Reproducible Experiments in Networking and Distributed Systems

Ecole automnale RESCOM de recherche du GDR RSD
October 2021

Fabrice Théoleyre   Arnaud Legrand
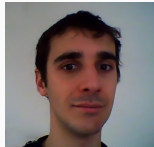
Bruno Donassolo   Guillaume Schreiner   Matthieu Simonin

Olivier Bonaventure   Simon Delamare   Philippe Bonnet   Morane Gruenpeter
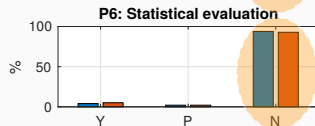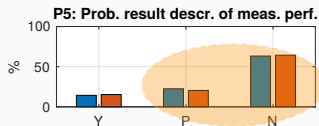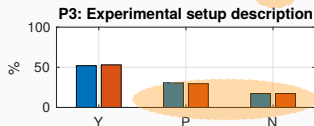
Toward Reproducible Computer Science Experiments ?

How are cloud performance currently obtained and reported?, Methodological Principles
for Reproducible Performance Evaluation in Cloud Computing, IEEE Trans. on Soft. Eng., July 2019
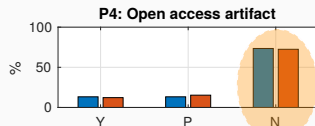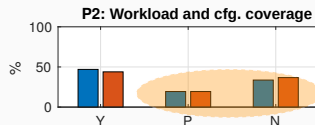
How are cloud performance currently obtained and reported?, Methodological Principles
for Reproducible Performance Evaluation in Cloud Computing, IEEE Trans. on Soft. Eng., July 2019

Research Reproducibility and Open Science has become
a major concern

Practices are quickly evolving

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, ... ?

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, ... ?

Bad computer/statistic/publication practices "broke science" 😊

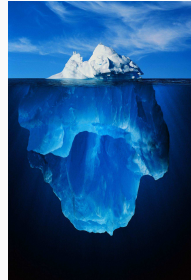- Ensure articles, data, and code are available
  - Open Access, Open Data, FAIR and DMP

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, ... ?

Bad computer/statistic/publication practices "broke science" 🙂

- Ensure articles, data, and code are available
    - Open Access, Open Data, FAIR and DMP
- *Ensure* the experiment design and analysis is not broken
    - ~~data torture and hacking~~, pre-registration, education, ...

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, ... ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles, data, and code are available
    - Open Access, Open Data, FAIR and DMP
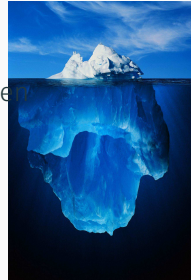- *Ensure* the experiment design and analysis is not broke
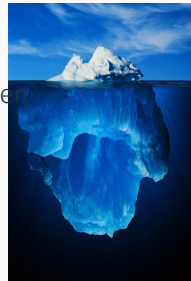   - ~~data torture and hacking~~, pre-registration, education, ...
- Explain how the computations unfold and why
    - Notebooks (that can quickly get out of control)
    - Simple computational workflows

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, … ?

Bad computer/statistic/publication practices "broke science" 🙂

- Ensure articles, data, and code are available
  - Open Access, Open Data, FAIR and DMP
- *Ensure* the experiment design and analysis is not broke~~n~~
  - ~~data torture and hacking~~, pre-registration, education, …
- Explain how the computations unfold and why
  - Notebooks (that can quickly get out of control)
  - Simple computational workflows
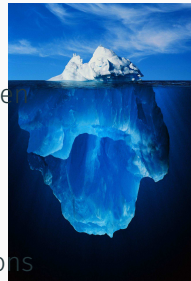- Allow to <u>inspect</u>, <u>rerun</u>, <u>make variations</u> on computations
  - Access to code, data, …

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, … ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles, data, and code are available
  - Open Access, Open Data, FAIR and DMP
- *Ensure* the experiment design and analysis is not broken
  - ~~data torture and hacking~~, pre-registration, education, …
- Explain how the computations unfold and why
  - Notebooks (that can quickly get out of control)
  - Simple computational workflows
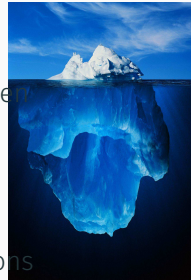- Allow to inspect, rerun, make variations on computations
  - Access to code, data, … options/parameters, environment, resources, testbeds ?

This requires:

- First class software engineering practices (not prototypes)
- Decent understanding of Statistics and Experimental Design

We will both

- Promote some best (???) practices   Hypothesis
- Apply them on a representative panel 😊   Experiment
- Evaluate all together how reproducible it is   Analysis

     We will send you a form afterward to get some feedback

Remember the important URLs:

- https://rsd-ecole.cnrs.fr/
- https://rsdschool21.gitlabpages.inria.fr/website/
- Mattermost to chat
- CodiMD to take notes