COMP90050, Winter term, 2023
Project Description

Report submission due: **July 24, 2023 Monday, 11:59pm** Melbourne Time

This project forms 40% of your final marks. The project is about creating **a survey on a recent topic in Databases as a presentation and a report, where the report will also include your reflections (as a group and individual reflections)**. Writing a survey is a cornerstone activity that we would like you to learn as well as teamwork. In our sector, this is an activity you would need to do regularly to keep up-to-date with developments and in most cases present to your supervisors/companies that you work for in the form of a white paper/position paper and a company presentation. Most IT companies have brownbag lunch hours where people present such new topics to co-workers.

This is a group project of **4** students. **The topic needs to be chosen from the list provided at the end of this document.** Here are the details of the project components -

1. **Forming a group:** This is an intensive semester of only 4 teaching weeks, so you need to form a group by the end of week 1. Please make use of your tutorial time and the Ed discussion in Canvas to find group members. **Students from different tutorial slots can be in the same group. You need to create a group with your group members in the People section of Canvas.** Forming a group in People section of Canvas help https://community.canvaslms.com/docs/DOC-10516-421264913.

2. **Choosing a topic:** Choose a topic as a team from the list of hot topics in databases **given below** to start the project. There is no advantage of choosing one topic to the other. Just a discussion among team members about their liking of an area is all that is needed to choose a topic. Then, **by the end of week 1 (July 2, 2023 11:59pm), you need to submit your group information and topic in Canvas->Assignments->A. Group topic submission.**

   Submission format for group topic submission: Only one member from each group needs to submit the following information - Group ID (after joining in a group in the People section of Canvas you will see a group ID), Name and student ID of the group members, and the topic chosen by the group (the exact formatting of the text is not important). If the topic of your group changes, you may re-submit. However, the first submission needs to be made by the due date **July 2, 2023 11:59pm.**

3. **Starting the work and expectations:** For your topic of choice, you should initially start reading Wikipedia, newspaper/magazine, and similar articles/webpages to get a high-level idea. These articles are not adequate for a survey though, but good for a start. After this initial phase, you should use **scholar.google.com or similar scholarly search engines** for performing a more detailed background search and do further reading. **You should distribute the work evenly among team members.** These specialized search engines give you the papers you will cite for your survey that will finally count. Generally, highly cited papers are more important ones in an area. You should use multiple keywords/searches to cover a topic to find all the related papers. If you login to our library with your student credentials, you will be able to access most papers and some books for free.

You are encouraged to find other survey papers on your topic. You do not need to start your survey from scratch. There would be many surveys out there potentially. Find the one that is most recent. Better, find many and you will see authors look at similar but not the same set of algorithms/papers. They may also have different classifications and organization. These should give you an idea on the common/popular methods and the key comparison parameters. They are also good examples on how to write surveys. **You cannot use other surveys or other papers at large to copy directly into your survey!** These should not be the sole source of your work. After you read and refer to these, you need to go to individual key papers mentioned in those surveys and your searches and start reviewing them to form your own opinions, judgements, and categorisations. You also need to check more recent works that may not appear in earlier surveys (although many of the categories would likely be similar to older survey papers of the topic/area.)

Y**ou are expected to not only list top papers in an area one after the other, but also categorise these developments/approaches and compare/critique them**. This is at the core of a survey. A list of papers with comments attached is called an "annotated bibliography" and is not a survey and is not the purpose of this project. It is important to note that we do not expect you to learn every paper in detail and to be perfectly comprehensive about the topic to cover all the papers. But rather cover the key papers/directions and classifications/parameters. Digest the key directions and ideas!

You should prepare your presentation in tandem with your report that pretty much presents the key points in the report. After this exercise students are expected to have a good idea in the topic and be able to answer questions during their presentation.

Following **report section-headings/structure needs to be followed (with recommended approximate page counts per section mentioned and that aspect could vary to a degree)**:

- Identification info for students/title/abstract/etc (1 cover page)
- Introduction to the Topic Area (1 page)
- Related Work Details (papers covered explained in brief in some structured manner, categorisation, how a paper extended an approach from a previous paper, etc.) (6 pages)
- Comparison of Key Approaches/Papers (benefits and disadvantages from various aspects, when some approach is more applicable than other, etc.) (2 pages) [Note: This part can be presented combinedly with related work]
- Conclusions/Discussions and Future Directions (1.5 pages)
- Group reflection (0.5 page) [details below]
- Individual reflection (maximum 200 words per student in the group) [details below]
- References (1 page)

Top conferences that you can find papers from on these topics in Databases are SIGMOD, VLDB, ICDE, SIGKDD, SIGIR, SIGSPATIAL, ICDM among others. There are also top journals we can recommend, TKDE, VLDBJ, ACM TODS. Other top tier publication venues are also acceptable. About **15** papers are generally enough to cover the basics of an area (not including other surveys), but this number is not a hard requirement for the project. There are numerous conference and journal rankings you can check for venue rankings in computer science. You also need to cover recent publications. As particular requirement, **you need to cite at least 2 recent publications (published on or after 2021).**

When reading the papers, please note that a technical paper is not read like a novel, i.e., not read from cover to cover sequentially or slowly, but is read in a manner that you can quickly grasp the key ideas, benefits/disadvantages. (Although technical papers can be read to the very detail or even one can contact authors for implementations, but we do not need that level of reading per method/paper from you for your project/survey.)

Group reflection: For each group report, a group reflection needs to be included (only one per report) of approximately **0.5 page**. Each group reflection should answer the following questions: As a team, how did you work on finding the reading materials for this survey? How did you decide which references to include in your survey? After completing this group project, if your team is asked to write another survey on a different topic, would you do something differently to find the reading materials? Why or why not?

Individual reflection: In each group report, each student from the group needs to include maximum 200 words on their individual reflection, clearly marked with their name and student ID. Each individual reflection should include – What are the references of this report that you were responsible for (simply mentioning the reference numbers is sufficient)? If you were allowed to pick only one paper from those references (the ones that you were responsible for) to include in the report, which reference would you choose, and why?

4. **Presentation: The presentation will be done as a team (with all members presenting) at the end of the semester based on a schedule announced by the staff member closer to the dates.** Given the dual delivery mode, the presentations can be done as Zoom sessions if needed. Structure of the presentation should follow the report as well, i.e., in terms of the key sections it involves. **You do not need to include group or individual reflection in your presentation.** The presentations will be allocated **no more than 25 mins** per group including questions and setup time etc. Thus, we recommend no more than 14-18 slides for presentations. **The presentation file needs to be submitted before your presentation in Canvas->Assignments->B. Group project presentation as a ppt, pptx, or pdf file** by only one member from the group.

5. **Report submission:** The deadline of this project is July 24, 2023 11:59 pm. Late submissions will get a penalty of 10% per day. **The report must be submitted as a PDF file on LMS->Assignments->C. Group project report.** Handwritten reports are not accepted in any form. The report should be submitted by only one person in your group, but all students' IDs should be visible in the cover page.

The report should be in **A4 size, 11-point Times New Roman font, single column,** main text with **1.5 line spacing** with **1-inch margins**. The report **should not exceed 16 pages (but also no less than 10 pages) including everything**. Figures and Table(s) (e.g., for comparing methods) is highly recommended and crucial in some cases. Text that describes methods is important to understand the algorithms. **Indicative number of words is 3500 words (excluding references), but may vary based on the informative tables and figures.**

Plagiarism check: All explanations and writing should be your own words and proper citations should be used when needed. Teams should work independently of other teams and plagiarism as usual will be checked by markers and our systems. Not sharing information about papers you found is also important as finding papers is a part of the project. Submissions are checked in our systems with other submissions, including other subjects, offerings from

different semesters, websites, papers in search engines, etc. through a professional plagiarism checking system that **includes AI writing detection.**

List of topics

Here we give a short list of hot **topics that you can choose from**. **No other topics are acceptable.** Topics that you can choose from are:

- Crowdsourced and Collaborative Data Management (In a nutshell: Databases where many people contribute to the data and workings of the system)
- Self-driving Databases (In a nutshell: Databases that take over functions from the Database administrators such as tuning, index selection, improve performance, etc.)

You need to focus on Database aspects of these topics as at times they relate to other areas of Computer Science as well. The database aspects are included but not limited to storage, indexing, index selection, query processing, efficient joins, query operations, reduction factor/selectivity/cardinality estimation, transaction management, query planning, database reliability, concurrency, etc. **You can pick 2-3 database aspects of your topic and go deeper on those aspects** (for compare, categorisation, contrast, etc.).

Marking

The presentation constitutes 15% of your final marks. The **students in the same group will receive the same mark** for the presentation. Each presentation will be marked in three aspects:

- Knowledge: understanding of topic and comprehensiveness of discussion
- Delivery: clarity and engagement
- Teamwork: time management, flow, and distribution of workload

Each aspect will be **marked from 0 to 5**. You do not need to present group reflection or individual reflection during the presentation. The mark of your presentation will be calculated as the sum of the three aspects. Detailed marking criteria for presentation:

	0	**…**	**5**
Knowledge	• Does not cover any representative work related to the topic • Content is focused on a wrong topic		• Covers the representative publications/products related to the topic • Compares different approaches (methods/ algorithms/ products/etc.) • Covers the importance/potential of the existing work

Delivery	• Difficult to understand, e.g., full of technical jargons • Content is totally unorganized		• Slides and oral presentation are easy to understand for general audience • Presentation is clear and captures the audience
Teamwork	• Unbalanced presentation time between team members • Content from different team members is unrelated with each other • Time length of the whole presentation exceeds the limit		• Workload is well balanced between team members • Content from different team members tells a full story with fluid presentation flow • Time length of the whole presentation is well controlled • Can handle questions from the audience

The final report is 25% of your final marks. **Students from the same group can receive different marks for the 'individual reflection' component, but will receive the same marks for all the other group components of the report**. The marking is similar to presentation marking above in many ways. Detailed marking criteria for report:

	0	**…**	**5**
Knowledge Coverage	• Report is focused on a wrong topic • Does not cover any representative work related to the topic	…	• Shows a comprehensive survey of the work related to the topic. • Presents representative works in the body of the report and in the references. • Adequate number of significant publications on the topic are selected, including recent papers, highly cited papers, papers from top-tier venues. Note that, you need to include at least 2 papers from 2021 or later.
Related Work Structure	• Papers are covered in a totally unorganized way • Development of the area is not visible to the reader	…	• Papers are well organized and categorised • Reader can see how the area has developed and/or covered by papers in subareas

Writing and formatting	• Contains many grammatical errors • Writing is difficult to understand • Writing style is not academic, e.g., using an informal tone • Format is not consistent with what is prescribed • Layout is awkward • Style of references is not consistent and academic across the reference section	…	• Grammatically sound • Readers with minimal knowledge of the topic can understand the content • Uses academic writing style • Shows figures/diagrams that help readers understand the content • Format is consistent • Uses correct layout All the references use the same proper style
Reflection (2.5 for group reflection + 2.5 for individual reflection)	• No consistency among the report and the reflection • No critical reflection	…	• Demonstrates critical reflection, based on own experience • Consistent with the references mentioned in the report
Critical Analysis & Comparisons	• Plainly lists all the approaches without analysis	…	• Shows motivation, practical use and/or potential of the approaches • Compares, contrasts, groups different approaches • Analysis from multiple perspectives

… End of Project Description..!