Bachelor Thesis Topic
Adaptive parameters for Workflow Engines

Motivation and Background

Workflow engines (e.g. Nextflow[1]) are already a pretty much integral part in scientific computing[2]. Workflow engines take task descriptions in a domain specific language, parse these and then execute the corresponding workflow. These workflows usually are comprised of a set of third-party tools which take several parameters. Finding the correct value for the parameters is a typically fully manual task, that takes a long time. Researchers typically have an idea for feasible parameter ranges/sets which are currently not actively used in an automatic manner. Related work (see [3]) only manages the hyper parameters of execution, not the concrete parameters of the tools.

Goals

The goal of this thesis is defining a data structure to enable parameter adaptations for workflows. The student shall assume that anomalies are detected, so that the only work is on trying to recover from anomalies. The final goal is evaluating the effectiveness of the approach on a set of provided samples.

Description of the Task

The specific tasks are:

-   Creating an adaptable, bounded parameter structure
-   Implementing search strategies upon those parameters
-   Evaluating the results against Nextflow

Research Type

Theoretical Aspects: ★★☆☆☆

Industrial Relevance: ★★★★☆

Implementation   ★★★★☆

Prerequisite

The student should be enrolled in the bachelor of computer science program, and has completed the required course modules to start a bachelor thesis (or similar).

Skills required

Programming skills in Java/Groovy, willingness to learn the required tool, basic knowledge of runtime complexity.

Contacts

Jan Arne Sparka (sparkaar@informatik.hu-berlin.de)

Software Engineering Group, Institut für Informatik, Humboldt-Universität zu Berlin

References

[1] Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). *Nextflow enables reproducible computational workflows*. Nature Biotechnology, 35(4), 316–319. doi:10.1038/nbt.3820

[2] Leser, U., Hilbrich, M., Draxl, C. *et al.* The Collaborative Research Center FONDA. *Datenbank Spektrum* **21,** 255–260 (2021). https://doi.org/10.1007/s13222-021-00397-5

[3] Fernando Chirigati, Vítor Silva, Eduardo Ogasawara, Daniel de Oliveira, Jonas Dias, Fábio Porto, Patrick Valduriez, and Marta Mattoso. 2012. Evaluating parameter sweep workflows in high performance computing. In Proceedings of the 1st ACM SIGMOD Workshop on Scalable Workflow Execution Engines and Technologies (SWEET '12). Association for Computing Machinery, New York, NY, USA, Article 2, 1–10. https://doi.org/10.1145/2443416.2443418