

National College of Ireland

**BSc. (Honours) In Computing, Year 2 (BSHCIFSC, BSHCE)**

**Higher Certificate in Computing, Year 2(HCCOMP, HCCOMPE)**

**Release Date-Time:**

**Submission Date-Time:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Advanced Databases**

**William Clifford**

**Enda Stafford**

**Instructions**

1. This is an open book Timed TABA.
2. You are allowed to use your class notes. If the lecture notes are used as a resource for this assignment, do not copy them directly but rather paraphrase them (write in your own words).
3. **You are not allowed to discuss your solution with any person other than the lecturer or an approved member of staff during the examination. If it is found that a student has discussed his/ her solution with anyone else, the case will be referred to disciplinary committee for further actions.**
4. This is a Turnitin assignment, and the plagiarism will be checked based on Turnitin database. It will be used to check whether a text is copied from Internet, any other source or peer students.
5. You should need to submit the solution of this assignment on your Moodle page at the TERMINAL ASSIGNMENT Link. You should submit the PDF (or DOC) file by the end of your exam time.
6. You can also draw a diagram/ figure on a piece of paper, take a picture with your mobile phone camera then you submit the photo on Moodle. Please ensure that you write below your drawing the question number the drawing corresponds to.
7. Use a single column layout document.
8. Font size for the body of the text should be 12-point Times New Roman/ Arial.
9. Include student name, student ID and course name at the top of the first page.
10. The question number being addressed must be clearly indicated in the document.
11. Attempt **all** 5 Questions, each is worth 20 marks and add to a total of 100 marks.

**Question 1.**

1. Concurrency control has 2 common problems discussed in our lecture, listed below. Go through the definitions of each and provide an example that can occur in the real world (be careful not to reproduce an example done in class).
   1. Inconsistent analysis problem
   2. Phantom read problem.
2. **marks)**
3. Write code similar to the example below to represent one of the examples you discussed above. **(4 marks)**

A close-up of a computer code

Description automatically generated

1. Construct a precedence graph to show if the code you created for part (b) is serializable. Explain with words and make use of the graph to show whether the two programs are serializable or not.

**(6 marks)**

**Question 2.**

1. Using the first three sentences in your answer to Question 1 (a), construct the binary term document incidence matrix associated with the entire contents of the three documents Doc1, Doc2 and Doc3. 
   1. **marks)**
2. Evaluate the term vectors for two of the words that you have included in your documents.
   1. **marks)**
3. Using the same sentences from q 2(i) construct an inverted index. Explain which technique (document incidence matrix vs inverted index) is more appropriate for the sets of data being processed.

**(7.5 marks)**

1. Evaluate a Boolean query over two terms from your inverted indexed terms.

**(2.5 marks)**

**Question 3.**

1. Discuss some computer-based controls used to ensure data on computers are reliably kept safe from damage and protected from malicious 3rd parties.

**(5 marks)**

1. Discuss current data protection and privacy laws used today within the EU and United States of America. How might these effect the development in an online analytical processing (OLAP) environment both in terms of data collection and analysis.
   * 1. **arks)**
2. Data can be physically stored using different methods. In class we covered heap, ordered, and hashed storage. Define and describe two of these methods of storing data.

**(5 marks)**

**Question 4.**

Using the last 2 digits in your student number select two database types from the table below. Get the sum of the last two digits in your student number and depending on whether this is odd or even this will give you a different set of areas of analysis to select from the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Database Type** |  | **Areas of Analysis** |
| 9 | MongoDB | Even | Transactions  Recoverability  ACID/BASE Properties  Concurrency  Scalability |
| 9 | Redis |
| 7 | Cassandra |
| 6 | Neo4j |
| 5 | HBase |
| 4 | CouchDB | Odd | Types of Data  Query Language  DBaaS  Cost  Security |
| 3 | InfluxDB |
| 2 | Riak |
| 1 | Memcached |
| 0 | Realm |

i) Using your database types and areas of analysis compare and contrast how these databases perform in each area of analysis you have been assigned.

**(12.5 marks)**

ii) Identify what tasks one database would be more appropriate for over another.

**(7.5 marks)**

**Question 5.**

Using the last two digits of your student number select an industry and business area.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Industry** |  | **Business Area (Entity relationship)** |
| **9** | Education | **0** | Productivity (Projects – Tasks) |
| **7** | Agriculture | **2** | Customer Support (Customers – Cases) |
| **5** | Sports | **4** | Invoicing (Invoice – Items) |
| **3** | Health | **6** | Purchase Orders (PO – Items) |
| **1** | Advertising | **8** | Logistics (Parcel – Contents) |
| **0** | Leisure | **1** | Inventory (Item – Categories) |
| **2** | Retail | **3** | Payroll (Employee – Entries) |
| **4** | Finance | **5** | Recruiting (Candidate – Skills) |
| **6** | Airlines | **7** | Training (Module – Students) |
| **8** | Food | **0** | Sales (Orders – Items) |

To create an OLAP (online-analytical processing) database, you will have to implement a data mart corresponding to the above industry and business area you have been assigned.

i) Create a dimensional model (star or snowflake schema) for the data mart to support whatever business requirements you believe are required (highlight what these business requirements will be below the schema to reinforce your design choices).

**(7.5 marks)**

ii) Write SQL scripts to create a fact table and dimensional tables matching your above schema.

**(5 marks)**

iii) Write an SQL script that loads data into this Data Mart (you can use Sakila Database) and provide some evidence of either manual or automatic loading of data into the data mart.

**(5 marks)**

iv) Write an SQL script to create a data cube to enable a multi-dimensional analysis of the data you have created.

**(2.5 marks)**