

National College of Ireland

**Advanced Databases**

**(BSHCIFSC2\_A, BSHCIFSC2\_B, HCCOMP2, BSHCIFSC2\_ColaisteDhulaigh\_Rathmines, BSHDS2, HCDS\_2, BSHCE2, HCCOMPE2)**

**Release Date: 15/10/2024**

**Submission Deadline: 15/11/2024**

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

**Continuous Assessment (CA) Type: Project 40%**

**Dr William Clifford, Dr David Hamill, Dr Anh Duong Trinh**

**Weight:** The assignment will be marked out of 100.

**Instructions:**

1. There are two parts to the Assessment, see a description below for each section.
2. Students will work **in teams of two to three** for the first section of this assignment.

**SUBMISSION DETAILS**:

1. Section 1 will take the form of a verbal presentation, where the pair of students will present the database they have designed and their rationale for certain design choices. A single student from each team should submit any supplementary materials to Moodle, i.e. SQL files, or scripts. This student should also submit the PowerPoint presentation they used for their live presentation to Moodle.

1. Section 2 will be a written report and this submission should be submitted to TurnItIn on Moodle. Each student should submit their report for this section separately.
2. **Late submissions will not be penalized if the student applied for an extension through NCI360 and it was approved.**
3. Please note that the **lecturer cannot grant extensions**. Extensions can only be granted if approved after the submission of a personal circumstances form. Please consult NCI360 or programmed coordinator for further information.
4. **All submissions will undergo electronic screening to detect any signs of academic misconduct, such as plagiarism, collusion, or misrepresentation. Any submission found to contain evidence of such misconduct will be reported to the college's Disciplinary committee for further disciplinary action.**

**Module Learning Outcomes**

On successful completion of this module the learner will be able to:

|  |  |
| --- | --- |
| LO2 | Implement techniques for query processing and optimization in relational DBMSs. |
| LO6 | Critically assess the suitability of novel data models in different contexts in order to implement effective data management solutions |

**Section 1 (60% weight)**

Students will construct a database. They will identify an important business process and associated business requirements. The students may use a database management system of their choice including but not limited to MySQL, Neo4J, MongoDB, Cassandra etc. **Note:** if you use a basic database management system, like Access it will be difficult to justify it’s use for a large business application.

The business process will be chosen from the list of industries and business areas found in the table below. Both group members student numbers should be used to select the industry and business area.

Student 1 (the student with the lowest student number): using the last digit of your student id, please select a topic from the industry column.

Student 2 (the student with the highest student number): using the last digit of your student id, please select a Business Area.

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

Students must complete the following tasks in relation to constructing their database and include each of these points in their final presentation:

1. Students must identify at least one important business requirement for each member of their team. This should be relevant, based on their selected industry and business area. Here’s an example below **(individual marks)**:

Education payroll: the payroll department for a school or college may want to analyse multiple dimensions to identify when the school was most in need of additional teaching staff. To do this they could examine a time dimension (table) alongside the number of staff they employed at that given time. Another dimension table that logged tickets with the teaching support centre could be used to highlight the need for additional teaching staff for a particular module. Based on these dimensions’ payroll could highlight what stages in the year additional teaching staff were required for specific modules.

1. Students should highlight the relationships that exist between either tables or document structures that were used for storing data in their database **(shared marks)**.
2. Students should show examples of creation statements used for storing data to the database **(shared marks).**
3. Students should show examples of insertion statements used for loading data to the database **(shared marks).**
4. Students should show at least one query that relates to the business requirement they each outlined in part 1 **(individual marks)**.
5. Finally, students should justify why they selected their database management system. They should relate this to their chosen business requirements from part 1 **(individual marks)**.

**Section 2 (40% Weight):**

Students must write the following prompt into ChatGPT:

“I have chosen to use <Insert database management system here> to build a database. In 300 words, can you describe the following areas of interest in relation to this database management system <Insert properties here>”

For each student in the team, they should include the following properties:

|  |  |
| --- | --- |
| Student with lowest student number | Database as a Service  Recoverability |
| Student with highest student number | ACID/BASE properties  Partitioning |
| The other student, if applicable | Scalability  Security |

After ChatGPT provides a response, it will be the job of the student to identify what parts of ChatGPT’s statements are factual. To do this the students will have to find where in the database management system’s documentation they state these things. If it is fabricated, then the student will have to state and cite what the truth is. The final report with the citations should be submitted to Moodle.

Students will be awarded marks in this section for identifying factual information provided by ChatGPT and backing this up with citations. Any statements provided by ChatGPT that cannot be verified with a citation will remove a percentage of the mark in proportion to the total number of overall statements made by ChatGPT. For example, if ChatGPT said the following:

**Recoverability**

1. **Data Durability**: MongoDB provides options for ensuring data durability through journaling. This means that write operations are logged in a journal before being applied, allowing for recovery in case of a failure.
2. **Replica Sets**: MongoDB uses replica sets to enhance data availability and recoverability. A replica set consists of a primary node and one or more secondary nodes. If the primary node fails, a secondary can be elected as the new primary, ensuring continuous availability.
3. **Backups**: MongoDB supports various backup strategies, including full backups and incremental backups. Tools like MongoDB Atlas provide automated backups, while on-premise setups can utilize tools like mongodump and mongorestore.

There would have to be at least three citations to either backup or refute these points. If there were only three statements made, then each citation would be worth 33%. Before editing ChatGPT’s report to add citations, you will be required to show the prompts you have given ChatGPT and the response it provided shown in screenshots.