

TERM PROJECT

VERSION 1.0

DUE DATES:

- [TP-D0] Friday, February 23rd, 2024, @8:00 pm [Team composition]
- [TP-D1] Friday, March 1st, 2024, @ 8:00 pm [Term Project Presentation Slides]
- [TP-D2] Wednesday April 24th, 2024 @ 8:00 pm [Software Submission]
- [TP-D3] Friday, April 26th, 2024 @ 8:00 pm [Report]
- [TP-D4] Friday, April 26th, 2024 @ 8:00 pm [PowerPoint Presentation]

OBJECTIVE

The objective of this assignment is to perform analytics at scale over spatial datasets (i.e., datasets that have a spatial component to them).

This assignment may be modified to clarify any questions (and the version number incremented), but the crux of the assignment and the distribution of points will not change.

Grading: This assignment will account for **10 points** towards your cumulative course grade. There are several components to this assignment, and the points-breakdown is listed in the remainder of the text. This assignment is to be done as part of a team. The scoring process will involve a one-to-one interview session where you will demonstrate the functionality of your project. You are required to make two presentations and reports as part of this effort. The slots for these interview sessions and presentations will be posted a few days prior to the submission deadline.

TEAM COMPOSITION: The term project is a team effort. CS555 students are restricted to be in teams of 2, CS455 students can be in teams of 2-3.

1 Requirements

As part of your term project you are required to design a system that solves a problem. The requirements are the following:

1. Your project should involve distributed analytics over datasets that have a spatial component to them. Spatial datasets are available in a diverse set of domains (census, ecology, climate, topographical, and atmospheric among others) and they now account for a large percentage of the data that are being generated. You are free to use any spatial dataset that you would like. A helpful resource is the Sustain project (<https://urban-sustain.org>) which includes several spatial datasets that are available for download.
 - a. Examples of unacceptable projects include: designing a new cryptographic encryption/decryption algorithm.
2. Your system or application should execute on a minimum of 10 machines. The problem should be data-intensive and/or compute-intensive. You are allowed to use distributed frameworks such as Hadoop, Spark, TensorFlow, PyTorch, etc.
 - a. Examples of unacceptable projects include: Using MPI, Fortran, or shared-memory based systems to solve the problem.
3. Either the problem or the solution you propose has to be original. Examples of unacceptable projects include:
 - a. Assignments from Distributed Systems (x55 series at CSU) or Big Data (x35 series at CSU) courses at CSU, MOOCS or other universities.
 - b. Implementation of existing systems: You cannot, for example, say you will be implementing Pastry, BitTorrent, etc.
 - c. Term projects from previous courses that you have taken (or are currently taking). The submission has to be an original submission that you have done specifically for CS555.

2 Third-party libraries and restrictions:

You are allowed to use libraries from the Hadoop, Spark, TensorFlow, and PyTorch ecosystem. Once you have chosen the libraries that you will use, you are responsible for coping with issues that you encounter with them i.e. there is no assistance for debugging why your PyTorch cluster is not working the way it should.

3 Deliverables

TP-DO: DELIVERABLE ZERO [0 points]

This deliverable requires submission of the composition of your group. Any problems that you are having with your non-performing or non-responsive teammate should be reported 2-3 weeks after Deliverable Zero. Timely intervention will allow us to resolve problems before they fester.

TP-D1: TERM PROJECT PITCH [1 point]: In class presentation (or a 5 minute-video) of your proposed term project.

The **PRESENTATION** is restricted to be for 5 minutes. You should have **exactly** 5 slides for your presentation and the content of these slides is tightly regulated to ensure that presenters stay on topic during the presentation.

- Slide1: Project Title and names of all the team members
- Slide 2: Problem characterization
 - This is a technical description of the problem. Your audience is your peers so express it in a way that they can appreciate.
- Slide 3: Why the problem is important
- Slide 4: Proposed solution and implementation strategy
 - Methodology summary and the libraries that you are planning to use.
- Slide 5: Evaluating the effectiveness of your solution
 - What are numbers that will use to assess how good your solution is? Examples of these include: accuracy, turnaround times, throughputs, number of false positive or false negatives, mean squared errors, etc.

TP-D2: CODE SUBMISSION & DEMO [4 points]

The software must be ready for demonstration of project functionality. This will also include a one-on-one interview session. The interview schedule will be posted once we get closer to the submission deadline.

TP-D3: FINAL REPORT FOR TERM PROJECT [3 points]

Please see the next section on the required elements in the Term Project report.

TP-D4: IN-CLASS PRESENTATION [2 points]: The regulated content and number of slides in your presentation has been posted on the course schedule page.

4 Anatomy of the Term Project Report

The term project report *must include* several elements, each of which will be a separate section. These include:

- Introduction
- Problem characterization
- Dominant approaches to the problem
- Methodology
- Experimental Benchmarks
- Insights Gleaned
- How the problem space will look like in the future
- Conclusions
- Bibliography

There are several pitfalls that you must avoid when you are writing technical articles. Avoid cringe-inducing marketing lingo and hearsay *e.g.*, "My teammate Tony Stark thinks ...". Quoting Professors and researchers in the University is not allowed. You are allowed to speculate, but these should be based on reasoned arguments. Avoid using words that are not part of your normal vocabulary – it is easy to know if someone had the thesaurus handy. Technical writing is meant to be clear while being accessible to those in the area.

Word Counts: The word counts set aside for each element of your term paper are specified below. Please do not try to skew the word limits for these sections so that you can reach the requisite word counts. Such skews are easy to spot and will be penalized. You are also not allowed to quote from cited papers just to pad the word counts.

4.1 Introduction

This section describes why the problem is important, where this research is being used, how this technology plays a role in our daily lives, etc. The introduction section is also a concise summary of your paper that outlines the rationale, organization, and key contributions of the term paper. It should be possible for a reader to know all the key aspects of your term paper just by reading your Introduction and Conclusions section.

You can also briefly inform this section with your past experience. Describe how you think your chosen area would be applicable to a project that you are working on or have worked on in the past.

Word count: 400 words

4.2 Problem characterization

This is a technical description of the problem. Your audience is your peers so express it in a way that they can understand and appreciate.

This section should describe the theoretical, physical, social, and/or engineering aspects that make the problem particularly challenging. A clear discussion of the challenge also makes the reader look forward to reading the remainder of the paper.

Word count: 400 words

4.3 Dominant approaches to the problem

The section must contrast and identify possible approaches and also identify inefficiencies in each of these schemes. For each work that you cite you need to describe the advantages, disadvantages, and the scope of the work. Your objective is not to defend any work, rather you should let the facts speak for themselves. Finally, in your write-up you need to describe each reference in the context of the overall narrative.

If you compare features across two systems and say that one system outperforms the other include a citation for this. List what the comparison point is. This could be latency, throughput, scaling, efficiency, accuracy, price, etc.

Citations have a specific purpose. They: (1) relate to work that you are surveying, (2) substantiate your claims, and (3) could be used by readers to delve deeper. Remember to number your references and list them in your bibliography in the order that you referenced it. If an article is in your bibliography it must be cited in the main text. Citing at the right location indicates what your source is for a particular piece of information, and also demonstrates that you have read the article. Make sure that you cite all your references including Wikipedia and online lecture notes that you may have perused. References that are not cited should not be in your bibliography.

Word count: 300 words

4.4 Your Methodology

Describe your methodology. This includes the tasks you have performed, the approach that you have taken, and the justification for your methodology. For example if your approach involves fitting models to the data, describe the rationale for your choice of the model fitting algorithm.

Word count: 900 words

4.5 Experimental Benchmarks

Include a description of the benchmarks that you performed. Your performance metrics must be amenable to quantitatively assessing the quality of your solution. Examples of these include: accuracy, turnaround times, throughputs, number of false positive or false negatives, mean squared errors, Area under the curve for the receiver operating characteristic, etc.

Word count: 400 words

4.6 Insights Gleaned

These are things that you *did not* know before you started this project. The best solutions are the ones that you may have not thought of, but seem incredibly obvious once you have come up with them.

Word count: 400 words

4.7 How will the problem space transform in the future

This is a thought experiment. You will be looking ahead and visualizing a future where there could be proliferation of certain types of devices, new types of services, changes in usage patterns, etc. You must describe the forces that you think will drive this change. Once you have these forces in place, identifying how the problem space will evolve in the future should be easier. Ultimately, you are describing what technology advancements and the way we interact with services will affect the problem space of your research area.

Word count: 300 words

4.8 Conclusions

A conclusion is not a summary. You must make a set of assertions about your work.

Word count: 400 words

4.9 Bibliography

The final term paper must have at least 8-10 references. All references must be cited in the paper. Citations must be numbered and sorted in the order that they appear. **The Bibliography is NOT included in your word count.**

4.10 Overall Word Count

Excluding the bibliography, the total word count for your term project report is 3500 words. Please do not exceed, or be under, the **total** word count limit by more than 200 words.

5 Submission deadline:

Please submit all deliverables by 8:00 pm on the due date using **CANVAS**. We will rely on the honor system: please do not make any modifications after the submission deadline has elapsed. There will be steep deductions for making modifications after you have submitted (especially, your source codes). You may resubmit within the 2 day late-submission window – this will result in the usual penalty

Nota Bene: Please do not e-mail the source codes to the Professor or the GTAs – there will be a **2 point deduction** for doing this.

6 Change History

Version	Date	Change
1.0	2/16/2024	First public release of the document describing the term project.