Phone number;
Note: Any item of data that can plausibly be stored in a school database and referring to the parents is acceptable.   **[1]**

(c) *Award up to [2 max].*
Award [1] for identifying the technique to reduce data input errors and [1] for a development/elaboration.

Data validation;
Checks that the input is well-formed/appropriately composed/suitable so to be valid for the program specification;
For example, checking that a phone number that is input contains only numbers and asking to re-input it, should it contain other (non-valid) characters / type check (num or char) / range check (months with 30 or 31 days) / length check (prefixed length of some values);

Data verification;
Is the process that checks the accuracy of the data in the database/ checks that the software is correct given the specification / checks the integrity of the database in comparison to the data provided in input / checks in the database the presence of the record associated to the input;

    [2]

(d) *Award up to [5 max].*
Note: The question asks for interactions between parent and DBMS while processing a payment: this should remain the focus during marking.

Example 1:
Award [3 max] if non-specific language is used
*Award [1] for **each** description*

Notification sent to parent*:*
DBMS used to retrieve parent's detail to enable mailing system to send out invoice;

Authentication / Validation*:*
Parents log in to DBMS with their credentials for <u>authentication</u> / <u>validation</u> of credentials by DBMS (the DBMS checks that data input are well-formed / appropriate);

Verification:
<u>Data verification</u> /The DBMS <u>verifies</u> that the parent's details are stored, as this could be a new parent who was provided details to log in, but the DBMS has not been updated. *NOTE accept any message on screen of the kind "contact the school manager", if verification fails*);

Restricted Access granted:
(Upon success of data verification) the DBMS provides access to reserved portion of data in database where parent can perform the operations (e.g. search the activity/period of time) / DBMS ensures access to the part of the Database storing child's activities for payment;

Preparation of payment:
Parent provides detail that enable submitting a payment / system requires data from parent (new and from DB) to proceed with payment / system triggers the processing of payment using different part of the system, the DBMS, and data provided by parents about bank account;

System update:
DBMS is used to update accounting record system **AND** to produce the receipt to parents. *Note: the second part of the sentence is needed to address the interaction) /* DBMS is used to update accounting record system **IF/WHEN** the parent performs a top-up of their balance;

Confirmation by email:
(Credentials of) open session on DBMS are passed to mailing system to send out email to parents with notification of payment/receipt;

Log-out /Close session/Time-out closure:
Parent closes session on DBMS with the effect of closing access to dedicated portion of database (in search and/or reading) / if parent does not log out, DBMS automatically closes access after time-out;

Example 2:

Note: *Example of answer worth **[1]** mark.*
It identifies the working of the payment system and it allows to detect some form of interaction.

- Parent authenticated by the DBMS
- Transaction is initiated
- Payment details added / entered
- If payment details and payment can be processed by DBMS then
    - transaction is completed by DBMS
    - Account is credited
    - Automatic e-mail notification of payment sent to parent
  Else
    - Transaction is rolled back
    - Notification of the failed transaction sent to the parent   **[5]**

(e)   *Award up to **[3 max]**.*
        Note: *One of the marks should address the fact that the workspace is partitioned, so if this is not addressed, **[2 max]**.*

Multiuser/concurrent access in reading/searching/querying of database;
But records in database are partitioned, to ensure privacy in access / each parent can access only the part of the database relative to his family data / Database has a row-locking mechanism, so that access is restricted to individuals only in permitted areas / there is no conflict/competition in accessing the same area of database;
(This/The partition/the row locking/the partial access) ensures **isolation** / **separation of data** (to present to transaction processing);
Therefore, multiple transactions (from different parents) may safely occur simultaneously on non-overlapping portions of the database during payment;
And this simplifies the management of system updates/recovery of (separated) transactions;

[3]

2.   (a)   *Award up to [1 max].*
         Collections of database tables that are linked via the fields / attributes / keys;
         A database structured in a way so that the data can be recognised / accessed /
         retrieved by using relations;   **[1]**

        (b)   *Award up to [4 max].*
         ***Do not credit** a description of what can be seen in an ERD, the command term is
         "explain".*

         The role of an ERD is to provide a compact view/graphical representation of
         entities/sets/actions/attributes/cardinality of entities/relations among entities;
         (So) it facilitates the planning/extension of a database because it provides a
         unique (and common) abstract representation of features independently from the
         actual implementation/nature of the database;
         (So) it informs the construction of the system at the logical level/represents the
         logical structure of the information system/DB;
         Enabling a structured and systematic development that can be better undertaken
         either by individuals or by teams of developers;
         And it informs over time how to model the data (in terms of its logical
         organisation) such as effect of system upgrades/ new formats needed to extend
         the old one);   **[4]**

        (c)   *Award up to [2 max].*
         *Award [1 max] for identifying an issue (I) and [1 max] for an expansion/example
         (consequence – C).*

         ***NOTE**: **There must be reference to the given scenario**. The wording also
         suggests also the "use" of the tables as they are (i.e. **with no
         deletions/additions/updates**).*

         *Award [1 max] for correct descriptions that lack reference to the given scenario,
         but still remain within the scope of "use".*

         Answers may include:

         I: Referential integrity is violated in TABLE_2 by the lowermost record, because
             there is no entry in TABLE_1 referring to Student_ID= 21;
         C: Therefore, any query built on a JOIN of the two tables, and referring to this
             value of Student_ID, cannot be answered;

         I: We want to use both tables to retrieve the email address in of the student
             whose ID is 21 in TABLE_2;
         C: But we cannot possibly do this, and the query cannot be answered, because
             the table will not be complete/will not be inhabited by the needed values,
             because referential integrity is broken on the value 12 of Student_ID in TABLE_2

         I: We want to use both tables to retrieve the email addresses of whomever is
             involved with the chess club;
         C: But we cannot possibly do this, and the query cannot be answered, because
             the table will not be complete/will not be inhabited by the needed values,
             because referential integrity is broken on the value 12 of Student_ID in
             TABLE_2
                                                                                                     [2]

(d)	(i)	*Award up to **[2 max]**.*
*Award **[1 max]** for identifying an issue that could occur after the deletion and **[1 max]** for an expansion,*
***Note:** Do not credit generic or non-specific responses or the use of referential integrity with no development (i.e. just being mentioned).*

Example 1:
We may not be able to find the email address of the Captain of the chess club;
Because the deletion of the whole record in TABLE_1 does not allow us to search in TABLE_2 with the value Student_ID=12;

Example 2:
We lose the value 12 of the primary key in TABLE_1 (losing entity integrity);
Therefore, a later successive insertion of a new record with primary key 12 in TABLE_1 may link with that value for the foreign key in TABLE_2, yet the meaning of the relation might be compromised;

Example 3:
We lose the value 12 of the primary key in TABLE_1 (losing entity integrity);
Therefore, the DBMS needs to have a method to decide what to do with the value 12 that appears in the foreign-key in TABLE_2;

[2]

(ii) *Award up to **[3 max]**.*

*Award **[2 max]** for evidence of **understanding referential integrity**.*
*Award **[1]** for an example, clearly referring to the <u>primary and foreign keys</u>
OR an expansion on how to limit the unwanted effects;*
***Do not credit** the term referential integrity being used with no development
(i.e. just being mentioned).*
***NOTE:** Be aware of possible repetitions with marks awarded in 2c(i).*

<u>Understanding:</u>
Data (in the relational database) should satisfy the property of referential
integrity;

Which requires that **each value of any field in some relation**/table exists
also as value in some field of another relation/table
OR
That requires that the values of the foreign key in one table, completely
relates to all values of the primary key of another table;

Examples:
Student_ID is a primary key in TABLE_1 and foreign key in TABLE_2 and
deleting a value in the primary key will have an effect on the foreign key;

The system must have procedure to restrict the effects of the violation, e.g.
by setting the affected values in the foreign key to a default value "null" (so
Student_ID in TABLE_2 will get value "null");
OR
by cascading (deleting) in all referencing foreign keys;
OR
by checking in advance whether a violation of referential integrity may
happen and not allow the deletion at the start;
OR
by recommending completion with null values and do normalisation, to reduce
incurring into violation of the property;   **[3]**

3. (a) (i) *Award up to **[2 max]***
***Note:** The question explicitly asks to <u>focus on the data structure as
displayed</u>*

Example 1:
The proposed data structure/sequence/list used to store SMS messages
has no determinate length/has unbounded length;
Therefore, we would have a table with an undefined number of columns
(and only the first two values are fixed in size for all subscribers);

Example 2:
The <u>values</u> in the two data structures show some repetitions /
redundancy (**Note:** do **not** accept duplication of data);
Not addressing this in an implementation of a sequence/list may eventually
generate tables that are bigger than necessary;

                                                                            [2]

(ii)   *Award up to **[4 max]**.*
The table must store each SMS separately as a row.
***Do not award any marks** if more than one table is constructed.*

***[1]** for Primary key or Primary Composite key, fully correct*
***[1]** presence of fields to store Sender's details (2 columns SenderPhNr and SenderName)*
***[1]** presence of fields to store Receiver's details and text messages (3 columns)*
***[1]** the table has three rows of records (in addition to the headers of the fields)*

Example 1:

SenderPhone Number	Sender Name	ReceiverPhone Number	Receiver Name	TextMessage
1111	fred	1234	Smith	Hi there
1111	fred	1234	Smith	Call me
1234	smith	1111	Fred	Later

*Award **[3 max]** if contents of table are correct <u>but</u> the composite key is incorrect or not identified.*

Example 2:
Table with extra unique identifier for messages, as primary key.

MsgCode	SenderPhone Number	Sender Name	ReceiverPhone Number	Receiver Name	TextMessage
A0001	1111	fred	1234	Smith	Hi there
A0002	1111	fred	1234	Smith	Call me
A0003	1234	smith	1111	Fred	Later

Example 3:
*Award **[2 max]** for an un-normalized representation of a table*

SenderPhone Number	Sender Name	ReceiverPhone Number	Receiver Name	TextMessage
1111	fred	1234	Smith	Hi there
		1234	Smith	Call me
1234	smith	1111	Fred	Later

[4]

(b)   *Award up to **[2 max]**.*
Data redundancy is the repetition of the same data field/attribute (or just data values that may populate two attributes with different names) within the database;
This may happen either across a number of tables in the database or within the same table (if there is just one);

[2]

(c) *Award up to [4 max].*
The tables must relate to the table created in 3a(ii) so follow through errors must not be penalized.

[1] for a table SUBSCRIBERS, correct, with two fields only;
[1] for indicating the primary key in SUBSCRIBERS;

*[1] for a table SMS that contains exactly **three** fields, **AND** only one element for the receiver (ether name or phone) and only one element for the sender (either name or phone) are indicated;*

OR
*[1] for a table SMS that contains exactly **four** fields, **AND** only one element for the receiver (ether name or phone) and only one element for the sender (either name or phone) is indicated;*

[1] to correctly indicate the primary key in the SMS table;

Example 1:
Note: *Accept all those variants that indicate in Subscribers the key as Name, and use it in the SMS, or that use in this latter table the name of the receiver. A variant is using TextMsg as primary key.)*

Possible answer deriving from Example 1 and Example 3 in 3a(ii)

SMS

SenderPhoneNumber	ReceiverPhoneNumber	TextMessage

SUBSCRIBERS

(Sender)PhoneNumber	Name

Example 2:
Note: *Accept all those variants that indicate in Subscribers the key as Name, and use it in the SMS, or that use in this latter table the name of the receiver. A variant is using TextMsg as primary key.*

Possible answer deriving From Example 2 in 3a(ii)

SMS

MsgCode	SenderPhoneNumber	ReceiverPhoneNumber	TextMessage

SUBSCRIBERS

(Sender)PhoneNumber	Name

Example 3:
Award [2 max] If three tables are constructed.
Award [1] if the three tables are correctly constructed.
Award [1] if the key is correctly identified in each table.

The SMS table may or may not include the MessageID field depending whether it was derived from Example 1 or Example 2 in 3a(ii).

SMS

SenderPhoneNumber	ReceiverPhoneNumber	TextMessage

SUBSCRIBERS

(Sender)PhoneNumber	Name

RECEIVERS

(Receiver)PhoneNumber	Name

[4]

(d) *Award up to **[3 max]**.*
*Award **[1]** for evidence of using **TWO**/**BOTH** tables (with a JOIN).*
*Award **[1]** for evidence that the filtering is triggered by the TextMessage.*
*Award **[1]** for evidence of correct referencing through receiver and phone number across the two chosen tables.*

__Note:__ SQL is not required, it is used here for shorthand notation to illustrate the structure of the types of response expected.
__Note:__ The question asks for the "steps" – evidence of computational thinking/procedural thinking is required.

__NOTE__: There can be variations in style of providing the answer, yet the guidance for the Markscheme is general enough to observe the steps to retrieve the Name of the Receiver given the message.

__Example 1:__
If two tables are used in 3(c):

SELECT ReceiverName
FROM RECEIVER JOIN SMS
WHERE
SMS.textMessage="thismessage"
AND
SMS.ReceiverPhoneNr=RECEIVER.PhoneNr

__Example 2:__
If three tables are used in 3(c):

SELECT Name
FROM RECEIVER JOIN SMS
WHERE
SMS.textMessage="thismessage"
AND
SMS.RecPhoneNr=SUBSCRIBERS.PhoneNr **[3]**

(e) *Award up to [2 max].*
Award [1] for identifying the <u>function</u> (underlined) and
Award [1] for an appropriate development / elaboration.

Ensure <u>backups are run regularly</u> / <u>m</u>aintaining archived data in backups /
<u>b</u>acking up and <u>restoring databases</u>;
To reduce the risk of losing data;

Ensure <u>consistency checks are run</u> on the database, (e.g. such as maintaining
indexes / managing and controlling data replication);
To ensure no data is inconsistent (across the related tables);

Ensure <u>access rights are set correctly</u> / setting <u>appropriate access rights at
various levels</u> administrator, user, etc;
To limit the risk of misuse across different users;

Check the need to <u>install and upgrade database software </u>/ ensure database is
up-to-date to handle <u>hardware/software changes / m</u>onitoring the database
performance;
To be sure that the data do not become vulnerable to external attacks because of
obsolescence;

[2]

(f) *Award up to [3 max]:*
Award [1] for the identification an ethical issue, and up to [2 max] for discussion/elaboration.

NOTE: The focus (conclusion) is usually on the trade-off between the rights of the individual to privacy/anonymity vs the needs for society to security / the greater good).

Example:
Personal information may be used by third party without the person knowing, therefore violating their privacy / anonymity;
However, this violation may be justified by security reasons, for example national security / personal safety / monitoring of threats or hate speech;
For example, from SMS one may extract information about locations/habits (gambling)/sensitive information (banks communications);

This enables third parties to:
- build up a profile of the individual through data aggregation (e.g the individual is a game-addict, but they didn't want to make this public);

- use for fraudulent actions over time (monitoring how often the bank transactions happen through a credit card, and generate hoax/impersonations/ further sell these data);

- use this data for marketing purposes (e.g. spamming messages);

- identify the location where an individual is, at particular times / surveillance / identify where and when an individual is not at home;

[3]

4. (a) *Award up to [2 max].*

Award [1 max] if only features from List 2 are mentioned;
Award [2 max] if at least 4 features altogether, but including some from List 1, are mentioned;

List 1:
Central storage/repository;
integrated data / data from different sources;
collected over time;
time-dependent (storage/data) / time-stamped (data);
presence of both historic data and current data;
non-volatile/permanent data;
static storage;
non-frequent access in retrieval;
no deletion of data;

List 2:
large quantities of data;
stored by large organisation/companies;
data subject to mining/analysis to explore trends;
data aggregation;
uses in business analytics/decision making;

[2]

(b)   *Award up to **[4 max]**.*

   Note: *the response should highlight the steps that occur within the ELT process.*
   Note: *mentioning Extract, Load, Transform (ELT) alone is not sufficient and gets **[0]***

   *Award **[1]** for **each of the three phases** described.*
   *Award **[1]** for any additional detail provided in any of the three phases.*

   Extraction:
   The relevant data from each school/at the source should be identified and
   prepared in view of the transfer to the data warehouse using the principles of the
   ETL);
   The relevant data from the school database is chosen/selected, by using the
   methods, schemata, relations proper of the database that the school uses
   (source);
   And then,
   They are stored temporarily (locally);

   Transformation:
   The data are transformed in a format that is suitable for the use required (by the
   target applications) / the data are transformed following the schema that the
   target application requires for processing;

   For example:
   Data records and field structures cleaned and checked for errors/missing data
   OR
   The data type is changed (integer into double, double into currency);
   OR
   Raw data are given (without normalisation), as the target database may
   preferably need to have un-normalised data to run analytics, and can still better
   normalise all data from all schools together once available (global view);

   Loading:
   Data from each school transmitted/transferred to the data warehouse to be
   loaded by the target databases (that will operate on data using target
   applications);
   Moreover, *t*he transfer may happen at scheduled times not to hinder other
   operations;

   [4]

(c)   *Award up to **[4 max]**.*
   Note: *Responses in the form of a SQL statement or similar **are not acceptable**.*

   Deviation detection is used;
   Access the marks for Mathematics from the data warehouse;
   For a given age group (and for a particular period / given a range for timestamp);
   Calculate the average for each school in Mathematics for that age group;
   Identify the outliers/low scores of schools after running the deviation detection
   (multiple if-then-else statements) algorithm;
   List the names of those schools;

   [4]

(d)   Data mining is the process of knowledge discovery in large data sets;
      Obtained by analysing data in a number of way to the purpose of detecting
      patterns/relationships;

      OR

      Data mining is the analysis of existing large data sets;
      Aiming in <u>discovering</u> (new) relationship/patterns among data (and hence
      knowledge);

                                                                                    [2]

(e)   *Award up to* **[4 max]**.
      NOTE: *Award* **[2 max]** *if the response does not make an explicit reference to the
      scenario.*

      Award **[1]** *for demonstrating an understanding of cluster analysis.*
      Award **[1]** *for demonstrating an understanding of association analysis*

      NOTE: *This question requires a comparison that relates to the scenario,
      "understanding the distribution of scores in grade 6 reading tests over the past 10
      years".*

      Award an additional **[2 max]** *for an in-depth comparison of the use of both
      techniques that explicitly relate to the scenario.*

      Features of cluster analysis:
      Cluster analysis groups schools or pupils by score / according to different factors,
      such as region/location;
      Therefore, it enables comparison between groups;

      Features of association analysis:
      Association analysis breaks up data sets by variables / by variables, such as
      gender, location, age;
      Therefore, it enables the investigation of patterns;

      Example of a comparison related to the scenario:
      Since the age group is fixed (grade 6), cluster analysis will allow us to see the
      geographic distribution of scores, and to analyse trends over time, suggesting (for
      example) whether an improvement in some geographic areas/locations is
      observable;
      whereas
      Association analysis may be used (even after cluster analysis) to detect some
      patterns, for example whether, independently from the geographic region,
      females perform better than males in reading / whether wearing corrective
      glasses (from health insurance records) may affect the proficiency in reading
      (accept similar) / whether being a second-language speaker may affect the
      proficiency;

                                                                                    [4]

(f)　　*Award up to **[4 max]**.*
__Note:__ The question requires the negative effects of cluster analysis and association analysis to be addressed.

*Award **[1]** for identifying a negative social impact*
Negative social impacts may include:
Type casting students / low morale / students or teachers losing faith in the assessment process;
__Note:__ Accept other negative social impacts.

*Award up to **[3 max]** for an elaboration that addresses **both** cluster analysis **and** association analysis*
*Award up to **[2 max]** for an elaboration that addresses **either** cluster analysis **or** association analysis*

__Negative effect of cluster analysis:__
Cluster analysis partitions the data into groups that exhibit similar characteristics / allows schools to be grouped based on performance;
Which may indicate one region has performed poorly over time;
If such information is used inappropriately it could have deleterious effects on students/schools;
It may impact on the ability of teachers and headteachers to maintain the school appealing to future students and this impoverishes the neighbourhood;

__Negative effect of association analysis:__
Relationships found with association analysis may not necessarily causal / based on correlation rather than causation;
Therefore, inferences made may not be correct / could be harmful if results are released/
may not provide decision makers with appropriate data;
And this over time may mean that decisions taken can implement procedures that do not allow alignment to desired targets;
May polarise/orient some choices/perceptions across students (e.g. the perceptions that Engineering is not for girls, or nursing is mostly for girls);

[4]

Option B — Modelling and simulation

5. (a) 0.1 **[1]**

(b) *Award **[1]** for using the two given probabilities and **[1]** for a correct expression.*
$1 - (\text{PROB}[i, j] + \text{PROB}[j, I])$ or $1 - \text{PROB}[i, j] - \text{PROB}[j, i]$ **[2]**

(c) 30 **[1]**

(d)

	A	B	C	D
1	30 – (14 + 10)	5	0	5
2	2			
3	9			
4	3			

Add a team's wins and losses;
Subtract from 30 and insert in correct diagonal;
   (*eg* A1 = 30 - (sum(A2;A3;A4) + sum(B1;C1;D1)))
    or even A1= 30 - A2 - A3 - A4 - B1 - C1 - D1
    … both examples cover the first two marking points)
Repeat for other teams;

Example 1 (based on candidates work):

M1 = 30-(2+5)-(9+0)-(3+5) = 6
M2 = 30-(2+5)-(1+8)-(5+5) = 4
M3 = 30-(9+0)-(1+8)-(4+4) = 4
M4 = 30-(3+5)-(5+5)-(4+4) = 4

Example 2 (based on candidates work):

[3]

(e) *Award [1] for each condition which increments the correct cell
(accept the use of the macro in part (d)).*

For example:
```
if m>n then
   PLAYED[i, j] = PLAYED[i, j] + 1
else if n>m
   PLAYED[j, i] = PLAYED[j, i] + 1
else {
      PLAYED [i, i] =  PLAYED [i, i]+1
      PLAYED [j,  j] =  PLAYED [j, j]+1
     }
end if
```

Accept if … if … if.

*If the student has used the wrong parameters in a consistent manner, award [2]
max on FT.
If the student has not used the correct increment, award [1] max for a correct
nested if...else structure or if … if … if.*   **[4]**

(f) A mathematical model reduces a situation to mathematical expressions and
rules;
In the example, the results of matches and points gained is expressed
mathematically / table PROB;
A simulation applies the model in order to investigate or measure the result;
In this case the probabilities were applied to give expected season results / table
PROB10;   **[4]**

(g) If the results are very different;
Something has changed from last year;
Can investigate any change in players/trainer;
Can check the calculations are correct;
Can be used to identify improvements and unexpected failures from specific
teams;
Implies changing the probabilities for the following year;
That takes this year's results;
And ignores the previous year;
Or give up using past form as it is no indication for the following year;   **[4]**

6.   (a)   *V < D* or *V<= D* or *V - D < 0*   **[1]**

(b)   *Award* **[1]** *for each appropriate factor identified and [1] for a description up to*
[2 max] for each of the **three** *variables.*

Mark as 2 and 2 and 2

For Input (*R*)
Measure river / rain fluctuation;
Over a period of time to establish average values / the maximum and minimum
expected values / outliers;

For Output (T)
Measure the amount used by the town;
at different times (day/night, weekday/weekend, workday/holiday, summer/winter,
etc);
to establish average values / the maximum and minimum expected values /
outliers;
(Do not award this marking point if already awarded for R.)

For overflows/outflows (E)
Use the data gathered from the analysis of Input (*R*), Output (O);
To plan for sufficient overflow capacity;

NOTE: for T and E do not accept the size of the pipes or the outlets since the
purpose of the simulation is to establish those. Also, the stem indicates that the
pipe will be large enough to fulfil the water needs of the town.
For R do not accept size of 'inlets' - there are no inlets.

NOTE: *establishing average values was originally not intended, but most*
students provide this and it is a reasonable factor to consider … however, do not
award this marking point for both R and T. The average does not apply to E at all
- since outlets are man-made and
controlled.

[6]

(c)   *Award marks as follows, up to **[6 max]**.*
*Award **[3]** for statements concerning the creation of the graphs;*
*Award **[3]** for statements concerning the interpretation of the graphs;*

For example*:*
Setting up a suitable relationship (like $V_{current} = V_{initial} + R - T - E$);
Graphing the values of V for fixed values of T and E, as R goes from minimum to maximum levels;
Including a horizontal line for danger level (D);

This should identify combinations of R, T and E that are problematic;
By showing situations where the level of V exceeds D;
Indicating the maximum required value of E;

***Note**: Many alternative systems are possible, for example more complex ones that graph V over time for fixed / changing values of R and T, leading to required values of overflow (E). These should be credited if described / discussed well. A much simpler (alternative) system which checks that R = T + E and works on minimum and maximum values of R should be credited in the same way for SL.*

[6]

(d)   *Award marks that clearly indicate that the physical model will be designed to meet the needs identified in the previous simulation of water flow.*

For example:
Knowing the maximum volumes of E that are needed;
Will help with the planning of the overflow outlets;
Overflow outlets can be placed at the appropriate water level to function in time;
The simulation could include future developments (like growth of the town);
In order to ensure that the reservoir design will not be outdated in the foreseeable future;

[4]

7.   (a)   *Award **[1]** for a suitable application and **[1]** for an expansion, up to **[2 max]**.*

For example, a map showing the journey from one point to another;
When only the route needs to be shown on a phone or otherwise;
As 3D version could take up too much space;
If no internet connection would be too slow;

Note: *Accept situations that are 2D in nature, like graphing a v-t diagram.*   **[2]**

(b)   A series of sketches could show the design from a limited number of angles;
Whereas the CAD system would allow an initial wire frame to be constructed which
could be rotated;
Changes could be (easily) made and the effect of the 3D shape from other angles seen immediately;
Easy to undo a change to the model, rather than erase part of the sketch;
Easy to scale the design;
CAD files are less likely to get damaged than physical drawings;

Note: *Do not accept vague statements like "CAD is more precise".*   **[3]**

(c)   Rendering would take place;
To apply light and texture to the model;
Using such techniques as ray tracing;
To the (wireframe) 2D representation;
To produce a realistic view of the final product in the chosen material;   **[4]**

8.   (a)   *Award up to **[2 max]** each, for two advantages.*
The quantity of material is too large;
It would be impossible for any effective human analysis of all the words;

Analysis is too complex;
As there is likely too many groups/multiple group memberships;   **[4]**

(b)   *Award up to **[2 max]**.*
Mistyping/misspelling;
Different use of words from different regions/countries;
Different styles, for instance modern online SMS text language;
Some of the posts may be written in different languages/need translating;
Acronyms *eg* LOL.
***Note**: Accept other reasonable answers.*   **[2]**

(c)   *Award up to **[2 max]**.*
Manually analyse a small section/sample of the opinions;
Decide or define the required outcomes *eg* frequency of specific technological words;
Construct the fitness function from this data;   **[2]**

(d)   *Award up to **[6 max]**.*
Diagram with input layer, minimum two hidden layers and an output layer;
Similar number of nodes with weights;
Application of nodes in hidden layers;
Mention of repeated application;
Until pattern fits fitness function within acceptable limits;
Back propagation likely and shown;
Genetic algorithms will take more time with more layers;   **[6]**

(e)   *Award **[2]** for a clear understanding of supervised learning, with an example of when it is appropriate.*
*Award **[2]** for a clear understanding of unsupervised learning, with an example of when it is appropriate.*
*Award **[2]** for suggesting when in this situation unsupervised would be better.*
***Note**: Do not accept simplistic answers such as "because the answer is not known".*

For example:
Supervised learning implies that there is a desired result to be found;
Such as categorizing the data into sets of identified technological advances;
Unsupervised learning is used in a situation where no desired output is known;
Such as the discovery of connections and patterns;
In this case it could be that opinions on one technological advance were related to multiple specific factors/strengths of opinion that were not foreseen;
With such a large data set it is unlikely that an appropriate fitness function could have been found;   **[6]**

Option C — Web science

9.   (a)   *Award up to [4 max].*
For each of the two reasons:
award [1 max] for identifying a reason (R) and [1 max] for an explanation (E).

R: The banner will be the same in all pages of the site / on top of all pages;
E: Making the site easily recognizable among many other merchants / easily
   identifiable / making the logo of the merchant visible;

R: The banner usually contains a hotspot button to go back to the homepage /
   (*Note: accept the logo itself as hotspot button)*;
E: So that navigation becomes easy;
   This is the place where the hotspot button is expected to be found /
   This is consistent use (normal practice) with other sites /
   This is the place which can be viewed first on all digital devices (Laptop,
   Mobile, Tablet etc.) therefore useful for scrolling;   **[4]**

(b)   *Award up to [1 max].*
Copyright notice/privacy statement/link to terms & conditions/link to commercial
laws;   **[1]**

(c)   (i)   *Award up to [1 max].*
           HTML / *XHTML / XMTML*;   **[1]**

(ii)   *Award up to [3 max].*
       Award [3] if the three actions are correct and their order is correct.
       Award [2] if three actions are correctly identified but their order is incorrect.
       Award [1] if two actions are correctly identified, no matter the order.
       Award [0] in all other cases

       User **selects** one among Spikes/Trail/Endurance (*accept:* "one of the
       three/available options" or words to that effect);
       Then **inputs** the size;
       Then **clicks on a Submit** button (*accepts* submits) to send out the form;
                                                                              [3]

(d)   (i)   *Award up to **[5 max]**.*
*Award **[1]** for each of the processing steps indicated below.*
Required elements are underlined.

- Server uses process.php to process information input via the form;

- Server uses <u>process.php</u> to process <u>item</u> and <u>size</u> input via the submitted form;

- Size is <u>checked</u> for being a <u>numeric value</u> **AND** in the <u>range [34,46] included</u>;
  OR
  Size is tested for <u>type-checking</u> (well-formedness) **AND** <u>range-check</u> (legitimate values), i.e. number between 34 and 46, extremes included;

- Check fails: (appropriate) error messages are <u>output</u> **/** <u>echoed</u> / <u>displayed</u> back to user;

- Check ok: server performs a <u>protected access</u> to the internal database <u>CatalogDB</u> in <u>localhost</u>;
  OR
- Check ok: server performs an access, protected by <u>username and pwd</u>, to the internal database <u>CatalogDB</u> in <u>localhost</u>;

- Server <u>performs a query</u> on CatalogDB with the <u>two parameters instantiated</u> from the input form;
  OR
  Server <u>performs a query</u> on CatalogDB with the (actual) values of the parameters <u>item</u> and <u>size</u> from the input form;

- <u>Nothing is found</u> on CatalogDB matching the query: return/output/echo appropriate error message
  (**Note**: *do not accept 'the query fails'; the point is that the query succeeds but the selection is empty, accept words to that effect*).

- Result of query is non-empty (**Note**: *accept words to that effect*): return/echo/display ALSO <u>image</u> and <u>price</u> from CatalogDB (in addition to item and size)

- Server <u>closes connection</u> to database;   **[5]**

(ii)   *Award up to **[2 max]**.*
        *Award **[1]** for an example and **[1]** for an explanation (to address clients'*
        side scripting)

   ***Example*:**
   add a script/Javascript (using `<script> .. </script>`) to the form to
   magnify text / to change background colour / to use specific fonts/
   to provide animation;

   Explanation:
   Features depending on user's taste/needs are processed faster on clients'
   side and would heavily clog transmission/generate bottleneck if they were
   depending on the server;

   Note: *Accept alternative examples.*
   Note: *Candidates are not expected to write JavaScript code.*   **[2]**

10.   (a)   *Award up to [4 max].*
Award up to [2] in relation to posting in blogs and forums
Award up to [2] in relation to commenting in blogs and forum.

***Note:** award **max 2 marks** for responses that lists evidence of understanding generic posting / commenting but with no explicit connection between the blog and the forum.*

Posting:
In a blog only the owner can post an article / open a thread of discussion / start a theme;
<u>Whereas</u> in a forum all registered participants can post an article / open a thread;

Commenting:
In a blog registered visitors may be allowed to comment but the blog owner may moderate
<u>Whereas</u> in a forum all registered users are allowed to comment (without moderation);

Example 1:
A blog is associated to / led by one person (the blog owner/the journalist) who drives / leads / coordinates the discussion;

Only the owner can post a new thread and the registered users may comment just to animate the debate, even they have different views/opinions;

However, registered users may be subject to moderation by the blog owner, especially if this is the journalist publishing in a major newspaper, to the purpose of filtering out comments that are inappropriate / offensive / out of scope / non-suitable for the newspaper policies / ensuring quality of content / relevance of posting / quality of discussions;

This is quite different from what happens in a discussion forum, where registered users are not subject to moderations and therefore can both open/post threads and comments in all threads;

With the effect of delivering a sense of participation / enforce a spirit of community / aggregate people on general ideas;

[4]

(b)   *Award up to [5 max].*

Tags are primarily important <u>to help users navigate</u> the blog site;
By <u>relating posts</u> within the same category/tag/theme <u>in the blog</u>;
Hence tags / they contribute <u>links from web pages inside</u> the blog;
Tags are <u>visible</u>, so they can be easily <u>accessed by web crawlers</u>;
That will be <u>guided to discover the site</u> content following the links;
To <u>avoid the risk of isolated pages</u> in the blog that would not be indexed.   **[5]**

(c)   *Award up to [2 max]*.

It has not been accessed enough, yet (possibly because there aren't many registered users, yet);

It contains keywords that are out-of-fashion/in disuse that nobody currently uses for their searches;

The whole site might be hosted by a company without a good reputation (and this penalizes ranking);

The site can only be accessed by browsing and does not have a feeder for mobile devices / automatic updates (*eg* RSS, Twitter, *etc*);   **[2]**

11.   (a)   (i)   *Award up to [1 max]*.

It is the idea of performing computing anytime and anywhere (pervasive) using any device / (using in everyday objects);

Note: *Accept* "Computing embedded into everyday objects for healthcare purposes"

[1]

(ii) *Award up to **[4 max]**.*
*For each of the **two** examples:*
*Award **[1]** for identifying an object in which the technology is embedded.*
*Award **[1]** for the use to which this is put.*

Answers may include:

Obj - (Digital) control/transmitters that are embedded in domestic appliances such as fridges / cookers / microwave or doors *etc* /smart appliances;
Use - So that the different objects can communicate among them to avoid a surcharge in power demand/avoid a power blackout;

Obj - (semiautonomous) wheelchair with sensors/camera;
Use - so to avoid specific obstacles when used in a closed environment / 'knows' the path to traverse a certain area of the flat/hall/living environment

Obj - smart white stick (for sight-impaired people);
Use - the tip of the stick has awareness of the environment and sends back, to the handle, impulses that code information to the user, for example it may signal a wet floor

Obj - A (smart) biometric watch;
Use - That routinely reports the vital parameters / elaborates statistics / checks the pulse;

Obj - voice recognition systems integrated on appliances/doors;
Use - for example a user with limited mobility may just say 'open the door' to the main door, but the recognition must be calibrated so to recognise only the owner

Obj - An app on a smartphone/computer/interactive screen;
Use - that prompts the user to regularly make some movements for rehabilitative physiotherapy, at home, on their own (soft gym/physiotherapy)

Obj - sensors in the floor/bathroom;
Use - detect whether the person has fallen and this can be integrated with other services to alert the nurse;

Obj - infrared camera
Use - to detect movements/presence/whether the person is sleeping or not/ body temperature therefore signalling parameters for possible illnesses;

Obj - weight sensor embedded in the bed
Use - to determine changes in weight that can be transmitted to other services;

[4]

(b)   *Award up to **[5 max]**.*
*Award up to **[2 max]** in relation to the public cloud.*
*Award up to **[2 max]** in relation to the private cloud.*
*Award **[1 max]** for a conclusion.*

Note*: Do not accept "Access for doctors/nurses..." as it is included in the stem of the question.*

Example one:
<u>Public</u> cloud could be accessed by many people;
Although access is restricted to their authorized area;
<u>Private</u> cloud provides dedicated infrastructure for one client;
Which makes it more secure;
Because resources are not shared;
(A breach of security in a private cloud means less data loss than if it happened in a public cloud);
(<u>*Conclusion*</u>) As this data is sensitive, it should therefore be stored in a private cloud (and encrypted);

Example two:
<u>Public </u>clouds are usually maintained by expert people;
SW updates to prevent security risks are very sophisticated;
<u>Private</u> clouds are usually maintained by people with less expertise;
When updates are needed in a private cloud it is harder to ensure the same level of security (as in a public cloud) across the private clouds, making it more vulnerable;
(<u>*Conclusion*</u>) This makes the private cloud not necessarily more secure,
Therefore, encryption is needed to prevent eavesdropping;   **[5]**

(c)   *Award up to [2 max].*
Award [1] for identifying a risk and [1 max] for an expansion/description.

Note: *Do not accept "(Quality of) security of data storage"*

Risk: Malfunctioning / reliability of the device (reliability of Hardware);
        The device fails to communicate (to send the signal) and the supervisor
        might not intervene on time / user might be induced to think the situation is
        OK;

Risk: Wearables could be lost / misplaced;
        Consequently, there is an interruption in monitoring some vital parameters;

Risk: Compatibility among devices;
        The user might be provided cheaper equipment that is not conform to
        standards, raising problems of interoperability with existing equipment;

Risk: Network may be down / stuttering / low bandwidth;
        The user might become isolated and unable to set it back up on their own
        (don't know how to make a WiFi work);

Risk: Privacy / protection and personal safety;
        The user's habits at home could be determined by monitoring the vital
        parameters (sleep/awake or pulse) and this information may be used for
        unlawful / unethical
        purposes;   **[2]**

(d)   CDMA / 3G / 4G / UMTS / GAN / GSM / iDEN / WiMAX / Wi-Fi.   **[1]**

(e)   *Award up to [5 max].*
Award up to [2 max] for describing the role of protocol and standards and
up to [3 max] for an explanation (examples may be included).

Protocols and standards provide tested and verified techniques and technologies;
To support data protection/security/privacy;
Protocols and standards are essential in a complex system that includes different
subsystems/components that interact;

Possible lines to develop an explanation:

* For example, in **transmission** (SSL/TLS):
  So that all information is encrypted/is protected for eavesdropping or
  man-in-the-middle attack/has Integrity and authenticity protection;
  For example, for the transmission of data/signals from wearables to the cloud;

* For example, in **storage and access**:
  In relation to Availability/Integrity/authenticity/Access control/logging;
  For example, to prevent risks for patients due to inappropriate data monitoring;

* For example, in relation to **evolution of the system**:
  In relation to the expansion of the system or changed architecture of
  the system;
  Which might include further/future devices/services involving the cloud /
  so that new devices are (backward) compatible with the existing ones;   **[5]**

12. (a) *Award up to [3 max].*
The essential features that distinguish the semantic web from the web are underlined here.

The semantic web is an <u>extension</u> of the web;
(Whereby) information is structured using a <u>common format on resources</u> (based on URI)/ by <u>adhering to a standard</u> set of rules/languages;

Semantic web permits <u>software to be used on the web</u> for query / analysis / extraction / it allows to interpret data to get information, not just be browsed;

So that information becomes <u>automatically understandable</u> by computer software (and therefore supports automatic knowledge discovery / machine learning / big data applications)

Languages for the semantic web use <u>annotations</u>/relations on top of <u>XML</u> described resources/data of the web;

Some possible responses include:
The semantic web provides interchange of data/resources,
<u>whereas</u> the web supports interchange of documents;

On the semantic web, one can perform queries/analyses/information extraction (*eg* SPARQL/OWL based tools/GRDDL),
<u>whereas</u> the web can just be browsed;

Information is structured to a common format (based on URI) and specific languages (logic-based or rule-based) are used,
<u>whereas</u> the data on the web are retrieved essentially by URL and by other standards for the language.

The semantic web makes the resources available on the web to be ***automatically*** understood by software (***Note***: idea of accessibility/understandable by machine/ enabling big data application/knowledge discovery), providing meaning to those resources;
<u>whereas</u> the web just retrieves data.

Web pages are typically written in HTML,
<u>whereas</u> for semantic purposes they should be at least with XML;
So to enable further annotations (with RDF/OWL) that have the ability of defining and describing the relations among any two data (resources) on the web; **[3]**

(b) *Award up to **[2 max]**.*

Elements of focus to award marks:
*Award **[1]** for showing evidence of understanding that edges/links are directed;*
*Award **[1]** for showing evidence of understanding that (sub)-graph is a mathematical object, and web-graph is the interpretation in terms of web terminology of a (bigger) mathematical object.*
Note: *an example is required*

Example 1:
Web-graph: interpretation in terms of URL/webpages and links of the elements (node and edges) of a <u>directed</u> graph
OR
a <u>directed</u> graph (set of nodes, set of <u>directed</u> edges), whose nodes are interpreted as resources/URL/web site and whose <u>directed</u> edges are interpreted as links between resources/URL/sites;

Sub-graph (of any graph): subset of nodes and a subset of edges of a given graph, preserving direction of edges;
Therefore, a web-graph may be the entire web, whereas a sub-graph of the directed graph can be the mathematical structure associated to a specific web site, for example the entire IB portal (all its nodes and its the links)

Example 2:
For example: Wikipedia
All URLs are nodes of a graph, and all links among URLs are directed edges among the nodes: this determines a directed graph which is only part of the large structure that collects all the WWW, which is the whole web-graph. Therefore, being part of it, it is a sub-graph of it.

Note: *Accept drawings that shows, with a few nodes linked and some terminology, evidence of such understanding.*

[2]

(c)   *Award up to [2 max].*
Award [1] for evidence that it retrieves multimedia
Award [1] for evidence that the input for search can be multimedia

The multimedia-web is the (Evolution of the text-based) web specialised/oriented
to **retrieve multimedia files** (audio / video / photo...);
And such that the **search** can be performed **with input from different media**
and of different nature (audio / video / photo) (as opposed to text based search
only);   **[2]**

(d)   *Award up to [4 max].*
*Note: do not accept search of music that starts by giving text (the title of a song,
for example) as input!*

The search engine accepts soundwaves (or MiDi);
Search **can be driven by** melody / rhythm / beat / pitch / presence of specific
scales or harmonics / sounds associated to specific instruments;
Therefore, soundwaves are converted into formats so to extract data
representing some characteristics that guide the search;
To compare the audio / physical characteristics extracted with the samples
soundwave in libraries / databases /specialist music (sound) archives;
And To returns the most likely matches, for example in the form of either single
music excerpt or aggregation of similar sounds;
                                                                                                    [4]

(e)   *Award up to [5 max].*
Award [1] for the <u>popularity</u> of the movie.
Award [1] for the <u>cache in P2P</u>.
Award [1] for the possibility of <u>parallelising download</u>.
Award [1] for why <u>Zipf's law</u> matters/what it suggests.
Award [1] for the <u>search in servers</u> as opposed to P2P.
Award [1] for a well constructed argument.

Note*: Award [1 max] for responses that describe only what a power law is, without any relation to the context of the question.*

Example 1:
If a movie is a "rare" / "niche" one, then it is unlikely that it is available by many nodes in the P2P network;
Hence, it requires more time to be <u>searched for;</u>
Because it has to searched elsewhere, <u>in repository servers;</u>
Besides, once it is found, its downloading may involve just one server;
Therefore, not only the search itself is slow, rather the downloading is from a single node, for an entire movie, the server may easily become congested, slowing down the overall process;
However, popular movies are more likely available within the P2P network;
And since P2P networks store entire movies/fragments in the cache;
The downloading itself can be more parallelized by downloading the smaller fragments and recombining the fragments later on;
Henceforth, overall downloading time of the movie decreases in a P2P network;

Example 2:
A power law applies (with some approximation) to the popularity of movies for downloading;
If movies are ranked by popularity, with 1 being the most popular, then the movie with rank r has a number of downloads expected to be proportional to 1/r, for Zipf's law;
OR
The most popular movie is downloaded at least twice as the second most popular movie, at least three times the third most popular movie, and so on, for Zipf's law;
OR
In a large sample of movies, the frequency of movie in a P2P network (the likelihood to find it in the P2P net) is inversely proportional to the rank of the movie itself (in the frequency table).

P2P networks have caches, where multiple copies of all or part of downloads are kept;
Therefore, the more often a file is downloaded, the more likely it is that at any one time there will be multiple copies cached on different servers/members of the P2P network;
Downloads will then be faster, because a closer server can be chosen, or different parts can be downloaded in parallel;
The 100th most popular movie download, for example, will only be downloaded about 1/100th as often as the most popular one, if Zipf's law applies;
So it is much less likely to be cached, and will take longer to download;

[5]

(f)   Award marks as follows up to **[4 max]**.
        *Award **[1]** for evidence of understanding the concept of collective intelligence,*
        *Award **[1]** for stating a relevant example, in the context of online buying habits*
        *Award **[2 max]** for a further expansion/explanation.*

        Note: *Do not award marks for "Collective intelligence affects the online buying habits of the general public": this is the repetition of the question itself!*

        Note: *Given the question, the response has to refer to the context of online buying habits. Do not accept examples that do not address the context.*

        <u>Collective intelligence</u> **[1 max]**
        Collective intelligence is the intelligence that emerges from a group that shares knowledge, and emerges from the collaboration, competition and coordination of (many of) the individuals of the group;
        OR
        Collective intelligence includes intelligence of a group / shared knowledge / collective coordination;

        (Therefore,) the group acquires knowledge at a much faster rate than an individual, and this allows the group to make better decision, faster, for the interest of the group;

        <u>Some ideas for contexts for the example (list not exhaustive)</u>
        - buying second-hand clothes /
        - freecycles communities /
        - vegetarian products / specific diets
        - culture / publishing
        - crowdsourcing / crowdfunding services;

        <u>Stating a Possible Example</u> **[1 max]**
        A group of people has an interest in minimizing food waste and organizes a network online to re-distribute their goods that are approaching the use-by date;

        People contributing the reviews to online vendors/marketplaces and possibly on web-sites that monitor people's browsing activity/preferences;

        People regularly selling pre-owned goods on Ebay (online) / members of an online community of artists using a specific technique;

        Serious-gamers that use their experience in distributed game strategies to act collectively and influence the price/bids for some product that is available online, for example, tickets for concerts/sport events, for re-sale;

        Popular subscriptions to services like Netflix or to torrents for downloading, in comparison to purchasing DVD or going to cinemas/theatres/concerts, with an impact on the way artists may sell/market their products and protect their intellectual property;

        Crowdfunding the publication of a study on some not-well-known ceramic artist, with the effect of increasing the visibility of both the artist and the author of the essay as expert in the area, affecting the reputation of both;

Possible lines to develop an expansion: [2 max]

- **Technical expansion**
  Can include/address use of preferred links/folksonomies for an interest group;
  Can include use of P2P or dedicated networks;
  Can include crowdfunding and crowdsourcing;

- **Sociological / economical / environmental expansion**
  Groups sharing the same ethical principles can merge into a bigger network;
  This may have (sometimes) an impact on small communities by using only the products that are produced locally/ by recycling locally, therefore reducing carbon footprint;
  These small groups may create a burgeoning niche market for their products, and can lead to increased economies of scale and price reductions that may stimulate the interest in other buyers (outside the immediate group);

- **Cognitive expansion**
  Likes and dislikes for some products/use of reviews;
  Trends in the market can be predicted by observing the habits on the net;
  Possibly knowing these habits (sentiment analysis) may lead to changes in some policies (for example, pressure groups to obtain some life-saving vaccinations for free rather than buying them privately);

[4]

Option D — Object-oriented programming

13. (a) *Award up to [3] marks.*
Note: An example is required with detailed explanation (up to [2 max];
Award [1 max] if there is lack of detail).

A class describes/provides the template for a particular object;
Whereas an instantiation of a class actually creates an object for that class;
For example:
An employee class would contain characteristics/actions that describe
an employee, and an instantiation would represent a particular employee;   **[3]**

(b) *Award marks as follows up to [4 max]:*
*Award up to [2 max] for <u>each</u> of the **two** required examples (award [1] mark for correct
example but lack of terminology/detail/competence)*
*Note: Examples should show competent use of terminology in describing the use of
inheritance, to get full marks.*

The responses should focus on these aspects, with keywords underlined:
- Several <u>sub-classes</u> can <u>inherit</u> characteristics from a <u>superclass</u>;
- Whilst <u>adding specific characteristics</u> particular to themselves;

Examples can then expand or be described to include the features above.

For example:
There could be a "GUI" superclass that contains the main elements for the UIs
with the GUI for each screen inheriting from its parent;

For example:
An "Employee" superclass that allows the different types of employees to <u>inherit</u>
<u>standard</u> characteristics;   **[4]**

(c) *Award up to [3 max]:*

Library programs are programs that have already been written;
They are standard/frequently used programs;
Development time will be saved if they can be incorporated into the design;
They have been tested / Will be free of bugs;

*Note: Allow [1 mark] for responses that include examples that expand on any of the aspects
listed above.*   **[3]**

14. (a) *Award [max 2 marks]:*
[1 mark] for each of the two correct assignments.

```
public SalesPerson (String id)
{
  this.id = id;
  salesHistory = new Sales [100];
}
```

Note: Allow any reasonable value for the number of Sales objects.   **[2]**

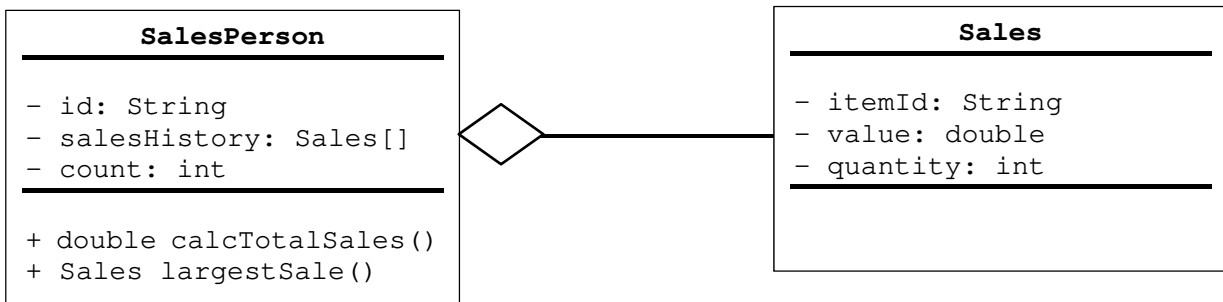(b)   *Award up to **[3 max]**.*
        ***Note**: Use of competent/detailed terminology (underlined) is required.*

        As the variables are declared as private;
        They/their values cannot be <u>accessed</u> directly from <u>outside</u> this class;
        Therefore, <u>public</u> accessor methods are needed to permit access;   **[3]**

        ***Note**: Either accessor **or** access required.*

(c)   (i)   *Award up to **[4 max]** as follows (there are 5 marking points):*
              ***[1]** for two classes BOTH with basic three-tier structure layout;*
              ***[1]** for correct dependency shown (eg diamond arrow) /*
              "has a" from `SalesPerson` to `Sales`/
              1:N relation.
              ***[1]** in `SalesPerson`, for all vars and methods (contentwise) with type*
              (allow if constructor/ accessor/mutator methods are present);
              ***[1]** in `Sales`, for all vars and methods (contentwise) with type*
              (allow if constructor/ accessor/mutator methods are present);
              ***[1]** for correct use of +/- for public/private;*

```
        SalesPerson                                    Sales
  ─────────────────────                         ─────────────────────
  − id: String                                  − itemId: String
  − salesHistory: Sales[]    ◇──────────        − value: double
  − count: int                                  − quantity: int
  ─────────────────────
  + double calcTotalSales()
  + Sales largestSale()
```
                                                                                          [4]

        (ii)   *Award up to **[2 marks]**.*

              Any change in the `Sales` class might require a rewriting of some
              `SalesPerson` methods/might not work;
              For example, the `calcTotalSales()` is dependent upon/uses data from
              the `Sales`
              class;   **[2]**

              ***Note**: Accept any reasonable example.*

(d)   *Award up to **[4 marks]**.*

        102;
        2;
        2550.40;
        5000.00; (***Note**: Accept, only in this case, 5000)*   **[4]**

(e)   *Award marks up to **[5 max]** as follows:*
*   **[1]** for assigning an appropriate value to total;*
*   **[1]** for correct loop;*
*   **[2]** for correctly updating total OR **[1]** for a reasonable effort;*
*   **[1]** for <u>returning</u> the total;*

```
public double calcTotalSales() // returns total value of all
                               // of the sales
{
  double total = 0;
  for (int x = 0; x<count; x++)
  {
    total = total + salesHistory[x].getValue()
                        *salesHistory[x].getQuantity();
  }
return total;
}
```
                                                                      [5]

(f) *Award marks as follows up to* **[5 max]**:
    [1] *for correct initialization;*
    [1] *for correct loop;*
    [1] *for correct comparison AND* **[1]** *for correct reassignment of total;*
    Note: *Award* **[1 max]** *for these two parts if dot notation is not used.*
    Award **[1]** *for correct re-assignment of index;*

```java
public String highest()
{
  double total = 0;
  int index = −1;
  for (int x=0; x<100; x++)
  {
    if (SalesPeople[x].calcTotalSales() > total)
    {
      total = SalesPeople[x].calcTotalSales();
      index = x;
    }
  }
  return SalesPeople[index].getId();
}
```

[5]

Note: *There will be other correct versions, but the markscheme shown above should still apply.*

(g)   *Award marks up to **[4 max]** as follows:*
***[1]** for <u>correct</u> loop;*
***[1]** for <u>correct</u> comparison (either as part of the loop condition, or within a for -*
loop);
***[1]** for assignment (with or without use of set);*
***[1]** for correct use of get / set methods.*
Note: *Allow for a "follow through" (FT) if the loop is incorrect*

Example of solution with while-loop

```
public void addSales(Sales s, String id)
{
   int x = 0;
   while (salesPeople[x].getId() != id)
   {
     x++;
   }
   salesPeople[x].setSalesHistory(s);
}
```

Example of solution with for-loop

```
public void addSales(Sales s, String id)
{
   for(int i=0;i<salesPeople.length; i++)
   {
     if(salesPeople[i].getId().equals(id))
     {
       salesPeople[i].setSalesHistory(s);
       break;
     }
   }
}
```
[4]

(h)   *Award up to **[3 max]**:*

A date variable must be added to the `Sales` class;
And included in its constructor;
Get/set methods added for the date;
The `calcTotalSales()` would have to be rewritten;   **[3]**

Note*: The spirit of the question is to address the question from a programmer/developer point of view. We are looking for elements of programming/functionality of the programs. Please take this perspective when marking vague responses.*

(i)   *Award up to **[3 max]***:

This is shown by the two constructors in the `SalesPerson` class;
Which is an example of overloading;
The constructor that is executed (by the compiler);
Is the one in which the parameters match up with the arguments in the calling
statement;   **[3]**

15. (a) *Award **[2]** marks for a correct answer (with any variable), and **[1]** if `<Sales>` is missing, but otherwise correct.*

```
private Linkedlist<Sales> salesHistory;
```
[2]

(b) *Answer using an **index**:*
*Award up to **[5 max]** as follows (there are 6 marking points).*
***[1]** for correct initialization.*
***[1]** for correct calculation of size (or equivalent).*
***[1]** for correct loop.*
***[1]** for correct use of `get()`.*
***[1]** for correct comparison (allow FT if no dot notation).*
***[1]** for correctly updating `largest`.*

```
public Sales largestSale()
{ double largest = -1;
  Sales maxSale = null;
  int a = salesHistory.size();
  for (int x=0; x<a; x++)
  { Sales s = salesHistory.get(x);
    if (s.getValue()*s.getQuantity() > largest)
    { largest = (s.getValue()*s.getQuantity());
      maxSale = s;
    }
  }
  return maxSale;
}
```

*Answer using an **iterator**:*
*Award up to **[5 max]** as follows (there are 6 marking points).*
***[1]** for correct initialization of `largest` and `maxSale`.*
***[1]** for correct calculation & initialization of the iterator.*
***[1]** for correct loop.*
***[1]** for correct movement through list.*
***[1]** for correct comparison (allow FT if no dot notation).*
***[1]** for correctly updating `largest`;*

```
public Sales largestSale()
{ double largest = -1;
  Sales maxSale = null;
  ListIterator<Sales> it = salesHistory.listIterator();
  while (it.hasNext())
  { Sales curr = it.next();
    double currAmt = curr.getValue() * curr.getQuantity();
    if (currAmt > largest)
    { largest = currAmt;
      maxSale = curr;
    }
  }
  return maxSale;
}
```
[5]

16. (a)   *Award up to **[3 max]**.*

Both the variables `left` and `right` are examples of reference variables;
They point to the address of another `Node` object;
Or to null;   **[3]**

(b)   *Award **[1]** for clearly showing different <u>levels of recursion</u> (labels not required).*
*Award **[1]** for showing <u>both</u> <u>recursive calls</u>.*
*Award **[1]** for indicating the different <u>nodes addressed</u>.*
*Award up to **[2 max]** for a correct trace showing <u>correct output order</u>*
*   (OR **[1]** for a good attempt that returns a partially correct order)*

Note: *The focus of this exercise is in addressing dynamical aspects of recursion.*

Note: *Award up to **[2 max]** for graphical descriptions that do not address dynamical aspects of computation **but** give evidence of **[1]** five instances of null/empty node, **[1]** correct sequence in output in the order 14, 12, 50, 33.*

Example one:

The table develops incrementally top down;
Each row is filled up once and the length of the row indicates the level of nesting of recursive calls/control;
It shows the calls with the correct argument (accept variants: node2 may be identified as node1.left);
It contains the actions to be performed (return, print(argument), output(argument));
The order of output is identified by the number in the leftmost column (step);

step	**Nesting of recursion (grows on the right)**			
1	print(root)/ print(node1)			
2		print(node1.left)/ print(node2)		
3			print(node2.left)	
4				return
5			print(node2.right)/ print(node3)	
6				print(node3.left)
7					return
8				print(node3.right)
9					return
10				Output 14
11			Output 12	
12		print(node1.right)/ print(node4)		
13			print(node4.left)	
14				return
15			print(node4.right)	
16				return
17			Output 50	
18		Output 33		

Example two:

This is a more informed and detailed variant of Example one
Note*: Not all of the detail shown below is needed and there are various ways of (correctly) showing this trace.*

Level 0	Level 1	Level 2	Level 3
node = node 1 node = null? false node.left ⟶	node = node 2 node = null? false node.left ⟶	node = null node = null? true return ⟵	
	node.right ⟶	node = node 3 node = null? false node.left ⟶	node = null node = null? true return ⟵
		node.right ⟶	node = null node = null? true return ⟵
		output 14 ⟵	
node.right ⟶	**output 12** ⟵ node = node 4 node = null? false node.left ⟶	node = null node = null? true return ⟵	
	node.right ⟶	node = null node = null? true return ⟵	
output 33	**output 50** ⟵		

Example three:

Table constructed top down (implicitly it indicates the step);
It focusses on arguments for the recursive calls and the result of the tests;
the level of recursion is implicit;
It produces the correct output in correspondence of the correctly called arguments;
The sequence of outputs is the last columns read top-down;
Both arguments for the recursive call are displayed;

Node addressed	node == null?	output
node 1	false	–
node 2	false	–
Null	true	–
node 4	false	–
Null	true	–
Null	true	14
node 2	–	12
node 3	false	–
Null	true	–
Null	true	50
node 1	–	33

Example four:

*This diagram allows only to retrieve 5 accesses to null (**[1]** mark); there is no way to reconstruct the correct sequence in output.*

node.left	node.right	node.data + " "
12	50	null
null	14	null
null	null	33

(c) *Award [5] marks as follows, depending on a recursive or non-recursive answer:*

Example, non-recursive answer:

Award marks as follows:
[1] correct method signature
[1] declaration/initialisation of current node
[1] correct loop (with correct condition)
[1] update current node with the node on the left
[1] correct data returned (with use of dot notation)

```
public int smallest(Node node)
{
   Node current = node;
   while (current.left != null)
   {
     current = current.left;
   }
   return(current.data);
}
```

Example, recursive answer:
Award marks as follows:
[1] correct method signature
[1] correct condition for checking node on the left (conditional statement)
[1] update of node with the node on the left (see NOTE below)
[1] recursive call on smallest
*(**Note:** The recursive call may incorporate the call to the left node)*
[1] correct data returned (with use of dot notation)

```
   public int smallest(Node node)
   {
     if (node.left == null)
     {
       return node.data;
     }
     node=node.left;
     return smallest(node);
   }
```

Example where Note needs to be applied

```
public int smallest(Node node)
{
   if (node.left == null)
   {
     return node.data;
   }
   return smallest(node.left);
}
```

[5]