

Associate Team acronym: ReDaS**Period of activity:** 2020**Principal Investigator (Inria):** Guillaume Huard (Polaris)**Principal investigator (Partner Institution):** Lucas Mello Schnorr (UFRGS)

1 Future of the Associate Team

Would you like to pursue this Associate Team for one more year? Yes No

2 Website of the Associate Team

<https://associatedteam.gitlabpages.inria.fr/redas>

3 List of participants

Name	Team	Status	Main expertise
Guillaume Huard	INRIA	Associate Professor	Performance evaluation
Jean-Marc Vincent	INRIA	Associate Professor	Statistics & Perf. Evaluation
Arnaud Legrand	INRIA	Researcher	Reproducible Research
Lucas Mello Schnorr	UFRGS	Adjunct Professor	Performance Evaluation
João Comba	UFRGS	Associate Professor	Information Visualization
Alexis Janon	Polaris	PhD Candidate	HW Counters Co-Design
Lucas Nesi	UFRGS	PhD Candidate	Task-based Performance Analysis
Marcelo Miletto	UFRGS	Master student	Sparse Task-based Performance Analysis
Ana Veroneze	UFRGS	Master student	Architecture-dependent Load-balancing
Guilherme Alles	UFRGS	Master student	Phenology Visualization

4 Achievements and Planned activities

Due to the COVID-19 situation, several travels that had been planned for the last year did not happen. Nevertheless, despite this difficult situation, the exchanges between the participants remain active and the common works that were planned have been pursued :

- Guillaume Huard and Alexis Janon have been to UFRGS, respectively during 14 and 19 days, in december 2019. They have worked with Lucas Schnorr to automate the experiment setup, deployment, execution and collect for the Ondes3D application and to study a first batch of results.
- Lucas Schnorr has been to LIG during 10 days in March 2020. He has worked with Guillaume Huard and Arnaud Legrand on several topics that are detailed below. His mission has been funded by the FAPERGS project he coordinates.
- Lucas Nesi, a brazilian PhD student advised by Lucas Schnorr, visited Arnaud Legrand during November 2019. Lucas Nesi has been working on static data placement strategies to improve the load balancing of linear algebra operations in hybrid clusters. The first results have been published in the 2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS) [2]. Another investigation involves the Exascale GeoStatistics project (ExaGeoStat) application, in which Nesi studied the interplay of static data placement strategies for the different phases of

the application's main loop. The mid-term goal is to establish a co-tutelle agreement between UFRGS and the UGA with Nesi being advised by Arnaud Legrand starting at September 2021. We also expect a whole year visit of Lucas Nesi at POLARIS funded by a CAPES/Cofecub project currently on hold between UFRGS and the LIG.

- Guilherme Alles has defended his master thesis in July 2020. His work has been strongly influenced by Jean-Marc Vincent during his repeated visits to UFRGS in the past years. Guilherme worked on the phenology visualization analysis (mathematical model to tackled measurement uncertainty) and his results have been published in the Ecological Informatics journal [1] joint publication.
- Marcelo Miletto (master student) has been working under the guidance of Lucas Schnorr and Arnaud Legrand about trace visualization and task modeling to carry out performance analysis of task-based sparse factorization operations (using the `qr_mumps` solver as application). Results are so far encouraging and we expect to submit a manuscript to the Future Generation Computer Systems journal as soon as possible. Marcelo's masters also enabled him to include the `qr_mumps` solver in the RAFEM parallel application, with very good results. This part of his masters has been submitted to the PDP 2021 conference and is under review.
- Ana Veroneze has started her Brazilian master in March 2020. She is currently investigating the behavior of dynamic load imbalances in different architectures (Intel, AMD, ARM) using the Grid5000 infrastructure and the BRGM's Ondes3D application. This investigation works by exchanging several ideas with Guillaume Huard. We expect, using various measurements collected during the execution, to identify different interference schemes that depend on the type of architecture. Ana's project for her masters also involves the performance analysis of the training phase of deep learning frameworks when these are executed in a distributed manner. If everything works smoothly, we expect a visit of Ana Veroneze at the LIG by the end of 2021.
- Regarding the collaboration between Joao Comba and Sihem Amer-Yahia, Cícero Pahins was expected to spent a year at Grenoble starting June 2020. But, due to the COVID-19 pandemic, the visit has been postponed. The collaboration effort, which has enabled the publication of some results [5] [3] [4], is ongoing in a virtual fashion.

In 2021, if the sanitary situation improves sufficiently, we expect to resume scientific missions between partners. In particular, Lucas Schnorr is expected to spent ≈ 10 days in Grenoble in the beginning of November 2021 (funded by the CAPES/Brafitec E4S project), and Ana Veroneze is expected to spent at least one month in Grenoble to work with Guillaume Huard (funded by the Associated Team). Afterwards, we will plan visits of Arnaud Legrand and Guillaume Huard to UFRGS.

5 Budget request for the coming year

Description	Budget
3 air tickets (1200€ each)	3600 €
1-month accomodation expenses in France (for a Brazilian student)	1600 €
33 accomodation expenses for one person in Brazil (112 € per day)	3696 €
Workshop organization	1200 €
TOTAL FOR EVERY YEAR	10096 €

In Brazil, the Brazilian coordinator has been involved the E4S CAPES/Brafitec that has been approved by CAPES in December 2019. We also have an active cooperation project with FAPERGS (the Rio Grande

do Sul research funding agency) for 3-years 2019-2021 using a similar title: “Reproducible Data Science: Analysis Techniques and Workflow Methodologies”. The renewing request of the CAPES/Cofecub project “Group Formation, Analysis, and Visualization in Big Data”, led by Joao Comba, has already been approved in December 2019. We expect to use these fundings for additional visits of students and researchers.

On the french side, the associated team is our main funding solution on the precise research theme of reproducible data science. Nevertheless, our members stay alerted about Brazil-Europe calls that could be used to get extra funding to support our research.

6 References

6.1 Joint publications of the partners

- [1] Guilherme Rezende Alles, João L.D. Comba, Jean-Marc Vincent, Shin Nagai, and Lucas Mello Schnorr. Measuring phenology uncertainty with large scale image processing. *Ecological Informatics*, 59:101109, 2020.
- [2] Lucas Leandro Nesi, Lucas Mello Schnorr, and Arnaud Legrand. Communication-aware load balancing of the lu factorization over heterogeneous clusters. In *2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS)*, Los Alamitos, CA, USA, dec 2020. IEEE Computer Society.
- [3] Behrooz Omidvar-Tehrani, Sihem Amer-Yahia, Eric Simon, Fabian Colque Zegarra, João LD Comba, and Viviane Moreira. Userdev: A mixed-initiative system for user group analytics. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, pages 1–8, 2019.
- [4] Cicero AL Pahins, Behrooz Omidvar-Tehrani, Sihem Amer-Yahia, Valérie Siroux, Jean-Louis Pepin, Jean-Christian Borel, and Joao LD Comba. Coviz: a system for visual formation and exploration of patient cohorts. *Proceedings of the VLDB Endowment*, 12(12):1822–1825, 2019.
- [5] Fabian Colque Zegarra, Juan C Carbajal Ipenza, Behrooz Omidvar-Tehrani, Viviane P Moreira, Sihem Amer-Yahia, and João LD Comba. Visual exploration of rating datasets and user groups. *Future Generation Computer Systems*, 105:547–561, 2020.