Parallel Computing for Data Science

Introduction

Jairo Cugliari

Master Informatique

Parcours Data Mining

Université Lumière Lyon 2

Université de Lyon

Organisation pédagogique

Organisation

- Chaque séance : 1h CM + 2h Lab

Responsable

- Jairo Cugliari (Jairo.Cugliari@univ-lyon2.fr)

MCC

- Projet / Cours / Dossier en groupes (50%)
- Examen individuel (avec ordinateur 50%)

Ressources

Books

- N. Matloff. `Parallel Computing for Data Science: With Examples in R, C++ and CUDA` (Chapman & Hall/CRC The R Series) Open textbook version
- M. Gorelick and I. Ozvald. `High Performance Python`. (O'Reilly)
- H. Wickley. `Advanced R` (Chapman & Hall/CRC The R Series) Open textbook version

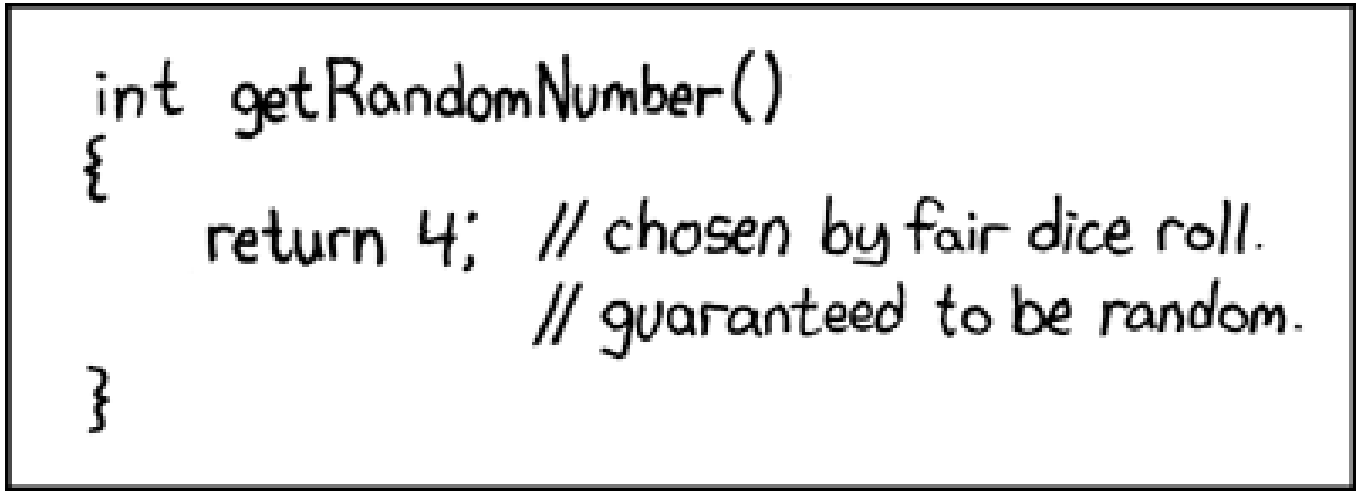
Ressources

E-learning

A tsunami of ressources, a glimpse:

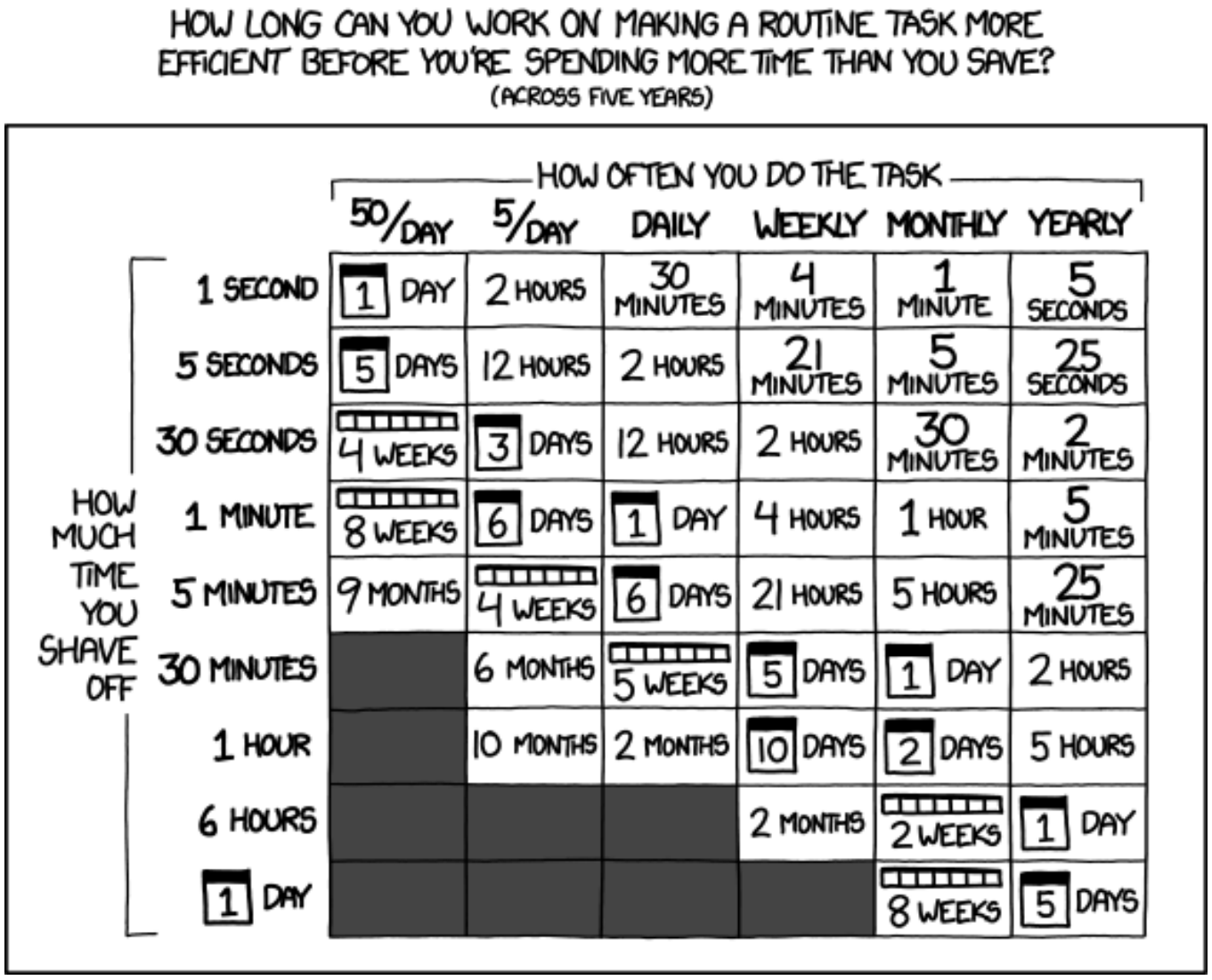
- Blog by J. VanderPlas
- Blog by T. Oliphant
- Course by V. Miele (in French)

Why ?



Source: xkcd.com

What ?


HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)
HOW OFTEN YOU DO THE TASK
50/DAY
5/DAY
DAILY
WEEKLY
MONTHLY
YEARLY
HOW MUCH TIME YOU SHAVE OFF
1 SECOND
1 DAY
2 HOURS
30 MINUTES
4 MINUTES
1 MINUTE
5 SECONDS
5 SECONDS
5 DAYS
12 HOURS
2 HOURS
21 MINUTES
5 MINUTES
25 SECONDS
30 SECONDS
4 WEEKS
3 DAYS
12 HOURS
2 HOURS
30 MINUTES
2 MINUTES
1 MINUTE
8 WEEKS
6 DAYS
1 DAY
4 HOURS
1 HOUR
5 MINUTES
5 MINUTES
9 MONTHS
4 WEEKS
6 DAYS
21 HOURS
5 HOURS
25 MINUTES
30 MINUTES
6 MONTHS
5 WEEKS
5 DAYS
1 DAY
2 HOURS
1 HOUR
10 MONTHS
2 MONTHS
10 DAYS
2 DAYS
5 HOURS
6 HOURS
2 MONTHS
2 WEEKS
1 DAY
1 DAY
8 WEEKS
5 DAYS

How ?


DO WE HAVE ANY ACTIONABLE ANALYTICS FROM OUR BIG DATA IN THE CLOUD?
YES, THE DATA SHOWS THAT MY PRODUCTIVITY PLUNGES WHENEVER YOU LEARN NEW JARGON.
MAYBE IN-MEMORY COMPUTING WILL ACCEL-ERATE YOUR APPLICA-TIONS.
PLUNGE, PLUNGE, PLUNGE.
Dilbert.com DilbertCartoonist@gmail.com
1-9-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick

Organisation

Main topics

1. Intro. Overview. Case study
2. Good practices for coding
3. Bottlenecks: seek & destroy
4. Some mathematics of parallel data analysis

Speakers

- Jairo Cugliari
- maybe an invited speaker also