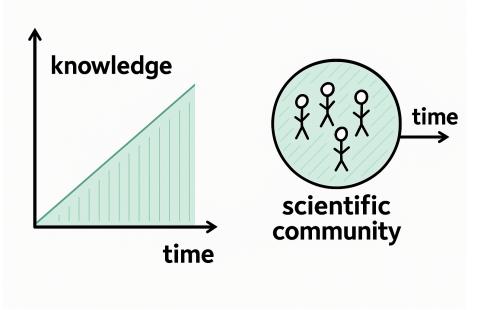
How to Write a Paper

Why do we Write Papers?



1

Reading papers is a critical skill:

• Understand latest developments before they are covered by other media like books

Reading papers is a critical skill:

- Understand latest developments before they are covered by other media like books
- Learn how to write and communicate ideas

Reading papers is a critical skill:

- Understand latest developments before they are covered by other media like books
- Learn how to write and communicate ideas
- Learn to read critically:
 Ask the right questions, challenge assumptions

Reading papers is a critical skill:

- Understand latest developments before they are covered by other media like books
- Learn how to write and communicate ideas
- Learn to read critically:
 Ask the right questions, challenge assumptions
- Synthesize new ideas: inspiration comes more often than not when reading the work of others

• Research question: Central message

• Research question: Central message

• Impact: Motivation, relevance, and impact

- Research question: Central message
- **Impact**: Motivation, relevance, and impact
- **Contributions**: What is new compared to previous work? How applicable is approach in general?

- Research question: Central message
- Impact: Motivation, relevance, and impact
- Contributions: What is new compared to previous work?
 How applicable is approach in general?
- New info for yourself: What can you learn from this paper?
 (e.g. good summary of related work)

- Research question: Central message
- Impact: Motivation, relevance, and impact
- Contributions: What is new compared to previous work?
 How applicable is approach in general?
- New info for yourself: What can you learn from this paper?
 (e.g. good summary of related work)
- Conclusions: Takeaways: Can we build upon this work? If so, how? Ideas for future work?

You should read a paper in three passes:

1. Get general idea (5-10 minutes)

4

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now

You should read a paper in three passes:

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)

4

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now
- 3. Understand in depth (>= 4 hours)

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now
- 3. Understand in depth (>= 4 hours)
 - Fully understand everything, pay attention to all details

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now
- 3. Understand in depth (>= 4 hours)
 - Fully understand everything, pay attention to all details
 - Check related work

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now
- 3. Understand in depth (>= 4 hours)
 - Fully understand everything, pay attention to all details
 - Check related work
 - Imagine your own implementation of the solution

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now
- 3. Understand in depth (>= 4 hours)
 - Fully understand everything, pay attention to all details
 - Check related work
 - Imagine your own implementation of the solution
 - Question everything

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know what the paper is about now
- 2. Understand Content (>= 1 hour)
 - Read full paper, ignore details (e.g. proofs)
 - Find key points, take notes
 - Understand figures
 - Mark references for further reading
 - You should understand the key points now
- 3. Understand in depth (>= 4 hours)
 - Fully understand everything, pay attention to all details
 - Check related work
 - Imagine your own implementation of the solution
 - Question everything
 - Generate new ideas for your own work

1. Get general idea (5-10 minutes)

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes
 - Understand figures

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes
 - Understand figures
 - You should understand the key points now

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes
 - Understand figures
 - You should understand the key points now
- 3. Review Report (>= 20 min)

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes
 - Understand figures
 - You should understand the key points now
- 3. Review Report ($\geq 20 \text{ min}$)
 - Check each of the required points in the text

How to Read for a Review in This Course

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes
 - Understand figures
 - You should understand the key points now
- 3. Review Report (>= 20 min)
 - Check each of the required points in the text
 - Take notes

How to Read for a Review in This Course

- 1. Get general idea (5-10 minutes)
 - Read abstract, introduction, headings and subheadings, and conclusion
 - You should know the main goal of the paper now
- 2. Understand Content (>= 30 min)
 - Read full report
 - Find key points, take notes
 - Understand figures
 - You should understand the key points now
- 3. Review Report (>= 20 min)
 - Check each of the required points in the text
 - Take notes
 - You should have all information for your review now

What You Might Write

• Thesis: Bachelor, Seminar, Master, PhD

What You Might Write

- Thesis: Bachelor, Seminar, Master, PhD
- Research: Research paper, project proposal

What You Might Write

- Thesis: Bachelor, Seminar, Master, PhD
- Research: Research paper, project proposal
- Industry: Documentation, design document, white paper, website

• Writing forces you to make thoughts concrete

7

- Writing forces you to make thoughts concrete
- Writing organizes thoughts

7

- Writing forces you to make thoughts concrete
- Writing organizes thoughts
- Clear writing is explaining

- Writing forces you to make thoughts concrete
- Writing organizes thoughts
- Clear writing is explaining
- Helps you to understand

- Writing forces you to make thoughts concrete
- Writing organizes thoughts
- Clear writing is explaining
- Helps you to understand
- Along the process you will have new ideas

- Writing forces you to make thoughts concrete
- Writing organizes thoughts
- Clear writing is explaining
- Helps you to understand
- Along the process you will have new ideas
- Will even improve your code

- Writing forces you to make thoughts concrete
- Writing organizes thoughts
- Clear writing is explaining
- Helps you to understand
- Along the process you will have new ideas
- Will even improve your code
- Writing is hard

Good Writing is Important

• More and more papers are written and published

Good Writing is Important

- More and more papers are written and published
- Nobody likes badly written papers

Good Writing is Important

- More and more papers are written and published
- Nobody likes badly written papers
- Nobody reads badly written papers

• Papers aren't written from beginning to end in one session.

- Papers aren't written from beginning to end in one session.
- Typical progress

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ..
 - Final paper

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft

 - Final paper
- The earlier the draft, the less time you should invest in polishing

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ...
 - Final paper
- The earlier the draft, the less time you should invest in polishing
- This helps a lot with avoiding writers block

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ...
 - Final paper
- The earlier the draft, the less time you should invest in polishing
- This helps a lot with avoiding writers block
- Think about writing early draft in a simple text editor

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ..
 - Final paper
- The earlier the draft, the less time you should invest in polishing
- This helps a lot with avoiding writers block
- Think about writing early draft in a simple text editor
- Iterations...

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ...
 - Final paper
- The earlier the draft, the less time you should invest in polishing
- This helps a lot with avoiding writers block
- Think about writing early draft in a simple text editor
- Iterations...
 - Require extreme concentration

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ..
 - Final paper
- The earlier the draft, the less time you should invest in polishing
- This helps a lot with avoiding writers block
- Think about writing early draft in a simple text editor
- Iterations...
 - Require extreme concentration
 - Maybe do it every morning for an hour

- Papers aren't written from beginning to end in one session.
- Typical progress
 - First draft (structure your thoughs)
 - 1. revised draft (e.g., after feedback)
 - 2. revised draft
 - ..
 - Final paper
- The earlier the draft, the less time you should invest in polishing
- This helps a lot with avoiding writers block
- Think about writing early draft in a simple text editor
- Iterations...
 - Require extreme concentration
 - Maybe do it every morning for an hour
- Kill your darlings: don't be afraid of deleting parts that you invested efforts into but don't fit
 anymore

• When reading, it is hard to view your own text form the perspective of a reader

- When reading, it is hard to view your own text form the perspective of a reader
- You can try the following:

- When reading, it is hard to view your own text form the perspective of a reader
- You can try the following:
- Dead Trees: Print your paper and read at another physical location than you write

- When reading, it is hard to view your own text form the perspective of a reader
- You can try the following:
- Dead Trees: Print your paper and read at another physical location than you write
- Random: Don't read from start. Jump to random sections and start revising

- When reading, it is hard to view your own text form the perspective of a reader
- You can try the following:
- Dead Trees: Print your paper and read at another physical location than you write
- Random: Don't read from start. Jump to random sections and start revising
- Read Aloud: Helps catching issues you are otherwise blind to

- When reading, it is hard to view your own text form the perspective of a reader
- You can try the following:
- Dead Trees: Print your paper and read at another physical location than you write
- Random: Don't read from start. Jump to random sections and start revising
- Read Aloud: Helps catching issues you are otherwise blind to
- Still: Feedback from other persons is most valuable!

Title

• The title of a paper is the first impression people have and will use for filtering

- The title of a paper is the first impression people have and will use for filtering
- Should be catchy, intriguing, and meaningful

- The title of a paper is the first impression people have and will use for filtering
- Should be catchy, intriguing, and meaningful
- Naming things is very important: (components, algorithms, methods, products, ...)

- The title of a paper is the first impression people have and will use for filtering
- Should be catchy, intriguing, and meaningful
- Naming things is very important: (components, algorithms, methods, products, ...)
 - Makes it easier to talk about things

- The title of a paper is the first impression people have and will use for filtering
- Should be catchy, intriguing, and meaningful
- Naming things is very important: (components, algorithms, methods, products, ...)
 - Makes it easier to talk about things
 - Makes it easier to think about things

- The title of a paper is the first impression people have and will use for filtering
- Should be catchy, intriguing, and meaningful
- Naming things is very important: (components, algorithms, methods, products, ...)
 - Makes it easier to talk about things
 - Makes it easier to think about things
 - Makes it easier to remember things

- The title of a paper is the first impression people have and will use for filtering
- Should be catchy, intriguing, and meaningful
- Naming things is very important: (components, algorithms, methods, products, ...)
 - Makes it easier to talk about things
 - Makes it easier to think about things
 - Makes it easier to remember things
- The title should give the reader an idea what your paper is about

• Readers will only read title and (maybe) abstract to decide whether to read your paper

- Readers will only read title and (maybe) abstract to decide whether to read your paper
- Abstract should summarize the whole paper

- Readers will only read title and (maybe) abstract to decide whether to read your paper
- Abstract should summarize the whole paper
 - What is it about

- Readers will only read title and (maybe) abstract to decide whether to read your paper
- Abstract should summarize the whole paper
 - What is it about
 - Motivation

- Readers will only read title and (maybe) abstract to decide whether to read your paper
- Abstract should summarize the whole paper
 - What is it about
 - Motivation
 - Core ideas, methods, and solutions

- Readers will only read title and (maybe) abstract to decide whether to read your paper
- Abstract should summarize the whole paper
 - What is it about
 - Motivation
 - Core ideas, methods, and solutions
 - Impact, conclusions, and findings

- Readers will only read title and (maybe) abstract to decide whether to read your paper
- Abstract should summarize the whole paper
 - What is it about
 - Motivation
 - Core ideas, methods, and solutions
 - Impact, conclusions, and findings
- Should be as short and expressive as possible

 Your text has to provide value to the reader Why bother writing it otherwise?

- Your text has to provide value to the reader Why bother writing it otherwise?
- You cannot express everything you have to say in one text

- Your text has to provide value to the reader Why bother writing it otherwise?
- You cannot express everything you have to say in one text
- Define one clear goal that your text has to achieve (e.g. answer a research question)

- Your text has to provide value to the reader Why bother writing it otherwise?
- You cannot express everything you have to say in one text
- Define one clear goal that your text has to achieve (e.g. answer a research question)
- **Kill your darlings**: Remove text that does not help to achieve the goal See Chekhov's Gun: (If the gun is there it will be important)

- Your text has to provide value to the reader Why bother writing it otherwise?
- You cannot express everything you have to say in one text
- Define one clear goal that your text has to achieve (e.g. answer a research question)
- **Kill your darlings**: Remove text that does not help to achieve the goal See Chekhov's Gun: (If the gun is there it will be important)
- Know your audience: The text should be optimized for the reader

• It's natural to start writing small pieces of text

- It's natural to start writing small pieces of text
 - You need to figure out how to combine them

- It's natural to start writing small pieces of text
 - You need to figure out how to combine them
 - Makes it hard to get a good overall structure

- It's natural to start writing small pieces of text
 - You need to figure out how to combine them
 - Makes it hard to get a good overall structure
- Plan your paper top down

- It's natural to start writing small pieces of text
 - You need to figure out how to combine them
 - Makes it hard to get a good overall structure
- Plan your paper top down
 - Start with a very rough outline

- It's natural to start writing small pieces of text
 - You need to figure out how to combine them
 - Makes it hard to get a good overall structure
- Plan your paper top down
 - Start with a very rough outline
 - Iteratively specify section contents with increasing detail

- It's natural to start writing small pieces of text
 - You need to figure out how to combine them
 - Makes it hard to get a good overall structure
- Plan your paper top down
 - Start with a very rough outline
 - Iteratively specify section contents with increasing detail
 - Works for whole paper as well as individual sections

• Sometimes you need to explain several independent things to lead to the next point

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail
 - Add several layers: Each layer may go down further into detail

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail
 - Add several layers: Each layer may go down further into detail
- This holds for sections as well as the whole paper:

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail
 - Add several layers: Each layer may go down further into detail
- This holds for sections as well as the whole paper:
 - Title

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail
 - Add several layers: Each layer may go down further into detail
- This holds for sections as well as the whole paper:
 - Title
 - Abstract

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail
 - Add several layers: Each layer may go down further into detail
- This holds for sections as well as the whole paper:
 - Title
 - Abstract
 - Introduction

- Sometimes you need to explain several independent things to lead to the next point
- Use the onion principle:
 - Start with the main point / complete overview but do not go into detail
 - Add several layers: Each layer may go down further into detail
- This holds for sections as well as the whole paper:
 - Title
 - Abstract
 - Introduction
 - Main part (may contain several layers)

Avoid the Wall of Text

• Reader looses attention when reading (or even seeing) long text

Avoid the Wall of Text

- Reader looses attention when reading (or even seeing) long text
- Use subsections

Avoid the Wall of Text

- Reader looses attention when reading (or even seeing) long text
- Use subsections
- Use small titles

Avoid the Wall of Text

- Reader looses attention when reading (or even seeing) long text
- Use subsections
- Use small titles
- Use short paragraphs

Avoid the Wall of Text

- Reader looses attention when reading (or even seeing) long text
- Use subsections
- Use small titles
- Use short paragraphs
- Use bullet points

• Help reader to check whether they understood something

- Help reader to check whether they understood something
- Should be minimal but interesting

- Help reader to check whether they understood something
- Should be minimal but interesting
- May be used in whole paper, also as motivation

- Help reader to check whether they understood something
- Should be minimal but interesting
- May be used in whole paper, also as motivation
- Best case: One simple, but sufficient example used throughout the entire paper

- Help reader to check whether they understood something
- Should be minimal but interesting
- May be used in whole paper, also as motivation
- Best case: One simple, but sufficient example used throughout the entire paper
- Finding good examples is extremly difficult, but worth it!

• When writing about an approach it may be useful to compare against a well-known baseline

- · When writing about an approach it may be useful to compare against a well-known baseline
- Problems of baseline lead to solution

- When writing about an approach it may be useful to compare against a well-known baseline
- Problems of baseline lead to solution
- This might help with the whole story of a paper

- · When writing about an approach it may be useful to compare against a well-known baseline
- Problems of baseline lead to solution
- This might help with the whole story of a paper
- Example:

- When writing about an approach it may be useful to compare against a well-known baseline
- Problems of baseline lead to solution
- This might help with the whole story of a paper
- Example:
 - Linear search on data

- When writing about an approach it may be useful to compare against a well-known baseline
- Problems of baseline lead to solution
- This might help with the whole story of a paper
- Example:
 - Linear search on data
 - Improvement: Sort data + perform binary search

- When writing about an approach it may be useful to compare against a well-known baseline
- Problems of baseline lead to solution
- This might help with the whole story of a paper
- Example:
 - Linear search on data
 - Improvement: Sort data + perform binary search
 - Improvement: Use sorted search-tree

Repeat Important Points

• Readers may read superficially or only parts of the text

Repeat Important Points

- Readers may read superficially or only parts of the text
- Important points should be clear to them too

Repeat Important Points

- Readers may read superficially or only parts of the text
- Important points should be clear to them too
- Repeat them in: (abstract), introduction, main part, conclusion

Why is Good Writing so Difficult?

• Text is linear - Humans think associatively

Why is Good Writing so Difficult?

- Text is linear Humans think associatively
- Curse of knowledge:

Why is Good Writing so Difficult?

- Text is linear Humans think associatively
- Curse of knowledge:
 - \bullet Once you understand something, it is hard to identify with someone who doesn't

• Our lack of knowledge about local conditions precluded determination of committee action effectiveness in fund allocation to those areas in greatest need of assistance.

- Our lack of knowledge about local conditions precluded determination of committee action effectiveness in fund allocation to those areas in greatest need of assistance.
- Because we knew nothing about local conditions, we could not determine how effectively the committee had allocated funds to areas that most needed assistance.

- Our lack of knowledge about local conditions precluded determination of committee action effectiveness in fund allocation to those areas in greatest need of assistance.
- Because we knew nothing about local conditions, we could not determine how effectively the committee had allocated funds to areas that most needed assistance.
- Why is the second version easier to understand?

- Our lack of knowledge about local conditions precluded determination of committee action effectiveness in fund allocation to those areas in greatest need of assistance.
- Because we knew nothing about local conditions, we could not determine how effectively the committee had allocated funds to areas that most needed assistance.
- Why is the second version easier to understand?
- Unclear who does what!

• People think in stories

- People think in stories
- Who does what?

- People think in stories
- Who does what?
 - Clear actors (subjects)

- People think in stories
- Who does what?
 - Clear actors (subjects)
 - Clear actions (verbs)

Actors should be clear: "Knuth developed TeX"

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:
 - "The compiler tells you..." ,

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:
 - "The compiler tells you..."
 - "The community observes..." 🗸

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:
 - "The compiler tells you..."
 - "The community observes..." 🗸
- Avoid hidden actors:

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:
 - "The compiler tells you..." 🔽,
 - "The community observes..." 🔽
- Avoid hidden actors:
 - "In your paper there is an explanation for ..." X

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:
 - "The compiler tells you..." 🔽,
 - "The community observes..." 🔽
- Avoid hidden actors:
 - "In your paper there is an explanation for ..." X

Actors - Subjects

- Actors should be clear: "Knuth developed TeX"
- Things or representations can be actors:
 - "The compiler tells you..." 🔽,
 - "The community observes..." 🔽
- Avoid hidden actors:
 - "In your paper there is an explanation for ..." X
 - "You explain ... in your paper." 🔽

• Avoid nouns that are hidden verbs:

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." 🗶

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." 🗶

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." 🗙
 - "We analyzed the behavior of ..." 🗸

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." 🗙
 - ullet "We analyzed the behavior of ..." ${f V}$
- Avoid nouns that are hidden adjectives:

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." X
 - "We analyzed the behavior of ..." 🔽
- Avoid nouns that are hidden adjectives:
 - "The implementation of this data structure poses difficulties"

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." X
 - "We analyzed the behavior of ..." 🔽
- Avoid nouns that are hidden adjectives:
 - "The implementation of this data structure poses difficulties"

- Avoid nouns that are hidden verbs:
 - "We made an analysis of the behavior of ..." 🗙
 - "We analyzed the behavior of ..."
- Avoid nouns that are hidden adjectives:
 - "The implementation of this data structure poses difficulties"
 - "The implementation of this data structure is difficult"

Use nouns to:

• Refer to the previous sentence:

- Refer to the previous sentence:
 - "These arguments are convincing" V

- Refer to the previous sentence:
 - "These arguments are convincing" V
- Create an object instead of a complicated construct:

- Refer to the previous sentence:
 - These arguments are convincing
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" 🗸

- Refer to the previous sentence:
 - "These arguments are convincing"
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" 🔽
 - "I do not understand either what he means or what she intends" X



- Refer to the previous sentence:
 - "These arguments are convincing"
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" 🗸
 - "I do not understand either what he means or what she intends"
- Identify the correct actor (subject):

- Refer to the previous sentence:
 - "These arguments are convincing" <a>V
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" 🗸
 - "I do not understand either what he means or what she intends"
- Identify the correct actor (subject):
 - "The fact that I set the correct flags was crucial" X

- Refer to the previous sentence:
 - "These arguments are convincing" <a>V
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" 🗸
 - "I do not understand either what he means or what she intends"
- Identify the correct actor (subject):
 - "The fact that I set the correct flags was crucial"
 - "My correct choice of flags was crucial"

- Refer to the previous sentence:
 - "These arguments are convincing" \(\sqrt{} \)
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" V
 - "I do not understand either what he means or what she intends"
- Identify the correct actor (subject):
 - "The fact that I set the correct flags was crucial"
 - "My correct *choice* of flags was crucial"
- Name repeated concepts (create new actors):

- Refer to the previous sentence:
 - "These arguments are convincing" 🔽
- Create an object instead of a complicated construct:
 - "I do not understand either his meaning or her intention" 🗸
 - "I do not understand either what he means or what she intends"
- Identify the correct actor (subject):
 - "The fact that I set the correct flags was crucial" X
 - "My correct choice of flags was crucial"
- Name repeated concepts (create new actors):
 - "Anyblob is a download manager" 🔽

• Be careful with passive voice

- Be careful with passive voice
- Write specific and concrete

- Be careful with passive voice
- Write specific and concrete
- Fewer prepositions

- Be careful with passive voice
- Write specific and concrete
- Fewer prepositions
- Shorter sentences

- Be careful with passive voice
- Write specific and concrete
- Fewer prepositions
- Shorter sentences
- Logical order

- Be careful with passive voice
- Write specific and concrete
- Fewer prepositions
- Shorter sentences
- Logical order
- Clear logical relationships

• Sentences often start with a transition or evaluation:

- Sentences often start with a transition or evaluation:
 - Hence, but, fortunately, importantly

- Sentences often start with a transition or evaluation:
 - Hence, but, fortunately, importantly
- The beginning of a sentence should contain known easy to recognize things

- Sentences often start with a transition or evaluation:
 - Hence, but, fortunately, importantly
- The beginning of a sentence should contain known easy to recognize things
- The end of a sentence should contain newest and significant information

- Sentences often start with a transition or evaluation:
 - Hence, but, fortunately, importantly
- The beginning of a sentence should contain known easy to recognize things
- The end of a sentence should contain newest and significant information
- You automatically stress the end of a sentence

Balance Clarity and Flow

- Sometimes clarity and flow are in conflict
- 1. "A black hole is created by the collapse of a dead star into a point perhaps no larger than a marble."
- 2. "The collapse of a dead star into a point perhaps no larger than a marble creates a black hole."

Balance Clarity and Flow

- Sometimes clarity and flow are in conflict
- 1. "A black hole is created by the collapse of a dead star into a point perhaps no larger than a marble."
- 2. "The collapse of a dead star into a point perhaps no larger than a marble creates a black hole."
- (2) is more clear
- "Some astonishing questions about the nature of the universe have been raised by scientists exploring the nature of black holes in space. (1) or (2) So much matter compressed into so little volume changes the fabric of space around it in profoundly puzzling ways."

Balance Clarity and Flow

- Sometimes clarity and flow are in conflict
- 1. "A black hole is created by the collapse of a dead star into a point perhaps no larger than a marble."
- 2. "The collapse of a dead star into a point perhaps no larger than a marble creates a black hole."
- (2) is more clear
- "Some astonishing questions about the nature of the universe have been raised by scientists exploring the nature of black holes in space. (1) or (2) So much matter compressed into so little volume changes the fabric of space around it in profoundly puzzling ways."
- (1) connects better to the previous sentence: "black hole"
- (1) connects better to the next sentence: "... no larger than a marble." "So much matter compressed"

Be Concise

• Use as few words as possible to express what you mean

Be Concise

- Use as few words as possible to express what you mean
 - "In my personal opinion, we must listen to and think over in a punctilious manner each and every suggestion that is offered to us." X

Be Concise

- Use as few words as possible to express what you mean
 - "In my personal opinion, we must listen to and think over in a punctilious manner each and every suggestion that is offered to us." X
 - "We must consider each suggestion carefully." 🔽



Be Concise

- Use as few words as possible to express what you mean
 - "In my personal opinion, we must listen to and think over in a punctilious manner each and every suggestion that is offered to us."
 - "We must consider each suggestion carefully."
- Another example:

Be Concise

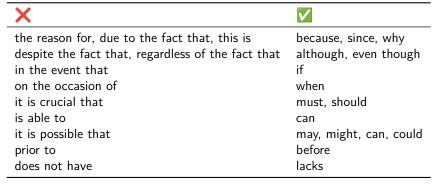
- Use as few words as possible to express what you mean
 - "In my personal opinion, we must listen to and think over in a punctilious manner each and every suggestion that is offered to us."
 - "We must consider each suggestion carefully."
- Another example:
 - "Imagine a picture of someone engaged in the activity of trying to learn the rules for playing the game of chess."

Be Concise

- Use as few words as possible to express what you mean
 - "In my personal opinion, we must listen to and think over in a punctilious manner each and every suggestion that is offered to us."
 - "We must consider each suggestion carefully."
- Another example:
 - "Imagine a picture of someone engaged in the activity of trying to learn the rules for playing the game of chess."
 - "Imagine someone trying to learn the rules of chess."

Avoid Wordy Phrases

Avoid wordy phrases:



• Figures are an integral part of communication



Amazon Redshift and the Case for Simpler Data Warehouses

Anurag Gupta, Deepak Agarwal, Derek Tan, Jakub Kulesza, Rahul Pathak, Stefano Stefani, Vidhya Srinivasan Amazon Web Senvissa

Abstract

Amazou Rodoliti is a fast, fully managod, petalyst-scale data surerbouse solution that makes it simple and cost-efficiency to officiently analyze large volumes of data using existing business intelligence tools. Since launching in February 2011, it has been Amazou Web Service's (AWS) fastest growing service, with many flowassist of customers and many petalysis of data under management.

Amazous Endolsti's pace of adoption has been a surprise to many

participates in the data wardoowing community. While Amazon Rachilit was priced desprisely at least, Available for as little as \$1000TH/year, those are many open-source data wardoowing schoologies and many commercial data wardoowing outputs that provide froe address for development or under some usage fruit. While Amazon Rachilit provides as models and PPC, colorance, While Amazon Rachilit provides as models and PPC, colorance, origines. And, while Amazon Rachilit is variable in the AWS cloud, one can build data wardoowes using ECI intensects and the database capitor of one's choice with either local or networkanched stonger.

In this paper, we discuss an ord-verlooked differentiating dataselemics of Antaron Reddshift - simpley. Our goal with Annans Radshift was more to compete with other data suchowing angains, but so conject with not-consumption. We believe the vast majority of data is collected but not analyzed. We believe, but they are majority of data is collected but not analyzed. We believe the test majority of the date, and the confidence in models' concerned better of the confidence in models' concerned better often or data set time and company size. And, we believe the media used to precess and consume analysis technology need to support experimentation and evaluation. Annance Reddshift was designed to bring data surchoosing to a runn metals by harding it may to by, care your boy, care

1. Introduction

Manufactures are their transaction processing database continues the data superhouse for approxing and analysis. Analysis outrained the data waveformer market segment as 13 of the overall national database market segment as 13 of the overall lecenses and support), with an 8-11% compound annual growth are (CACRE). While this is a strong growth rate for a large, market market, over the past ten years, analysts also estimate data forcage at a pignic enterprise growing at 30-49% CACRE. Over

Permission to make digital or band copies of part or all of this work for presental or climerous not in granted without Fee provided that copies are not made or distributed for profit or communical advantage and that copies both this notice and that illustration on the fore page. Copyrights for thisiputy components of this work must be honored. For all other uses, commer the Owner/Andre Copyright is held by the continuation of all other conceptions. In the Commercial Conference of the Conference of the



Figure 1: Data Analysis Gap in the Enterprise [10]

This implies most data is an enterprise in "dat's data", das that is collected but not easily analyzed, We see this se unfortunate. If our customers defin't see this data as having value, they would not retain it. Many, companies are trying no become increasily distriction. And yet, not only is most data already dath, the overall data landcape; to only petting dathers. Sowing this data in NoSQL, stores and/or Hadoop is one way to bridge the gap for certain use cases. However in doesn't address all scenarios.

In our discussions with customers, we found the "analysis gap' between data being collected and data available for analysis a due to four major causes.

- Cost Most commercial database solutions capable of analyzing data at scale require significant up-front expense. This is hard to justify for large datasets with unclear value.
- This is hard to justify for large datasets with unclear value.

 2. Complexity Database provisioning, maintenance, backup, and tuning are complex tasks requiring specialized skills. They require IT involvement and cannot easily be performed.
- by line of business data scientists or analysts.

 3. Performance It is difficult to grow a data warehous without negatively impacting query performance. Once balls
- IT teams sometimes discourage augmenting data or adding queries as a way of protecting current reporting SLAs.

 Biology Meet discharge and host and stable second
- Rigidity Most databases work best on highly structured relational data. But a large and increasing percentage of data consists of machine-generated logs that matate over time, audo and video, not madely accorded to relational analysis.

We see each of the above issues only increasing with data set size. To take one large-scale customer example, the Amuson Retail team collects about 5 billion web log records daily (2TBiday,

¹ Forecast data combined from IDC, Gartner, and 451 Research ² http://www.gartner.com/it-glossary/dark-data

- Figures are an integral part of communication
- Many people only look at figures and captions



Amazon Redshift and the Case for Simpler Data Warehouses

Anurag Gupta, Deepak Agarwal, Derek Tan, Jakub Kulesza, Rahul Pathak, Stefano Stefani, Vidihya Srinivasan Amazon Web Senional

Abstract

Armann Reddrift in a flort, fully managed, petalyte-scale data warehouse solution that makes it simple and cost-efficiency to efficiently analyze large volumes of data using existing business intelligence tools. Since launching in February 2011, it has been Armann Web Service's (AWS) fastest growing service, with many fluorands of customers and many petalytics of data under management. Armann Reddrift's pace of adoption has been a surprise to many

participants in the data worknowing community. Which Amazon Rachility stay priced descriptively all under, variable for as tittle as \$1000 TB/year, those are many open-source data warehousing stochoologies and many commenculation associated supportant provide free editions for development or under some usage limit. Which Amazon Rachility provides a modern MPP, colorancy works and the stochastic provides a modern MPP, colorancy engines. And, while Amazone Rachilit is resibility in the AWS cloud, one can held that worknowers using ECI intenances and the durbative angine of one's choice with either local or networkateded and the stock of the stock of the stock of the stock of many control of the stock of the sto

In this space, we discuss an ord-overhoods differentiating characteristic of Amazon Roddill—empire, Our goal with Amazon Raddill was more than the company, with other data washousing ouglass, that to compare with more consequent. We believe the washer of the compared on the company of the c

1. Introduction

Manufactures are their transaction processing database continues the data superhouse for approxing and analysis. Analysis outrained the data waveformer market segment as 13 of the overall national database market segment as 13 of the overall lecenses and support), with an 8-11% compound annual growth are (CACRE). While this is a strong growth rate for a large, market market, over the past ten years, analysts also estimate data forcage at a pignic enterprise growing at 30-49% CACRE. Over

Permission to make digital or hard copies of part or all of this work for presental or discouse me is granted without free provided the copies are interested as the copies are made or distributed for prote or commercial advantage and that copies that the native and that distributes on the first party components of this work must be honored. For all where was, contact the Osciet's America Copyright is held by the constraints of all where was, contact the Osciet's America Copyright is held by the constraints of SCHADO'S May 31—Same 4, 2015, Melbourne, Ventria, Australia. ACM 970-1-601-529-98-1505.



the past 12-18 months, new market research has begun to show an

Figure I: Data Analysis Gap in the Enterprise [10]

This implies most data is no enterprise in "fark data", data this collected but not easily analysed, We see this as understante. If our customers didn't see this data as having valler, they would not retain it. Many companies are trying to become increasily districtiven. And yet, not only is most data already dark, the overall data landscape is only getting darkers. Sowing this data is Invited to early great and the collection of the property of the contraction of the collection of the collect

In our discussions with customers, we found the "analysis gap" between data being collected and data available for analysis as due to four major causes.

- Cost Most commercial database solutions capable of analyzing data at scale require significant up-front expense. This is hard to justify for large datasets with unclear value.
- Complexity Database provisioning, maintenance, backup, and tuning are complex tasks requiring specialized skills. They require IT involvement and cannot easily be performed.
- by line of business data scientists or analysts.

 Performance It is difficult to grow a data warehous without nonstrody inspection many performance. Once held
- Without negativery impacting query personnance. Once than, IT teams sometimes discourage augmenting data or adding queries as a way of protecting current reporting SLAs.
- Rigidity Most databases work best on highly structured relational data. But a large and increasing percentage of data consists of machine-generated logs that matate over time, audio and video, not readily accessible to relational analysis.

We see each of the above issues only increasing with data set size. To take one large-scale customer example, the Amason Retail seam collects about 5 billion web log records daily (2TB/day,

¹ Forecast data combined from IDC, Gartner, and 451 Research.
² http://www.gartner.com/it-glossary/datk-data

- Figures are an integral part of communication
- Many people only look at figures and captions
- Papers are recognized by graphs: Add a first page figure



Amazon Redshift and the Case for Simpler Data Warehouses

Anurag Gupta, Deepak Agarwal, Derek Tan, Jakub Kulesza, Rahul Pathak, Stefano Stefani, Vidihya Srinivasan Amazon Web Senional

Abstract

Amazou Rodoliti is a fast, fully managod, petalyst-scale data surerbouse solution that makes it simple and cost-efficiency to officiently analyze large volumes of data using existing business intelligence tools. Since launching in February 2011, it has been Amazou Web Service's (AWS) fastest growing service, with many flowassist of customers and many petalysis of data under management.

Amazous Endolsti's pace of adoption has been a surprise to many

participates in the data varietousing community. While Annauce Rodolft was precide descriptively at lause-Apathlo for as littles \$1000 THysar, there are many open-source data varietousing activation of the conference of the common temperature of the provide three different for development or under some using limit. While Annauce Rodolft provides a roution MPV, columnar, while Annauce Rodolft provides a roution MPV, columnar, on the conference of the common temperature of the ATS cloud, one can be talked an available in the ATS cloud, one can be talked an available using ICI intensice and the database cupies of one's choice with either local or networkanced designs.

In this paper, we discous an oth-overlooked differentiating characteristic of Amazon Reddill—employ: Our goal with Amazon Raddill was more than the company, with other data washowing engines, that to conject with not-consumption. We believe the wast majority of data is collected but not analyzed. We believe, but the continuous most of the continuous models conceived in models' conceived private and company size. And, we believe the models used to precess and company size. And, we believe the models used to precess and contenue analyzes tochologie past of to appear congenitation and evaluation. Amazone Raddill was designed to bring data warehousing to a runn model by handing a year to by, care year.

1. Introduction

Many comparies augment their transaction-processing database. Many comparies augment their transaction-processing database contrasts the data wavebosse for apprinting and majors. Analysis outstands the data wavebosse market segment as 13 of the overall excisions distallates market segment (814 to 1, 8486 for software locuses and support), with an 8-13% compound annual growth see (CACR). While their is a strong growth rate for a large, market market, over the past ten years, analysis also entirate data stronge as a higher destrates growing at 20-49% CACR. Over

Permission to make digital or hard outloss of part or all of this work for proceed or characters are in granted without fee provided that copies are not made or distributed for ports or communical advantage and that outloss have their notice and that distations on the first page. Chypylink for thispany components of this work must be honored for all other uses, contact the Osward-Andre Corpytight is held by the communication; \$500ADO '15' May 11-lane 4, 30%, Melhouran, Ventria, Australia ACM 970-1403-1599-1500.



Figure I: Data Analysis Gap in the Enterprise [10]

This implies most data is no enterprise in "fart, data", data "his collected but not easily analyses. We see this as unfortunate. If our customers didn't see this data as having valler, they would not retain it. Many companies are nying to becreen increasily data-driven. And yet, not only is most data already data, the overall data landscape is only getting dathers. Sowing this data is Invited to only getting dathers. Sowing this data is Invited to a collection of the data in Invited to the only and the sould increasion.

Some of the data is the data of the data in the data i

In our discussions with customers, we found the "analysis gap between data being collected and data available for analysis a due to four major causes.

- Cost Most commercial database solutions capable of analyzing data at scale require significant up-front expense. This is hard to justify for large datasets with unclear value.
- Complexity Database provisioning, maintenance, backup, and maing are complex tasks requiring specialized skills. They require IT involvement and cannot easily be performed by line of business data scientists or analysts.
- Performance It is difficult to grow a data warehous without negatively impacting query performance. Once built IT teams sometimes discourage augmenting data or adding
- IT teams sometimes discourage augmenting data or adding queries as a way of protecting current reporting SLAs.

 4. Rinidry – Most databases work best on highly structures
- Rigistry Noos estatutates were note on taginty structure relational data. But a large and increasing percentage of data consists of machine-generated logs that matate over time, audio and video, not readily accessible to relational analysis.

To take one large-scale customer example, the Amazon Retal team collects about 5 billion web log records daily (2TH/day ¹Forecast data combined from IDC, Gartner, and 451 Research. ² http://www.gartner.com/st-glossary/dark-data

- Figures are an integral part of communication
- Many people only look at figures and captions
- Papers are recognized by graphs: Add a first page figure
- Figures are equally important as text



Amazon Redshift and the Case for Simpler Data Warehouses

Anurag Gupta, Deepak Agarwal, Derek Tan, Jakub Kulesza, Rahul Pathak, Stefano Stefani, Vidhya Srinivasan Amaron Web Servicios

Abstract

Armson Redshift is a fast, fully managed, petabyte-scale data warehouse solution that makes it simple and cont-efficiently analyze large volunces of data using existing business intelligence teels. Since lumching in February 2013, it has been Armson Web Service's (AMS) fastest growing service, with many thousands of customers and many petabytes of data under resonancement.

Amount Robbill's pace of adoption has been a surprise to many puricipants in the data surdensing constraint; While Amazon Robbill was priced disreptively a launch, available for as little as \$10000THyere, these are many open-securities and survelessing engines that was commerced after surchessing engines that while Amazon Robbill provides a non-densing origins that While Amazon Robbill provides a non-density engine and proposed and provides and the commerced and the commerced

In this paper, we discous an oth-everlooked differentiating detactions of Antaron Reddill—empire, Our goal with Antaion Raddill was not to compare with other data sunchossing engines, that to compare with mote-consergions. We believe the sunch of the s

1. Introduction

Many comparies augment their transaction-processing database. Many comparies augment their transaction-processing database contrasts the data wavebosse for apprinting and majors. Analysis outstands the data wavebosse market segment as 13 of the overall excisions distallates market segment (814 to 1, 8486 for software locuses and support), with an 8-13% compound annual growth see (CACR). While their is a strong growth rate for a large, market market, over the past ten years, analysis also entirate data stronge as a higher destrates growing at 20-49% CACR. Over

Permission to make digital or hard outpics of part or all of this work for personal or characters are in granted without fee provided the copies are not made or distributed for port or communical advantage and that outpic but rife notice and the full distation on the first page. Chryspitals for thisputy components of this work must be honored For all other uses, common the Osciet'Assist, Corpytials is held by the communication; \$GOMOD'S, May 31—Sume 4, 3045, Melhouran, Victoria, Australia ACM 970-1403-2559-01505.



Figure 1: Data Analysis Gap in the Enterprise [10]

This implies most data is no enterprise in "fart, data", data "his collected but not easily analyses. We see this as unfortunate. If our customers didn't see this data as having value, they would not retain it. Many companies are nying to become increasily districtives. And yet, not only it must data alterdy date, the overall data landscape is only getting datelers. Storing this data in NoSQS, stores and/or Hadoop is one way to bridge the gap for certain use cases. However it down't address all recentries.

In our discussions with customers, we found the "analysis gap' between data being collected and data available for analysis as due to four major causes.

- Cost Most commercial database solutions capable of analyzing data at scale require significant up-front expense. This is hard to justify for large datasets with unclear value.
- This is hard to justify for large datasets with unclear value.

 2. Complexity Database provisioning, maintenance, backup and tuning are complex tasks requiring specialized skills
- They require IT involvement and cannot easily be performed by line of business data scientists or analysts. Performance — It is difficult to grow a data warehouse without negatively impacting query performance. Once built,
- Without negatively impacting query performance. Once tour, IT teams sometimes discourage augmenting data or adding queries as a way of protecting current reporting SLAs.
- Rigidity Most databases work best on highly structured relational data. But a large and increasing percentage of data consists of machine-generated logs that matate over time, audio and video, not readily accessible to relational analysis.

To take one large-scale customer example, the Amazon Retai team collects about 5 billion web log records daily (2TB/day

¹ Forecast data combined from IDC, Gartner, and 451 Research ² http://www.gartner.com/st-glossars/datk-data

- Figures are an integral part of communication
- Many people only look at figures and captions
- Papers are recognized by graphs: Add a first page figure
- Figures are equally important as text
- Text should discuss the takeaways of a figure



Amazon Redshift and the Case for Simpler Data Warehouses

Anurag Gupta, Deepak Agarwal, Derek Tan, Jakub Kulesza, Rahul Pathak, Stefano Stefani, Vidhya Srinivasan Amaron Web Reviews

Abstract

Armson Redshift is a fast, fully managed, petabyte-scale data warehouse solution that makes it simple and cont-efficiently analyze large volunces of data using existing business intelligence teels. Since lumching in February 2013, it has been Armson Web Service's (AMS) fastest growing service, with many thousands of customers and many petabytes of data under resonancement.

Amazon Rabbili's pass of adoption has been a surprise to many participates in the data wardensing consensative. Which Amazon Rabbili was proof disreptively at lanceb, resibilité fers a litric as 1000 Thiyan; also are many open-second data wardensing suchossipps and many commercial data wardensing segmen that Major and the surprise of the surprise of the surprise of the Major and Salability freededs a medien MPP, colarany, scale-out architecture, as is one many other data wardensing sugars. And, which Amazon Rabbilit is resibled in the APS support of the surprise of th

In this paper, we discous an oth-everlooked differentiating detactions of Antaron Reddill—empire, Our goal with Antaion Raddill was not to compare with other data sunchossing engines, that to compare with mote-consergions. We believe the sunch of the s

1. Introduction

Monty companies augment their transaction-processing database. Monty companies augment their transaction-processing database and a support of the support of the support of the support outsined database market segment at 15 of the overall actional database market segment (54 the x x-850 for software locuses and support), with an 8-11% compound annual growth are for a few portions of the support of the support and (CAGR). While this is a storage growth rate for a large, market market, over the past ten years, analysis also estimate data success as a high categorie growing at 29-49% CAGR. Over success as a long of the support of the support success as a long of the support of the support success as a long of the support of the support success as a long of the support success as a long of the support success as a long of the support success success

Permission to make digital or hard outpics of part or all of this work for personal or characters are in granted without fee provided the copies are not made or distributed for port or communical advantage and that outpic but rife notice and the full distation on the first page. Chryspitals for thisputy components of this work must be honored For all other uses, common the Osciet'Assist, Corpytials is held by the commissionality. \$GOMOD'15. May 31—Sume 4, 3015, Melhouran, Victoria, Australia ACM 970-1460-2559-01505.



Figure 1: Data Analysis Gap in the Enterprise [10]

This implies most data is no enterprise in "fark data", data that is collected but not easily analysed. We see this as unferturate. If our customers didn't see this data as having value, they would not retain it. Many companies are typing to become increasily districtives. And yet, not only it most data already dark, the overall data landages is only griting datafer. Sowing this data in NASUR, stores and/or Hadoop is one way to bridge the gap for certain use cases. However it doesn't address all recentries.

In our discussions with customers, we found the "analysis gap' between data being collected and data available for analysis as due to four major causes.

- Cost Most commercial database solutions capable of analyzing data at scale require significant up-front expense. This is bard to instiff for large datasets with unclear value.
- Complexity Database provisioning, maintenance backup and tuning are complex tasks requiring specialized skills They require IT involvement and cannot easily be performed.
- by line of business data scientists or analysts.

 3. Performance It is difficult to grow a data warehous without negatively impacting query performance. Once built
- IT teams sometimes discourage augmenting data or adding queries as a way of protecting current reporting SLAs.

 4. Rigidity — Most databases work best on highly structures
- 4. Rejustry Most entirenses work best on taginty structures relational data. But a large and increasing percentage of data consists of machine-generated logs that metate over time, audio and video, not readily accessible to relational analysis. We see each of the above issues only increasing with data set size. To take one large-scale customer example, the Armson Retail.

team collects about 5 billion web log records daily (2TB/day

Forecast data combined from IDC, Gartner, and 451 Research.

 It should be possible to roughly understand a figure without the text

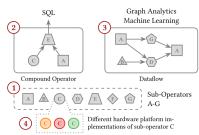


Figure 1: Sub-operators (1) build more complex data operations (2) or dataflows (3), where each sub-operator can be implemented on multiple hardware platforms (4).



- It should be possible to roughly understand a figure without the text
- Use captions to explain figures

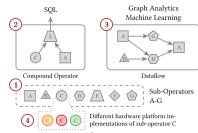


Figure 1: Sub-operators (1) build more complex data operations (2) or dataflows (3), where each sub-operator can be implemented on multiple hardware platforms (4).



- It should be possible to roughly understand a figure without the text
- Use captions to explain figures
- Captions may be several lines long

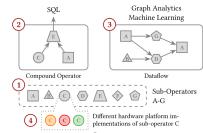


Figure 1: Sub-operators (1) build more complex data operations (2) or dataflows (3), where each sub-operator can be implemented on multiple hardware platforms (4).



- It should be possible to roughly understand a figure without the text
- Use captions to explain figures
- Captions may be several lines long
- Use text within figures to explain

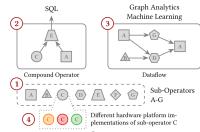


Figure 1: Sub-operators (1) build more complex data operations (2) or dataflows (3), where each sub-operator can be implemented on multiple hardware platforms (4).



- It should be possible to roughly understand a figure without the text
- Use captions to explain figures
- Captions may be several lines long
- Use text within figures to explain
- But: every figure needs to have a purpose for the text and must be referenced!

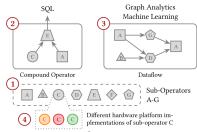


Figure 1: Sub-operators (1) build more complex data operations (2) or dataflows (3), where each sub-operator can be implemented on multiple hardware platforms (4).



- It should be possible to roughly understand a figure without the text
- Use captions to explain figures
- Captions may be several lines long
- Use text within figures to explain
- But: every figure needs to have a purpose for the text and must be referenced!
- For complex graphics, also use the circledsteps package to (1) reference figure parts

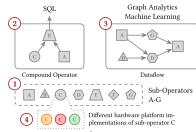


Figure 1: Sub-operators (1) build more complex data operations (2) or dataflows (3), where each sub-operator can be implemented on multiple hardware platforms (4).



Figures Should be Simple

• Make figures as simple as possible

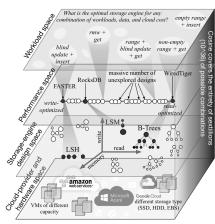
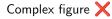


Figure 1: Fixed-design systems capture only a small fraction of the possible storage-engine design space on the cloud.



Figures

Figures Should be Simple

- Make figures as simple as possible
- Better use two figures than one complex figure

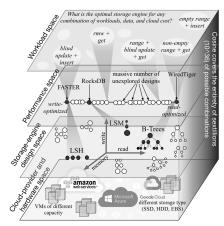
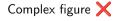


Figure 1: Fixed-design systems capture only a small fraction of the possible storage-engine design space on the cloud.



• Good coloring can make a graph more understandable

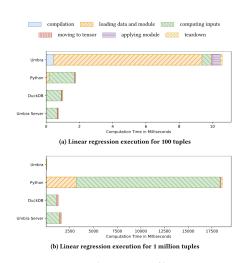


Figure 10: Runtime decomposition of linear regression.



- Good coloring can make a graph more understandable
- Should be accessible to colorblind and people with black and white printers

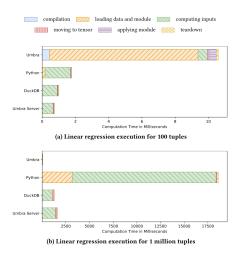


Figure 10: Runtime decomposition of linear regression.



- Good coloring can make a graph more understandable
- Should be accessible to colorblind and people with black and white printers
 - Important to many, ignored by many

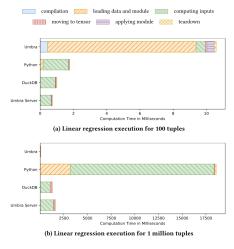


Figure 10: Runtime decomposition of linear regression.



- Good coloring can make a graph more understandable
- Should be accessible to colorblind and people with black and white printers
 - Important to many, ignored by many
- Our advice:

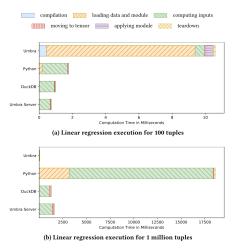


Figure 10: Runtime decomposition of linear regression.



- Good coloring can make a graph more understandable
- Should be accessible to colorblind and people with black and white printers
 - Important to many, ignored by many
- Our advice:
 - Use color to make figures as expressive as possible

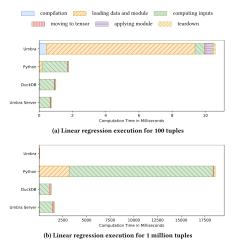


Figure 10: Runtime decomposition of linear regression.



- Good coloring can make a graph more understandable
- Should be accessible to colorblind and people with black and white printers
 - Important to many, ignored by many
- Our advice:
 - Use color to make figures as expressive as possible
 - Use shades (<= 4), shapes, or text to present the same information without color

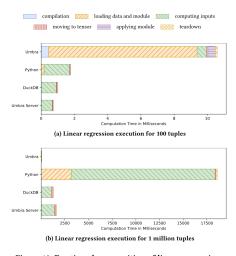


Figure 10: Runtime decomposition of linear regression.



- Good coloring can make a graph more understandable
- Should be accessible to colorblind and people with black and white printers
 - Important to many, ignored by many
- Our advice:
 - Use color to make figures as expressive as possible
 - Use shades (<= 4), shapes, or text to present the same information without color
 - Ignore this rule if it is impossible



Figure 10: Runtime decomposition of linear regression.



Revise Figures

 Just as for text, you should spend most of your time revising a figure

Revise Figures

- Just as for text, you should spend most of your time revising a figure
- Start with sketches of figures or quick plots

Revise Figures

- Just as for text, you should spend most of your time revising a figure
- Start with sketches of figures or quick plots
- If they stay in the paper, you polish them later

• Usually, graphs are better than tables

Official ScanNet Benchmark

Method	avg IoU	Chair	Floor	Other Furniture	Picture	Sofa	Table	Wall	
${\bf PointNet}{+}{+}$	0.339	0.360	0.677	0.183	0.117	0.346	0.232	0.523	
PointNet++ 1	0.481	0.686	0.931	0.299	0.102	0.580	0.470	0.711	
Additional Features (ours)	0.557	0.744	0.946	0.376	0.205	0.643	0.497	0.756	



- Usually, graphs are better than tables
- Tables allow to perform lots of comparisons

Official ScanNet Benchmark

Method	avg IoU	Chair	Floor	Other Furniture	Picture	Sofa	Table	Wall	
PointNet++	0.339	0.360	0.677	0.183	0.117	0.346	0.232	0.523	
PointNet++ 1	0.481	0.686	0.931	0.299	0.102	0.580	0.470	0.711	
Additional Features (ours)	0.557	0.744	0.946	0.376	0.205	0.643	0.497	0.756	



- Usually, graphs are better than tables
- Tables allow to perform lots of comparisons
 - You can provide a lot of detail in tables

Official ScanNet Benchmark

Method	avg IoU	Chair	Floor	Other Furniture	Picture	Sofa	Table	Wall	
PointNet++	0.339	0.360	0.677	0.183	0.117	0.346	0.232	0.523	
PointNet++ 1	0.481	0.686	0.931	0.299	0.102	0.580	0.470	0.711	
Additional Features (ours)	0.557	0.744	0.946	0.376	0.205	0.643	0.497	0.756	



- Usually, graphs are better than tables
- Tables allow to perform lots of comparisons
 - You can provide a lot of detail in tables
- Use visual hints in tables:

Official ScanNet Benchmark

Method	avg IoU	Chair	Floor	Other Furniture	Picture	Sofa	Table	Wall	
PointNet++	0.339	0.360	0.677	0.183	0.117	0.346	0.232	0.523	
PointNet++ 1	0.481	0.686	0.931	0.299	0.102	0.580	0.470	0.711	
Additional Features (ours)	0.557	0.744	0.946	0.376	0.205	0.643	0.497	0.756	



- Usually, graphs are better than tables
- Tables allow to perform lots of comparisons
 - You can provide a lot of detail in tables
- Use visual hints in tables:
 - Bold for best

Official ScanNet Benchmark

Method	avg IoU	Chair	Floor	Other Furniture	Picture	Sofa	Table	Wall	
PointNet++	0.339	0.360	0.677	0.183	0.117	0.346	0.232	0.523	
PointNet++ 1	0.481	0.686	0.931	0.299	0.102	0.580	0.470	0.711	
Additional Features (ours)	0.557	0.744	0.946	0.376	0.205	0.643	0.497	0.756	



- Usually, graphs are better than tables
- Tables allow to perform lots of comparisons
 - You can provide a lot of detail in tables
- Use visual hints in tables:
 - Bold for best
 - Color for good/neutral/bad

Official ScanNet Benchmark

Method	avg IoU	Chair	Floor	Other Furniture	Picture	Sofa	Table	Wall	
$\mathbf{PointNet} + +$	0.339	0.360	0.677	0.183	0.117	0.346	0.232	0.523	
PointNet++ 1	0.481	0.686	0.931	0.299	0.102	0.580	0.470	0.711	
Additional Features (ours)	0.557	0.744	0.946	0.376	0.205	0.643	0.497	0.756	

Pseudocode

• Pseudocode can be very helpful

Pseudocode

- Pseudocode can be very helpful
- Describes algorithms very accurately

Pseudocode

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible
 - De-clutter syntax (Python-like syntax can be good)

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible
 - De-clutter syntax (Python-like syntax can be good)
 - Use syntax highlighting

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible
 - De-clutter syntax (Python-like syntax can be good)
 - Use syntax highlighting
 - Name variables well

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible
 - De-clutter syntax (Python-like syntax can be good)
 - Use syntax highlighting
 - Name variables well
 - Name algorithms / functions / methods

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible
 - De-clutter syntax (Python-like syntax can be good)
 - Use syntax highlighting
 - Name variables well
 - Name algorithms / functions / methods
 - Add descriptive captions

- Pseudocode can be very helpful
- Describes algorithms very accurately
 - Be precise on details, there is no room for interpretation
- Is difficult to understand
 - Reduce cognitive load as much as possible
 - De-clutter syntax (Python-like syntax can be good)
 - Use syntax highlighting
 - Name variables well
 - Name algorithms / functions / methods
 - Add descriptive captions
 - Specify input and output

Use LLMs for

- Use LLMs for
 - checking spelling and grammar, get feedback

- Use LLMs for
 - checking spelling and grammar, get feedback
 - getting inspiration for how to write a certain sentence or section

- Use LLMs for
 - checking spelling and grammar, get feedback
 - getting inspiration for how to write a certain sentence or section
- Do not use LLMs for

- Use LLMs for
 - checking spelling and grammar, get feedback
 - getting inspiration for how to write a certain sentence or section
- Do not use LLMs for
 - structure your paper or work on your very first draft (e.g., bulletpoints) of a section

- Use LLMs for
 - checking spelling and grammar, get feedback
 - getting inspiration for how to write a certain sentence or section
- Do not use LLMs for
 - structure your paper or work on your very first draft (e.g., bulletpoints) of a section
 - writing your (entire paper), the writing style tends to be not on point, also external requirements

- Use LLMs for
 - checking spelling and grammar, get feedback
 - getting inspiration for how to write a certain sentence or section
- Do not use LLMs for
 - structure your paper or work on your very first draft (e.g., bulletpoints) of a section
 - writing your (entire paper), the writing style tends to be not on point, also external requirements
- If you don't want to give some company all your data, you can also self-host small models I like ollama with lama3-8b

- Use LLMs for
 - checking spelling and grammar, get feedback
 - getting inspiration for how to write a certain sentence or section
- Do not use LLMs for
 - structure your paper or work on your very first draft (e.g., bulletpoints) of a section
 - writing your (entire paper), the writing style tends to be not on point, also external requirements
- If you don't want to give some company all your data, you can also self-host small models I like ollama with lama3-8b
- If we notice that your paper includes any text you did not understand or find long sections that just paraphrase sources you will not pass this course

• LaTeX composes visually nice results (visual appeal is important)

- LaTeX composes visually nice results (visual appeal is important)
- Often, you have to use templates anyway

- LaTeX composes visually nice results (visual appeal is important)
- Often, you have to use templates anyway
- Write one sentence per line

- LaTeX composes visually nice results (visual appeal is important)
- Often, you have to use templates anyway
- Write one sentence per line
 - Check length of sentences

- LaTeX composes visually nice results (visual appeal is important)
- Often, you have to use templates anyway
- Write one sentence per line
 - Check length of sentences
 - Easier reasoning about sentences (each sentence should have a purpose)

- LaTeX composes visually nice results (visual appeal is important)
- Often, you have to use templates anyway
- Write one sentence per line
 - Check length of sentences
 - Easier reasoning about sentences (each sentence should have a purpose)
 - Works well with version control

• Always use git even if you are working alone

- Always use git even if you are working alone
- Helps with collaboration

- Always use git even if you are working alone
- Helps with collaboration
- Synchronizes across devices

- Always use git even if you are working alone
- Helps with collaboration
- Synchronizes across devices
- Shows diffs

- Always use git even if you are working alone
- Helps with collaboration
- Synchronizes across devices
- Shows diffs
- Makes it harder to loose data

- Always use git even if you are working alone
- Helps with collaboration
- Synchronizes across devices
- Shows diffs
- Makes it harder to loose data
- Commit often, push often

- Always use git even if you are working alone
- Helps with collaboration
- Synchronizes across devices
- Shows diffs
- Makes it harder to loose data
- Commit often, push often
- Check in all stuff necessary for building (text and figures) but not outputs

• For quick iteration try to avoid LaTeX for initial draft(s)

- For quick iteration try to avoid LaTeX for initial draft(s)
 - Text: Simple editor without much distraction (e.g. Word, Google Docs, Plain Editor)

- For quick iteration try to avoid LaTeX for initial draft(s)
 - Text: Simple editor without much distraction (e.g. Word, Google Docs, Plain Editor)
 - Tables: Any software that can edit tables (Word, Excel, or alternatives)

- For quick iteration try to avoid LaTeX for initial draft(s)
 - Text: Simple editor without much distraction (e.g. Word, Google Docs, Plain Editor)
 - Tables: Any software that can edit tables (Word, Excel, or alternatives)
 - Images: avoid perfectionism (excalidraw.com, Paper)

- For quick iteration try to avoid LaTeX for initial draft(s)
 - Text: Simple editor without much distraction (e.g. Word, Google Docs, Plain Editor)
 - Tables: Any software that can edit tables (Word, Excel, or alternatives)
 - Images: avoid perfectionism (excalidraw.com, Paper)
- use \usepackage{booktabs} for tables

Citations with BibTeX

Citing with BibTeX is easy

```
___ sample.bib _
@Article{Abril07,
 author = "Patricia S. Abril and...",
 title
             = "The patent holder's...",
             = "Communications of the ACM",
 journal
 volume
             = "50",
 number
              = "1",
 month
             = jan,
             = "2007",
 year
             = "36--44",
 pages
 doi
             = "10.1145/1188913.1188915",
             = "http://doi.acm.org/...",
 url
```

```
main.tex
\section{Citations}
Some references~\cite{Abril07,Xy} ...
```

Citations with BibTeX

- Citing with BibTeX is easy
- Get correct .bib files from dblp.org

```
___ sample.bib _
@Article{Abril07,
 author = "Patricia S. Abril and...",
 title
              = "The patent holder's...",
              = "Communications of the ACM",
 iournal
 volume
              = "50",
 number
              = "1",
 month
              = jan,
              = "2007",
 year
              = "36--44",
 pages
 doi
              = "10.1145/1188913.1188915",
              = "http://doi.acm.org/...",
 url
```

- Citing with BibTeX is easy
- Get correct .bib files from dblp.org
 - Be sure that you cite the correct version of the paper

```
__ sample.bib _
@Article{Abril07,
 author
              = "Patricia S. Abril and...",
 title
              = "The patent holder's...",
              = "Communications of the ACM",
 iournal
 volume
              = "50",
 number
              = "1",
 month
              = jan,
              = "2007",
 vear
              = "36--44",
 pages
 iob
              = "10.1145/1188913.1188915".
              = "http://doi.acm.org/...",
 url
```

```
main.tex
\section{Citations}
Some references~\cite{Abril07,Xy} ...
```

- Citing with BibTeX is easy
- Get correct .bib files from dblp.org
 - Be sure that you cite the correct version of the paper
 - Arxiv is one of the worst sources, try to find a source from a journal or conference

```
_ sample.bib _
@Article{Abril07.
 author
               = "Patricia S. Abril and...".
 title
               = "The patent holder's...",
 iournal
               = "Communications of the ACM",
 volume
               = "50",
 number
               = "1",
 month
               = jan,
               = "2007",
 vear
               = "36--44".
 pages
 iob
               = "10.1145/1188913.1188915",
               = "http://doi.acm.org/...",
 url
```

```
main.tex
\section{Citations}
Some references~\cite{Abril07,Xy} ...
```

- Citing with BibTeX is easy
- Get correct .bib files from dblp.org
 - Be sure that you cite the correct version of the paper
 - Arxiv is one of the worst sources, try to find a source from a journal or conference
 - Especially on google scholar you will often find bad BibTeX files

```
__ sample.bib _
@Article{Abril07.
 author
               = "Patricia S. Abril and...".
 title
               = "The patent holder's...",
 iournal
               = "Communications of the ACM",
 volume
               = "50",
 number
               = "1",
 month
               = ian.
               = "2007",
 vear
               = "36--44".
 pages
 iob
               = "10.1145/1188913.1188915".
               = "http://doi.acm.org/...",
 url
```

```
main.tex \( \section\{\text{Citations}\} \)
Some references~\( \text{cite}\{\text{Abril07,Xy}\} \) \( \text{...} \)
```

- Citing with BibTeX is easy
- Get correct .bib files from dblp.org
 - Be sure that you cite the correct version of the paper
 - Arxiv is one of the worst sources, try to find a source from a journal or conference
 - Especially on google scholar you will often find bad BibTeX files
- Before submission: make sure any warnings are gone and all citations render!

```
_ sample.bib _
@Article{Abril07.
 author
               = "Patricia S. Abril and...".
 title
               = "The patent holder's...",
 iournal
               = "Communications of the ACM",
 volume
               = "50",
 number
               = "1",
 month
               = ian.
               = "2007",
 vear
               = "36--44".
 pages
 iob
               = "10.1145/1188913.1188915".
 url
               = "http://doi.acm.org/...",
```

```
\section{Citations}
Some references~\cite{Abril07,Xy} ...
```

• New terms should be written once in italics and then explained

- New terms should be written once in italics and then explained
- Make sure each figure is referenced in text

- New terms should be written once in *italics* and then explained
- Make sure each figure is referenced in text
- Position figures when you are finished with the rest

- New terms should be written once in italics and then explained
- Make sure each figure is referenced in text
- Position figures when you are finished with the rest
- Optimize line breaks

- New terms should be written once in italics and then explained
- Make sure each figure is referenced in text
- Position figures when you are finished with the rest
- Optimize line breaks
- Check references

- New terms should be written once in italics and then explained
- Make sure each figure is referenced in text
- Position figures when you are finished with the rest
- Optimize line breaks
- Check references
- Spell check

- New terms should be written once in italics and then explained
- Make sure each figure is referenced in text
- Position figures when you are finished with the rest
- Optimize line breaks
- Check references
- Spell check
- Use your favorite LLVM for spell and grammar checks

Further References

- Deirdre Nansen McCloskey, Economical Writing, Third Edition, 2019
- Joseph Williams, Style: Toward Clarity and Grace, Univ. of Chicago Press, 1990
- Justin Zobel, Writing for Computer Science, Springer, Third Edition, 2014
- Larry McEnerney, The Craft of Writing Effectively [link]
- Lorenz Froihofer, Tips for scientific writing (for Germans) [link]
- Lorenz Froihofer, How to write a computer science paper [link]

• Good writing is hard but important

- Good writing is hard but important
- Each sentence needs to contribute to your goal

- Good writing is hard but important
- Each sentence needs to contribute to your goal
- Revise more than you write

- Good writing is hard but important
- Each sentence needs to contribute to your goal
- Revise more than you write
- Write clear and concise

- Good writing is hard but important
- Each sentence needs to contribute to your goal
- Revise more than you write
- Write clear and concise
- A perfect text is one where you cannot omit anything