CSC 343 H5S

Read all instructions before starting the report

This Final Report document includes 3 pages and one question (i.e. the Report question).

Deadline: Friday April 9th, 2021 by 11:59PM EST
NO EXTENSIONS WILL BE GRANTED.

Instructions:

1. This report can be completed in a pair of two, or alone. Students who choose to form a partnership must do so on MarkUs, those who fail to accept the partnership request by the due date will receive a grade of 0. Partners will be assigned the same grade regardless of contribution. It is up to you to pick a responsible partner with the same academic goals/values. If your partner drops the course, becomes ill, or is otherwise unavailable this does not relieve you from this report. Furthermore, each student is responsible for the integrity of their submission and must make sure that credit is given where it is due.
2. This is an open-book report; however, you **cannot** communicate with anyone except for the Course Staff and your partner. Communication is limited to the Course Instructors via email and Course Staff via **private** Piazza Post. Any form of alternative communication will be considered a case of academic dishonesty – this includes public posts on Piazza, external discussion boards/rooms, Chegg, Stack Overflow, Stack Exchange, etc...
3. All students are required to complete the MCS Academic Integrity Declaration Form. Students are bound by the University of Toronto's Code of Behaviour on Academic Matters.
4. The report will have a 5-page maximum. This includes any illustrations, diagrams, etc... This excludes references (i.e. references can be on separate pages). Additionally, the report must adhere to all formatting standards set by the `skeleton.tex` file. Further, references are to be formated using ACM's Citation Style and Reference Formats.
5. You are responsible for ensuring that you submit all the proper files to the Course Instructors on MarkUs by the deadline of your final report (Friday April 9th, 2021 by 11:59PM EST). All documents submitted must have both partner's names, student numbers, and UtorIDs.
6. The Final Report will be submitted on MarkUs. You have been provided skeleton code that must be used. Your submission must be typed in LaTeX. The `PDF` and all source files (`.tex`, `.bib`, `.png`, etc.) must be uploaded to MarkUs. Exact naming conventions will be specified there. Reports that are not typed in LaTeX will be heavily penalized; those hand-written will not be marked and awarded a grade of 0.

category	mark	out of
content		70
g.s.p.w.[a]		15
references		15
total		100

[a] grammar, spelling, punctuation, writing

Final Report Question [100 marks]

You have been selected by Professors Dema and Liut to aid them in an investigative project in the field of Databases and Data Centres. You will write a report. Before diving into the nitty-gritty, you must provide an introduction to the space. Specifically, you are asked to introduce and define the relationship between Databases and Data Centres. Additionally, you must find at least 2 of the industrial *big players* (companies, not DBMSs, e.g., Amazon, Google, IBM, Microsoft, Oracle, etc.) in these two spaces and discuss their contribution to the field as part of the literature review.

You have been tasked to take your CSC343 knowledge and apply it to the following areas:

i. Efficiency
ii. Scalability
iii. Social Impact

Your report must include each of the three areas above in a deep form of comparison between, and within, Databases and Data Centres. To help you get started, Professors Dema and Liut have suggested some areas that you can use to get started with each category, but you are not limited by these suggestions. Note: `DB` represents a Database suggestion and `DC` represents a Data Centre suggestion.

Efficiency

DB: Storage versus redundancy (e.g., having more storage versus the amount of redundancy), and the trade-offs that exist within this space.

DB: Power/CPU utilization versus query optimization, and the trade-offs that exist within this space.

DC: Power consumption, specifically the tension between performance and cost.

DC: How does sustainability get a say in the geographical location and how can the source of energy used affect the planet.

DB/DC: How does an individual's choice of using databases/cloud storage directly impact the environment. How does this relate to the broader topic of sustainability.

Scalability

DB: How does the choice of a DBMS affect the scalability of industrial solutions. How can foundational database understanding impact this.

DC: Horizontal versus vertical scaling and its impact economically, environmentally, etc.

- Horizontal Scaling (Sharding): using existing data centres and offloading portions of it to other locations rather than dismantling the data centre and upgrading it.
- Vertical Scaling (Upgrading): upgrading the data centre by gutting its components, wiring, and redesigning/reconstructing the space. Even relocation of the data centre.

Social Impact

DB: How our choices to use software services that fundamentally rely on databases impact our lives and the world around us.

DC: The economic and social impact of dismantling and relocating entire data centres. Expand on the trade-offs that exist.

DB/DC: The cost of sustainability versus economic versus evolutionary and factors that are (versus should be) prioritized.

As academics, your Professors are sticklers for references, so you must ensure that your report contains substantial citations from reputable sources. Remember, you are required to draw your own conclusions and discuss these topics but should not copy (nor be using substantial excerpts) from your sources, rather use them to form your justifications. There is no binary solution, this is an investigative report that requires you to use references to justify your answers and challenge the existing ideas in the field. Think critically, your Professors are expecting you to offer your own insights into these topics, challenge existing views that your see in literature, ask questions, and to draw meaningful conclusions. Your justifications must use sound arguments and reputable sources.

Reports without valid content, justification, or sources, will receive a grade of 0. Reports with insufficient content will be penalized accordingly.

To help you with this process your Professors have provided you with a few supporting links:

- UTSC's "Writing Research Essays", Grammarly's "How to Write a Research Paper", Wisconsin-Madison's "Writing a Research Paper", and St. George's "INI104: Writing Reports"
- LaTeX online collaboration tool: Overleaf

Also, some references to get you started:

- Sustainability: Global Reporting Initiative (GRI) and the GRI Sustainability Disclosure Database
- How much energy do data centers use?, David Mytton.
  Note: this is a blog that is well cited, the references herein are worth investigating further.
- Microsoft's Underwater Data Centres, John Roach.
- Data Center Energy Consumption Modeling: A Survey, Dayarathna et al.
- Analyzing the Energy Efficiency of a Database Server, Tsirogiannis et al.
- Accordion: Elastic Scalability for Database Systems Supporting Distributed Transactions, Serafini et al.
- Data centre sustainability – Beyond energy efficiency, Flucker et al.
- Slicer: Auto-Sharding for Datacenter Applications, Adya et al.
- F1 - The Fault-Tolerant Distributed RDBMS Supporting Google's Ad Business, Shute et al.
- Sharding the Shards: Managing Datastore Locality at Scale with Akkio, Annamalai et al.
- For additional references, consider using: UofT's Library Search