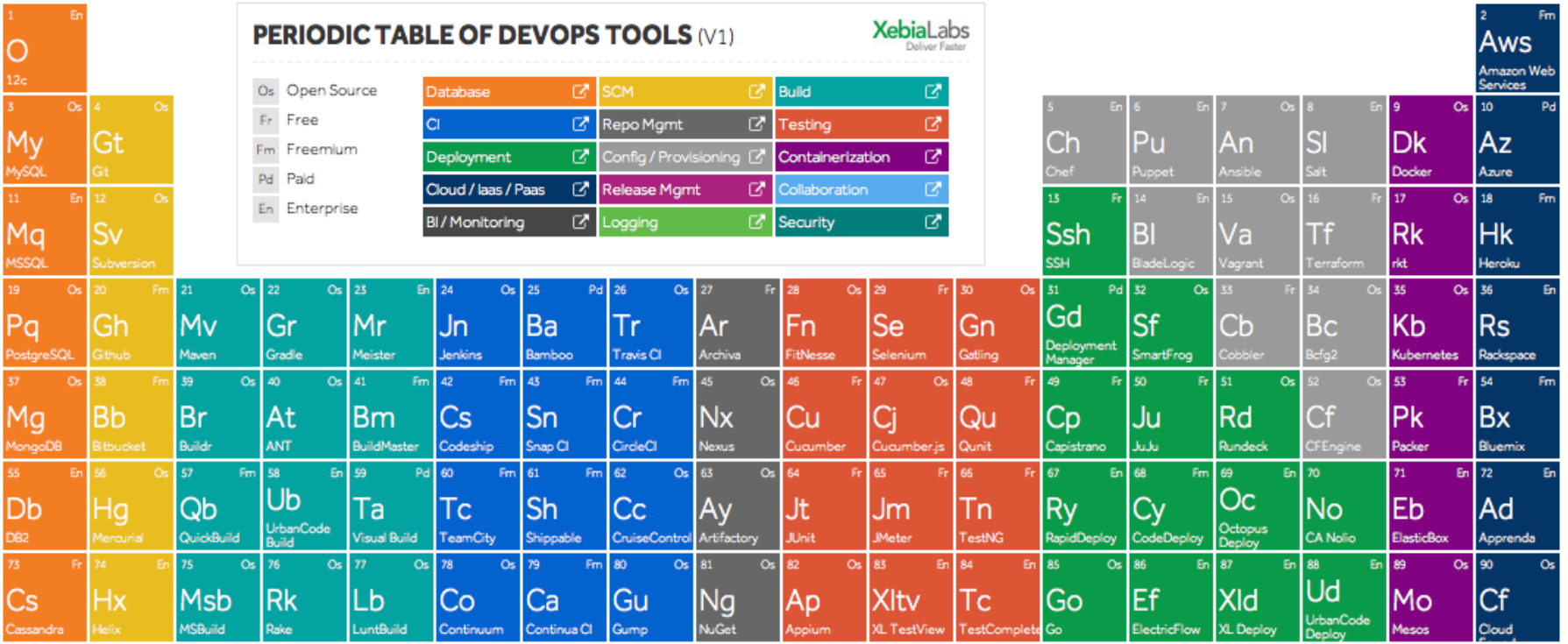
Introduction to Scientific Software Deployment and Development

damien.francois@uclouvain.be
October 2021

http://www.ceci-hpc.be/training.html



https://xebialabs.com/periodic-table-of-devops-tools/

Goal of this session:

“Promote the *tools*
the professionals are using for
developing and **deploying** programs,
to make them **correct**, **maintainable**, **shareable**, and **fast**,
efficiently.”

“...to make them **correct** and **maintainable**, …, *efficiently*”

Paul F. Dubois. 1999. **Ten Good Practices in Scientific Programming**. *Computing in Science and Eng.* 1, 1 (January 1999), 7-11. DOI=10.1109/MCISE.1999.743610 http://dx.doi.org/10.1109/MCISE.1999.743610

Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. (2014) **Best Practices for Scientific Computing**. *PLoS Biol* 12(1): e1001745. doi:10.1371/journal.pbio.1001745

Dubois PF, Epperly T, Kumfert G (2003) **Why Johnny can't build (portable scientific software)**. *Comput Sci Eng* 5: 83–88. doi: 10.1109/mcise.2003.1225867

Prlić A, Procter JB (2012) **Ten Simple Rules for the Open Development of Scientific Software**. *PLoS Comput Biol* 8(12): e1002802. doi:10.1371/journal.pcbi.1002802

Victor R. Basili, Jeffrey C. Carver, Daniela Cruzes, Lorin M. Hochstein, Jeffrey K. Hollingsworth, Forrest Shull, Marvin V. Zelkowitz, "**Understanding the High-Performance-Computing Community: A Software Engineer's Perspective**," *IEEE Software*, vol. 25, no. 4, pp. 29-36, July/August, 2008

Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) **Good enough practices in scientific computing**. *PLoS Comput Biol* 13(6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510

“...to make them **correct** and **maintainable**, …, *efficiently*”

Follow programming good practices:

1. Write for humans, not for computers
2. Use the appropriate language
3. Organize for change and make incremental changes
4. Follow good coding principles
5. Plan for mistakes, automate testing
6. Use modern source-code management system
7. Document the design and purpose, not the implementation
8. Optimize only when it works already
9. Debug cleverly

1. Write for humans, not for computers

"Indeed, the ratio of time spent reading versus writing is well over 10 to 1. We are constantly reading old code as part of the effort to write new code. ...[Therefore,] making it easy to read makes it easier to write."

Robert C. Martin *Clean code A Handbook of Agile Software Craftsmanship,* 2009

1. Write for humans, not for computers

"Indeed, the ratio of time spent reading versus writing is well over 10 to 1. We are constantly reading old code as part of the effort to write new code. ...[Therefore,] making it easy to read makes it easier to write."

- Do not be afraid of being too explicit
- Choose names carefully
- Structure your code

Let the compiler do the writing for the computer

Robert C. Martin *Clean code A Handbook of Agile Software Craftsmanship,* 2009

1. Write for humans, not for computers

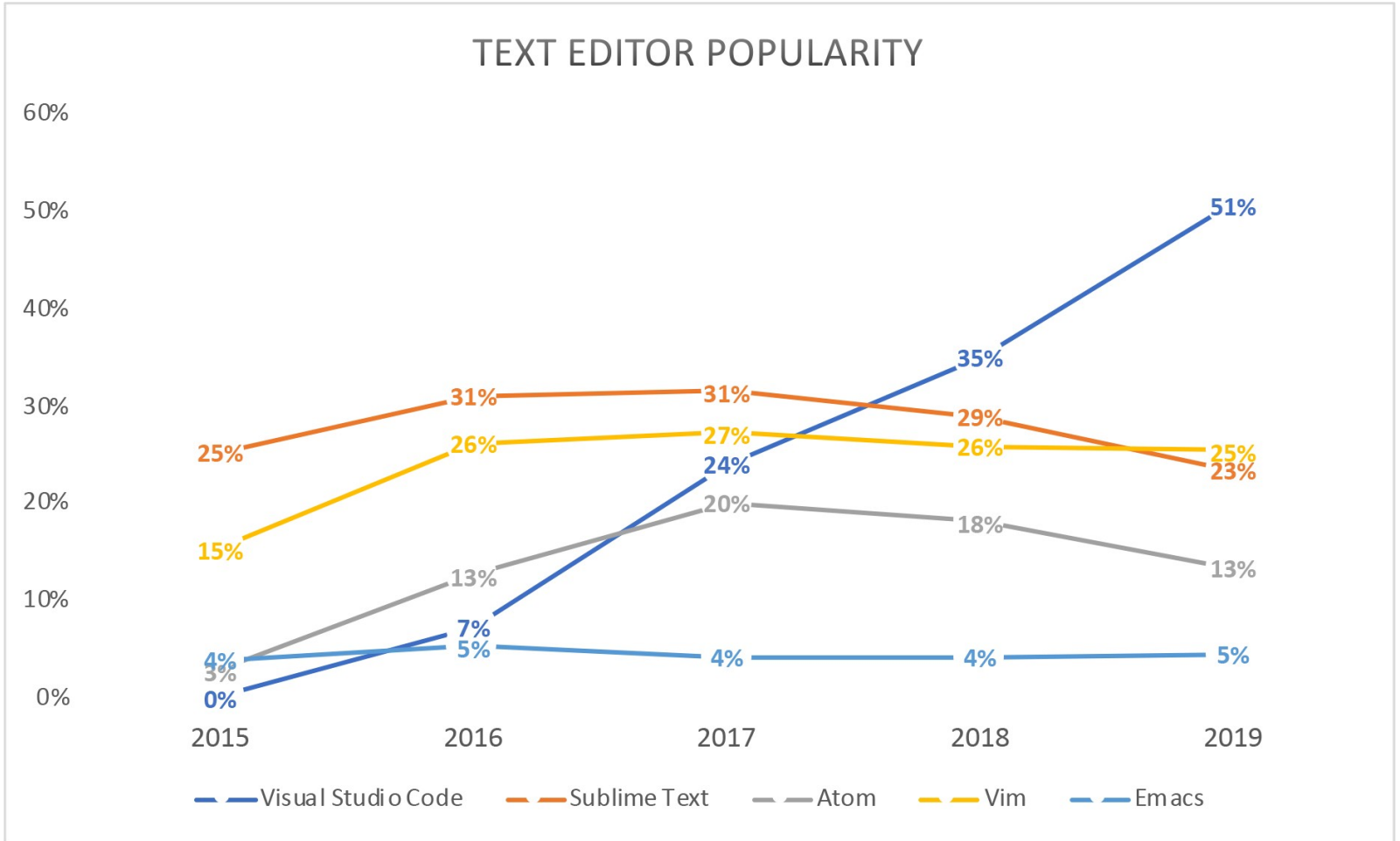
```
for i in range(n):
    for j in range(m):
        for k in range(l):
            temp_value = X[i][j][k] * 12.5
            new_array[i][j][k] = temp_value + 150
```

vs

```
PIXEL_NORMALIZATION_FACTOR = 12.5
PIXEL_OFFSET_FACTOR = 150

for row_index in range(row_count):
    for column_index in range(column_count):
        for color_channel_index in range(color_channel_count):
            normalized_pixel_value = (
                original_pixel_array[row_index][column_index][color_channel_index]
                * PIXEL_NORMALIZATION_FACTOR
            )
            transformed_pixel_array[row_index][column_index][color_channel_index] = (
                normalized_pixel_value + PIXEL_OFFSET_FACTOR
            )
```

https://towardsdatascience.com/data-scientists-your-variable-names-are-awful-heres-how-to-fix-them-89053d2855be

1. Write for humans, not for computers (and learn your editor)



https://blog.robenkleene.com/2020/09/21/the-era-of-visual-studio-code/

2. Use the appropriate language

are all valid choices in a scientific context.

2. Use the appropriate language

What they have in common:

- Computation efficiency concern
- Optimized libraries available for linear algebra, signal processing, learning, etc.
- Support for parallel computing
- Extensions/libraries for using GPUs

26 Oct	Olivier Mattelaer, "Introduction to Object-Oriented programming with C++"
26 Oct	Olivier Mattelaer, "Introduction to C programming language"
21 Oct	Damien François, "Introduction to parallel computing"
21 Oct	Damien François, "Introduction to scripting and interpreted languages (Python, R, Octave)"
21 Oct	Pierre-Yves Barriat, "Introduction to structured programming with Fortran"
20 Oct	Bernard Van Renterghem, "Introduction to compilers and compiling, and optimized libraries"
20 Oct	Bernard Van Renterghem, "Introduction to modules and software on a CÉCI cluster"

2. Use the appropriate language

“Functional programming”

Very close to mathematical formulation

Imposes constraints that make code less prone to bugs and easier to make parallel

Not very popular in HPC (yet)

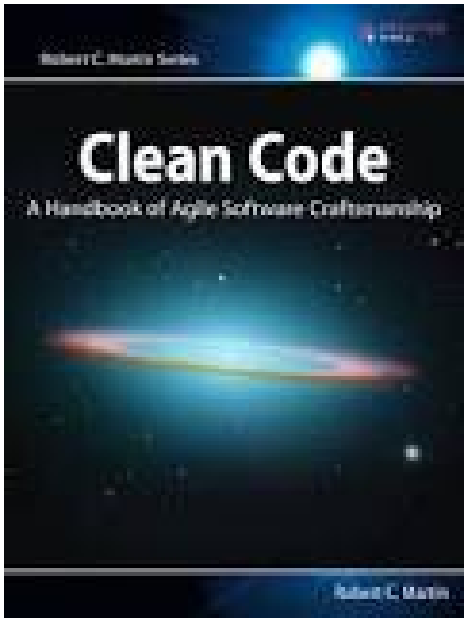
3. Organize for change and make incremental changes

- Scientific software specifications are always changing
- Work from working state to another working state
- Document the changes and why they were made
- (And sometimes restart from scratch)

Keyword: **modularity:** small independent interchangeable building blocks (e.g. functions)

4. Follow good coding principles

- Don't repeat yourself (DRY)
- Keep it simple, Stupid (KISS)
- One level of abstraction
- Single responsibility principle
- Separation of concern
- Avoid premature optimization
- Follow style guidelines
- Many others...



> Bill Mitchell View profile More options Sep 26 1991, 1:57 am In article <5...@ksr.com> j...@ksr.com (John F. Woods) writes:
>
> > [...] Always code as if the guy who ends up maintaining your code will be a violent psychopath who knows where you live. Code for readability.
>
> Damn right!

Clean Code: A Handbook of Agile Software Craftsmanship, R. C. Martin, Prentice Hall, 2008

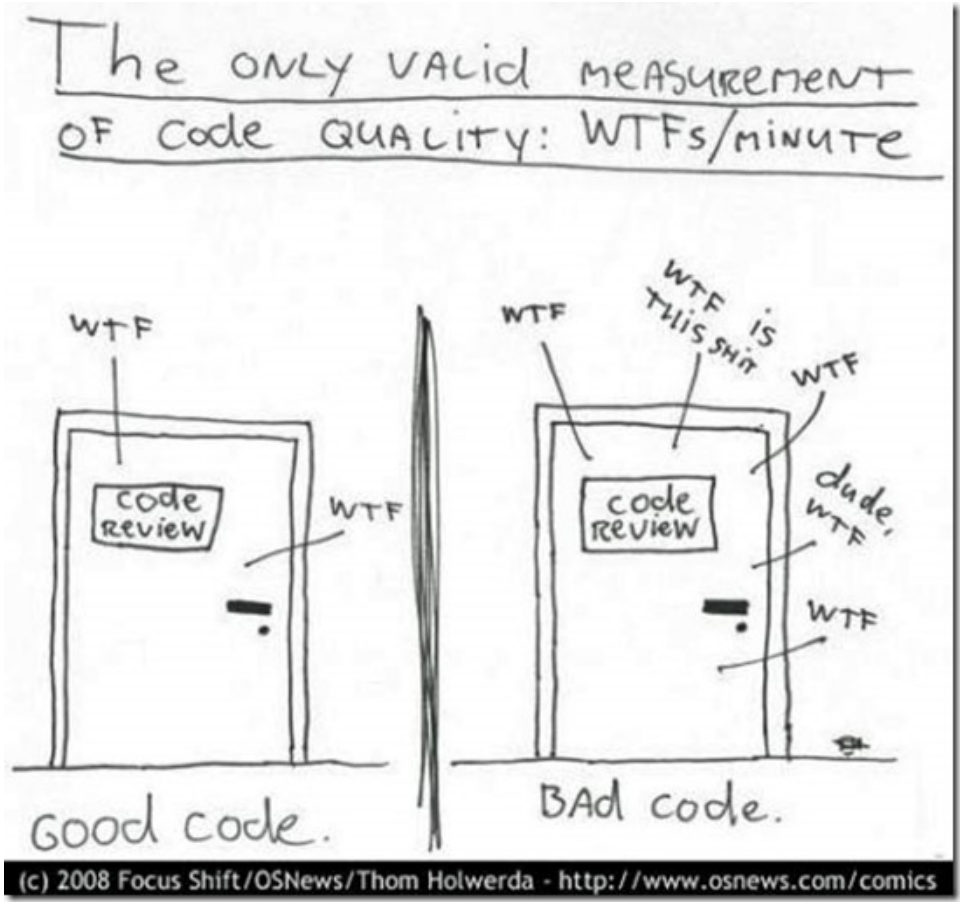
4. Follow good coding principles and style

https://stackoverflow.com/questions/1642028/what-is-the-operator-in-c

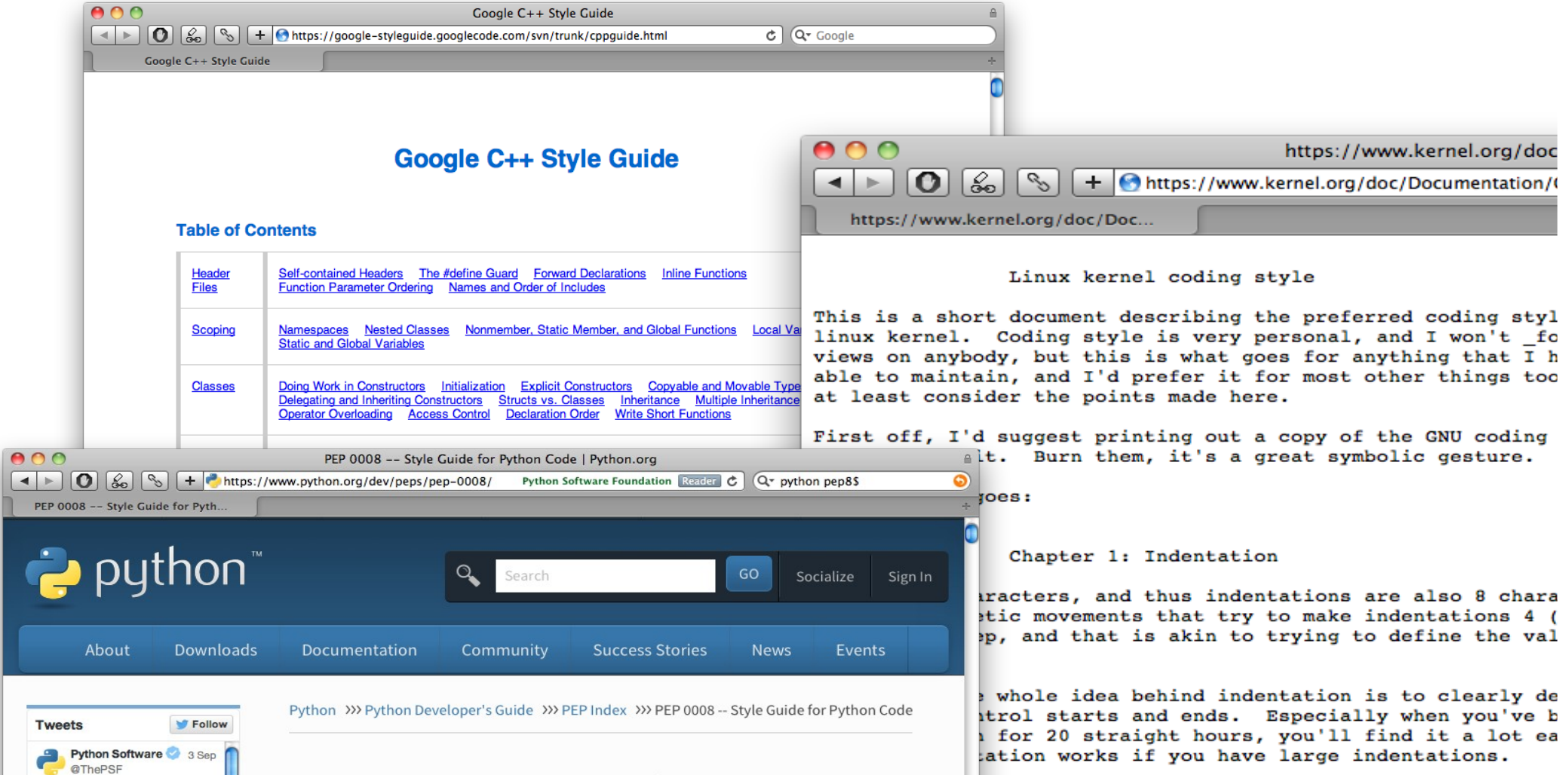
4. Follow good coding principles and style

https://stackoverflow.com/questions/1642028/what-is-the-operator-in-c

4. Follow good coding principles and style

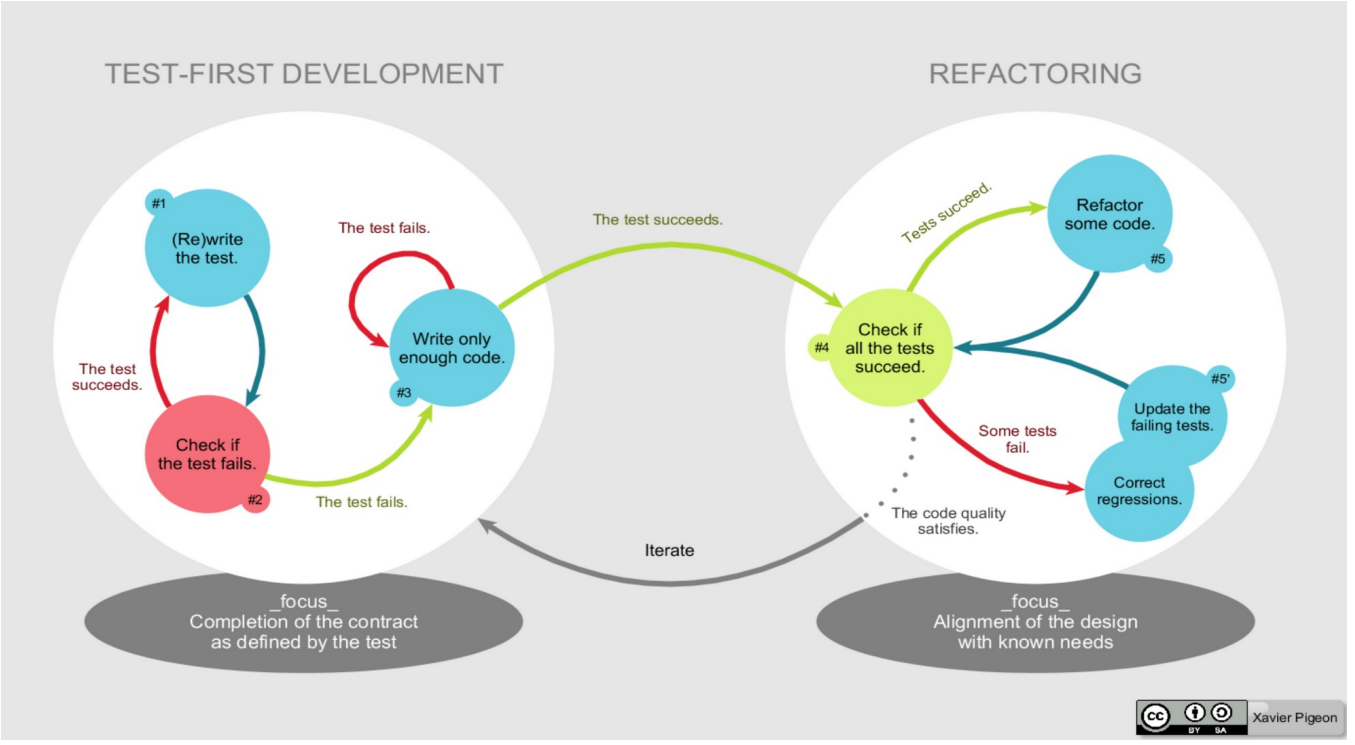


4. Follow good coding principles and style



5. Plan for mistakes, automate testing; Test-driven development

Write the tests before you even write the code



https://en.wikipedia.org/wiki/Test-driven_development

6. Use modern source-code management system

for your code, papers, thesis, etc.

2021

Introduction to code versioning

by Dr Olivier Mattelaer (UCLouvain/CISM)

Wednesday 27 Oct 2021, 09:00 → 12:00 Europe/Brussels

comodal (louvain-la-neuve or remote)

Description

Code versioning is very important to master, even for non programmers. It allows tracking the changes made to a submission script, a piece of code, a configuration file, or event a dataset and propagate the changes in a consistent and systematic way to all clusters.

Contents:

- Notions of code versioning
- Working as a team with code versioning
- Using git to access code from others
- Publishing code

Prerequisite:

- Being able to use SSH with private keys
- Being familiar with a text editor
- Mastering the Linux command line and the GNU utilities (mkdir, cp, scp, etc.)

Type: Hands-on
Target audience: Rookie programmer
Must: This session is a must-have for anyone not familiar with code versioning or git.

Organized by UCLouvain/CISM

Registration Participants 11 / 60 Register

Contact egs-cism@listes.uclouvain.be
0494424767

7. Document the purpose and design, not the implementation

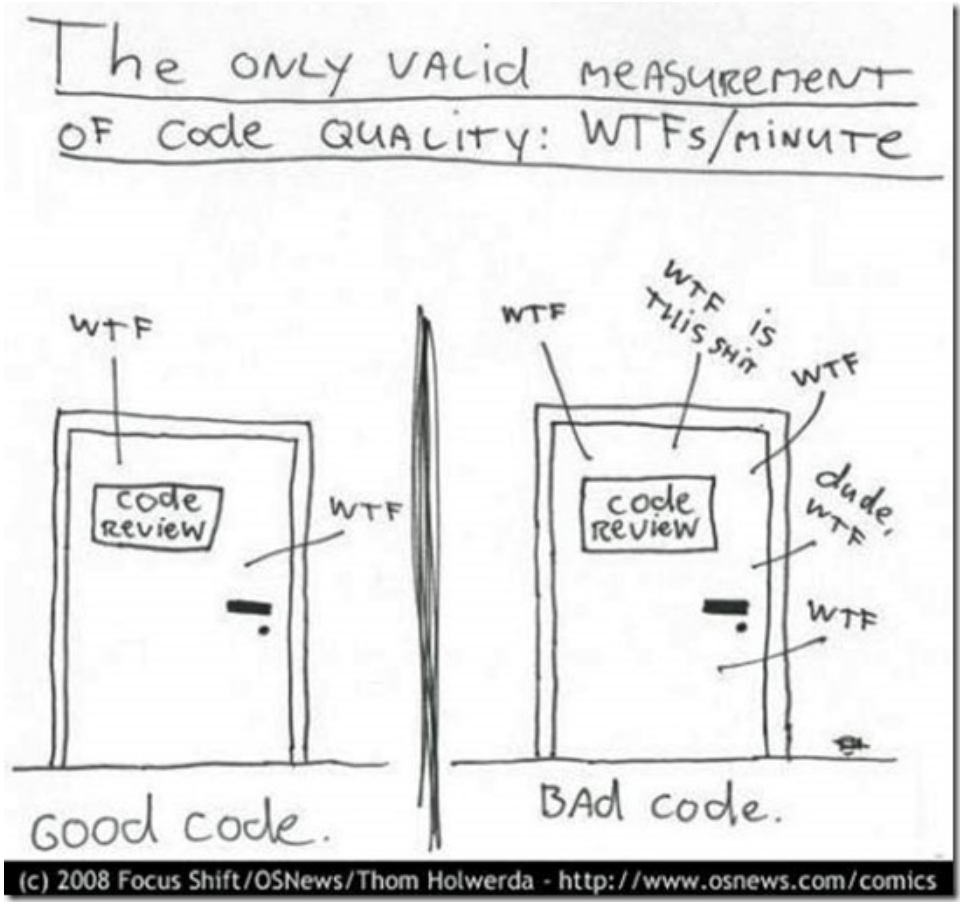
```
function res = f(base, num)
% Assign base to res
res = base
% loop from 2 to num
for i=2:num
    % multiply current res by base
    res=base*res;
end
```

vs

```
function res = pow(base, num)
% compute base^num by iterative multiply for baseline check
res = base
for i=2:num
    res=res*base;
end
```

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006561

7. Document the purpose and design, not the implementation

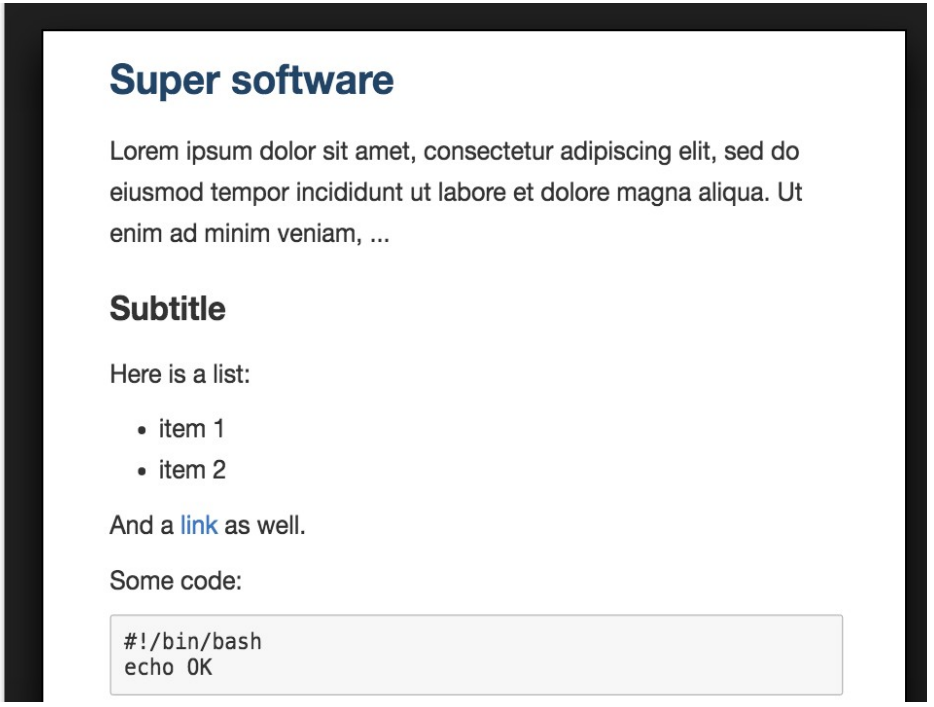


again

7. Document the purpose and design, not the implementation

Learn Markdown or RestructuredText

```
Super software
==========

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do
eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut
enim ad minim veniam, ...

Subtitle
---------

Here is a list:

  - item 1
  - item 2

And a [link](http://www.google.com) as well.

Some code:

    #!/bin/bash
```



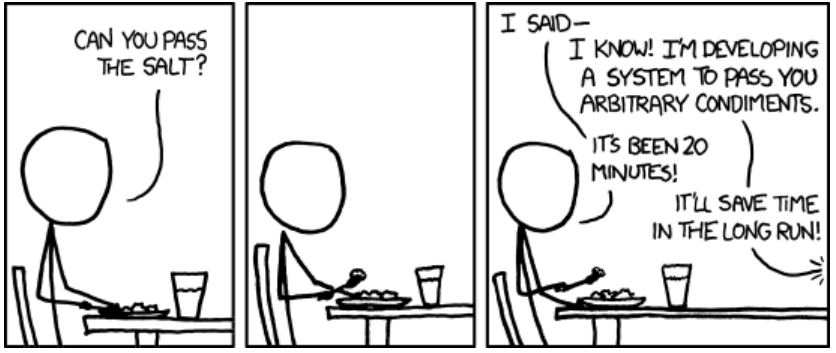
https://daringfireball.net/projects/markdown/

8. Optimize when it works already

- Do not try to make it fast when it is not working yet

  (I can always make your code faster by commenting out some of it)

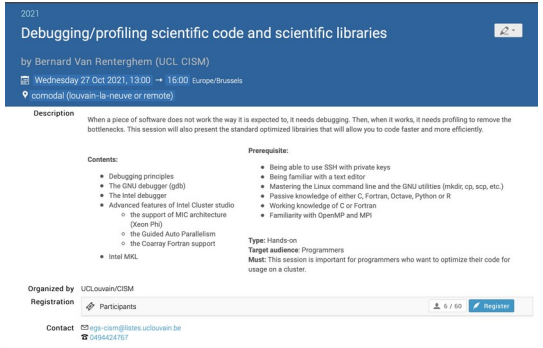
- Do not try to make it universal for future needs at the beginning



https://xkcd.com/974/

8. Optimize when it works already

Use a profiler


2021

Debugging/profiling scientific code and scientific libraries

by Bernard Van Renterghem (UCL CISM)

Wednesday 27 Oct 2021, 13:00 → 16:00 Europe/Brussels

comodal (louvain-la-neuve or remote)

Description

When a piece of software does not work the way it is expected to, it needs debugging. Then, when it works, it needs profiling to remove the bottlenecks. This session will also present the standard optimized libraries that will allow you to code faster and more efficiently.

Contents:

- Debugging principles
- The GNU debugger (gdb)
- The Intel debugger
- Advanced features of Intel Cluster studio
  - the support of MIC architecture (Xeon Phi)
  - the Guided Auto Parallelism
  - the Coarray Fortran support
- Intel MKL

Prerequisite:

- Being able to use SSH with private keys
- Being familiar with a text editor
- Mastering the Linux command line and the GNU utilities (mkdir, cp, scp, etc.)
- Passive knowledge of either C, Fortran, Octave, Python or R
- Working knowledge of C or Fortran
- Familiarity with OpenMP and MPI

Type: Hands-on

Target audience: Programmers

Must: This session is important for programmers who want to optimize their code for usage on a cluster.

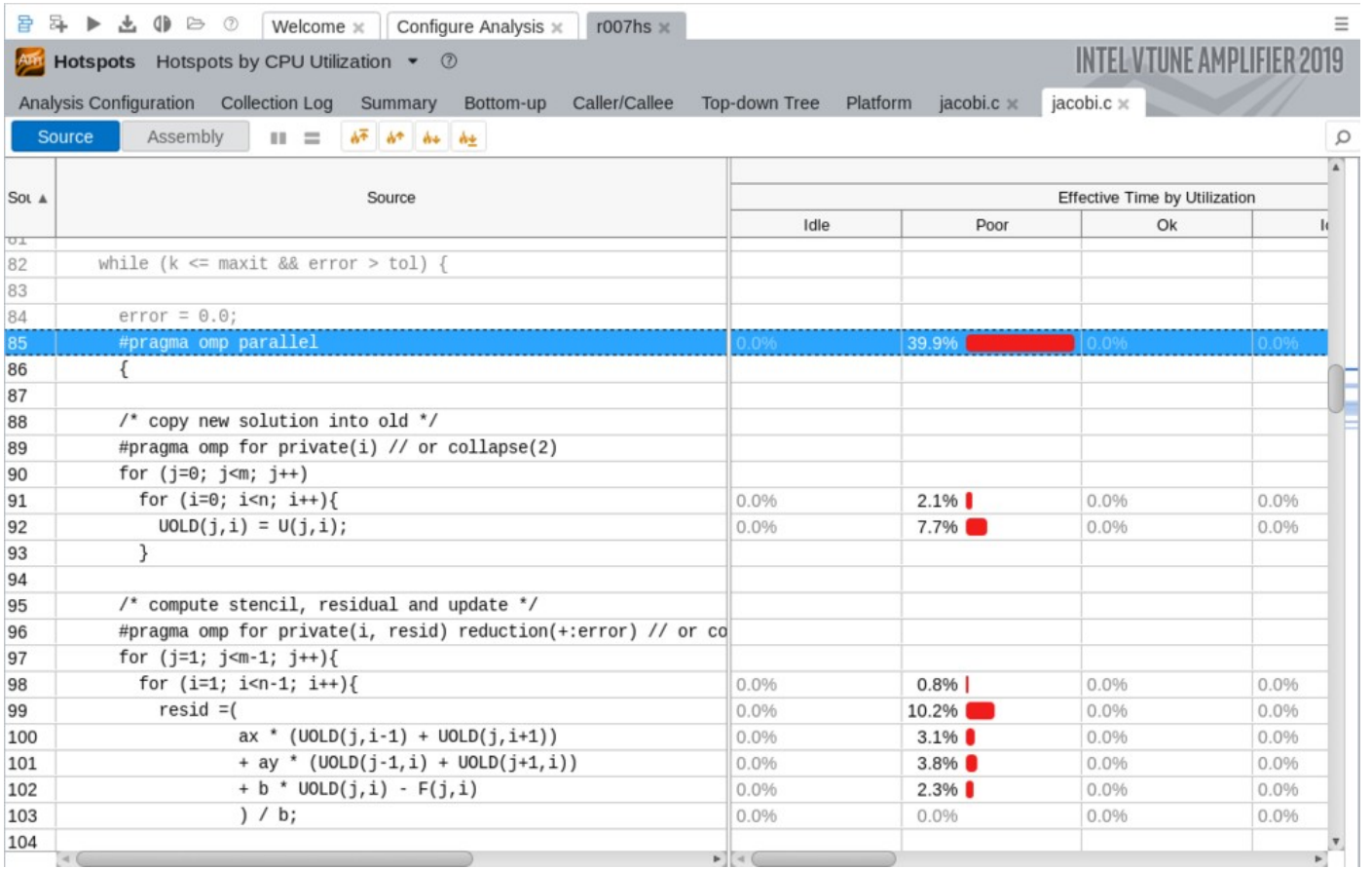
Organized by UCLouvain/CISM

Registration Participants 6 / 60 Register

Contact egs-cism@listes.uclouvain.be

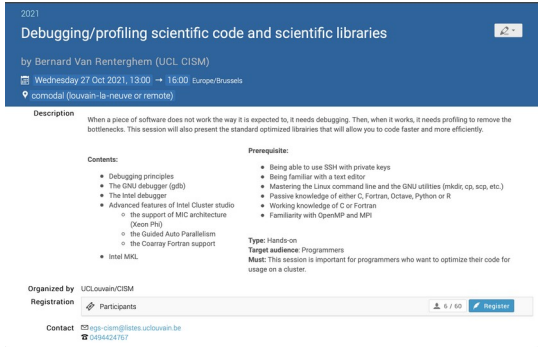
0494424767

Incorporate benchmarks in your tests



9. Debug cleverly

Use a debugger


2021

Debugging/profiling scientific code and scientific libraries

by Bernard Van Renterghem (UCL CISM)

Wednesday 27 Oct 2021, 13:00 → 16:00 Europe/Brussels

comodal (louvain-la-neuve or remote)

Description

When a piece of software does not work the way it is expected to, it needs debugging. Then, when it works, it needs profiling to remove the bottlenecks. This session will also present the standard optimized libraries that will allow you to code faster and more efficiently.

Contents:

- Debugging principles
- The GNU debugger (gdb)
- The Intel debugger
- Advanced features of Intel Cluster studio
  - the support of MIC architecture (Xeon Phi)
  - the Guided Auto Parallelism
  - the Coarray Fortran support
- Intel MKL

Prerequisite:

- Being able to use SSH with private keys
- Being familiar with a text editor
- Mastering the Linux command line and the GNU utilities (mkdir, cp, scp, etc.)
- Passive knowledge of either C, Fortran, Octave, Python or R
- Working knowledge of C or Fortran
- Familiarity with OpenMP and MPI

Type: Hands-on

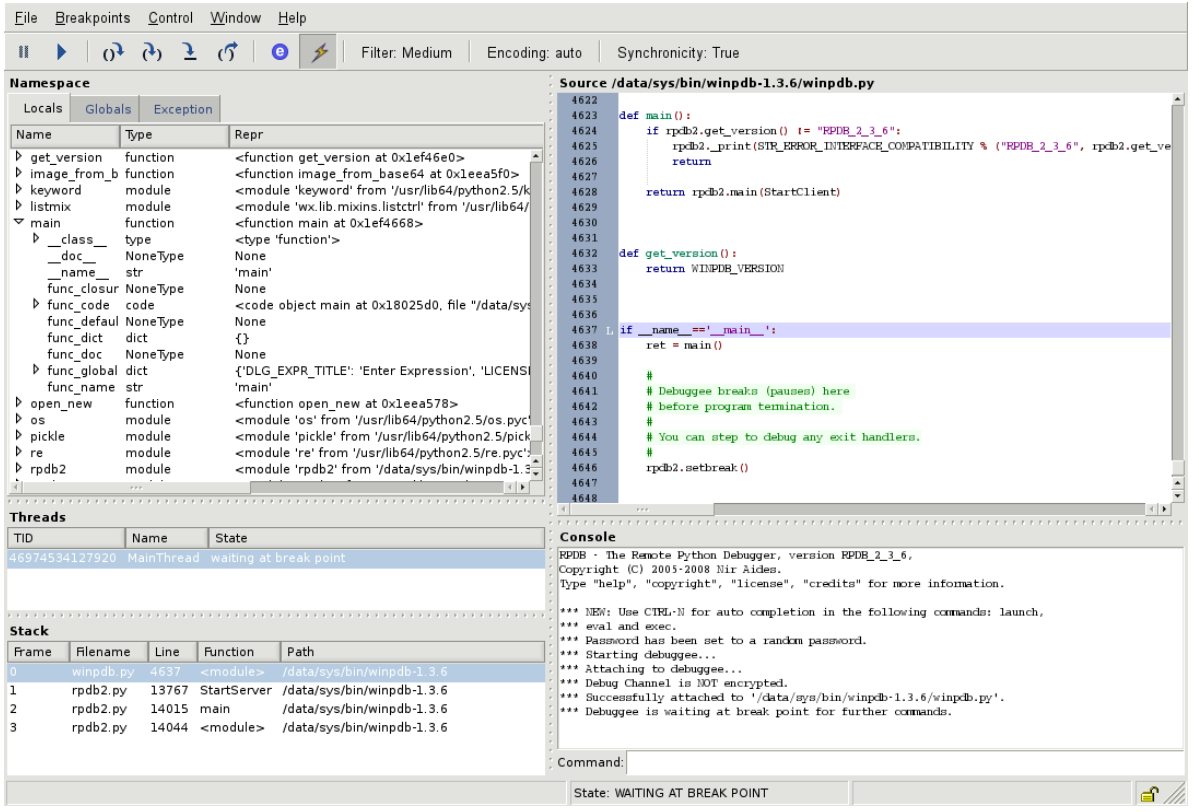
Target audience: Programmers

Must: This session is important for programmers who want to optimize their code for usage on a cluster.

Organized by UCLouvain/CISM

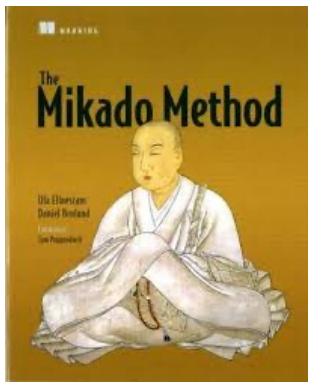
Registration Participants 6 / 60 Register

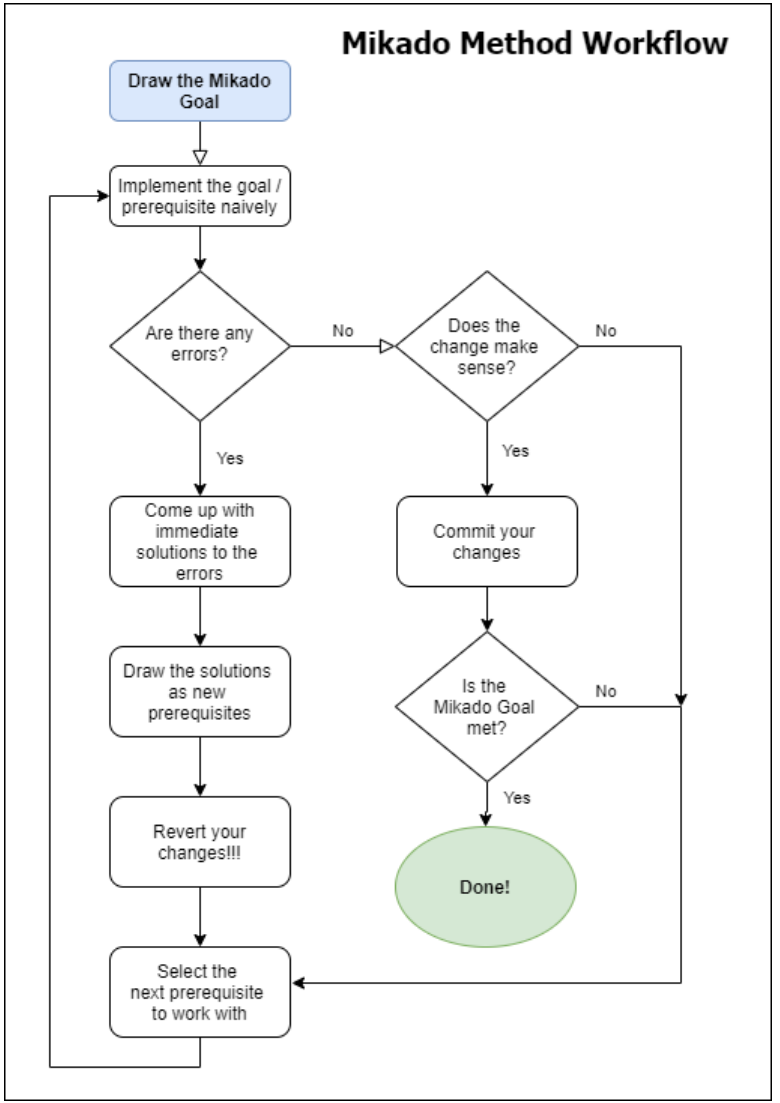
Contact egs-cism@listes.uclouvain.be 0494424767



9. Debug cleverly

Use a method



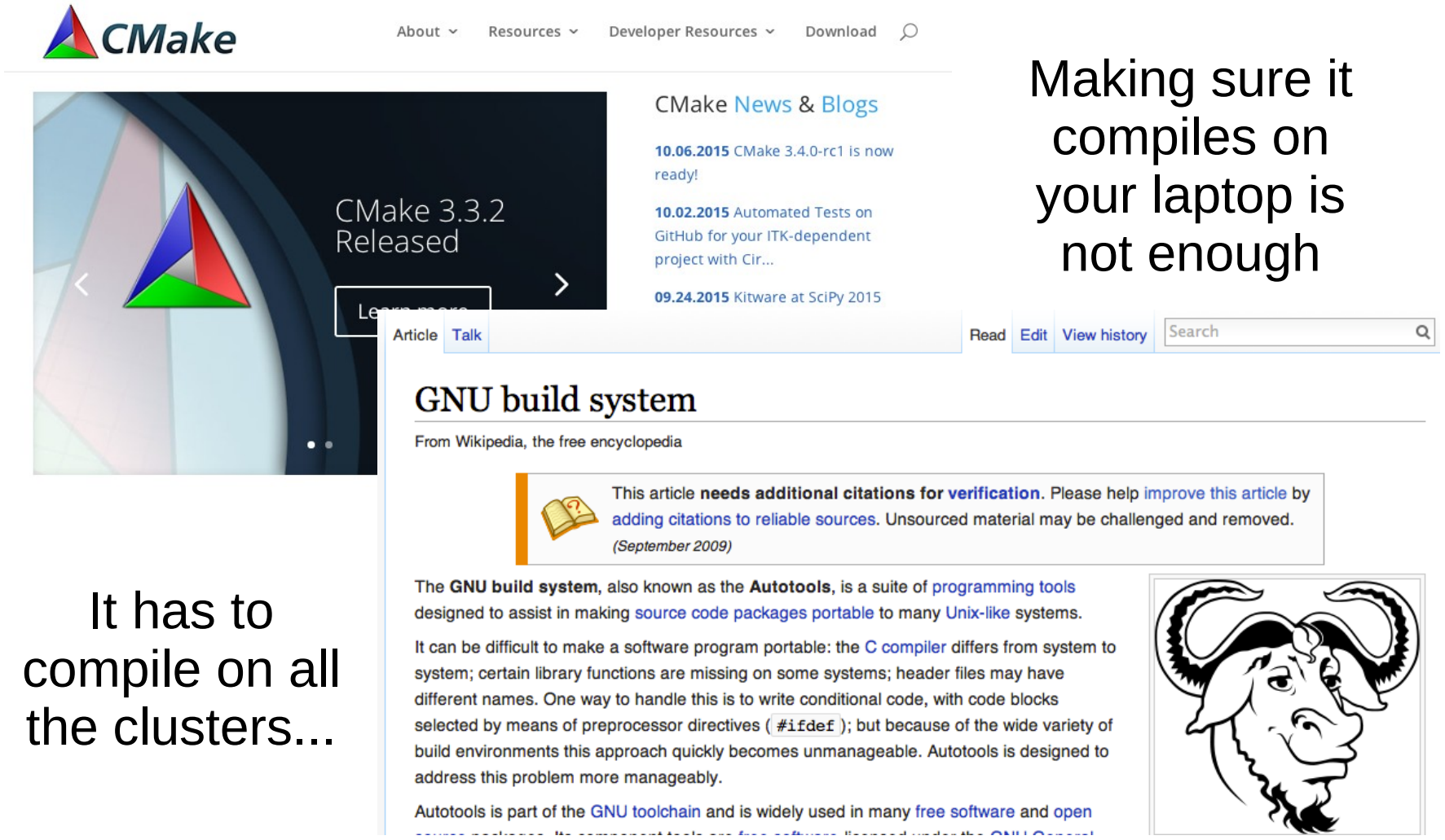


The Mikado Method, O. Ellnestam & D. Brolund, Manning, 2014

“… to make them … **shareable** …, *efficiently*”

1. Automate the compiling process
2. Learn about containers
3. License your code

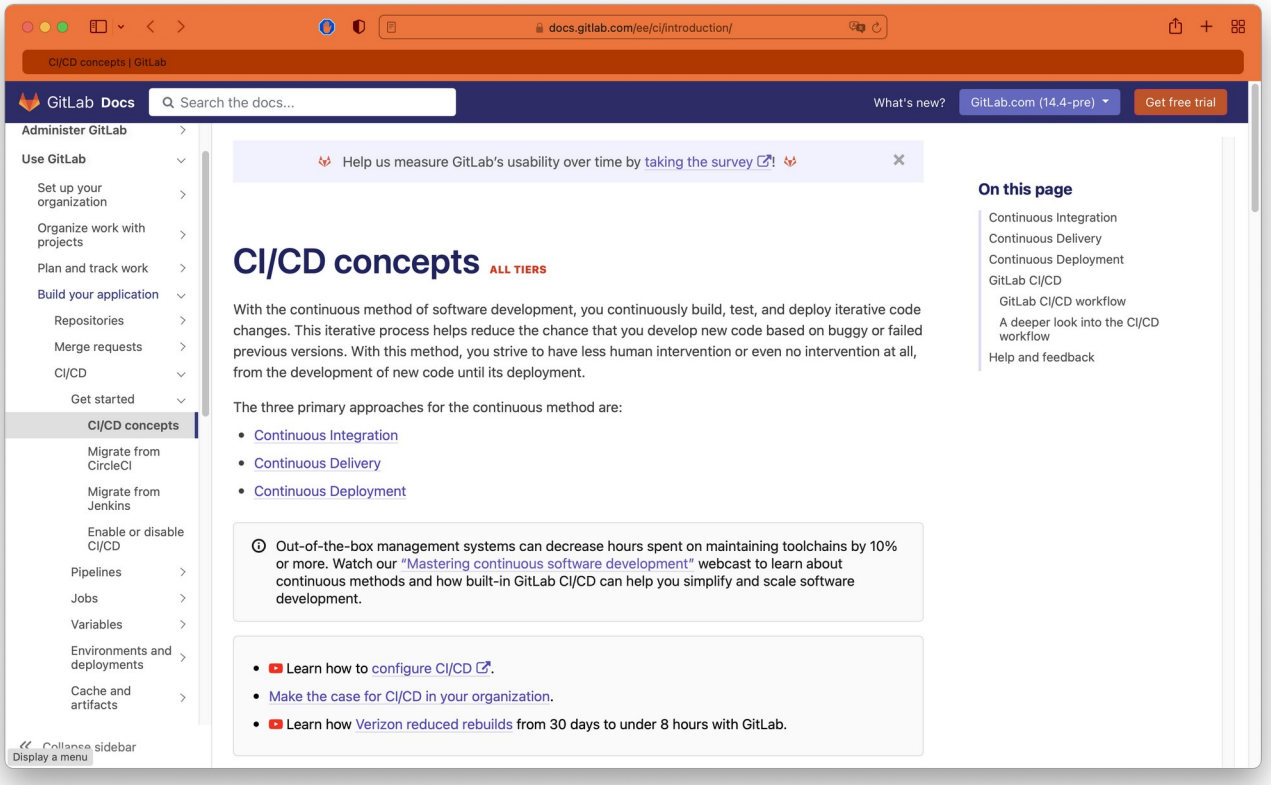
1. Automate the compiling process



Making sure it compiles on your laptop is not enough

It has to compile on all the clusters...

1. Automate the compiling process



https://docs.gitlab.com/ee/ci/introduction/

2. Learn about containers

2021

Packaging software in portable containers with Singularity

by Dr Olivier Mattelaer (UCLouvain/CISM)

Tuesday 16 Nov 2021, 09:45 → 12:45 Europe/Brussels

comodal (louvain-la-neuve or remote)

Description

Singularity is a container solution for HPC. Containers help with reproducibility as they nicely package software and data dependencies, along with libraries that are needed. It allows users to install and run software that required root access to be installed on clusters where they only have regular user permissions. The rationale is to perform all the software installation in a container image (a kind of lightweight virtual machine, that can use a different Linux distribution than the one on the compute nodes!) on a machine where you have root access and then transfer and run that image on the machine on which you do not have root access. Images can be built from recipes shared by others, and from recipes made for Docker, the leader container solution outside the HPC world.

Contents:

- Container concepts and benefits
- Starting a Singularity container on the cluster
- Accessing the cluster filesystems
- Building a container image from a recipe
- Building a container image from scratch
- Singularity hub

Prerequisite:

- Being able to use SSH with private keys
- Being familiar with a text editor
- Mastering the Linux command line and the GNU utilities (mkdir, cp, scp, etc.)
- Basic knowledge of Linux system administration

Type: Hands-on
Target audience: Advanced user
Must: This session is a must-have for anyone dealing with software that only installs on Ubuntu...

Organized by UCLouvain/CISM

Registration Participants 4 / 60 Register

Contact egs-cism@listes.uclouvain.be
0494424767

3. License your code: Why?

- **Commercial reason** :
  - you want to make money out of it – forbid distribution
    – forbid reverse engineering
- **Scientific reason** :
  - you want to it to be used and get citations
    – you need to allow usage, and/or modification, etc.
    – you require others to cite your work
  - you want to protect yourself from liability claims

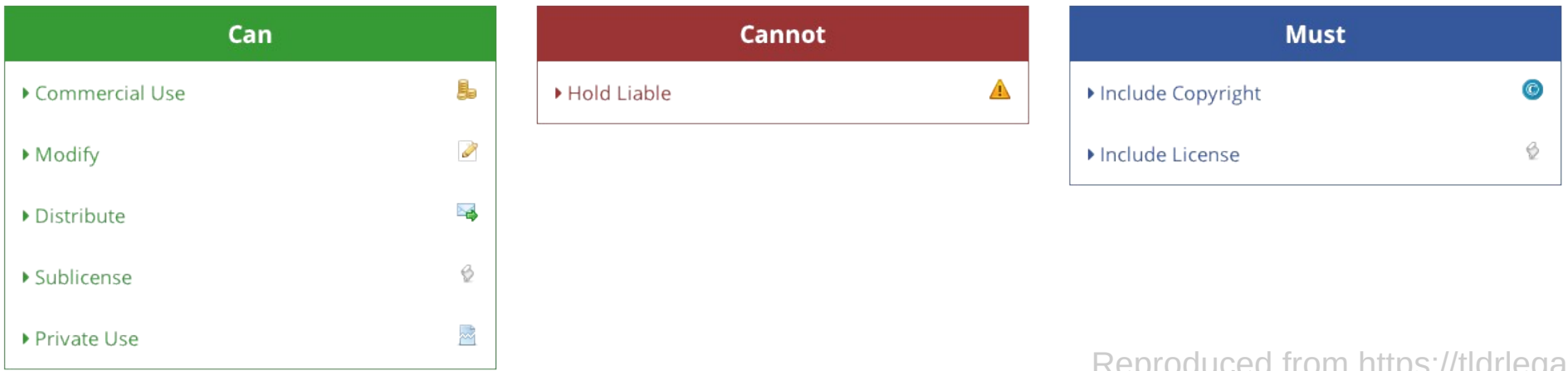
3. License your code: e.g. MIT

Copyright (c) <year> <copyright holders>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Softwarre.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.



3. License your code: e.g. BSD, GPL

BSD

Can	Cannot	Must
Commercial Use	Use Trademark	Include Copyright
Modify	Hold Liable	Include License
Distribute		
Place Warranty		

GPL

Can	Cannot	Must
Commercial Use	Sublicense	Include Original
Modify	Hold Liable	State Changes
Distribute		Disclose Source
Place Warranty		Include License
Use Patent Claims		Include Copyright
		Include Install Instructions

3. License your code: finding help

LIEU – Network of Knowledge Transfer offices

Coordinator

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
TECHNOLOGY TRANSFER OFFICE (LTTO)
Sébastien ADAM
✆ +32 10 47 24 43
sebastien.adam@uclouvain.be

UNIVERSITÉ DE LIÈGE
TECHNOLOGY TRANSFER OFFICE
Jérémie FAYS
✆ + 32 4 349 85 21
j.fays@uliege.be

UNIVERSITÉ LIBRE DE BRUXELLES
TECHNOLOGY TRANSFER OFFICE (ULB-TTO)
Kévin DEPLUS
✆ +32 650 23 15
kevin.deplus@ulb.ac.be

UNIVERSITÉ DE NAMUR
TECHNOLOGY TRANSFER OFFICE
Nathalie FABRY
✆ +32 81 72 53 36
nathalie.fabry@unamur.be

UNIVERSITÉ DE LIÈGE
TECHNOLOGY TRANSFER OFFICE
Thi Tuyet Minh NGUYEN
✆ +32 349 85 10
ttm.nguyen@uliege.be

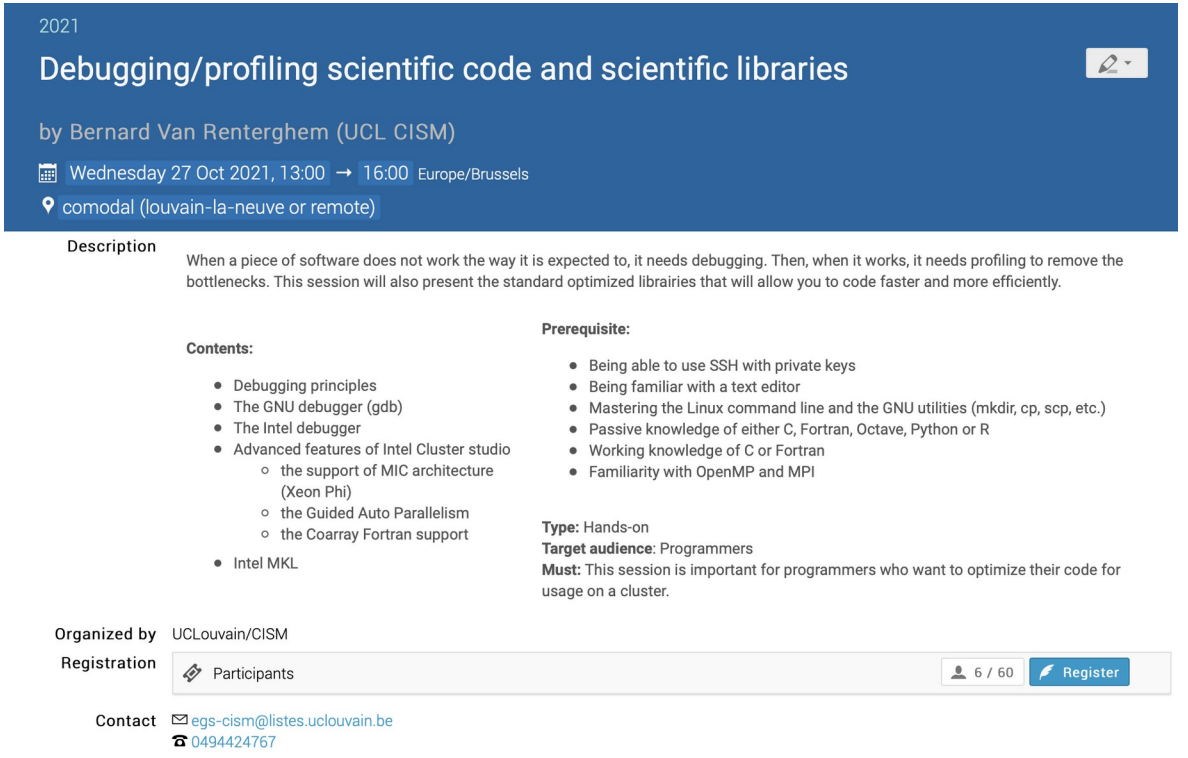
UNIVERSITÉ DE MONS
TECHNOLOGY TRANSFER OFFICE
Sandrine BROGNAUX
✆ +32 65 37 47 97
sandrine.brognaux@umons.ac.be

Slide credit: Sébastien ADAM

“… to make them … **fast** …, *efficiently*”

1. Use optimized libraries
2. Choose the right storage system
3. Think parallel from the start
4. Integrate checkpoint/restart from the start

1. Use optimized libraries



2021

Debugging/profiling scientific code and scientific libraries

by Bernard Van Renterghem (UCL CISM)

Wednesday 27 Oct 2021, 13:00 → 16:00 Europe/Brussels

comodal (louvain-la-neuve or remote)

Description

When a piece of software does not work the way it is expected to, it needs debugging. Then, when it works, it needs profiling to remove the bottlenecks. This session will also present the standard optimized librairies that will allow you to code faster and more efficiently.

Contents:

- Debugging principles
- The GNU debugger (gdb)
- The Intel debugger
- Advanced features of Intel Cluster studio
  - the support of MIC architecture (Xeon Phi)
  - the Guided Auto Parallelism
  - the Coarray Fortran support
- Intel MKL

Prerequisite:

- Being able to use SSH with private keys
- Being familiar with a text editor
- Mastering the Linux command line and the GNU utilities (mkdir, cp, scp, etc.)
- Passive knowledge of either C, Fortran, Octave, Python or R
- Working knowledge of C or Fortran
- Familiarity with OpenMP and MPI

Type: Hands-on
Target audience: Programmers
Must: This session is important for programmers who want to optimize their code for usage on a cluster.

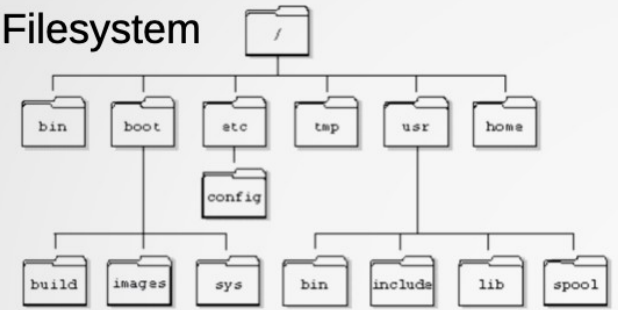
Organized by UCLouvain/CISM

Registration Participants 6 / 60 Register

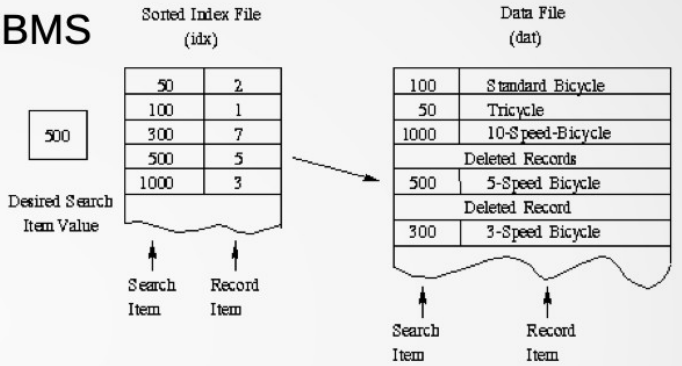
Contact egs-cism@listes.uclouvain.be
0494424767

2. Choose the right storage system

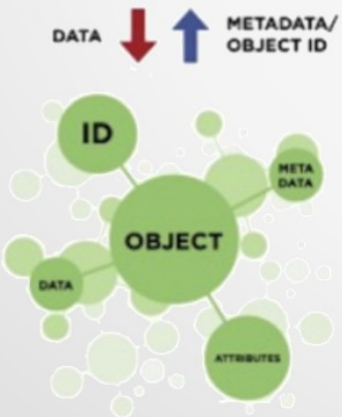
Filesystem



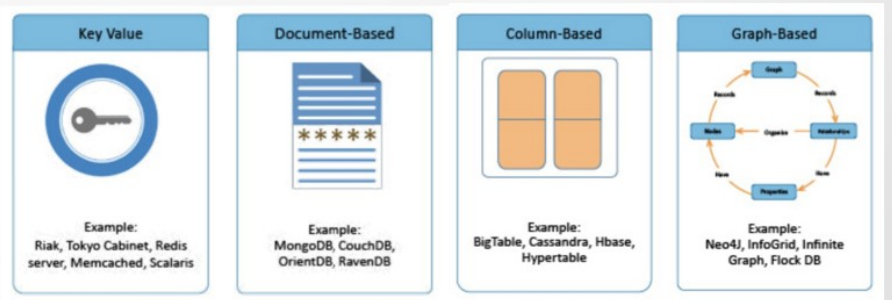
RDBMS



Objects store



NoSQL



3. Think parallel from the start

1. Identify data flows and independent tasks
2. Make data decomposition easy
3. Make work decomposition easy

```
data = [(x1,y1), (x2,y2), .., (x10,y10)]
for i = 1 .. 10
  s[i] = ( data[xi]  + data[yi] )
  ss[i] = ( data[xi]^2  + data[yi]^2 )
end
```

```
begin=0, end=10
data = [(x1,y1), (x2,y2), .., (x10,y10)]
for i = begin .. end
  s[i] = ( data[xi]  + data[yi] )
end
for i = begin .. end
  ss[i] = ( data[xi]^2  + data[yi]^2 )
end
```

3. Think parallel from the start

- 10 Nov Orian Louant, "Directive Based Parallel programming on GPU"
- 10 Nov Olivier Mattelaer, "Parallel programming on GPU with CUDA"
- 09 Nov Orian Louant, "Parallel programming with OpenMP"
- 29 Oct Orian Louant, "Parallel programming with MPI (Part II)"
- 29 Oct Orian Louant, "Parallel programming with MPI (Part I)"

4. Integrate checkpoint/restart from the start

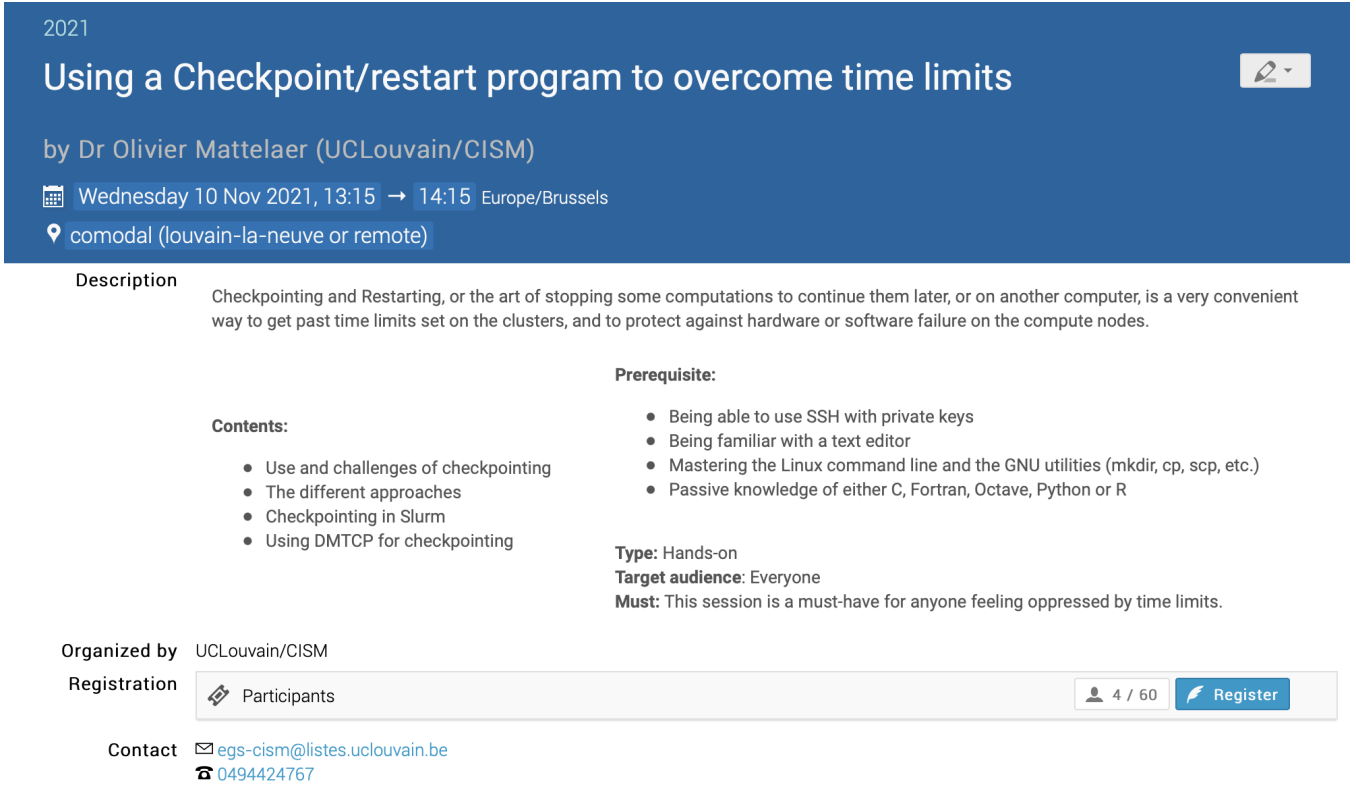
1. Allow starting from a non-initial state
2. Save variables to disk frequently

```
data = [(x1,y1), (x2,y2), .., (x10,y10)]
for i = 1 .. 10
  res[i] = ( data[xi]^2  + data[yi]^2 )
end
```

```
if exists(i) and exists(res)
    begin=load(i)
    res=load(res)
else
    begin = 1
end=10
data = [(x1,y1), (x2,y2), .., (x10,y10)]
for i = begin .. end
  res[i] = ( data[xi]^2  + data[yi]^2 )
  save(res, i)
end
```

4. Integrate checkpoint/restart from the start

Because sometimes your code will not be fast enough….



2021

Using a Checkpoint/restart program to overcome time limits

by Dr Olivier Mattelaer (UCLouvain/CISM)

Wednesday 10 Nov 2021, 13:15 → 14:15 Europe/Brussels

comodal (louvain-la-neuve or remote)

Description

Checkpointing and Restarting, or the art of stopping some computations to continue them later, or on another computer, is a very convenient way to get past time limits set on the clusters, and to protect against hardware or software failure on the compute nodes.

Contents:

- Use and challenges of checkpointing
- The different approaches
- Checkpointing in Slurm
- Using DMTCP for checkpointing

Prerequisite:

- Being able to use SSH with private keys
- Being familiar with a text editor
- Mastering the Linux command line and the GNU utilities (mkdir, cp, scp, etc.)
- Passive knowledge of either C, Fortran, Octave, Python or R

Type: Hands-on
Target audience: Everyone
Must: This session is a must-have for anyone feeling oppressed by time limits.

Organized by UCLouvain/CISM

Registration Participants 4 / 60 Register

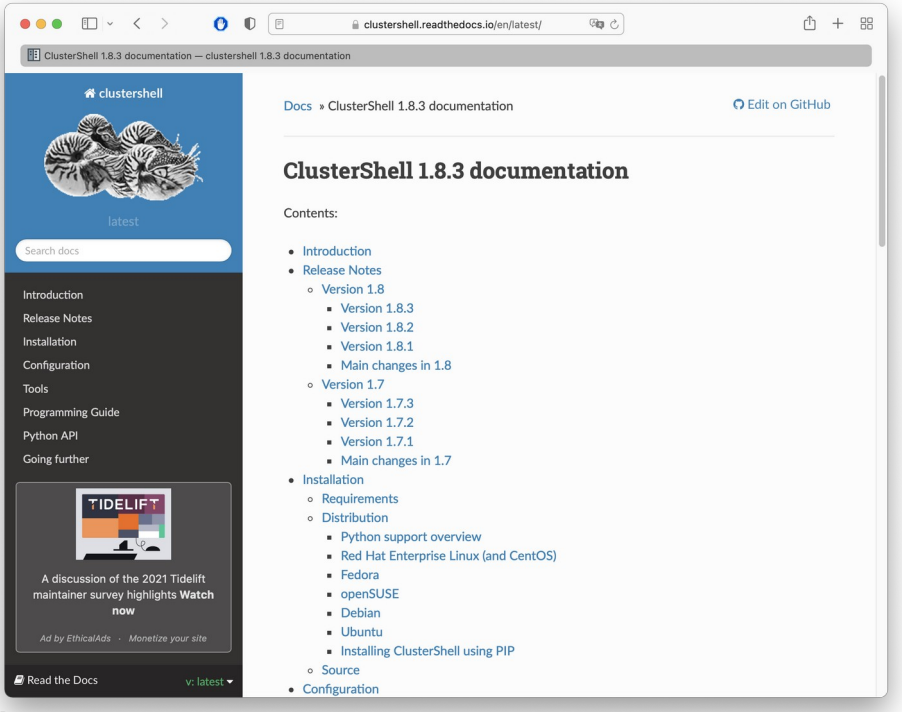
Contact egs-cism@listes.uclouvain.be
0494424767

“ …, *efficiently*”

1. Perform “multi-host” SSH
2. Master configuration management
3. Use terminal multiplexing
4. Install software like a boss
5. BACKUPS!

1. Perform “multi-host” SSH

```
❯ clush -Bw lemaitre3,hercules,nic5,dragon2 "emacs --version"
---------------
dragon2,hercules,lemaitre3 (3)
---------------
GNU Emacs 24.3.1
Copyright (C) 2013 Free Software Foundation, Inc.
GNU Emacs comes with ABSOLUTELY NO WARRANTY.
You may redistribute copies of Emacs
under the terms of the GNU General Public License.
For more information about these matters, see the file named COPYING.
---------------
nic5
---------------
GNU Emacs 26.1
Copyright (C) 2018 Free Software Foundation, Inc.
GNU Emacs comes with ABSOLUTELY NO WARRANTY.
You may redistribute copies of GNU Emacs
under the terms of the GNU General Public License.
For more information about these matters, see the file named COPYING.
❯ clush -Bw lemaitre3,hercules,nic5,dragon2 "scontrol version"
---------------
dragon2,lemaitre3 (2)
---------------
slurm 20.02.7
---------------
hercules
---------------
slurm 20.02.6
---------------
nic5
---------------
slurm 20.02.3
❯ clush -w lemaitre3,hercules,nic5,dragon2 "squeue -tPD | wc -l"
nic5: 1420
lemaitre3: 288
dragon2: 145
hercules: 102
❯
```



https://clustershell.readthedocs.io/en/latest/

2. Master configuration management


```
❯ ansible -i lemaitre3,nic5 'all' -m lineinfile -a "dest=myfile line='Contents' create=true"
nic5 | CHANGED => {
    "ansible_facts": {
        "discovered_interpreter_python": "/usr/libexec/platform-python"
    },
    "backup": "",
    "changed": true,
    "msg": "line added"
}
lemaitre3 | SUCCESS => {
    "ansible_facts": {
        "discovered_interpreter_python": "/usr/bin/python"
    },
    "backup": "",
    "changed": false,
    "msg": ""
}
```

2. Master configuration management


```
❯ cat inventory playbook.yml myfile
```

File: **inventory**

```
1  [all]
2
3  lemaitre3 short_name="lm3"
4  nic5      short_name="nic5"
```

File: **playbook.yml**

```
1  ---
2  - hosts:
3      - lemaitre3
4      - nic5
5    tasks:
6      - name: Upload templated file
7        template: src=myfile dest=. mode=700
```

File: **myfile**

```
1  This cluster's short name is {{ short_name }}
```

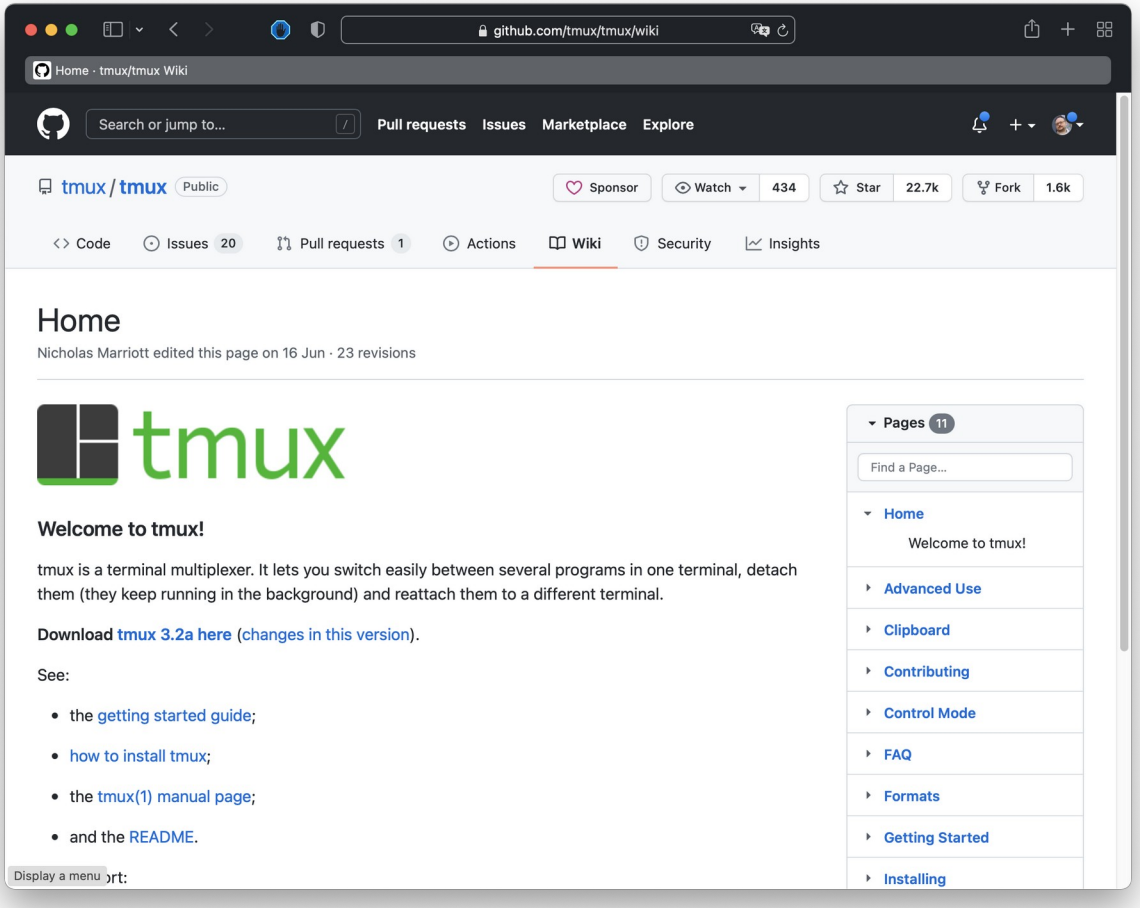
https://www.ansible.com

2. Master configuration management


```
❯ ansible-playbook -i inventory playbook.yml --diff

PLAY [lemaitre3,nic5] ***********************************************************************************************

TASK [Gathering Facts] **********************************************************************************************
ok: [nic5]
ok: [lemaitre3]

TASK [Upload templated file] ****************************************************************************************
--- before: ./myfile
+++ after: /Users/dfrancois/.ansible/tmp/ansible-local-40594k4ng0q9q/tmpd689k7ap/myfile
@@ -1 +1 @@
-Contents
+This cluster's short name is nic5

changed: [nic5]
--- before: ./myfile
+++ after: /Users/dfrancois/.ansible/tmp/ansible-local-40594k4ng0q9q/tmpa5narr9m/myfile
@@ -1 +1 @@
-Contents
+This cluster's short name is lm3

changed: [lemaitre3]

PLAY RECAP **********************************************************************************************************
lemaitre3                  : ok=2    changed=1    unreachable=0    failed=0    skipped=0    rescued=0    ignored=0
nic5                       : ok=2    changed=1    unreachable=0    failed=0    skipped=0    rescued=0    ignored=0

❯ clush -w lemaitre3,nic5 "cat myfile"
nic5: This cluster's short name is nic5
lemaitre3: This cluster's short name is lm3
```

3. Use terminal multiplexing

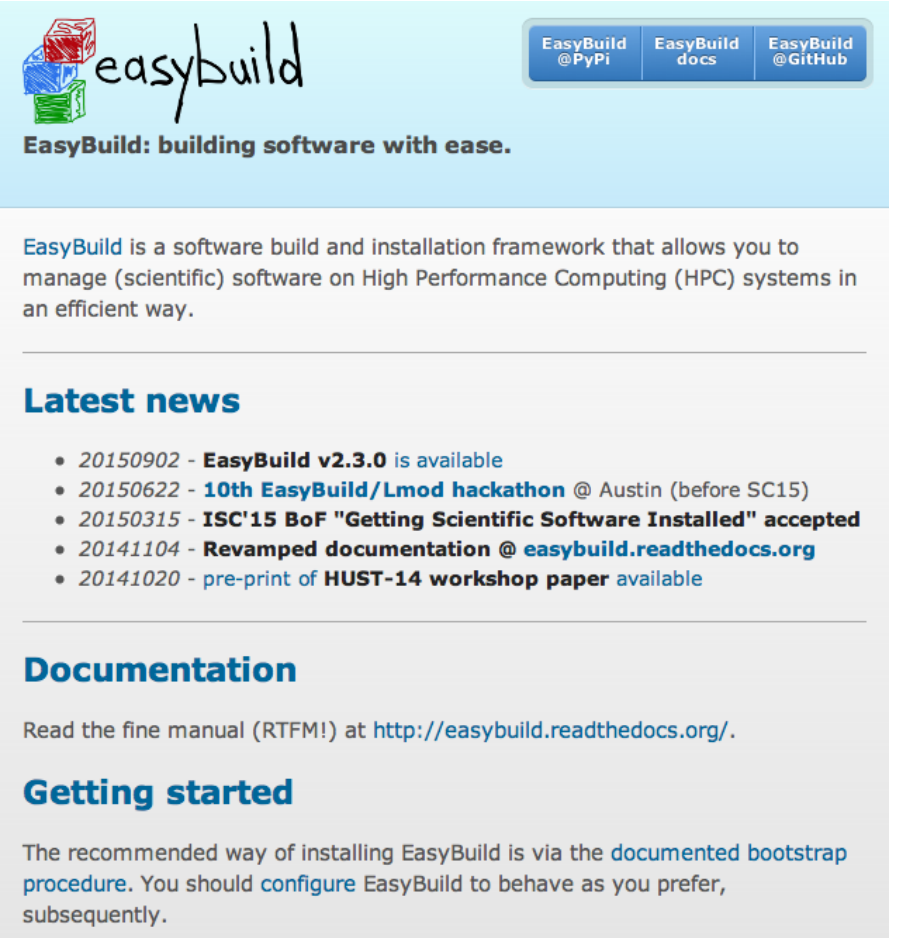
Do not let SSH disconnections harm your workflow
(and much more)



https://github.com/tmux/tmux/wiki

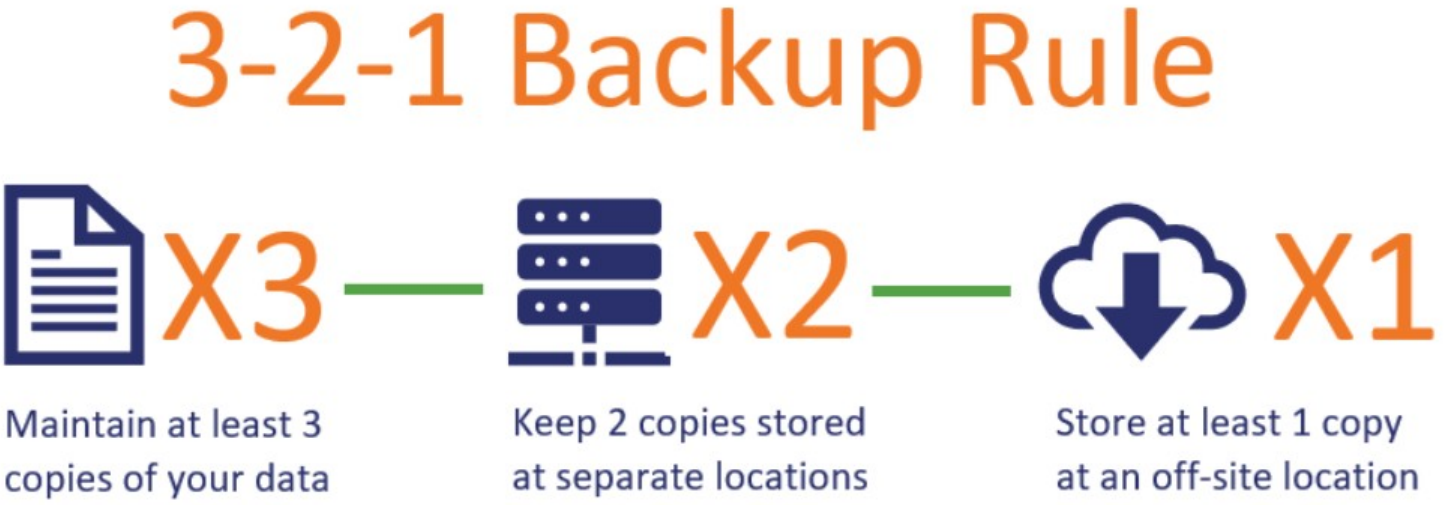
4. Install software like a boss

```
[dfr@lemaitre3 ~]$ eb --search emacs
== found valid index for /usr/easybuild/easyconfigs, so using it...
== found valid index for /usr/easybuild/easyconfigs, so using it...
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.3-GCC-4.8.3-bare.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.3-GCC-4.8.3.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.4-GCC-4.9.2.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.5-GCC-4.9.3-2.25.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.1-foss-2016a.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.3-GCCcore-6.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.3-GCCcore-6.4.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.3-GCCcore-7.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-26.3-GCCcore-8.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-27.1-GCCcore-9.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-27.1-GCCcore-10.2.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.3-GCC-4.8.3-bare.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.3-GCC-4.8.3.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.4-GCC-4.9.2.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-24.5-GCC-4.9.3-2.25.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.1-foss-2016a.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.3-GCCcore-6.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.3-GCCcore-6.4.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-25.3-GCCcore-7.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-26.3-GCCcore-8.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-27.1-GCCcore-9.3.0.eb
 * /usr/easybuild/easyconfigs/e/Emacs/Emacs-27.1-GCCcore-10.2.0.eb
```



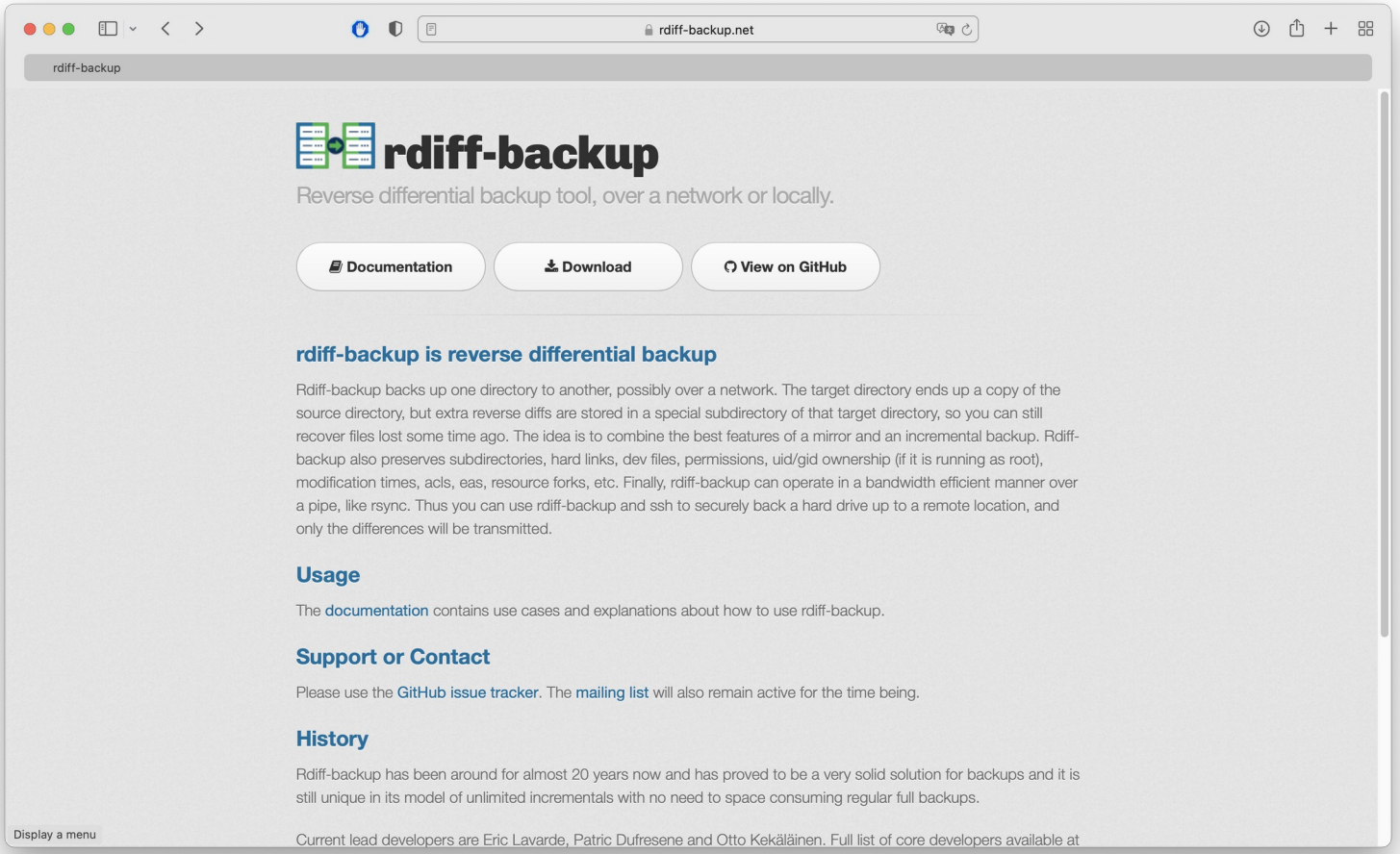
https://easybuild.readthedocs.io/en/latest/

5. BACKUPS!!!



https://securityboulevard.com/2020/05/3-2-1-backup-rule-the-rule-of-thumb-to-solve-your-data-loss-problems/

5. BACKUPS!!!



https://rdiff-backup.net

This was:

"A short catalog of *tools*
the professionals are using for
developing and **deploying** programs,
to make them **correct**, **maintainable**, **shareable**, and **fast**,
efficiently."

We discussed:

- good practices
- important choices
- useful tools
- practical references

The “Phillip test” (by Philip Guo)

12 simple questions
ordered by 'difficulty'
measures quality of organization
for research programming

If you do not score at least a 7
there is room for improvement
using the tools presented here

1. Do you have reliable ways of taking, organizing, and reflecting on notes as you're working?
2. Do you have reliable to-do lists for your projects?
3. Do you write scripts to automate repetitive tasks?
4. Are your scripts, data sets, and notes backed up on another computer?
5. Can you quickly identify errors and inconsistencies in your raw data sets?
6. Can you write scripts to acquire and merge together data from different sources and in different formats?
7. Do you use version control for your scripts?
8. If you show analysis results to a colleague and they offer a suggestion for improvement, can you adjust your script, re-run it, and produce updated results within an hour?
9. Do you use `assert` statements and test cases to sanity check the outputs of your analyses?
10. Can you re-generate any intermediate data set from the original raw data by running a series of scripts?
11. Can you re-generate all of the figures and tables in your research paper by running a single command?
12. If you got hit by a bus, can one of your lab-mates resume your research where you left off with less than a week of delay?

Work faster & more reliably