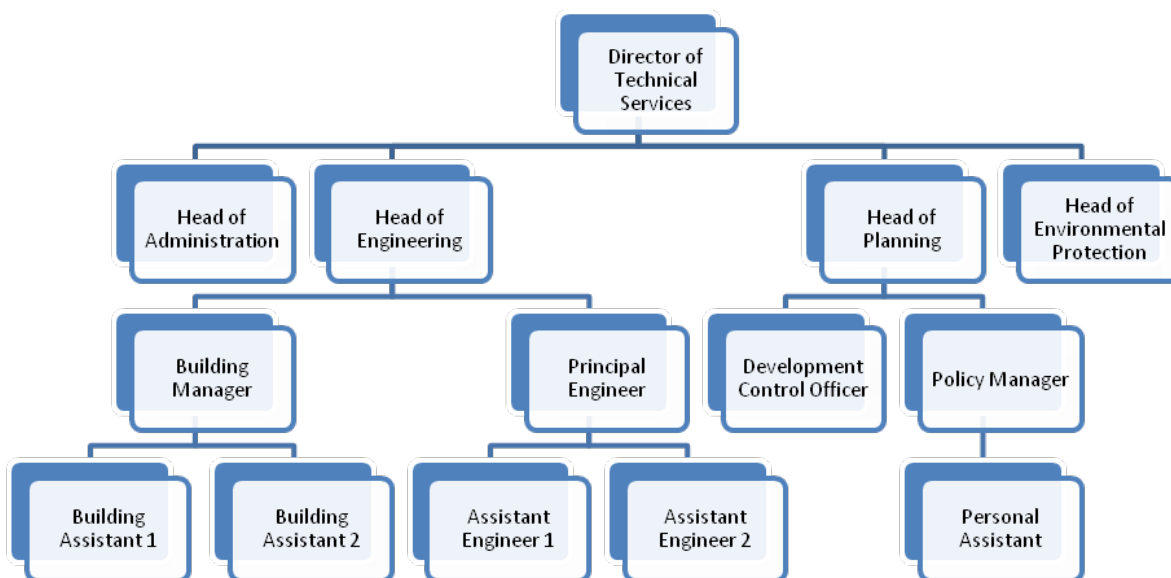**Paper 2 – Option A - Databases**

**SL/HL Core**

A1. A company has a division called Technical Services.  This division is organised into departments. The following chart shows how the various staff are organised.



The Human Resources Department maintains a flat file database to store information about all the staff members for salary, holiday and promotion purposes. A sample of the stored information is shown below.

| Staff Number | Surname | Forename | Job Title | Line Manager | Salary |
|---|---|---|---|---|---|
| 948 | Kent | McKenzie | Head of Planning | Adrienne Mathis | 90875 |
| 156 | Barker | Rae | Head of Administration | Adrienne Mathis | 30465 |
| 815 | Riddle | Kevyn | Development Control Officer | Kent McKenzie | 34768 |
| 580 | Figueroa | Rina | Policy Manager | Kent McKenzie | 45078 |
| 871 | Mathis | Adrienne | Director of Technical Services | Russell Z. Harrell | 45800 |
| 457 | Neal | Paul | Head of Environmental Protection | Adrienne Mathis | 80670 |
| 297 | Oliver | Ralph | Principal Engineer | Daria Gilmore | 50796 |
| 51 | Gilmore | Daria | Head of Engineering | Adrienne Mathis | 90655 |

The Human Resources Department is experiencing many errors when compiling reports of personnel and has been advised that it should change to a relational database.

(a) (i) Define the term *database*. [1 mark]

A database is a store of digital information that is kept in an organized way for simplified retrieval of data.

①

(ii) Describe the difference between a *flat file database* and a *relational database*. [2 marks]

In a relational database data is organised into relevant tables and refers to other tables for the full meaning. In a flat file database all information relevant to a row is stored in one table, in the row itself.

Benefit of doubt given – could be expressed a bit more clearly.

②

(b) (i) Identify two problems caused by *data redundancy*. [2 marks]

Data redundancy is when data is duplicated in a database. It makes errors more likely as it is possible that data could be left in an inconsistent state when records are updated or deleted.

②

(ii) Outline using examples how data redundancy in this database could lead to errors. [2 marks]

In this database the name of the line manager is stored for each relevant row. If the name of a line manager changes, e.g. through marriage or the employee being replaced, then the name will also need to be updated for anyone under the supervision of that particular line manager and it is possible that the name will not be changed in every instance and so leave an employee with the wrong line manager.

②

(c) The departments usually have many ongoing projects that occupy their team members.  The Human Resources Department needs to produce lists showing who is involved in current projects.

The following information has been provided:

**Project**: Project_Ref, Project_Name, Budget

**Department**: Dept_Ref, Dept_Name,

**Staff**: Staff_Number, Forename, Surname, Job_Title, Salary

The projects are owned by a department.

Each member of staff is a member of a department.

(i) Construct an Entity Relationship Diagram to show how a relational database could be designed to make this possible with the least chance of producing errors caused by data redundancy.      [4 marks]
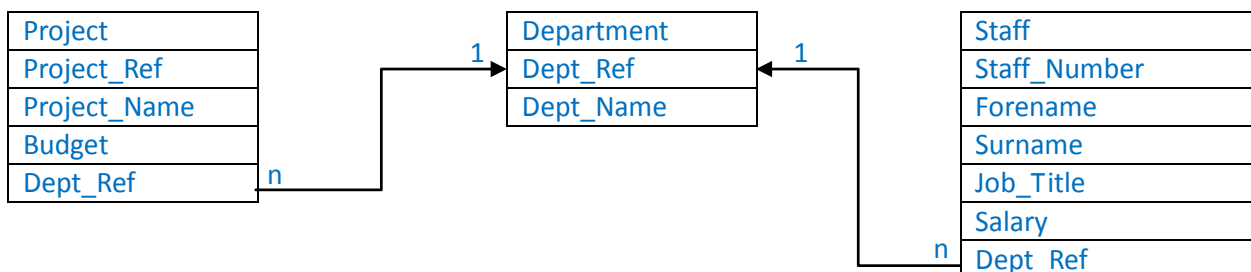


4

(ii) Identify the steps to create a query that could be constructed to produce a list of staff members whose salary is greater than $50000 who are involved in the project called "Relocation_2012". [4 marks]

Select the following fields:  Forename, Surname, Salary from **Staff**, Project_Name from **Project**
Join the tables: **Staff** to **Department**, **Department** to **Project**
Where the criteria are met:  Salary > $50000, Project_Name = "Relocation_2012"

The steps are clearly indicated and there is no requirement to use SQL.

4

A2. A bank wants to transfer money from one account to another. Both accounts are held in the bank's database management system (DBMS). The account DBMS carries out a series of separate operations in order to achieve this transaction.  These include removing of money from one account and adding it to the other.  An error may occur if this process is interrupted, for example due to a power failure.  This could compromise the integrity of the account database.

(a) (i) Identify two characteristics of a *transaction* in terms of a database. [2 marks]

An indivisible action performed on a database and is performed independently of other transactions.

The use of a list for this type of question is preferred.                               (2)

(ii) Distinguish between a database *schema* and a database *state*. [2 marks]

A database schema is the definition of the structure of the database whereas the state of the database includes the data contained within the database as well as any constraints placed on the data.

(2)

(iii) Explain why the interuption described above could compromise the **integrity** of the account database. [2 marks]

The databases state includes the total amount of money, specified across accounts. If one account is debited and the relevant account isn't credited then the database is left in an inconsistent state because the overall amount of money has changed.

Clear explanation.                                                                  (2)

(b) Identify the steps that should be taken by the software in order to maintain the **integrity** of the account database. [4 marks]

The state of the database should be checked before the changes by the software.

The transaction should then take place.

The software should then check the state of the database again

If the changes have resulted in an inconsistent state, then the changes should be undone.

If this happens then the change should either be attempted again or flagged for human intervention.

The use of a list for this type of question is preferred.                               (4)

(c)   The bank account data is shared between various applications, such as the bank's internal accounting systems and the transaction system.  Explain how problems of concurrency can arise and how they can be avoided.                                    [6 marks]

If the data is duplicated between multiple applications then at some point the data must be synchronized between the applications, there will be some latency involved with this synchronization and so different applications will have slightly different versions of the data. If a transaction occurs before the data is updated then an account could be debited with funds that have also been debited elsewhere. Also, if two transactions occur at the same time then a race condition occurs and the final value of the data will depend on which transaction occurs last.

In order to avoid concurrency issues any transactions should take place as soon as possible so that any future transactions are based on the correct current values, also, while a transaction takes place the data should be locked so that other processes cannot edit the data or get incorrect values.

The data should be stored in a single location rather than different applications having their own individual local copy as the presence of local copies gives rise to the possibility of different copies of the data existing and as many transactions will be based on current data values this will lead to incorrect values being used as the basis of transactions.

The response shows a good understanding of concepts and presents a clear description of the types of errors that could occur and explains how they can be avoided.

6

A3. (a) (i) Identify two functions of a  database management system (DBMS).          [2 marks]

A database management system is responsible for the creation and the use [too vague] of a database, or set of databases.

A list would have been preferable.

( 1 )

(ii) List two tools usually provided with a DBMS. [2 marks]

A query construction tool
Data import/export tool

( 2 )

(b) Database management systems allow data sharing and multiple views.  Explain why these features are important to an organisation that uses databases.                                        [4 marks]

Within an organisation there are likely to be many people employees who need to access the same data, this makes data sharing important so that employees can easily and simultaneously access needed information. Having multiple views is important so that only the relevant information is presented to people accessing the data. There is likely to be information in the database that is either irrelevant or confidential for some employees and so this information should be hidden from viewers that do not need to, or shouldn't, see the data.

A well explained response.

( 4 )

(c) Explain why it is important for data to be independent of the application software that manipulates it.                                                                                          [6 marks]

The same data could be used in many different within an organization and so different applications would be used in order to view and edit the data. If the data is only compatible with a particular piece of software then all interaction with the data must occur through that software. Also, software can have a limited lifespan and if support ends for a particular piece of software or a new feature is needed then it may not be possible if there is a heavy reliance between the software and the data.

 If, however, the data is stored independently then it will continue to be valid despite different software being used. Also, the converse is true as if the structure of the data is changed, for example by adding another column to a table then the independence of data and software will ensure that the change will not break the software.

This was awarded 6 marks.

Although the response is brief, the student presents a clear and detailed explanation.

( 6 )

**HL extension**

**A4**. Meubles de France is a large company that manufactures and sells furniture. It maintains a large factory and offices in Provence and over a hundred retail outlets. It employs hundreds of workers. It holds huge amounts of data on separate computer systems to handle:

- customer orders
- employees
- sales data
- production data
- finance
- budgeting.

All these systems have been acquired from different software suppliers. The management of Meubles de France has been advised that a data warehouse would be of great benefit to the business.

(a) (i) Identify two characteristics of a data warehouse.                                        [2 marks]

A data warehouse provides a way to aggregate information from different sources within an organisation. It also provides relevant views and reports based on the data.

A list would be preferred.

2

(ii) Identify four features of Meubles de France's business that indicate the need for a data warehouse.
                                                                                                [4 marks]

Multiple sources of data.

Data from different computer systems which may contain unlinked but related data.

As the operational systems come from different software vendors it is not likely to be possible to simply combine the data without the assistance of a data warehouse to interpret the data.

A data warehouse would also provide a central location to generate reports on the business to give better indications of the state of the company.

The use of a list for this type of question is preferred.

4

(iii) Identify four transformations that may be necessary in order to produce usable data in a data warehouse.                                                                [4 marks]

Data from different systems will have to be converted so that any symbolic information is in the same form.

Certain parts of the data will need to be merged

Any duplicated columns will need to be removed to stop redundant data being present.

The data may also need to be validated

The use of a list for this type of question is preferred.                    ④

(b) Meubles de France uses data mining to uncover patterns of consumer spending so they can optimize their marketing for future products and services.

Contrast the use of association and cluster analysis for this purpose.                    [4 marks]

Association looks at the way that events are connected. For example, it could look at the typical products that are purchased together and inform the management of Meubles de France on methods of marketing these products in the future. For example, finding that customers who buy garden furniture are also likely to buy garden plants.

Cluster analysis can organize the consumers of Meubles de France into particular groups or clusters from the complete dataset held in the data warehouse based on a number of variables. For example, a cluster of people who shop in the garden centre are also found to be over 50 years old and provide the majority of the lunchtime revenue in the restaurant during weekdays. These groupings can to be used by the managers to move the restaurant close to the garden centre.

Clearly and concisely explained.                                              ④

Many organizations such as Meubles de France use data mining to build a complex profile of its customers

(c) Explain why civil liberty groups may be concerned about Meubles de France having such detailed information.                                                          [6 marks]

Data mining can allow Meubles de France to build up a complex data profile of the customer. However, once the data has been collected it could be used by other companies in a way that the customers of Meubles de France did not think about when they provided the data originally. Civil liberty groups may also be concerned that other companies who may access this data may use it for purposes that may lead to unexpected issues for the customers of Meubles de France, for example, if information may be passed from the third party to other organizations which could lead to increased costs in healthcare insurance or to Government agencies that may infringe the customers privacy.

Clearly and comprehensively explained.

6