ib Diploma Programme
Programme du diplôme
Programa del Diploma

Markscheme

May 2015

Information technology in a global society

Higher level

Paper 3

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Examiners should be aware that in some cases, candidates may take a different approach, which if appropriate should be rewarded. If in doubt, check with your Team Leader.

If candidates answer more than the prescribed number of 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*.

1. (a) Identify **two** characteristics of a data warehouse. **[2]**

Answers may include:
- integrating or collecting data from one or more different data sources
- a central repository of data for a business, organization or enterprise or government
- data warehouses store current and historical data
- data warehouses store a variety of data such as sales data, customer data that have a common focus, eg from a business
- data in a data warehouse is added to but not changed, and often not deleted but archived.
- data is a warehouse is stored in a non-normalised form (which is easier for summative data)
- May use a non-relational database which allows statistical querying eg an olap / cube
- Data in a warehouse may be anonymised or have non-relevant details excluded
- stored a huge volume of data?
- Often outsourced to/managed by a third party

NB: Answers that identify the use/purpose of a data warehouse are not characteristics, eg support management decisions, for data mining, etc, but the answer may include characteristics.

Simple answers such as "a collection of data", "central repository of data" and "it's where most of the data analysis takes place"are too general and do not identify the integrated storage aspect of the data, the sources of the data or the type of data that separate the data warehouse from other data storages such as personal hard drives, USB memory devices and backups/archives of data.

(b) Outline **one** reason why an extract, transform, load (ETL) process should be used when moving data from one computer system to another. **[2]**

Answers may include:
- the main reason is to transform/convert data from its previous format so that it can be placed into another database with a different format/structure
- data from a source database cannot be processed or stored in another database (simple, RDMS, special) unless it is in a format that matches the receiving database, *eg* a customer's name, NAME (George Smith), stored in one database may need to be separated into FIRSTNAME (George) and FAMILY NAME (Smith) in another database
- an automated ETL process is required because modifying the format or structure of data in the process of moving it from one database to another is too difficult and time consuming by hand
- data in one database may not be "clean", *i.e.* corrupt, not validated, incomplete, ill-defined, non-standard format, or inaccurate, for a number of reasons so it cannot be processed easily. Therefore "cleaning" the data so it can be used by moving it to another database is required
- saves time, compared to writing a program(eg i a DML such as SQL) which does this
- ETL is configurable for many different data structures and can easily be modified by non-technical users
- an ETL tool understands all common data formats by default (databases, flat files, etc) therefore can extract data from many different systems , including legacy systems, transform it to a standard format and load it into the data warehouse format.

NB: The answer must address the transformation process as the extract and load processes are included in the question. The details of how the transformation happens using lookup table and processing rules is not required.

*Award **[1]** for a basic comment related to the transformation process.*
*Award **[2]** for outline of the reason the process is needed.*

2.   (a)   Distinguish between data mining and data querying.   **[2]**

Answers may include:
- data mining is a tool that analyses data from different perspectives and summarizes it into useful information; data querying is more focused on retrieving information from a database which then needs to be analysed
- in contrast to a query which simply returns the data that fulfills certain constraints as expressed in a database query statement (usually in SQL format), data mining constructs models of the data in question. The models produced by data mining can be viewed as high-level summaries of the underlying data
- data querying is a low-level extraction of information from a database; data mining is a high-level extraction of information that produces more complex and comprehensive results than a single query. Data querying is a tool used in data mining and the results are then processed further to look for patterns and formulas
- the results of a set of data queries performed on a database by a person can be summarized into a pattern or conclusion by that person. Data mining techniques automate this process
- data queries produce a set of data matching a specific set of constraints usually expressed in a database query language such as SQL. Data mining produces statistical summaries, patterns of data and formulas that explain patterns.

NB: Essentially these answers are the same but focus on a slightly different type of comparison: results, methods, purposes. Basically data querying is a sub-set/process of data mining and there is significant overlap as the data querying becomes more complex – either automated or by hand.

*Award [1] if data mining and/or data querying are described in isolation up to [2 marks].
Award [1] for each distinction made between the two up to [2 marks]*

(b) "ASI [*Asociación de Supermercados Independientes*] management are concerned that according to data privacy laws there must be appropriate access controls to protect customer and store data when accessed from the database system" (lines 119–121).

Explain **one** control that can be applied to staff access, and **one** control that can be applied to customer access. **[4]**

Answers may include:
General
- managers and employees need to have system logon names and passwords (or other form of security such as fingerprint, eye-scans or cards that need scanning) to enter system
- audit logs of activity on the database that can be examined for unusual/suspicious activity at unusual times and that track the type and extent of data accessed
- access levels to the data and the data mining software depending on the needs of the managers and employees
- legal agreements, user agreements and policies are an acceptable way of control (ie. the answer does not need to just focus on technical solutions).

Accessed in real-time via a web-browser/smartphone dashboard app by store managers
- access to secure website using logon name and password
- limited search and access capabilities from the web-browser/app to summary data only. Detailed search and analysis capabilities restricted to use while on the network in their offices
- set up a VPN that can use the full capabilities of the search and analysis software using client software, or receiving results directly from the database software
- access can be from a locally connected LAN computer only, or limited by a white list of MAC addresses
- access could also be limited by time, ie during their specified working hours.

Accessed by customers through the loyalty card scheme website and smartphones
- users need a logon (loyalty card number, logon name or both) and password to access a secure site using HTTPS
- having a set of security questions that can be used to identify the customer
- registering the identity (IP address, MAC number) of the computer/smartphone that is accessing the site. If accessed from an unknown device a security question is asked
- using a one-time password sent by SMS to the phone or via email each time access is attempted.

Reponses need to explain how the security will be achieved by the measures described.

For each unique control method, one for staff and one for customers:
*Award **[1]** for identification of the control method used.*
*Award **[2]** for explanation of control method and how data will be protected.*

*If the control method outlined is identical for staff AND customers, award a maximum of **[2 marks]***

3. The information obtained by SUSA Analytics revealed that the number of staff working at various times was not always appropriate to the level of sales (lines 96–98). Based on this information, ASI management recommended that patterns of staff working hours should be reviewed by the store managers.

Discuss whether store managers should use this information to decide the hours the staff should work. **[8]**

Answers may include:
General
- sales may be seasonal or promotions so data could be misleading
- if based on value of sales rather than volume then perhaps a lot of work but little profit or vice versa, ie. do larger but more expensive items sell at a certain time of day which would imply needing more staff but workload may be lower.
- different types of staff - more customers may mean more salespeople required but not necessarily more managers, office staff etc. so a proper ratio would need to be found
- examine the data collected initially that was used to decide the number of staff at various working times Train staff to develop skills in areas that need development.

Considerations for:
Staff
- greater recognition of skills and abilities in a more organized and efficient workplace
- higher wages due to greater recognition and less staff required – productivity increase benefits.

Managers/Business
- less staff needed to pay – save money
- more productivity from each staff member as less staff used.

Considerations against:
Staff
- reduced working hours as less staff may be needed
- higher work load as minimal staff allocated each shift
- less flexible working hours due to decreased staff and more rigid shift requirements
- forced to work at certain times due to need for specialized staff in each shift.

Managers/Business
- unhappy staff who have the problems above
- need to spend more time and effort organizing rosters of staff
- need to negotiate with staff, especially those with other commitments such as children
- minimal staff could be a problem if something unusual happens
- managers need to be aware of the specific skill sets of staff so that departments are appropriately staffed.

Note to examiners: *Please use the following simplified version of the markband*

*If there is **no** awareness of the **different** impacts (positive and negative) on staff and/or the managers/business award **[0]**.*

*To enter the 1/2 markband the response needs to present **different** types of considerations, but **without** an attempt at a comparison or a conclusion.*

*To enter the 3/4 markband, responses need to have an attempt at a conclusion that is based on at least two **different** types of considerations for **either** staff **or** the managers/business.*

*To enter the higher markbands, 5/6 and 7/8, the conclusion needs to be based on **explicit** considerations **for** and **against** for **both** the staff **and** the managers/business. However three types of considerations covering **both** staff and managers/business and used to support a detailed analysis may warrant entry to the 5/6 markband – best fit for detailed analysis.*

At least one for each type of the four types of considerations is needed to enter markband 7/8 and are used to support a conclusion.

SL and HL paper 1 part (c) and HL paper 3 question 3 markband

No marks	• No knowledge or understanding of the relevant ITGS issues and concepts. • No ITGS terminology.
Basic **1–2 marks**	• *Shows only a little ITGS knowledge* • *Makes at least one argument.* • *May not have any comparison/conclusion*
Adequate **3–4 marks**	• *Shows a little more ITGS knowledge but still weak.* • *Has more arguments, (at least two) and possibly from different stakeholders.* • *Has a conclusion or judgements which are probably not backed by much reasoning*
Competent **5–6 marks**	• *Shows good ITGS knowledge and detail* • *Has more arguments and they are balanced (+ and -) and for different stakeholders* • *Conclusion/judgements are supported by the arguments and is well thought out*
Proficient **7–8 marks**	• *Shows very good ITGS knowledge* • *Arguments are very balanced and detailed.* • *Conclusion is based completely on the arguments*

4.   With reference to your independent research, to what extent do the benefits for ASI's customers of receiving targeted advertising outweigh the ethical concerns that these customers may have about the impacts on their privacy, security and anonymity (lines 108–117)?   **[12]**

Answers may include:

Positives
- profile of customer purchases used to send advertising to customers may be seen by some customers as a benefit as it provides them with access to the latest product and sales information in a timely manner without the need to do extra research
- targeted personal advertising can remind them of a need without having to make an effort to check
- access to products before others have the chance to purchase them
- access to special deals that others may not be offered
- targeted advertising is cheaper than bulk advertising and more effective so the business will make more money and hopefully pass this on to the customer
- by enrolling in the loyalty card scheme the customers have opted into the target advertising and expect to benefit from it with rewards.

Negatives
- invasion of privacy – profile of customer purchases used to send advertising to customers may be a concern to some customers
  - "How much do they know about me?"
  - "What else will they do with the information?"
- intrusion into personal life – targeted, personal advertising may be intrusive as it will be pushed out to the customer who may not want that amount or type of advertising
- the advertising may be targeted but not fully appropriate to the needs of the customer due to the lack of detail in the profile and the changing needs of the customer which could be annoying
- the customer may prefer to research their own purchases themselves
- is it wrong to force advertising on customers when clearly it is being pushed out to the customer and not to some others?
- is it wrong to use information in the profile without asking specific permission of the customer? Has this been done?
- customers may want to be able to cancel the use of these advertisements so this facility will need to be included in the website. Cancelling requires an effort from the customer
- customers who are targeted with extra advertising may feel pressured to purchase when not able to control their spending or they do not have the money – such as the elderly, children and immature adults for whatever reason
- the customer profile may be sold or hacked into which could have negative impacts on the customer such as unexpected advertising from unknown companies
- also it may be stored in another country with different laws and security arrangements which could lead to the data being unlawfully accessed, sold and used
- by enrolling in the loyalty card scheme the customers have opted into the target advertising but may not have realized the extent to which it would impact them
- customers may be concerned that their purchases may be used to infer other things about them (medical insurance if health products are purchased etc)
- customers might be concerned that their credit card details may be compromised if stored with the profile.

The response must focus on the customer. When business orientated impacts and actions are included they need to be linked to a positive or negative impact on the customer, eg "loss of market share due to customers being annoyed at the advertising" should focus on the annoyance for the customer and not the loss impact for the company.

The focus of the response must be a judgement about the decision that has been made: overall, is the decision good or bad for the customer, or a mixture of both. The judgement needs to be defended by comparing the impacts.

HL paper 3 question 4 markband

Marks	Level descriptor
No marks	• *A response with no knowledge or understanding of the relevant ITGS issues and concepts.* • *A response that includes no appropriate ITGS terminology.*
Basic **1–3 marks**	• *A response with minimal knowledge and understanding of the relevant ITGS issues and concepts.* • *A response that includes minimal use of appropriate ITGS terminology.* • *A response that has no evidence of judgments, conclusions or future strategies.* • *No reference is made to the information in the case study or independent research in the response.* • *The response may be no more than a list.*
Adequate **4–6 marks**	• *A descriptive response with limited knowledge and/or understanding of the relevant ITGS issues and/or concepts.* • *A response that includes limited use of appropriate ITGS terminology.* • *A response that has evidence of conclusions, judgments or future strategies that are no more than unsubstantiated statements. The analysis underpinning them may also be partial or unbalanced.* • *Implicit references are made to the information in the case study or independent research in the response.*
Competent **7–9 marks**	• *A response with knowledge and understanding of the relevant ITGS issues and/or concepts.* • *A response that uses ITGS terminology appropriately in places.* • *A response that includes conclusions and/or judgments that have limited support and are underpinned by a balanced analysis.* • *Explicit references to the information in the case study or independent research are made at places in the response.*
Proficient **10–12 marks**	• *A response with a detailed knowledge and understanding of the relevant ITGS issues and/or concepts.* • *A response that uses ITGS terminology appropriately throughout.* • *A response that includes conclusions, judgments or future strategies that are well supported and underpinned by a balanced analysis.* • *Explicit references are made appropriately to the information in the case study and independent research throughout the response.*