Doctoral thesis (title of the thesis): AUTOMATIC SCHEDULING, EXECUTION AND MONITORING OF COMPUTATIONAL WORKFLOWS ON DISTRIBUTED SYSTEMS

Name of the doctoral student (name and surname): Marta Jaroš

Name and institution of the reviewer (full name of the reviewer, full name and country of the institution):

Jan Martinovič, VSB - Technical University of Ostrava, Czech Republic

---

Please state your opinion on the following aspects of (I) the candidate's doctoral thesis and (II) the candidate's overall achievements, and (III) state your conclusion (a minimum of approx. 300 characters for each item? point below is recommended):

I. Doctoral Thesis

Appropriateness and Relevance

Is the area addressed by the thesis appropriate to the particular scientific discipline of the doctoral thesis and does the thesis address relevant problems within the chosen area?

The orchestration of complex workflows on heterogeneous hardware is an essential topic for supercomputing/data centres nowadays, together with the definition of how to use their infrastructure by the HPC-as-a-Service approaches. For example, several ongoing European projects in high-performance computing and big data management domains tackle these problems. Many new scientific challenges relate to these topics, and this thesis deals with several of them.

A summary of the Contributions of the Thesis

From your point of view, please summarize what the goal of the thesis is, what the main contributions of the thesis are, and whether the thesis has achieved the chosen goal.

Please indicate also specific contributions of the doctoral student.

I see effective workflow execution planning based on multi-objective optimisation as the main goal of this thesis. The main novelty of this thesis is in the proposed optimisation heuristics and fitness functions. Precisely, a workflow execution planning algorithm for k-Dispatch is presented together with its implementation based on the optimisation by genetic algorithms. Several optimisation strategies were proposed by the student together with experiments and discussion. The proposed approach considers handling incomplete performance data and workflow quality evaluation. Additionally, the student proposes the Tetrisator simulator, which can simulate the execution of the workflows coming to the system in a sequential order.

The workflow manager system called k-Dispatch is presented as part of the thesis. This tool was developed with the motivation to have a lightweight platform for experiments. This objective was overcome, and k-Dispatch is a stand-alone, already commercialised tool. A minor shortcoming of the proposed design of the k-Dispatch solution from the security point of view is the presented HPC-as-a-Service deployment

concept in which the SSH credentials are stored in a docker volume located outside the data centre used for the computation. Comparison and discussion related to other HPC-as-a-Service solutions presented in Chapter 2 is missing.

Novelty and Significance:

Please assess the level of novelty of the results and their significance for the given scientific area, for its further development, and if applicable for possible applications in practice.

The proposed approach for the optimisation of moldable ultrasound workflow executions has the potential to be generalised to different use case domains and complex workflow orchestration solutions. The thesis presented the proposed solution's impact: (i) k-Dispatch is already commercialised; (ii) k-Dispatch is integrated into k-Plan modelling tool, which has 35 active users.

Evaluation of the Formal Aspects of the Thesis:

Please evaluate formal qualities of the doctoral thesis and its language level.

The language level of this thesis is of good quality. Nevertheless, I have several comments about the formal aspects of the thesis:

-   Page 6: Direct Acyclic Graph definition seems incomplete – it has to be clearly stated that the graph is without cycles connecting other edges.
-   Page 9: several appropriate software tools are missing: e.g., Galaxy, PyCOMPSs or COMPSs.
-   Page 9: It would be better to have a paragraph about HPC-as-a-Service as a separate subsection.
-   Page 20: software licenses are mentioned, but it is unclear what kind of licenses the author is talking about.
-   Page 21: T_s, T_p, and T are not defined properly. Formula $C = P \times T$ contains undefined T.
-   Figure 4.3: unclear definition of C (page 21) and Computation Cost in the Figure 4.3 where units are core-milliseconds / steps.
-   Page 26: the first sentence is unclear. The word "strategy" is mentioned only on this page and page 35. The first Algorithm is defined in Chapter 5, but this sentence refers to algorithms described in Chapter 4.
-   Algorithm 1 (Presumptions 3): M is not defined.
-   Page 32: The definition of the CriticalPath is missing.
-   Page 36: a_i, b_i are not defined precisely.

The comments mentioned above complicate the readability of the thesis.

Quality of Publications

Has the core of the thesis been published at an appropriate level? Please judge the quantity and quality of the publications. When judging the quality, please take into account internationally recognized standards (WoS/Scopus quartiles, CORE ranks, specific knowledge of flagship publication channels of a given community, etc.) in a way appropriate for the given area of the thesis.

The thesis is based on five core conference publications. Three of them were published as conference papers at the conferences with ranking B (ERA). This is acceptable result, even though I would appreciate

to see already published journal paper presenting the core of the thesis. This is a small weakness related to publication activities of the student.

Student is also author/co-author of additional five conference papers. Other conference and journal papers are under preparation. I would like to highlight that student is a co-author of the poster presented at the international SuperComputing conference.

II. Candidate's Overall Achievements

<u>Overall R&D Activities Evaluation</u>:

Does the student's doctoral thesis, the results included into it, and possible other scientific achievements listed in the list of scientific activities indicate that he/she is a person with scientific erudition and creative abilities?

YES

<u>Assessment of Other Candidate Characteristics</u> (optional):

More characteristics of the doctoral student may be added here in a separate paragraph (e.g., awards, grant participation, international collaboration, etc.).

Based on the received documents, I can state that the student shows competencies to be a valuable researcher during the doctoral study. For example, the student gave several talks at international events, won the PRACE summer of HPC Ambassador Award 2016, spent a year as a Research Assistant at the Department of Medical Physics and Biomedical Engineering at University College London, and participated in several research projects.

III. Conclusion

The conclusion should contain an explicit statement saying whether, in your opinion, the doctoral thesis and the student´s achievements until now meet the generally accepted requirements for the award of an academic degree (in accordance with Section 47 of Act No. 111/1998 Coll., on higher education institution).*

* Short overview of both the Act and corresponding internal BUT regulations is enclosed.

In my opinion, this thesis has met a valid scientific contribution for the award of the PhD degree.

Place Ostrava 30.08.2023