



Degree Project in the Field of Technology Computer Science and the Main Field of Study
Data Science

Second cycle, 30 credits

Accelerating Feature Store performance with Apache Iceberg

A Python-based alternative delivering speed and scalability without
Spark dependency

SEBASTIANO MENEGHIN

Accelerating Feature Store performance with Apache Iceberg

**A Python-based alternative delivering speed and
scalability without Spark dependency**

SEBASTIANO MENECHIN

Master's Programme, ICT Innovation, 120 credits

Date: January 22, 2025

Supervisors: Sina Sheikholeslami, Fabian Schmidt, Jim Dowling

Examiner: Vladimir Vlassov

School of Electrical Engineering and Computer Science

Host company: Hopsworks AB

Swedish title: Snabbare prestanda för Feature Store med Apache Iceberg

Swedish subtitle: Ett Python-baserat alternativ som ger snabbhet och skalbarhet
utan Spark-beroende

Abstract

Keep in mind that most of your potential readers are only going to read your title and abstract. This is why the abstract must give them enough information so that they can decide if this document is relevant to them or not. Otherwise, the likely default choice is to ignore the rest of your document.

An abstract should stand on its own, i.e., no citations, cross-references to the body of the document, acronyms must be spelled out,

Write this early and revise as necessary. This will help keep you focused on what you are trying to do.

Write an abstract that is about 250 and 350 words (1/2 A4-page) with the following components:

- What is the topic area?
- (optional) Introduces the subject area for the project.
- Short problem statement
- Why was this problem worth a Master's thesis project? (*i.e.*, why is the problem both significant and of a suitable degree of difficulty for Master's thesis project? Why has no one else solved it yet?)
- How did you solve the problem? What was your method/insight?
- Results/Conclusions/Consequences/Impact: What are your key results/conclusions? What will others do based on your results? What can be done now that you have finished - that could not be done before your thesis project was completed?

Keywords

Canvas Learning Management System, Docker containers, Performance tuning

Formatting the keywords:

- The first letter of a keyword should be set with a capital letter and proper names should be capitalized as usual.
- Spell out acronyms and abbreviations.

- Avoid "stop words" - as they generally carry little or no information.
- List your keywords separated by commas (",").

Sammanfattning

Inside the following scontents environment, you cannot use a `\includefilename` as it will not end up in the for diva information. Additionally, you should not use a straight double quote character in the abstracts or keywords, use two single quote characters instead.

Enter your Swedish abstract or summary here!

If you are writing your thesis in English, you can leave this until the draft version that goes to your opponent for the written opposition. In this way, you can provide the English and Swedish abstract/summary information that can be used in the announcement for your oral presentation.

If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.

Do not use the `\glspl{}` command in an abstract that is not in English, as my programs do not know how to generate plurals in other languages. Instead, you will need to spell these terms out or give the proper plural form. In fact, it is a good idea not to use the glossary commands at all in an abstract/summary in a language other than the language used in the `acronyms.tex` file - since the glossary package does **not** support use of more than one language.

Nyckelord

Canvas Lärplattform, Dockerbehållare, Prestandajustering

If you are an exchange student, use the relevant language or languages for abstracts for your home university, as this will often avoid the need for writing another thesis for your home university.

If you are fluent in other languages, feel free to add the abstracts in one or more of them.

Note that you may need to augment the set of languages used in `polyglossia` or `babel` (see the file `kththesis.cls`). The following languages include those languages that were used in theses at KTH in 2018-2019, except for one in Chinese.

Remove those versions of abstracts that you do not need.

If you add a new language, when specifying the language for the abstract, use the three-letter ISO 639-2 Code – specifically the "B" (bibliographic) variant of these codes (note that this is the same language code used in DiVA).

Sommarior

Sommarior in italiano.

Parole Chiave

5-6 parole chiave

Acknowledgments

It is nice to acknowledge the people that have helped you. It is also necessary to acknowledge any special permissions that you have gotten – for example, getting permission from the copyright owner to reproduce a figure. In this case, you should acknowledge them and this permission here and in the figure’s caption.

I would like to thank xxxx for having yyyy. Or in the case of two authors: We would like to thank xxxx for having yyyy.

Stockholm, January 2025
Sebastiano Meneghin

Contents

1	Introduction	1
1.1	Background	2
1.2	Problem	2
1.2.1	Original problem and definition	2
1.2.2	Scientific and engineering issues	3
1.3	Purpose	3
1.4	Goals	3
1.5	Ethics and Sustainability	4
1.6	Research Methodology	4
1.7	Delimitations	4
1.8	Structure of the thesis	5
2	Background	7
2.1	Major background area 1	7
2.1.1	Subarea 1.1	8
2.1.2	Subarea 1.1.2	8
2.1.3	Subarea 1.1.2	9
2.1.4	Link layer Encapsulation	9
2.1.5	IP packet headers	9
2.1.6	Test for accessibility of formulas	9
2.2	Major background area 2	10
2.2.1	WLAN Security	10
2.2.2	Network layer security	10
2.3	Related work area	10
2.3.1	Major related work 1	10
2.3.2	Major related work n	11
2.3.3	Minor related work 1	11
2.3.4	Minor related work n	11
2.4	Summary	11

3	Method	13
3.1	Research Process	15
3.2	Research Paradigm	15
3.3	Data Collection	15
3.3.1	Sampling	16
3.3.2	Sample Size	16
3.3.3	Target Population	16
3.4	Experimental design/Planned Measurements	16
3.4.1	Test environment/test bed/model	16
3.4.2	Hardware/Software to be used	16
3.5	Assessing reliability and validity of the data collected	16
3.5.1	Validity of method	16
3.5.2	Reliability of method	16
3.5.3	Data validity	16
3.5.4	Reliability of data	17
3.6	Planned Data Analysis	17
3.6.1	Data Analysis Technique	17
3.6.2	Software Tools	17
3.7	Evaluation framework	17
3.8	System documentation	17
4	Implementation	19
4.1	Hardware/Software design .../Model/Simulation model & parameters/...	19
4.2	Implementation .../Modeling/Simulation/...	20
4.2.1	Some examples of coding	20
4.2.2	Some examples of figures in tikz	21
4.2.2.1	Azure's Form Recognizer	21
4.2.2.2	Hyper-V with Containers	22
4.2.2.3	VM versus Containers	22
5	Results and Analysis	25
5.1	Major results	25
5.2	Reliability Analysis	28
5.3	Validity Analysis	28
6	Discussion	29

7	Conclusions and Future work	31
7.1	Conclusions	31
7.2	Limitations	31
7.3	Future work	32
7.3.1	What has been left undone?	32
7.3.1.1	Cost analysis	32
7.3.1.2	Security	32
7.3.2	Next obvious things to be done	32
7.4	Reflections	33
	References	35
A	Supporting materials	35
B	Something Extra	37
B.1	Just for testing KTH colors	37
C	Main equations	39
C.1	A simple example	39
C.2	An even simpler example	39

List of Figures

2.1	Lots of stars (Inspired by Figure x.y on page z of [xxx])	8
2.2	Ethernet data link layer protocol encapsulated into a IEEE 802.3 MAC packet	9
2.3	IPv4 datagram header	9
2.4	IPv6 datagram header	10
3.1	Research Process	15
4.1	Homepage icon	20
4.2	The processing of key-value extraction from a PDF document using Azure's Form Recognizer	22
4.3	Hyper-V with containers	22
4.4	Virtual machines (VMs) versus Containers	23
5.1	A GNUplot figure	26
5.2	Rust types distribution for the compiler, crates.io, and lib.rs. (percentage) - appears here with the permission of the author - see the thesis at https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akth%3Adiva-332124	28

List of Tables

2.1	xxx characteristics	8
4.1	Configurations tested	19
5.1	Delay measurement statistics	25
5.2	RTT for 4 hosts	26
5.3	Median values of sandwich attributes	27

Listings

4.1	Hello world in C code	20
4.2	Using a python program to access the KTH API to get all of the programs at KTH	20

List of acronyms and abbreviations

This document is incomplete. The external file associated with the glossary ‘acronym’ (which should be called `thesis.acr`) hasn’t been created.

Check the contents of the file `thesis.acn`. If it’s empty, that means you haven’t indexed any of your entries in this glossary (using commands like `\gls` or `\glsadd`) so this list can’t be generated. If the file isn’t empty, the document build process hasn’t been completed.

Try one of the following:

- Add `automake` to your package option list when you load `glossaries-extra.sty`. For example:

```
\usepackage[automake]{glossaries-extra}
```

- Run the external (Lua) application:

```
makeglossaries-lite.lua "thesis"
```

- Run the external (Perl) application:

```
makeglossaries "thesis"
```

Then rerun \LaTeX on this document.

This message will be removed once the problem has been fixed.

The list of acronyms and abbreviations should be in alphabetical order based on the spelling of the acronym or abbreviation.

Chapter 1

Introduction

Often the problem and the problem owner come from industry where a specific solution to a specific problem is desired. This is often “too narrowly” defined and often provides a “too narrow” solution for the result to be interesting from a more general engineering perspective and with “new” experiences as a result. Consider together with the project stakeholders (student, problem owner and academia) how the current problem/proposal could be used to investigate some engineering aspect and whose results could provide new or complementary experience to the engineering community and science.

The experience comes from a question that the thesis tries to answer with previous and other experience, own or modified methods that give a result that can be used to discuss an answer to the research question. This paragraph should thus, in addition to the original “narrow” problem, contain what is to be investigated to create new engineering experience and/or science.

The first paragraph after a heading is not indented, all of the subsequent paragraphs have their first line indented.

This chapter describes the specific problem that this thesis addresses, the context of the problem, the goals of this thesis project, and outlines the structure of the thesis.

Give a general introduction to the area. (Remember to use appropriate references in this and all other sections.)

We use the *bibtex* package to handle our references. We, therefore, use the command `\cite{farshin_make_2019}`. For example, Farshin, *et al.*, described how to improve LLC cache performance in [?] in the context of links running at 200 Gbps.

Use the *glossaries* package to help yourself and your readers. Add the acronyms and abbreviations to `lib/acronyms.tex`. Some examples are shown below:

In this thesis, we will examine the use of **Local Area Networks (LANs)**. In this thesis, we will assume that **LANs** include **Wireless Local Area Networks (WLANs)**, such as **Wireless Fidelity (Wi-Fi)**.

1.1 Background

Present the background for the area. Set the context for your project – so that your reader can understand both your project and this thesis. (Give detailed background information in Chapter 2 - together with related work.) Sometimes it is useful to insert a system diagram here so that the reader knows what are the different elements and their relationship to each other. This also introduces the names/terms/... that you are going to use throughout your thesis (be consistent). This figure will also help you later delimit what you are going to do and what others have done or will do.

As one can find in RFC 1235 [?] multicast is useful for xxxx. A number of different **operating systems (OSes)** have been used in this work, such as the following **OSes**: UNIX, Linux, Windows, etc. The main focus will be on one **OS**, namely Linux.

1.2 Problem

Longer problem statement

If possible, end this section with a question as a problem statement.

1.2.1 Original problem and definition

Some text

1.2.2 Scientific and engineering issues

some text

1.3 Purpose

Distinguish between purpose and goal! Purpose is to change something for the better. In thesis work there are often two aspects to this. On the one hand, the problem owner (the company) wants to have its problem solved for the better, but academia and the engineering community also want to gain new experiences and knowledge. Describe a purpose that satisfies both these aspects. There is also another purpose that may be worth considering and that is that you as a student will graduate and you will have to prove, in your thesis, that you fulfill the degree objectives. These objectives coincide with the course objectives of the thesis course. State the purpose of your thesis and the purpose of your degree project. Describe who benefits and how they benefit if you achieve your goals. Include anticipated ethical, sustainability, social issues, etc. related to your project. (Return to these in your reflections in Section 7.4.)

1.4 Goals

State the goal/goals of this degree project. Distinguish between purpose and goal. The purpose is to bring about a change in something. Objectives are the concrete things to be done to achieve the desired change if possible (purpose).

The goal of this project is XXX. This has been divided into the following three sub-goals:

1. Subgoal 1

för att tillfredsställa problemägaren – industrin?

2. Subgoal 2

för att tillfredsställa ingenjörssamfundet och vetenskapen – akademien)

3. Subgoal 3

eventuellt, för att uppfylla kursmålen – du som student

In addition to presenting the goal(s), you might also state what the deliverables and results of the project are.

1.5 Ethics and Sustainability

1.6 Research Methodology

This is where you specify the overall research strategy or methodology you will use to try to answer the academic question and at the same time solve the original problem. Often you can use “solving the original problem” as a case study of an academic question. You investigate some interesting issue in a “live” situation and gather results and experience from this. Keep in mind that sometimes the company has to stand back in its desire and expectation of the project outcome in favor of new or complementary engineering experience and science (your thesis). It is you as a student who decides and solves the distribution between these two interests but make sure that everyone is informed. Introduce your choice of methodology/methodologies and method/methods – and the reason why you chose them. Contrast them with and explain why you did not choose other methodologies or methods. (The details of the actual methodology and method you have chosen will be given in Chapter 3. Note that in Chapter 3, the focus could be research strategies, data collection, data analysis, and quality assurance.) In this section you should present your philosophical assumption(s), research method(s), and research approach(es).

1.7 Delimitations

Describe the boundary/limits of your thesis project and what you are explicitly not going to do. This will help you bound your efforts – as you have clearly defined what is out of the scope of this thesis project. Explain the delimitations. These are all the things that could affect the study if they were examined and included in the degree project.

1.8 Structure of the thesis

Chapter 2 presents relevant background information about xxx. Chapter 3 presents the methodology and method used to solve the problem. ...

Chapter 2

Background

When you do your literature study, you should have a nearly complete Chapters 1 and 2.

You may also find it convenient to introduce the future work section into your report early – so that you can put things that you think about but decide not to do now into this section.

Note that later you can move things between this future work section and what you have done as you may change your mind about what to do now versus what to put off to future work.

What does a reader (another x student – where x is your study line) need to know to understand your report? What have others already done? (This is the “related work”.) Explain what and how prior work/prior research will be applied on or used in the degree project/work (described in this thesis). Explain why and what is not used in the degree project and give valid reasons for rejecting the work/research.

This chapter provides basic background information about xxx. Additionally, this chapter describes xxx. The chapter also describes related work xxxx.

2.1 Major background area 1

There are xxx characteristics that distinguish yyy from other information and communication technology (ICT) system, as shown in Figure 2.1. Table 2.1 summarizes these characteristics.

Egenskaper

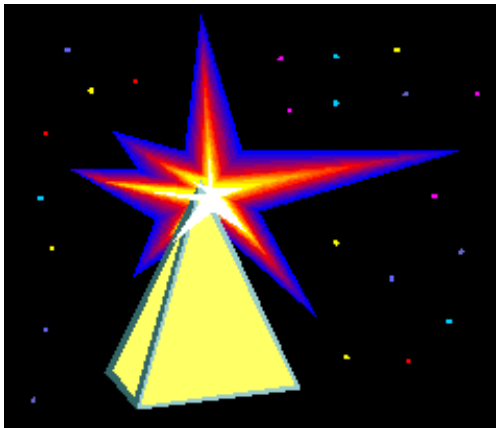


Figure 2.1: Lots of stars (Inspired by Figure x.y on page z of [xxx])

Table 2.1: xxx characteristics

Characteristics	Description
α	β
1	1 110.1
2	10.1
3	23.113 231

Beskrivning

2.1.1 Subarea 1.1

Entangled states are an important part of quantum cryptography, but also relevant in other domains. This concept might be relevant for neutrinos, see for example [?].

2.1.2 Subarea 1.1.2

Computational methods are increasingly used as a third method of carrying out scientific investigations. For example, computational experiments were used to find the amount of wear in a polyethylene liner of a hip prosthesis in [?]. ...

2.1.3 Subarea 1.1.2

Using the nearest data center may improve performance, see [?]

2.1.4 Link layer Encapsulation

See Figure 2.2 which uses the `bytefield` L^AT_EX package.

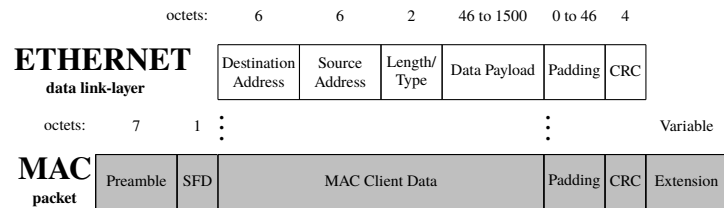


Figure 2.2: Ethernet data link layer protocol encapsulated into a IEEE 802.3 MAC packet

2.1.5 IP packet headers

The data link layer will receive a packet from the IP layer. The layout of an IPv4 packet is shown in Figure 2.3. This should be contrasted with the IPv6 header shown in Figure 2.4.

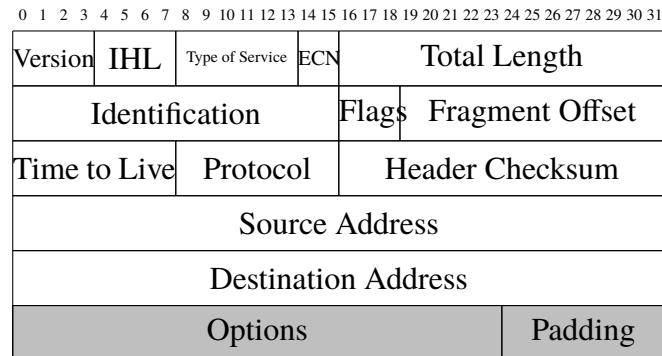


Figure 2.3: IPv4 datagram header. Light grey coloured fields are optional.

2.1.6 Test for accessibility of formulas

As can be seen in these equations: $c = 2 \cdot \pi \cdot r$ or

$$\int_a^b x^2 dx$$

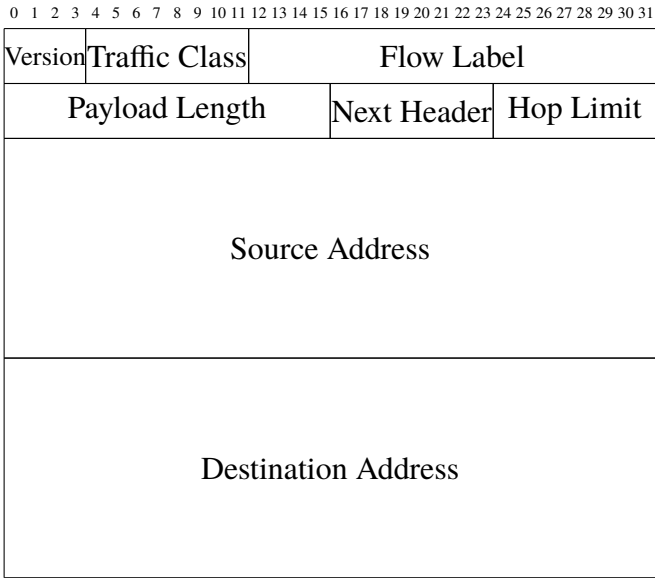


Figure 2.4: IPv6 datagram header

a chemical formula: $(C_5O_2H_8)_n \dots$

2.2 Major background area 2

...

2.2.1 WLAN Security

2.2.2 Network layer security

...

2.3 Related work area

2.3.1 Major related work 1

Carrier clouds have been suggested as a way to reduce the delay between the users and the cloud server that is providing them with content. However, there is a question of how to find the available resources in such a carrier cloud. One approach has been to disseminate resource information using an extension to OSPF-TE, see Roozbeh, Sefidcon, and Maguire [?].

2.3.2 Major related work n

2.3.3 Minor related work 1

...

2.3.4 Minor related work n

2.4 Summary

Sammanfattning

It is nice to have this chapter conclude with a summary. For example, you can include a table that summarizes other people's ideas and benefits and drawbacks with each - so as later you can compare your solution to each of them. This will also help you define the variables that you will use for your evaluation.

Chapter 3

Method

This chapter is about Engineering-related content, Methodologies and Methods. Use a self-explaining title.
The contents and structure of this chapter will change with your choice of methodology and methods.

Describe the engineering-related contents (preferably with models) and the research methodology and methods that are used in the degree project.

Give a theoretical description of the scientific or engineering methodology you are going to use and why have you chosen this method. What other methods did you consider and why did you reject them?

In this chapter, you describe what engineering-related and scientific skills you are going to apply, such as modeling, analyzing, developing, and evaluating engineering-related and scientific content. The choice of these methods should be appropriate for the problem. Additionally, you should be conscious of aspects relating to society and ethics (if applicable). The choices should also reflect your goals and what you (or someone else) should be able to do as a result of your solution - which could not be done well before you started. What scientific or engineering methodology did you use and why did you choose this methodology. What other methods did you consider and why you reject them. What are your objectives (what should you be able to do as a result of your solution - which cannot be done well before you started) What you are going to do? How to do it? Why are you doing it? For example, if you have implemented an artifact what did you do and why? How will you evaluate it.

The purpose of this chapter is to give an overview of the research methodology used in this thesis. Section 3.1 describes the research process. Section 3.2 describes the research paradigm in detail. Section 3.3 focuses on the data collection techniques used for this research. Section 3.4 describes the experimental design. Section 3.5 explains the techniques used to assess the the reliability and validity of the data collected. Section 3.6 describes the methodology used for the data analysis. Finally, Section 3.7 describes the framework chosen to evaluate xxx.

Often a number of follow-up questions can be linked to the research question and problem solving e.g. (1) What process should be used for the construction of the solution and what process should be linked to it to answer the research question? (2) How and what results (quantities) should be presented both to report answers to the research question (results chapter in this report) and to report results of the problem solution (prototype, often documents as appendices but which documents and why?) (3) Which theory/technology should be chosen and used both for the investigation (taxonomy, mathematics, graphs, quantities etc.) and problem solving (UML, UseCases, Java etc.) and why?

(4) What do you as a student need to deliver to achieve high quality (minimum requirements) or very high quality of the thesis?

(5) The questions link to the following subchapters. (6) The reasoning is based on the fact that students in the hing program often have to design something for the problem owner and that this must be linked to an interesting engineering question. There is always a dualism between these aspects in the thesis.

3.1 Research Process

Figure 3.1 shows the steps carried out to implement Describe, preferably with an activity diagram (UML?), your research process and development process. You need to link the academic interest (research process) with the original problem (development process) this research. Activity diagram from e.g. UML standard



Figure 3.1: Research Process

Example of using customized item labels.

Some steps in the process:

- Step 1** plan experiment,
- Step 2** conduct experiment,
- Step 3** analyze data from the experiment, and
- Step 4** discuss the results of the analysis.

3.2 Research Paradigm

For example

Positivistic (what/how does it work?) qualitative case study with a deductive (pre-determined) chosen approach and an inductive (gradually emerging data areas and data) collection of data and experiences.

3.3 Data Collection

This should also show that you are aware of the social and ethical concerns that might be relevant to your data collection method.

3.3.1 Sampling

3.3.2 Sample Size

3.3.3 Target Population

3.4 Experimental design and Planned Measurements

3.4.1 Test environment/test bed/model

Describe everything that someone else would need to reproduce your test environment/test bed/model/... .

3.4.2 Hardware/Software to be used

3.5 Assessing reliability and validity of the data collected

3.5.1 Validity of method

How will you know if your results are valid? Remember that validity is about the *accuracy* of a measurement while reliability is about the *consistency* of the measurement values under the same conditions (*i.e.*, repeatability).

3.5.2 Reliability of method

How will you know if your results are reliable? How good are your methods, are there better methods? How can you improve them?

3.5.3 Data validity

How do you know if your results are valid? Are your results accurate?

3.5.4 Reliability of data

How do you know if your results are reliable? How good are your results?

3.6 Planned Data Analysis

3.6.1 Data Analysis Technique

3.6.2 Software Tools

3.7 Evaluation framework

Method for evaluation, comparison etc. Links to chapters ??

3.8 System documentation

With which documents and how should a constructed prototype be documented? These often become appendices to the report and what the problem owner of the original problem (industry) often wants. These annexes often include, and according to some specified standard, requirements documents, architecture documents, design documents, implementation documents, operational documents, test protocols, etc. If this is going to be a complete document consider putting it in as an appendix, then just put the highlights here.

Chapter 4

Implementation

Choose your own chapter title to describe this

What have you done? How did you do it? What design decisions did you make? How did what you did help you to meet your goals? Choose the appropriate heading (“Implementation”, “Construction”, “Development” or other

4.1 Hardware/Software design .../Model/Simulation model & parameters/...

Figure 4.1 shows a simple icon for a home page. The time to access this page when served will be quantified in a series of experiments. The configurations that have been tested in the test bed are listed in Table 4.1. In 7.0 % of cases, there was an error indicating xxxxx.

Table 4.1: Configurations tested

Configuration	Description
1	Simple test with one server
2	Simple test with one server

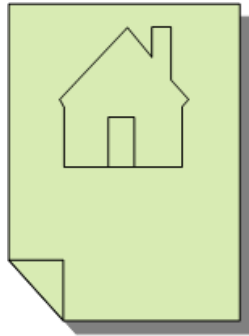


Figure 4.1: Homepage icon

4.2 Implementation .../Modeling/Simulation/...

Two commonly used simulators are:

- Mininet** This simulator uses traffic control (tc) to simulate network devices connected by links with specific bandwidth, packet loss rates, qdisc methods, etc.
- ns-2 or ns-3 simulator** These simulators are very useful for simulating wireless communication links between moving devices. You can specify the mobility patterns of the nodes.

4.2.1 Some examples of coding

This section is simply to show some example of how you can include code in your thesis - this is not a section you would have in your thesis.

Listing 4.1 shows an example of a simple program written in C code.

Listing 4.1: Hello world in C code

```
int main() {
    printf("hello , \world");
    return 0;
}
```

In contrast, Listing 4.2 is an example of code in Python to get a list of all of the programs at KTH.

Listing 4.2: Using a python program to access the KTH API to get all of the programs at KTH

```
KOPPSbaseUrl = 'https://www.kth.se '
```

```
def v1_get_programmes():
    global Verbose_Flag
    #
    # Use the KOPPS API to get the data
    # note that this returns XML
    url = "{0}/api/kopps/v1/programme".format(KOPPSbaseUrl)
    if Verbose_Flag:
        print("url:_" + url)
    #
    r = requests.get(url)
    if Verbose_Flag:
        print("result_of_getting_v1_programme:_" + r.text)
    #
    if r.status_code == requests.codes.ok:
        return r.text          # simply return the XML
    #
    return None
```

4.2.2 Some examples of figures in tikz

This section is simply to show some example of how you can draw your own figures for in your thesis - this is not a section you would have in your thesis.

These figures are just some examples to show that you can draw your own figures for in your thesis. This has two advantages: (i) you do not have to worry about copyrights – as these are your own figures and (ii) the text is now readable and not simply a picture of text – so screen readers can read the figure’s contents to someone who is listening to the contents of your thesis.

4.2.2.1 Azure’s Form Recognizer

Figure 4.2 shows the processing of key-value extraction from a PDF document using Azure’s Form Recognizer.

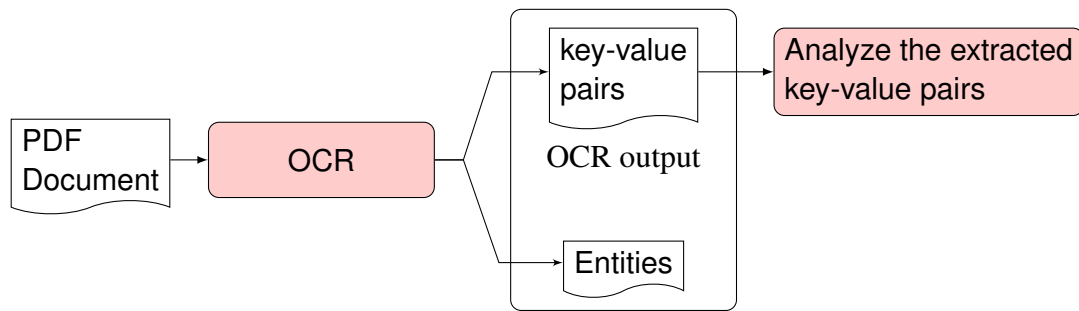


Figure 4.2: The processing of key-value extraction from a PDF document using Azure's Form Recognizer

4.2.2.2 Hyper-V with Containers

Figure 4.3 shows how Hyper-V deals with containers.

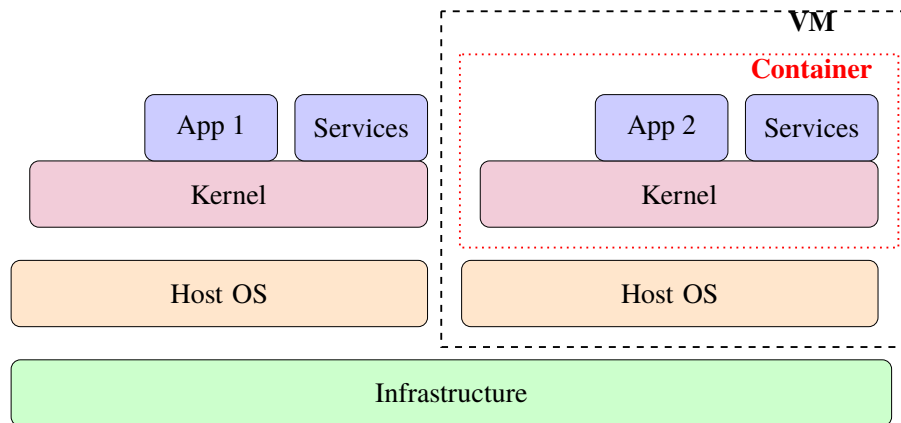


Figure 4.3: Hyper-V with containers

4.2.2.3 VM versus Containers

Figure 4.4 shows a comparison of virtual machines (VMs) versus containers.

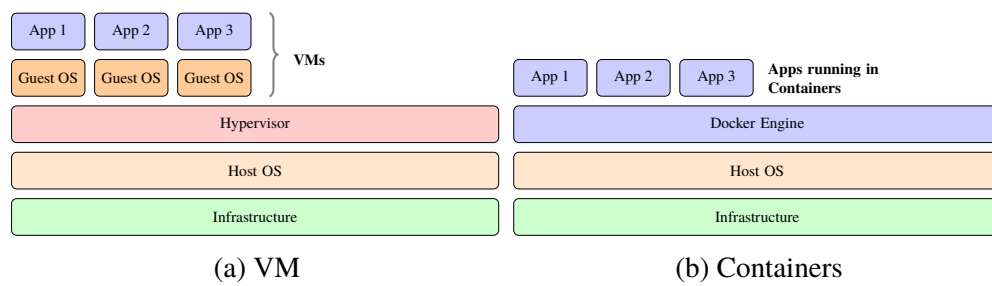


Figure 4.4: Virtual machines (VMs) versus Containers

Chapter 5

Results and Analysis

Sometimes this is split into two chapters.
 Keep in mind: How you are going to evaluate what you have done?
 What are your metrics?
 Analysis of your data and proposed solution
 Does this meet the goals which you had when you started?

In this chapter, we present the results and discuss them.

5.1 Major results

Some statistics of the delay measurements are shown in Table 5.1. The delay has been computed from the time the GET request is received until the response is sent.

Table 5.1: Delay measurement statistics

Configuration	Average delay (ns)	Median delay (ns)
1	467.35	450.10
2	1 687.5	901.23

Table 5.2 shows the measurement of round trip times from four hosts to and from a server.

Table 5.2: Result for the ping measurements of RTT for 4 hosts

Host	host to server RTT in ms			
	min	avg	max	mdev
h1	5.625	5.625	5.625	0.0
h2	2.909	2.909	1.909	0.0
h3	5.007	5.007	5.007	0.0
h4	2.308	2.308	2.308	0.0

Fördröj mätstatistik

Konfiguration | Genomsnittlig fördröjning (ns) | Median fördröjning (ns)

Figure 5.1 shows an example of the performance as measured in the experiments.

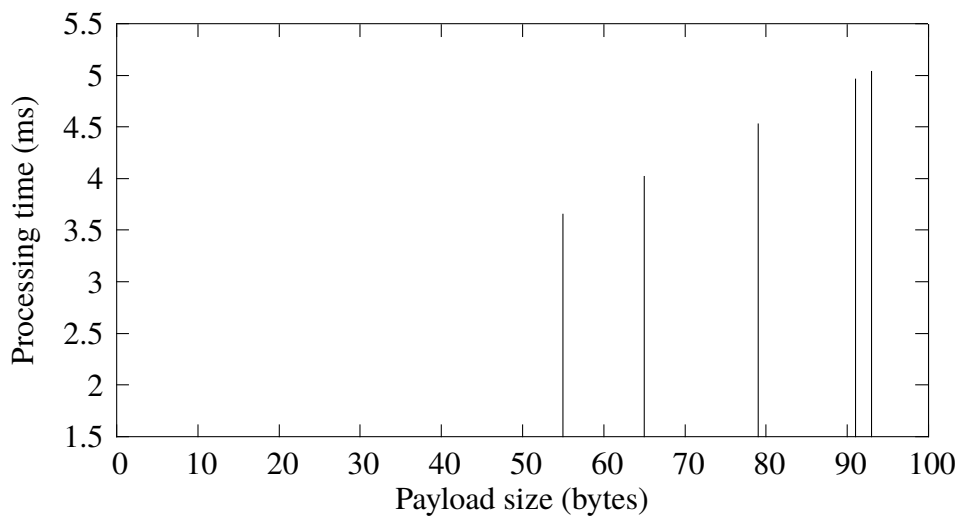


Figure 5.1: Processing time vs. payload length

Given these measurements, we can calculate our processing bit rate as the inverse of the time it takes to process an additional byte divided by 8 bits per byte:

$$\text{bit rate} = \frac{1}{\frac{\text{time}_{\text{byte}}}{8}} = 20.03 \text{ kb/s}$$

Table 5.3 shows another table in which some values have been set in bold (using \B) to emphasize them. Note how the S formatting has been modified so that it considers the weight of the characters and this is able to decimal align even these hold-faced numbers with the numbers in the column above them.

Table 5.3: Median values of sandwich attributes

Attribute	sites	
	A	B
price (in SEK)	36.5	71.3
protean (g)	97.2	100.0
salt (mg)	9.7	9.3
Average customer rating in %	82.2	89.9

Figure 5.2 shows a stacked bar chart using pgfplots. It illustrates how easy it is to take a set of data and make a stacked bar plot. One of the features is the shifted values – this is very useful when the bar itself is too small to put the value into.

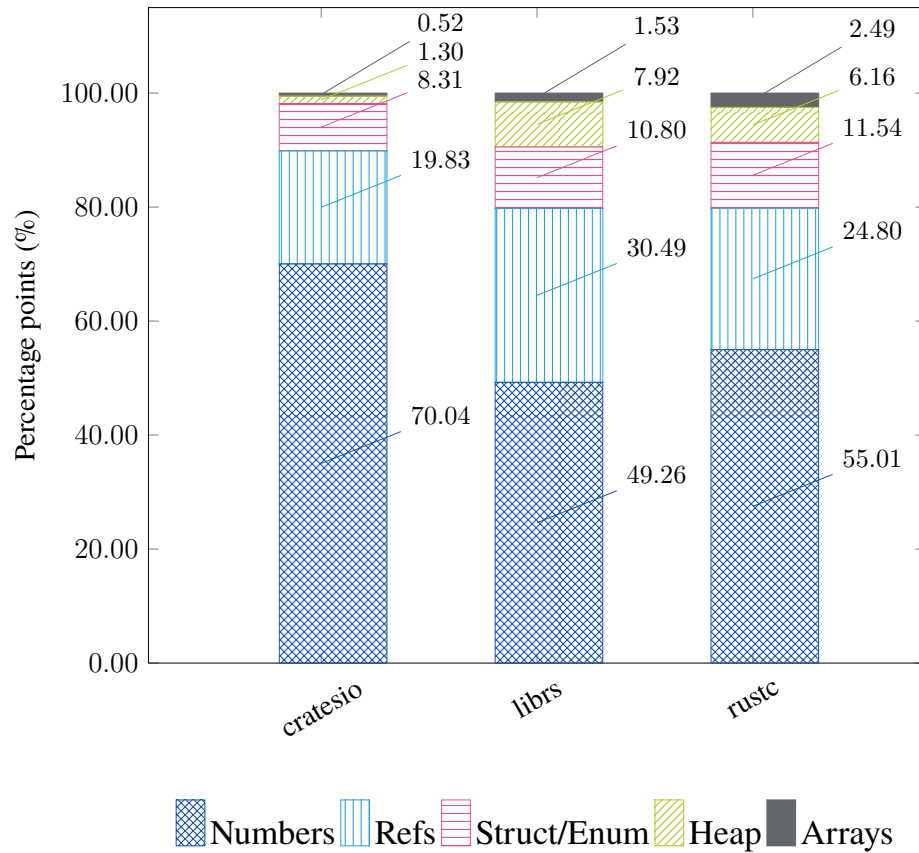


Figure 5.2: Rust types distribution for the compiler, crates.io, and lib.rs. (percentage) - appears here with the permission of the author - see the thesis at <https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akh%3Adiva-332124>

5.2 Reliability Analysis

5.3 Validity Analysis

Chapter 6

Discussion

This can be a separate chapter or a section in the previous chapter.

Chapter 7

Conclusions and Future work

Add text to introduce the subsections of this chapter.

7.1 Conclusions

Describe the conclusions (reflect on the whole introduction given in Chapter 1).

Discuss the positive effects and the drawbacks.
Describe the evaluation of the results of the degree project.
Did you meet your goals?
What insights have you gained?
What suggestions can you give to others working in this area?
If you had it to do again, what would you have done differently?

7.2 Limitations

What did you find that limited your efforts? What are the limitations of your results?

7.3 Future work

Describe valid future work that you or someone else could or should do. Consider: What you have left undone? What are the next obvious things to be done? What hints can you give to the next person who is going to follow up on your work?

Due to the breadth of the problem, only some of the initial goals have been met. In these section we will focus on some of the remaining issues that should be addressed in future work. ...

7.3.1 What has been left undone?

The prototype does not address the third requirment, *i.e.*, a yearly unavailability of less than 3 minutes; this remains an open problem. ...

7.3.1.1 Cost analysis

Example of a missing component

The current prototype works, but the performance from a cost perspective makes this an impractical solution. Future work must reduce the cost of this solution; to do so, a cost analysis needs to first be done. ...

7.3.1.2 Security

Example of a missing component

A future research effort is needed to address the security holes that results from using a self-signed certificate. Page filling text mass. Page filling text mass. ...

7.3.2 Next obvious things to be done

In particular, the author of this thesis wishes to point out xxxxxx remains as a problem to be solved. Solving this problem is the next thing that should be done. ...

7.4 Reflections

What are the relevant economic, social, environmental, and ethical aspects of your work?

One of the most important results is the reduction in the amount of energy required to process each packet while at the same time reducing the time required to process each packet.

The thesis contributes to the **United Nations (UN) Sustainable Development Goals (SDGs)** numbers 1 and 9 by xxxx.


In the references, let Zotero or other tool fill this in for you. I suggest an extended version of the IEEE style, to include URLs, DOIs, ISBNs, etc., to make it easier for your reader to find them. This will make life easier for your opponents and examiner.

IEEE Editorial Style Manual: https://www.ieee.org/content/dam/ieee-org/ieee/web/org/conferences/style_references_manual.pdf

Appendix A

Supporting materials

Here is a place to add supporting material that can help others build upon your work. You can include files as attachments to the PDF file or indirectly via URLs. Alternatively, consider adding supporting material uploaded as separate files in DiVA.

The BibTeX references used in this thesis are attached. 

Some source code relevant to this project can be found at <https://github.com/ggmaquirejr/E-learning> and <https://github.com/ggmaquirejr/Canvas-tools>.

Your reader can access the attached (embedded) files using a PDF tool such as Adobe Acrobat Reader using the paperclip icon in the left menu, as shown in ?? or by right-clicking on the push-pin icon in the PDF file and then using the menu to save the embedded file as shown in ??.

An argument for including supporting material in the PDF file is that it will be available to anyone who has a copy of the PDF file. As a result, they do not have to look elsewhere for this material. This comes at the cost of a larger PDF file. However, the embedded files are encoded into a compressed stream within the PDF file; thus, reducing the number of additional bytes. For example, the references.bib file that was used in this example is 10 617 B in size but only occupies 4 261 B in the PDF file.

DiVA is limited to ≈ 1 GB for each supporting file. If you have very large amounts of supporting material, you will probably want to use one of the data repositories. For additional help about this, contact KTH Library via researchdata@kth.se.

Appendix B

Something Extra

B.1 Just for testing KTH colors

You have selected to optimize for print output

- Primary color

- kth-blue 

- kth-blue80 

- Secondary colors

- kth-lightblue 

- kth-lightred 

- kth-lightred80 

- kth-lightgreen 

- kth-coolgray 

- kth-coolgray80 

black 

Appendix C

Main equations

This appendix gives some examples of equations that are used throughout this thesis.

C.1 A simple example

The following example is adapted from Figure 1 of the documentation for the package `nomencl` (<https://ctan.org/pkg/nomencl>).

$$a = \frac{N}{A} \tag{C.1}$$

The equation $\sigma = ma$ follows easily from Equation (C.1).

C.2 An even simpler example

The formula for the diameter of a circle is shown in Equation (C.2) area of a circle in eq. (C.3).

$$D_{circle} = 2\pi r \tag{C.2}$$

$$A_{circle} = \pi r^2 \tag{C.3}$$

Some more text that refers to (C.3).

€€€€ For DIVA €€€€

```
{
  "Author1": { "Last name": "Meneghin",
    "First name": "Sebastiano",
    "Local User Id": "u16dn7xj",
    "E-mail": "meneghin@kth.se",
    "organisation": { "L1": "School of Electrical Engineering and Computer Science",
    }
  },
  "Cycle": "2",
  "Course code": "DA258X",
  "Credits": "30.0",
  "Degree1": { "Educational program": "Master's Programme, ICT Innovation, 120 credits"
    , "programcode": "TIVNM"
    , "Degree": "Master's Programme, Distributed Systems and Data Mining for Big Data, 120 credits"
    , "subjectArea": "Computer Science"
  },
  "Title": {
    "Main title": "Accelerating Feature Store performance with Apache Iceberg",
    "Subtitle": "A Python-based alternative delivering speed and scalability without Spark dependency",
    "Language": "eng"
  },
  "Alternative title": {
    "Main title": "Snabbare prestanda för Feature Store med Apache Iceberg",
    "Subtitle": "Ett Python-baserat alternativ som ger snabbhet och skalbarhet utan Spark-beroende",
    "Language": "swe"
  },
  },
  "Supervisor1": { "Last name": "Sheikholeslami",
    "First name": "Sina",
    "Local User Id": "u1znylhw",
    "E-mail": "sinash@kth.se",
    "organisation": { "L1": "School of Electrical Engineering and Computer Science",
    "L2": "Computer Science" }
  },
  "Supervisor2": { "Last name": "Schmidt",
    "First name": "Fabian",
    "Local User Id": "u1znylhw",
    "E-mail": "fschm@kth.se",
  },
  "Supervisor3": { "Last name": "Dowling",
    "First name": "Jim",
    "E-mail": "jim@hopsworx.ai",
    "Other organisation": "Hopsworx AB"
  },
  "Examiner1": { "Last name": "Vlassov",
    "First name": "Vladimir",
    "Local User Id": "u19yb2c8",
    "E-mail": "vladv@kth.se",
    "organisation": { "L1": "School of Electrical Engineering and Computer Science",
    "L2": "Computer Science" }
  },
  "Cooperation": { "Partner_name": "Hopsworx AB"},
  "National Subject Categories": "10201, 10205, 10206",
  "Other information": { "Year": "2025", "Number of pages": "1,41"},
  "Copyrightleft": "copyright",
  "Series": { "Title of series": "TRITA – EECS-EX", "No. in series": "2024:0000" },
  "Opponents": { "Name": "A. B. Normal & A. X. E. Normalè"},
  "Presentation": { "Date": "2022-03-15 13:00"
    , "Language": "eng"
    , "Room": "via Zoom https://kth-se.zoom.us/j/ddddddddd"
    , "Address": "Isafjordsgatan 22 (Kistagången 16)"
    , "City": "Stockholm"
  },
  "Number of lang instances": "3",
  "Abstract[eng ]": €€€€
}
```

Write an abstract that is about 250 and 350 words (1/2 A4-page) with the following components:

- What is the topic area?
- (optional) Introduces the subject area for the project.
- Short problem statement
- Why was this problem worth a Master's thesis project? (*i.e.*, why is the problem both significant and of a suitable degree of difficulty for Master's thesis project? Why has no one else solved it yet?)
- How did you solve the problem? What was your method/insight?

- Results/Conclusions/Consequences/Impact: What are your key results/ conclusions? What will others do based on your results? What can be done now that you have finished - that could not be done before your thesis project was completed?

€€€€,
 "Keywords[eng]": €€€€
 Canvas Learning Management System, Docker containers, Performance tuning €€€€,
 "Abstract[swe]": €€€€

Enter your Swedish abstract or summary here!

If you are writing your thesis in English, you can leave this until the draft version that goes to your opponent for the written opposition. In this way, you can provide the English and Swedish abstract/summary information that can be used in the announcement for your oral presentation.

If you are writing your thesis in English, then this section can be a summary targeted at a more general reader. However, if you are writing your thesis in Swedish, then the reverse is true – your abstract should be for your target audience, while an English summary can be written targeted at a more general audience.

Do not use the `\glspl{}` command in an abstract that is not in English, as my programs do not know how to generate plurals in other languages. Instead, you will need to spell these terms out or give the proper plural form. In fact, it is a good idea not to use the glossary commands at all in an abstract/summary in a language other than the language used in the `acronyms.tex` file - since the glossary package does **not** support use of more than one language.

€€€€,
 "Keywords[swe]": €€€€
 Canvas Lärplattform, Dockerbehållare, Prestandajustering €€€€,
 "Abstract[ita]": €€€€
 Sommario in italiano. €€€€,
 "Keywords[ita]": €€€€
 5-6 parole chiave €€€€,
 }

acronyms.tex

```
%%% Local Variables:
%%% mode: latex
%%% TeX-master: t
%%% End:
% The following command is used with glossaries-extra
\setabbreviationstyle{acronym}{long-short}
% The form of the entries in this file is \newacronym{label}{acronym}{phrase}
%                                     or \newacronym[options]{label}{acronym}{phrase}
% see "User Manual for glossaries.sty" for the details about the options, one example is shown below
% note the specification of the long form plural in the line below
\newacronym[longplural={Debugging Information Entities}]{DIE}{DIE}{Debugging Information Entity}
%
% The following example also uses options
\newacronym[shortplural={OSes}, firstplural={operating systems (OSes)}]{OS}{OS}{operating system}

% note the use of a non-breaking dash in long text for the following acronym
\newacronym{IQL}{IQL}{Independent Q28091Learning}

\newacronym{KTH}{KTH}{KTH Royal Institute of Technology}

\newacronym{LAN}{LAN}{Local Area Network}
\newacronym{VM}{VM}{virtual machine}
% note the use of a non-breaking dash in the following acronym
\newacronym{WiFi}{Wi28091Fi}{Wireless Fidelity}

\newacronym{WLAN}{WLAN}{Wireless Local Area Network}
\newacronym{UN}{UN}{United Nations}
\newacronym{SDG}{SDG}{Sustainable Development Goal}
```