

## **INFS4205/7205 Assignment 2 (20%)**

**Due dates:** Various (see details below)  
*All submissions/reports to UQ Blackboard only.*

In this assignment, students will be asked to form a group of 5-6 students to complete a project on a topic of their choice related to theoretical or practical aspects of advanced techniques for high dimensional data. Each group needs to choose **either** a research-focused task (Type A) or an implementation-focused task (Type B) and propose a topic of appropriate scope and relevance which will be approved by teaching staff.

### **Forming Groups**

Students should first attempt to form groups of 5-6 students by **5pm on the 3rd of May (Friday)**. Once they have formed a group, an email should be sent to the tutor via [t.zhou@uq.edu.au](mailto:t.zhou@uq.edu.au) listing the name and student number of all students in that group. Emails forming groups of less than 5 students will be ignored. Students looking for groups or recruiting group members should do so at the weekly lecture, consultation or via Piazza. Groups containing both INFS4205 and INFS7205 students are allowed.

Students who are not part of a complete group by this deadline will be randomly assigned to a group with preference to course code. No changes to grouping are allowed after this point of time.

### **Type A: Research Focused Tasks**

In this course we have discussed some fundamental concepts of high dimensional data management and processing. There are still many challenges and open issues in this area in terms of both efficiency and effectiveness, such as high dimensional indexing and search, kNN and skyline query processing, spatio-temporal query processing and mining, and retrieval of multimedia and social media data. Research results on these topics can be found from many recent papers published in leading conferences, such as SIGMOD, VLDB, ICDE and ACM Multimedia conferences.

If a group decides to do a Type A project, they must choose *at least three research papers* published on a topic of their choice that is related to a problem in high dimensional databases. The research papers must have been published in one of the four conferences listed above (SIGMOD, VLDB, ICDE, ACM Multimedia). Each group will need to submit the following assessment items via Blackboard:

1. A research proposal by **5pm on the 13th of May (Monday)**.
2. A research report by **5pm on the 31st of May (Friday)**.

Only a single group member needs to submit these assessment items. The following subsections describe each assessment item in further depth.

### **Type A Research Proposal (due on the 13th of May at 5pm)**

The research proposal is a document up to 2 pages long that contains:

1. A clear and concise explanation (around a paragraph long) of the proposed topic and the scope of research report
2. A brief description of each of your chosen research papers and how they are relevant to the chosen topic
3. A justification of why the chosen topic is of suitable complexity and how it is relevant to high dimensional databases
4. An initial allocation/delegation of tasks to group members

The topic should be chosen carefully, based on discussions among team members and possibly also with the Lecturer/Tutor. *[Proceedings of these conferences can be found online (using Google or DBLP), and full text of the papers can often be downloaded freely. Should a password be needed, you can (1) find an alternative source which does not require password; (2) access from UQ Library through UQ's site licence; and (3) talk to your Lecturer immediately after failing other options.]*

### **Type A Research Report (due on the 31st of May at 5pm)**

The research report is a document up to 6 pages long that includes/discusses:

1. A clear and concise explanation of the problem
2. Application scenarios for the problem
3. Related work prior to the chosen research papers
4. New solutions proposed in the papers and their relationship with each other as well as other related works
5. Methods used for evaluating the proposed solutions and experimental results
6. Potential issues still remaining for that problem

## **Type B: Implementation Focused Tasks**

In this course we have studied spatial databases and spatial indexes in depth. Groups doing a Type B project must complete a technical task which aims to implement and test the performance or behaviour of spatial query processing (e.g., to support what is discussed in Assignment 1, or to test the performance of query processing using different types of spatial indexes).

Implementations can be done using existing commercial database management systems, such as Oracle Spatial. Alternatively, students may prefer to develop their own implementation, for example Z-value based indexes or R-tree indexes. No support is provided from the course staff on how to use your chosen system. Each group will need to submit the following assessment items via Blackboard:

1. A research proposal by **5pm on the 13th of May (Monday)**.
2. A research report by **5pm on the 31st of May (Friday)**. Any other documents, such as database design, source code and test data should also be submitted with the research report in a .zip file.

Only a single group member needs to submit these assessment items. The following subsections describe each assessment item in further depth.

### **Type B Research Proposal (due on the 13th of May at 5pm)**

The research proposal is a document up to 2 pages long that contains:

1. A clear and concise explanation (around a paragraph long) of what you are trying to implement and test (e.g. behaviour or performance of spatial indexes).
2. A proposed plan of implementation and testing tasks that will be required to test this with brief justifications.
3. An initial allocation/delegation of tasks to group members
4. Initial design choices such as software, platforms or datasets.

The topic should be chosen carefully, based on discussions among team members and possibly also with the Lecturer/Tutor.

### **Type B Research Report (due on the 31st of May at 5pm)**

The research report is a document of up to 4 pages long that includes:

1. A clear and concise explanation of the problem and motivations for implementation and testing
2. Documentation of the implementation used including justification for design choices
3. Testing methodology

4. Clearly presented experimental results
5. A discussion and conclusion based on these results

For reiteration, any other supporting documents, such as the database design, source code, test data or data references, and key screen dumps should also be submitted with the research report in a .zip file. **If your research report is not supported by the submitted supporting documents (e.g., missing or incorrect source code), then your final report can be nullified at the discretion of the markers.**

## Presentation

Documents should be submitted as a PDF in a format similar to this document and many of the provided weekly reading research papers. Documents should be properly referenced using the IEEE referencing style.

## Assessment

The same score will be given to all members of a group. A peer assessment form will be made available after the final assessment due date.

**Research Proposal [3 marks]:** description clarity of your selected task [1], proper justification [1] and consideration of required resources [1].

**Research Report [13 marks]:** the report will be marked as a whole package. The completeness of your work (in comparison to the scope established in the research proposal, unless modified with written permission from Lecturer/Tutor) can affect the overall marks of your report. This report is expected to have a good general introduction to the problem and overall cohesion towards your topic [3]; a clear, easy-to-read and self-contained technical section [6] to describe (i) (for Type A tasks) the problem definition and variations as studied in your selected papers, an outline of their key ideas in your own language and their relationship; or (ii) (for Type B tasks) the system you choose to use, the design of your work, the objectives you plan to achieve, and the methodology you choose to follow; and a findings session [4] to outline how evaluations are done and the results of your study.

**Presentation [4 marks]:** presentation marks will be given at marker discretion based factors including but not limited to report structure, readability, clarity, and referencing.

**Please design the project conservatively and manage your group work and time carefully to ensure your team can complete this assignment in time. Each group is welcome to discuss with the Lecturer/Tutor when doing this assignment via provided communication channels (contact hours, Piazza, email).**

## **Academic Misconduct**

All submitted assessment items should be original work. Students are expected to be familiar with the School's student misconduct guidelines. Further questions regarding what is acceptable should be directed to course staff via the course's provided communication channels.