Vrije Universiteit Amsterdam
https://www.overleaf.com/read/mbzddrhpdbvr#e96bec

Honours Programme, Problem Statement

Precision Capacity Planning: Simulating the Cost-Energy Trade-offs of LLM Training on Homogeneous Clusters

---

Author: Thijs van den Heuvel (2849429)

1st supervisor: Prof. Dr. Ir. Cav. Alexandru Iosup
daily supervisor: MSc. Dante Niewenhuis

A report submitted in fulfillment of the requirements for the Honours Programme, which is an excellence annotation to the VU Bachelor of Science degree in Computer Science/Artificial Intelligence/Information Sciences
version 1.0

December 8, 2025

Abstract

The training of Large Language Models (LLMs) has evolved into a massive industrial operation hello asdf asdf asdf asdlfjasdlkf jklasdf jklas dfjklas djfklas jfklas jfklasd jf

1 intro

2 Introduction

Explain the research project. Also include here the personal value you hope to derive from this project.

Explain at least:

1. The context of this research project. How broad do you see the impact of a good result? (Will you change the world? The science of Europe? The industry of the Netherlands?)
2. The key terms addressed in this research project. You will expand on this element in Section 3.
3. The main problem addressed in this research project. You will expand on this element in Section 4.
4. The key prior work related to this research project. You will expand on this element in Section 5.
5. The main research question, possibly paraphrased. You will expand on this element in Section 6. (If possible, also indicate the core of the approach, or an insight that can lead to it. You will expand on this element in Section 7.)
6. The expected contribution of this research, for the scientific community and/or for your employer. You will expand on this element in Sections 6, 7, and 8.
7. Expected contribution of this research, for yourself. How will this project develop you? How will it develop your career?

For example, consider the project leading to publication [**?**]:

1. Context: datacenters, the backbone of cloud computing and our digital economy.
2. Key terms: datacenters, scheduling, reference architecture.
3. Problem: understanding and improving the process of scheduling in datacenters.
4. Key prior work: research on scheduling in large-scale systems, scheduling practices in Big Tech companies (Google, Microsoft, Alibaba, etc.)
5. Main research question: How to design a good abstraction for datacenter scheduling? Key insight: a unified reference architecture is a good abstraction for the scheduling process.

6. Expected contribution, community: a survey, a reference architecture, an analysis of existing systems as mapped to the new reference architecture, a simulator implementing the reference architecture as the scientific instrument, experiments in simulation, description of a process for others to use the reference architecture, analysis of threats to validity. Plus: a technical report accompanying the publication[1], various public talks, etc. (The team also went for and obtained the ACM reproducibility badge, which among others requires publishing FOS software and FAIR data.)

7. Expected contribution, personal: development into an independent researcher.

3 Background

Explain the key concepts needed to understand this work. See also Section II of [**?**].

4 Problem

Explain in this section the main problem addressed in this work. The goal is to emphasize the value of a research project that addresses the problem. See also Sections I and III.A of [**?**].

Notes:

1. Define the scope of the problem.

2. Refer back to the background (see Section 3) for key terms.

5 Related Work

Explain in this section related work on the problem explained in Section 4. The goal is to emphasize the extent and the key elements of related work. See also Sections I and VII of [**?**].

Notes:

1. At this stage of your research career, this part will include a brief survey of the state-of-the-art, guided by the project supervisor.

2. Review and summarize the related work. What is known already? What should be known but isn't?

6 Research Question(s)

Explain in this section the core research of the project. The goal is to show that the research is sufficiently balanced and broad. See also Sections I and the short formulations (e.g., "we investigate...") in the following sections of [**?**].

Notes:

[1]The technical report is published as open science: `https://arxiv.org/pdf/1808.04224.pdf`

1. Formulate the main research question.

2. Define the scope of the project. Typically, the scope of the project is much smaller than the scope of the problem (defined in Section 4).

3. Define detailed research questions. For each, explain at least: *Why?*, *Why important?*, and *Why challenging?*

7 Approach

Explain in this section how you anticipate you can answer the question(s) formulated in Section 6. The goal is to show that the research is feasible. For this reason, this section is mainly methodological; the pragmatic plans on how to complete all this work follow, in Section 8. See, for example, Sections I (overview) and V.A (experiment design) of [**?**].

Notes:

1. Describe the approach, for each research question. Emphasis on method(s) – What? Expected contribution.

2. Introduce intuition about the key innovation and/or conceptual contribution.

3. Try to explain why the approach would work. Explain the expected technical contribution.

8 Plan

Explain in this section how you expect to complete the parts defined in Section 7. The goal is to show the work is feasible in the allocated time.

Notes: Notes:

1. Understand this is a preliminary plan.

2. Try to define at least the large components of the project. To do this, discuss with the project supervisor and/or consult a good article published recently in the field. For the running example, consult [**?**].

3. Try to plan tasks with a granularity of at most one week, and ideally with a granularity of a day. Try to make the near-future tasks SMART. Plan tasks long into the future of the project as *slack*.

4. Try to attach milestones and key deliverables to the most important tasks. Make sure deliverables include the final report (or article) and at least one presentation (hopefully, in a major scientific venue).

5. Revisit the plans as soon as you complete a task, but especially after the first few tasks of a kind, e.g., a literature review task (you read a new article), a design iteration (you made or improved a design), an implementation task (you coded a new feature), an experiment task (you conducted one experiment).

For the running example, the research plan included:

```
‘‘‘
I plan to take the first two research questions in one step, since
they are closely related:

To build a representative abstraction, I need to survey the
existing approaches in the field. This way, the validation step
is combined with the design step. This combined stage I
intend to work on in the coming three months, and
have a first report on my results ready by late January 2017.

After this stage is completed, I will begin integrating it in the
OpenDC project [n.b., the simulator].
Because I can imagine that this step will take a
substantial amount of time,  I plan to have produced a first,
full prototype of this integration by May 2017.

I will try to keep the paper writing process parallel to
these two stages as much as possible. However, knowing that
this is difficult, I am allocating the time from June to
July of 2017 to tie together the pieces and get
this paper ready for publication.
‘‘‘
```

9 Conclusion

Revisit the context, problem statement, related work, and research design. See, for example, Section VIII of [**?**].

References